Statistics and Its Interface Volume 16 (2023) 199–216

Least absolute deviations estimation for nonstationary vector autoregressive time series models with pure unit roots

Yao Zheng*, Jianhong Wu, Wai Keung Li, and Guodong Li

This paper derives the asymptotic distribution of the least absolute deviations estimator for nonstationary vector autoregressive time series models with pure unit roots under mild conditions. As this distribution has a complicated form, many commonly used bootstrap techniques cannot be directly applied. To tackle this problem, we propose a novel hybrid bootstrap method by combining the classical wild bootstrap and the method in [17]. We establish the asymptotic validity of the proposed method and further apply it to construct three bootstrapping panel unit root tests. Monte Carlo experiments support the validity of our inference procedure in finite samples. The usefulness of the proposed panel unit root tests is demonstrated via analyses of real economic and financial data sets.

AMS 2000 subject classifications: Primary 62F40, 62H15, 91B84; secondary 62F35.
Keywords and phrases: Bootstrap, Least absolute deviation, Panel unit root test, Vector autoregression.

1. INTRODUCTION

[20, 21] first considered tests for unit root hypothesis in panels. One of their motivations is to increase the power since unit root tests in univariate time series usually lack power in small samples. As these tests assume independence along the cross-sectional dimension, they belong to the so-called first generation tests. However, when cross-sectional dependence is present, these tests usually exhibit large size distortions [28]. As a result, many second generation panel unit root tests have recently been proposed to handle different kinds of cross-sectional dependence. [7] discussed the interdependence of individual errors and considered the generalized least squares estimation in constructing the panel unit root tests. [2] introduced spatial dependence into panel unit root tests, and [1] considered common factors for the cross-sectional dependence when the number of individuals is large; see also, e.g., [34], [6] and [30]. As one of many ways to introduce cross-sectional dependence, the general nonstationary vector autoregressive (AR) time series model with pure unit roots is considered in this paper.

*Corresponding author.

There is plenty of empirical evidence that financial time series are usually heavy-tailed [27, 26]. The least absolute deviations (LAD) estimation has many advantages over the ordinary least squares (OLS) approach. It requires less stringent moment conditions and is robust to heavy-tailed data [32, 22, 36]. [16] first studied the LAD-based unit root tests for univariate time series, and [23] discussed the LAD estimation for unit root AR processes with generalized autoregressive conditional heteroscedastic (GARCH) errors. This paper considers the LAD estimation for nonstationary vector AR models with pure unit roots. The technical conditions on the error term are weaker than those commonly used in two aspects. First, it permits a constant, possibly nonzero, conditional median of the error term. While the error term of nonstationary time series models with unit roots necessarily has a zero mean, the LAD technique additionally requires a zero median for it. Fortunately, owing to a redundant parameter in the vector AR model discussed in this paper, we can avoid the restrictive assumption that both the mean and the median are zero. Second, we allow the error term to be conditional heteroscedastic, which is an important feature of financial time series; see [11] and [5]. These two relaxations introduce no more difficulties into the technical proofs, but will affect the subsequent bootstrap mechanism.

In contrast to the OLS method, the LAD estimator in this paper has a complicated asymptotic distribution, which can be difficult to approximate even in simple cases; see Section 2 for more details. Therefore, it is naturally desirable to use the bootstrap method; see [10]. The most commonly used bootstrap method for unit root tests is to resample the residuals [7, 31], but it may not work well in the presence of conditional heteroscedasticity since the time order will be destroyed by the resampling operation. [30] considered a block-wise bootstrap for the panel unit root tests, yet it may be challenging to select the block size. [29] showed by simulation that the block-wise bootstrap does not seem to be particularly powerful in unit root tests for univariate time series. Although Wu's [1986] wild bootstrap can be an appealing method since the time order is kept, it may not be good enough for the LAD estimator and cannot be directly applied to the asymptotic distribution in this paper; see Section 3 for more details. [17] proposed a bootstrap method (henceforth called the JYW method) by perturbing the minimand of the objective function, which was applied to the LAD-based ANOVA analysis by [9]. [8] extended this method to the estimating equations when the objective function is derivable. Note that there is no need to generate the bootstrapped sequences in this method, so it is especially attractive when the time series or error sequence has too many constraints to be bootstrapped; see also [37]. A drawback of this method is that it may not be able to approximate distributions more complicated than normality; see Corollary 3.1 in Section 3 for more details. This may be the reason why the JYW method has not attracted much attention in this area. To approximate the asymptotic distribution of the LAD estimator for nonstationary vector AR models with pure unit roots, in Section 3 we propose a novel bootstrap method, which is a hybrid version of the wild bootstrap and the JYW method.

There are two sources of cross-sectional dependence in vector AR models: structural dependence and interdependence of individual errors. In Section 4, we consider the interdependence of individual errors in nonstationary vector AR models with pure unit roots as in [7], and we then propose three bootstrapping panel unit root tests by applying the methods in Sections 2 and 3. Monte Carlo simulations and real data applications are presented in Sections 5 and 6, respectively. Section 7 concludes with a brief discussion. All technical proofs are relegated to the appendix. Throughout the paper, $o_p(1)$ denotes a series of random variables (vectors) converging to zero in probability, $O_p(1)$ denotes a series of random variables (vectors) that are bounded in probability, $D = D[0,1]$ denotes the space of functions on $[0,1]$, which is defined and equipped with the Shorokhod topology [4], and $\Rightarrow$ denotes weak convergence on D. In addition, the notations $o_p^*(1)$ and $O_p^*(1)$ are used for the bootstrapped space.

2. LEAST ABSOLUTE DEVIATIONS ESTIMATION

The vector AR model can be equivalently written as the following error correction model,

$$(1)\quad \Delta\mathbf{y}_t = \boldsymbol{\mu} + \Phi\mathbf{y}_{t-1} + \sum_{j=1}^{p}\Psi_j\Delta\mathbf{y}_{t-j} + \mathbf{e}_t, \qquad 1 \le t \le T,$$

where $\mathbf{y}_t = (y_{1,t}, ..., y_{N,t})^{\mathrm{T}}$, $\mathbf{e}_t = (e_{1,t}, ..., e_{N,t})^{\mathrm{T}}$, $\Delta\mathbf{y}_t = \mathbf{y}_t - \mathbf{y}_{t-1}$, $\boldsymbol{\mu}$ is an N-dimensional vector of parameters, Φ and Ψ_j are $N \times N$ matrices of parameters, and $\mathbf{e}_t$ is the error term with $E(\mathbf{e}_t) = 0$; see Chapter 18 of [15].

Assumption 2.1. *All the roots of the polynomial* $det(\mathbf{I}_N - \sum_{j=1}^{p}\Psi_j z^j) = 0$ *lie outside the unit circle, where* $\mathbf{I}_N$ *is the* N*-dimensional identity matrix and* $det(A)$ *is the determinant of the matrix* A.

Assumption 2.1 is necessary for $\{\mathbf{y}_t\}$ to be either stationary or stationary after differencing, i.e., $I(0)$ or $I(1)$. Let $r = rank(\Phi)$ be the rank of the matrix Φ. Then $\{\mathbf{y}_t\}$ is cointegrated with r linearly independent cointegrating vectors if $0 < r < N$, whereas it is not cointegrated but has pure unit roots if $r = 0$ (i.e., $\Phi = 0$); see [12]. Note that this paper is focused on statistical inference for the latter case, i.e., $\{\mathbf{y}_t\}$ has pure unit roots.

From model (1), when $\Phi = 0$, we have $\mathbf{y}_t = (1 - \sum_{j=1}^{p}\Psi_j)^{-1}\boldsymbol{\mu}\cdot t + \sum_{j=1}^{t}\mathbf{u}_t + \mathbf{y}_0$, where $\mathbf{u}_t = (\mathbf{I}_N - \sum_{j=1}^{p}\Psi_j B^j)^{-1}\mathbf{e}_t$ is a stationary N-dimensional time series; see the proof of Lemma A.1. This shows that when $\boldsymbol{\mu} \neq 0$, $\{\mathbf{y}_t\}$ has an extra linear trend. Note that asymptotic theory derived under a linear trend will be very different from those without it; see [33]. Throughout this paper, we consider $\{\mathbf{y}_t\}$ without a linear trend. Thus, concisely, we assume that $\{\mathbf{y}_t\}$ follows model (1) with $\boldsymbol{\mu} = 0$ and $\Phi = 0$.

Let $\boldsymbol{\mu} = (\mu_1, ..., \mu_N)^{\mathrm{T}}$, $\Phi = (\phi_1, ..., \phi_N)^{\mathrm{T}}$, and $\Psi_j = (\Psi_{1,j}, ..., \Psi_{N,j})^{\mathrm{T}}$ for $1 \le j \le p$. Denote the parameter vector of model (1) by $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^{\mathrm{T}}, ..., \boldsymbol{\theta}_N^{\mathrm{T}})^{\mathrm{T}}$, where $\boldsymbol{\theta}_i = (\phi_i^{\mathrm{T}}, \mu_i, \psi_i^{\mathrm{T}})^{\mathrm{T}}$ and $\psi_i = (\Psi_{i,1}^{\mathrm{T}}, ..., \Psi_{i,p}^{\mathrm{T}})^{\mathrm{T}}$ for $1 \le i \le N$. These notations are used for both the parameters and their true values below without confusion. For $\{\mathbf{y}_t\}$ generated by model (1) with $\boldsymbol{\mu} = 0$, $\Phi = 0$, and initial value $\mathbf{y}_0 = 0$, we consider the LAD estimation,

$$\begin{aligned}\widehat{\boldsymbol{\theta}} &= (\widehat{\boldsymbol{\theta}}_1^{\mathrm{T}}, ..., \widehat{\boldsymbol{\theta}}_N^{\mathrm{T}})^{\mathrm{T}} \\ &= \operatorname{argmin}\sum_{t=p+2}^{T}\|\Delta\mathbf{y}_t - \boldsymbol{\mu} - \Phi\mathbf{y}_{t-1} - \sum_{j=1}^{p}\Psi_j\Delta\mathbf{y}_{t-j}\|_1,\end{aligned}$$

where $\|x\|_1 = \sum_{i=1}^{N}|x_i|$ for an N-dimensional vector $x = (x_1, ..., x_N)^{\mathrm{T}}$. Equivalently, we can perform the above optimization separately for each $1 \le i \le N$, i.e.,

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_i &= (\widehat{\phi}_i^{\mathrm{T}}, \widehat{\mu}_i, \widehat{\psi}_i^{\mathrm{T}})^{\mathrm{T}} \\ &= \operatorname{argmin}\sum_{t=p+2}^{T}|\Delta y_{i,t} - \mu_i - \phi_i^{\mathrm{T}}\mathbf{y}_{t-1} - \sum_{j=1}^{p}\Psi_{i,j}^{\mathrm{T}}\Delta\mathbf{y}_{t-j}|.\end{aligned}$$

While these estimators all depend on the sample size T, we suppress the subscript T for simplicity.

Assumption 2.2. *The error sequence* $\{\mathbf{e}_t\}$ *is a strictly stationary and ergodic martingale difference sequence with* $E(\mathbf{e}_t) = 0$, $E(\mathbf{e}_t\mathbf{e}_t^{\mathrm{T}}) = \Sigma_{\mathbf{e}} < \infty$, *and* $\max_{1\le j\le N} E|e_{j,t}|^{2+\delta} < \infty$ *for a* $\delta > 0$.

Apparently, the parameter vector $\boldsymbol{\mu}$ is redundant from the viewpoint of estimation. However, it is used to absorb the conditional median of the error term $\mathbf{e}_t$. This allows us to avoid the stringent condition that both the mean and the median of $\mathbf{e}_t$ are zero. Let the conditional median of $\mathbf{e}_t$ be $\mathbf{m} = (m_1, ..., m_N)^{\mathrm{T}}$, i.e., $P(e_{i,t} > m_i|\mathcal{F}_{t-1}) = 0.5$ for $1 \le i \le N$, where $\mathcal{F}_t$ is the σ-field generated by $\{\mathbf{e}_t, \mathbf{e}_{t-1}, ...\}$. Let $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, ..., \varepsilon_{N,t})^{\mathrm{T}} = \mathbf{e}_t - \mathbf{m}$, and let $f_{i,t}(x)$, $1 \le i \le N$, be the density of $\varepsilon_{i,t}$ conditional on $\mathcal{F}_{t-1}$. The following condition is needed for the LAD technique.

Assumption 2.3. *For all $1 \leq i \leq N$, there exists a small value $\pi > 0$ such that $f_{i,t}(x)$ is bounded and continuous on $(-\pi, \pi)$ with probability one, uniformly for all t.*

Assumptions 2.2 and 2.3 include two important cases. The first one is that $\{\mathbf{e}_t\}$ are independent and identically distributed (*i.i.d.*) with continuous marginal densities at the median. The second is that $\{\mathbf{e}_t\}$ follow a multivariate GARCH model such that $\mathbf{e}_t$ conditional on $\mathcal{F}_{t-1}$ has a normal distribution with mean zero and time-varying variance $\boldsymbol{\sigma}_t^2$; for the forms of $\boldsymbol{\sigma}_t^2$ under different multivariate GARCH models, see Chapter 11 of [13]. As mentioned in Section 1, the commonly used bootstrap techniques may not work well for both cases; see also Corollary 3.1 in Section 3.

Let $\Sigma_{\mathbf{e}} = E(\mathbf{e}_t\mathbf{e}_t^{\mathrm{T}})$, $\Sigma_{\mathbf{s}} = E[\mathrm{sgn}(\boldsymbol{\varepsilon}_t)\,\mathrm{sgn}(\boldsymbol{\varepsilon}_t)^{\mathrm{T}}]$, $\mathbf{z}_t = (\Delta\mathbf{y}_{t-1}^{\mathrm{T}}, ..., \Delta\mathbf{y}_{t-p}^{\mathrm{T}})^{\mathrm{T}}$, $\bar{\mathbf{z}}_t = (1, \mathbf{z}_t^{\mathrm{T}})^{\mathrm{T}}$, $\Sigma_{\mathbf{z}} = E(\mathbf{z}_t\mathbf{z}_t^{\mathrm{T}})$, $\Sigma_{\varepsilon,\mathbf{ez}} = E[\mathrm{sgn}(\boldsymbol{\varepsilon}_t)^{\mathrm{T}} \otimes (\mathbf{e}_t\mathbf{z}_t^{\mathrm{T}})]$, and

$$\Omega = \begin{pmatrix} \Sigma_{\mathbf{e}} & E[\mathbf{e}_t\,\mathrm{sgn}(\boldsymbol{\varepsilon}_t)^{\mathrm{T}}] & \Sigma_{\varepsilon,\mathbf{ez}} \\ E[\mathrm{sgn}(\boldsymbol{\varepsilon}_t)\mathbf{e}_t^{\mathrm{T}}] & \Sigma_{\mathbf{s}} & 0 \\ \Sigma_{\varepsilon,\mathbf{ez}}^{\mathrm{T}} & 0 & \Sigma_{\mathbf{s}} \otimes \Sigma_{\mathbf{z}} \end{pmatrix},$$

where 0 is a zero matrix, and $\otimes$ represents the Kronecker product. It can be shown that, under Assumptions 2.1 and 2.2, the matrices, $\Sigma_{\mathbf{e}}$, $\Sigma_{\mathbf{s}}$, $\Sigma_{\mathbf{z}}$ and Ω, are all positive definite. Let $\mathbf{B}(\tau) = [\mathbf{B}_1^{\mathrm{T}}(\tau), ..., \mathbf{B}_{N+2}^{\mathrm{T}}(\tau)]^{\mathrm{T}}$ be a $(pN^2 + 2N)$-dimensional Brownian motion with covariance matrix $\tau\Omega$, where the first two components are N-dimensional, and the other N components are (pN)-dimensional.

Theorem 2.1. *For all $1 \leq i \leq N$, under Assumptions 2.1–2.3, it holds that*

$$\begin{pmatrix} T\widehat{\phi}_i \\ \sqrt{T}(\widehat{\mu}_i - m_i) \\ \sqrt{T}(\widehat{\psi}_i - \psi_i) \end{pmatrix} \Rightarrow 0.5\Gamma_i^{-1} \begin{pmatrix} \Upsilon g(\mathbf{B}) \\ \mathbf{B}_2^{(i)}(1) \\ \mathbf{B}_{i+2}(1) \end{pmatrix},$$

where ψ_i is the corresponding true parameter vector, $\Upsilon = (\mathbf{I}_N - \Psi_1 - \cdots - \Psi_p)^{-1}$, $g(\mathbf{B}) = \int_0^1 \mathbf{B}_1(\tau)d\mathbf{B}_2^{(i)}(\tau)$, $\mathbf{B}_2^{(i)}(\tau)$ is the ith element of $\mathbf{B}_2(\tau)$, $\Gamma_{11,i} = E[f_{i,t}(0)]\Upsilon\int_0^1 \mathbf{B}_1(\tau)\mathbf{B}_1^{\mathrm{T}}(\tau)d\tau\Upsilon^{\mathrm{T}}$, $\Gamma_{12,i} = \Upsilon\int_0^1 \mathbf{B}_1(\tau)d\tau E[f_{i,t}(0)\bar{\mathbf{z}}_t^{\mathrm{T}}]$, and

$$\Gamma_i = \begin{pmatrix} \Gamma_{11,i} & \Gamma_{12,i} \\ \Gamma_{12,i}^{\mathrm{T}} & E[f_{i,t}(0)\bar{\mathbf{z}}_t\bar{\mathbf{z}}_t^{\mathrm{T}}] \end{pmatrix}.$$

Let $\widehat{\Phi}$, $\widehat{\boldsymbol{\mu}}$ and $\widehat{\Psi}_j$, $1 \leq j \leq p$, be the corresponding components of the estimator $\widehat{\boldsymbol{\theta}}$. The above theorem implies that the estimator $\widehat{\boldsymbol{\mu}}$ actually provides information about the median $\mathbf{m}$. The asymptotic distributions of the $\widehat{\boldsymbol{\theta}}_i$'s may depend on each other, although they can be estimated separately.

Similar to univariate AR processes with unit roots, we may also consider the case without the intercept (i.e., the parameter vector $\boldsymbol{\mu}$ in model (1) is suppressed) and the case with a linear trend (i.e., there is one more term $\mathbf{c} \cdot t$ with $\mathbf{c} = (c_1, ..., c_N)^{\mathrm{T}}$). It can be shown that the estimators will be biased for the first case, unless both the mean and the median of $\mathbf{e}_t$ are zero. The asymptotic results for the second case can be derived similarly and hence are omitted here.

If the errors $\{\mathbf{e}_t\}$ are *i.i.d.*, then each density $f_{i,t}(0)$ will be a constant independent of t, and we may denote it by $f_i(0)$. As a result,

$$T\widehat{\Phi} \Rightarrow 0.5\mathbf{f}^{-1}\Lambda_{\mathbf{B}}^{\mathrm{T}}\Upsilon^{-1},$$

$$\sqrt{T}(\widehat{\boldsymbol{\mu}} - \mathbf{m}) \Rightarrow 0.5\mathbf{f}^{-1}\left[\mathbf{B}_2(1) - \Lambda_{\mathbf{B}}^{\mathrm{T}}\int_0^1 \mathbf{B}_1(\tau)d\tau\right],$$

and

$$\sqrt{T}\{\mathrm{vec}[(\widehat{\Psi}_1, ..., \widehat{\Psi}_p)^{\mathrm{T}}] - \mathrm{vec}[(\Psi_1, ..., \Psi_p)^{\mathrm{T}}]\} \Rightarrow N\{0, 0.25(\mathbf{f}^{-1} \otimes \Sigma_{\mathbf{z}}^{-1})(\Sigma_{\mathbf{s}} \otimes \Sigma_{\mathbf{z}})(\mathbf{f}^{-1} \otimes \Sigma_{\mathbf{z}}^{-1})\},$$

where

$$\Lambda_{\mathbf{B}} = \left[\int_0^1 \mathbf{B}_1(\tau)\mathbf{B}_1^{\mathrm{T}}(\tau)d\tau - \int_0^1 \mathbf{B}_1(\tau)d\tau\int_0^1 \mathbf{B}_1^{\mathrm{T}}(\tau)d\tau\right]^{-1} \left[\int_0^1 \mathbf{B}_1(\tau)d\mathbf{B}_2^{\mathrm{T}}(\tau) - \int_0^1 \mathbf{B}_1(\tau)d\tau\mathbf{B}_2^{\mathrm{T}}(1)\right]$$

is a random matrix, $\mathbf{f} = \mathrm{diag}\{f_1(0), ..., f_N(0)\}$ is an $N \times N$ diagonal matrix, and $\mathrm{vec}(A)$ is the vectorization of the matrix A. The first two asymptotic distributions above are complicated, and to approximate the asymptotic distribution in Theorem 2.1, bootstrapping techniques may be desirable.

To obtain the estimates $\widehat{\boldsymbol{\theta}}$, we can rewrite model (1) into a linear regression form, and then use any standard program, such as *lad* in MATLAB, *rq* in the QUANTREG package of R, etc., for the optimization.

3. HYBRID BOOTSTRAP APPROXIMATION

To approximate the asymptotic distributions in the previous section, we introduce a hybrid bootstrap method which combines the wild bootstrap and the JYW method.

The first step is to construct the bootstrapped sequence as in the wild bootstrap. Let $\{\widehat{\mathbf{e}}_1, \ldots, \widehat{\mathbf{e}}_T\}$ be the residual sequence of model (1) fitted by the LAD method, i.e.,

$$\widehat{\mathbf{e}}_t = \Delta\mathbf{y}_t - \widehat{\Phi}\mathbf{y}_{t-1} - \sum_{j=1}^{p}\widehat{\Psi}_j\Delta\mathbf{y}_{t-j}$$

for $p + 2 \leq t \leq T$, and $\widehat{\mathbf{e}}_t = 0$ for $1 \leq t \leq p + 1$, where $\widehat{\Phi}$ and $\widehat{\Psi}_j$ are the components of $\widehat{\boldsymbol{\theta}}$ defined in the previous section. Note that $\widehat{\boldsymbol{\mu}}$ is not involved in the calculation of the residuals since it is an estimator of the median $\mathbf{m}$. Let $\{\omega_t\}$ be a sequence of *i.i.d.* positive random variables with $E\omega_t^{2+\delta} < \infty$ for a $\delta > 0$ and both mean and variance equal to one. Then the new residual sequence $\{\mathbf{e}_t^*\}$ can be generated by $\mathbf{e}_t^* = (\omega_t - 1)\widehat{\mathbf{e}}_t$, $1 \le t \le T$. Construct the bootstrapped sequence by

$$\mathbf{y}_t^* = \mathbf{y}_{t-1}^* + \sum_{j=1}^{p} \widehat{\Psi}_j \Delta \mathbf{y}_{t-j}^* + \mathbf{e}_t^*, \tag{2}$$

where the initial values of $\mathbf{y}_t^*$ for $1 \le t \le p+1$ can be set to zero. By Theorem 18.2 in [4] and the Beveridge-Nelson representation in the proof of Lemma A.1, we can show that, conditional on $\mathbf{y}_1, \ldots, \mathbf{y}_T$,

$$\frac{1}{\sqrt{T}} \mathbf{y}_{[T\tau]}^* = \frac{1}{\sqrt{T}} \cdot \widehat{\Upsilon} \sum_{t=1}^{[T\tau]} \mathbf{e}_t^* + R_T(\tau) \Rightarrow \Upsilon \mathbf{B}_1^*(\tau), \tag{3}$$

where $[T\tau]$ is the integer part of $T\tau$, $\widehat{\Upsilon} = (\mathbf{I} - \widehat{\Psi}_1 - \cdots - \widehat{\Psi}_p)^{-1}$, the remainder term $R_T(\tau)$ satisfies $\sup_{0\le\tau\le 1} \|R_T(\tau)\|_1 = o_p^*(1)$, and $\mathbf{B}_1^*(\tau)$ is an N-dimensional Brownian motion with covariance matrix $\tau\Sigma_{\mathbf{e}}$. Note that $\mathbf{B}_1^*(\tau)$ has the same covariance matrix as $\mathbf{B}_1(\tau)$.

The second step is to approximate the asymptotic distribution in Theorem 2.1 by the JYW method. Consider two auxiliary estimators as follows,

$$\widehat{\boldsymbol{\theta}}_1^* = \operatorname{argmin} \sum_{t=p+2}^{T} \|\Delta \mathbf{y}_t - \boldsymbol{\mu} - \Phi \mathbf{y}_{t-1}^* - \sum_{i=1}^{p} \Psi_i \Delta \mathbf{y}_{t-i}\|_1$$

and

$$\widehat{\boldsymbol{\theta}}_2^* = \operatorname{argmin} \sum_{t=p+2}^{T} \omega_t \|\Delta \mathbf{y}_t - \boldsymbol{\mu} - \Phi \mathbf{y}_{t-1}^* - \sum_{i=1}^{p} \Psi_i \Delta \mathbf{y}_{t-i}\|_1,$$

where $\widehat{\boldsymbol{\theta}}_j^* = (\widehat{\boldsymbol{\theta}}_{1,j}^{*\mathrm{T}}, \ldots, \widehat{\boldsymbol{\theta}}_{N,j}^{*\mathrm{T}})^{\mathrm{T}}$ and $\widehat{\boldsymbol{\theta}}_{i,j}^* = (\widehat{\phi}_{i,j}^{*\mathrm{T}}, \widehat{\mu}_{i,j}^*, \widehat{\psi}_{i,j}^{*\mathrm{T}})^{\mathrm{T}}$ with $j = 1$ and 2. The next theorem shows that the quantity $\widehat{\boldsymbol{\theta}}_2^* - \widehat{\boldsymbol{\theta}}_1^*$ can be used to approximate the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$.

Let $\mathbf{B}^*(\tau) = [\mathbf{B}_1^{*\mathrm{T}}(\tau), \ldots, \mathbf{B}_{N+2}^{*\mathrm{T}}(\tau)]^{\mathrm{T}}$ be a $(pN^2 + 2N)$-dimensional Brownian motion with covariance matrix $\tau\Omega$, where the first two components are N-dimensional, and the other N components are (pN)-dimensional.

Theorem 3.1. *For $1 \le i \le N$, under Assumptions 2.1–2.3, it holds that, conditional on $\mathbf{y}_1, \ldots, \mathbf{y}_T$,*

$$\begin{pmatrix} T(\widehat{\phi}_{i,2}^* - \widehat{\phi}_{i,1}^*) \\ \sqrt{T}(\widehat{\mu}_{i,2}^* - \widehat{\mu}_{i,1}^*) \\ \sqrt{T}(\widehat{\psi}_{i,2}^* - \widehat{\psi}_{i,1}^*) \end{pmatrix} \Rightarrow 0.5(\Gamma_i^*)^{-1} \begin{pmatrix} \Upsilon g(\mathbf{B}^*) \\ \mathbf{B}_2^{*(i)}(1) \\ \mathbf{B}_{i+2}^*(1) \end{pmatrix},$$

in probability, where $g(\mathbf{B}^*) = \int_0^1 \mathbf{B}_1^*(\tau) d\mathbf{B}_2^{*(i)}(\tau)$,

$$\Gamma_i^* = \begin{pmatrix} \Gamma_{11,i}^* & \Gamma_{12,i}^* \\ \Gamma_{12,i}^{*\mathrm{T}} & E[f_{i,t}(0)\bar{\mathbf{z}}_t \bar{\mathbf{z}}_t^{\mathrm{T}}] \end{pmatrix},$$

$\Gamma_{11,i}^* = E[f_{i,t}(0)]\Upsilon \int_0^1 \mathbf{B}_1^*(\tau)\mathbf{B}_1^{*\mathrm{T}}(\tau) d\tau \Upsilon^{\mathrm{T}}$, $\Gamma_{12,i}^* = \Upsilon \int_0^1 \mathbf{B}_1^*(\tau) d\tau E[f_{i,t}(0)\bar{\mathbf{z}}_t^{\mathrm{T}}]$, $\mathbf{B}_2^{*(i)}(\tau)$ is the ith element of $\mathbf{B}_2^*(\tau)$, and Ω, $f_i(x)$ and Υ are defined as in Theorem 2.1.

Combining Theorems 2.1 and 3.1, we establish the asymptotic validity of the proposed hybrid bootstrap method. Note that, in the above procedure, the bootstrapped sequence $\{\mathbf{y}_t^*\}$ is only used to generate $\mathbf{B}_1^*(\tau)$; see the proof of Theorem 3.1. From (3), we can alternatively generate this sequence by $\mathbf{y}_t^* = \widehat{\Upsilon} \sum_{j=1}^{t} \mathbf{e}_j^*$, and they are equivalent asymptotically.

The wild bootstrap method may not work well by itself since the asymptotic distribution of the LAD estimator depends on two important quantities of the error term: the median and the density at the median. It is very hard to maintain both of them at the same time for the bootstrapped residual sequence.

One may also consider using the JYW method to approximate the asymptotic distribution in the previous section, i.e., using the quantity $\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}}$ to approximate the distribution of $\widehat{\boldsymbol{\theta}}$, where

$$\widehat{\boldsymbol{\theta}}^* = \operatorname{argmin} \sum_{t=p+2}^{T} \omega_t \|\Delta \mathbf{y}_t - \boldsymbol{\mu} - \Phi \mathbf{y}_{t-1} - \sum_{i=1}^{p} \Psi_i \Delta \mathbf{y}_{t-i}\|_1;$$

see [17]. The following result is a direct consequence of the proofs of Theorems 2.1 and 3.1, and we state it without proof.

Corollary 3.1. *For the simplest case, i.e., $N = 1$, $p = 0$, $\mathbf{m} = 0$ and $\{\mathbf{e}_t\}$ are i.i.d., if Assumptions 2.1–2.3 hold, then, conditional on $\mathbf{y}_1, \ldots, \mathbf{y}_T$,*

$$\begin{pmatrix} T(\widehat{\Phi}^* - \widehat{\Phi}) \\ \sqrt{T}(\widehat{\boldsymbol{\mu}}^* - \widehat{\boldsymbol{\mu}}) \end{pmatrix} \Rightarrow N\left\{0, \frac{1}{4f^2(0)} \cdot \Omega_{JYW}^{-1}\right\}$$

in probability, where $f(\cdot)$ is the density of $\mathbf{e}_t$, $\widehat{\boldsymbol{\theta}}^* = (\widehat{\Phi}^*, \widehat{\boldsymbol{\mu}}^*)^{\mathrm{T}}$, $\widehat{\boldsymbol{\theta}} = (\widehat{\Phi}, \widehat{\boldsymbol{\mu}})^{\mathrm{T}}$, and

$$\Omega_{JYW} = \lim_{T\to\infty} \begin{pmatrix} T^{-2} \sum \mathbf{y}_{t-1}^2 & T^{-3/2} \sum \mathbf{y}_{t-1} \\ T^{-3/2} \sum \mathbf{y}_{t-1} & 1 \end{pmatrix}.$$

This shows that the limiting distribution is normal, so the JYW method by itself fails to work here.

To compute the estimates $\widehat{\boldsymbol{\theta}}_1^*$ and $\widehat{\boldsymbol{\theta}}_2^*$, we can rewrite the corresponding models into a linear regression form, and then use a standard program to obtain their values as in the previous section.

4. APPLICATIONS TO PANEL UNIT ROOT TESTS

4.1 Three panel unit root tests

In this section, we construct three bootstrapping panel unit root tests by applying the methods in Sections 2 and

3. Two of them are for heterogeneous unit roots, and one is for homogeneous unit roots.

We first introduce the tests for heterogeneous unit roots. Consider the following special case of model (1),

$$\text{(4)} \quad \Delta y_{i,t} = \mu_i + \phi_i y_{i,t-1} + \sum_{j=1}^{p_i} \psi_{i,j} \Delta y_{i,t-j} + e_{i,t},$$

for $1 \le i \le N$ and $1 \le t \le T$, where $\mathbf{e}_t = (e_{1,t}, \ldots, e_{N,t})^{\mathrm{T}}$ is defined as in Section 2, and the parameters ϕ_i may take different values; see [7]. Comparing this with model (1), the interdependence of the individual errors $\mathbf{e}_t$ is kept, while the structural dependence is ignored; i.e., each individual time series depends on its past values only. Note that model (4) can be written as model (1) by letting $\Phi = \operatorname{diag}\{\phi_1, \ldots, \phi_N\}$, $\Psi_j = \operatorname{diag}\{\psi_{1,j}, \ldots, \psi_{N,j}\}$ and $p = \max\{p_1, \ldots, p_N\}$. We redefine the parameter vectors $\Phi = (\phi_1, \ldots, \phi_N)^{\mathrm{T}}$ and $\psi_i = (\psi_{i,1}, \ldots, \psi_{i,p_i})^{\mathrm{T}}$ for $1 \le i \le N$ in this section.

The hypotheses for the tests for heterogeneous unit roots can be formalized as

$$H_0^{HE} : \phi_1 = \cdots = \phi_N = 0 \quad \text{vs} \quad H_1^{HE} : \text{at least one } \phi_i \text{ is less than zero.}$$

Consider the LAD estimation of model (4),

$$\widehat{\boldsymbol{\theta}}_i = \operatorname{argmin} \sum_{t=p_i+2}^{T} |\Delta y_{i,t} - \phi_i y_{i,t-1} - \mu_i - \sum_{j=1}^{p_i} \psi_{i,j} \Delta y_{i,t-j}|,$$

for $1 \le i \le N$, where $\widehat{\phi}_i$ are the estimated unit roots. Let $\widehat{\Phi} = (\widehat{\phi}_1, \ldots, \widehat{\phi}_N)^{\mathrm{T}}$. A simple F-type test statistics can be constructed as

$$F_{HE1} = T^2 \widehat{\Phi}^{\mathrm{T}} \widehat{\Phi} = T^2 \sum_{i=1}^{N} \widehat{\phi}_i^2,$$

and we reject the null hypothesis if the value of F_{HE1} is too large.

Note that the $\widehat{\phi}_i$'s may have different variances and may be correlated with each other. To take this into account, we first find a matrix that standardizes the estimator $\widehat{\Phi}$ as for the OLS-based unit root tests [7]. When the errors $\{\mathbf{e}_t\}$ are *i.i.d.*, we have

$$\widehat{\phi}_i = \frac{1}{2 f_i(0) B_i} \sum_{t=p_i+2}^{T} (y_{i,t-1} - \mathbf{A}_i \bar{\mathbf{z}}_{i,t}) \operatorname{sgn}(e_{i,t}) + o_p(T^{-1}),$$

for $1 \le i \le N$, where $f_i(0) = f_{i,t}(0)$ is independent of t, $\bar{\mathbf{z}}_{i,t} = (1, \Delta y_{i,t-1}, \ldots, \Delta y_{i,t-p_i})^{\mathrm{T}}$,

$$\mathbf{A}_i = \sum_{t=p_i+2}^{T} y_{i,t-1} \bar{\mathbf{z}}_{i,t}^{\mathrm{T}} \left(\sum_{t=p_i+2}^{T} \bar{\mathbf{z}}_{i,t} \bar{\mathbf{z}}_{i,t}^{\mathrm{T}} \right)^{-1},$$

and

$$B_i = \sum_{t=p_i+2}^{T} y_{i,t-1}^2 - \sum_{t=p_i+2}^{T} y_{i,t-1} \bar{\mathbf{z}}_{i,t}^{\mathrm{T}} \left(\sum_{t=p_i+2}^{T} \bar{\mathbf{z}}_{i,t} \bar{\mathbf{z}}_{i,t}^{\mathrm{T}} \right)^{-1} \sum_{t=p_i+2}^{T} y_{i,t-1} \bar{\mathbf{z}}_{i,t};$$

see the proof of Theorem 4.1 for details. Note that it may be difficult to obtain consistent estimators of the densities $f_1(0), \ldots, f_N(0)$ especially when the number of individuals N is large. Fortunately, as they do not affect the correlation structure of $\widehat{\Phi}$, we can ignore their differences and define the (i,j)th element of the standardizing matrix $\widehat{\operatorname{var}}(\widehat{\Phi})$ by

$$\widehat{\operatorname{var}}(\widehat{\Phi})_{(i,j)} = \frac{\widehat{\Sigma}_{\mathbf{s}(i,j)}}{B_i B_j} \sum_{t=p+2}^{T} (y_{i,t-1} - \mathbf{A}_i \bar{\mathbf{z}}_{i,t})(y_{j,t-1} - \mathbf{A}_j \bar{\mathbf{z}}_{j,t}),$$

where $p = \max\{p_1, \ldots, p_N\}$, $\widehat{\Sigma}_{\mathbf{s}} = T^{-1} \sum_{t=p+2}^{T} \operatorname{sgn}(\widehat{\mathbf{e}}_t - \widehat{\boldsymbol{\mu}}) \operatorname{sgn}(\widehat{\mathbf{e}}_t - \widehat{\boldsymbol{\mu}})^{\mathrm{T}}$ is a consistent estimator of $\Sigma_{\mathbf{s}}$, and $\widehat{\Sigma}_{\mathbf{s}(i,j)}$ is the (i,j)th element of $\widehat{\Sigma}_{\mathbf{s}}$. This leads to another F-type test statistic

$$F_{HE2} = \widehat{\Phi}^{\mathrm{T}} [\widehat{\operatorname{var}}(\widehat{\Phi})]^{-1} \widehat{\Phi}.$$

Theorem 4.1 will show that the test statistic F_{HE2} is valid for the general case, although the standardizing matrix is derived for a simple case. Moreover, $\widehat{\operatorname{var}}(\widehat{\Phi}) = O_p(T^{-2})$ under H_0^{HE}, while $\widehat{\operatorname{var}}(\widehat{\Phi}) = O_p(T^{-1})$ under H_1^{HE}, and hence the test F_{HE1} is supposed to be more powerful as T is larger; see also [23].

In the literature of panel unit root tests, the test for homogeneous unit roots is usually employed to improve the power. Consider the model

$$\text{(5)} \quad \Delta y_{i,t} = \mu_i + \phi y_{i,t-1} + \sum_{j=1}^{p_i} \psi_{i,j} \Delta y_{i,t-j} + e_{i,t},$$

for $1 \le i \le N$ and $1 \le t \le T$, where all individual time series share the same unit root parameter ϕ. The hypotheses are

$$H_0^{HO} : \phi = 0 \quad \text{vs} \quad H_1^{HO} : \phi < 0,$$

and the corresponding LAD estimator is

$$\widetilde{\boldsymbol{\theta}} = \operatorname{argmin} \sum_{i=1}^{N} \sum_{t=p_i+2}^{T} |\Delta y_{i,t} - \phi y_{i,t-1} - \mu_i - \sum_{j=1}^{p_i} \psi_{i,j} \Delta y_{i,t-j}|.$$

The estimated coefficient $\widetilde{\phi}$ can be directly used to construct the test, resulting in the third test statistic

$$F_{HO} = T \widetilde{\phi}.$$

We next state the asymptotic null distributions of the three panel unit root test statistics.

For $1 \leq i \leq N$, denote $\varphi_i = (1 - \sum_{j=1}^{p_i} \psi_{i,j})^{-1}$, $\mathbf{z}_{i,t} = (\Delta y_{i,t-1}, \ldots, \Delta y_{i,t-p_i})^{\mathrm{T}}$, $\bar{\mathbf{z}}_{i,t} = (1, \mathbf{z}_{i,t}^{\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{\mu}_P(i) = E[f_{i,t}(0)\bar{\mathbf{z}}_{i,t}]$, and $\Sigma_P(i) = E[f_{i,t}(0)\bar{\mathbf{z}}_{i,t}\bar{\mathbf{z}}_{i,t}^{\mathrm{T}}]$. Let

$$\Omega_P = \begin{pmatrix} \Omega_{11,P} & \Omega_{12,P} \\ \Omega_{12,P}^{\mathrm{T}} & \Omega_{22,P} \end{pmatrix},$$

where

$$\Omega_{11,P} = \begin{pmatrix} \Sigma_{\mathbf{e}} & E[\mathbf{e}_t \operatorname{sgn}(\boldsymbol{\varepsilon}_t)^{\mathrm{T}}] \\ E[\operatorname{sgn}(\boldsymbol{\varepsilon}_t)\mathbf{e}_t^{\mathrm{T}}] & \Sigma_{\mathbf{s}} \end{pmatrix},$$

$$\Omega_{12,P} = \begin{pmatrix} E[\operatorname{sgn}(\varepsilon_{1,t})\mathbf{e}_t\mathbf{z}_{1,t}^{\mathrm{T}}] & \cdots & E[\operatorname{sgn}(\varepsilon_{1,t})\mathbf{e}_t\mathbf{z}_{1,t}^{\mathrm{T}}] \\ 0 & \cdots & 0 \end{pmatrix},$$

$$\Omega_{22,P} = \begin{pmatrix} \Sigma_{\mathbf{s}(1,1)}E(\mathbf{z}_{1,t}\mathbf{z}_{1,t}^{\mathrm{T}}) & \cdots & \Sigma_{\mathbf{s}(1,N)}E(\mathbf{z}_{1,t}\mathbf{z}_{N,t}^{\mathrm{T}}) \\ \vdots & \ddots & \vdots \\ \Sigma_{\mathbf{s}(N,1)}E(\mathbf{z}_{N,t}\mathbf{z}_{1,t}^{\mathrm{T}}) & \cdots & \Sigma_{\mathbf{s}(N,N)}E(\mathbf{z}_{N,t}\mathbf{z}_{N,t}^{\mathrm{T}}) \end{pmatrix},$$

$\Sigma_{\mathbf{s}(i,j)}$ is the (i,j)th element of the matrix $\Sigma_{\mathbf{s}} = E[\operatorname{sgn}(\boldsymbol{\varepsilon}_t)\operatorname{sgn}(\boldsymbol{\varepsilon}_t)^{\mathrm{T}}]$ and $\Sigma_{\mathbf{e}} = E(\mathbf{e}_t\mathbf{e}_t^{\mathrm{T}})$. Let $\mathbf{W}(\tau) = [\mathbf{W}_1^{\mathrm{T}}(\tau), \ldots, \mathbf{W}_{N+2}^{\mathrm{T}}(\tau)]^{\mathrm{T}}$ be a $(2N + \sum_{i=1}^N p_i)$-dimensional Brownian motion with covariance matrix $\tau\Omega_P$, where $\mathbf{W}_1(\tau)$ and $\mathbf{W}_2(\tau)$ are N-dimensional and $\mathbf{W}_{2+i}(\tau)$ is p_i-dimensional for $1 \leq i \leq N$. Let

$$\Lambda_{\mathbf{W}} = \int_0^1 \mathbf{W}_1(\tau)\mathbf{W}_1^{\mathrm{T}}(\tau)d\tau - \int_0^1 \mathbf{W}_1(\tau)d\tau \int_0^1 \mathbf{W}_1^{\mathrm{T}}(\tau)d\tau,$$

and $\zeta_{\mathbf{W}} = [\zeta_{\mathbf{W}}(1), \ldots, \zeta_{\mathbf{W}}(N)]^{\mathrm{T}}$, where

$$\begin{aligned}\zeta_{\mathbf{W}}^U(i) = &\int_0^1 \mathbf{W}_1^{(i)}(\tau)d\mathbf{W}_2^{(i)}(\tau) \\ &- \boldsymbol{\mu}_P^{\mathrm{T}}(i)\Sigma_P^{-1}(i)[\mathbf{W}_2^{(i)}(1), \mathbf{W}_{i+2}^{\mathrm{T}}(1)]^{\mathrm{T}} \int_0^1 \mathbf{W}_1^{(i)}(\tau)d\tau,\end{aligned}$$

$$\begin{aligned}\zeta_{\mathbf{W}}^L(i) =&2\varphi_i\{\int_0^1 [\mathbf{W}_1^{(i)}(\tau)]^2 d\tau E[f_{i,t}(0)] \\ &- [\int_0^1 \mathbf{W}_1^{(i)}(\tau)d\tau]^2 \boldsymbol{\mu}_P^{\mathrm{T}}(i)\Sigma_P^{-1}(i)\boldsymbol{\mu}_P(i)\},\end{aligned}$$

$\zeta_{\mathbf{W}}(i) = \zeta_{\mathbf{W}}^U(i)/\zeta_{\mathbf{W}}^L(i)$, and $\mathbf{W}_j^{(i)}(\tau)$ is the ith element of $\mathbf{W}_j(\tau)$ with $j = 1$ and 2.

Theorem 4.1. *Under H_0^{HE} or H_0^{HO}, if Assumptions 2.1–2.3 hold, then,*

$$F_{HE1} \Rightarrow \zeta_{\mathbf{W}}^{\mathrm{T}}\zeta_{\mathbf{W}}, \qquad F_{HE2} \Rightarrow \zeta_{\mathbf{W}}^{\mathrm{T}}\Sigma_{\widehat{\phi}}^{-1}\zeta_{\mathbf{W}},$$

and

$$F_{HO} \Rightarrow \zeta_{HO}^U/\zeta_{HO}^L,$$

where

$$\begin{aligned}\zeta_{HO}^U =&\sum_{i=1}^N \varphi_i\{\int_0^1 \mathbf{W}_1^{(i)}(\tau)d\mathbf{W}_2^{(i)}(\tau) \\ &- \boldsymbol{\mu}_P^{\mathrm{T}}(i)\Sigma_P^{-1}(i)[\mathbf{W}_2^{(i)}(1), \mathbf{W}_{i+2}^{\mathrm{T}}(1)]^{\mathrm{T}} \int_0^1 \mathbf{W}_1^{(i)}(\tau)d\tau\},\end{aligned}$$

$$\begin{aligned}\zeta_{HO}^L =&2\sum_{i=1}^N \varphi_i^2\{\int_0^1 [\mathbf{W}_1^{(i)}(\tau)]^2 d\tau E[f_{i,t}(0)] \\ &- [\int_0^1 \mathbf{W}_1^{(i)}(\tau)d\tau]^2 \boldsymbol{\mu}_P^{\mathrm{T}}(i)\Sigma_P^{-1}(i)\boldsymbol{\mu}_P(i)\},\end{aligned}$$

the (i,j)th element of the matrix $\Sigma_{\widehat{\phi}}$ is $\Sigma_{\widehat{\phi}(i,j)} = \Sigma_{\mathbf{s}(i,j)}\Lambda_{\mathbf{W}(i,j)}/[\varphi_i\varphi_j\Lambda_{\mathbf{W}(i,i)}\Lambda_{\mathbf{W}(j,j)}]$.

The asymptotic distributions in the above theorem are too complicated to use in practice. The hybrid bootstrap method in Section 3 offers a convenient way to approximate the corresponding critical values and p-values.

4.2 Bootstrap approximation

We first apply the hybrid bootstrap method to approximate the critical values and p-values of the two test statistics for heterogeneous unit roots, F_{HE1} and F_{HE2}.

Let $\mathbf{y}_t^* = (y_{1,t}^*, \ldots, y_{N,t}^*)^{\mathrm{T}}$ and $y_{i,t}^* = \widehat{\varphi}_i \sum_{j=1}^t (\omega_j - 1)\widehat{e}_{i,j}$, where $\widehat{\varphi}_i = (1 - \sum_{j=1}^{p_i} \widehat{\psi}_{i,j})^{-1}$, $\{\omega_t\}$ is an *i.i.d.* positive sequence defined as in Section 3, and $\widehat{e}_{i,t} = \Delta y_{i,t} - \widehat{\phi}_i y_{i,t-1} - \sum_{j=1}^{p_i} \widehat{\psi}_{i,j}\Delta y_{i,t-j}$ for $t \geq p_i + 2$ and is zero otherwise. Let

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{i,1}^* = \operatorname{argmin} \sum_{t=p_i+2}^T &|\Delta y_{i,t} - \phi_i y_{i,t-1}^* - \mu_i \\ &- \sum_{j=1}^{p_i} \psi_{i,j}\Delta y_{i,t-j} - \widehat{\phi}_i y_{i,t-1}|\end{aligned}$$

for $1 \leq i \leq N$, and

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{i,2}^* = \operatorname{argmin} \sum_{t=p_i+2}^T &\omega_t|\Delta y_{i,t} - \phi_i y_{i,t-1}^* - \mu_i - \\ &\sum_{j=1}^{p_i} \psi_{i,j}\Delta y_{i,t-j} - \widehat{\phi}_i y_{i,t-1}|\end{aligned}$$

for $1 \leq i \leq N$, and we denote by $\widehat{\phi}_{i,1}^*$ and $\widehat{\phi}_{i,2}^*$ the corresponding estimated unit roots.

For $1 \leq i, j \leq N$, define the following quantities: $\mathbf{A}_i^* = \sum_{t=p_i+2}^T y_{i,t-1}^*\bar{\mathbf{z}}_{i,t}^{\mathrm{T}}(\sum_{t=p_i+2}^T \bar{\mathbf{z}}_{i,t}\bar{\mathbf{z}}_{i,t}^{\mathrm{T}})^{-1}$,

$$\begin{aligned}B_i^* = &\sum_{t=p_i+2}^T y_{i,t-1}^{*2} \\ &- \sum_{t=p_i+2}^T y_{i,t-1}^*\bar{\mathbf{z}}_{i,t}^{\mathrm{T}}\left(\sum_{t=p_i+2}^T \bar{\mathbf{z}}_{i,t}\bar{\mathbf{z}}_{i,t}^{\mathrm{T}}\right)^{-1} \sum_{t=p_i+2}^T y_{i,t-1}^*\bar{\mathbf{z}}_{i,t},\end{aligned}$$

and

$$\widehat{\operatorname{var}}(\widehat{\Phi})_{(i,j)}^* = \frac{\widehat{\Sigma}_{\mathbf{s}(i,j)}}{B_i^*B_j^*} \sum_{t=p+2}^T (y_{i,t-1}^* - \mathbf{A}_i^*\bar{\mathbf{z}}_{i,t})(y_{j,t-1}^* - \mathbf{A}_j^*\bar{\mathbf{z}}_{j,t}),$$

where $\widehat{\Sigma}_{\mathbf{s}(i,j)} = T^{-1}\sum_{t=p+2}^{T}\operatorname{sgn}(\widehat{e}_{i,t}-\widehat{\mu}_i)\operatorname{sgn}(\widehat{e}_{j,t}-\widehat{\mu}_j)$ with $p = \max\{p_1,\dots,p_N\}$. Consider

$$F_{HE1}^* = T^2(\widehat{\Phi}_2^* - \widehat{\Phi}_1^*)^{\mathrm{T}}(\widehat{\Phi}_2^* - \widehat{\Phi}_1^*)$$

and

$$F_{HE2}^* = (\widehat{\Phi}_2^* - \widehat{\Phi}_1^*)^{\mathrm{T}}[\widehat{\mathrm{var}}(\widehat{\Phi})^*]^{-1}(\widehat{\Phi}_2^* - \widehat{\Phi}_1^*),$$

where $\widehat{\Phi}_1^* = (\widehat{\phi}_{1,1}^*,\dots,\widehat{\phi}_{N,1}^*)^{\mathrm{T}}$, $\widehat{\Phi}_2^* = (\widehat{\phi}_{1,2}^*,\dots,\widehat{\phi}_{N,2}^*)^{\mathrm{T}}$, and the (i,j)th element of $\widehat{\mathrm{var}}(\widehat{\Phi})^*$ is $\widehat{\mathrm{var}}(\widehat{\Phi})^*_{(i,j)}$. It may be expected that the conditional distributions of F_{HE1}^* and F_{HE2}^* are the same as the distributions of F_{HE1} and F_{HE2} respectively in the asymptotic sense.

We next consider the bootstrap approximation for the test statistic for homogeneous unit roots, F_{HO}. Let $\mathbf{y}_t^* = (y_{1,t}^*,\dots,y_{N,t}^*)^{\mathrm{T}}$ and $y_{i,t}^* = \widetilde{\varphi}_i\sum_{j=1}^{t}(\omega_j - 1)\widetilde{e}_{i,j}$, where $\widetilde{\varphi}_i = (1-\sum_{j=1}^{p_i}\widetilde{\psi}_{i,j})^{-1}$ and $\widetilde{e}_{i,t} = \Delta y_{i,t} - \widetilde{\phi}y_{i,t-1} - \sum_{j=1}^{p_i}\widetilde{\psi}_{i,j}\Delta y_{i,t-j}$ for $t > p_i+2$ and is zero otherwise. Consider

$$\widetilde{\boldsymbol{\theta}}_1 = \operatorname{argmin}\sum_{i=1}^{N}\sum_{t=p_i+2}^{T}|\Delta y_{i,t} - \phi y_{i,t-1}^* - \mu_i - \sum_{j=1}^{p_i}\psi_{i,j}\Delta y_{i,t-j} - \widetilde{\phi}y_{i,t-1}|$$

and

$$\widetilde{\boldsymbol{\theta}}_2 = \operatorname{argmin}\sum_{i=1}^{N}\sum_{t=p_i+2}^{T}\omega_t|\Delta y_{i,t} - \phi y_{i,t-1}^* - \mu_i - \sum_{j=1}^{p_i}\psi_{i,j}\Delta y_{i,t-j} - \widetilde{\phi}y_{i,t-1}|.$$

Let

$$F_{HO}^* = T(\widetilde{\phi}_2^* - \widetilde{\phi}_1^*),$$

where $\widetilde{\phi}_1^*$ and $\widetilde{\phi}_2^*$ are the estimated unit roots in $\widetilde{\boldsymbol{\theta}}_1$ and $\widetilde{\boldsymbol{\theta}}_2$ respectively. Let $\mathbf{W}^*(\tau) = [\mathbf{W}_1^{*\mathrm{T}}(\tau),\dots,\mathbf{W}_{N+2}^{*\mathrm{T}}(\tau)]^{\mathrm{T}}$ be a $(2N+\sum_{i=1}^{N}p_i)$-dimensional Brownian motion with covariance matrix $\tau\Omega_P$, where $\mathbf{W}_1^*(\tau)$ and $\mathbf{W}_2^*(\tau)$ are N-dimensional and $\mathbf{W}_{2+i}^*(\tau)$ is p_i-dimensional for $1\le i\le N$. Let

$$\Lambda_{\mathbf{W}}^* = \int_0^1\mathbf{W}_1^*(\tau)\mathbf{W}_1^{*\mathrm{T}}(\tau)d\tau - \int_0^1\mathbf{W}_1^*(\tau)d\tau\int_0^1\mathbf{W}_1^{*\mathrm{T}}(\tau)d\tau$$

and $\zeta_{\mathbf{W}}^* = [\zeta_{\mathbf{W}}^*(1),\dots,\zeta_{\mathbf{W}}^*(N)]^{\mathrm{T}}$, where

$$\zeta_{\mathbf{W}}^{*U}(i) = \int_0^1\mathbf{W}_1^{*(i)}(\tau)d\mathbf{W}_2^{*(i)}(\tau) - \boldsymbol{\mu}_P^{\mathrm{T}}(i)\Sigma_P^{-1}(i)[\mathbf{W}_2^{*(i)}(1),\mathbf{W}_{i+2}^{*\mathrm{T}}(1)]^{\mathrm{T}}\int_0^1\mathbf{W}_1^{*(i)}(\tau)d\tau,$$

$$\zeta_{\mathbf{W}}^{*L}(i) = 2\varphi_i\{\int_0^1[\mathbf{W}_1^{*(i)}(\tau)]^2d\tau E[f_{i,t}(0)] - [\int_0^1\mathbf{W}_1^{*(i)}(\tau)d\tau]^2\boldsymbol{\mu}_P^{\mathrm{T}}(i)\Sigma_P^{-1}(i)\boldsymbol{\mu}_P(i)\},$$

$\zeta_{\mathbf{W}}^*(i) = \zeta_{\mathbf{W}}^{*U}(i)/\zeta_{\mathbf{W}}^{*L}(i)$, and $\mathbf{W}_j^{*(i)}(\tau)$ is the ith element of $\mathbf{W}_j^*(\tau)$ with $j=1$ and 2.

Theorem 4.2. *Under Assumptions 2.1–2.3, conditional on $\mathbf{y}_1,\dots,\mathbf{y}_T$, if H_0^{HE} or H_1^{HE} holds, then,*

$$F_{HE1}^* \Rightarrow \zeta_{\mathbf{W}}^{*\mathrm{T}}\zeta_{\mathbf{W}}^* \quad \textit{and} \quad F_{HE2}^* \Rightarrow \zeta_{\mathbf{W}}^{*\mathrm{T}}\Sigma_{\widehat{\phi}}^{*-1}\zeta_{\mathbf{W}}^*$$

in probability, and if H_0^{HO} or H_1^{HO} holds, then

$$F_{HO}^* \Rightarrow \zeta_{HO}^{*U}/\zeta_{HO}^{*L}$$

in probability, where

$$\zeta_{HO}^{*U} = \sum_{i=1}^{N}\varphi_i\{\int_0^1\mathbf{W}_1^{*(i)}(\tau)d\mathbf{W}_2^{*(i)}(\tau) - \boldsymbol{\mu}_P^{\mathrm{T}}(i)\Sigma_P^{-1}(i)[\mathbf{W}_2^{*(i)}(1),\mathbf{W}_{i+2}^{*\mathrm{T}}(1)]^{\mathrm{T}}\int_0^1\mathbf{W}_1^{*(i)}(\tau)d\tau\},$$

$$\zeta_{HO}^{*L} = 2\sum_{i=1}^{N}\varphi_i^2\{\int_0^1[\mathbf{W}_1^{*(i)}(\tau)]^2d\tau E[f_{i,t}(0)] - [\int_0^1\mathbf{W}_1^{*(i)}(\tau)d\tau]^2\boldsymbol{\mu}_P^{\mathrm{T}}(i)\Sigma_P^{-1}(i)\boldsymbol{\mu}_P(i)\},$$

the (i,j)th element of the matrix $\Sigma_{\widehat{\phi}}^*$ is

$$\Sigma_{\widehat{\phi}(i,j)}^* = \Sigma_{\mathbf{s}(i,j)}\Lambda_{\mathbf{W}(i,j)}^*/[\varphi_i\varphi_j\Lambda_{\mathbf{W}(i,i)}^*\Lambda_{\mathbf{W}(j,j)}^*].$$

The above theorem establishes the asymptotic validity of the corresponding bootstrapping procedure for the panel unit root tests.

There are two types of bootstrapping unit root tests for univariate time series: residual-based and difference-based tests; see [31] and [29]. The three test statistics in this section are residual-based. We can construct the corresponding difference-based tests by deleting $\widehat{\phi}_i y_{i,t-1}$ (or $\widetilde{\phi}_i y_{i,t-1}$) from the objective functions of $\widehat{\boldsymbol{\theta}}_{i,1}^*$ and $\widehat{\boldsymbol{\theta}}_{i,2}^*$ (or $\widetilde{\boldsymbol{\theta}}_1$ and $\widetilde{\boldsymbol{\theta}}_2$), and the same asymptotic distribution is expected under H_0^{HE} or H_0^{HO}.

5. SIMULATION STUDIES

We conduct three simulation experiments to examine the finite-sample performance of the proposed hybrid bootstrap method and bootstrapping panel unit root tests. All simulation results are based on 1000 replications, with 1000 bootstrapped sequences for each replication.

In the first simulation experiment, we examine the performance of the hybrid bootstrap method in Section 3, and the data generating process is

$$\Delta \mathbf{y}_t = \boldsymbol{\mu} + \Phi \mathbf{y}_{t-1} + \mathbf{e}_t,$$

where

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = 0 \quad \text{and} \quad \Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} = 0.$$

The error terms $\{\mathbf{e}_t\}$ are either *i.i.d.* random vectors

$$\mathbf{e}_t = \begin{pmatrix} 0.2 & 0 \\ 0.3 & 0.3 \end{pmatrix} \begin{pmatrix} \chi_1^2 - 1 \\ \chi_4^2 - 4 \end{pmatrix}$$

or generated from a vector GARCH process,

$$(6) \qquad \mathbf{e}_t = V_t^{1/2} \boldsymbol{\varepsilon}_t, \quad V_t = 0.1\Sigma + 0.7\mathbf{e}_{t-1}\mathbf{e}_{t-1}^{\mathrm{T}} + 0.2V_{t-1},$$

where χ_1^2 and χ_4^2 are independent chi-squared distributed random variables with one and four degrees of freedom respectively, $\{\boldsymbol{\varepsilon}_t\}$ are *i.i.d.* N-dimensional standard normal random vectors with $N = 2$, and

$$\Sigma = \begin{pmatrix} 0.08 & 0.12 \\ 0.12 & 0.90 \end{pmatrix}.$$

Note that $\{\mathbf{e}_t\}$ have the same unconditional covariance Σ for both cases, and the distribution of $\mathbf{e}_t$ is skewed with median $(-0.1059, -0.1912)^{\mathrm{T}}$ in the *i.i.d.* case. We consider three perturbing distributions of $\{\omega_t\}$ in this experiment: (i) the standard exponential distribution, (ii) a two-point distribution, which takes the value 0 or 2, each with probability 0.5, (iii) the log-normal distribution with its logarithm following the normal distribution with mean and variance being $-0.5\log(2)$ and $\log(2)$ respectively. The two-point distribution was employed by [38] and [24] for bootstrapping approximations, while the exponential and log-normal distributions are commonly used for nonnegative continuous random variables with both mean and variance equal to one. We consider three series lengths, $T = 50$, 100 and 200, and the hybrid bootstrap method in Section 3 is applied to calculate the 95% confidence intervals of the parameters for each replication. The empirical coverages are listed in Table 1, and it can be observed that the log-normal distribution outperforms the other two distributions.

The second experiment focuses on the bootstrapping panel unit root tests in Section 4. The data generating process is

$$\Delta y_{i,t} = \mu_i + \phi_i y_{i,t-1} + e_{i,t}, \quad 1 \le i \le N,$$

where $\mu_i = 0$ and $\phi_i = 0$ correspond to the size, while $\phi_i = -0.05$ and -0.10 correspond to the power. The error term $\mathbf{e}_t = (e_{1,t}, \ldots, e_{N,t})^{\mathrm{T}}$ follows the vector GARCH process in (6), and the matrix Σ is generated as in [7]:

(i) Generate an $N \times N$ matrix U whose entries are independent, following $U[0,1]$, the uniform distribution on $[0,1]$, and then construct an orthogonal matrix $H = U(U'U)^{-1/2}$.

Table 1. Empirical coverage rate of the bootstrapped 95% confidence intervals with sample size $T = 50$, 100 and 200

	Skewed *i.i.d.* errors			Vector GARCH errors		
	50	100	200	50	100	200
	Exponential distribution					
μ_1	0.948	0.952	0.934	0.871	0.924	0.947
μ_2	0.916	0.930	0.941	0.877	0.914	0.927
$\phi_{1,1}$	0.972	0.944	0.900	0.896	0.879	0.908
$\phi_{1,2}$	0.987	0.968	0.949	0.974	0.952	0.945
$\phi_{2,1}$	0.974	0.963	0.954	0.969	0.959	0.953
$\phi_{2,2}$	0.905	0.901	0.880	0.890	0.893	0.896
	Two-point distribution					
μ_1	0.956	0.938	0.923	0.885	0.927	0.937
μ_2	0.913	0.928	0.934	0.899	0.917	0.929
$\phi_{1,1}$	0.937	0.905	0.860	0.899	0.879	0.904
$\phi_{1,2}$	0.977	0.951	0.939	0.968	0.943	0.956
$\phi_{2,1}$	0.955	0.943	0.937	0.967	0.957	0.947
$\phi_{2,2}$	0.887	0.884	0.880	0.893	0.893	0.887
	Log-normal distribution					
μ_1	0.951	0.947	0.952	0.860	0.909	0.941
μ_2	0.909	0.937	0.959	0.863	0.920	0.923
$\phi_{1,1}$	0.965	0.942	0.928	0.880	0.884	0.907
$\phi_{1,2}$	0.990	0.978	0.965	0.975	0.975	0.950
$\phi_{2,1}$	0.974	0.967	0.966	0.977	0.970	0.950
$\phi_{2,2}$	0.901	0.907	0.910	0.893	0.905	0.907

(ii) Generate N eigenvalues $\eta_1, \ldots, \eta_N$, where $\eta_1 = 0.1, \eta_N = 1$ and η_i are independent $U[0.1, 1]$ for $i = 2, \ldots, N-1$.

(iii) Let $\Sigma = H \operatorname{diag}\{\eta_1, \ldots, \eta_N\} H'$.

The resulting vector GARCH processes have different underlying structures for different replications. The number of individuals is $N = 2$, and we consider three series lengths, $T = 100$, 200 and 300. The three perturbing distributions for the hybrid bootstrap method in the first experiment are employed. Table 2 presents the empirical sizes and powers of the three proposed panel unit root tests, F_{HE1}^*, F_{HE2}^* and F_{HO}^*, at the significance levels of 1%, 5% and 10%. From this table, we have five findings: (i) the empirical sizes are all close to the corresponding nominal values even when the series length is as small as $T = 100$; (ii) the empirical powers increase quickly as the series length increases from $T = 100$ to 300; (iii) the test for homogeneous unit roots, F_{HO}^*, is more powerful than those for heterogeneous unit roots, F_{HE1}^* and F_{HE2}^*; (iv) the test F_{HE2}^* is slightly better than F_{HE1}^* for most cases with shorter length $T = 100$ and smaller departure $\phi_i = 0.05$, and otherwise F_{HE1}^* is better; and (v) there is no significant difference among the three perturbing distributions.

The third experiment aims to compare the tests proposed in Section 4 with the OLS-based tests, F_{OT}^* and t_{OT}^*, in [7]. The data generating process is

$$\Delta y_{i,t} = \mu_i + \phi_i y_{i,t-1} + \psi_i \Delta y_{i,t-1} + e_{i,t}, \quad 1 \le i \le N,$$

Table 2. *Empirical size and power of the panel unit root tests, F^*_{HE1}, F^*_{HE2} and F^*_{HO}, with three perturbing distributions*

		$T = 100$			$T = 200$			$T = 300$		
Test	ϕ	1%	5%	10%	1%	5%	10%	1%	5%	10%
					Exponential distribution					
F^*_{HE1}	0.0	0.009	0.046	0.094	0.010	0.045	0.096	0.010	0.055	0.097
	-0.05	0.096	0.324	0.519	0.403	0.838	0.931	0.818	0.982	0.999
	-0.10	0.332	0.744	0.914	0.917	0.989	0.996	0.997	0.999	1.000
F^*_{HE2}	0.0	0.011	0.052	0.103	0.012	0.056	0.094	0.011	0.047	0.118
	-0.05	0.093	0.372	0.553	0.358	0.741	0.885	0.655	0.922	0.977
	-0.10	0.254	0.686	0.841	0.775	0.958	0.983	0.961	0.996	0.998
F^*_{HO}	0.0	0.011	0.050	0.091	0.009	0.046	0.095	0.009	0.051	0.097
	-0.05	0.192	0.595	0.780	0.703	0.949	0.983	0.954	0.995	1.000
	-0.10	0.562	0.920	0.970	0.973	0.992	0.996	0.998	1.000	1.000
					Two-point distribution					
F^*_{HE1}	0.0	0.007	0.046	0.108	0.011	0.051	0.097	0.010	0.053	0.094
	-0.05	0.069	0.323	0.578	0.420	0.859	0.950	0.818	0.984	0.995
	-0.10	0.340	0.775	0.903	0.922	0.992	0.997	0.999	1.000	1.000
F^*_{HE2}	0.0	0.011	0.045	0.104	0.009	0.049	0.093	0.011	0.058	0.098
	-0.05	0.060	0.329	0.528	0.371	0.754	0.887	0.656	0.913	0.970
	-0.10	0.252	0.664	0.821	0.751	0.953	0.989	0.959	0.999	0.999
F^*_{HO}	0.0	0.009	0.049	0.105	0.009	0.048	0.098	0.009	0.044	0.097
	-0.05	0.169	0.599	0.797	0.726	0.956	0.990	0.941	0.995	0.997
	-0.10	0.601	0.902	0.958	0.973	0.994	0.998	1.000	1.000	1.000
					Log-normal distribution					
F^*_{HE1}	0.0	0.010	0.054	0.100	0.010	0.042	0.094	0.009	0.045	0.097
	-0.05	0.068	0.322	0.545	0.388	0.828	0.935	0.797	0.981	0.998
	-0.10	0.295	0.744	0.899	0.917	0.998	0.998	1.000	1.000	1.000
F^*_{HE2}	0.0	0.010	0.050	0.114	0.008	0.055	0.097	0.010	0.054	0.092
	-0.05	0.101	0.388	0.576	0.371	0.759	0.907	0.660	0.913	0.969
	-0.10	0.294	0.704	0.850	0.778	0.969	0.994	0.956	0.996	1.000
F^*_{HO}	0.0	0.010	0.055	0.108	0.008	0.048	0.095	0.008	0.051	0.107
	-0.05	0.169	0.589	0.779	0.676	0.945	0.982	0.936	0.996	0.998
	-0.10	0.569	0.919	0.973	0.979	0.997	0.998	0.998	1.000	1.000

where $\mu_i = 0$, the ϕ_i's are *i.i.d.* following $U[-c_0, 0]$, and the ψ_i's are *i.i.d.* following $U[0.2, 0.4]$. The constant c_0 is zero for the size and 0.1, 0.2, 0.4 or 0.6 for the power. The errors $\{\mathbf{e}_t\}$ are either *i.i.d.* normal random vectors with covariance matrix Σ or generated from the vector GARCH process in (6), where the matrix Σ is defined as in the previous experiment. We consider three combinations of the number of individuals and the series length, $(N, T) = (2, 100)$, $(2, 200)$ and $(5, 100)$, and use the log-normal perturbing distribution for $\{\omega_t\}$. Tables 3 and 4 report the rejection rates for the case of *i.i.d.* errors and for the case of vector GARCH errors, respectively, at the significance levels of 5% and 10%. It can be observed that the OLS-based tests have distorted sizes for the vector GARCH error case, while the empirical sizes of the proposed tests are close to the nominal rates for both cases. This is not surprising since the OLS-based bootstrapping tests in [7] are based on resampling the residuals, which are expected to handle the conditional heteroscedasticity poorly. Furthermore, the proposed tests are more powerful than the OLS-based tests, even for the case of *i.i.d.* errors. Note that F^*_{HE1} and F^*_{HO} are actually coefficient statistics while the two OLS-based tests are t-statistics, and the coefficient statistics are usually more powerful; see [23]. Another possible reason might be the powerfulness of the JYW method; see [8]. Finally, compared with the two tests for heterogeneous unit roots, the test F^*_{HO} is powerful for small values of c_0, and it is also noteworthy that the values of the ϕ_i's are closer to each other in such cases.

Lastly, we briefly report the computation time of the proposed procedure. All recorded time is for running our Matlab programs on a laptop with the Intel® Core™ i7-9750H processor (CPU@2.60GHz, 16.0GB RAM). The recorded time includes that spent on generating the time series data. Table 5 displays the average computation time per replication for constructing the bootstrapped confidence intervals in Table 1. Table 6 reports that for conducting the proposed tests F^*_{HE1}, F^*_{HE2} and F^*_{HO} under the settings of Table 3 when $\phi_i = 0$ and the significance level is 5%. Generally speaking, the computation speed is relatively insensitive to the perturbing and error distributions. We also investigated more settings of (N, T) than those reported here. We found that the computation time tends to increase linearly with T, and it increases faster as N gets larger.

Table 3. Rejection rate of the proposed tests, F^*_{HE1}, F^*_{HE2} and F^*_{HO}, and two OLS-based bootstrapping panel unit root tests, F^*_{OT} and t^*_{OT}, for the case of $i.i.d.$ errors with $(N,T) = (2,100)$, $(2,200)$ and $(5,100)$

	$(2,100)$		$(2,200)$		$(5,100)$	
Test	5%	10%	5%	10%	5%	10%
			$\phi_i = 0$			
F^*_{HE1}	0.054	0.108	0.050	0.098	0.040	0.086
F^*_{HE2}	0.053	0.112	0.048	0.099	0.049	0.096
F^*_{HO}	0.060	0.101	0.054	0.108	0.051	0.098
F^*_{OT}	0.060	0.122	0.077	0.140	0.059	0.138
t^*_{OT}	0.045	0.087	0.067	0.133	0.062	0.108
			$\phi_i \sim U[-0.1, 0]$			
F^*_{HE1}	0.512	0.675	0.837	0.917	0.686	0.852
F^*_{HE2}	0.354	0.531	0.621	0.768	0.560	0.731
F^*_{HO}	0.577	0.715	0.808	0.877	0.808	0.886
F^*_{OT}	0.261	0.380	0.448	0.552	0.365	0.495
t^*_{OT}	0.270	0.374	0.453	0.554	0.432	0.542
			$\phi_i \sim U[-0.2, 0]$			
F^*_{HE1}	0.812	0.897	0.956	0.983	0.978	0.994
F^*_{HE2}	0.594	0.756	0.834	0.910	0.930	0.975
F^*_{HO}	0.783	0.858	0.901	0.932	0.927	0.951
F^*_{OT}	0.433	0.535	0.650	0.718	0.596	0.716
t^*_{OT}	0.406	0.495	0.531	0.620	0.575	0.659
			$\phi_i \sim U[-0.4, 0]$			
F^*_{HE1}	0.948	0.968	0.987	0.992	1.000	1.000
F^*_{HE2}	0.847	0.914	0.964	0.979	0.993	0.999
F^*_{HO}	0.894	0.932	0.953	0.968	0.961	0.970
F^*_{OT}	0.587	0.674	0.769	0.828	0.850	0.904
t^*_{OT}	0.523	0.584	0.686	0.744	0.660	0.729
			$\phi_i \sim U[-0.6, 0]$			
F^*_{HE1}	0.991	0.994	0.995	0.996	1.000	1.000
F^*_{HE2}	0.930	0.965	0.982	0.988	1.000	1.000
F^*_{HO}	0.941	0.961	0.966	0.980	0.981	0.988
F^*_{OT}	0.702	0.772	0.833	0.886	0.919	0.948
t^*_{OT}	0.568	0.648	0.731	0.787	0.696	0.760

Table 4. Rejection rate of the proposed tests, F^*_{HE1}, F^*_{HE2} and F^*_{HO}, and the two OLS-based bootstrapping panel unit root tests, F^*_{OT} and t^*_{OT}, for the case of vector GARCH errors with $(N,T) = (2,100)$, $(2,200)$ and $(5,100)$

	$(2,100)$		$(2,200)$		$(5,100)$	
Test	5%	10%	5%	10%	5%	10%
			$\phi_i = 0$			
F^*_{HE1}	0.042	0.081	0.040	0.083	0.042	0.091
F^*_{HE2}	0.047	0.091	0.041	0.092	0.045	0.099
F^*_{HO}	0.047	0.093	0.049	0.099	0.044	0.095
F^*_{OT}	0.069	0.130	0.090	0.170	0.064	0.131
t^*_{OT}	0.087	0.146	0.102	0.157	0.092	0.138
			$\phi_i \sim U[-0.1, 0]$			
F^*_{HE1}	0.528	0.709	0.893	0.942	0.655	0.837
F^*_{HE2}	0.462	0.612	0.777	0.869	0.630	0.799
F^*_{HO}	0.633	0.758	0.851	0.896	0.823	0.903
F^*_{OT}	0.273	0.352	0.458	0.579	0.350	0.470
t^*_{OT}	0.306	0.401	0.472	0.556	0.419	0.537
			$\phi_i \sim U[-0.2, 0]$			
F^*_{HE1}	0.843	0.918	0.970	0.983	0.958	0.989
F^*_{HE2}	0.723	0.849	0.904	0.953	0.919	0.971
F^*_{HO}	0.815	0.886	0.917	0.941	0.923	0.960
F^*_{OT}	0.386	0.484	0.634	0.714	0.590	0.725
t^*_{OT}	0.394	0.488	0.552	0.639	0.530	0.614
			$\phi_i \sim U[-0.4, 0]$			
F^*_{HE1}	0.955	0.976	0.994	0.996	1.000	1.000
F^*_{HE2}	0.881	0.933	0.973	0.988	0.993	1.000
F^*_{HO}	0.894	0.933	0.952	0.964	0.966	0.975
F^*_{OT}	0.569	0.683	0.734	0.797	0.774	0.865
t^*_{OT}	0.509	0.612	0.654	0.727	0.670	0.691
			$\phi_i \sim U[-0.6, 0]$			
F^*_{HE1}	0.979	0.990	0.995	0.997	1.000	1.000
F^*_{HE2}	0.945	0.973	0.982	0.995	0.999	1.000
F^*_{HO}	0.942	0.965	0.966	0.975	0.976	0.980
F^*_{OT}	0.686	0.762	0.831	0.873	0.886	0.928
t^*_{OT}	0.581	0.642	0.742	0.803	0.697	0.774

6. REAL DATA EXAMPLES

We demonstrate the usefulness of the proposed panel unit root tests in two real data applications: monthly real exchange rates and daily trading volumes of stocks.

In economics, empirical tests of purchasing power parity (PPP) are typically formulated as tests of stationarity of the real exchange rate (RER) in an I(1) modeling framework. This is an important application of panel unit root testing, and support of PPP corresponds to rejection of the unit root null hypothesis. However, empirical results have been mixed dependent on different data sets and methodologies; see [3] for a recent survey. In our first application, we investigate if PPP holds among a subgroup of the world's most advanced economies using the monthly country average RER from January 1970 to December 2020. The data set uses the U.S. as the based country and is obtained from the Agricultural Exchange Rate Data Set of the U.S. Department of Agriculture (USDA) Economic Research Service; see https://www.ers.usda.gov/data-products/agricultural-exchange-rate-data-set/. Note that the RER is different from the nominal exchange rate (NER) as the former is adjusted by the relative rates of inflation: specifically, for the USDA data,

$$\mathrm{RER}_{i,t} = \frac{\mathrm{NER}_{i,t}}{\mathrm{LCCPI}_{i,t}/\mathrm{USCPI}_{i,t}},$$

Table 5. Average computation time (in seconds) per replication for constructing bootstrapped 95% confidence intervals of $\mu_1, \mu_2, \phi_{1,1}, \phi_{1,2}, \phi_{2,1}$ and $\phi_{2,2}$ under the settings of Table 1 with $T = 50$, 100 and 200

	i.i.d. errors			GARCH errors		
	50	100	200	50	100	200
Exponential	1.7952	2.5711	5.5951	1.6246	2.1997	4.9739
Two-point	1.7190	2.5094	5.5953	1.5729	2.2215	4.7462
Log-normal	1.8064	2.5610	5.8437	1.6341	2.4676	4.8876

Table 6. *Average computation time (in seconds) per replication for conducting the proposed tests F^*_{HE1}, F^*_{HE2} and F^*_{HO} under the settings of Table 3 when $\phi_i = 0$, the significance level is 5% and $(N, T) = (2, 100)$, $(2, 200)$ or $(5, 100)$*

	$(2, 100)$	$(2, 200)$	$(5, 100)$
All three tests	4.7311	10.4988	14.5598
F^*_{HE1} & F^*_{HE2}	2.6421	5.7982	7.4263
F^*_{HO} only	2.1121	4.5489	7.9297

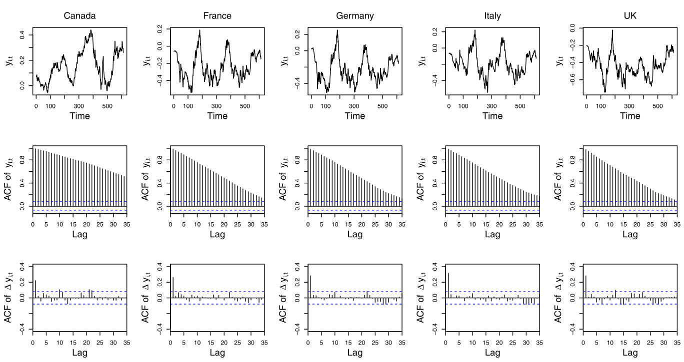

Figure 1. Monthly logarithm transformed real exchange rates $\{\mathbf{y}_t\}$ (top) and empirical autocorrelation functions of $\{\mathbf{y}_t\}$ (middle) and $\{\Delta\mathbf{y}_t\}$ (bottom) for five countries from January 1970 to December 2020. The dashed lines represent 95% confidence bands.

where $\mathrm{NER}_{i,t}$ is the nominal dollar exchange rate in terms of country i's local currency, and $\mathrm{LCCPI}_{i,t}$ and $\mathrm{USCPI}_{i,t}$ are the Consumer Price Indexes for country i and the U.S., respectively; for details, see the documentation on the above website for the data.

We consider five countries ($N = 5$), including Canada, France, Germany, Italy, and the U.K., and define $\mathbf{y}_t \in \mathbb{R}^5$ as the logarithm transformed sequence of the monthly RER; i.e., $y_{i,t} = \log(\mathrm{RER}_{i,t})$. There are $T = 612$ observations. The time series plots of $\{y_{i,t}\}$ and empirical autocorrelation functions (ACFs) of $\{y_{i,t}\}$ and $\{\Delta y_{i,t}\}$ are displayed in Figure 1, for $1 \le i \le 5$. Since the ACF of $\{y_{i,t}\}$ substantially exceeds the upper confidence limit and decays very slowly for all countries, we strongly suspect that there exist unit roots. In addition, the ACF of $\{\Delta y_{i,t}\}$ suggests that only lag 1 of $\Delta y_{i,t}$ is significant for all countries. Thus, we choose the lag orders $p_i = 1$ for $1 \le i \le 5$. Then we conduct the proposed panel unit root tests based on 1000 bootstrapped sequences and the log-normal perturbing distribution defined as in the previous section. The total computation time is 230.53 seconds using the same laptop for reporting the time in Section 5. The resulting p-values are 0.3020 for F^*_{HE1}, 0.8180 for F^*_{HE2}, and 0.3590 for F^*_{HO}, all suggesting that the unit root hypothesis is not rejected, i.e., PPP is not supported by this data set.

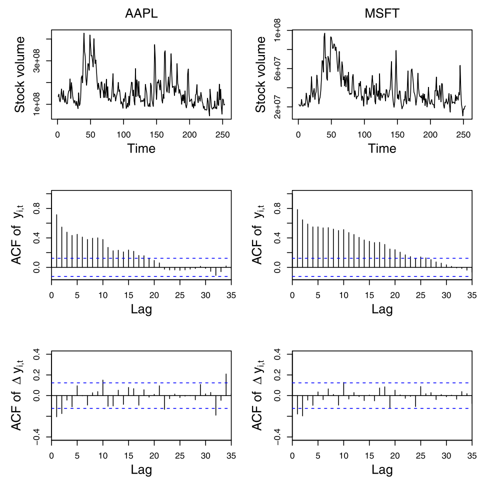

Figure 2. Daily trading volumes (top) and empirical autocorrelation functions of the normalized series $\{\mathbf{y}_t\}$ (middle) and first difference $\{\Delta\mathbf{y}_t\}$ for two stocks from January 2, 2020 to December 31, 2020. The dashed lines represent 95% confidence bands.

Next we consider an application in finance: testing the unit root behavior in stock trading volumes time series. For illustration, we picked two top technology companies, Apple (AAPL) and Microsoft (MSFT), whose stock data are among the list of most popular datasets on the Historical Data webpage of Nasdaq Inc.; see https://www.nasdaq.com/market-activity/quotes/historical, which provides real-time daily historical stock prices and volumes data to the public. We focus on daily trading volumes of the two stocks from January 2, 2020 to December 31, 2020, where there are $T = 253$ trading days; see the top panel of Figure 2. Since the raw volume observations are in the order of 10^7, we define each $\{y_{i,t}\}$ for $i = 1, 2$ as the correspondingly normalized series with mean zero and variance one. Note that the normalization does not affect the autocorrelations. Unlike the middle panel of Figure 1, the empirical ACF of $\{y_{i,t}\}$ in Figure 2 decays only moderately slowly, so it is much less clear whether there exist unit roots. Moreover, in contrast to Figure 1, the empirical ACF of $\{\Delta y_{i,t}\}$ in Figure 2 is still prominent at some fairly large lags, especially in the case of AAPL. Thus, we rely on the Akaike information criterion (AIC) to select the lag order p_i based on fitting an AR model for each $\{\Delta y_{i,t}\}$, where $i = 1$ corresponds to AAPL and $i = 2$ to MSFT. The AIC suggests $p_1 = 5$ and $p_2 = 7$. Accordingly, following the same procedure as in the previous example, we obtain p-values of the proposed tests as follows: 0.0020 for F^*_{HE1}, 0.0230 for F^*_{HE2}, and less than 10^{-4} for F^*_{HO}. The total computation time is 16.08 seconds. Clearly, the unit root hypothesis is rejected. To demonstrate the advantage of the panel unit root tests over the univariate approach, we conduct the augmented Dickey-Fuller (ADF) test for each $\{y_{i,t}\}$. The p-values are 0.0279 for AAPL and

0.2682 for MSFT, both much larger than those of the panel unit root tests. This suggests that testing unit roots in panels can indeed significantly increase the power of the test. Note that the panel data approach will detect the violation of the unit root null hypothesis if any component series has a unit root. The ability to jointly test unit root behavior is especially useful for financial data where the analysis of large panels is commonly of interest.

7. CONCLUSION AND DISCUSSION

This paper studies the LAD estimation for vector AR models with pure unit roots. The technical conditions on the error term are mild, and two important cases are included: *i.i.d.* errors with the median possibly not equal to zero and conditional heteroscedastic errors. A novel bootstrap method is proposed for the approximation of the complicated asymptotic distributions, and its validity is verified by both theory and simulations. We further construct three bootstrapping panel unit root tests whose usefulness is demonstrated by real applications.

Besides the pure unit roots, the nonstationary vector AR model is usually employed to study cointegration, which is another important feature in financial time series; see Section 2 for details. The methodology proposed in this paper can be readily extended to this case. Furthermore, the proposed bootstrap method may be extended to other estimation methods such as the quantile regression. We will leave these for future research.

APPENDIX A. TECHNICAL DETAILS

Lemma A.1. *Under Assumptions 2.1 and 2.2, the following results are jointly satisfied:*

$$(i)\quad \frac{1}{T^2}\sum_{t=p+2}^{T}\mathbf{y}_{t-1}\mathbf{y}_{t-1}^{\mathrm{T}} \Rightarrow \Upsilon\int_0^1 \mathbf{B}_1(\tau)\mathbf{B}_1^{\mathrm{T}}(\tau)d\tau\Upsilon^{\mathrm{T}},$$

$$(ii)\quad \frac{1}{T\sqrt{T}}\sum_{t=p+2}^{T}\mathbf{y}_{t-1} \Rightarrow \Upsilon\int_0^1 \mathbf{B}_1(\tau)d\tau,$$

$$(iii)\quad \frac{1}{T}\sum_{t=p+2}^{T}\mathrm{sgn}(\varepsilon_{i,t})\mathbf{y}_{t-1} \Rightarrow \Upsilon\int_0^1 \mathbf{B}_1(\tau)d\mathbf{B}_2^{(i)}(\tau), \qquad i=1,\dots,N,$$

$$(iv)\quad \frac{1}{\sqrt{T}}\sum_{t=p+2}^{T}\mathrm{sgn}(\boldsymbol{\varepsilon}_t) \Rightarrow \mathbf{B}_2(1),$$

$$(v)\quad \frac{1}{\sqrt{T}}\sum_{t=p+2}^{T}\mathrm{sgn}(\varepsilon_{i,t})\mathbf{z}_t \Rightarrow \mathbf{B}_{i+2}(1), \qquad i=1,\dots,N,$$

where $\Upsilon = (\mathbf{I}_N - \Psi_1 - \cdots - \Psi_p)^{-1}$, $\mathbf{B}(\tau) = [\mathbf{B}_1^{\mathrm{T}}(\tau),\dots,\mathbf{B}_{N+2}^{\mathrm{T}}(\tau)]^{\mathrm{T}}$ *is the* (pN^2+2N)*-dimensional Brownian motion as defined in Theorem 2.1, and* $\mathbf{B}_2^{(i)}(\tau)$ *is the ith element of* $\mathbf{B}_2(\tau)$.

Proof. Let

$$\zeta_t = [\mathbf{e}_t^{\mathrm{T}}, \mathrm{sgn}(\boldsymbol{\varepsilon}_t)^{\mathrm{T}}, \mathrm{sgn}(\varepsilon_{1,t})\mathbf{z}_t^{\mathrm{T}},\dots,\mathrm{sgn}(\varepsilon_{N,t})\mathbf{z}_t^{\mathrm{T}}]\lambda,$$

where λ is a (pN^2+2N)-dimensional constant vector with $\lambda^{\mathrm{T}}\lambda \neq 0$. It is noteworthy that $\{\zeta_t, t\in\mathbb{Z}\}$ is a martingale difference sequence with respect to the filtration $\{\mathcal{F}_t, t\in\mathbb{Z}\}$, and $E(\zeta_t^2) = \lambda^{\mathrm{T}}\Omega\lambda$, where Ω is defined as in Theorem 2.1 and $0<\lambda^{\mathrm{T}}\Omega\lambda<\infty$.

Let $S_i = T^{-1/2}\sum_{t=1}^{i}\zeta_t$. Note that the sequences $\{\zeta_t\}$ and $\{E({\zeta_t}^2|\mathcal{F}_{t-1})\}$ are both strictly stationary and ergodic, and $E{S_T}^2 = \lambda^{\mathrm{T}}\Omega\lambda$. Hence, it can be verified that

$$\frac{1}{T}\sum_{t=1}^{T}\frac{E({\zeta_t}^2|\mathcal{F}_{t-1})}{E{S_T}^2} \to 1 \tag{7}$$

in the almost surely sense and, for any $\epsilon>0$,

$$\frac{1}{T}\sum_{t=1}^{T}E[{\zeta_t}^2 I(\zeta_t \geq \sqrt{T\mathrm{var}(\zeta_t)}\epsilon)] \to 0, \tag{8}$$

as $T\to\infty$. The invariance principle for martingales [14], together with (7) and (8), implies that

$$S_{[T\tau]} = \frac{1}{\sqrt{T}}\sum_{t=1}^{[T\tau]}\zeta_t \Rightarrow W(\tau),$$

where $W(\tau)$ is a Brownian motion with variance $\tau\lambda^{\mathrm{T}}\Omega\lambda$. Thus, by Cramér's device,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[T\tau]}[\mathbf{e}_t^{\mathrm{T}}, \mathrm{sgn}(\boldsymbol{\varepsilon}_t)^{\mathrm{T}}, \mathrm{sgn}(\varepsilon_{1,t})\mathbf{z}_t^{\mathrm{T}},\dots,\mathrm{sgn}(\varepsilon_{N,t})\mathbf{z}_t^{\mathrm{T}}]^{\mathrm{T}} \Rightarrow \mathbf{B}(\tau), \tag{9}$$

and then we obtain (iv) and (v).

By Assumption 2.1, $\Pi(B) = (\mathbf{I}_N - \sum_{j=1}^{p}\Psi_j B^j)^{-1} = \sum_{j=0}^{\infty}\Pi_j B^j$ and $\sum_{j=0}^{\infty} j|\Pi_j|<\infty$, where B is the back shift operator and the Π_j's are $N\times N$ real matrices; see Chapter 18 of [15]. Note that $\boldsymbol{\mu}=0$. Let $\mathbf{u}_t = \Delta\mathbf{y}_t$, and then we have $\mathbf{u}_t = \Pi(B)\mathbf{e}_t$. We consider the Beveridge-Nelson representation as follows,

$$\mathbf{u}_t = \Upsilon\mathbf{e}_t + (\bar{\mathbf{u}}_{t-1} - \bar{\mathbf{u}}_t),$$

where $\Upsilon = \Pi(1)$, $\bar{\Pi}_j = \sum_{i=j+1}^{\infty}\Pi_i$ and $\bar{\mathbf{u}}_t = \sum_{j=0}^{\infty}\bar{\Pi}_j\mathbf{e}_{t-j}$. Thus,

$$\mathbf{y}_t = \Upsilon\sum_{i=1}^{t}\mathbf{e}_i + (\bar{\mathbf{u}}_0 - \bar{\mathbf{u}}_t). \tag{10}$$

It holds that $\sum_{j=0}^{\infty}|\bar{\Pi}_j|<\infty$, and then $\{\bar{\mathbf{u}}_t\}$ is an N-dimensional time series with strict stationarity. In view of (9) and (10), we complete the proof of (i)–(iii), and Theorem 2.2 in [19] makes sure that (i)–(v) hold jointly. □

Proof of Theorem 2.1. Define the objective functions

$$Q_T(\boldsymbol{\theta}) = \sum_{i=1}^{N} Q_{i,T}(\boldsymbol{\theta})$$

and

$$Q_{i,T}(\boldsymbol{\theta}) = \sum_{t=p+2}^{T} |\Delta y_{i,t} - \phi_i^{\mathrm{T}}\mathbf{y}_{t-1} - \mu_i - \psi_i^{\mathrm{T}}\mathbf{z}_t|,$$

where $i = 1, \ldots, N$ and $\mathbf{z}_t = (\Delta\mathbf{y}_{t-1}^{\mathrm{T}}, \ldots, \Delta\mathbf{y}_{t-p}^{\mathrm{T}})^{\mathrm{T}}$. For any $\mathbf{v}_i = (\mathbf{v}_{i1}^{\mathrm{T}}, \mathbf{v}_{i2}^{\mathrm{T}})^{\mathrm{T}}$, $\mathbf{v}_{i1} \in \mathbb{R}^N$ and $\mathbf{v}_{i2} \in \mathbb{R}^{pN+1}$, denote by $Q_{i,T}(\mathbf{v}_i)$ the function $Q_{i,T}(\boldsymbol{\theta}_i)$ evaluated at $\boldsymbol{\theta}_{0i} + (\mathbf{v}_{i1}^{\mathrm{T}}/T, \mathbf{v}_{i2}^{\mathrm{T}}/\sqrt{T})^{\mathrm{T}}$, where $\boldsymbol{\theta}_{0i} = (0, m_i, \psi_i^{\mathrm{T}})$ is the true value of the parameter vector $\boldsymbol{\theta}_i$, and 0 here denotes an N-dimensional zero vector. Let $\mathbf{v} = (\mathbf{v}_1^{\mathrm{T}}, \ldots, \mathbf{v}_N^{\mathrm{T}})^{\mathrm{T}}$ and $Q_T(\mathbf{v}) = \sum_{i=1}^{N} Q_{i,T}(\mathbf{v}_i)$.

Notice that

$$|x-y| - |x| = -y\,\mathrm{sgn}(x) + 2\int_0^y I(x \le s) - I(x \le 0)ds,$$

for $x, y \in \mathbb{R}$ and $x \neq 0$, where $\mathrm{sgn}(x)$ is equal to 1 for $x > 0$ and -1 for $x < 0$; see [18]. Thus, $Q_T(\mathbf{v}) - Q_T(\boldsymbol{\theta}_0) = \sum_{i=1}^{N}[Q_{i,T}(\mathbf{v}) - Q_{i,T}(\boldsymbol{\theta}_0)]$ and

$$Q_{i,T}(\mathbf{v}) - Q_{i,T}(\boldsymbol{\theta}_0) = \xi_i - \left(\frac{1}{T}\sum_{t=p+2}^{T}\mathbf{y}_{t-1}^{\mathrm{T}}\,\mathrm{sgn}(\varepsilon_{i,t}), \frac{1}{\sqrt{T}}\sum_{t=p+2}^{T}\bar{\mathbf{z}}_t^{\mathrm{T}}\,\mathrm{sgn}(\varepsilon_{i,t})\right)^{\mathrm{T}}\mathbf{v}_i,$$

where $\bar{\mathbf{z}}_t = (1, \mathbf{z}_t^{\mathrm{T}})^{\mathrm{T}}$ and

$$\xi_i = 2\sum_{t=p+2}^{T}\int_0^{\mathbf{v}_{i1}^{\mathrm{T}}\mathbf{y}_{t-1}/T + \mathbf{v}_{i2}^{\mathrm{T}}\bar{\mathbf{z}}_t/\sqrt{T}} I(\varepsilon_{i,t} \le s) - I(\varepsilon_{i,t} \le 0)ds.$$

Let

$$\xi_{1i} = 2\sum_{t=p+2}^{T}\int_0^{\mathbf{v}_{i1}^{\mathrm{T}}\mathbf{y}_{t-1}/T + \mathbf{v}_{i2}^{\mathrm{T}}\bar{\mathbf{z}}_t/\sqrt{T}} F_{i,t}(s) - F_{i,t}(0)ds$$

and

$$\xi_{2i} = \mathbf{v}_i^{\mathrm{T}}\begin{pmatrix} \widehat{\Gamma}_{11,i} & \widehat{\Gamma}_{12,i} \\ \widehat{\Gamma}_{12,i}^{\mathrm{T}} & T^{-1}\sum_t f_{i,t}(0)\bar{\mathbf{z}}_t\bar{\mathbf{z}}_t^{\mathrm{T}} \end{pmatrix}\mathbf{v}_i,$$

where $\widehat{\Gamma}_{11,i} = T^{-2}\sum_t f_{i,t}(0)\mathbf{y}_{t-1}\mathbf{y}_{t-1}^{\mathrm{T}}$, $\widehat{\Gamma}_{12,i} = T^{-3/2}\sum_t f_{i,t}(0)\mathbf{y}_{t-1}\bar{\mathbf{z}}_t^{\mathrm{T}}$, and $F_{i,t}(x)$ is the conditional distribution function of $\varepsilon_{i,t}$, i.e., $F_{i,t}(x) = P(\varepsilon_{i,t} \le x|\mathcal{F}_{t-1})$. We next show that $\xi_i = \xi_{1i} + o_p(1)$ and $\xi_{1i} = \xi_{2i} + o_p(1)$.

Define $\iota_T(t) = \mathbf{v}_{i1}^{\mathrm{T}}\mathbf{y}_{t-1}/T + \mathbf{v}_{i2}^{\mathrm{T}}\bar{\mathbf{z}}_t/\sqrt{T}$. By (10), $E(\mathbf{y}_t\mathbf{y}_t^{\mathrm{T}}) = t\Upsilon\Sigma_{\mathbf{e}}\Upsilon^{\mathrm{T}} + O(1)$, and then

$$\begin{aligned} &TE[\iota_T^2(t)] \\ &\le 2\mathbf{v}_{i1}^{\mathrm{T}}\Upsilon\Sigma_{\mathbf{e}}\Upsilon^{\mathrm{T}}\mathbf{v}_{i1} + 2\mathbf{v}_{i2}^{\mathrm{T}}\,\mathrm{diag}\{1, \Sigma_{\mathbf{z}}\}\mathbf{v}_{i2} + o(1) \\ &\le C_1 \end{aligned} \tag{11}$$

for $t > n_0$, where C_1 is a constant independent of T and n_0 is a large but fixed positive number. Note that the quantity $\xi_i - \xi_{1i}$ is the summation of a martingale difference sequence with respect to the filtration $\{\mathcal{F}_t, t \in \mathbb{Z}\}$. Thus, for any $\delta > 0$,

$$\begin{aligned} &0.25E(\xi_i - \xi_{1i})^2 \\ &\le \sum_{t=p+2}^{T} E\left\{\int_0^{\iota_T(t)}[I(\varepsilon_{i,t} \le s) - I(\varepsilon_{i,t} \le 0)]ds\right\}^2 \\ &= a_T(\delta) + \sum_{t=p+2}^{T} E\{\int_0^{\iota_T(t)}[I(\varepsilon_{i,t} \le s) - I(\varepsilon_{i,t} \le 0)]ds \\ &\qquad \cdot I(|\iota_T(t)| > \delta)\}^2 \\ &\le a_T(\delta) + TE[\iota_T^2(t)I(|\iota_T(t)| > \delta)], \end{aligned} \tag{12}$$

where

$$a_T(\delta) = \sum_{t=p+2}^{T} E\left\{\int_0^{\iota_T(t)}[I(\varepsilon_{i,t} \le s) - I(\varepsilon_{i,t} \le 0)]dsI(|\iota_T(t)| \le \delta)\right\}^2.$$

It is implied by (11) that, for any fixed $\delta > 0$,

$$TE[\iota_T^2(t)I(|\iota_T(t)| > \delta)] \to 0, \tag{13}$$

as $T \to \infty$. Furthermore, it can be verified that

$$\int_0^y I(x \le s) - I(x \le 0)ds = (y-x)I(0 < x < y) + (x-y)I(y < x < 0).$$

Hence, when $0 < \delta < \pi$, we can obtain that

$$\begin{aligned} a_T(\delta) &= \sum_{t=p+2}^{T} E[\iota_T(t) - \varepsilon_{i,t}]^2[I(0 < \varepsilon_{i,t} < \iota_T(t) \le \delta) \\ &\qquad + I(0 > \varepsilon_{l,t} > \iota_T(t) \ge -\delta)] \\ &\le \delta \cdot \frac{2}{3}TE[\iota_T^2(t)] \cdot C_2 \le \delta \cdot \frac{2}{3}C_1C_2, \end{aligned}$$

where C_1 is given as in (11) and $\sup_{|x| \le \pi, t \in \mathbb{Z}} f_{i,t}(x) \le C_2$ by Assumption 2.3. In light of (12) and (13), we can show that $E(\xi_i - \xi_{1i})^2 = o(1)$, which implies $\xi_i = \xi_{1i} + o_p(1)$. Note that, for $0 < \delta < \pi$,

$$\begin{aligned} \xi_{1i} - \xi_{2i} &= 2\sum_{t=p+2}^{T}\int_0^{\iota_T(t)} F_{i,t}(s) - F_{i,t}(0) - f_{i,t}(0)sds \\ &\le \sup_{|x| \le \delta, t \in \mathbb{Z}} |f_{i,t}(x) - f_{i,t}(0)| \sum_{t=p+2}^{T}\iota_T^2(t)I(|\iota_T(t)| \le \delta) \\ &\quad + \sum_{t=p+2}^{T}[|\iota_T(t)| + f_{i,t}(0)\iota_T^2(t)]I(|\sigma_{i,t}^{-1}\iota_T(t)| > \delta). \end{aligned}$$

Similarly, we can show that $\xi_{1i} = \xi_{2i} + o_p(1)$.

By the ergodic theorem, we have $T^{-1}\sum_{t=p+2}^T f_{i,t}(0)\bar{\mathbf{z}}_t\bar{\mathbf{z}}_t^{\mathrm{T}} = E[f_{i,t}(0)\bar{\mathbf{z}}_t\bar{\mathbf{z}}_t^{\mathrm{T}}] + o_p(1)$. By a method similar to the proof of Theorem 3.1 in [25], we can show that

$$\frac{1}{T^2}\sum_{t=p+2}^T f_{i,t}(0)\mathbf{y}_{t-1}\mathbf{y}_{t-1}^{\mathrm{T}} = E[f_{i,t}(0)]\cdot\frac{1}{T^2}\sum_{t=p+2}^T \mathbf{y}_{t-1}\mathbf{y}_{t-1}^{\mathrm{T}} + o_p(1)$$

and

$$\frac{1}{T\sqrt{T}}\sum_{t=p+2}^T f_{i,t}(0)\bar{\mathbf{z}}_t\mathbf{y}_{t-1}^{\mathrm{T}} = \frac{1}{T\sqrt{T}}E[f_{i,t}(0)\bar{\mathbf{z}}_t]\cdot\sum_{t=p+2}^T \mathbf{y}_{t-1}^{\mathrm{T}} + o_p(1),$$

which together with Lemma A.1 implies that

$$Q_{i,T}(\mathbf{v}_i) - Q_{i,T}(\boldsymbol{\theta}_0) \Rightarrow -\left[\int_0^1 \mathbf{B}_1^{\mathrm{T}}(\tau)d\mathbf{B}_2^{(i)}(\tau)\Upsilon^{\mathrm{T}}, \mathbf{B}_2^{(i)}(1), \mathbf{B}_{i+2}^{\mathrm{T}}(1)\right]^{\mathrm{T}}\mathbf{v}_i + \mathbf{v}_i^{\mathrm{T}}\Gamma_i\mathbf{v}_i,$$

and then

$$Q_T(\mathbf{v}) - Q_T(\boldsymbol{\theta}_0) \Rightarrow -\sum_{i=1}^N\left[\int_0^1 \mathbf{B}_1^{\mathrm{T}}(\tau)d\mathbf{B}_2^{(i)}(\tau)\Upsilon^{\mathrm{T}}, \mathbf{B}_2^{(i)}(1), \mathbf{B}_{i+2}^{\mathrm{T}}(1)\right]^{\mathrm{T}}\mathbf{v}_i + \sum_{i=1}^N \mathbf{v}_i^{\mathrm{T}}\Gamma_i\mathbf{v}_i.$$

Note that $Q_T(\boldsymbol{\theta})$ is a convex function with respect to $\boldsymbol{\theta}$. Hence, we complete the proof by following [18]. □

Proof of Theorem 3.1. By a method similar to the proof of Lemma A.1, we can show that, condition on $\mathbf{y}_1,\dots,\mathbf{y}_T$,

(14)
$$\frac{1}{T^2}\sum_{t=p+2}^T \mathbf{y}_{t-1}^*\mathbf{y}_{t-1}^{*\mathrm{T}} \Rightarrow \Upsilon\int_0^1 \mathbf{B}_1^*(\tau)\mathbf{B}_1^{*\mathrm{T}}(\tau)d\tau\Upsilon^{\mathrm{T}},$$
$$\frac{1}{T\sqrt{T}}\sum_{t=p+2}^T \mathbf{y}_{t-1}^* \Rightarrow \Upsilon\int_0^1 \mathbf{B}_1^*(\tau)d\tau,$$
$$\frac{1}{T}\sum_{t=p+2}^T(\omega_t-1)\operatorname{sgn}(\varepsilon_{i,t})\mathbf{y}_{t-1}^* \Rightarrow \Upsilon\int_0^1 \mathbf{B}_1^*(\tau)d\mathbf{B}_2^{*(i)}(\tau), \quad 1\le i\le N,$$
$$\frac{1}{\sqrt{T}}\sum_{t=p+2}^T(\omega_t-1)\operatorname{sgn}(\boldsymbol{\varepsilon}_t) \Rightarrow \mathbf{B}_2^*(1),$$
$$\frac{1}{\sqrt{T}}\sum_{t=p+2}^T(\omega_t-1)\operatorname{sgn}(\varepsilon_{i,t})\mathbf{z}_t \Rightarrow \mathbf{B}_{i+2}^*(1), \quad 1\le i\le N,$$

in probability, where $\mathbf{B}^*(\tau) = [\mathbf{B}_1^{*\mathrm{T}}(\tau),\dots,\mathbf{B}_{N+2}^{*\mathrm{T}}(\tau)]^{\mathrm{T}}$ is the (pN^2+2N)-dimensional Brownian motion defined in Theorem 3.1.

Let

$$Q_T^*(\boldsymbol{\theta}) = \sum_{i=1}^N Q_{i,T}^*(\boldsymbol{\theta})$$

and

$$Q_{i,T}^*(\boldsymbol{\theta}) = \sum_{t=p+2}^T \omega_t|\Delta y_{i,t} - \phi_i^{\mathrm{T}}\mathbf{y}_{t-1}^* - \mu_i - \psi_i^{\mathrm{T}}\mathbf{z}_t|.$$

We first show that

(15)
$$Q_{i,T}^*(\mathbf{v}_i) - Q_{i,T}^*(\boldsymbol{\theta}_0) = -\left(\frac{1}{T}\sum_{t=p+2}^T \omega_t\mathbf{y}_{t-1}^{*\mathrm{T}}\operatorname{sgn}(\varepsilon_{i,t}), \frac{1}{\sqrt{T}}\sum_{t=p+2}^T \omega_t\bar{\mathbf{z}}_t^{\mathrm{T}}\operatorname{sgn}(\varepsilon_{i,t})\right)^{\mathrm{T}}\mathbf{v}_i + \xi_{2i}^* + o_p^*(1),$$

where both $\mathbf{v} = (\mathbf{v}_1^{\mathrm{T}},\dots,\mathbf{v}_N^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\theta}_0$ are defined as in the proof of Theorem 2.1, $\widehat{\Gamma}_{11,i}^* = T^{-2}\sum_t \omega_t f_{i,t}(0)\mathbf{y}_{t-1}^*\mathbf{y}_{t-1}^{*\mathrm{T}}$, $\widehat{\Gamma}_{12,i}^* = T^{-3/2}\sum_t \omega_t f_{i,t}(0)\mathbf{y}_{t-1}^*\bar{\mathbf{z}}_t^{\mathrm{T}}$, and

$$\xi_{2i}^* = \mathbf{v}_i^{\mathrm{T}}\begin{pmatrix}\widehat{\Gamma}_{11,i}^* & \widehat{\Gamma}_{12,i}^* \\ \widehat{\Gamma}_{12,i}^{*\mathrm{T}} & T^{-1}\sum_t \omega_t f_{i,t}(0)\bar{\mathbf{z}}_t\bar{\mathbf{z}}_t^{\mathrm{T}}\end{pmatrix}\mathbf{v}_i.$$

Denote the σ-field $\mathcal{F}_t = \sigma(\boldsymbol{\varepsilon}_t,\boldsymbol{\varepsilon}_{t-1},\dots)$ and $\mathcal{F}_t^* = \sigma(\omega_t,\omega_{t-1},\dots,\omega_1)$. Let $\iota_T^*(t) = \mathbf{v}_{i1}^{\mathrm{T}}\mathbf{y}_{t-1}^*/T + \mathbf{v}_{i2}^{\mathrm{T}}\bar{\mathbf{z}}_t/\sqrt{T}$,

$$\xi_i^* = 2\sum_{t=p+2}^T \omega_t\int_0^{\iota_T^*(t)} I(\varepsilon_{i,t}\le s) - I(\varepsilon_{i,t}\le 0)ds,$$

and

$$\xi_{1i}^* = 2\sum_{t=p+2}^T \omega_t\int_0^{\iota_T^*(t)} F_{i,t}(s) - F_{i,t}(0)ds,$$

where $F_{i,t}(x)$ is the conditional distribution function of $\varepsilon_{i,t}$. Then

$$\begin{aligned}TE\{[\iota_T^*(t)]^2|\mathcal{F}_T^*\} &\le 2\mathbf{v}_{i1}^{\mathrm{T}}\widehat{\Upsilon}\left(t^{-1}\sum_{i=1}^t \widehat{\mathbf{e}}_i\widehat{\mathbf{e}}_i^{\mathrm{T}}\right)\widehat{\Upsilon}^{\mathrm{T}}\mathbf{v}_{i1} + 2\mathbf{v}_{i2}^{\mathrm{T}}\bar{\mathbf{z}}_t\bar{\mathbf{z}}_t^{\mathrm{T}}\mathbf{v}_{i2} \\ &= 2\mathbf{v}_{i1}^{\mathrm{T}}\Upsilon\Sigma_{\mathbf{e}}\Upsilon^{\mathrm{T}}\mathbf{v}_{i1} + 2\mathbf{v}_{i2}^{\mathrm{T}}\operatorname{diag}\{1,\Sigma_{\mathbf{z}}\}\mathbf{v}_{i2} + o_p(1),\end{aligned}$$

which implies that

$$TE\{[\iota_T^*(t)]^2 I(|\iota_T^*(t)| > \delta)|\mathcal{F}_T^*\} = o_p(1) \tag{16}$$

for any $\delta > 0$. By a method similar to the proof of Theorem 2.1, we can show that

$$\begin{aligned} a_T^*(\delta) = \sum_{t=p+2}^{T} E\{\int_0^{\iota_T^*(t)} [I(\varepsilon_{i,t} \le s) - I(\varepsilon_{i,t} \le 0)]ds \\ I(|\iota_T^*(t)| \le \delta)\}^2 \\ \le \delta \cdot \frac{2}{3} TE[\iota_T^*(t)]^2 \cdot \sup_{|x|\le\pi, t\in\mathbb{Z}} f_{i,t}(x) \end{aligned} \tag{17}$$

for $0 < \delta < \pi$. Note that the quantity $\xi_i - \xi_{1i}$ is the summation of a martingale difference sequence with respect to the filtration $\{\mathcal{F}_t \cup \mathcal{F}_T^*, t \in \mathbb{Z}\}$. In view of (16) and (17), we have that, for any $\delta > 0$,

$$\begin{aligned} 0.25E[(\xi_i^* - \xi_{1i}^*)^2|\mathcal{F}_T^*] \le a_T(\delta) + TE[\iota_T^2(t) I(|\iota_T(t)| > \delta)] \\ = o_p(1) \end{aligned}$$

and hence $\xi_i^* = \xi_{1i}^* + o_p^*(1)$. Similarly, it can be shown that $\xi_{1i}^* = \xi_{2i}^* + o_p^*(1)$, and we finish the proof of (15).

For the components of the matrix ξ_{2i}^*, we have

$$\begin{aligned} \frac{1}{T^2}\sum_{t=p+2}^{T} \omega_t f_{i,t}(0)\mathbf{y}_{t-1}^*\mathbf{y}_{t-1}^{*\mathrm{T}} \\ = E[f_{i,t}(0)]\frac{1}{T^2} \cdot \sum_{t=p+2}^{T} \mathbf{y}_{t-1}^*\mathbf{y}_{t-1}^{*\mathrm{T}} + o_p^*(1), \end{aligned}$$

$$\begin{aligned} \frac{1}{T\sqrt{T}}\sum_{t=p+2}^{T} \omega_t f_{i,t}(0)\bar{\mathbf{z}}_t\mathbf{y}_{t-1}^* \\ = E[f_{i,t}(0)\bar{\mathbf{z}}_t] \cdot \frac{1}{T\sqrt{T}} \sum_{t=p+2}^{T} \mathbf{y}_{t-1}^* + o_p^*(1) \end{aligned}$$

and

$$\frac{1}{T}\sum_{t=p+2}^{T} \omega_t f_{i,t}(0)\bar{\mathbf{z}}_t\bar{\mathbf{z}}_t^{\mathrm{T}} = E[f_{i,t}(0)\bar{\mathbf{z}}_t\bar{\mathbf{z}}_t^{\mathrm{T}}] + o_p^*(1);$$

see the proof of Theorem 3.1 in [25]. By (15), we have $\xi_{2i}^* = \mathbf{v}_i^{\mathrm{T}}\Gamma_{i,T}^*\mathbf{v}_i$ and

$$\begin{pmatrix} T\widehat{\phi}_{i,2}^* \\ \sqrt{T}(\widehat{\mu}_{i,2}^* - m_i) \\ \sqrt{T}(\widehat{\psi}_{i,2}^* - \psi_i) \end{pmatrix} = 0.5(\Gamma_{i,T}^*)^{-1} \begin{pmatrix} T^{-1}\sum_{t=p+2}^{T} \omega_t \mathbf{y}_{t-1}^{*\mathrm{T}} \operatorname{sgn}(\varepsilon_{i,t}) \\ T^{-1/2}\sum_{t=p+2}^{T} \omega_t \bar{\mathbf{z}}_t^{\mathrm{T}} \operatorname{sgn}(\varepsilon_{i,t}) \end{pmatrix} + o_p^*(1), \tag{18}$$

where $\Gamma_{11,i,T}^* = E[f_{i,t}(0)] \cdot T^{-2}\sum_t \mathbf{y}_{t-1}^*\mathbf{y}_{t-1}^{*\mathrm{T}}$, $\Gamma_{12,i,T}^* = T^{-3/2}\sum_t \mathbf{y}_{t-1}^* \cdot E[f_{i,t}(0)\bar{\mathbf{z}}_t^{\mathrm{T}}]$, and

$$\Gamma_{i,T}^* = \begin{pmatrix} \Gamma_{11,i,T}^* & \Gamma_{12,i,T}^* \\ \Gamma_{12,i,T}^{*\mathrm{T}} & E[f_{i,t}(0)\bar{\mathbf{z}}_t\bar{\mathbf{z}}_t^{\mathrm{T}}] \end{pmatrix},$$

Similarly, it can be shown that

$$\begin{pmatrix} T\widehat{\phi}_{i,1}^* \\ \sqrt{T}(\widehat{\mu}_{i,1}^* - m_i) \\ \sqrt{T}(\widehat{\psi}_{i,1}^* - \psi_i) \end{pmatrix} = 0.5(\Gamma_{i,T}^*)^{-1} \begin{pmatrix} T^{-1}\sum_{t=p+2}^{T} \mathbf{y}_{t-1}^{*\mathrm{T}} \operatorname{sgn}(\varepsilon_{i,t}) \\ T^{-1/2}\sum_{t=p+2}^{T} \bar{\mathbf{z}}_t^{\mathrm{T}} \operatorname{sgn}(\varepsilon_{i,t}) \end{pmatrix} + o_p^*(1).$$

Combining the above two equations, together with (14), we finish the proof of this theorem. □

Proof of Theorem 4.1. We first prove the result for the test statistics F_{HE1} and F_{HE2}. By Lemma A.1, the following results hold jointly for $1 \le i \le N$:

$$\begin{aligned} &\frac{1}{T\sqrt{T}}\sum_{t=p_i+2}^{T} y_{i,t-1} \Rightarrow \varphi_i \int_0^1 \mathbf{W}_1^{(i)}(\tau)d\tau, \\ &\frac{1}{\sqrt{T}}\sum_{t=p_i+2}^{T} \operatorname{sgn}(\varepsilon_{i,t}) \Rightarrow \mathbf{W}_2^{(i)}(1), \\ &\frac{1}{\sqrt{T}}\sum_{t=p_i+2}^{T} \operatorname{sgn}(\varepsilon_{i,t})\mathbf{z}_{i,t} \Rightarrow \mathbf{W}_{i+2}(1), \\ &\frac{1}{T}\sum_{t=p_i+2}^{T} \operatorname{sgn}(\varepsilon_{i,t})y_{i,t-1} \Rightarrow \varphi_i \int_0^1 \mathbf{W}_1^{(i)}(\tau)d\mathbf{W}_2^{(i)}(\tau), \\ &\frac{1}{T^2}\sum_{t=p_i+2}^{T} \mathbf{y}_{t-1}\mathbf{y}_{t-1}^{\mathrm{T}} \Rightarrow \operatorname{diag}\{\varphi_1, \dots, \varphi_N\} \\ &\qquad \int_0^1 \mathbf{W}_1(\tau)\mathbf{W}_1^{\mathrm{T}}(\tau)d\tau \operatorname{diag}\{\varphi_1, \dots, \varphi_N\}, \end{aligned} \tag{19}$$

where $\mathbf{y}_t = (y_{1,t}, \dots, y_{N,t})^{\mathrm{T}}$, $\varphi_i = (1 - \sum_{j=1}^{p_i} \psi_{i,j})^{-1}$, $\mathbf{z}_{i,t} = (\Delta y_{i,t-1}, \dots, \Delta y_{i,t-p_i})^{\mathrm{T}}$, and $\mathbf{W}(\tau) = [\mathbf{W}_1^{\mathrm{T}}(\tau), \dots, \mathbf{W}_{N+2}^{\mathrm{T}}(\tau)]^{\mathrm{T}}$ is the Brownian motion defined in Theorem 4.1. Thus,

$$\begin{aligned} &\frac{1}{T^2}B_i \Rightarrow \varphi_i^2 \left\{\int_0^1 [\mathbf{W}_1^{(i)}(\tau)]^2 d\tau - \left[\int_0^1 \mathbf{W}_1^{(i)}(\tau)d\tau\right]^2\right\}, \\ &\frac{1}{\sqrt{T}}\mathbf{A}_i \Rightarrow \varphi_i \left(\int_0^1 \mathbf{W}_1^{(i)}(\tau)d\tau, 0, \dots, 0\right), \\ &\frac{1}{T^2}\sum_{t=p+2}^{T} (y_{i,t-1} - \mathbf{A}_i\bar{\mathbf{z}}_{i,t})(y_{j,t-1} - \mathbf{A}_j\bar{\mathbf{z}}_{j,t}) \\ &\qquad \Rightarrow \varphi_i\varphi_j \left\{\int_0^1 \mathbf{W}_1^{(i)}(\tau)\mathbf{W}_1^{(j)}(\tau)d\tau\right. \end{aligned}$$

$$-\int_0^1 \mathbf{W}_1^{(i)}(\tau)d\tau \int_0^1 \mathbf{W}_1^{(j)}(\tau)d\tau\biggr\},$$

where $\bar{\mathbf{z}}_{i,t} = (1, \mathbf{z}_{i,t}^{\mathrm{T}})^{\mathrm{T}}$. Then it follows that

$$\frac{1}{T^2}\widehat{\mathrm{var}}(\widehat{\Phi}) \Rightarrow \Sigma_{\widehat{\phi}}. \tag{20}$$

Denote the objective functions of model (4) by

$$Q_T(\boldsymbol{\theta}) = \sum_{i=1}^{N} Q_{i,T}(\boldsymbol{\theta})$$

and

$$Q_{i,T}(\boldsymbol{\theta}) = \sum_{t=p+2}^{T} |\Delta y_{i,t} - \phi_i y_{i,t-1} - \mu_i - \psi_i^{\mathrm{T}}\mathbf{z}_{i,t}|,$$

where $1 \le i \le N$ and $\psi_i = (\psi_{i,1}, \ldots, \psi_{i,p_i})^{\mathrm{T}}$. For any $\mathbf{v}_i = (v_{i1}^{\mathrm{T}}, \mathbf{v}_{i2}^{\mathrm{T}})^{\mathrm{T}}$, $v_{i1} \in \mathbb{R}$ and $\mathbf{v}_{i2} \in \mathbb{R}^{p_i+1}$, denote by $Q_{i,T}(\mathbf{v}_i)$ the value of the function $Q_{i,T}(\boldsymbol{\theta}_i)$ at $\boldsymbol{\theta}_{0i} + (v_{i1}^{\mathrm{T}}/T, \mathbf{v}_{i2}^{\mathrm{T}}/\sqrt{T})^{\mathrm{T}}$, where $\boldsymbol{\theta}_{0i}$ is the true value of the parameter vector of model (4). Let $\mathbf{v} = (\mathbf{v}_1^{\mathrm{T}}, \ldots, \mathbf{v}_N^{\mathrm{T}})^{\mathrm{T}}$ and $Q_T(\mathbf{v}) = \sum_{i=1}^N Q_{i,T}(\mathbf{v}_i)$. By (19), [18] and a method similar to the proof of Theorem 2.1, we have

$$\begin{aligned}
&Q_{i,T}(\mathbf{v}_i) - Q_{i,T}(\boldsymbol{\theta}_{0i}) \\
&\quad= (v_{i1}, \mathbf{v}_{i2}^{\mathrm{T}}) \begin{pmatrix} T^{-2}\sum_t y_{i,t-1}^2 f_{i,t}(0) & \mathbf{H}_i \\ \mathbf{H}_i^{\mathrm{T}} & \mathbf{G}_i \end{pmatrix} \begin{pmatrix} v_{i1} \\ \mathbf{v}_{i2} \end{pmatrix} \\
&\qquad - v_{i1}\frac{1}{T}\sum_t y_{i,t-1}\,\mathrm{sgn}(\varepsilon_{i,t}) \\
&\qquad - \mathbf{v}_{i2}^{\mathrm{T}}\frac{1}{\sqrt{T}}\sum_t \bar{\mathbf{z}}_{i,t}\,\mathrm{sgn}(\varepsilon_{i,t}) + o_p(1),
\end{aligned}$$

and then $T\widehat{\phi}_i \Rightarrow \zeta_{\mathbf{W}}(i)$, where $\mathbf{H}_i = T^{-3/2}\sum_t y_{i,t-1}\bar{\mathbf{z}}_{i,t}^{\mathrm{T}} f_{i,t}(0)$ and $\mathbf{G}_i = T^{-1}\sum_t \bar{\mathbf{z}}_{i,t}\bar{\mathbf{z}}_{i,t}^{\mathrm{T}} f_{i,t}(0)$. In view of (20), we complete the proof for the test statistics F_{HE1} and F_{HE2}.

For the asymptotic distribution of the test statistic F_{HO}, denote by $Q_T(\mathbf{v})$ the corresponding objective function, where $v_1 = v_{i1}$ for $1 \le i \le N$ and $\mathbf{v} = (v_1, \mathbf{v}_{12}^{\mathrm{T}}, \ldots, \mathbf{v}_{N2}^{\mathrm{T}})^{\mathrm{T}}$. By a method similar to the proof of Theorem 2.1, we can show that

$$\begin{aligned}
&Q_T(\mathbf{v}) - Q_T(\boldsymbol{\theta}_0) \\
&\quad= \mathbf{v}\begin{pmatrix} \sum_{i=1}^N T^{-2}\sum_t y_{i,t-1}^2 f_{i,t}(0) & \mathbf{H} \\ \mathbf{H}^{\mathrm{T}} & \mathbf{G} \end{pmatrix}\mathbf{v}^{\mathrm{T}} \\
&\qquad - v_1\sum_{i=1}^N \frac{1}{T}\sum_t y_{i,t-1}\,\mathrm{sgn}(\varepsilon_{i,t}) \\
&\qquad - \sum_{i=1}^N \mathbf{v}_{i2}^{\mathrm{T}}\frac{1}{\sqrt{T}}\sum_t \bar{\mathbf{z}}_{i,t}\,\mathrm{sgn}(\varepsilon_{i,t}) + o_p(1),
\end{aligned}$$

where $\mathbf{H} = (\mathbf{H}_1, \ldots, \mathbf{H}_N)$ and $\mathbf{G} = \mathrm{diag}\{\mathbf{G}_1, \ldots, \mathbf{G}_N\}$. Note that the function $Q_T(\mathbf{v})$ is still convex with respect to $\mathbf{v}$. By (19), [18] and elementary calculations, we complete the proof of the theorem. □

Proof of Theorem 4.2. Compared with the bootstrap mechanism in Section 3, there is one more term, $\widehat{\phi}_i y_{i,t-1}$ or $\widetilde{\phi} y_{i,t-1}$, involved in this theorem. We first verify two important facts to handle the extra term, say $\widehat{\phi}_i y_{i,t-1}$.

The first one is that, for $1 \le i \le N$, $\widehat{\phi}_i = O_p(T^{-1})$ and $\max_{1\le t\le T}|y_{i,t}| = O_p(T^{-1/2})$ if H_0 holds; see Theorem 4.1 and [23]. However, if H_1 holds, then $\widehat{\phi}_i - \phi_i = O_p(T^{-1/2})$, where $\phi_i \ne 0$ is the true value of the corresponding parameter.

The second one provides a way of decomposition. For example, for $1 \le i \le N$, define the objective function related to the estimator $\widehat{\boldsymbol{\theta}}_{i,2}^*$ as

$$Q_{i,T}^*(\boldsymbol{\theta}) = \sum_{t=p+2}^{T} \omega_t|\Delta y_{i,t} - \phi_i y_{i,t-1}^* - \mu_i - \psi_i^{\mathrm{T}}\mathbf{z}_{i,t} - \widehat{\phi}_i y_{i,t-1}|.$$

Thus,

$$\begin{aligned}
&Q_{i,T}^*(\mathbf{v}_i) - Q_{i,T}^*(\boldsymbol{\theta}_0) \\
&\quad= \sum_{t=p_i+2}^{T}\omega_t\Big|\varepsilon_{i,t} - \frac{v_{i1}}{T}y_{i,t-1}^* - \frac{\mathbf{v}_{i2}}{\sqrt{T}}\bar{\mathbf{z}}_{i,t} - \widehat{\phi}_i y_{i,t-1}\Big| \\
&\qquad\qquad - \sum_{t=p_i+2}^{T}\omega_t|\varepsilon_{i,t} - \widehat{\phi}_i y_{i,t-1}| \\
&\quad= \left(\sum_{t=p_i+2}^{T}\omega_t\Big|\varepsilon_{i,t} - \frac{v_{i1}}{T}y_{i,t-1}^* - \frac{\mathbf{v}_{i2}}{\sqrt{T}}\bar{\mathbf{z}}_{i,t} - \widehat{\phi}_i y_{i,t-1}\Big| - \sum_{t=p_i+2}^{T}\omega_t|\varepsilon_{i,t}|\right) \\
&\qquad - \left(\sum_{t=p_i+2}^{T}\omega_t|\varepsilon_{i,t} - \widehat{\phi}_i y_{i,t-1}| - \sum_{t=p_i+2}^{T}\omega_t|\varepsilon_{i,t}|\right),
\end{aligned}$$

where $\boldsymbol{\theta}_0$ is the true value of the parameter vector, and $\mathbf{v}_i = (v_{i1}, \mathbf{v}_{i2}^{\mathrm{T}})^{\mathrm{T}}$ is defined as in the proof of Theorem 4.1. By a method similar to the proof of Theorem 3.1, together with the first fact mentioned at the beginning of this proof, we establish a result similar to (18). Repeatedly using the above two facts, we complete the proof by a method similar to that of Theorem 3.1. □

ACKNOWLEDGEMENTS

We are deeply grateful to the Editor, the Associate Editor and two referees for their valuable comments that led to substantial improvement in the quality of this paper. This research was partially supported by the National Nature

Science Foundation of China (72173086 and 72033002) and Hong Kong RGC grant 17306519.

Received 1 March 2021

REFERENCES

[1] Bai, J. and S. Ng (2004). A panic attack on unit roots and cointegration. *Econometrica 72*, 1127–1177. MR2064710

[2] Baltagi, B. H., G. Bresson, and A. Pirotte (2007). Panel unit root tests and spatial dependence. *Journal of Applied Econometrics 22*, 339–360. MR2364409

[3] Banerjee, A. and M. Wagner (2009). *Palgrave Handbook of Econometrics*, Volume Volume 2: Applied Econometrics, Chapter 13. Panel Methods to Test for Unit Roots and Cointegration. Palgrave Macmillan UK.

[4] Billingsley, P. (1999). *Convergence of Probability Measures* (2nd ed.). New York: Wiley. MR0233396

[5] Bollerslev, T. (1986). Generalized autoregression conditional heteroscedasticity. *Journal of Econometrics 31*, 307–327. MR0853051

[6] Breitung, J. and S. Das (2008). Testing for unit roots in panels with a factor structure. *Econometric Theory 24*, 88–108. MR2408860

[7] Chang, Y. (2004). Bootstrap unit root tests in panels with cross-sectional dependency. *Journal of Econometrics 120*, 263–293. MR2058890

[8] Chatterjee, S. and A. Bose (2005). Generalized bootstrap for estimating equations. *The Annals of Statistics 33*, 414–436. MR2157808

[9] Chen, K., Z. Ying, H. Zhang, and L. Zhao (2008). Analysis of least absolute deviation. *Biometrika 95*, 107–122. MR2394499

[10] Efron, B. and R. Tibshirani (1993). *An Introduction to the Bootstrap*. Boca Raton: Chapman & Hall/CRC. MR1270903

[11] Engle, R. F. (1982). Autoregression conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica 50*, 987–1008. MR0666121

[12] Engle, R. F. and C. W. J. Granger (1987). Co-integrated and error correction: representation, estimation and testing. *Econometrica 55*, 251–276. MR0882095

[13] Francq, C. and J. M. Zakoian (2010). GARCH *models*. New York: Wiley. MR3186556

[14] Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and Its Application*. New York: Academic Press. MR0624435

[15] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press. MR1278033

[16] Herce, M. A. (1996). Asymptotic theory of LAD estimation in a unit root process with finite variance errors. *Econometric Theory 12*, 129–153. MR1396383

[17] Jin, Z., Z. Ying, and L. J. Wei (2001). A simple resampling method by perturbing the minimand. *Biometrika 88*, 381–390. MR1844838

[18] Knight, K. (1998). Limiting distributions for l_1 regression estimators under general conditions. *The Annals of Statistics 26*, 755–770. MR1626024

[19] Kurtz, T. G. and P. Protter (1991). Weak limit theorems to stochastic integrals and stochastic differential equations. *The Annals of Probability 19*, 1035–1070. MR1112406

[20] Levin, A. and C. Lin (1992). Unit root test in panel data: asymptotic and finite sample properties. Discussion paper, *University of California at San Diego*, 92–93.

[21] Levin, A. and C. Lin (1993). Unit root test in panel data: new results. Discussion paper, *University of California at San Diego*, 93–56.

[22] Li, G. and W. K. Li (2008). Least absolute deviation estimation for fractionally integrated autoregressive moving average time series models with conditional heteroscedasticity. *Biometrika 95*, 399–414. MR2521590

[23] Li, G. and W. K. Li (2009). Least absolute deviation estimation for unit root processes with GARCH errors. *Econometric Theory 25*, 1208–1227. MR2540498

[24] Li, G. and W. K. Li (2011). Testing a linear time series models against its threshold extension. *Biometrika 98*, 243–250. MR2804225

[25] Ling, S. and W. K. Li (2003). Asymptotic inference for unit root processes with GARCH(1,1) errors. *Econometric Theory 19*, 541–564. MR1997932

[26] Mikosch, T. and C. Starica (2000). Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *The Annals of Statistics 28*, 1427–1451. MR1805791

[27] Mittnik, S., S. T. Rachev, and M. S. Paolella (1998). Stable paretian models in finance: some empirical and theoretical aspects. In R. J. Adler, R. E. Feldman, and M. S. Taqqu (Eds.), *A Practical Guide to Heavy Tails*. Boston: Birkhauser. MR1652284

[28] O'Connell, P. G. J. (1998). The overvalution of purchasing power parity. *Journal of International Economics 44*, 1–19.

[29] Palm, F. C., S. Smeekes, and J. P. Urbain (2008). Bootstrap unit-root tests: comparison and extensions. *Journal of Time Series Analysis 29*, 371–401. MR2392778

[30] Palm, F. C., S. Smeekes, and J. P. Urbain (2011). Cross-sectional dependence robust block bootstrap panel unit root tests. *Journal of Econometrics 163*, 85–104. MR2803668

[31] Paparoditis, E. and D. N. Politis (2005). Bootstrapping unit root tests for autoregressive time series. *Journal of the American Statistical Association 100*, 545–553. MR2160558

[32] Peng, L. and Q. Yao (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika 90*, 967–975. MR2024770

[33] Phillips, P. C. B. and P. Perron (1988). Testing for a unit root in time series regression. *Biometrika 75*, 335–346. MR0946054

[34] Phillips, P. C. B. and D. Sul (2003). Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometric Journal 6*, 217–259. MR1993440

[35] Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics 14*, 1261–1350. MR0868303

[36] Y. Zheng, W. K. Li, and G. Li. A robust goodness-of-fit test for generalized autoregressive conditional heteroscedastic models. *Biometrika*, 105:73–89, 2018. MR3768866

[37] Y. Zheng, Q. Zhu, G. Li, and Z. Xiao. Hybrid quantile regression estimation for time series models with conditional heteroscedasticity. *Journal of the Royal Statistical Society: Series B*, 80:975–993, 2018. MR3874306

[38] Zhu, L. X. and K. W. Ng (2003). Checking the adequacy of a partial linear model. *Statistica Sinica 13*, 763–781. MR1997173

Yao Zheng
Department of Statistics
University of Connecticut
215 Glenbrook Rd. U-4120, Storrs, CT
US
E-mail address: yao.zheng@uconn.edu

Jianhong Wu
Department of Mathematics
Shanghai Normal University
100 Guilin Road, Shanghai
China
E-mail address: wujianhong@shnu.edu.cn

Wai Keung Li
Department of Mathematics & Information Technology
Education University of Hong Kong
10 Lo Ping Road, Tai Po, New Territories, Hong Kong
China
E-mail address: waikeungli@eduhk.hk

Guodong Li
Department of Statistics & Actuarial Science
University of Hong Kong
Pokfulam Road, Hong Kong
China
E-mail address: gdli@hku.hk