Statistica Sinica Preprint No: SS-2020-0415	
Title	Tests of Unit Root Hypothesis With Heavy-Tailed Heteroscedastic Noises
Manuscript ID	SS-2020-0415
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202020.0415
Complete List of Authors	Rui She
Corresponding Author	Rui She
E-mail	rshe@swufe.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

Tests of Unit Root Hypothesis with Heavy-tailed Heteroscedastic Noises

Rui She

Southwestern University of Finance and Economics

Abstract: This study examines unit-root testing with unspecified and heavy-tailed heteroscedastic noise. A new weighted least squares estimation (WLSE) is designed for the Dickey–Fuller (DF) test, the asymptotic normality of which is verified. However, the performance of the DF test relies strongly on the estimation accuracy of the asymptotic variance, which is not stable for dependent time series. To overcome this issue, we develop two novel unit-root tests by applying the empirical likelihood technique to the WLSE score equation. We show that both empirical likelihood-based tests converge weakly to a chi-squared distribution with one degree of freedom. Furthermore, the limiting theory is extended to the weighted M-estimation score equation. In contrast to existing unit-root tests for heavy-tailed time series, empirical likelihood tests do not involve any estimators of the unknown parameters or any restrictions on the tail index of the noise. This makes them appealing in practice, with wide applications in finance and econometrics. Extensive simulation studies are conducted to examine the effectiveness of the proposed methods.

Key words and phrases: G/ARCH type noise; Heavy-tailed; Unit-root; Empirical likelihood.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

1. Introduction

Consider the following AR(1) model:

$$y_t = \phi y_{t-1} + \varepsilon_t, \tag{1.1}$$

where the noise $\{\varepsilon_t\}$ is a sequence of stationary random variables. We are interested in detecting a possible unit root in model (1.1); that is, we test the null hypothesis $H_0 : \phi = 1$ versus the alternative $H_1 : |\phi| < 1$. There is an extensive and relatively complete body of literature on unit-root estimation and testing when $E\varepsilon_t^2$ is finite. When the noise $\{\varepsilon_t\}$ is an independent and identically distributed (i.i.d.) random variable, Dickey and Fuller (1979) and Evans and Savin (1981) proposed the classical Dickey–Fuller (DF) test and Student's t test, respectively, based on the ordinary least squares estimator (LSE) of the regression parameter. Phillips (1987) further studied these tests and established the corresponding limiting theory when the noise is strong-mixing. For a concise review on this topic, see Chan (2009).

In the past two decades, a growing number of empirical studies have documented heavy-tailed noise in financial markets. Koedijk and Kool (1992) studied the exchange rate returns for three East European currencies, and found that their tail indices are smaller than two. Francq and Zakoïan (2013) investigated nine major financial markets, arguing that time series modeling driven by heavy-tailed noise may be more appropriate for financial data analyses; see Rachev (2003) and She and Ling (2020), among many others. All previous findings show that there is a practical and

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

urgent need to study heavy-tailed time series. Moreover, unit-root detection for models with heavy-tailed innovations is of practical importance.

However, when the noise is heavy-tailed (i.e., $E\varepsilon_t^2 = \infty$), a unit-root inference is much more complicated and challenging, even for the i.i.d. case. For instance, Chan and Tran (1989) studied the DF test when ε_t lies in the domain of attraction of a stable law with tail index $\alpha < 2$, such that they have an infinite variance. They found that, compared with the finite-variance case, the limiting distributions of the classical DF tests are no longer pivotal, because they depend on the unknown tail index of the noise, which is very difficult to estimate properly in practice (Resnick, 1997). To bypass the problem in heavy-tailed time series, one popular approach is to use the bootstrap or subsampling method to approximate the critical values. For example, Cavaliere, Georgiev, and Taylor (2018) proposed a sieve wild bootstrap method to obtain the null distribution of the augmented DF (ADF) test when the noise is a linear process driven by i.i.d. heavy-tailed innovations; see Horváth and Kokoszka (2003) and Moreno and Romo (2012) for early work. Zhang and Chan (2020) extended the results in Cavaliere, Georgiev, and Taylor (2018) to the case where the noise is from a standard GARCH model. However, their simulation results indicate that the aforementioned wild bootstrap method cannot deal with heavy-tailed GARCH noise. Recently, Huang et al. (2020) proposed a novel empirical-likelihood-based method to construct a unified test for model (1.1), with

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

the noise following the standard GARCH(p, q) model, namely,

$$\varepsilon_t = \eta_t h_t, \quad h_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}^2, \tag{1.2}$$

where $\omega > 0, \alpha_i \geq 0$, and $\beta_j \geq 0$. The core part of their work is to bound the possible heavy tail of h_t by some weighting or normalizing function so that the heavy tail effect in h_t is eliminated, leading to the robustness of their test. Using the empirical likelihood technique, their test also removes the estimation for the nuisance parameters. Nonetheless, the form of their normalizer may rely on the specific structure of h_t in (1.2), making it infeasible in cases without a priori knowledge on the structure of h_t. Furthermore, the conditions imposed on the moments of η_t and h_t (e.g., $E\eta_t^2 < \infty$ and $Eh_t < \infty$) are relatively restrictive, and exclude the classical heavy-tailed i.i.d. case ($E\eta_t^2 = \infty$ and $h_t = 1$). The heavy-tailed case with $Eh_t = \infty$ is also indispensable, even under the condition $E\eta_t^2 < \infty$, as illustrated by the domain D_2 in Figure 1. Therefore, these issues identified in Huang et al. (2020) motivate us to construct a unified unit-root test for model (1.1) or its extensions that is free of strong moment conditions on the noise and does not require a priori information on the structure of h_t.

In this study, we examine the unit-root process with unspecified and heavy-tailed heteroscedastic noise. To address the foregoing issues, a new weighted least squares estimation (WLSE) is proposed and embedded in the traditional DF test. We show that the derived DF-type test converges in distribution to a normal distribution un-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

der the null hypothesis, and to negative infinity under the alternative. However, the performance of the new DF test is vulnerable to the estimation accuracy of the asymptotic variance, which is not stable for strongly dependent time series. We thus develop two novel unit-root tests by applying the empirical likelihood technique to the WLSE score equations. Both empirical likelihood-based tests are shown to be asymptotically chi-squared with power approaching one. The corresponding asymptotic theory is also extended to the general weighted M-estimation score equations. As expected, our unit-root tests remove the estimations of the regression parameters, tail index of the noise, and structure of the heteroscedasticity, and thus have a broader application in finance and econometrics. In addition, the proposed tests can be used in more general settings, such as the unit root with a constant term and the unit root in the AR(r) model. A simulation study is conducted to demonstrate the performance of the proposed tests.

The rest of the paper is organized as follows. Section 2 gives a fundamental assumption and studies the DF-type test based on the new WLSE. Section 3 derives the asymptotic properties of the proposed unit-root tests using the standard empirical likelihood method and the adjusted empirical likelihood method. Extensions to more general unit-root models are presented in Section 4. The results of the simulation studies and a comparison with existing tests are summarized in Section 5. The technical proofs of the main results are given in the Supplementary Material.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

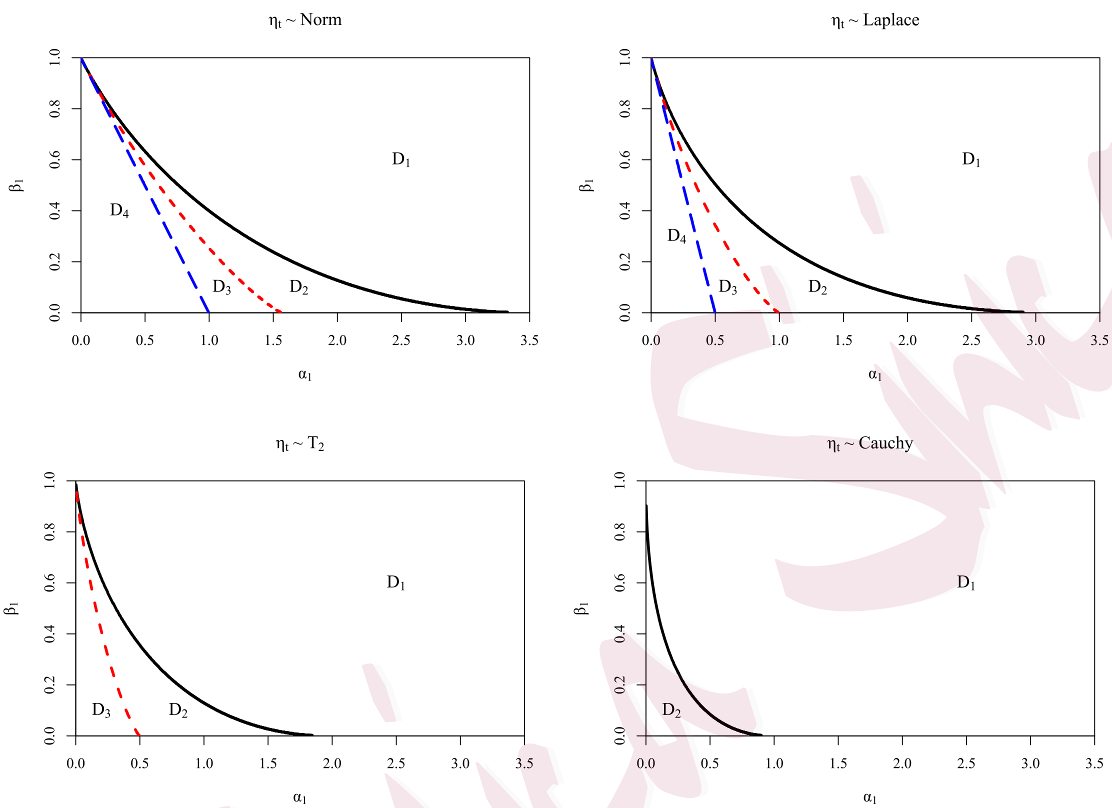

Figure 1: The regions of the tail index α of h_t in the GARCH(1,1) model when $\eta_t \sim N(0,1)$, Laplace$(0,1)$, t_2, or Cauchy distribution, where D_1 means the domain of no stationary solution, D_2 means the domain of tail index $\alpha \in (0,1)$, D_3 means the domain of tail index $\alpha \in (1,2)$, and D_4 means the domain of tail index $\alpha \in (2,\infty)$.

2. Assumption and the WLSE

2.1 Assumption

Throughout this paper, we focus on the noise satisfying the heteroscedastic form:

$$\varepsilon_t = \eta_t h_t \quad \text{and} \quad h_t = h(\eta_{t-1}, \eta_{t-2}, \cdots), \tag{2.1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

where the innovation $\{\eta_t\}$ is a sequence of i.i.d. symmetric random variables with $P(\eta_t \neq 0) > 0$, and $h(\cdot)$ is a measurable positive function. Because the structure of h_t is not specified, it is general enough, and many popular G/ARCH-type models are included in model (2.1), such as the absolute value GARCH model in Taylor (1986) and Schwert (1989), nonlinear GARCH model in Engle (1990), GJR model in Glosten, Jagannathan, and Runkle (1993), threshold GARCH model in Zakoïan (1994), quadratic ARCH model in Sentana (1995), and volatility switching GARCH model in Fornari and Mele (1997). We make the following fundamental assumption on $\{\varepsilon_t\}$.

Assumption 2.1. *There exists some strictly positive deterministic sequence* $\{a_n\}$*, such that, in the Skorohod space* $\mathbb{D}[0,1]$ *equipped with an* S*-topology,*

$$S_n(\tau) = \frac{1}{a_n}\sum_{t=1}^{[n\tau]}\varepsilon_t \xrightarrow{w} S(\tau),$$

where $a_n \to \infty$ *and* $\int_0^1 S^2(\tau)d\tau > 0$ *almost surely.*

Assumption 2.1 is actually a very mild condition that allows for both the finite-variance case ($\alpha > 2$) and the infinite-variance case ($\alpha < 2$). For a better illustration, we now provide several examples and conditions under which Assumption 2.1 holds, especially for those commonly used in the literature and the model used in the simulation.

Note that the S-topology is a sequential topology on the Skorohod space $\mathbb{D}[0,1]$

proposed by Jakubowski (1997). By Proposition 3.1 and Theorem 3.5 in that paper, for Assumption 2.1 to hold, it is sufficient to show that

$$(S_n(\tau_1), \cdots, S_n(\tau_k)) \xrightarrow{d} (S(\tau_1), \cdots, S(\tau_k)), \forall k \in \mathbb{N} \text{ and } \tau_i \in [0,1], \tag{2.2}$$

$$\|S_n\| = O_p(1), \text{ and for any } a < b, \text{ we have } N^{a,b}(S_n) = O_p(1), \tag{2.3}$$

where $\|S_n\| = \sup_{\tau \in [0,1]} |S_n(\tau)|$, and $N^{a,b}(S_n)$ is the usual number of up-crossing defined by the following relation: $N^{a,b}(S_n) \geq l$ if and only if there exist numbers $0 \leq \tau_1 < \tau_2 < \cdots < \tau_{2l-1} < \tau_{2l} \leq 1$ such that $S_n(\tau_{2i}) > b$ and $S_n(\tau_{2i-1}) < a$, for all $i = 1, \cdots, l$. Although the weak convergence in an S-topology is much weaker than that in a J_1-topology (or uniform topology), it has been proved that the well-known almost sure Skorohod representation theorem still holds; see Jakubowski (1997) for details. Thus, the S-topology can be widely used in many scenarios, especially for heavy-tailed data.

Remark 2.1. The convergence in (2.2) is well known as the convergence of a finite-dimension distribution, and is found in many G/ARCH-type processes. For example, consider one representative class of G-GARCH processes in Zhang and Ling (2015), with

$$h_t^2 = \omega + c(\eta_{t-1}) h_{t-1}^2,$$

where $\omega > 0$, and $c(\cdot)$ is a nonnegative function with $c(0) < 1$. Zhang and Ling (2015) show that there exists a unique $\alpha \in (0, 2k_0]$ such that $E(c(\eta_t))^{\alpha/2} = 1$, and

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

the noise ε_t is a regular variation with tail index α (i.e., $P(|\varepsilon_t| > x) \sim x^{-\alpha}$), under the following conditions:

(a) $E\log(c(\eta_t)) < 0$;

(b) There exists a $k_0 > 0$ such that $E(c(\eta_t))^{k_0} \geq 1$ and $E[(c(\eta_t))^{k_0}\log^+(c(\eta_t))] < \infty$, and $E(|\eta_t|^{2k_0}) < \infty$, where $\log^+(x) = \max\{0, \log(x)\}$;

(c) The density $f(x)$ of η_t is positive in the neighbourhood zero.

Furthermore, using similar arguments to those for Theorem 2.1 in Chan and Zhang (2010), it is straightforward to obtain condition (2.2), in which $a_n = \sqrt{n}$ for $\alpha > 2$ (light-tail) and $a_n = n^{1/\alpha}$ for $\alpha < 2$ (heavy-tail).

Remark 2.2. On the other hand, one can easily show that condition (2.3) is satisfied for the two foregoing cases. In the first case, the tail index $\alpha > 2$ and $a_n = \sqrt{n}$. Here, because $S_n(\tau)$ is martingale and by Doob's inequality, it follows that, for any $M > 0$,

$$P(\|S_n\| > M) \leq 3M^{-1}\sup_{\tau\in[0,1]} E|S_n(\tau)|, \quad EN^{a,b}(S_n) \leq \frac{1}{b-a}(|a| + \sup_{\tau\in[0,1]} E|S_n(\tau)|).$$

Then, condition (2.3) holds from $\sup_n \sup_{\tau\in[0,1]} E|S_n(\tau)| \leq (E\varepsilon_t^2)^{1/2} < \infty$. In the second case, the tail index $\alpha < 2$ and $a_n = n^{1/\alpha}$. Rewrite

$$S_n(\tau) = \sum_{t=1}^{[n\tau]} \varepsilon_t 1_{(|\varepsilon_t| < a_n)}/a_n + \sum_{t=1}^{[n\tau]} \varepsilon_t 1_{(|\varepsilon_t| \geq a_n)}/a_n = S_{n1}(\tau) + S_{n2}(\tau).$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

Note that $S_{n1}(\tau)$ is still martingale, and by Karamata's theorem, we have

$$\sup_{\tau\in[0,1]} ES_{n1}^2(\tau) = \frac{nE\varepsilon_t^2 1_{(|\varepsilon_t|<a_n)}}{a_n^2} \longrightarrow \frac{\alpha}{2-\alpha}, \quad \text{as } n\to\infty.$$

Then, condition (2.3) holds for $S_{n1}(\tau)$. Furthermore, choose $\rho \in (0, \alpha\wedge 1)$. Then, by Karamata's theorem again, we have

$$E\|S_{n2}\|^\rho \le \frac{nE|\varepsilon_t|^\rho 1_{(|\varepsilon_t|\ge a_n)}}{a_n^\rho} \longrightarrow \frac{\alpha}{\alpha-\rho}, \quad \text{as } n\to\infty,$$

which implies that $\|S_{n2}\| = O_p(1)$. For $N^{a,b}(S_{n2})$, because $N^{a,b}(S_{n2}) \le \sum_{t=1}^n 1_{(|\varepsilon_t|\ge a_n)}$, $\limsup_n EN^{a,b}(S_{n2}) \le \limsup_n nP(|\varepsilon_t| \ge a_n) < \infty$. Thus, (2.3) holds for $S_{n2}(\tau)$. As a result, condition (2.3) holds for $S_n(\tau) = S_{n1}(\tau) + S_{n2}(\tau)$.

2.2 The WLSE approach

Now, we investigate the estimation of ϕ. The ordinary LSE is defined as

$$\hat{\phi}_{lse} = \arg\min \sum_{t=1}^n (y_t - \phi y_{t-1})^2 = \frac{\sum_{t=1}^n y_t y_{t-1}}{\sum_{t=1}^n y_t^2}.$$

Conventionally, when $\phi = 1$ (i.e., under H_0), Assumption 2.1 may imply that

$$n(\hat{\phi}_{lse} - 1) = n\frac{\sum_{t=1}^n \varepsilon_t y_{t-1}}{\sum_{t=1}^n y_t^2} \xrightarrow{d} \frac{\int_0^1 S^-(\tau)dS(\tau)}{\int_0^1 S^2(\tau)d\tau},$$

where $S^-(\tau)$ denotes the left-hand limit of $S(\tau)$. When the variance of ε_t is infinite, $S(\tau)$ is always a stable process with a tail index smaller than two, as in Chan and Zhang (2010) and the references therein. In this case, the above limiting distribution is not pivotal, and the existing bootstrap methods are very sensitive to the structure and tail index of ε_t (Cavaliere, Georgiev, and Taylor, 2018;

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

Zhang and Chan, 2020). It seems infeasible to use a unified bootstrap method to deal with this issue.

Inspired by the main ideas of Chan, Li, and Peng (2012) and Huang et al. (2020), we define the WLSE as

$$\hat{\phi}_{wlse} = \arg\min \sum_{t=1}^{n} \frac{(y_t - \phi y_{t-1})^2}{(1+y_{t-1}^2)^{1/2}[1+(\Delta y_t)^2]^{1/2}},$$

where $\Delta y_t = y_t - y_{t-1}$. Then, under the null hypothesis, it is easy to get that

$$\sum_{t=1}^{n} \frac{y_{t-1}^2}{(1+y_{t-1}^2)^{1/2}[1+(\Delta y_t)^2]^{1/2}}(\hat{\phi}_{wlse} - 1) = \sum_{t=1}^{n} \frac{y_{t-1}\varepsilon_t}{(1+y_{t-1}^2)^{1/2}[1+(\Delta y_t)^2]^{1/2}}.$$

Thus, we can derive the DF-type test statistic

$$T_n = n^{-1/2} \sum_{t=1}^{n} \frac{y_{t-1}^2}{(1+y_{t-1}^2)^{1/2}[1+(\Delta y_t)^2]^{1/2}}(\hat{\phi}_{wlse} - 1).$$

The asymptotic properties of T_n are given in Theorem 2.1.

Theorem 2.1. *Suppose that Assumption 2.1 holds. Under H_0, it follows that*

$$T_n \xrightarrow{d} N(0, \sigma^2),$$

where $\sigma^2 = E[\varepsilon_t^2/(1+\varepsilon_t^2)]$. Under H_1, it follows that $T_n \xrightarrow{p} -\infty$.

Remark 2.3. In practice, we can replace σ^2 with $\hat{\sigma}^2 = n^{-1}\sum_{t=1}^{n}(\Delta y_t)^2/[1+(\Delta y_t)^2]$ and develop the WLSE-based test as

$$\tilde{T}_n = T_n/\hat{\sigma}.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

Because $\hat{\sigma}$ is a consistent estimator of σ under H_0, by (2.1) and the ergodic theorem, it is obvious that $\tilde{T}_n$ converges in distribution to $N(0,1)$. At the same time, because $\hat{\sigma}$ is bounded by one, we still have $\tilde{T}_n \xrightarrow{p} -\infty$ under H_1. Then, at the given significant level α, the null hypothesis should be rejected when $\tilde{T}_n < u_\alpha$, where u_α denotes the αth quantile of the standard normal distribution. Under this criterion, the power approaches one as $n \to \infty$.

However, it is well known that the estimation of the asymptotic variance is not always stable, especially for strongly dependent time series or a small sample size. As a result, the WLSE-based test $\tilde{T}_n$ may suffer from a serious size distortion, as shown in Section 5. In the next section, we attempt to bypass the estimation of the nuisance parameters by using the empirical likelihood technique in Owen (2001), which has been found to be very useful in many scientific fields.

3. Empirical Likelihood Methods

3.1 Empirical likelihood test

Recall that the proposed WLSE is based on the core idea that, under the null hypothesis, the heavy-tailed term ε_t (i.e.,Δy_t) can be bounded by $[1+(\Delta y_t)^2]^{1/2}$, whereas y_{t-1} can be bounded by $(1+y_{t-1}^2)^{1/2}$. Thus, we consider the score function

$$Z_t(\phi) = \frac{y_{t-1}(y_t - \phi y_{t-1})}{(1+y_{t-1}^2)^{1/2}[1+(y_t-\phi y_{t-1})^2]^{1/2}}. \tag{3.1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

The empirical likelihood function is given by

$$L(\phi) = \sup\left\{\prod_{t=1}^{n}(np_t) : \sum_{t=1}^{n} p_t = 1, \sum_{t=1}^{n} p_t Z_t(\phi) = 0, p_t > 0, t = 1, \cdots, n\right\}.$$

Using the Lagrange multiplier technique, we can show that

$$L(\phi) = \prod_{t=1}^{n} \frac{1}{1 + \lambda Z_t(\phi)},$$

where the Lagrange multiplier λ is the solution of

$$\sum_{t=1}^{n} \frac{Z_t(\phi)}{1 + \lambda Z_t(\phi)} = 0.$$

At the same time, the empirical log-likelihood ratio is

$$l(\phi) = -2\log L(\phi) = 2\sum_{t=1}^{n} \log[1 + \lambda Z_t(\phi)].$$

Now, we give the asymptotic results for $l(\phi)$ in the following theorem.

Theorem 3.1. *Suppose that Assumption 2.1 holds. Under H_0, it follows that*

$$l(1) \xrightarrow{d} \chi_1^2,$$

as $n \to \infty$. Under H_1, we have $l(1) \xrightarrow{p} \infty$ and $l(\phi) \xrightarrow{d} \chi_1^2$, as $n \to \infty$.

According to Theorem 3.1, we reject the null hypothesis at the significance level α if $l(1) > \chi^2_{1,1-\alpha}$, where $\chi^2_{1,1-\alpha}$ denotes the $(1-\alpha)$th quantile of a chi-squared distribution with one degree of freedom. After rejecting H_0, Theorem 3.1 implies that the confidence interval for ϕ at level $1-\alpha$ can be constructed as

$$I_{1-\alpha} = \{\phi : l(\phi) < \chi^2_{1,1-\alpha}\}.$$

Remark 3.1. The new empirical likelihood test (ELT) based on the score function $Z_t(\phi)$ in (3.1) is essentially distinct from the ELT proposed by Huang et al. (2020) in terms of both model settings and methodology. In our settings, under the null,

$$Z_t(1) = \frac{y_{t-1}\Delta y_t}{(1+y_{t-1}^2)^{1/2}[1+(\Delta y_t)^2]^{1/2}}.$$

Because $\Delta y_t = \varepsilon_t$ must be bounded by the normalizer $[1+(\Delta y_t)^2]^{1/2}$, all the heavy-tailed effects in the noise ε_t cancel out. As a result, our method does not rely on any specific form of $h(\cdot)$ and no moment condition on h_t or η_t is required, making it applicable to many popular G/ARCH-type models, such as the standard GARCH, nonlinear GARCH, and GJR, among many others. In contrast, Huang et al. (2020) mainly considered noise ε_t satisfying the standard GARCH(p,q) model, as in (1.2), and their ELT method depends on the functional of $h(\cdot)$. For instance, when $q = 0$ (ARCH model), under the null, they use the following score equation:

$$\begin{aligned} Y_t(1) &= \frac{y_{t-1}\Delta y_t}{(1+y_{t-1}^2)^{1/2}[1+\sum_{k=1}^m (\Delta y_{t-k})^2]^{1/2}} \\ &= \eta_t \times \frac{y_{t-1}h_t}{(1+y_{t-1}^2)^{1/2}[1+\sum_{k=1}^m (\Delta y_{t-k})^2]^{1/2}}, \end{aligned}$$

where m is chosen to be larger than p to guarantee the inequality

$$h_t^2 \leq \max\{\omega, \alpha_1, \cdots, \alpha_p\}\Big[1+\sum_{k=1}^m (\Delta y_{t-k})^2\Big].$$

Therefore, a priori information on the model structure $h(\cdot)$ is indispensable, albeit typically unknown in practice, especially when checking the stationarity of a time

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

series. On the other hand, the normalizer $[1+\sum_{k=1}^{m}(\Delta y_{t-k})^2]^{1/2}$ in the denominator of $Y_t(1)$ is only able to remove the heavy-tailed effect in h_t, resulting in the inefficiency for the case with $E\eta_t^2=\infty$, as shown in Section 5. See Figure 1, in which Huang et al. (2020) may not be able to handle the domain D_2.

3.2 Adjusted ELT

In order to improve the size performance of the proposed ELT, we further consider the adjusted empirical likelihood approach proposed by Chen, Variyath, and Abraham (2008). Define the additional term

$$Z_{n+1}(\phi) = -b_n n^{-1}\sum_{t=1}^{n} Z_t(\phi),$$

where $Z_t(\phi)$ is defined in (3.1) and b_n is some positive constant. Then, the adjusted empirical likelihood function is defined as

$$L^a(\phi) = \sup\left\{\prod_{t=1}^{n+1}(n+1)p_t : p_t > 0, t = 1, \cdots, n+1; \sum_{t=1}^{n+1} p_t = 1, \sum_{t=1}^{n+1} p_t Z_t(\phi) = 0\right\}.$$

Similarly, the corresponding adjusted empirical log-likelihood ratio is

$$l^a(\phi) = -2\log L^a(\phi) = 2\sum_{t=1}^{n+1}\log[1+\lambda Z_t(\phi)],$$

where the Lagrange multiplier λ is the solution of

$$\sum_{t=1}^{n+1}\frac{Z_t(\phi)}{1+\lambda Z_t(\phi)} = 0.$$

Then, the limiting theory of $l^a(\phi)$ can be derived as follows.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

Theorem 3.2. *Suppose that Assumption 2.1 holds and* $b_n/n + 1/b_n = o(1)$. *Under* H_0, *it follows that*

$$l^a(1) \xrightarrow{d} \chi_1^2,$$

as $n \to \infty$. *Under* H_1, *we have* $l^a(1) \xrightarrow{p} \infty$ *and* $l^a(\phi) \xrightarrow{d} \chi_1^2$, *as* $n \to \infty$.

Theorem 3.2 shows that we need to reject the null hypothesis at the significance level α if $l^a(1) > \chi^2_{1,1-\alpha}$ and the confidence interval of ϕ at level $1-\alpha$ is constructed as $I^a_{1-\alpha} = \{\phi : l^a(\phi) < \chi^2_{1,1-\alpha}\}$.

Remark 3.2. Here, we simply point out the difference between the two empirical likelihood methods based on $l(\phi)$ and $l^a(\phi)$. By the definition of $L(\phi)$, we can see that the necessary and sufficient condition for its existence is that the original point is an interior point of the convex hull of $\{Z_t(\phi), t \leq n\}$. Under some moment and dependence assumptions, this condition can hold with probability tending to one as $n \to \infty$ (Owen, 2001). However, for general time series or in the case of a small sample size, this may be a serious limitation (Chen, Variyath, and Abraham, 2008). Thus, the adjusted term $Z_{n+1}(\phi)$ is used to ensure that the original point is an interior point of the convex hull of $\{Z_t(\phi), t \leq n+1\}$ such that $L^a(\phi)$ is well defined. As shown in our simulations, $l^a(1)$ has better size performance than that of $l(1)$. At the same time, b_n can be chosen as $\max\{1, \log(n)/2\}$, as recommended by Chen, Variyath, and Abraham (2008). For more discussions on the two ELTs, we refer to Zheng and Yu (2013).

We are now ready to extend the score function $Z_t(\phi)$ to a more general form

$$Z_t(\phi) = \frac{y_{t-1}}{(1+y_{t-1}^2)^{1/2}}\rho(y_t - \phi y_{t-1}), \quad t = 1, \cdots, n,$$

and $Z_{n+1}(\phi) = -b_n n^{-1} \sum_{t=1}^n Z_t(\phi)$, where $\rho(x)$ is a function on the real line. Using a proof similar to those of Theorems 3.1–3.2, it is not hard to obtain the following corollary.

Corollary 3.1. *Suppose that Assumption 2.1 holds and $b_n/n + 1/b_n = o(1)$. If $\rho(x)$ is a bounded odd monotonic function, and one of the following conditions holds:*

1. *$\rho(x)$ is strictly monotonic;*

2. *$\rho(x) \neq \rho(y)$, $\forall xy < 0$, and the density of η_t is positive in the neighborhood of zero.*

Then, all the limiting results in Theorems 3.1–3.2 still hold.

Remark 3.3. The traditional M-estimator $\hat{\phi}_M$ of ϕ is the solution of the equation

$$\sum_{t=1}^n y_{t-1}\rho(y_t - \phi y_{t-1}) = 0,$$

where $\rho(x)$ is typically the first derivative of some loss function. In this case, the statistical inference is based on the asymptotic property of $\hat{\phi}_M$. To guarantee the derived DF-type test to be asymptotically Gaussian under H_0, additional continuous conditions for $\rho(x)$ are often needed (Knight, 1991; Shin and So, 1999; Samarakoon and Knight, 2009). However, the proposed unit-root tests do not rely on any estimator of ϕ or need any continuous assumption for $\rho(\cdot)$. Thus, many widely used functions are incorporated in this framework, such as the Huber function $\rho(x) = \min[c, \max(-c, x)]$ and the sign function $\rho(x) = 1_{(x>0)} - 1_{(x<0)}$.

4. Extensions to other models

In this section, we further generalize the proposed empirical likelihood method to other unit-root models. We first study the unit-root model with a constant term, namely,

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t, \tag{4.1}$$

where μ is a constant and ε_t satisfies the heteroscedastic form (2.1). Recall that the LSE of the parameters (ϕ, μ) is the solution of the equations

$$\sum_{t=1}^{n} \varepsilon_t(\phi, \mu) = 0 \text{ and } \sum_{t=1}^{n} y_{t-1}\varepsilon_t(\phi, \mu) = 0,$$

where $\varepsilon_t(\phi, \mu) = y_t - \mu - \phi y_{t-1}$. Then, it is natural to consider the weighted LSE score equations $\mathbf{Z}_t(\phi, \mu) = (Z_{t,1}(\phi, \mu), Z_{t,2}(\phi, \mu))'$, with

$$Z_{t,1}(\phi, \mu) = \frac{\varepsilon_t(\phi, \mu)}{[1 + \varepsilon_t^2(\phi, \mu)]^{1/2}} \text{ and } Z_{t,2}(\phi, \mu) = \frac{y_{t-1}}{(1 + y_{t-1}^2)^{1/2}} Z_{t,1}(\phi, \mu).$$

Under H_0, because $|y_n|/(1 + y_n^2)^{1/2} \xrightarrow{p} 1$, it is not hard to show that

$$n^{-1/2} \sum_{t=1}^{n} Z_{t,i}(1, \mu_0) \xrightarrow{d} N(0, \sigma^2),$$

for $i = 1, 2$, where μ_0 is the true parameter and σ^2 is defined in Theorem 2.1. At the same time, $n^{-1/2}\sum_{t=1}^{n} Z_{t,2}(1, \mu_0)$ is asymptotically equivalent to $n^{-1/2}\sum_{t=1}^{n} Z_{t,1}(1, \mu_0)$, which implies that neither of them are bivariate normal. To overcome this degenerate issue, similarly to Li, Chan, and Peng (2014) and Huang et al. (2020), we add some independent samples into the score equations. Specifically, define $\tilde{\mathbf{Z}}_t(\phi, \mu) = (\tilde{Z}_{t,1}(\phi, \mu), \tilde{Z}_{t,2}(\phi, \mu))'$, with

$$\tilde{Z}_{t,1}(\phi, \mu) = \frac{\varepsilon_t(\phi, \mu)}{[1 + \varepsilon_t^2(\phi, \mu)]^{1/2}},$$
$$\tilde{Z}_{t,2}(\phi, \mu) = \frac{y_{t-1}}{(1 + y_{t-1}^2)^{\delta}} \tilde{Z}_{t,1}(\phi, \mu) + w_t,$$

where the constant $\delta > 1/2$ and w_t is a sequence of i.i.d. random variables with $P(w_t = \pm 1) = 1/2$. As suggested by Li, Chan, and Peng (2014), δ is usually set to 0.75. Then, the associated empirical likelihood function is given by

$$\tilde{L}(\phi, \mu) = \sup\left\{\prod_{t=1}^{n}(np_t) : p_t > 0, \sum_{t=1}^{n} p_t = 1, \sum_{t=1}^{n} p_t \tilde{\mathbf{Z}}_t(\phi, \mu) = \mathbf{0}\right\}. \quad (4.2)$$

Because the true parameter μ_0 is unknown, we need to consider the profile empirical likelihood function $\tilde{L}(\phi) = \max_{\mu} \tilde{L}(\phi, \mu)$ and put $\tilde{l}(\phi) = -2\log(\tilde{L}(\phi))$. The following theorem gives its limiting property.

Theorem 4.1. *Suppose that Assumption 2.1 holds and $a_n/n \to c \in [0, \infty]$. Then, under H_0, it follows that $\tilde{l}(1) \xrightarrow{d} \chi_1^2$, as $n \to \infty$. Furthermore, under H_1, it follows that $\tilde{l}(1) \xrightarrow{p} \infty$.*

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

Now, we investigate a more complicated unit-root AR(r) model with a constant term, namely,

$$y_t = \mu + \phi y_{t-1} + \sum_{j=1}^{r} \phi_j \Delta y_{t-j} + \varepsilon_t, \quad r \geq 1. \tag{4.3}$$

Denote $\boldsymbol{\theta} = (\mu, \phi_1, \cdots, \phi_r)'$ and $\varepsilon_t(\phi, \boldsymbol{\theta}) = y_t - \mu - \phi y_{t-1} - \sum_{j=1}^{r} \phi_j \Delta y_{t-j}$. Note that, under the null hypothesis, the term Δy_{t-j} is also likely to be heavy tailed, and thus the score equations are modified as follows:

$$\begin{aligned}
\bar{Z}_{t,1}(\phi, \boldsymbol{\theta}) &= \frac{\varepsilon_t(\phi, \boldsymbol{\theta})}{[1 + \sum_{j=1}^{r} (\Delta y_{t-j})^2]^{3/2} [1 + \varepsilon_t^2(\phi, \boldsymbol{\theta})]^{1/2}}, \\
\bar{Z}_{t,2}(\phi, \boldsymbol{\theta}) &= \frac{y_{t-1}}{(1 + y_{t-1}^2)^{\delta}} \bar{Z}_{t,1}(\phi, \boldsymbol{\theta}) + w_t, \\
\bar{Z}_{t,2+j}(\phi, \boldsymbol{\theta}) &= \frac{\Delta y_{t-j}}{[1 + (\Delta y_{t-j})^2]^{1/2}} \bar{Z}_{t,1}(\phi, \boldsymbol{\theta}), \text{ for } j = 1, \cdots, r,
\end{aligned}$$

where the additional term $[1 + \sum_{j=1}^{r} (\Delta y_{t-j})^2]^{3/2}$ in the denominator is used to bound $\partial^k \bar{Z}_{t,1}(\phi, \boldsymbol{\theta}) / \partial \boldsymbol{\theta}^k$, for $k = 1, 2, 3$. Let $\bar{\mathbf{Z}}_t(\phi, \boldsymbol{\theta}) = (\bar{Z}_{t,1}(\phi, \boldsymbol{\theta}), \bar{Z}_{t,2}(\phi, \boldsymbol{\theta}), \cdots, \bar{Z}_{t,2+r}(\phi, \boldsymbol{\theta}))'$ and the empirical likelihood function is

$$\bar{L}(\phi, \boldsymbol{\theta}) = \sup \left\{ \prod_{t=1}^{n} (np_t) : p_t > 0, \sum_{t=1}^{n} p_t = 1, \sum_{t=1}^{n} p_t \bar{\mathbf{Z}}_t(\phi, \boldsymbol{\theta}) = \mathbf{0} \right\}. \tag{4.4}$$

Similarly, define $\bar{L}(\phi) = \max_{\boldsymbol{\theta}} \bar{L}(\phi, \boldsymbol{\theta})$ and $\bar{l}(\phi) = -2 \log(\bar{L}(\phi))$. Note that Δy_t is a linear process with respect to the noise ε_t. Hence, before showing its limiting property, we need the following counterpart of Assumption 2.1.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

Assumption 4.1. *There exists some deterministic sequence* $\{\bar{a}_n\}$*, such that*

$$\bar{S}_n(\tau) = \frac{1}{\bar{a}_n}\sum_{t=1}^{[n\tau]} u_t \xrightarrow{w} \bar{S}(\tau),$$

where $\bar{a}_n \to \infty$ *and* $u_t = \sum_{l=0}^{\infty} \rho_l \varepsilon_{t-l}$ *with* $\rho_l = O(\rho^l)$*, for some* $\rho \in (0, 1)$*.*

Remark 4.1. When ε_t is a sequence of i.i.d. heavy-tailed noises, Avram and Taqqu (1992) show that, under the classical J_1-topology, Assumption 4.1 holds if and if only the linear process $\{u_t\}$ is independent. Assumption 4.1, though it seems to be unreasonable under the J_1-topology, does make sense under the S-topology. Indeed, one can easily derive Assumption 4.1 from Assumption 2.1 using the sufficient conditions in (2.2) and (2.3), if the following additional condition holds:

$$\lim_{H\to\infty} \limsup_{n\to\infty} P\Big(\sup_{0\le\tau\le1}\Big|a_n^{-1}\sum_{t=1}^{[n\tau]} u_{t,H}\Big| > \eta\Big) = 0,$$

where η is any positive number and $u_{t,H} = \sum_{l=H+1}^{\infty} \rho_l \varepsilon_{t-l}$. In this case, $\bar{S}(\tau) = (\sum_{l=0}^{\infty} \rho_l) S(\tau)$ and $\bar{a}_n = a_n$; see Zhang, Sin, and Ling (2015) for some examples.

Theorem 4.2. *Suppose that Assumption 4.1 holds and* $\bar{a}_n/n \to c \in [0, \infty]$*. Then, under* H_0*, if* $\{\Delta y_t\}$ *is strictly stationary, then it follows that* $\bar{l}(1) \xrightarrow{d} \chi_1^2$ *as* $n \to \infty$*. Furthermore, under* H_1*, if* $\{y_t\}$ *is strictly stationary , then it follows that* $\bar{l}(1) \xrightarrow{p} \infty$*.*

Theorems 4.1–4.2 show that the proposed profile ELTs are still asymptotically chi-squared, even in the more complicated unit-root models (4.1) and (4.3). In

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

general, time series $\{y_t\}$ is generated from the linear model

$$y_t = \sum_{i=1}^{k} \gamma_i f_i(t) + \phi y_{t-1} + \sum_{j=1}^{r} \phi_j \Delta y_{t-j} + \varepsilon_t + \sum_{l=1}^{m} \rho_l \varepsilon_{t-l},$$

where $\{f_i(t)\}$ are time trend functions. Let $\boldsymbol{\theta} = (\gamma_i, \cdots, \phi_j, \cdots, \rho_l, \cdots)'$ and define

$$\varepsilon_t(\phi, \boldsymbol{\theta}) = y_t - \sum_{i=1}^{k} \gamma_i f_i(t) - \phi y_{t-1} - \sum_{j=1}^{r} \phi_j \Delta y_{t-j} - \sum_{l=1}^{m} \rho_l \varepsilon_{t-l}(\phi, \boldsymbol{\theta}), \tag{4.5}$$

where $\varepsilon_t(\phi, \boldsymbol{\theta}) = 0$, for $t < 1$. Then, the weighted score functions denoted by $\mathbf{Z}_t(\phi, \boldsymbol{\theta})$ can be constructed in the same way as those in (4.2) and (4.4). Furthermore, if $n^{-1/2} \sum_{t=1}^{n} \mathbf{Z}_t(1, \boldsymbol{\theta}_0)$ is asymptotically normal and $\partial^k \mathbf{Z}_t(\phi, \boldsymbol{\theta})/\partial \boldsymbol{\theta}^k$ is uniformly bounded for $k = 1, 2, 3$, the asymptotically chi-squared property can be derived using a similar proof procedure to those of Theorems 4.1–4.2; see the Supplementary Material for technical details. Therefore, our empirical likelihood methods are feasible for many structures of unit-root models.

Remark 4.2. As suggested by one of the referees, we consider the unit-root testing problem when the noise is a linear process, namely,

$$y_t = \phi y_{t-1} + u_t = \phi y_{t-1} + \varepsilon_t + \sum_{l=1}^{\infty} \rho_l \varepsilon_{t-l}. \tag{4.6}$$

In contrast to all the foregoing models, infinite numbers of parameters $\{\rho_1, \rho_2, \cdots\}$ are involved in (4.6). Thus, it is not feasible to use the recursive definition for $\varepsilon_t(\phi, \boldsymbol{\theta})$ as (4.5), where m is finite. One possible method is to use the basic idea in

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

the ADF test, as in Zhang and Chan (2020). Specifically, under H_0, model (4.6) can be rewritten as

$$y_t = \phi y_{t-1} + \sum_{j=1}^{k} \beta_j \Delta y_{t-j} + \varepsilon_t + \rho_{t,k}, \tag{4.7}$$

where $(1-\sum_{j=1}^{\infty} \beta_j z^j)$ is the inverse function of $(1+\sum_{l=1}^{\infty} \rho_l z^l)$ and $\rho_{t,k} = \sum_{j=k+1}^{\infty} \beta_j u_{t-j}$. Then, we define $\varepsilon_t(\phi, \boldsymbol{\theta}) = y_t - \phi y_{t-1} - \sum_{j=1}^{k} \beta_j \Delta y_{t-j}$, with $\boldsymbol{\theta} = (\beta_1, \cdots, \beta_k)'$. Note that $\varepsilon_t(1, \boldsymbol{\theta}_0) = \varepsilon_t + \rho_{t,k}$ and, thus, it is necessary to select an appropriate $k \to \infty$ as $n \to \infty$. Nevertheless, this procedure involves a high-dimensional empirical likelihood, which is substantially different from the methodology in the fixed-dimensional case. Therefore, we leave this problem for future work.

5. Simulation Studies

To examine the finite-sample behavior of the proposed unit-root tests, we focus on model (1.1), with $\phi = 1$ corresponding to the null hypothesis (i.e., unit root), and $\phi \in \{0.95, 0.9, 0.85\}$ corresponding to the alternative (i.e., stationary). In all simulations, we take 1000 replications for each case, and the results are reported at the 5% significance level.

5.1 GARCH-type noise

In this section, we consider the following GARCH-type noise

$$\varepsilon_t = \eta_t h_t, \;\; h_t^2 = \omega + [\beta_1 + \alpha_1 \eta_{t-1}^2 + \gamma \eta_{t-1}^2 1_{(\eta_{t-1}<0)}] h_{t-1}^2, \tag{5.1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

$$\varepsilon_t = \eta_t h_t,\ h_t^2 = \omega + \{\beta_1 + \alpha_1[1 - 2\gamma \text{sign}(\eta_{t-1}) + \gamma^2]\eta_{t-1}^2\}h_{t-1}^2, \tag{5.2}$$

where $\omega = 0.1$ and $\gamma = 0.1$. Let $\boldsymbol{\theta} = (\beta_1, \alpha_1)$ be the unknown parameters, and α be the tail index of ε_t. By Remark 2.1, the following heavy-tailed cases are considered:

1. For $\alpha \in (1, 2)$, we uniformly take (i) $\boldsymbol{\theta} = (0.6, 0.4)$, with η_t being $N(0, 1)$; (ii) $\boldsymbol{\theta} = (0.5, 0.3)$, with η_t being a Laplace(0,1) distribution; (iii) $\boldsymbol{\theta} = (0.7, 0.1)$, with η_t being a t_3 distribution; and (iv) $\boldsymbol{\theta} = (0.5, 0.1)$, with η_t being a t_2 distribution.

2. For $\alpha \in (0, 1)$, we uniformly take (i) $\boldsymbol{\theta} = (0.6, 0.5)$, with η_t being $N(0, 1)$; (ii) $\boldsymbol{\theta} = (0.5, 0.4)$, with η_t being a Laplace(0,1) distribution; (iii) $\boldsymbol{\theta} = (0.65, 0.1)$, with η_t being a t_2 distribution; and (iv) $\boldsymbol{\theta} = (0.35, 0.1)$, with η_t being a Cauchy distribution.

For comparison, we also implement the ELT proposed by Huang et al. (2020) with the order $m = 2$ (see Remark 3.1).

Tables 1–2 report the size and power of the tests with $\alpha \in (1, 2)$ when $n = 100$ and $n = 300$, respectively. Under the null hypothesis ($\phi = 1$), it is apparent that the proposed test $\tilde{T}_n$ is always undersized. In particular, the size of $\tilde{T}_n$ is only 0.016 in model (5.2) when $\eta_t \sim N(0, 1)$ and $n = 100$. Instead, the proposed ELTs $l(1)$ and $l^a(1)$ present satisfactory size performance, whereas the ELT is oversized and the size distortion becomes more severe as the tail of η_t becomes heavier. In addition, all the

ELTs improve in terms of power when the sample size increases from 100 to 300. The proposed tests $l(1)$ and $l^a(1)$ are more powerful than the ELT in almost all cases, except for $\eta_t \sim N(0,1)$, where the ELT is better, but the gap is acceptable. Similar phenomena are observed in Tables 3–4, which summarize the associated simulation results for $\alpha \in (0,1)$.

For extremely heavy-tailed noise, Tables 3–4 show that the size and power performance of $l(1)$ and $l^a(1)$ are quite robust, with a stable size and a rational power. However, when η_t follows a Cauchy distribution, the ELT can suffer from serious size distortion and a severe loss in power, even for a large sample size. Therefore, the existing ELT is sensitive to the tail of η_t, and may not be used in unit-root testing when the tail index is unknown. In summary, our proposed ELTs are efficient and powerful for detecting a possible unit root, especially for models with heavy-tailed innovations.

5.2 The i.i.d. noise

We now conduct a simulation study to illustrate that the proposed tests are still valid when ε_t are i.i.d. (i.e., $h_t = 1$) with an infinite variance. Specifically, the noise ε_t is generated from the model

$$\varepsilon_t = |\eta_t|^{1/\alpha}\text{sign}(\eta_t),$$

where η_t ∼Cauchy distribution and α is the tail index. For comparison, we also

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

Table 1: Size and power of the unit-root tests (×100) with $\alpha \in (1,2)$ and $n = 100$

		ε_t ~model (5.1)				ε_t ~model (5.2)			
$\eta_t \sim$	ϕ	$\tilde{T}_n$	$l(1)$	$l^a(1)$	ELT	$\tilde{T}_n$	$l(1)$	$l^a(1)$	ELT
N(0,1)	1.00	2.4	4.8	4.7	5.7	1.6	4.5	4.9	6.7
	0.95	19.3	19.8	17.3	22.6	17.6	19.5	16.7	23.5
	0.90	34.1	36.0	33.9	47.0	32.6	39.0	37.0	49.3
	0.85	54.9	58.1	55.9	68.4	51.5	60.2	57.5	72.8
Laplace	1.00	3.4	5.3	4.8	6.5	2.9	4.9	4.8	6.2
	0.95	42.7	39.0	36.6	28.9	38.2	36.1	34.0	25.5
	0.90	64.1	62.3	59.9	47.5	59.7	60.1	57.0	48.7
	0.85	77.1	78.7	78.7	58.9	77.9	79.4	78.1	63.9
t_3	1.00	2.8	5.1	4.7	7.3	2.8	4.2	5.0	7.1
	0.95	35.2	28.6	26.9	21.5	27.9	26.9	25.2	23.0
	0.90	57.4	53.0	50.5	39.3	58.1	60.5	58.2	40.3
	0.85	77.9	76.1	73.9	52.1	80.8	84.7	82.8	55.9
t_2	1.00	4.3	5.5	4.9	8.8	3.6	5.6	5.1	8.8
	0.95	54.4	51.1	48.8	28.0	54.2	52.5	50.1	25.7
	0.90	80.2	80.3	77.9	41.7	81.9	82.7	81.4	43.4
	0.85	91.5	91.5	90.3	51.0	92.0	94.6	94.2	49.2

Table 2: Size and power of the unit-root tests (×100) with $\alpha \in (1,2)$ and $n = 300$

		ε_t ~model (5.1)				ε_t ~model (5.2)			
$\eta_t \sim$	ϕ	$\tilde{T}_n$	$l(1)$	$l^a(1)$	ELT	$\tilde{T}_n$	$l(1)$	$l^a(1)$	ELT
N(0,1)	1.00	3.6	5.5	5.3	5.7	3.6	4.6	4.7	5.4
	0.95	71.6	66.7	65.6	76.6	73.1	68.4	67.0	75.0
	0.90	94.2	92.3	91.6	96.6	96.1	96.5	95.4	97.6
	0.85	98.9	98.5	98.3	99.3	99.5	99.5	99.4	99.8
Laplace	1.00	3.7	4.9	4.7	6.4	3.3	4.3	4.9	5.7
	0.95	96.3	94.1	93.8	79.0	93.9	92.8	92.6	80.4
	0.90	99.6	99.9	99.6	92.8	99.7	99.9	99.7	94.7
	0.85	100	100	100	94.8	100	100	100	98.2
t_3	1.00	4.3	5.4	5.2	6.5	3.9	5.3	5.0	6.9
	0.95	91.4	86.3	85.5	84.4	91.9	88.8	88.9	59.9
	0.90	99.4	99.1	99.1	76.4	99.7	99.7	99.7	83.7
	0.85	100	100	100	89.3	100	100	100	91.9
t_2	1.00	3.6	4.7	4.9	8.0	3.5	4.4	4.8	7.7
	0.95	99.2	98.7	98.5	59.3	99.1	98.8	98.8	59.4
	0.90	100	100	100	73.1	100	100	100	75.3
	0.85	100	100	100	79.8	100	100	100	82.7

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

Table 3: Size and power of the unit-root tests (×100) with $\alpha \in (0,1)$ and $n = 100$

		ε_t ∼model (5.1)				ε_t ∼model (5.2)			
$\eta_t \sim$	ϕ	$\tilde{T}_n$	$l(1)$	$l^a(1)$	ELT	$\tilde{T}_n$	$l(1)$	$l^a(1)$	ELT
N(0,1)	1.00	3.9	5.0	4.8	5.6	3.2	5.0	4.7	6.1
	0.95	21.9	21.6	19.2	23.5	22.4	19.4	17.4	21.9
	0.90	37.9	31.4	30.0	39.7	36.9	34.0	31.1	44.5
	0.85	49.0	43.7	41.2	55.8	53.5	51.3	49.4	60.4
Laplace	1.00	4.1	5.5	4.7	6.3	3.3	4.6	4.9	5.9
	0.95	47.8	41.1	40.7	29.7	48.7	41.7	41.2	29.1
	0.90	66.8	58.8	56.7	43.8	66.9	61.0	58.2	45.3
	0.85	78.0	72.9	70.9	51.4	78.5	75.7	74.6	58.1
t_2	1.00	3.3	4.6	4.5	9.1	3.6	4.9	4.8	10.3
	0.95	51.9	41.5	41.8	26.5	51.7	44.3	43.7	24.0
	0.90	74.9	67.4	65.3	35.3	77.9	73.4	70.7	37.2
	0.85	82.9	78.5	76.7	41.4	91.6	90.2	89.1	45.4
Cauchy	1.00	4.2	4.8	4.5	21.6	4.6	5.2	4.7	19.5
	0.95	93.4	89.3	88.7	33.8	95.0	92.9	92.5	36.1
	0.90	98.2	96.2	95.8	35.5	99.2	98.1	97.8	35.3
	0.85	99.0	97.5	97.5	38.8	99.3	98.9	98.9	36.3

Table 4: Size and power of the unit-root tests (×100) with $\alpha \in (0,1)$ and $n = 300$

		ε_t ∼model (5.1)				ε_t ∼model (5.2)			
$\eta_t \sim$	ϕ	$\tilde{T}_n$	$l(1)$	$l^a(1)$	ELT	$\tilde{T}_n$	$l(1)$	$l^a(1)$	ELT
N(0,1)	1.00	4.4	4.8	4.8	7.2	4.7	4.9	4.8	6.6
	0.95	67.2	58.0	57.3	65.8	74.2	65.4	64.6	71.1
	0.90	85.6	78.4	77.3	84.4	90.6	85.8	85.4	90.1
	0.85	96.0	91.7	91.6	94.2	96.9	95.7	95.5	96.2
Laplace	1.00	5.8	5.1	5.0	6.6	4.8	4.9	4.7	6.5
	0.95	97.3	94.0	94.7	74.7	96.6	94.7	94.2	78.1
	0.90	99.4	98.9	98.9	84.0	99.8	99.6	99.6	89.3
	0.85	99.8	99.8	99.8	89.1	99.6	99.6	99.5	94.3
t_2	1.00	4.7	5.3	5.2	9.2	4.2	4.6	4.7	7.6
	0.95	98.1	95.9	95.7	49.8	98.7	97.6	97.4	50.5
	0.90	99.8	99.6	99.6	60.1	100	100	100	69.9
	0.85	100	100	100	66.8	100	100	100	72.9
Cauchy	1.00	4.2	5.4	5.2	20.5	4.4	4.5	4.6	22.6
	0.95	100	100	100	42.3	100	100	100	42.3
	0.90	100	100	100	39.1	100	100	100	39.9
	0.85	100	100	100	35.4	100	100	100	37.7

consider two common ADF-type tests:

$$R_{n,\kappa} = \frac{(n-k)(\hat{\phi}-1)}{1-\sum_{j=1}^{k}\hat{\beta}_j}, \;\; Q_{n,\kappa} = \frac{\hat{\phi}-1}{s(\hat{\phi})},$$

where $(\hat{\phi}, \hat{\beta}_1, \cdots, \hat{\beta}_k)$ is the LSE of the regression parameters in model (4.7), $s(\hat{\phi})$ is the usual standard error of $\hat{\phi}$, and $k = [\kappa(n/100)^{1/4}]$ is the selected lag length, with $\kappa = 4$ and 12. To implement these two tests in heavy-tailed cases, we employ the wild sieve bootstrap method proposed by Cavaliere, Georgiev, and Taylor (2018) with bootstrap size $b = 1000$, and denote the associated tests as $R^b_{n,\kappa}$ and $Q^b_{n,\kappa}$. The simulation results are presented in Tables 5–6.

The results show that the proposed ELTs $l(1)$ and $l^a(1)$ outperform all competitors in terms of power performance, with a stable size for nearly all cases. The wild bootstrap tests are sensitive to the selected lag length κ, where a smaller κ means higher power, which is consistent with that in Cavaliere, Georgiev, and Taylor (2018). In addition, it is interesting that the proposed tests become more powerful when the tail of the noise becomes heavier.

Supplementary Material

The Supplementary Material contains technical proofs for the results in Sections 2–4, and the simulation results for the confidence interval estimation under the alternative.

Acknowledgments

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

Table 5: Size and power of the unit-root tests (×100) for i.i.d. noise when $n = 100$

		Empirical likelihood				Wild bootstrap			
	ϕ	$\tilde{T}_n$	$l(1)$	$l^a(1)$	ELT	$Q^b_{n,4}$	$R^b_{n,4}$	$Q^b_{n,12}$	$R^b_{n,12}$
$\alpha = 2.0$	1.00	3.0	4.5	4.4	8.9	5.5	3.7	5.2	4.7
	0.95	32.1	30.4	27.8	23.6	33.0	31.8	26.9	17.9
	0.90	61.7	61.3	58.8	40.4	67.0	60.9	43.2	32.8
	0.85	80.4	82.3	80.2	53.1	84.4	82.3	55.0	40.7
$\alpha = 1.5$	1.00	4.6	5.6	5.2	10.5	5.1	4.6	6.3	5.2
	0.95	69.5	64.4	62.5	26.1	39.6	37.3	34.2	22.2
	0.90	91.2	89.7	88.9	33.2	69.9	67.6	51.9	37.7
	0.85	97.3	96.8	96.3	41.7	86.0	85.6	64.6	48.9
$\alpha = 1.0$	1.00	4.9	5.3	4.9	20.8	5.0	5.6	6.9	5.0
	0.95	98.6	97.7	97.2	29.5	56.0	50.0	45.8	35.5
	0.90	99.8	99.7	99.7	34.7	78.9	76.5	60.9	49.0
	0.85	100	100	100	34.5	88.3	88.9	72.3	59.5
$\alpha = 0.5$	1.00	4.0	5.2	4.7	44.7	4.5	4.8	9.2	6.7
	0.95	100	100	100	51.7	75.1	75.5	65.7	60.4
	0.90	100	100	100	50.0	87.4	86.2	78.4	67.3
	0.85	100	100	100	50.5	92.2	93.2	81.7	75.9

Table 6: Size and power of the unit-root tests (×100) for i.i.d. noise when $n = 300$

		Empirical likelihood				Wild bootstrap			
	ϕ	$\tilde{T}_n$	$l(1)$	$l^a(1)$	ELT	$Q^b_{n,4}$	$R^b_{n,4}$	$Q^b_{n,12}$	$R^b_{n,12}$
$\alpha = 2.0$	1.00	4.3	4.8	4.7	8.0	5.6	5.4	4.7	4.6
	0.95	89.1	84.9	84.2	53.9	91.8	93.7	78.9	76.7
	0.90	99.8	99.8	99.8	75.7	99.8	99.7	97.1	94.7
	0.85	100	100	100	84.5	100	100	98.2	95.9
$\alpha = 1.5$	1.00	3.5	4.7	4.9	11.3	4.7	5.4	4.9	4.3
	0.95	99.7	99.4	99.4	43.7	91.6	92.4	82.3	80.2
	0.90	100	100	100	54.2	99.7	99.5	95.4	94.7
	0.85	100	100	100	62.5	99.6	99.6	97.1	95.0
$\alpha = 1.0$	1.00	5.3	5.1	4.6	21.0	4.9	5.5	6.0	4.8
	0.95	100	100	100	33.5	94.4	92.7	86.6	84.9
	0.90	100	100	100	33.2	98.8	98.5	94.3	93.7
	0.85	100	100	100	35.3	99.0	99.8	97.0	95.5
$\alpha = 0.5$	1.00	4.8	5.2	4.9	46.4	4.7	4.4	5.2	5.8
	0.95	100	100	100	49.5	95.2	94.1	89.9	90.5
	0.90	100	100	100	47.9	97.7	98.9	94.1	95.5
	0.85	100	100	100	47.0	98.3	98.3	95.3	94.9

The author would like to thank the Editor, Associate Editor, and two anonymous reviewers for their insightful comments and suggestions. I also would like to thank Professor Shiqing Ling for his guidance on the GARCH process, and Professor Deyuan Li for helpful discussions on the empirical likelihood method. Rui She's research was supported by Fundamental Research Funds for the Central Universities (Nos. JBK1806002 and JBK190905) of China.

References

Avram, F. and Taqqu, M. (1992). Weak convergence of sums of moving averages in the α-stable domain of attraction. *Ann. Probab.* **20**, 483–503.

Billingsley, P. (1999). *Convergence of Probability Measures, 2nd Ed.* Wiley, New York.

Caner, M. (1998). Tests for cointegration with infinite variance errors. *J. Econom.* **86**, 155–175.

Cavaliere, G., Georgiev, I. and Taylor, A. M. R. (2018). Unit root inference for non-stationary linear processes driven by infinite variance innovations. *Econom. Theory* **34(02)**, 302–348.

Chan, N.H. (2009). Time series with roots on or near the unit circle. In *Handbook of Financial Time Series,* 695–707. Springer-Verlag, New York.

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

Chan, N. H. and Tran, L. T. (1989). On the first-order autoregressive process with infinite variance. *Econom. Theory* **5**, 354–362.

Chan, N. H. and Zhang, R.M. (2010). Inference for unit-root models with infinite variance GARCH errors. *Statistica Sinica*, **20**, 1363–1393.

Chen, J., Variyath, A. M. and Abraham, B. (2008). Adjusted empirical likelihood and its properties. *J. Comput. Graph. Stat.* **17**, 426–443.

Dickey, D.A. and Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* **74**, 427–431.

Engle, R.F. (1990). Discussion: Stock market volatility and the crash of 1987. *Rev. Financ. Stud.* **3**, 103–106.

Evans, G.B.A. and Savin, N.E. (1981). Testing for Unit Roots:1. *Econometrica*, **49**, 753–782.

Fornari, F. and Mele, A. (1997). Sign- and Volatility-switching ARCH models: theory and applications to international stock arkets. *J. Appl. Econom.* **12**, 1779–1801.

Francq, C. and Zakoïan, J. (2013). Estimating the marginal law of a time series with applications to heavy-tailed distributions. *J. Bus. Econ. Stat.* **31**, 412–425.

Glosten, L. R., Jagannathan, R. and Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *J. Finance* **48**, 1779–1801.

Horváth, L. and Kokoszka, P.S. (2003). A bootstrap approximation to a unit root test statistic for heavy tailed observations. *Stat. Probab. Lett.* **62**, 163–173.

Huang, H.T., Leng, X., Liu, X.H. and Peng, L. (2020). Unified inference for an AR process regardless of finite or infinite variance GARCH errors. *J. Financ. Econom.* **18(2)**, 425–470.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

Jakubowski, A. (1997). A non-Skorohod topology on the Skorohod space. *Electron. J. Probab.* **2**, 1–21.

Knight, K. (1991). Limit theory for M-estimates in an integrated infinite variance process. *Econom. Theory* **7**, 200–212.

Koedijk, K.G. and Kool, C.J.M. (1992). Tail estimates of East European exchange rates. *J. Bus. Econ. Stat.* **10**, 83–96.

Li, D., Chan, N. H. and Peng, L. (2014). Empirical likelihood test for causality for bivariate AR(1) processes. *Econom. Theory* **30**, 357–371.

Moreno M. and Romo J. (2012). Unit root bootstrap tests under infinite variance. *J. Time Ser. Anal.* **33**, 32–47.

Owen, A.B. (2001). *Empirical Likelihood*, New York: Chapman & Hall/CRC.

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

Phillips, P.C.B. (1990). Time series regression with a unit root and infinite-variance errors. *Econom. Theory* **6**, 44–62.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimation equations. *Ann. Statist.* **22**, 300–325.

Rachev, S.T. (2003). Handbook of Heavy Tailed Distributions in Finance. Elsevier/North-Holland.

Resnick, S. I. (1997). Heavy tail modeling and teletraffic data. (With discussion and rejoinder by author). *Ann. Statist.* **25**, 1805–1869.

Samarakoon, D.M.M. and Knight, K. (2009). A note on unit root tests with infinite variance noise. *Econom.*

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

Rev. **28(4)**, 314–334.

Schwert, G.W. (1989). Why does stock market volatility change over time? *J. Finance* **45**, 1129–1155.

Sentana, E. (1995). Quadratic ARCH models. *Rev. Econ. Stud.* **62**, 639–661.

She, R. and Ling, S. (2020). Inference in heavy-tailed vector error correction models. *J. Econom.* **214**, 433–450.

Shin, D.W. and So, B.S. (1999). New tests for unit roots in autoregressive processes with possibly infinite variance errors. *Stat. Probab. Lett.* **44**, 387–397.

Taylor, S. (1986). *Modelling Financial Time Series*, New York: Wiley.

Zakoïan, J. M. (1994). Threshold heteroskedastic models. *J. Econ. Dyn. Control* **18**, 931–955.

Zhang, R.M. and Chan, N. H. (2020). Nonstationary linear processes with infinite variance GARCH errors. *Econom. Theory*, **0**, 1–34.

Zhang, R. M. and Ling, S. (2015) Asymptotic inference for AR models with heavy-tailed G-GARCH noises. *Econom. Theory*, **31**, 880–890.

Zhang, R.M., Sin, C.Y. and Ling, S. (2015). On functional limits of short- and long-memory linear processes with GARCH(1,1) noises. *Stoch. Process. Their Appl.* **125**, 482–512.

Zheng, M. and Yu, W. (2013). Empirical likelihood method for multivariate Cox regressions. *Comput. Stat.* **28**, 1241–1267.

Rui She

Center of Statistical Research, School of Statistics,

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0415

Southwestern University of Finance and Economics, Chengdu, China.

E-mail: rshe@swufe.edu.cn