Econometric Theory, **0**, 2026, 1–48.
doi:10.1017/S0266466626100474

SELF-WEIGHTED ESTIMATION FOR LOCAL UNIT ROOT REGRESSIONS WITH APPLICATIONS

ZHISHUI HU
University of Science and Technology of China

NAN LIU
Xiamen University

PETER C. B. PHILLIPS
Yale University, University of Auckland - City Campus, and Singapore Management University

QIYING WANG
The University of Sydney

A new self-weighted (SW) least squares (LS) estimation approach is developed for local unit root (LUR) autoregression with conditionally heteroskedastic innovations. Asymptotically controlled trimming of the most extreme observations in LS regression is shown to lead to mixed Gaussian limit theory and studentization produces standard normality free of the unknown LUR parameter. 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 LUR test in augmented autoregression is developed using SW estimation, providing an alternative approach that employs Gaussian limit theory in contrast to the augmented Dickey–Fuller and other tests. Simulation results show that the test has good size control and power performance against mildly integrated alternatives. An empirical illustration is conducted evaluating the persistence of several market fundamentals as predictors of S&P 500 excess returns.

1. INTRODUCTION

Consider the local unit root (LUR) autoregression

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

The authors thank the Editor, Yixiao Sun, the Co-Editor, and two referees for helpful and extensive comments that have led to considerable improvements in revision. Address correspondence to Qiying Wang, School of Mathematics and Statistics, The University of Sydney, Sydney, Australia, e-mail: qiying.wang@sydney.edu.au

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

with initialization $y_0 = O_P(1)$, innovations $\{u_k\}_{k\geq 1}$, and constant localizing coefficient $\tau \in \mathbb{R}$.[1] 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 (Chan and Wei, 1987; Phillips, 1987a, 1987b, 1988) that

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

where $J_\tau = \{J_\tau(t) = \int_0^t e^{(t-s)\tau} dB_s\}_{t\geq 0}$ is throughout the article 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 (Hansen, 1999; Mikusheva, 2007, 2012; Phillips, 2014, 2023). To assist in resolving this difficulty, many different approaches have been introduced in the literature. The first-difference method in Phillips and Han (2008) 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 Han, Phillips, and Sul (2011), 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, So and Shin (1999) suggested the use of the Cauchy estimator, using the sign function as an IV. Chan, Li, and Peng (2012) generalized that approach by considering weighted estimation in model (1.1). An interesting feature of Chan, Li, and Peng (2012) 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.

Time series generated according to LUR processes of the type (1.1) but with more general serially dependent innovations play a major role in the study of cointegrating and predictive regressions. Using quite a different methodology, Phillips and Magdalinos (2009) introduced the IVX approach which was extended in Kostakis, Magdalinos, and Stamatogiannis (2015) and designed to provide a systematic method of estimation and inference in regressions that involve

[1] In Section 3, extensions of (1.1) that allow for an intercept and relax the $O_P(1)$ condition on the initialization y_0 are discussed—see footnote 2.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

near-unit-root processes. 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 IVs 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 Magdalinos and Petrova (2022), Yang et al. (2020), Demetrescu et al. (2022), Shi and Phillips (2022, 2023), and Liu and Phillips (2023).

The present article proposes a new estimator that has a mixed Gaussian limit by using self-weighted (SW) estimation in (1.1). The approach can be implemented in cointegrating and predictive regressions as well as LUR regressions such as (1.1). Noting that $y_n/\sqrt{n} = O_P(1)$ in (1.1), the key idea is that the function $I(|y_{k-1}| \leq b_n\sqrt{n})$, where $b_n \to 0$, can be used as a softening weight to attenuate the signal in ordinary LS estimation theory so that Wang's extended martingale limit theorem (Wang, 2014; Wang, 2015, Thm. 3.13) applies to establish the asymptotics. The idea is a subtle form of asymptotic trimming that attenuates nonstationarity in the signal sufficiently to remove nuisance parameter effects and produce mixed Gaussian limit behavior. In the context of (1.1), this article explicitly investigates the SW LS estimator $\widehat{\alpha}_{1n}$ of α defined by

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

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 behavior 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. The sequence $b_n \to 0$ softens the nonstationarity in the regressor y_{t-1} sufficiently to reduce the $O_P(n^2)$ signal and the associated $O_P(n)$ superconsistency of LS estimation but retains nonstationarity in the signal. This has the direct benefit of mixed Gaussian limit theory, independence of the localizing coefficient τ and standard normal asymptotics for the associated SW studentized statistic. The method therefore uses asymptotically controlled trimming of nonstationarity in the data rather than elimination of outliers as in traditional statistical methods to eliminate heavy tailed data in the estimation of a mean.

A similar truncation idea was used by Li, Tjostheim, and Gao (2016) in predictive regression with Harris recurrent Markov chains and by Jin and Wang (2024) with strong smoothing conditions. Methods of truncation and self-weighting estimation appear in many other contexts in statistical work, for example, in controlling outliers in regression (Ling, 2007) and testing (Fiorio, Hajivassiliou, and Phillips, 2010). In relation to the present article, a major obstacle in deriving asymptotics for $\widehat{\alpha}_{1n}$ is the development of limit theory for sample moments of the form

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

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

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 Wang and Phillips (2009a) are unsuitable, as explained in Section 2. Another contribution of this article is to establish convergence of V_n^2 under an alternative used in Wang and Phillips (2009a). 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 Wang and Phillips (2009a), the restriction on c_n is stronger, but it is sufficient for the purposes of this article, as explained in Section 2.

The article proceeds as follows. Section 2 explores convergence to local time of a Gaussian process, giving a significant extension of Wang and Phillips (2009a) by allowing for martingales and many other integrated dependent sequences. SW LS estimation theory for near-unit-root autoregression with heteroskedasticity is developed in Section 3. Section 4 develops a new LUR test based on SW estimation and introduces a new cross-validation method that involves iterative training sample testing to select the tuning parameter for data trimming. Simulations and a brief empirical application to several market fundamentals are reported in Sections 5 and 6. Section 7 concludes. Proofs are provided in Appendix A. Section S1 of the Supplementary Material gives a strong approximation result for martingales, which is required in the proof of Theorem 3.1. Some examples satisfying the equation (2.6) below are given in Section S2 of the Supplementary Material. Sections S3 and S4 of the Supplementary Material report additional simulation findings with corresponding tables and figures. Throughout the article, $C, C_1, C_2, \ldots$ are constants that may differ 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 from the rest of the article. The notation differs slightly from other sections to allow for more general applications.

2.1. Preliminaries

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

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

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

for all positive (or bounded) Borel function $h : \mathbb{R} \to \mathbb{R}$. 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, that is, 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 $[0,1] \times \mathbb{R}$ satisfying $\int_{-\infty}^{\infty} \sup_{0\le s\le 1} |g(s,x)| dx < \infty$ 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, for example, Geman and Horowitz (1976, 1980) and also Chapter 2 of Wang (2015).

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

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

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

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

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 Wang and Phillips (2009a), allowing for the array y_{nk} to have a certain structure depending on conditional arguments (see also Chapter 2.3 of Wang, 2015). Theorem 2.1 of Wang and Phillips (2009a) is quite general, so that the main results given by Borodin and Ibragimov (1995), Akonom (1993), Phillips and Park (1998), and Jeganathan (2004) 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 $\mathcal{F}_{nk}=\sigma(y_{n1},\ldots,y_{nk})$. In addition to some regularity conditions on d_{nkj}, Assumption 2.3 of Wang and Phillips (2009a) 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 article.

This section provides an alternative to Assumption 2.3 of Wang and Phillips (2009a). 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 $\{X_{nk}^*\}_{\substack{1\leq k\leq n\\ n\geq 1}}$ such that, for each $n\geq 1$, $\{X_{n1},\ldots,X_{nn}\}=_d\{X_{n1}^*,\ldots,X_{nn}^*\}$, and*

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

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 $\sup_{0\leq s\leq 1}|g_{i\delta}(s,x)|\leq C_\delta/(1+|x|^{1+b})$ for some $b>0$ and*

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

$C_\delta > 0$ 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

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

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 including martingales and partial sums of mixing sequences are given in Sections S1 and S2 of the Supplementary Material. It should be mentioned that Wang and Phillips (2009a) and Theorem 2.19 of Wang (2015) 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 illustrate, let $X_{nk} = \frac{1}{\sqrt{n}}\sum_{j=1}^{k} u_k$, 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

$$\{X_{n1}, \ldots, X_{nn}\} =_d \{B_{1/n}, \ldots, B_{n/n}\} =: \{X^*_{n1}, \ldots, X^*_{nn}\}$$

for each $n \ge 1$. Hence, by noting that $\sqrt{n}\sup_{0\le t\le 1}|B_{\lfloor nt\rfloor/n} - B_t| \to_P \infty$, to ensure that (2.4) is 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 Wang and Phillips (2009a, 2009b) and subsequent research, but it is sufficient for our purpose in this article.

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 $B = \{B_t\}_{t\ge 0}$ so that

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

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 Phillips (2007).

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

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 $\sup_{0\le s\le 1}|g(s,x)|\le C/(1+|x|^{1+b})$ cannot be materially improved. An example similar to Example IV. 2.3 of Borodin and Ibragimov (1995) 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

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

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*

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

(ii) $\sup_{k\ge1}Ev_k<\infty$ *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*

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

(iii) $|g(s,x)-g(s',x)|\le C|s-s'|/(1+|x|^{1+b})$ *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* $c_n/n\to0$, *we have*

$$S_{1n}=A_0S_n+o_P(1). \tag{2.7}$$

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

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

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 Hu, Kasparis, and Wang (2021), where the additional condition (i) is not required. When $c_n=1$, we

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

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 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

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

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

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 $A_0 = \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^n Ew_j$.

3. SELF-WEIGHTED NEAR-UNIT-ROOT ESTIMATION

We consider model (1.1) with innovations $\{u_k\}_{k\geq 1}$ having the following form[2]:

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

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). We assume that σ_k is a positive stationary α-mixing random 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 u_k is similar to that used in Andrews and Guggenberger (2012), allowing for GARCH models of the form defined in (5.1) (e.g., Athreya and Pantula, 1986). Furthermore, we allow for the volatility process σ_k to be nonlinear processes, such as TAR and Bilinear models or, more generally, stationary causal processes defined by $\sigma_k = l(\eta_k, \eta_{k-1}, \dots)$, where $\{\eta_k\}_{k\in\mathbb{Z}}$ is a sequence of i.i.d. random variables and $l(\cdot, \cdot, \dots)$ is a nonnegative real measurable function of its components (e.g., Wu, 2006). While (3.1) is a rather general stationary innovation

[2] In our main results, the model (1.1) may include an intercept δ_n, together with a relaxed condition on the initialization y_0, so that

$$y_k = \delta_n + \alpha y_{k-1} + u_k, \quad \alpha = 1 + \tau/n, \quad k = 1, 2, \dots, n,$$

where $\delta_n = O_P(n^{-r_1})$ and $y_0 = O_P(n^{r_2})$ with $r_1 > 1/2$ and $r_2 < 1/2$. This extension leads to no change in the limit theory and involves standard calculations rather than new methodology. The details are therefore omitted.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

sequence that is appropriate for many practical applications, it does rule out long memory and unit root innovations as considered in Wang, Lin, and Gulati (2003), Xu and Phillips (2008), Cavaliere and Taylor (2009), and Tu, Chan, and Wang (2020). SW near-unit-root estimation in a model such as (1.1) with long memory or unit root innovations would require quite a different asymptotic development and is left for future research.

Define $\widehat{\alpha}_{1n}$ as in (1.3). The asymptotic theory for $\widehat{\alpha}_{1n}$ under model (1.1) with u_k defined by (3.1) 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 $0 < b_n \to 0$ *satisfying* $b_n \log^K n \to \infty$ *for some* $K > 0$, *we have*

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

where $N =_d \mathcal{N}(0,1)$ *is 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*

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

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 rescaled and written as

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

Asymptotics for the LS estimator $\widehat{\alpha}_n$ under stationary conditional heteroskedasticity are widely considered in the past literature, for example, Kim and Schmidt (1993), Ling, Li, and McAleer (2003), and the references therein. As explained in Section 1, these existing results typically rely on the unknown localizing coefficient τ, which is identified but not consistently estimable. The new estimator $\widehat{\alpha}_{1n}$ has a distinctive pivotal standard normal limit which is convenient for inference and an advantage in applications. It should be mentioned that such a pivotal normal limit depends on the mutually uncorrelated error innovations u_k defined as in (3.1). As with the standard LS estimator $\widehat{\alpha}_n$, a bias term will appear if a correlated innovation process such as $w_k = \sum_{j=0}^{\infty}\psi_j u_{k-j}$ is used instead of u_k. Dealing with such correlated innovation processes is possible given that the parameter τ in the LUR model (1.1) is a constant localizing coefficient. But a careful design involving

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

some complicated calculations is required. Detailed analysis and discussion of that approach is left for future work.

Remark 3.2. In comparison with the usual LS estimator $\widehat{\alpha}_n$, the SW 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 behavior 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 as slow as possible to ensure that the SW 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 $b_n \log^K n \to \infty$ with some $K > 0$ is imposed[3] for the purpose that $b_n \to 0$ can be chosen with some flexibility to assist in particular applications. 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, finite sample simulations show that a balanced C_0 is required so that $b_n = C_0 \log^{-1} n$ is small enough (i.e., closer to 0) for the asymptotic normality condition in (3.4) to be satisfied, while at the same time being large enough to avoid excessive softening of the observations and to help reduce variance. This balance helps to achieve satisfactory finite sample performance.[4] It is explored in simulations of Section 5, where a cross-validation method of selecting the tuning parameter C_0 is introduced for practical applications.

Remark 3.3. Other weight functions $W(x)$ may be used in (1.3) instead of $I(|x| \le 1)$, giving an SW LS estimator $\widetilde{\alpha}_{1n}$ of α defined by

$$\widetilde{\alpha}_{1n} = \arg\min_{\alpha} \sum_{k=1}^{n} (y_k - \alpha y_{k-1})^2 W\left(\frac{y_{k-1}}{\sqrt{n}b_n}\right) = \frac{\sum_{k=1}^{n} y_k y_{k-1} W(y_{k-1}/\sqrt{n}b_n)}{\sum_{k=1}^{n} y_{k-1}^2 W(y_{k-1}/\sqrt{n}b_n)}.$$

The asymptotic theory for $\widetilde{\alpha}_{1n}$ matches that of Theorem 3.1 and is given in the next result, whose proof involves only simple modifications in that of Theorem 3.1 and is omitted.

THEOREM 3.2. *Suppose that for some* $\delta > 0$,

(a) $\sup_{k \ge 1} E(|\epsilon_k|^{2+\delta} | \mathcal{F}_{k-1}) \le C < \infty$ *and* $E\sigma_1^{2+\delta} < \infty$;
(b) *the mixing coefficients* $\alpha(n)$ *of* $\{\sigma_k\}_{k \ge 1}$ *satisfy* $\alpha(n) \le Cn^{-(1+\delta)}$;

[3] The requirement $b_n \log^K n \to \infty$ could be replaced by $b_n n^\delta \to \infty$ for some sufficiently small $\delta > 0$ to allow for the sequence $0 < b_n \to 0$ to be chosen within a wider interval. But this simplification is unnecessary because a slower rate $b_n \to 0$ provides an improved consistency rate in the SW estimator $\widehat{\alpha}_n$, as indicated in (3.3).

[4] When $b_n = C_0 \log^{-1} n$, it follows from (3.2) that

$$n(\log n)^{-3/2}\left(\widehat{\alpha}_{1n} - \alpha\right) \to_D \sigma^{3/2} C_0^{-3/2} \left[\frac{2}{3} L_{J_\tau}(1,0)\right]^{-1/2} N.$$

Increasing C_0 therefore lowers the conditional asymptotic variance of $\widehat{\alpha}$.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

(c) *$W(x)$ is a positive continuous function satisfying $W(x) \le C/(1+|x|^{1+b})$ and $|W(x)-W(y)| \le C|x-y|/(1+|x|^{1+b}+|y|^{1+b})$, for some $b>0$ so that $\int_{-\infty}^{\infty} x^2 W(x)dx < \infty$.*

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

$$n b_n^{3/2}\left(\widetilde{\alpha}_{1n}-\alpha\right) \to_D \sigma^{3/2}\Big[\int_{-\infty}^{\infty} x^2 W^2(x)dx\Big]^{1/2}\Big[\int_{-\infty}^{\infty} x^2 W(x)dx\Big]^{-1} L_{J_\tau}^{-1/2}(1,0)\,N, \tag{3.5}$$

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

$$\Big[\sum_{k=1}^{n} y_{k-1}^2 W(y_{k-1}/\sqrt{n}b_n)\Big]\Big[\sum_{k=1}^{n} y_{k-1}^2 W^2(y_{k-1}/\sqrt{n}b_n)\Big]^{-1/2}\left(\widetilde{\alpha}_{1n}-\alpha\right) \to_D \mathcal{N}(0,\sigma^2). \tag{3.6}$$

We further have

$$\hat{\sigma}_{1n}^{-1}\Big[\sum_{k=1}^{n} y_{k-1}^2 W(y_{k-1}/\sqrt{n}b_n)\Big]\Big[\sum_{k=1}^{n} y_{k-1}^2 W^2(y_{k-1}/\sqrt{n}b_n)\Big]^{-1/2}\left(\widetilde{\alpha}_{1n}-\alpha\right) \to_D \mathcal{N}(0,1), \tag{3.7}$$

where $\hat{\sigma}_{1n}^2$ is a natural consistent estimator of σ^2 as in Remark 3.1.

Remark 3.4. Condition (c) in Theorem 3.2 excludes the weight function $I(|x| \le 1)$ used in Theorem 3.1, indicating that (3.5) may hold for $W \in \Gamma$, where Γ is a wider function class in comparison with that required in condition (c). It may therefore be of interest to find an optimal function class Γ as well as an optimal $W_{opt} \in \Gamma$ for which (3.5) holds. This topic is beyond the scope of the present article and is left for future work.

4. SELF-WEIGHTED LOCAL UNIT ROOT TESTING

Consider model (1.1) with the innovations $\{u_k\}_{k\ge 1}$ given by (3.1). Theorem 3.1 implies that, upon normalization, $\widehat{\alpha}_{1n}$ can serve as a test statistic for the null and alternative hypotheses

$$\mathcal{H}_0 : \alpha = 1 + \frac{\tau}{n} \text{ for some unknown } \tau \in \mathcal{R} \quad vs. \quad \mathcal{H}_1 : \alpha = 1 - \frac{\tau_0}{k_n} \text{ with } \tau_0 > 0, \tag{4.1}$$

where k_n is a sequence of positive constants for which $k_n/n \to 0$. Note that, under alternative $\mathcal{H}_1$, the y_k generated by model (1.1) allows for a mildly integrated

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

process having near-stationary moderate deviations from a unit root of the type introduced in Phillips and Magdalinos (2007). Such a test seeks to confirm the presence of a possible unit root or local to unit root formulation as the null against the possibility of mildly integrated or near-stationary alternatives. This approach allows for a wider class of possible null hypotheses that are in line with practical experience in applications with nonstationary data that autoregressive roots may be close to unity in the LUR sense rather than strictly at unity. This formulation of the null hypothesis in nonstationarity testing seems new to the extensive unit root literature. We have the following result.

THEOREM 4.1. *Suppose that*

(a) *model (1.1) 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* $\alpha(n) \leq Cn^{-(1+\delta)}$.

For any $0 < b_n \to 0$ *satisfying* $b_n \log^K n \to \infty$ *with some* $K > 0$*, under the null hypothesis* $\mathcal{H}_0$*, we have*

$$T_n := \hat{\sigma}_{1n}^{-1} \left[\sum_{k=1}^{n} y_{k-1}^2 I(|y_{k-1}| \leq b_n \sqrt{n}) \right]^{1/2} \left(\widehat{\alpha}_{1n} - 1 \right) \to_D \mathcal{N}(0,1), \tag{4.2}$$

where $\hat{\sigma}_{1n}^2 = \frac{1}{n} \sum_{k=1}^{n} \left(y_k - \widehat{\alpha}_n y_{k-1} \right)^2$ *with* $\widehat{\alpha}_n$ *being the LS estimator of* α *in model (1.1). Furthermore, if in addition* $nb_n^2/k_n \to \infty$*, then, under* $\mathcal{H}_1$*,*

$$\lim_{n\to\infty} P\left(|T_n| \geq t_\delta \right) = 1, \tag{4.3}$$

for any $0 < \delta < 1$*, where* $\Phi(t_\delta) = 1 - \delta/2$ *and* $\Phi(.)$ *is the standard normal cdf.*

Remark 4.1. Result (4.3) indicates that our test of an LUR has non-trivial power even in the rather weak conditions that $k_n/n \to 0$ and $nb_n^2/k_n \to \infty$.[5] Recall that $b_n \to 0$ and $b_n \log^K n \to \infty$ for some $K > 0$. Such conditions are clearly satisfied for $k_n = n^\gamma$ for any $0 < \gamma < 1$ so that the rejection of the null allows for the mildly integrated or near-stationary alternatives. However, as in conventional methods, the alternatives in (4.1) are wider than mild integration or near stationarity. Hence, rejection of the null does not necessarily imply the acceptance of mildly integrated or near-stationary alternatives. To resolve such an issue, one referee suggested the construction of another test T_n' that employs the mildly integrated model as the null and the pure stationary model as the alternative (i.e., $k_n \to \infty$ satisfying $nb_n^2/k_n \to \infty$ against fixed $k_n > 0$). If T_n is rejected but T_n' is not, the rejection of

[5]The actual convergence rate of the proposed estimator $\widehat{\alpha}_{1n}$ is $(nb_n^{2/3})^{-1}$ under $\mathcal{H}_0$ by virtue of (3.2). Hence, the optimal power condition in terms of the magnitude of the local departure under $\mathcal{H}_1$ seems to be $nb_n^{2/3}/k_n \to \infty$ rather than $nb_n^2/k_n \to \infty$, as pointed out by the Editor, Yixiao Sun. Note that the LUR process under $\mathcal{H}_0$ has quite different properties from that of the local stationary process under $\mathcal{H}_1$. To achieve an optimal power, condition is therefore challenging and new techniques may be required to establish sharp asymptotics for the proposed estimator $\widehat{\alpha}_{1n}$ under $\mathcal{H}_1$. This analysis is beyond the scope of the present article and left for future work.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

T_n may be attributed to mild integration. If both are rejected, then evidence favors a fixed alternative. Such a test is interesting but beyond the scope of this article. It is therefore left for future work.

If we are only interested in standard unit root testing, namely the unit root null

$$\mathcal{H}_0 : \alpha = 1 \quad vs. \quad \mathcal{H}_1 : \alpha < 1,$$

model (1.1) with innovations (3.1) can be extended to ADF form as

$$y_k = \alpha y_{k-1} + \sum_{j=1}^{p-1} \theta_j (y_{k-j} - y_{k-j-1}) + u_k, \tag{4.4}$$

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

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

$$\sum_{k=1}^{n} \Big[y_k - \alpha y_{k-1} - \sum_{j=1}^{p-1} \theta_j \big(y_{k-j} - y_{k-j-1} \big) \Big] \big(y_{k-l} - y_{k-l-1} \big) = 0, \quad l = 1, \ldots, p-1.$$

Let $z_k = \big(y_{k-1} - y_{k-2}, \ldots, y_{k-p+1} - y_{k-p} \big)$ 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

$$\begin{pmatrix} \widehat{\alpha}_{2n} \\ \widehat{\theta}_n \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{n} y_{k-1} \widetilde{y}_{k-1} & \sum_{k=1}^{n} z_k \widetilde{y}_{k-1} \\ \sum_{k=1}^{n} z_k' y_{k-1} & \sum_{k=1}^{n} z_k' z_k \end{pmatrix}^{-1} \begin{pmatrix} \sum_{k=1}^{n} y_k \widetilde{y}_{k-1} \\ \sum_{k=1}^{n} z_k' y_k \end{pmatrix}. \tag{4.5}$$

Routine derivations 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 unit root testing in the augmented Dickey–Fuller (ADF) model (4.4).

THEOREM 4.2. *Suppose that*

(a) *model (4.4) 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 $\alpha(n) \leq Cn^{-(1+\delta)}$.*

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

$$\Big[\sum_{k=1}^{n} y_{k-1}^2 I(|y_{k-1}| \leq b_n \sqrt{n}) \Big]^{1/2} \big(\widehat{\alpha}_{2n} - \alpha \big) \to_D \mathcal{N}(0, \sigma^2), \tag{4.6}$$

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

and

$$\hat{\sigma}_{2n}^{-1}\left[\sum_{k=1}^{n} y_{k-1}^2 I(|y_{k-1}| \le b_n\sqrt{n})\right]^{1/2}\left(\widehat{\alpha}_{2n}-\alpha\right) \to_D \mathcal{N}(0,1), \tag{4.7}$$

where $\hat{\sigma}_{2n}^2 = \frac{1}{n}\sum_{k=1}^{n}\left(y_k - \widehat{\alpha}_{2n} y_{k-1} - z_k\widehat{\theta}_n\right)^2$.[6]

Remark 4.2. Since the original work of Dickey and Fuller (1979) 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 Phillips and Xiao (1998) and Herranz (2017). In recent years, the 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 So and Shin (1999), Phillips, Park, and Chang (2004), Magdalinos and Petrova (2022), 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 Han, Phillips, and Sul (2011).

5. MONTE CARLO SIMULATIONS

Simulations were conducted to explore the finite sample properties of the SW estimator in LUR testing. In implementing the truncation approach, we set $b_n = C_0 \log^{-1} n$, where the tuning parameter C_0 is selected via a new method of cross-validation. The procedure for implementing this data-driven choice is discussed below in Section 5.1.

We investigate the finite sample bias of the SW estimator $\widehat{\alpha}_{1n}$ and compare this bias with that of the conventional OLS estimator in Section 5.2. We also evaluate the empirical coverage probabilities of confidence intervals constructed based on the SW estimator in Section 5.3. The empirical size and local power properties of the SW test are analyzed in Sections 5.4 and 5.5, respectively.

Our numerical simulations follow the model (1.1) with initial value $y_0 = 0$. For each $k = 1, \ldots, n$, the innovation u_k admits a GARCH (1,1) representation[7]

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

[6] Simple calculations show that $\hat{\sigma}_{2n}^2$ is consistent and a natural estimator of $\sigma^2 = Eu_1^2$ under model (4.4). We may also estimate σ^2 by using $\widetilde{\sigma}_{2n}^2 = \frac{1}{n}\sum_{k=1}^{n}\left(y_k - \widetilde{\alpha}_n y_{k-1} - z_k\widetilde{\theta}_n\right)^2$, where $(\widetilde{\alpha}, \widetilde{\theta}_n')$ is the LS estimator of (α, θ') in model (4.4).

[7] Additional simulations were conducted for models with MA(1) errors and these are reported in Sections S3 and S4 of the Supplementary Material to which interested readers can refer.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

where $\varphi_0 = 1, \varphi_1 = 0.2$, and $\varphi_2 = 0.3$, and $\epsilon_k \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma_\epsilon^2\right)$ with $\sigma_\epsilon^2 \in \{0.25, 0.5, 1\}$ are considered in the simulation study. We test any fixed simple null against a wide class of mildly integrated alternatives:

$$\mathcal{H}_0 : \alpha = 1 + \frac{\tau}{n} \text{ for some } \tau \in \mathcal{R} \text{ vs. } \mathcal{H}_1 : \alpha = 1 - \frac{\tau_0}{n^\gamma} \text{ with } \tau_0 > 0 \text{ and } \gamma \in (0, 1). \tag{5.2}$$

The SW test statistic T_n is constructed as in (4.2). We first detail the cross-validation procedure used to select the tuning parameter C_0, and then examine the empirical size and local power of the SW test, comparing its performance with the conventional ADF unit root test.

5.1. Selection of the Tuning Parameter C_0

To determine the tuning parameter C_0 in the truncation rule $b_n = C_0 \log^{-1} n$, we employ a cross-validation procedure over a discrete grid within the interval $[\underline{C}_0, \overline{C}_0]$. To account for the temporal dependence in time-series data, we implement a rolling cross-validation scheme that preserves the ordering of the observations.

Given the time series $\{y_k\}_{k=1}^n$, we partition the data into K sequential folds. The initial fold is used for model training, and the immediate next fold is used for out-of-sample forecasting. Subsequently, the testing fold is added to the training set, and the procedure continues until the final fold is used. Let $[j]$ denote the index set of the jth fold, for $j = 1, \ldots, K$. The iterative training–testing proceeds as follows:

- Training: [1]; Testing: [2];
- Training: [1], [2]; Testing: [3];
- $\cdots$
- Training: [1], [2], $\ldots$, $[K-1]$; Testing: $[K]$.

In each iteration, we estimate the model by regressing y_k on y_{k-1} on the training data, applying a truncation to the lagged regressor based on the threshold $b_n = C_0 \log^{-1} n$, where C_0 varies across grid points in $[\underline{C}_0, \overline{C}_0]$. Using this estimate, out-of-sample predictions are produced for the testing data, and the corresponding mean squared prediction error (MSPE) is calculated. The process is repeated across folds, and the average MSPE (AMSPE) is computed for each candidate C_0. Specifically,

$$\text{AMSPE}(C_0) = \frac{1}{K-1} \sum_{j=2}^{K} \text{MSPE}_j(C_0) \text{ with } \text{MSPE}_j(C_0) = \frac{1}{n_j} \sum_{k \in [j]} \left(y_k - \widehat{y}_k^{(j)}(C_0) \right)^2,$$

where $\widehat{y}_k^{(j)}(C_0)$ denotes the predicted value of y_k using the model trained on the first $(j-1)$th fold with truncation level determined by C_0, and n_j is the number of observations in the jth testing fold. The optimal tuning parameter $\widehat{C}_0$ is then selected as the value of C_0 that minimizes AMSPE over the grid $[\underline{C}_0, \overline{C}_0]$,

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

$$\widehat{C}_0 = \arg \min_{C_0 \in [\underline{C}_0, \overline{C}_0]} \text{AMSPE}(C_0). \tag{5.3}$$

To implement this selection in our simulation study, we consider a sample size of $n = 200$ and perform 10,000 Monte Carlo replications for each data-generating process (DGP). Figures 1–3 present histograms of the selected $\widehat{C}_0$ across different settings, with normally distributed innovations ϵ_k having variances $\sigma_\epsilon^2 \in \{0.25, 0.5, 1\}$ and localizing parameters $\tau \in \{-0.1, -0.5, -1, 0.1, 0.5, 1\}$. The search for the optimal tuning parameter is conducted over the interval $[0.2\hat{\sigma}_{1n}, 10\hat{\sigma}_{1n}]$, where $\hat{\sigma}_{1n}$ denotes an estimate of the standard deviation of u_k, ensuring that the truncation levels considered are appropriate across the simulation designs.

In addition to the histograms of the selected tuning parameter, Figure 4 plots the cross-validated predicted mean squared error (PMSE) as a function of C_0 for a sequence of increasing sample sizes $n \in \{50, 100, 200, 500, 1000, 2000\}$, using three-fold cross-validation. Across all designs, the PMSE curves exhibit a well-defined and stable minimum, indicating that the prediction-based criterion yields a practical, identifiable choice of the tuning parameter. Moreover, as the sample size increases, the PMSE curves become increasingly concentrated around a common minimizer, providing empirical evidence that the selected tuning parameter stabilizes and converges as n grows. These results provide supporting finite-sample evidence for the proposed tuning procedure.

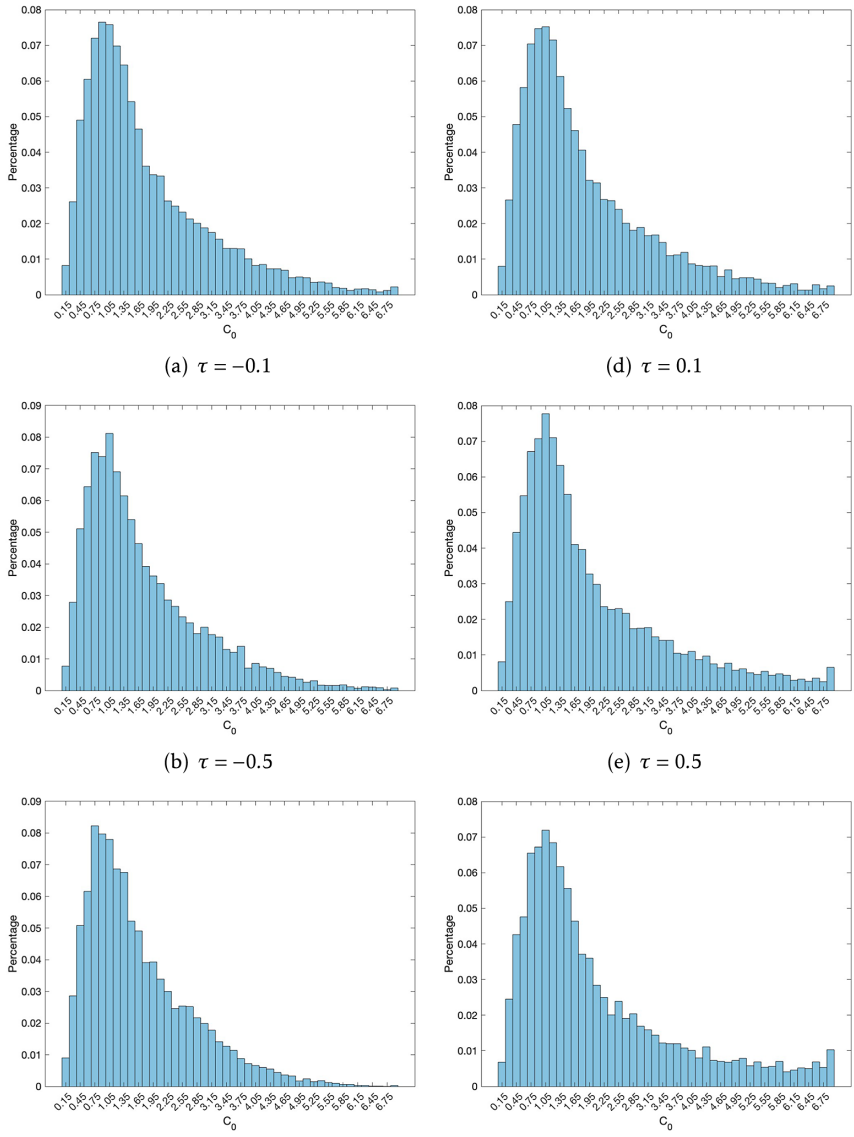
5.2. Finite Sample Comparison of SW and OLS Estimators

To illustrate the finite sample behavior of the OLS and SW estimators, we conduct simulations using the LUR model (1.1) with coefficient $\tau \in \{-1, -0.5, -0.1, 0, 0.1, 0.5, 1\}$, and innovation process following a GARCH(1,1) specification with variance $\sigma_\varepsilon^2 = 1$. The sample sizes are $n \in \{50, 100, 200, 500\}$, and each design is replicated 10,000 times. For each replication, we compute the OLS and SW estimators and evaluate their finite sample performance using bias and mean absolute error (MAE).

The OLS estimator is well-known to exhibit substantial finite sample bias in LUR settings. For the unit root case ($\tau = 0$), the bias of the OLS estimator is approximately $-1.72/n$, as shown in the literature (Phillips and Perron, 1988). In contrast, the SW estimator employs trimming of extreme values, which reduces the bias while retaining a Gaussian limit.

Table 1 reports the bias and MAE of the two estimators. The OLS estimator displays negative finite sample bias that diminishes slowly with sample size. The bias is significant for small n and in the unit-root case ($\tau = 0$). The SW estimator exhibits smaller finite sample bias, while its MAE is slightly larger than that of the OLS estimator, reflecting the effect of trimming. The MAE differences decrease with the sample size and are limited in magnitude, suggesting that trimming does not have a substantial finite sample cost.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press



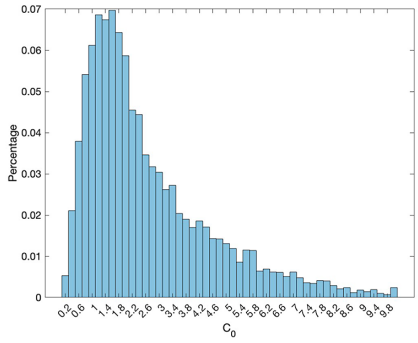
(a) $\tau = -0.1$ (d) $\tau = 0.1$

(b) $\tau = -0.5$ (e) $\tau = 0.5$

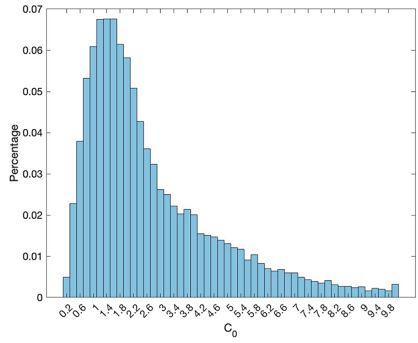
(c) $\tau = -1$ (f) $\tau = 1$

FIGURE 1. Histogram of choice of $\widehat{C}_0$ using cross-validation when $\epsilon_k \sim N(0, 0.25)$.
Note: The sample size $n = 200$. The number of replications is $10,000$.

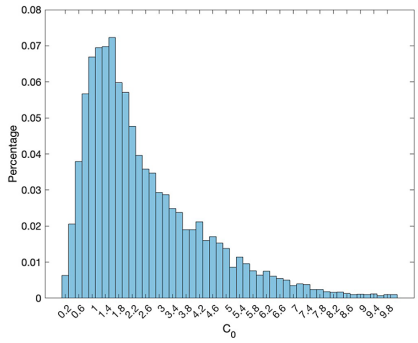
https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press



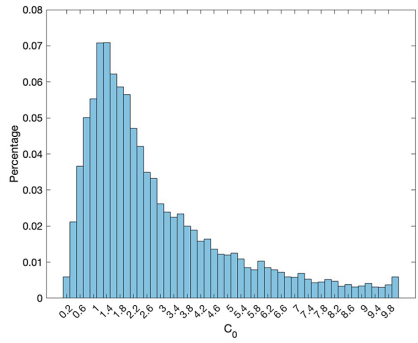
(a) $\tau = -0.1$



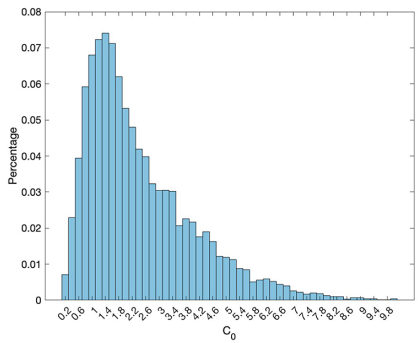
(d) $\tau = 0.1$



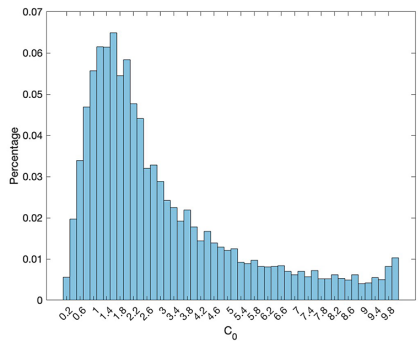
(b) $\tau = -0.5$



(e) $\tau = 0.5$



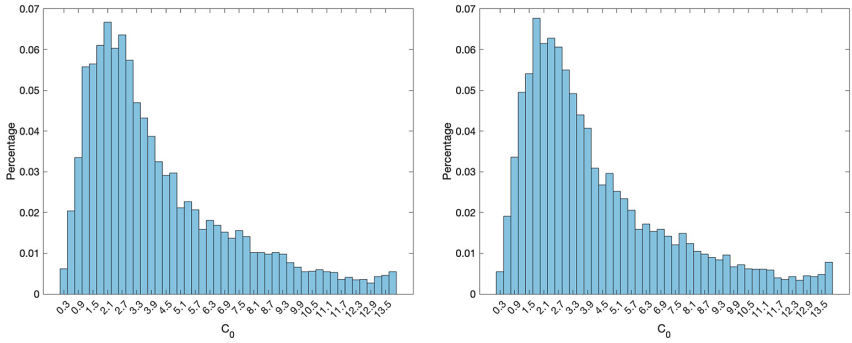
(c) $\tau = -1$



(f) $\tau = 1$

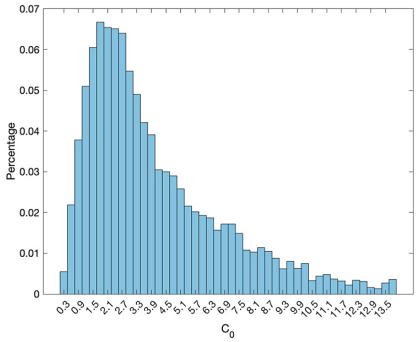
FIGURE 2. Histogram of choice of $\widehat{C}_0$ using cross-validation when $\epsilon_k \sim N(0, 0.5)$.
Note: The sample size $n = 200$. The number of replications is 10,000.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

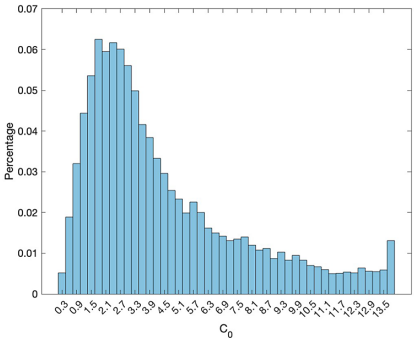


(a) $\tau = -0.1$

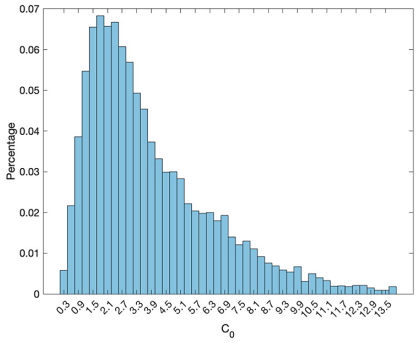
(d) $\tau = 0.1$



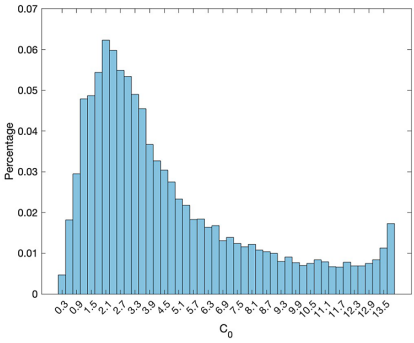
(b) $\tau = -0.5$



(e) $\tau = 0.5$



(c) $\tau = -1$



(f) $\tau = 1$

FIGURE 3. Histogram of choice of $\widehat{C}_0$ using cross-validation when $\epsilon_k \sim N(0,1)$.
Note: The sample size $n = 200$. The number of replications is 10,000.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

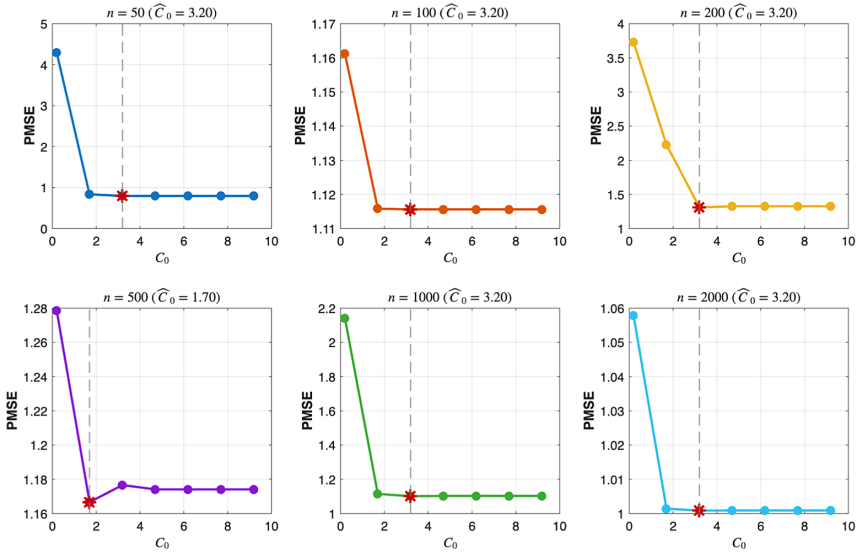


FIGURE 4. Cross-validation: PMSE versus C_0 for different sample sizes.
Note: The number of replications is 10,000. The figures show the three-fold cross-validated predicted mean squared error (PMSE) as a function of the tuning parameter C_0.

Figure 5 shows kernel density estimates of the sampling distributions of the OLS and SW t-statistics for the unit-root case ($\tau = 0$) with sample sizes $n \in \{100, 200, 500\}$. The OLS distribution is shifted away from the true parameter. The SW distribution is more centered and approximately symmetric, closely approximating the standard normal distribution, consistent with the theoretical results.

5.3. Empirical Coverage Probabilities of SW Estimator

Finite sample inference performance of the SW test is studied by examining the empirical coverage probabilities of confidence intervals for $\widehat{\alpha}_{1n}$ for different sample sizes and localizing parameters.

The simulations are based on the LUR model (1.1), where the localizing coefficient $\tau \in \{-1, -0.5, -0.2, -0.1, 0.1, 0.2, 0.5, 1\}$, and the innovation follows a GARCH(1,1) specification. The innovation variance $\sigma_\varepsilon^2 \in \{0.25, 0.5, 1\}$ determines the signal-to-noise ratio. We consider sample sizes $n \in \{50, 100, 200, 500\}$, with $S = 10,000$ Monte Carlo replications for each design.

For each replication, α is estimated using the proposed SW estimator with bandwidth $b_n = C_0 \log^{-1} n$, where C_0 is chosen by the cross-validation procedure described in Section 5.1. The resulting estimator $\widehat{\alpha}_{1n}$ is defined in (1.3), and its standard error $\widehat{\sigma}_{1n}$ is computed as in Remark 3.1. According to Theorem 3.1, a $(1-\eta)$-level confidence interval is constructed as $\mathrm{CI}_{1-\eta} = \left[\widehat{\alpha}_{1n} \pm z_{1-\eta/2}\widehat{\sigma}_{1n}\right]$,

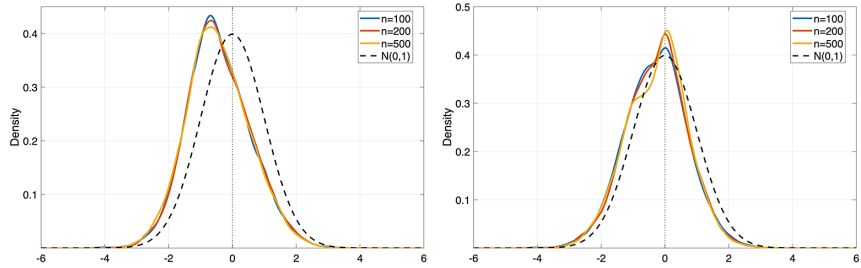
https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

Table 1. Bias and MAE of OLS and SW estimators.

T	$\tau=-1$		$\tau=-0.5$		$\tau=-0.1$		$\tau=0$		$\tau=0.1$		$\tau=0.5$		$\tau=1$	
	OLS	SW	OLS	SW	OLS	SW	OLS	SW	OLS	SW	OLS	SW	OLS	SW
Bias														
50	−0.0365	−0.0311	−0.0360	−0.0306	−0.0354	−0.0304	−0.0352	−0.0300	−0.0350	−0.0296	−0.0339	−0.0271	−0.0320	−0.0254
100	−0.0184	−0.0167	−0.0182	−0.0154	−0.0178	−0.0159	−0.0177	−0.0156	−0.0176	−0.0145	−0.0170	−0.0141	−0.0160	−0.0119
200	−0.0092	−0.0074	−0.0090	−0.0071	−0.0088	−0.0071	−0.0087	−0.0069	−0.0087	−0.0069	−0.0084	−0.0058	−0.0079	−0.0054
500	−0.0039	−0.0033	−0.0039	−0.0030	−0.0038	−0.0030	−0.0037	−0.0027	−0.0037	−0.0028	−0.0035	−0.0025	−0.0033	−0.0021
MAE														
50	0.0492	0.0656	0.0474	0.0626	0.0459	0.0603	0.0455	0.0604	0.0450	0.0600	0.0431	0.0572	0.0402	0.0528
100	0.0246	0.0336	0.0236	0.0329	0.0228	0.0319	0.0226	0.0322	0.0223	0.0315	0.0214	0.0288	0.0199	0.0267
200	0.0123	0.0170	0.0118	0.0166	0.0114	0.0156	0.0113	0.0150	0.0111	0.0151	0.0106	0.0145	0.0099	0.0138
500	0.0051	0.0073	0.0049	0.0069	0.0047	0.0067	0.0047	0.0066	0.0046	0.0064	0.0044	0.0065	0.0040	0.0059

Note: Entries report Monte Carlo bias and mean absolute error (MAE) for OLS and SW estimators. Number of Monte Carlo replications: $S = 10,000$. Bandwidth parameter: $b_n = \widehat{C}_0 \log^{-1} n$, where $\widehat{C}_0$ is selected by cross-validation.



(a) OLS t-statistic (b) SW t-statistic

FIGURE 5. Density estimates of the OLS and SW t-statistics for $\tau = 0$ and sample sizes $n = 100, 200, 500$.

Note: The number of replications is 10,000. For the SW estimator, we apply the truncation rate $b_n = \widehat{C}_0 \log^{-1} n$ with cross-validated $\widehat{C}_0$.

where $z_{1-\eta/2}$ denotes the $(1-\eta/2)$ quantile of the standard normal distribution. The empirical coverage probability is computed as

$$\text{Coverage Rate} = \frac{1}{S}\sum_{s=1}^{S} \mathbb{I}\left\{\alpha \in \text{CI}_{1-\eta}^{(s)}\right\}, \tag{5.4}$$

where $\text{CI}_{1-\eta}^{(s)}$ denotes the confidence interval in the s-th replication. We set $\eta = 0.05$, so that the nominal coverage level is 95%.

Table 2 reports the empirical coverage probabilities of the 95% confidence intervals for $\widehat{\alpha}_{1n}$ across different combinations of n, τ, and σ_ϵ^2. The empirical coverage rates are consistently close to the nominal level across all cases, indicating that the large-sample approximation developed in Theorem 3.1 captures the finite sample distribution of $\widehat{\alpha}_{1n}$ quite well.

5.4. Empirical Size for the SW Local Unit Root Test

The following Monte Carlo simulation is designed to evaluate the finite sample performance of the SW test relative to the ADF test. The DGP follows a near-unit-root AR(1) model, which allows for testing the local null hypothesis $\mathcal{H}_0 : \alpha = 1 + \frac{\tau}{n}$. The SW test statistic T_n is constructed according to (4.2) in Theorem 4.1. When testing the local null hypothesis $\mathcal{H}_0$, which is a moving target, the asymptotic distribution of T_n is derived under $\alpha = 1$. Consequently, the asymptotic normality of T_n is established in a manner analogous to the classical ADF framework. This construction ensures that the Monte Carlo comparison of the empirical sizes of the SW and ADF tests is both conceptually consistent and theoretically justified.

Specifically, size performance of the SW and ADF tests is evaluated for a nominal size of 5%, sample sizes $n \in \{50, 100, 200, 500\}$, innovation variance

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

Table 2. Empirical coverage probabilities of the 95% confidence intervals for $\widehat{\alpha}_{1n}$.

T	σ_ε^2	τ					
		-1	-0.5	-0.1	0.1	0.5	1
50	0.25	0.9643	0.9657	0.9661	0.9671	0.9686	0.9680
	0.50	0.9647	0.9660	0.9666	0.9663	0.9700	0.9711
	1.00	0.9644	0.9650	0.9655	0.9654	0.9694	0.9683
100	0.25	0.9686	0.9691	0.9690	0.9703	0.9723	0.9720
	0.50	0.9698	0.9695	0.9696	0.9722	0.9719	0.9731
	1.00	0.9690	0.9701	0.9711	0.9720	0.9742	0.9727
200	0.25	0.9711	0.9722	0.9736	0.9732	0.9752	0.9730
	0.50	0.9727	0.9734	0.9751	0.9746	0.9759	0.9744
	1.00	0.9740	0.9756	0.9771	0.9764	0.9745	0.9759
500	0.25	0.9712	0.9716	0.9727	0.9738	0.9733	0.9751
	0.50	0.9709	0.9718	0.9744	0.9734	0.9762	0.9739
	1.00	0.9736	0.9743	0.9721	0.9715	0.9724	0.9750

Note: Nominal coverage level is 95%. Number of Monte Carlo replications: $S = 10,000$. Bandwidth parameter: $b_n = \widehat{C}_0 \log^{-1} n$, where $\widehat{C}_0$ is selected by cross-validation.

$\sigma_\epsilon^2 \in \{0.25, 0.5, 1\}$, and localizing coefficient under the null $\tau \in \{-0.1, -0.5, -1, 0.1, 0.5, 1\}$, with 10,000 replications for each DGP. The simulation design included positive localizing coefficients $\tau > 0$ to assess properties of the procedure under the null in the explosive LUR region. The SW test uses the data-driven truncation parameter $\widehat{C}_0$ selected by the cross-validation method described in Section 5.1. For the ADF test, no lagged terms are included due to the mds design in the simulation. The results are shown in Table 3.

The results show that the SW test with cross-validated $\widehat{C}_0$ achieves robust size control across different combinations of $n \in \{50, 100, 200, 500\}$ and $\sigma_\epsilon^2 \in \{0.25, 0.5, 1\}$, provided that the localizing coefficient τ is not too negative. In particular, the empirical rejection rate of the SW test is close to the nominal 5% level across all settings, except when $\tau = -1$, where the test exhibits a mild degree of size distortion. Notably, size becomes more conservative under the null in the explosive LUR region $\tau > 0$.

The ADF test also exhibits size distortion when τ becomes more negative (e.g., $\tau = -0.5$ or -1), particularly for smaller sample sizes. In comparison, the SW test shows better size control overall and particularly when τ deviates further from zero in the negative direction, which shows some advantage of SW testing with cross-validation selection of $\widehat{C}_0$ in finite samples. Like the SW test, the ADF test is size conservative in the explosive LUR region, but notably not as conservative as the SW test.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

TABLE 3. Empirical size for the local unit root tests when $\epsilon_k \sim \mathcal{N}(0, \sigma_\epsilon^2)$.

n	σ_ϵ^2	$\tau = -1$		$\tau = -0.5$		$\tau = -0.1$		$\tau = 0.1$		$\tau = 0.5$		$\tau = 1$	
		SW	ADF	SW	ADF	SW	ADF	SW	ADF	SW	ADF	SW	ADF
50	0.25	0.0798	0.0826	0.0653	0.0675	0.0562	0.0552	0.0516	0.0497	0.0428	0.0421	0.0338	0.0330
	0.5	0.0794	0.0860	0.0645	0.0693	0.0530	0.0573	0.0504	0.0521	0.0427	0.0423	0.0320	0.0331
	1	0.0789	0.0926	0.0625	0.0762	0.0546	0.0642	0.0486	0.0575	0.0400	0.0465	0.0326	0.0371
100	0.25	0.0788	0.0789	0.0620	0.0640	0.0526	0.0528	0.0470	0.0477	0.0399	0.0405	0.0296	0.0315
	0.5	0.0773	0.0809	0.0606	0.0637	0.0507	0.0529	0.0464	0.0492	0.0378	0.0403	0.0290	0.0313
	1	0.0733	0.0870	0.0572	0.0693	0.0496	0.0575	0.0444	0.0519	0.0361	0.0424	0.0298	0.0330
200	0.25	0.0697	0.0758	0.0561	0.0615	0.0477	0.0514	0.0425	0.0457	0.0370	0.0400	0.0300	0.0329
	0.5	0.0722	0.0780	0.0543	0.0610	0.0433	0.0510	0.0418	0.0474	0.0339	0.0399	0.0287	0.0309
	1	0.0663	0.0815	0.0543	0.0650	0.0445	0.0530	0.0400	0.0482	0.0339	0.0410	0.0264	0.0317
500	0.25	0.0700	0.0860	0.0606	0.0716	0.0521	0.0594	0.0462	0.0551	0.0375	0.0450	0.0279	0.0323
	0.5	0.0728	0.0852	0.0591	0.0685	0.0487	0.0592	0.0451	0.0539	0.0351	0.0431	0.0285	0.0336
	1	0.0733	0.0845	0.0557	0.0687	0.0479	0.0571	0.0457	0.0511	0.0372	0.0430	0.0283	0.0314

Note: Nominal size 5%. The number of simulations is 10,000. For the SW estimator, we apply the truncation rate $b_n = \widehat{C}_0 \log^{-1} n$ with cross-validated $\widehat{C}_0$. For the ADF test, the lag length $p = 0$.

5.5. Local Power of the SW Local Unit Root Test

In the classical ADF framework, the sequence $\alpha = 1 + \frac{\tau}{n}$ is typically viewed as a local alternative to the unit root null hypothesis $\mathcal{H}_0 : \alpha = 1$, leading to the conventional notion of local asymptotic power. In contrast, our analysis adopts a broader perspective. We treat $\alpha = 1 + \frac{\tau}{n}$ as part of the null hypothesis, thereby considering a far wider null space than the strict unit root case. This formulation is practically motivated, as it reflects the empirical challenge of distinguishing exact unit roots from local-to-unity processes.

Within this framework, we use the term "local power" in a distinct sense, specifically, to refer to the finite sample rejection probabilities computed under sequences of local alternatives $\mathcal{H}_1 : \alpha = 1 + \frac{\tau}{n^\gamma}$ with $\gamma \in (0, 1]$, where $\gamma = 1$ corresponds to the empirical size and $\gamma < 1$ captures the local power in our setup. Further, local power performance of the SW and ADF tests was explored using size-corrected power curves for both tests under local alternatives $\mathcal{H}_1$ with values of the localizing exponent $\gamma \in (0, 1]$ and $\tau_0 > 0$, that is, under stationary LUR alternatives $\mathcal{H}_1$. For each DGP, 10,000 replications were employed with sample size $n = 200$.

Figures 6–8 present the size-corrected power curves of the SW and ADF tests, where the innovations $\epsilon_k \sim_{iid} \mathcal{N}(0, \sigma_\epsilon^2)$, with variances $\sigma_\epsilon^2 \in \{0.25, 0.5, 1\}$, and localizing coefficients $\tau \in \{-0.1, -0.5, -1, 0.1, 0.5, 1\}$, and $\tau_0 = |\tau| > 0$ under $\mathcal{H}_1$. Power performance of the SW test with data-driven truncation $\widehat{C}_0$ selected by cross-validation is compared with the conventional ADF test.

The size-corrected power function is evaluated over $\gamma \in (0, 1]$. When $\tau_0 = \tau > 0$, the case $\gamma = 1$ corresponds to the test size, while $\gamma < 1$ represents local alternatives. When $\tau_0 = -\tau > 0$, the power function for any $\gamma \in (0, 1]$ captures the local power performance. To ensure comparability, empirical critical values are computed under the null to deliver exact 5% rejection rates, and these are subsequently used to compute the power under alternatives. The results show that the SW test with cross-validated $\widehat{C}_0$ achieves power levels similar to the ADF test under mild local alternatives. Nonetheless, under more substantial deviations (i.e., smaller values of γ), the SW test exhibits some power deficiency relative to the ADF test. This is no doubt associated with a key property of the test that involves trimming out data, which reduces the effective sample size and, in doing so, thereby power. In some ways, this is an inevitable characteristic of trimming and a well-known implication of such procedures. Power reduction of this type is accentuated in right-sided testing of a UR against explosive alternatives where trimming of extreme observations in the explosive region has a direct impact on the power delivered by such observations. For right-sided tests, an alternative approach would be to formulate testing in a different manner wherein the null hypothesis is mildly explosive paper under which the limit theory is known (Phillips and Magdalinos, 2009) against the alternative of a UR or stationary LUR. That approach is left as a topic for future research.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

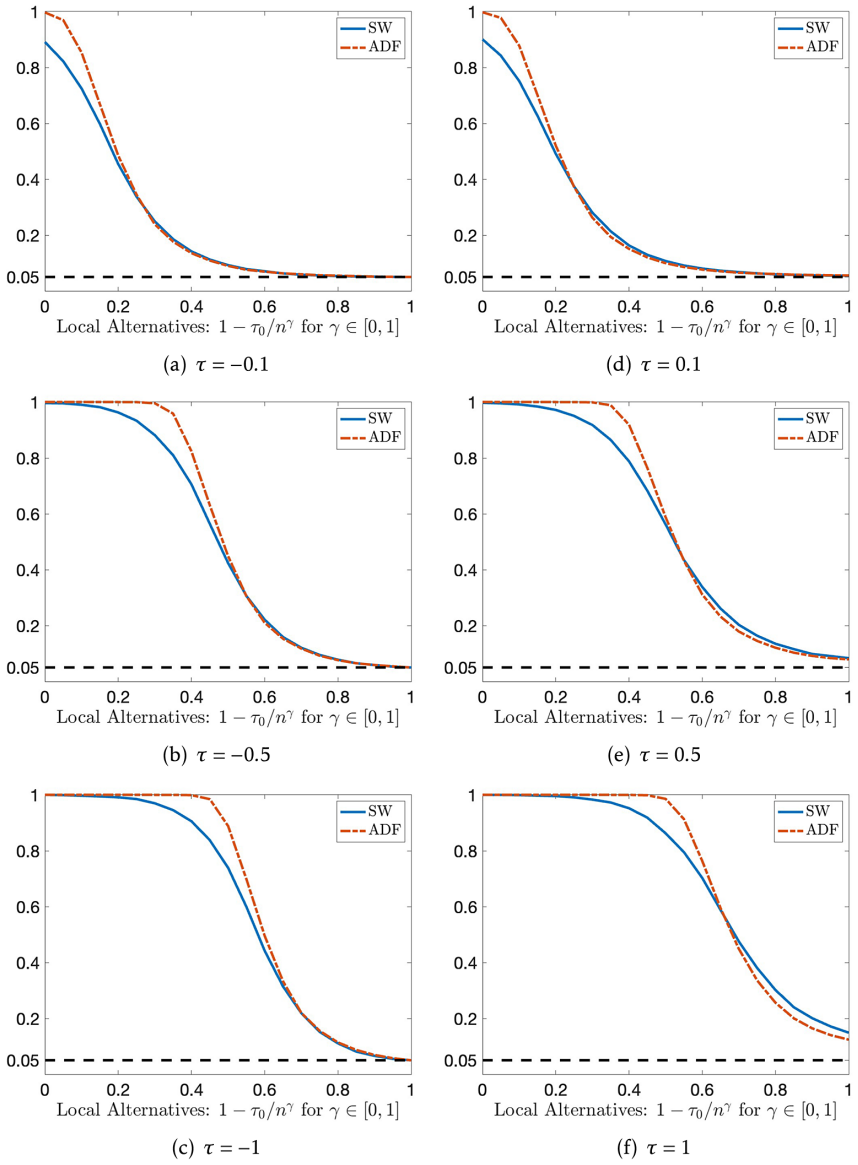


FIGURE 6. The local power of the SW and ADF unit root tests when $\epsilon_k \sim N(0, 0.25)$.
Note: The sample size $n = 200$. The number of replications is 10,000. The parameter $\tau_0 = |\tau| > 0$ under $\mathcal{H}_1$. Empirical critical values are computed under the null to deliver exact 5% rejection rates. For the SW estimator, we apply the truncation rate $b_n = \widehat{C}_0 \log^{-1} n$ with cross-validated $\widehat{C}_0$. For the ADF test, the lag length $p = 0$.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

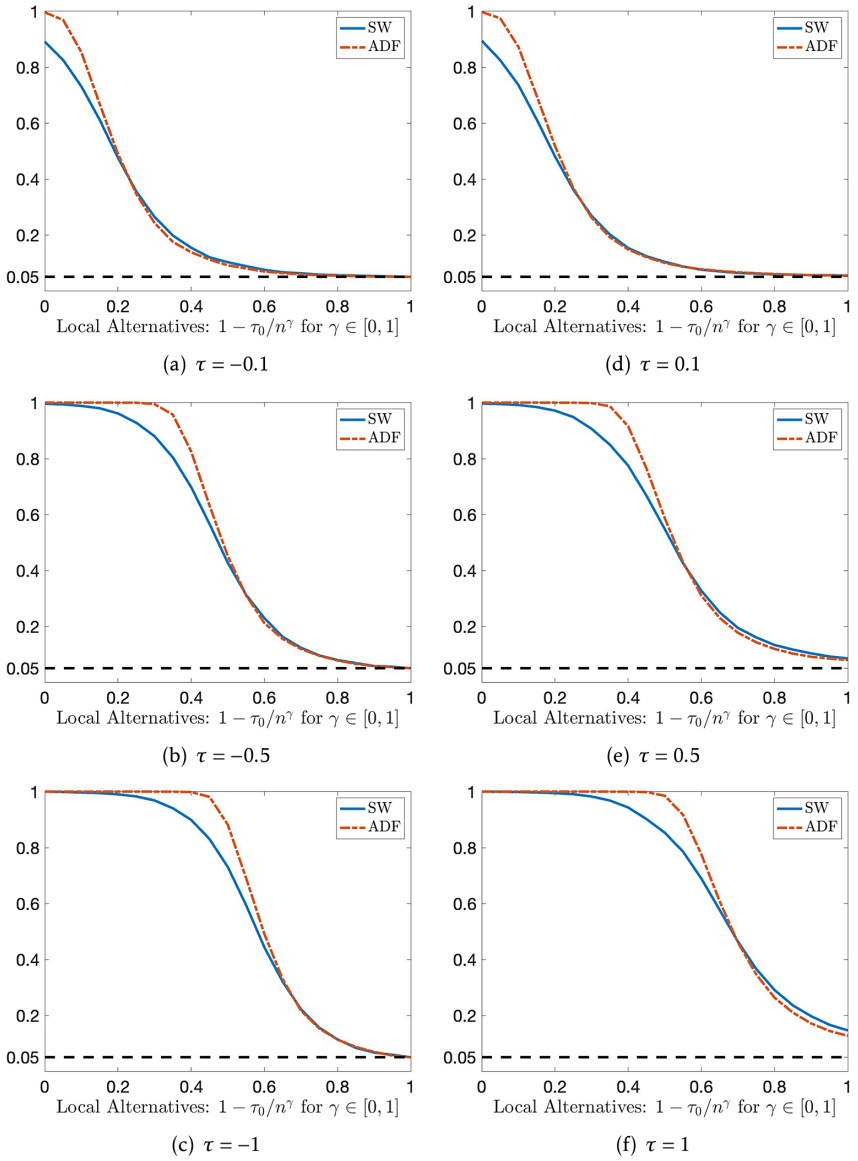


FIGURE 7. The local power of the SW and ADF unit root tests when $\epsilon_k \sim N(0, 0.5)$.
Note: The sample size $n = 200$. The number of replications is 10,000. The parameter $\tau_0 = |\tau| > 0$ under $\mathcal{H}_1$. Empirical critical values are computed under the null to deliver exact 5% rejection rates. For the SW estimator, we apply the truncation rate $b_n = \widehat{C}_0 \log^{-1} n$ with cross-validated $\widehat{C}_0$. For the ADF test, the lag length $p = 0$.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

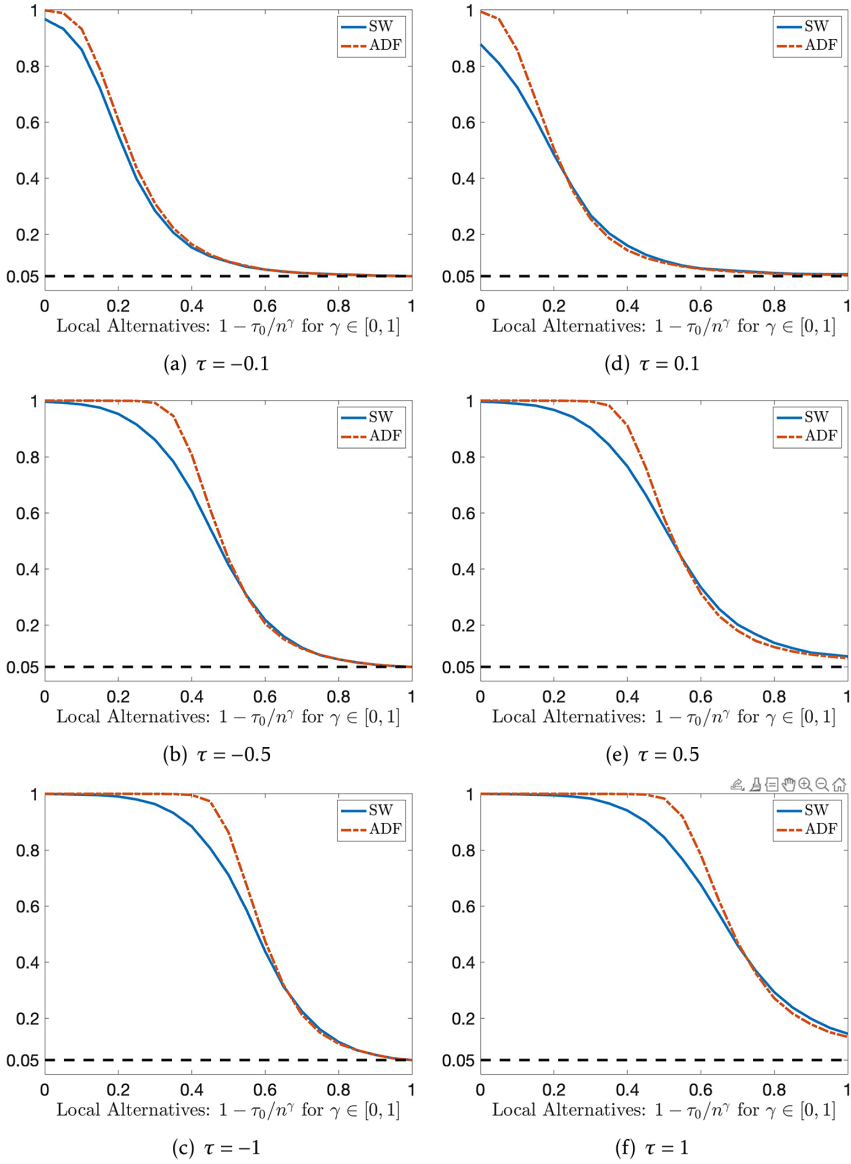


FIGURE 8. The local power of the SW and ADF unit root tests when $\epsilon_k \sim N(0,1)$.
Note: The sample size $n = 200$. The number of replications is 10,000. The parameter $\tau_0 = |\tau| > 0$ under $\mathcal{H}_1$. Empirical critical values are computed under the null to deliver exact 5% rejection rates. For the SW estimator, we apply the truncation rate $b_n = \widehat{C}_0 \log^{-1} n$ with cross-validated $\widehat{C}_0$. For the ADF test, the lag length $p = 0$.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

6. EMPIRICAL ILLUSTRATION

This section applies the SW test to examine the persistence of a set of market fundamentals that have been widely employed as predictors of S&P 500 excess returns in the literature (Kostakis, Magdalinos, and Stamatogiannis, 2015; Liu et al., 2025). We employ the dataset of Welch and Goyal (2008), covering the period from January 1960 to December 2022, at both monthly and quarterly frequencies. The predictors include the dividend payout ratio (d/e), long-term yield (lty), dividend yield (d/y), T-bill rate (tbl), earnings–price ratio (e/p), book-to-market ratio (b/m), net equity expansion (ntis), term spread (tms), and inflation rate (infl).

For the SW confidence interval and testing procedures, the truncation rate is specified as $b_n = C_0 \log^{-1}(n)$, where C_0 is selected by cross-validation, as discussed in Section 5.1. For comparison, we also report results from the ADF test, estimated from an AR(p) model with a constant term, where the lag length p is determined using the MIC proposed by Ng and Perron (2001).

Table 4 presents the results based on monthly data. Both the SW and ADF tests reject the LUR hypothesis for the dividend payout ratio (d/e) and term spread (tms) at the 1% significance level, and for net equity expansion (ntis) and inflation rate (infl) at the 5% level. For the earnings–price ratio (e/p), the ADF test rejects the LUR hypothesis at the 1% level, whereas the SW test does not. For the remaining variables, including long-term yield (lty), T-bill rate (tbl), and book-to-market ratio (b/m), both tests fail to reject the LUR null, indicating strong persistence in these predictors. Table 5 reports the corresponding results based on quarterly data. The SW test statistics suggest a lower degree of persistence relative to the monthly data, consistent with the intuition that lower-frequency data tend to exhibit reduced persistence.

Table 4. 95% confidence intervals and LUR tests for predictive regressors (monthly data).

Variables	95% CI		SW	ADF
Dividend payout ratio (d/e)	0.9678	0.9821	−6.8597***	−5.1761***
Long-term yield (lty)	0.9950	1.0052	0.0253	−1.3337
Dividend yield (d/y)	0.9870	1.0025	−1.3206*	−1.3180
T-bill rate (tbl)	0.9865	1.0061	−0.7403	−2.4452
Earnings–price ratio (e/p)	1.0147	1.0315	5.3712	−3.8489***
Book-to-market ratio (b/m)	0.9914	1.0023	−1.1296	−1.1896
Net equity expansion (ntis)	0.9707	1.0015	−1.7704**	−2.9942**
Term spread (tms)	0.9411	0.9726	−5.3698***	−3.9619***
Inflation rate (infl)	0.8081	0.9751	−2.5445***	−2.9836**

Note: The columns labeled "95% CI" report the lower and upper bounds of the confidence intervals constructed based on Theorem 3.1. The SW and ADF columns report the self-weighted local unit root and augmented Dickey–Fuller test statistics, respectively. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period spans from January 1960 to December 2022.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

Table 5. 95% confidence intervals and LUR tests for predictive regressors (quarterly data).

Variables	95% CI		SW	ADF
Dividend payout ratio (d/e)	0.7797	0.8723	−7.3597***	−4.8355***
Long-term yield (lty)	0.9724	1.0059	−1.2728	−1.4820
Dividend yield (d/y)	0.9547	1.0051	−1.5654*	−1.5560
T-bill rate (tbl)	0.9209	0.9926	−2.3660***	−2.3420
Earnings–price ratio (e/p)	1.0814	1.1469	6.8283	−2.9430**
Book-to-market ratio (b/m)	0.9637	1.0224	−0.4646	−1.3027
Net equity expansion (ntis)	0.9043	1.0068	−1.7015**	−1.8641
Term spread (tms)	0.8022	0.9462	−3.4261***	−3.3985**
Inflation rate (infl)	0.7221	0.9959	−2.0183**	−2.2736

Note: The columns labeled "95% CI" report the lower and upper bounds of the confidence intervals constructed based on Theorem 3.1. The SW and ADF columns report the self-weighted local unit root and augmented Dickey–Fuller test statistics, respectively. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period spans from 1960:Q1 to 2022:Q4.

7. CONCLUSION

This article provides a new SW LS approach to estimation and inference with LUR regressors. The procedure trims extreme data observations by means of a tuning parameter and has robust mixed Gaussian asymptotics, allowing for autocorrelated and heteroskedastic errors in near-unit root and predictive regression. A new rolling cross-validation training scheme is developed for the selection of the tuning parameter. With this framework, the method is used in the present work to provide a new approach to LUR testing that has the advantage of simple pivotal standard normal inference. Simulation findings indicate that the method works well against conventional tests, such as ADF in terms of both size control and local power.

The idea of trimming data of extreme observations has a long history in statistics, primarily in the context of removing outliers, for example, in mean estimation in the presence of heavy tailed data. The present work considers nonstationary data instead and the key idea is to soften the extremes of nonstationarity to facilitate inference by delivering standard normal limit theory and removing nuisance parameter dependencies. The approach reveals some of the utility of this idea in conducting univariate inference with nonstationary data. But the ideas that underly the approach are also applicable in wider multivariate contexts such as cointegrating regression. In particular, they can be used to provide an alternative to the popular IVX regression approach (Phillips and Magdalinos, 2009; Kostakis, Magdalinos, and Stamatogiannis, 2015) to inference. In fact, both SW regression and IVX rely on similar ideas: IVX by use of carefully constructed endogenous

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

instruments; and SW by showing that instrumentation is not needed with careful softening in the regression itself.[8]

SW estimation can also be used in nonlinear cointegrating regression. In this regard, a general model was considered recently in Jin and Wang (2024) using strong smoothing conditions, which are difficult to verify for some dependent sequences, including martingale differences. The new local time limit theory of this article avoids the use of strong smoothing conditions in the innovations and in doing so enables applications of SW estimation to nonlinear nonparametric cointegrating regression. A result of this type was given in Hu et al. (2024), which relates to the present article. But local time does not exist for limit processes like Brownian motion and linear diffusions of dimension that exceed two, so technical difficulties arise in establishing limit theory for SW estimation and testing in multiple regression with nonstationary time series, in contrast to the IVX approach (e.g., Phillips and Magdalinos, 2009; Kostakis, Magdalinos, and Stamatogiannis, 2015; Magdalinos and Petrova, 2022). Allowing for multiple regressors is therefore an important advantage of IVX in comparison with the SW approach where analysis is presently limited to the scalar regressor case. Potential extensions of SW estimation and testing to the multivariate setting are currently under exploration.

A. PROOFS

A.1. Proof of Theorem 2.1

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

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

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

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

[8] Notably, SW softens the data directly in the regression, whereas IVX softens only the instruments. This may explain why SW has difficulty doing better than IVX in simulation experiments we have conducted separate from the present article: in effect, the goal of facilitating inference by an approach that delivers standard limit theory can be achieved by softening only instrumentation in nonstationary regression rather than all of the data. Similar intuition explains why SW shows some evidence of power deficiency relative to ADF in testing the LUR hypothesis in the simulations reported in Section 5.5. Attention to a new general instrumental variable version of the methodology employed in the present article is being pursued in ongoing research.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

$$= \frac{1}{2\pi}\int_{-ac_n}^{ac_n}\int_0^1 \widehat{g}\Big(\frac{[nu]}{n}, -\frac{t}{c_n}\Big)e^{itX_u}du\,dt + o_{\mathrm{P}}(1)$$
$$= R_{1n}(A) + R_{1n}(A), \tag{A.1}$$

where

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

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 $n \to \infty$,

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

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

$$\mathrm{E}|R_{2n}(A)|^2 \to 0, \tag{A.3}$$

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

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

as $n \to \infty$ first and then $A \to \infty$. This proves (A.3) and also completes the proof of (2.5) when $g(s,x)$ is continuous by imposing **Con**.

We next remove the additional condition **Con** when $g(s,x)$ is continuous on $[0,1]\times\mathbb{R}$ satisfying $|g(s,x)| \le C/(1+|x|^{1+b})$. This is essentially the same as in the proof of Theorem IV 2.1 in Borodin and Ibragimov (1995) (see also Wang, 2015), and so we only provide an outline. Write, for $0 < b < 1$,

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

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

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 $c_2 > 0$,

$$c_1/(1+|x|^{1+b/2}) \le f(x) \le c_2/(1+|x|^{1+b/2}). \tag{A.4}$$

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

$$\sup_{\substack{0\le s\le 1\\ x\in R}} |g_\delta(s,x) - g(s,x)|$$

$$\le \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\sin^2 y}{y^2} \sup_{\substack{0\le s\le 1\\ x\in R}} \big|g(s,x+\delta y) - g(s,x)\big| dy \ \le\ \epsilon, \tag{A.5}$$

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

$$\le \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\sin^2 y}{y^2} \int_{-\infty}^{\infty} \sup_{0\le s\le 1} \big|g(s,x+\delta y) - g(s,x)\big| dxdy \ \le\ \epsilon, \tag{A.6}$$

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 $A \in \mathbb{R}$ satisfying $\int_{-\infty}^{\infty} \sup_{0\le s\le 1} \big|g(s,x)\big| dx < \infty$.

It follows from (A.4)–(A.6) that, for small $\epsilon > 0$, there exists an $\delta > 0$ such that

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

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

$$\text{both } g^+(s,x) \text{ and } g^-(s,x) \text{ satisfy } \mathbf{Con}. \tag{A.7}$$

Indeed, in terms of (A.7), it follows from the result proved above that

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

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

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 since ϵ is arbitrary.

The verification of (A.7) is simple. 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

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

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

we have

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

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

A.2. 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 condition **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 (A.1), we may write

$$\frac{c_n}{n}\sum_{k=1}^{n} g\Big(\frac{k}{n}, c_n X_{nk}\Big)v_k = A_0 \frac{c_n}{n}\sum_{k=1}^{n} g\Big(\frac{k}{n}, c_n X_{nk}\Big) + \frac{1}{2\pi}\int_{-ac_n}^{ac_n} A_n(t)dt, \tag{A.8}$$

where $A_n(t) = \frac{1}{n}\sum_{k=1}^{n} \widehat{g}\big(\frac{k}{n}, -\frac{t}{c_n}\big) e^{itX_{nk}}\widetilde{v}_k$ 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),

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

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

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 $\sup_{k\geq 1}Ev_k<\infty$, we have

$$\begin{aligned}|A_n(t)| &\leq \frac{1}{n}\sum_{j=0}^{T_n}\Big|\sum_{k=jm+1}^{(j+1)m}\widehat{g}\Big(\frac{k}{n},-\frac{t}{c_n}\Big)e^{itX_{nk}}\widetilde{v}_k\Big|+\frac{C_1}{n}\sum_{k=mT_n+1}^{n}|\widetilde{v}_k|\\ &\leq \frac{1}{n}\sum_{j=0}^{T_n}\Big|\sum_{k=jm+1}^{(j+1)m}\widehat{g}\Big(\frac{k}{n},-\frac{t}{c_n}\Big)\widetilde{v}_k\Big|+\frac{C_1}{n}\sum_{k=mT_n+1}^{n}|\widetilde{v}_k|\\ &\quad+C|t|\max_{0\leq j\leq T_n jm<k\leq(j+1)m}|X_{nk}-X_{n,jm}|\frac{1}{n}\sum_{k=1}^{n}|\widetilde{v}_k|\\ &\leq \frac{C_1}{n}\sum_{j=0}^{T_n}\Big|\sum_{k=jm+1}^{(j+1)m}\widetilde{v}_k\Big|+O_P(|t|)\sup_{0\leq u\leq 1}\sup_{|u-v|\leq m/n}|X_{n,[nu]}-X_{n,[nv]}|\\ &\quad+O_P\Big(\frac{|n-mT_n|+m}{n}\Big).\end{aligned}\tag{A.9}$$

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

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

This, together with condition (ii) and (A.9), yields $\sup_{|t|\leq ac_n}|A_n(t)|=o_P(c_n^{-1})$. Taking this estimate into (A.8), the required (2.7) is proved. □

A.3. Proof of Theorem 3.1

Write

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

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

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

$$\big(S_n,\ V_n^2\big)\to_D\Big(\Big[\sigma\int_{|x|\leq 1}x^2dxL_{J_\tau}(1,0)\Big]^{1/2}N,\ \sigma^{-1}\int_{|x|\leq 1}x^2dxL_{J_\tau}(1,0)\Big).\tag{A.10}$$

Indeed, by noting that

$$\widehat{\alpha}_{1n}-\alpha=\frac{\sum_{k=1}^{n}u_k y_{k-1}I(|y_{k-1}|\leq b_n\sqrt{n})}{\sum_{k=1}^{n}y_{k-1}^2I(|y_{k-1}|\leq b_n\sqrt{n})},\tag{A.11}$$

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

the required results follow from a simple algebra by using (A.10), $\int_{|x|\le 1} x^2 dx = 2/3$ and the continuous mapping theorem.

To prove (A.10), 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

$$\sigma^2 = Eu_k^2 \quad \text{and} \quad \sup_{k\ge 1} E|u_k|^{2+\delta} < \infty, \tag{A.12}$$

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 $\alpha(n) \le Cn^{-(1+\delta)}$, it follows from Lemma 2 of Shao (1993) with a simple algebra that there exists a positive constant K depending only on δ such that

$$P\big(\max_{1\le m\le n} \big|\sum_{k=1}^m \big(\sigma_{k,A} - E\sigma_{k,A}\big)\big| \ge x\big) \le CKnx^{-1-\delta/(4+3\delta)}, \tag{A.13}$$

for every $x \ge Kn^{1/2}\log n$. 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 (A.13) that, when n is sufficiently large,

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

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

$$\max_{1\le m\le n} \Big|\sum_{k=1}^m E(u_k^2|\mathcal{F}_{k-1}) - m\sigma^2\Big| = O_P(n^{1-\delta/4(1+\delta)}), \tag{A.14}$$

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

$$\max_{0\le j\le n-m} E\Big|\frac{1}{m}\sum_{k=j+1}^{j+m} \sigma_{k,A} - E\sigma_{1,A}\Big| = O(m^{-\delta/4(1+\delta)}). \tag{A.15}$$

Due to (A.12) and (A.14), the strong approximation result from Proposition S1.1 in the Supplementary Material shows that, on a richer probability space, there exists a Brownian motion $B = \{B_t\}_{t\ge 0}$ and a random array $\{z_{nk}^*, n \ge 1, k = 1, 2, \ldots, n\}$ such that $\{z_{nk}^*, k = 1, 2, \ldots, n\} =_d \{z_k, k = 1, 2, \ldots, n\}$ for all $n \ge 1$ and

$$\sup_{0\le t\le 1} \Big|\frac{1}{\sqrt{n}\sigma} z_{n,\lfloor nt\rfloor}^* - B_t\Big| = O_P(n^{-\eta}), \tag{A.16}$$

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

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

$$\frac{y_{\lfloor nt \rfloor}}{\sqrt{n}} = \frac{z_{\lfloor nt \rfloor}}{\sqrt{n}} + \tau \int_0^t e^{\tau(t-s)} \frac{z_{\lfloor ns \rfloor}}{\sqrt{n}} ds + R_n(t), \tag{A.17}$$

where $\sup_{0\le t\le 1} |R_n(t)| = O_P(n^{-1/2})$. This, together with (A.16), indicates that there exists a random array $\{y^*_{nk}, n \ge 1, k = 1, 2, \ldots, n\}$ such that

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

for all $n \ge 1$, and

$$\sup_{0\le t\le 1} \Big| \frac{1}{\sqrt{n}\sigma} y^*_{n,\lfloor nt \rfloor} - J_\tau(t) \Big| = O_P(n^{-\eta}), \tag{A.18}$$

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

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

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 (A.15)) that, for any $b_n > 0$ satisfying $b_n \to 0$ and $b_n \log^K n \to \infty$ for some $K > 0$,

$$\begin{aligned} &\frac{1}{nb_n} \sum_{k=2}^{n} \sigma_{k,A}\, h(y_{k-1}/b_n\sqrt{n}) I(|y_{k-1}| \le b_n\sqrt{n}) \\ &\quad = \frac{1}{nb_n} \sum_{k=2}^{n} \sigma_{k,A} H(y_{n,k-1}/b_n) = \frac{E\sigma_{1,A}}{nb_n} \sum_{k=1}^{n} H(y_{nk}/b_n) + o_P(1), \end{aligned} \tag{A.19}$$

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 (A.19) and Theorem 2.1, we have

$$\begin{aligned} &\Big(\frac{1}{\sqrt{n}\sigma} \sum_{i=1}^{\lfloor nt \rfloor} u_i,\ \frac{1}{nb_n} \sum_{k=2}^{n} \sigma_k^2 H(y_{n,k-1}/b_n) \Big) \\ &\quad = \Big(\frac{1}{\sqrt{n}\sigma} \sum_{i=1}^{\lfloor nt \rfloor} u_i,\ \frac{\sigma^2}{nb_n} \sum_{k=1}^{n} H(y_{nk}/b_n) \Big) + o_P(1) \\ &\quad \Rightarrow \big(B_t,\ \sigma \int_{|x|\le 1} h(x) dx L_{J_\tau}(1,0) \big), \end{aligned} \tag{A.20}$$

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

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

We are now ready to prove (A.10). Let $S_n = \sum_{k=1}^{n} x_{nk}u_k$, where

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

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

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

Since result (A.20) with $h(x) = x^2$ yields that, on $D_{\mathbb{R}^2}[0,1]$,

$$\Big(\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{\lfloor nt\rfloor} u_i,\ \widetilde{V}_n^2\Big) = \Big(\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{\lfloor nt\rfloor} u_i,\ \sigma^2 V_n^2\Big) + o_P(1)$$
$$\Rightarrow \Big(B_t,\ \sigma\int_{|x|\leq 1} x^2 dx L_{J_\tau}(1,0)\Big), \tag{A.21}$$

by using Wang's extended martingale limit theorem (e.g., Theorem 3.13 in Wang, 2015 with $X_{nk} = x_{nk}u_k$ and $Z_{nk} = u_k/\sqrt{n}$), (A.10) will follow if we prove

$$\frac{1}{n}\sum_{k=1}^{n} E\big[u_k^2 I(|u_k| \geq \epsilon\sqrt{n})|\mathcal{F}_{k-1}\big] \to_P 0, \quad \text{for any } \epsilon > 0, \tag{A.22}$$

$$\sum_{k=1}^{n} x_{nk}^2 E\big[u_k^2 I(|x_{nk}u_k| \geq \epsilon)|\mathcal{F}_{k-1}\big] \to_P 0, \quad \text{for any } \epsilon > 0, \tag{A.23}$$

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} |x_{nk}| E(u_k^2|\mathcal{F}_{k-1}) \to_P 0. \tag{A.24}$$

The proof of (A.24) is simple. In fact, by noting $nb_n \to \infty$, it follows from (A.20) with $h(x) = |x|$ $(H(x) = h(\sigma x)I(\sigma|x| \leq 1))$ that

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

as required. As for (A.23), 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 $E\big(\epsilon_k^2|\mathcal{F}_{k-1}\big) = 1$ that

$$\sum_{k=1}^{n} x_{nk}^2 E\big[u_k^2 I(|x_{nk}u_k| \geq \epsilon)|\mathcal{F}_{k-1}\big]$$
$$\leq \sum_{k=1}^{n} x_{nk}^2 E\big[u_k^2 I(|u_k| \geq A)|\mathcal{F}_{k-1}\big] + \sum_{k=1}^{n} x_{nk}^2 I(|x_{nk}| \geq \epsilon/A) E\big[u_k^2|\mathcal{F}_{k-1}\big]$$
$$\leq \sum_{k=1}^{n} x_{nk}^2\sigma_k^2 I(\sigma_k \geq \sqrt{A}) + \sup_{k\geq 1} E\big[\epsilon_k^2 I(|\epsilon_k| \geq \sqrt{A})|\mathcal{F}_{k-1}\big]\sum_{k=1}^{n} x_{nk}^2\sigma_k^2 + A^2\epsilon^{-2}\sum_{k=1}^{n} x_{nk}^4\sigma_k^2$$

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

$$= \frac{1}{nb_n}\sum_{k=1}^{n} H_1(y_{n,k-1}/b_n)\sigma_k^2 I(\sigma_k \geq \sqrt{A})$$
$$+ \sup_{k\geq 1} E\big[\epsilon_k^2 I(|\epsilon_k| \geq \sqrt{A})|\mathcal{F}_{k-1}\big]\frac{1}{nb_n}\sum_{k=1}^{n} H_1(y_{n,k-1}/b_n)\sigma_k^2$$
$$+ A^2\epsilon^{-2}\frac{1}{(nb_n)^2}\sum_{k=1}^{n} H_2(y_{n,k-1}/b_n)\sigma_k^2, \tag{A.25}$$

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 $nb_n \to \infty$ 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 (A.19) and (A.25) 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 (A.23). Similarly, for $\epsilon > 0$, we have

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

implying (A.22). Equation (A.10) is now proved and we also complete the proof of Theorem 3.1. □

A.4. Proof of Theorem 4.1

It follows from (3.4) that, under $\mathcal{H}_0 : \alpha = 1 + \tau/n$,

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

On the other hand, by using (A.21), we have

$$R_n := \frac{1}{n^2 b_n^3}\sum_{k=1}^{n} y_{k-1}^2 I(|y_{k-1}| \leq b_n\sqrt{n}) \to_D \sigma\int_{|x|\leq 1} x^2 dx L_{J_\tau}(1,0).$$

In terms of these estimates, it follows that, under $\mathcal{H}_0$,

$$T_n = A_n + \hat{\sigma}_{1n}^{-1}\Big[\sum_{k=1}^{n} y_{k-1}^2 I(|y_{k-1}| \leq b_n\sqrt{n})\Big]^{1/2}(\alpha - 1)$$
$$= A_n + \tau\, b_n^{3/2}\hat{\sigma}_{1n}^{-1} R_n^{1/2} \to_D \mathcal{N}(0,1),$$

due to $b_n \to 0$ and $\hat{\sigma}_{1n}^2 \to_P \sigma^2$. This proves (4.2).

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

We next prove (4.3). First note that, when the real model is (1.1) with the α being replaced by $\rho_n = 1 - \frac{\tau_0}{k_n}$ (i.e., under $\mathcal{H}_1$), we have

$$\Big[\sum_{k=1}^{n} y_{k-1}^2 I(|y_{k-1}| \le b_n\sqrt{n})\Big]^{1/2}\big(\widehat{\alpha}_{1n} - \alpha\big)$$
$$= \widetilde{T}_n + \Big[\sum_{k=1}^{n} y_{k-1}^2 I(|y_{k-1}| \le b_n\sqrt{n})\Big]^{1/2}(1-\rho_n),$$

where $\widetilde{T}_n = \Big[\sum_{k=1}^n y_{k-1}^2 I(|y_{k-1}| \le b_n\sqrt{n})\Big]^{-1/2} \sum_{k=1}^n y_{k-1} I(|y_{k-1}| \le b_n\sqrt{n}) u_k$. Since $\hat{\sigma}_{1n}^2 \to_P \sigma^2$, it suffices to show that, under $\mathcal{H}_1$,

$$\widetilde{T}_n \to_D \mathcal{N}(0, \sigma^2), \tag{A.26}$$

$$\frac{1-\rho_n^2}{n}\sum_{k=1}^{n} y_{k-1}^2 I(|y_{k-1}| \le b_n\sqrt{n}) \to_P \sigma^2. \tag{A.27}$$

Indeed, it follows from (A.26) and (A.27) that

$$(k_n/n)^{1/2}|T_n| = (k_n/n)^{1/2}\widetilde{T}_n + \frac{k_n^{1/2}|1-\rho_n|}{(1-\rho_n^2)^{1/2}}\Big[\frac{1-\rho_n^2}{n}\sum_{k=1}^{n} y_{k-1}^2 I(|y_{k-1}| \le b_n\sqrt{n})\Big]^{1/2}$$
$$\to_P (\tau_0/2)^{1/2}\sigma,$$

implying (4.3) due to $k_n/n \to 0$.

We next prove (A.26) and (A.27). To this end, write $S_n = \sum_{k=1}^n y_{k-1}u_k$ and $\sigma_n^2 = \sum_{k=1}^n y_{k-1}^2$. Recall that $y_k = \sum_{j=1}^k \rho_n^{j-k} u_j$ with $\rho_n = 1 - \frac{\tau_0}{k_n}$ under $\mathcal{H}_1$. We have the following proposition.

PROPOSITION A.1. *For $y_0 = O_P(1)$ and any $k_n > 0$ so that $n\rho_n \to \infty$, we have*

(a) $\frac{y_{\lfloor nt\rfloor}}{\sqrt{k_n}} \Rightarrow \frac{\sigma}{\sqrt{\tau_0}} B_t$ *on* $D[0,1]$,

(b) $\frac{(1-\rho_n^2)^{1/2}}{\sqrt{n}} S_n \to_D N(0, \sigma^4)$, *and*

(c) $\frac{1-\rho_n^2}{n}\sigma_n^2 \to_P \sigma^2$.

The proof of Proposition A.1 is similar to that of Giraitis and Phillips (2006, Lemmas 1 and 2) and hence the details are omitted. In terms of (b) and (c) in Proposition A.1, under $\mathcal{H}_1$, we have $S_n/\sigma_n \to_D N(0,\sigma^2)$. Hence, (A.26) follows from the fact that, by (a) of Proposition A.1 and $nb_n^2/k_n \to \infty$,

$$P\big(\widetilde{T}_n \ne S_n/\sigma_n\big) \le P\big(\max_{1\le k\le n} |y_k|/\sqrt{k_n} \ge b_n\sqrt{n/k_n}\big) \to 0.$$

Similarly, (A.27) follows from

$$P\big[\sigma_n^2 \ne \sum_{k=1}^{n} y_{k-1}^2 I(|y_{k-1}| \le b_n\sqrt{n})\big] \le P\big(\max_{1\le k\le n} |y_k|/\sqrt{k_n} \ge b_n\sqrt{n/k_n}\big) \to 0.$$

The proof of (4.3) is complete. □

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

A.5. Proof of Theorem 4.2

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 $|z| \le 1$ so that

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

It follows from Phillips and Solo (1992) that

$$\widetilde{u}_k = \Theta^{-1}(1)\, u_k + \hat{u}_{k-1} - \hat{u}_k, \tag{A.28}$$

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 $u_{-i} \ge 0$ if $i \ge 0$)

$$\max_{1\le k\le n} |\hat{u}_k| \le \sum_{i=0}^{\infty} |\hat{\pi}_i| \max_{1\le k\le n} |u_k| = O_P(n^{1/(2+\delta)}), \tag{A.29}$$

where we have used the facts that $\sum_{i=0}^{\infty} |\hat{\pi}_i| \le \sum_{i=0}^{\infty} i|\pi_i| < \infty$ and, due to (A.12),

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

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

$$\max_{0\le j\le n-m} E\Big|\sum_{k=j+1}^{j+m} \widetilde{u}_k\Big| \le C m^{1/2}, \tag{A.30}$$

for some constant $C > 0$. Furthermore, when $\alpha = 1$, we may rewrite (4.4) as

$$y_k = y_{k-1} + \widetilde{u}_k, \tag{A.31}$$

so that $z_k = (\widetilde{u}_{k-1}, \ldots, \widetilde{u}_{k-p})$ (recall $\widetilde{u}_j = 0$ if $j < 0$). See Chang and Park (2002), for instance.

The proof of (4.6) now follows from the following lemmas. Define $||z|| = (z_1^2 + \cdots + z_p^2)^{1/2}$ for a vector $z = (z_1, \ldots, z_p)$ and $||A|| = \max_z ||Az||/||z||$ for a $p \times p$ matrix A.

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

Proof. (a) and (b) follow from Lemma 3.2 of Chang and Park (2002). See also Berk (1974). (c) is obvious by using the martingale properties. We next prove (d), starting with

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

some preliminaries. It follows from (A.28) and (A.31) that

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

Since $|\hat{u}_0| + \max_{1\le k\le n}|\hat{u}_k| = O_P(n^{1/(2+\delta)})$ for some $\delta > 0$ by using (A.29), on a richer probability space, there exists a Brownian motion $B = \{B(t)\}_{t\ge 0}$ so that, for some $\eta > 0$,

$$\sup_{0\le t\le 1}\left| y_{n,\lfloor nt\rfloor} - \frac{\Theta^{-1}(1)B(nt)}{\sqrt{n}}\right| = O_P(n^{-\eta}), \tag{A.32}$$

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

$$\begin{aligned}
&\sup_{0\le u\le 1}\sup_{|u-v|\le m/n}|y_{n,[nu]} - y_{n,[nv]}| \\
&\le \Theta^{-1}(1)\sup_{0\le u\le 1}\sup_{|u-v|\le m/n}\left|\frac{B(nu)}{\sqrt{n}} - \frac{B(nv)}{\sqrt{n}}\right| + o_P(n^{-\eta}) \\
&= O_P(1)(m/n)^{1/2}\log(m/n) + o_P(n^{-\eta}).
\end{aligned} \tag{A.33}$$

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

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

Recall that $z_k = (\widetilde{u}_{k-1}, \ldots, \widetilde{u}_{k-p})$. In terms of (A.30) and (A.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

$$\begin{aligned}
&\left|\left|\sum_{k=1}^{n} z_k H_n(b_n^{-1}y_{nk})\right|\right| \le \sum_{j=0}^{T_n}\left|\left|\sum_{k=jm+1}^{(j+1)m} z_k H_n(b_n^{-1}y_{nk})\right|\right| + C\sum_{k=mT_n+1}^{n}||z_k|| \\
&\le C\sum_{j=0}^{T_n}\left|\left|\sum_{k=jm+1}^{(j+1)m} z_k\right|\right| + C\sum_{k=mT_n+1}^{n}||z_k|| \\
&\quad + \max_{0\le j\le T_n}\max_{jm<k\le(j+1)m}|H_n(b_n^{-1}y_{nk}) - H_n(b_n^{-1}y_{n,(j+1)m})|\sum_{k=1}^{n}||z_k|| \\
&\le O_P\big[nm^{-1/2} + (n - mT_n)\big] + O_P(nb_n^{-2})\sup_{0\le u\le 1}\sup_{|u-v|\le m/n}|y_{n,[nu]} - y_{n,[nv]}| \\
&= O_P(nb_n).
\end{aligned} \tag{A.34}$$

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

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

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

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

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

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 (A.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

$$R_n := \Big|\Big|\sum_{k=1}^{n} z_k l_{2n}(b_n^{-1} y_{nk})\Big|\Big| = O_P(nb_n). \tag{A.35}$$

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

$$\begin{aligned} R_n &\le \sum_{k=1}^{n} I(1-b_n \le |b_n^{-1} y_{n,k}| \le 1+b_n)||z_k|| \\ &\le \Big(\sum_{k=1}^{n} I(1-b_n \le |b_n^{-1} y_{nk}| \le 1+b_n) \sum_{k=1}^{n} ||z_k||^2\Big)^{1/2} \\ &= O_P(n^{1/2})\Big(\sum_{k=1}^{n} I(1-b_n \le |b_n^{-1} y_{nk}| \le 1+b_n)\Big)^{1/2}. \end{aligned} \tag{A.36}$$

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

$$\begin{aligned} &\sum_{k=1}^{n} I(1-b_n \le |b_n^{-1} y_{nk}| \le 1+b_n) \\ &\quad = n\int_0^1 I(1-b_n \le |b_n^{-1} y_{n,\lfloor nt\rfloor}| \le 1+b_n)dt + O_P(1) \\ &\quad \le n\int_0^1 I(1-2b_n \le |b_n^{-1} B_n(t)| \le 1+2b_n)dt + O_P(1) = O_P(nb_n^2), \end{aligned} \tag{A.37}$$

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

$$\begin{aligned} &\int_0^1 I(1-2b_n \le |b_n^{-1} B_n(t)| \le 1+2b_n)dt \\ &\overset{d}{=} \int_0^1 I(1-2b_n \le |b_n^{-1} B(a_0 t)| \le 1+2b_n)dt \\ &= a_0^{-1}\int_0^{a_0} I(1-2b_n \le |b_n^{-1} B(t)| \le 1+2b_n)dt \end{aligned}$$

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

$$= a_0^{-1} \int_{-\infty}^{\infty} I(1-2b_n \le |b_n^{-1}x| \le 1+2b_n) L_B(a_0,x)dx$$
$$= a_0^{-1} b_n \int_{-\infty}^{\infty} I(1-2b_n \le |x| \le 1+2b_n) L_B(a_0,xb_n)dx$$
$$= 8a_0^{-1} b_n^2 L_B(a_0,0)(1+o_P(1)).$$

Note that (A.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 (A.35) follows from (A.36) and (A.37) and hence the proof of (d) is complete. □

LEMMA A.2. *Recall* $\widetilde{y}_k = y_k I(|y_k| \le b_n\sqrt{n})$. *When* $\alpha = 1$, *we have*

$$\Big(\frac{1}{\sqrt{n^2 b_n^3}} \sum_{k=1}^{n} \widetilde{y}_{k-1}\, u_k,\ \frac{1}{n^2 b_n^3} \sum_{k=1}^{n} \widetilde{y}_{k-1}^2\Big) \to_D \big(\sigma\, Z^{1/2} N,\ Z\big), \tag{A.38}$$

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 (A.10) by using (A.32) and Theorem 2.2. The details are omitted. □ □

Returning to the proofs of (4.6) and (4.7), we note that, when $\alpha = 1$,

$$\widehat{\alpha}_{2n} - 1 = A_n / B_n, \tag{A.39}$$

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

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

Using Lemmas A.1 and A.2, simple calculations show that

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

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

due to $b_n \to 0$ and $b_n \log^K n \to \infty$ for some $K > 0$. This proves (4.6). The proof of (4.7) is simple and details are omitted. □

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at https://doi.org/10.1017/S0266466626100474.

FUNDING STATEMENT

Qiying Wang acknowledges funding support for the Open Access of this article from the University of Sydney. Peter C. B. Phillips acknowledges research support from the NSF (Grant No. SES 18-50860), a Kelly Fellowship at the University of Auckland, and an LKC Fellowship at Singapore Management University.

COMPETING INTERESTS

The authors declare that no competing interests exist.

REFERENCES

Akonom, J. (1993). Comportement asymptotique du temps d'occupation du processus des sommes partielles. *Annales de l'IHP Probabilités et Statistiques*, 29, 57–81.

Andrews, D. W., & Guggenberger, P. (2012). Asymptotics for LS, GLS, and feasible GLS statistics in an AR(1) model with conditional heteroskedasticity. *Journal of Econometrics*, 169, 196–210.

Athreya, K. B., & Pantula, S. G. (1986). A note on strong mixing of ARMA processes. *Statistics & Probability Letters*, 4(4), 187–190.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Annals of Statistics*, 2(3), 489–502.

Borodin, A. N., & Ibragimov, I. A. (1995). *Limit theorems for functionals of random walks*. American Mathematical Society.

Cavaliere, G., & Taylor, A. M. R. (2009). Heteroskedastic time series with a unit root. *Econometric Theory*, 25(5), 1228–1276.

Chan, N. H., Li, D., & Peng, L. (2012). Toward a unified interval estimation of autoregressions. *Econometric Theory*, 28(3), 705–717.

Chan, N. H., & Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics*, 15(3), 1050–1063.

Chang, Y., & Park, J. Y. (2002). On the asymptotics of ADF tests for unit roots. *Econometric Reviews*, 21(4), 431–447.

Demetrescu, M., Georgiev, I., Rodrigues, P. M., & Taylor, A. R. (2022). Testing for episodic predictability in stock returns. *Journal of Econometrics*, 227(1), 85–113.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427–431.

Fiorio, C. V., Hajivassiliou, V. A., & Phillips, P. C. B. (2010). Bimodal t-ratios: The impact of thick tails on inference. *The Econometrics Journal*, 13(2), 271–289.

Geman, D., & Horowitz, J. (1976). Occupation-times for functions with countable level sets and the regeneration of stationary processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 35(3), 189–211.

Geman, D., & Horowitz, J. (1980). Occupation densities. *The Annals of Probability*, 8(1), 1–67.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

Giraitis, L., & Phillips, P. C. B. (2006). Uniform limit theory for stationary autoregression. *Journal of Time Series Analysis*, 27(1), 51–60.

Han, C., Phillips, P. C. B., & Sul, D. (2011). Uniform asymptotic normality in stationary and unit root autoregression. *Econometric Theory*, 27(6), 1117–1151.

Hansen, B. E. (1999). The grid bootstrap and the autoregressive model. *Review of Economics and Statistics*, 81(4), 594–607.

Herranz, E. (2017). Unit root tests. *Wiley Interdisciplinary Reviews: Computational Statistics* 9(3), e1396.

Hu, Z., Kasparis, I., & Wang, Q. (2021). *Chronological trimming methods for nonlinear predictive regressions with persistent data*. University of Cyprus.

Hu, Z., Liu, N., Phillips, P. C. B., & Wang, Q. (2024). Self-weighted estimation for local unit root regression with applications. Cowles Foundation discussion paper no. 2400, Cowles Foundation for Research in Economics, Yale University.

Jeganathan, P. (2004). Convergence of functionals of sums of r.v.s to local times of fractional stable motions. *The Annals of Probability*, 32(3A), 1771–1795.

Jin, C., & Wang, Q. (2024). Weighted nonlinear regression with nonstationary time series. *Statistica Sinica*, 34, 1765–1800.

Kim, T. H., & Schmidt, P. (1993). Unit root test with conditional heteroskedasticity. *Journal of Econometrics*, 59(3), 287–300.

Kostakis, A., Magdalinos, T., & Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *The Review of Financial Studies*, 28(5), 1506–1553.

Li, D., Tjostheim, D., & Gao, J. (2016). Estimation in nonlinear regression with Harris recurrent Markov chains. *Annals of Statistics*, 44(4), 1957–1987.

Ling, S. (2007). Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics*, 140(2), 849–873.

Ling, S., Li, W. K., & McAleer, M. (2003). Estimation and testing for unit root process with GARCH(1,1) errors: Theory and Monte Carlo evidence. *Econometric Reviews*, 22(2), 179–202.

Liu, N., Liu, Y., Phillips, P. C. B., & Zhang, Y. (2025). Robust inference for time-varying predictability: A Sieve-IVX approach. Working Paper.

Liu, Y., & Phillips, P. C. B. (2023). Robust inference with stochastic local unit root regressors in predictive regressions. *Journal of Econometrics*, 235(2), 563–591.

Magdalinos, T., & Petrova, K. (2022). Uniform and distribution-free inference with general autoregressive processes. Economics Working Papers 1837, Department of Economics and Business, Universitat Pompeu Fabra.

Mikusheva, A. (2007). Uniform inference in autoregressive models. *Econometrica*, 75(5), 1411–1452.

Mikusheva, A. (2012). One-dimensional inference in autoregressive models with the potential presence of a unit root. *Econometrica*, 80(1), 173–212.

Ng, S., & Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69(6), 1519–1554.

Phillips, P. C. B. (1987a). Time series regression with a unit root. *Econometrica*, 55(2), 277–301.

Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika*, 74(3), 535–547.

Phillips, P. C. B. (1988). Regression theory for near-integrated time series. *Econometrica*, 56(5), 1021–1043.

Phillips, P. C. B. (2007). Unit root log periodogram regression. *Journal of Econometrics*, 138(1), 104–124.

Phillips, P. C. B. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica*, 82(3), 1177–1195.

Phillips, P. C. B. (2023). Estimation and inference with near unit roots. *Econometric Theory*, 39(2), 221–263.

Phillips, P. C. B., & Han, C. (2008). Gaussian inference in AR(1) time series with or without a unit root. *Econometric Theory*, 24(3), 631–650.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press

Phillips, P. C. B., & Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, 136(1), 115–130.
Phillips, P. C. B., & Magdalinos, T. (2009). Econometric inference in the vicinity of unity. Singapore Management University, CoFie Working Paper 7, 35.
Phillips, P. C. B., & Park, J. Y. (1998). Nonstationary density estimation and kernel autoregression. Cowles Foundation Discussion Paper 1181.
Phillips, P. C. B., Park, J. Y., & Chang, Y. (2004). Nonlinear instrumental variable estimation of an autoregression. *Journal of Econometrics*, 118(1–2), 219–246.
Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2), 335–346.
Phillips, P. C. B., & Solo, V. (1992). Asymptotics for linear processes. *Annals of Statistics*, 20(2), 971–1001.
Phillips, P. C. B., & Xiao, Z. (1998). A primer on unit root testing. *Journal of Economic Surveys*, 12(5), 423–470.
Shao, Q. M. (1993). Complete convergence for α-mixing sequences. *Statistics & Probability Letters*, 16(4), 279–287.
Shi, S., & Phillips, P. C. B. (2022). Econometric analysis of asset price bubbles. Working Paper.
Shi, S., & Phillips, P. C. B. (2023). Diagnosing housing fever with an econometric thermometer. *Journal of Economic Surveys*, 37(1), 159–186.
So, B. S., & Shin, D. W. (1999). Cauchy estimators for autoregressive processes with applications to unit root tests and confidence intervals. *Econometric Theory*, 15(2), 165–176.
Tu, Y., Chan, N., & Wang, Q. (2020). Testing for a unit root with nonstationary nonlinear heteroskedasticity. *Econometric Reviews*, 39(9), 904–929.
Wang, Q. (2014). Martingale limit theorem revisited and nonlinear cointegrating regression. *Econometric Theory*, 30(3), 509–535.
Wang, Q. (2015). *Limit theorems for nonlinear cointegrating regression*. World Scientific.
Wang, Q., Lin, Y., & Gulati, C. M. (2003). Asymptotics for general fractionally integrated processes with applications to unit root tests. *Econometric Theory*, 19(1), 143–164.
Wang, Q., & Phillips, P. C. B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, 25(3), 710–738.
Wang, Q., & Phillips, P. C. B. (2009b). Structural nonparametric cointegrating regression. *Econometrica*, 77(6), 1901–1948.
Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4), 1455–1508.
Wu, W. B. (2006). Unit root testing for functionals of linear processes. *Econometric Theory*, 22(1), 1–14.
Xu, K. L., & Phillips, P. C. B. (2008). Adaptive estimation of autoregressive models with time-varying ce. *Journal of Econometrics*, 142(1), 265–280.
Yang, B., Long, W., Peng, L., & Cai, Z. (2020). Testing the predictability of U.S. housing price index returns based on an IVX-AR model. *Journal of the American Statistical Association*, 115(532), 1598–1619.

https://doi.org/10.1017/S0266466626100474 Published online by Cambridge University Press