Biometrika (2022), **109**, 2, *pp.* 543–550
doi: 10.1093/biomet/asab034
Advance Access publication 14 September 2021

Testing for unit roots based on sample autocovariances

BY JINYUAN CHANG

School of Statistics, Southwestern University of Finance and Economics, 555 Liutai Avenue, Chengdu, Sichuan 611130, China
changjinyuan@swufe.edu.cn

GUANGHUI CHENG

School of Economics and Statistics, Guangzhou University, 230 Waichengxi Road, Guangzhou, Guangdong 510006, China
chenggh845@nenu.edu.cn

AND QIWEI YAO

Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, U.K.
q.yao@lse.ac.uk

SUMMARY

We propose a new unit-root test for a stationary null hypothesis H_0 against a unit-root alternative H_1. Our approach is nonparametric as H_0 assumes only that the process concerned is $I(0)$, without specifying any parametric forms. The new test is based on the fact that the sample autocovariance function converges to the finite population autocovariance function for an $I(0)$ process, but diverges to infinity for a process with unit roots. Therefore, the new test rejects H_0 for large values of the sample autocovariance function. To address the technical question of how large is large, we split the sample and establish an appropriate normal approximation for the null distribution of the test statistic. The substantial discriminative power of the new test statistic is due to the fact that it takes finite values under H_0 and diverges to infinity under H_1. This property allows one to truncate the critical values of the test so that it has asymptotic power 1; it also alleviates the loss of power due to the sample-splitting. The test is implemented in R.

Some key words: Autocovariance; Integrated process; Normal approximation; Power-one test; Sample-splitting.

1. INTRODUCTION

Models with unit roots are frequently used for nonstationary time series. The importance of the unit-root concept stems from the fact that many economic, financial, business and social-domain data exhibit segmented trend-like or random wandering phenomena. While the random-walk-like behaviour of stock prices was noticed much earlier, for example by Jules Regnault, a French broker, in 1863 and by Louis Bachelier in his 1900 PhD thesis, the development of statistical inference for unit roots started only in the late 1970s. Nevertheless, the literature on unit-root tests is by now immense and diverse. We review only a selection of important developments below, leading naturally to the new test presented in this paper.

The Dickey–Fuller tests (Dickey & Fuller, 1979, 1981) deal with Gaussian random walks with independent errors. Efforts to relax the condition of independent Gaussian errors have led to, among others, the augmented Dickey–Fuller tests (Said & Dickey, 1984; Elliott et al., 1996), which deal with autoregressive errors, and the Phillips–Perron test (Phillips, 1987; Phillips & Perron, 1988), which estimates the long-run variance of the error process nonparametrically. The augmented Dickey–Fuller tests have been further extended to deal with structural breaks in trend (Zivot & Andrews, 1992), long memory processes (Robinson, 1994), seasonal unit roots (Chan & Wei, 1988; Hylleberg et al., 1990), bootstrap unit-root tests (Paparoditis & Politis, 2005), nonstationary volatility (Cavaliere & Taylor, 2007), panel data (Pesaran, 2007) and local stationary processes (Rho & Shao, 2019); see the survey papers by Stock (1994) and Phillips & Xiao (1998), and the monographs by Hatanaka (1996) and Maddala & Kim (1998) for further references.

The Dickey–Fuller tests and their variants are based on regression of a time series on its first lag, in which the existence of a unit root is postulated as a null hypothesis in the form of the regression coefficient being equal to 1. This null hypothesis is tested against a stationary alternative hypothesis that the regression coefficient is smaller than 1. This setting leads to innate indecisive inference for ascertaining the existence of unit roots, as a statistical test is incapable of accepting a null hypothesis. To place the assertion of unit roots on firmer ground, Kwiatkowski et al. (1992) adopted a different approach: their proposed test considers a stationary null hypothesis against a unit-root alternative. It is based on a plausible representation of possible nonstationary time series in which a unit root is represented as an additive random-walk component. Then, under the null hypothesis, the variance of the random-walk component is zero. The test of Kwiatkowski et al. (1992) is the one-sided Lagrange multiplier test for testing the variance being zero against it being greater than zero.

Despite the many exciting developments mentioned above, testing for the existence of unit roots remains a challenge in time series analysis, as most available methods suffer from lack of accurate size control and low power. In this paper we propose a new test that is based on a radically different idea from existing approaches. Our setting is similar in spirit to that of Kwiatkowski et al. (1992), in that we test a stationary null hypothesis H_0 against a unit-root alternative H_1. However, our approach is nonparametric as H_0 assumes only that the process concerned is $I(0)$, without specifying any parametric forms. The new test is based on the simple fact that under H_0 the sample autocovariance function converges to the finite population autocovariance function, while under H_1 it diverges to infinity. Therefore, we can reject H_0 for large absolute values of the sample autocovariance function. To address the technical question of how large is large, we split the sample and establish an appropriate normal approximation for the null distribution of the test statistic. Our sample autocovariance function-based test statistic offers substantial discriminative power as it takes finite values under H_0 and diverges to infinity under H_1. This property allows us to truncate the critical values determined by the normal approximation to ensure that the test has asymptotic power 1; furthermore, it alleviates the loss of power due to the sample-splitting, so that our test outperforms the test of Kwiatkowski et al. (1992) in a power comparison simulation. Another advantage of the new method is that it has remarkable discriminative power, being able to tell the difference between, for example, a random walk and an AR(1) with autoregressive coefficient close to but still smaller than 1, a case in which most available unit-root tests, including the method of Kwiatkowski et al. (1992), suffer from weak discriminative power. Admittedly, the new test is technically sophisticated, which we argue is inevitable in order to gain improvement over existing methods. Nevertheless, we have developed an `R` (R Development Core Team, 2022) function `ur.test` in the package `HDTSA` that implements the test in an automatic manner.

2. Main results

2.1. *A power-one test*

A time series $\{Y_t\}$ is said to be $I(0)$, denoted by $Y_t \sim I(0)$, if $E(Y_t) \equiv \mu$, $E(Y_t^2) < \infty$ and $\sum_{k=0}^{\infty} |\gamma(k)| < \infty$ where $\gamma(k) \equiv \mathrm{cov}(Y_{t+k}, Y_t)$. Let $\nabla Y_t = Y_t - Y_{t-1}$, $\nabla^0 Y_t = Y_t$ and $\nabla^d Y_t = \nabla(\nabla^{d-1} Y_t)$ for any integer $d \geqslant 1$. A time series $\{Y_t\}$ is said to be $I(d)$, denoted by $Y_t \sim I(d)$, if $\{\nabla^d Y_t\}$ is $I(0)$ and $\{\nabla^{d-1} Y_t\}$ is not $I(0)$. An $I(d)$ process is also called a unit-root process with integration order d. Given the observations $\{Y_t\}_{t=1}^{n}$, we are interested in testing the hypotheses

$$H_0 : Y_t \sim I(0) \quad \text{versus} \quad H_1 : Y_t \sim I(d) \text{ for some integer } d \geqslant 1. \tag{1}$$

Write $\bar{Y} = n^{-1}\sum_{t=1}^{n} Y_t$ and denote the sample autocovariance function at lag k by $\hat{\gamma}(k) = n^{-1}\sum_{t=1}^{n-k}(Y_{t+k} - \bar{Y})(Y_t - \bar{Y})$, which is a consistent estimator for $\gamma(k)$ under H_0. When $Y_t \sim I(d)$, Wold's decomposition for the purely nondeterministic $I(0)$ process gives

$$\nabla^d Y_t = \mu_d + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \tag{2}$$

where $\mu_d = E(\nabla^d Y_t)$ is a constant, $\psi_0 = 1$ and $\{\epsilon_t\}$ is white noise. Proposition 1 indicates that $\hat{\gamma}(k)$ diverges to infinity under H_1. Thus we can reject H_0 for large values of $|\hat{\gamma}(k)|$.

Proposition 1. *Let Y_t satisfy (2) with independent $\epsilon_t \sim (0, \sigma_\epsilon^2)$ and $\sum_{j=1}^{\infty} j|\psi_j| < \infty$. Write $a = \sum_{j=0}^{\infty}\psi_j$ and $V_{d-1}(t) = F_{d-1}(t) - \int_0^1 F_{d-1}(t)\,\mathrm{d}t$ where $F_{d-1}(t)$ is the scalar multi-fold integrated Brownian motion defined recursively by $F_j(t) = \int_0^t F_{j-1}(x)\,\mathrm{d}x$ for any $j \geqslant 1$, with $F_0(t)$ the standard Brownian motion. For any given integer $k \geqslant 0$, as $n \to \infty$ we have that* (i) *$n^{-(2d-1)}\hat{\gamma}(k) \to a^2\sigma_\epsilon^2\int_0^1 V_{d-1}^2(t)\,\mathrm{d}t$ in distribution if $\mu_d = 0$, and* (ii) *$n^{-2d}\hat{\gamma}(k) \to \phi_{d,k}\mu_d^2$ in probability if $\mu_d \neq 0$, where $\phi_{d,k} > 0$ is a bounded constant depending only on d and k.*

By Proposition 1, one may consider rejecting H_0 for large values of $T_{\text{naive}} = \sum_{k=0}^{K_0}|\hat{\gamma}(k)|^2$ with a prescribed integer $K_0 \geqslant 0$, as T_{naive} converges to $\sum_{k=0}^{K_0}|\gamma(k)|^2 < \infty$ under H_0. Unfortunately, there are two obstacles to using T_{naive}: (i) to determine the critical values one has to derive the null distribution of $a_n\{T_{\text{naive}} - \sum_{k=0}^{K_0}|\gamma(k)|^2\}$ with some $a_n \to \infty$; (ii) one needs a consistent estimator for $\sum_{k=0}^{K_0}|\gamma(k)|^2$ under H_0, which is not readily available as we do not know whether or not H_0 holds in practice. To overcome these two obstacles, we implement the idea of data-splitting. Let $N = \lfloor n/2 \rfloor$. Define $\hat{\gamma}_1(k) = N^{-1}\sum_{t=1}^{N-k}(Y_{t+k} - \bar{Y})(Y_t - \bar{Y})$ and $\hat{\gamma}_2(k) = N^{-1}\sum_{t=N+1}^{2N-k}(Y_{t+k} - \bar{Y})(Y_t - \bar{Y})$. The test statistic for (1) is defined as

$$T_n = \sum_{k=0}^{K_0}|\hat{\gamma}_2(k)|^2,$$

where $K_0 \geqslant 0$ is a prescribed integer that controls the amount of information from different time lags to be used. Although our theory allows K_0 to diverge with the sample size n, the simulation results reported in § 3 indicate that the finite-sample performance of the test is robust with respect to different values of K_0 and the test works well even with small K_0.

Formally, we reject H_0 at the significance level $\phi \in (0, 1)$ if $T_n > \text{cv}_\phi$, where cv_ϕ is the critical value satisfying $\text{pr}_{H_0}(T_n > \text{cv}_\phi) \to \phi$. As we will see in (3), $\{\hat{\gamma}_1(k)\}_{k=0}^{K_0}$ are used to determine the critical value cv_ϕ. One obvious concern with splitting the sample into two halves is loss of testing power. However, the fact that T_n takes finite values under H_0 and diverges to infinity under H_1 implies that T_n has strong discriminative power to tell H_1 apart from H_0, which is enough to provide more power than, for example, the test of Kwiatkowski et al. (1992). Our simulation results indicate that the sample-splitting works well even for sample size $n = 80$. Under H_0, write $y_{t,k} = 2\{(Y_t - \mu)(Y_{t+k} - \mu) - \gamma(k)\}\text{sgn}(k + t - N - 1/2)$. For $\ell \geqslant 1$ define $B_\ell^2 = E\{(\sum_{t=1}^{\ell} Q_t)^2\}$ where $Q_t = \sum_{k=0}^{K_0}\xi_{t,k}$ with $\xi_{t,k} = 2y_{t,k}\gamma(k)$. The following regularity conditions are needed; see the Supplementary Material for a discussion of their validity.

Condition 1. Under H_0, $\max_{1 \leqslant t \leqslant n} E(|Y_t|^{2s_1}) \leqslant c_1$ for two constants $s_1 \in (2, 3]$ and $c_1 > 0$.

Condition 2. Under H_0, $\{Y_t\}$ is α-mixing with $\alpha(\tau) = \sup_t \sup_{A \in \mathcal{F}_{-\infty}^t, B \in \mathcal{F}_{t+\tau}^{\infty}} |\text{pr}(AB) - \text{pr}(A)\text{pr}(B)| \leqslant c_2\tau^{-\beta_1}$ for any $\tau \geqslant 1$, where $\mathcal{F}_{-\infty}^t$ and $\mathcal{F}_{t+\tau}^{\infty}$ denote the σ-fields generated by $\{Y_u\}_{u \leqslant t}$ and $\{Y_u\}_{u \geqslant t+\tau}$, respectively, and $c_2 > 0$ and $\beta_1 > 2(s_1 - 1)s_1/(s_1 - 2)^2$ are two constants, with s_1 as specified in Condition 1.

Condition 3. Under H_0, there is a constant $c_3 > 0$ such that $B_\ell^2 \geqslant c_3\ell$ for any $\ell \geqslant 1$.

THEOREM 1. *Suppose that H_0 holds with Conditions 1–3 satisfied, and let $K_0 = o\{n^{\xi(\beta,s_1)}\}$ with $\xi(\beta, s_1) = \min[(s_1-2)/(4s_1), (\beta-1)(s_1-2)/\{(2\beta+2)s_1\}]$, where s_1 and β_1 are specified, respectively, in Conditions 1 and 2, and $\beta = \beta_1(s_1-2)^2/\{2s_1(s_1-1)\}$. Then, as $n \to \infty$,*

$$\sup_{u>0}\left|\text{pr}\left\{n^{1/2}T_n > u + n^{1/2}\sum_{k=0}^{K_0}|\hat{\gamma}_1(k)|^2\right\} - 1 + \Phi\left(\frac{2Nu}{B_{2N-K_0}n^{1/2}}\right)\right| \to 0.$$

One may select the critical value as $\text{cv}_{\phi,\text{naive}} = z_{1-\phi}\hat{B}_{2N-K_0}/(2N) + \sum_{k=0}^{K_0}|\hat{\gamma}_1(k)|^2$, where $z_{1-\phi}$ is the $(1-\phi)$-quantile of the standard normal distribution $\mathcal{N}(0, 1)$ and $\hat{B}_{2N-K_0}$ is an estimate of B_{2N-K_0} that satisfies the condition $\hat{B}_{2N-K_0}/B_{2N-K_0} \to 1$ in probability under H_0, as then the rejection probability of the test under H_0 converges to ϕ; see § 2.3 and Theorem 3. Unfortunately, $\sum_{k=0}^{K_0}|\hat{\gamma}_1(k)|^2$ diverges to infinity under H_1, and this leads to substantial power loss. To remedy this defect, we apply the truncation idea from Chang et al. (2017, § 2.3). More precisely, we set the critical value as

$$\text{cv}_\phi = \text{cv}_{\phi,\text{naive}}\, I(\mathcal{T}) + \kappa_n I(\mathcal{T}^c), \tag{3}$$

where $\kappa_n = 0.1 \log N$ with $N = \lfloor n/2 \rfloor$ and the event $\mathcal{T}$ satisfies the conditions $\text{pr}_{H_0}(\mathcal{T}) \to 1$ and $\text{pr}_{H_1}(\mathcal{T}^c) \to 1$ as $n \to \infty$. Observe that $\text{pr}_{H_0}(\text{cv}_\phi = \text{cv}_{\phi,\text{naive}}) \to 1$ and $\text{pr}_{H_1}(\text{cv}_\phi = \kappa_n) \to 1$; the former ensures that the rejection probability of the proposed test under H_0 converges to the nominal level ϕ. Proposition 1 shows that $\text{pr}_{H_1}\{|\hat{\gamma}_2(0)|^2 > \kappa_n\} \to 1$. Owing to $T_n \geqslant |\hat{\gamma}_2(0)|^2$, we have $\text{pr}_{H_1}(T_n > \kappa_n) \to 1$, which entails that the proposed test has power 1 asymptotically. We will describe in § 2.2 how to specify a qualified event $\mathcal{T}$.

THEOREM 2. *Let cv_ϕ be defined by (3) with $\mathcal{T}$ satisfying $\text{pr}_{H_0}(\mathcal{T}) \to 1$ and $\text{pr}_{H_1}(\mathcal{T}^c) \to 1$, and suppose that $\hat{B}_{2N-K_0}/B_{2N-K_0} \to 1$ in probability under H_0 as $n \to \infty$. Then* (i) *$\text{pr}_{H_0}(T_n > \text{cv}_\phi) \to \phi$ if the conditions of Theorem 1 hold, and* (ii) *$\text{pr}_{H_1}(T_n > \text{cv}_\phi) \to 1$ if Y_t satisfies (2) with independent $\epsilon_t \sim (0, \sigma_\epsilon^2)$ and $\sum_{j=1}^{\infty} j|\psi_j| < \infty$.*

2.2. *Determining the event $\mathcal{T}$ in (3)*

The critical value cv_ϕ defined in (3) depends on the event $\mathcal{T}$. Let $X_t = \nabla Y_t$ and $\hat{\gamma}_x(k) = (n-1)^{-1}\sum_{t=2}^{n-k}(X_{t+k} - \bar{X})(X_t - \bar{X})$ for $k \geqslant 0$, where $\bar{X} = (n-1)^{-1}\sum_{t=2}^{n} X_t$. To avoid the effect of the innovation variance σ_ϵ^2, we consider the ratio $R = \{\hat{\gamma}(0) + \hat{\gamma}(1)\}/\{\hat{\gamma}_x(0) + \hat{\gamma}_x(1)\}$. Notice that $R = O_\text{p}(1)$ under H_0 and $\text{pr}_{H_1}(R \geqslant C_* N^{3/5}) \to 1$ for any fixed constant $C_* > 0$. We define $\mathcal{T}$ in (3) by

$$\mathcal{T} = \{R < C_* N^{3/5}\}. \tag{4}$$

To use $\mathcal{T}$ with finite samples, C_* must be specified according to the underlying process.

PROPOSITION 2. *Let $Y_t \sim I(1)$ satisfy (2) with independent $\epsilon_t \sim (0, \sigma_\epsilon^2)$ and $\sum_{j=1}^{\infty} j|\psi_j| < \infty$. Write $\eta = \sum_{j=0}^{\infty}\psi_j^2 + \sum_{j=0}^{\infty}\psi_j\psi_{j+1}$. As $n \to \infty$, we have that* (i) *$n^{-1}R \to 2a^2\eta^{-1}\int_0^1 V_0^2(t)\,\text{d}t$ in distribution if $\mu_1 = 0$, where a and $V_0(t)$ are defined in Proposition 1, and* (ii) *$n^{-2}R \to 6^{-1}\sigma_\epsilon^{-2}\eta^{-1}\mu_1^2$ in probability if $\mu_1 \neq 0$.*

Proposition 2 shows that R with $\mu_1 \neq 0$ diverges faster than R with $\mu_1 = 0$. Thus, for any given $C_* > 0$, the requirement $\text{pr}_{H_1}(\mathcal{T}^c) \to 1$ is satisfied more readily with $\mu_1 \neq 0$. Hence, we focus on the cases with $\mu_1 = 0$ only. Recall that $X_t = \nabla Y_t = \mu_1 + \sum_{j=0}^{\infty}\psi_j\epsilon_{t-j}$. Then $a^2\eta^{-1} = \lambda^{-1}(1+\rho)^{-1}$, where $\rho = (\sum_{j=0}^{\infty}\psi_j^2)^{-1}\sum_{j=0}^{\infty}\psi_j\psi_{j+1}$ is the first-order autocorrelation coefficient and $\lambda = \sigma_\text{S}^2/\sigma_\text{L}^2$ with the short-run variance $\sigma_\text{S}^2 = \sigma_\epsilon^2\sum_{j=0}^{\infty}\psi_j^2$ and the long-run variance $\sigma_\text{L}^2 = \sigma_\epsilon^2(\sum_{j=0}^{\infty}\psi_j)^2$. Write $\hat{\sigma}_\text{S}^2 = \hat{\gamma}_x(0)$.

Applying the estimation method for the long-run variance suggested in § 2.3, we can obtain $\hat{\sigma}_{\mathrm{L}}^2$, the kernel-type estimate of σ_{L}^2, based on $\{X_t - \bar{X}\}_{t=2}^n$. Then we can estimate λ and ρ by $\hat{\lambda} = \hat{\sigma}_{\mathrm{S}}^2/\hat{\sigma}_{\mathrm{L}}^2$ and $\hat{\rho} = \hat{\gamma}_x(1)/\hat{\gamma}_x(0)$. As $E\{\int_0^1 V_0^2(t)\,\mathrm{d}t\} = 1/6$, we now specify the model-dependent constant C_* in (4) as

$$C_* = 2c_\kappa/\{\hat{\lambda}(1+\hat{\rho})\} \tag{5}$$

for some constant $c_\kappa > 1/6$. Our extensive simulation results indicate that this specification of C_* with $c_\kappa \in [0.45, 0.65]$ works well across a range of models.

Although the above specification was derived for $Y_t \sim I(1)$, our simulation results indicate that it also works well for $I(2)$ processes. Testing $I(0)$ against $I(d)$ with $d > 1$ is easier than doing so with $d = 1$, as the autocovariances are of order at least n^{2d-1} for $I(d)$ processes; hence the difference between the values of T_n under H_1 and those under H_0 increases as d increases.

2.3. *Estimation of $B_{2N-K_0}^2$*

Write $m = 2N - K_0$. Recall that $Q_t = \sum_{k=0}^{K_0} \xi_{t,k}$ with $\xi_{t,k} = 2y_{t,k}\gamma(k)$. Let V_m be the long-run variance of the sequence $\{Q_t\}_{t=1}^m$. We then have $B_{2N-K_0}^2 = mV_m$. Define $\tilde{Q}_t = \sum_{k=0}^{K_0} \tilde{\xi}_{t,k}$ with $\tilde{\xi}_{t,k} = 2\tilde{y}_{t,k}\hat{\gamma}(k)$, where $\tilde{y}_{t,k} = 2\{(Y_t - \bar{Y})(Y_{t+k} - \bar{Y}) - \hat{\gamma}(k)\}\mathrm{sgn}(k + t - N - 1/2)$. Let $\tilde{G}_j = m^{-1}\sum_{t=j+1}^m \tilde{Q}_t\tilde{Q}_{t-j}$ if $j \geqslant 0$ and $\tilde{G}_j = m^{-1}\sum_{t=-j+1}^m \tilde{Q}_{t+j}\tilde{Q}_t$ otherwise. We can estimate V_m by $\tilde{V}_m = \sum_{j=-m+1}^{m-1} \mathcal{K}(j/b_m)\tilde{G}_j$ with a kernel $\mathcal{K}(\cdot)$ and bandwidth b_m. Let

$$\hat{B}_{2N-K_0} = (m\tilde{V}_m)^{1/2}.$$

Andrews (1991) found that the quadratic spectral kernel is optimal for such an estimation. We suggest using this kernel in practice by calling the function `lrvar` of the `R` package `sandwich` (Zeileis et al., 2021) with the default bandwidth specified in the function. To state the required asymptotic property for $\hat{B}_{2N-K_0}$ with general kernels, we need the following regularity conditions.

Condition 4. The kernel function $\mathcal{K}(\cdot) : \mathbb{R} \to [-1, 1]$ is continuously differentiable on $\mathbb{R}$ and is such that (i) $\mathcal{K}(0) = 1$, (ii) $\mathcal{K}(x) = \mathcal{K}(-x)$ for any $x \in \mathbb{R}$, and (iii) $\int_{-\infty}^{\infty} |\mathcal{K}(x)|\,\mathrm{d}x < \infty$. Let $K_* = K_0 + 2$ satisfy $K_*^{13}\log K_* = o(n^{1-2/s_2})$ with s_2 as specified in Condition 5. The bandwidth $b_m \to \infty$ as $n \to \infty$ satisfies $b_m = o\{n^{1/2-1/s_2}(K_*^5 \log K_*)^{-1/2}\}$ and $K_*^4 = o(b_m)$.

Condition 5. Under H_0, $\max_{1\leqslant t\leqslant n} E(|Y_t|^{2s_2}) \leqslant c_4$ for two constants $s_2 > 4$ and $c_4 > 0$, and the α-mixing coefficients $\{\alpha(\tau)\}_{\tau\geqslant 1}$ satisfy $\alpha(\tau) \leqslant c_5\tau^{-\beta_2}$ for two constants $c_5 > 0$ and $\beta_2 > \max\{2s_2/(s_2-2), s_2/(s_2-4)\}$, where $\alpha(\tau)$ is as defined in Condition 2.

THEOREM 3. *Suppose that Conditions 4 and 5 hold. Then, as $n \to \infty$, $\hat{B}_{2N-K_0}/B_{2N-K_0} \to 1$ in probability under H_0.*

2.4. *Implementation of the test*

Based on § 2.2 and § 2.3, Algorithm 1 outlines the steps of performing our test, which includes two tuning parameters. The algorithm is implemented in an `R` function `ur.test` in the package `HDTSA` (Lin et al., 2021). To perform the test using function `ur.test`, one merely needs to input the time series $\{Y_t\}_{t=1}^n$ and the nominal level ϕ. The package sets the default value $c_\kappa = 0.55$ and returns the five testing results for $K_0 = 0, 1, \ldots, 4$. One can also set (c_κ, K_0) subjectively. We recommend using $c_\kappa \in [0.45, 0.65]$ and $K_0 \in \{0, 1, 2, 3, 4\}$.

To illustrate robustness with respect to the choice of (c_κ, K_0), we apply our test to 14 U.S. annual economic time series (Nelson & Plosser, 1982) that are often used for testing unit roots in the literature. The results with $c_\kappa \in \{0.45, 0.55, 0.65\}$ and $K_0 \in \{0, 1, 2, 3, 4\}$ are exactly the same for each of the 14 time series; see the Supplementary Material for details.

Algorithm 1. Sample autocovariance function-based unit-root test.

Input: Time series $\{Y_t\}_{t=1}^n$, nominal level ϕ, and two (optional) tuning parameters (c_κ, K_0).
Step 1. Compute $\hat{\gamma}(k)$, $\hat{\gamma}_1(k)$, $\hat{\gamma}_2(k)$ and $\hat{\gamma}_x(k)$. Put $\hat{\rho} = \hat{\gamma}_x(1)/\hat{\gamma}_x(0)$.
Step 2. Call function `lrvar` from the `R` package `sandwich`, with the default bandwidth of the function, to compute the long-run variances of $\{\tilde{Q}_t\}$ and $\{X_t\}$, denoted by $\tilde{V}_{2N-K_0}$ and $\hat{\sigma}_{\mathrm{L}}^2$, respectively, where $\tilde{Q}_t$ is defined in § 2.3. Put $\hat{\lambda} = \hat{\gamma}_x(0)/\hat{\sigma}_{\mathrm{L}}^2$.
Step 3. Calculate the test statistic $T_n = \sum_{k=0}^{K_0} |\hat{\gamma}_2(k)|^2$ and the critical value CV_ϕ as in (3) with $\hat{B}_{2N-K_0} = (2N - K_0)^{1/2}\tilde{V}_{2N-K_0}^{1/2}$ and $\mathcal{T}$ given in (4) for C_* specified in (5).
Step 4. Reject H_0 if $T_n > \mathrm{CV}_\phi$.

3. SIMULATION STUDY

We investigate the finite-sample properties of our test T_n by simulation with $K_0 \in \{0, 1, 2, 3, 4\}$ and $c_\kappa \in \{0.45, 0.55, 0.65\}$. We also consider T_n with the untruncated critical value $\mathrm{CV}_{\phi,\mathrm{naive}}$, i.e., $c_\kappa = \infty$ in (5). Hualde & Robinson (2011) proposed the pseudo maximum likelihood estimator $\hat{d}$ for the integration order d in the autoregressive fractionally integrated moving average models that can be used to construct a t-statistic $\hat{d}/\mathrm{sd}(\hat{d})$ for $H_0 : d = 0$ versus $H_1 : d \geqslant 1$. We call the test that rejects H_0 if $\hat{d}/\mathrm{sd}(\hat{d}) > z_{1-\phi}$ the HR test, where $z_{1-\phi}$ is the $(1-\phi)$-quantile of $\mathcal{N}(0, 1)$. For comparison, we include the test of Kwiatkowski et al. (1992) and the HR test in our experiments. We set $N = 40, 70, 100$ and repeat each setting 2000 times. To examine the rejection probability of the tests under H_0, we consider the following three models.

Model 1: $Y_t = \rho Y_{t-1} + \epsilon_t$.
Model 2: $Y_t = \epsilon_t + \phi_1\epsilon_{t-1} + \phi_2\epsilon_{t-2}$.
Model 3: $Y_t - \rho_1 Y_{t-1} - \rho_2 Y_{t-2} = \epsilon_t + 0.5\epsilon_{t-1} + 0.3\epsilon_{t-2}$.

To examine the rejection probability of the tests under H_1, we consider the following four models.

Model 4: $\nabla Y_t = Z_t, Z_t = \rho Z_{t-1} + \epsilon_t$.
Model 5: $\nabla Y_t = Z_t, Z_t = \epsilon_t + \phi_1\epsilon_t + \phi_2\epsilon_{t-1}$.
Model 6: $\nabla Y_t = Z_t, Z_t - \rho_1 Z_{t-1} - \rho_2 Z_{t-2} = \epsilon_t + 0.5\epsilon_t + 0.3\epsilon_{t-1}$.
Model 7: $\nabla^2 Y_t = Z_t, Z_t = \epsilon_t + \phi_1\epsilon_t + \phi_2\epsilon_{t-1}$.

Unless specified otherwise, we always assume that $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$ independently with $\sigma_\epsilon^2 = 1$ or 2 and set the nominal level ϕ to 5%. The results with different (c_κ, K_0) are similar, indicating once again that our test is robust with respect to the choice of (c_κ, K_0). We list the results with $K_0 = 0$ and $\sigma_\epsilon^2 = 1$ in Table 1, and report other results and the $\epsilon_t \sim t(2)$ and $\epsilon_t \sim t(5)$ cases in the Supplementary Material.

Overall the rejection probabilities of our test under H_0 are close to the nominal level $\phi = 5\%$, especially when n is large, such as $N = 100$. The performance of our test is stable across different models with different parameters, different K_0 and different innovation distributions, whereas that of Kwiatkowski et al.'s test and of the HR test vary and are adequate only in some settings. Table 1 indicates that our test works well for Model 1 with both positive and negative ρ, while Kwiatkowski et al.'s test and the HR test perform poorly when $\rho < 0$ and even worse when $\rho > 0$. Kwiatkowski et al.'s test and the HR test completely fail when $\rho = 0.9$, as the rejection probabilities are at least 46.7%. This is due to the fact that when ρ is close to 1, Kwiatkowski et al.'s test and the HR test have difficulties distinguishing ρ from 1, which is a unit root; see also Table 3 of Kwiatkowski et al. (1992). Our test does not suffer from this closeness to 1, as the order of the magnitude of the autocovariance function matters. Our test and that of Kwiatkowski et al. (1992) work well for Model 2, while the HR test is too conservative. For Model 3, the rejection probabilities of our test and the HR test are close to 5%, while Kwiatkowski et al.'s test does not work as its rejection probabilities range from 16.6% to 26.2%. Our test with $c_\kappa = \infty$ has no power, which shows that the truncation step for the critical value in (3) is necessary. The test of Kwiatkowski et al. (1992) has impressive power owing to the fact that it has a tendency to overestimate the rejection probability under H_0, leading to inflated power. Nevertheless, our test exhibits greater power in most cases. The HR test has good power for Models 4 and 5, but performs poorly for Model 6. The power-one property of our test is observable in the simulation since the rejection probability tends to 1 as N increases. Comparing the results of Models 5 and 7, we find that

Table 1. *Rejection probabilities (%) of the proposed test T_n with $K_0 = 0$ and $c_\kappa = 0.45, 0.55, 0.65, \infty$, the test of Kwiatkowski et al. (1992), and the* HR *test; the nominal level is* 5%

		Model 1								Model 4					
ρ	N	∞	0.45	0.55	0.65	KPSS	HR	ρ	N	∞	0.45	0.55	0.65	KPSS	HR
0.5	40	6.0	6.0	6.0	6.0	10.4	5.7	0.5	40	11.7	94.2	88.4	84.0	84.2	96.4
	70	6.9	6.9	6.9	6.9	10.1	7.0		70	11.7	96.5	92.9	88.4	90.9	99.8
	100	6.1	6.1	6.1	6.1	10.2	8.4		100	11.3	98.0	95.5	92.2	95.5	100.0
0.9	40	7.2	41.9	30.0	20.3	51.2	46.8	0.9	40	13.1	99.2	97.3	94.6	91.1	98.9
	70	7.8	23.7	14.6	10.4	46.7	58.8		70	14.8	99.8	99.1	97.9	95.3	100.0
	100	8.5	12.7	9.4	8.6	49.2	61.1		100	16.4	99.9	99.5	99.1	97.2	100.0
−0.5	40	7.4	7.4	7.4	7.4	1.8	0.1	−0.5	40	5.6	82.2	75.1	67.6	81.5	99.7
	70	6.9	6.9	6.9	6.9	2.5	0.2		70	6.3	92.1	86.1	80.0	90.1	100.0
	100	6.4	6.4	6.4	6.4	1.8	0.3		100	5.8	94.2	89.5	85.2	94.5	100.0
		Model 2								Model 5					
(ϕ_1, ϕ_2)	N	∞	0.45	0.55	0.65	KPSS	HR	(ϕ_1, ϕ_2)	N	∞	0.45	0.55	0.65	KPSS	HR
(0.8, 0.3)	40	6.2	6.2	6.2	6.2	7.6	0.9	(0.8, 0.3)	40	11.8	94.3	88.8	82.3	82.0	99.4
	70	6.4	6.4	6.4	6.4	6.2	0.4		70	11.8	96.6	92.7	88.3	90.1	100.0
	100	7.2	7.2	7.2	7.2	7.0	0.4		100	12.1	98.4	95.4	91.8	95.3	100.0
(0.9, 0.5)	40	6.7	6.7	6.7	6.7	8.5	0.4	(0.9, 0.5)	40	11.8	95.3	90.0	84.2	83.5	99.8
	70	6.5	6.5	6.5	6.5	8.1	0.0		70	12.2	97.2	93.8	89.8	89.2	100.0
	100	5.6	5.6	5.6	5.6	7.4	0.0		100	11.6	98.6	96.4	92.7	94.8	100.0
(0.95, 0.9)	40	7.2	7.2	7.2	7.2	9.0	0.0	(0.95, 0.9)	40	13.1	95.0	90.0	83.9	83.0	99.6
	70	7.1	7.1	7.1	7.1	7.3	0.2		70	11.6	97.3	93.8	89.7	90.2	100.0
	100	5.5	5.5	5.5	5.5	8.1	0.0		100	13.7	99.0	96.4	92.3	95.2	100.0
		Model 3								Model 6					
(ρ_1, ρ_2)	N	∞	0.45	0.55	0.65	KPSS	HR	(ρ_1, ρ_2)	N	∞	0.45	0.55	0.65	KPSS	HR
(0.4, 0.2)	40	7.2	8.2	7.4	7.3	22.5	4.8	(0.4, 0.2)	40	14.8	98.0	95.2	90.6	85.9	29.2
	70	7.7	7.7	7.7	7.7	17.3	5.0		70	15.4	99.1	97.0	93.8	92.0	43.4
	100	7.2	7.2	7.2	7.2	18.0	5.1		100	16.6	99.6	98.8	96.5	96.5	54.5
(0.5, 0.1)	40	8.5	8.9	8.5	8.5	19.6	5.4	(0.5, 0.1)	40	14.2	99.1	95.9	91.3	84.7	30.2
	70	8.0	8.0	8.0	8.0	16.6	6.2		70	14.8	99.4	97.2	94.0	91.2	47.8
	100	6.3	6.3	6.3	6.3	17.4	5.9		100	15.0	99.6	98.5	96.2	95.5	60.9
(0.6, 0.1)	40	8.5	12.7	9.6	8.7	26.2	6.0	(0.6, 0.1)	40	14.5	99.2	97.1	93.3	87.2	27.6
	70	7.3	7.3	7.3	7.3	22.4	6.8		70	15.7	99.7	98.5	96.2	93.5	37.3
	100	7.6	7.6	7.6	7.6	20.3	7.0		100	16.4	99.8	99.1	97.7	95.7	44.0
		Model 7								Model 7					
(ϕ_1, ϕ_2)	N	∞	0.45	0.55	0.65	KPSS	HR	(ϕ_1, ϕ_2)	N	∞	0.45	0.55	0.65	KPSS	HR
(0.8, 0.3)	40	6.7	100.0	100.0	99.9	98.5	100.0	(0.9, 0.5)	40	7.0	100.0	100.0	100.0	98.4	100.0
	70	6.3	100.0	100.0	100.0	99.7	100.0		70	5.5	100.0	100.0	100.0	99.5	100.0
	100	7.0	100.0	100.0	100.0	99.8	100.0		100	5.9	100.0	100.0	100.0	99.9	100.0
(0.95, 0.9)	40	8.0	100.0	100.0	100.0	98.5	100.0								
	70	7.3	100.0	100.0	100.0	99.2	100.0								
	100	6.1	100.0	100.0	100.0	99.9	100.0								

KPSS, the test of Kwiatkowski et al. (1992); HR, the test that rejects H_0 if $\hat{d}/\mathrm{sd}(\hat{d}) > z_{1-\phi}$ where $z_{1-\phi}$ is the $(1-\phi)$-quantile of $\mathcal{N}(0, 1)$.

our test displays the power-one property more distinctly as our test statistic has more discriminative power between $I(2)$ and $I(0)$ than between $I(1)$ and $I(0)$.

Acknowledgement

All authors contributed equally to the paper. We thank the editor, associate editor and referees for their constructive comments. Chang and Cheng were supported by the National Natural Science Foundation of

China. Chang was also supported by the Center of Statistical Research and the Joint Lab of Data Science and Business Intelligence at Southwestern University of Finance and Economics. Yao was supported in part by the U.K. Engineering and Physical Sciences Research Council.

Supplementary material

Supplementary Material available at *Biometrika* online includes all the technical proofs and some additional numerical results.

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817–58.

Cavaliere, G. & Taylor, A. M. R. (2007). Testing for unit roots in time series models with non-stationary volatility. *J. Economet.* **140**, 919–47.

Chan, N. H. & Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* **16**, 367–401.

Chang, J., Zheng, C., Zhou, W.-X. & Zhou, W. (2017). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics* **73**, 1300–10.

Dickey, D. A. & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Statist. Assoc.* **74**, 427–31.

Dickey, D. A. & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* **49**, 1057–72.

Elliott, G., Rothenberg, T. J. & Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica* **64**, 813–36.

Hatanaka, M. (1996). *Time-Series-Based Econometrics: Unit Roots and Cointegration*. Oxford: Oxford University Press.

Hualde, J. & Robinson, P. M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Ann. Statist.* **39**, 3152–81.

Hylleberg, S., Engle, R. F., Granger, C. F. J. & Yoo, S. (1990). Seasonal integration and cointegration. *J. Economet.* **44**, 215–38.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *J. Economet.* **54**, 159–78.

Lin, C., Cheng, G., Chang, J. & Yao, Q. (2021). *HDTSA: High Dimensional Time Series Analysis Tools*. R package version 1.0.0, available at https://cran.r-project.org/web/packages/HDTSA/.

Maddala, G. S. & Kim, I.-M. (1998). *Unit roots, Cointegration and Structural Change*. Cambridge: Cambridge University Press.

Nelson, C. R. & Plosser, C. I. (1982). Trends versus random walks in macroeconomic time series: Some evidence and implications. *J. Monet. Econ.* **10**, 139–62.

Paparoditis, E. & Politis, D. N. (2005). Bootstrapping unit root tests for autoregressive time series. *J. Am. Statist. Assoc.* **100**, 545–53.

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Appl. Economet.* **22**, 265–317.

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

Phillips, P. C. B. & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, **75**, 335–46.

Phillips, P. C. B. & Xiao, Z. (1998). A primer on unit root testing. *J. Econ. Surv.* **12**, 423–69.

R Development Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org.

Rho, R. & Shao, X. (2019). Bootstrap-assisted unit root testing with piecewise locally stationary errors. *Economet. Theory* **35**, 143–66.

Robinson, P. M. (1994). Efficient tests of nonstationary hypothesis. *J. Am. Statist. Assoc.* **89**, 1420–37.

Said, S. E. & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* **71**, 599–608.

Stock, J. H. (1994). Unit roots, structural breaks and trends. In *Handbook of Econometrics*, vol. 4, R. F. Engle and D. L. McFadden (eds). vol. 4. Amsterdam: Elsevier.

Zeileis, A., Lumley, T., Graham, N. & Koell, S. (2021). *sandwich: Robust Covariance Matrix Estimators*. R package version 3.0-1, available at https://sandwich.R-Forge.R-project.org/.

Zivot, E. & Andrews, D. W. K. (1992). Further evidence on the great crash, the oil price shock, and the unit root hypothesis. *J. Bus. Econ. Statist.* **10**, 251–70.

[*Received on* 26 *August* 2020. *Editorial decision on* 10 *June* 2021]