Submitted to

STRICT STATIONARITY TESTING AND GLOBAL ROBUST QUASI-MAXIMUM LIKELIHOOD ESTIMATION OF DAR MODELS

By Shaojun Guo[§], Dong Li[†] and Muyi Li[‡]

Chinese Academy of Sciences and London School of Economics[§], *Tsinghua University*[†], *and Xiamen Unviersity*[‡]

The assumption of strict stationarity is pivotal when a double autoregressive (DAR) model is used. Yet, tractable tools are unavailable for testing strict stationarity in the DAR framework. In this article we attempt to provide a procedure for this test. We formulate such testing problem as testing if a top Lyapunov exponent is negative or not and introduce the t-type tests. To achieve this goal, a consistent estimator of the associate top Lyapunov exponent without strict stationarity assumption is presented and a random weighting approach is suggested for variance estimation. It is shown that such t-type tests are consistent and powerful, and the resampling method is very useful in capturing the sampling variability of the variance. We also propose a global robust quasi-maximum likelihood estimation (QMLE) for parameters of interest, which weakens key assumptions on the commonly used QMLE. All estimators, except for the intercept, are shown to be consistent and asymptotically normal in both stationary and explosive situations. Their asymptotic variances, although quite different in both cases, can be consistently estimated via random weighting approach in a unified framework. The finite sample performance of the proposed inference procedures are examined via Monte Carlo simulation studies and a real example is analyzed.

1. Introduction. Strict stationarity assumption is pivotal in modeling nonlinear time series and very helpful in obtaining asymptotic properties of inference on a given time series model. A variety of nonlinear models, such as popular autoregressive conditional heteroscedastic (ARCH) model, generalized ARCH (GARCH) model, double autoregressive (DAR) model and

[§]Partly supported by Key Laboratory of RCSDS in Chinese Academy Sciences and an EPSRC research grant in UK.

[†]Partly supported by the start-up fund of Tsinghua University (No.553310013) and NSFC (No.11401337 and No.11571348)

[‡]Partly supported by NSFC(No. 11301433) and Key Laboratory of Econometrics (Xiamen University), Ministry of Education

MSC 2010 subject classifications: Primary 60M10; secondary 60F05, 62F12

Keywords and phrases: DAR models, Nonstationarity, Quasi-Maximum Exponential Likelihood Estimation, Random Weighting, Strict Stationarity Testing.

their numerous variants, have been proposed, and the associate strict stationarity conditions have been extensively studied, see Douc, et al. (2014) for a general overview. In contrast, practitioners would often ask the question: *Is the strict stationarity assumption necessary when a time series model is applied to a group of data?* This is quite challenging to pursuit an affirmative answer generally. An ad hoc approach is to handle this problem on a class-by-class basis. Nevertheless, only few results are available in literature so far. A notable exception is that Francq and Zakoïan (2012, 2013) provided inference tools for testing strict stationarity and nonstationarity under the GARCH-based framework and paves a new way for this important topic. Besides GARCH, DAR model (see also Weiss (1984) and Ling (2004)) is another important conditional heteroscedastic model and recently receives much attention in statistics and economics. One aim of this paper is to develop tractable procedures for testing strict stationarity under the DAR framework.

The DAR model inherits merits of AR models, targeting more on the conditional mean given the past data, and of ARCH models, concentrating on the conditional variance under the historical data. The first-order DAR (hereafter DAR (1)) model is defined as

$$y_t = \phi_0 y_{t-1} + \eta_t \sqrt{\omega_0 + \alpha_0 y_{t-1}^2}, \quad t = 0, \pm 1, \pm 2, ..., \tag{1.1}$$

where $\phi_0 \in R$, $\omega_0 > 0$, $\alpha_0 > 0$, $\{\eta_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables and independent of $\{y_j, j < t\}$. Model (1.1) is a special case of the ARMA-ARCH models in Weiss (1984) and of the nonlinear AR models in Cline and Pu (2004), but it is different from the ARCH model in Engle (1982) if $\phi_0 \neq 0$. The high-order DAR model and its generalizations can be found in Weiss (1984), Tsay (1987), Lu (1998), Ling (2004, 2007), Zhu and Ling (2013), Li, Ling and Zakoïan (2015), Li, et al. (2015), Li, Ling and Zhang (2016), etc.

The strict stationarity condition for model (1.1) has been well conducted and is closely related to the top Lyapunov exponent

$$\gamma_0 = E \log |\phi_0 + \eta \sqrt{\alpha_0}|,$$

when η is symmetric, where η is a generic random variable with the same distribution as η_t and independent of η_t, see Borkovec and Klüppelberg (2001), Ling (2004, 2007), Chen, Li and Ling (2014) and among others. Particularly, Borkovec and Klüppelberg (2001) proved that $\gamma_0 < 0$ is sufficient for strict stationarity of model (1.1), while Chen, Li and Ling (2014) complemented that it is (almost) necessary. Thus, testing strict stationarity is equivalent to testing the following hypothesis:

$$H_0: \ \gamma_0 < 0 \quad v.s. \quad H_1: \ \gamma_0 \geq 0. \tag{1.2}$$

To test (1.2), one challenge is how to obtain a good estimator of γ_0 under the null and alternative hypotheses. Surprisingly, in contrast to the majority of literature focusing on inference of $(\omega_0, \phi_0, \alpha_0)$, there is limited attention to estimating γ_0. This may be partly because it is nonstandard and the associate asymptotic theory is much difficult. Fig. 1 shows the strictly stationary region of model (1.1) for three different distributions of η, indicating that γ_0 depends heavily on the innovation and intrinsic parameters.

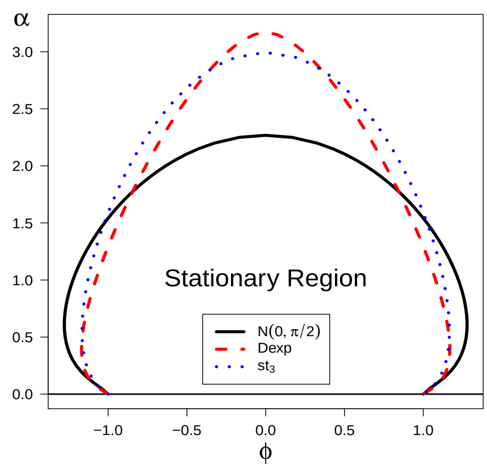

FIG 1. *The (strictly) stationary region determined by* $\{(\phi,\alpha) : E\log|\phi+\eta\sqrt{\alpha}| < 0\}$ *for* $\eta \sim N(0,\pi/2)$*, the Laplace distribution with density* $f(x) = \exp(-|x|)/2$*, and the standardized Student's* t_3 (st_3) *with density* $f(x) = 4\pi^2/(\pi^2+4x^2)^2$.

In recent years, the stationarity concern has been strongly voiced. Ling (2004) considered a testing problem:

$$H_0: \ (\phi_0, \alpha_0) = (\pm 1, 0) \quad v.s. \quad H_1: \ (\phi_0, \alpha_0) \neq (\pm 1, 0). \tag{1.3}$$

Under the null hypothesis H_0 in (1.3), $\{y_t\}$ is the standard unit root process. However, as the classical unit root tests, we cannot claim that $\{y_t\}$ is stationary even if H_0 in (1.3) is rejected. He also pointed out that how to test $\gamma_0 < 0$, i.e., (1.2), remains an interesting issue. Chan and Peng (2005) discussed the estimation of γ_0 in the stationary case via simulation studies so as to check the stationarity condition. To the best of our knowledge, they are among the first to formally consider estimation of γ_0. Nevertheless, they did not consider its asymptotic properties such as consistency and asymptotic normality, even in the stationary case, hampering the applicability of γ_0 in (1.2).

This paper has two major contributions. First, we propose data-driven procedures for testing (1.2). The basic ingredients for the procedures are as follows. In the first step, we provide a natural estimator of γ_0 and demonstrate that the proposed estimator shares desirable asymptotic properties without any stationarity assumption. In the second step, we develop a random weighting method to approximate its asymptotic covariance. The nice feature is that such a procedure does not rely on asymptotic expression of variance and turns out to be very helpful in capturing the sampling uncertainty adaptively. Based on these estimators, in the last step, we prefer t-type test statistic for testing strict stationarity and give the consistent critical region of such test. The performance of the procedure is illustrated largely via simulation examples. As a by-product, we discuss also the construction of confidence intervals and testing nonstationarity (3.8) in Section 3. Recently, Francq and Zakoïan (2012) proposed an inference tool for testing stationarity under the GARCH(1,1) framework. The basic idea is similar but the underlying asymptotic theories own large gaps, resulting that both testing procedures are different. Roughly speaking, their proof techniques fail to apply here. To solve this problem, we borrow the strength from modern empirical process theory of dependent data and provide new theoretical insights about testing stationarity problems in other scenarios. An intuitive discussion for their difference can be found in Section 3.

Second, we offer a unified framework for parameter estimation of model (1.1) in the stationary and explosive situations. This framework does not require the finite fourth moment of the innovation η_t and hence can be applied for heavy-tailed cases. Specifically, we shall propose an unconstrained global robust quasi-maximum likelihood estimation (QMLE) for $(\phi_0, \alpha_0, \omega_0)$ in the sense that the value of ω_0 is not fixed. Here, 'global' means that the convergence rate of the estimator is first obtained and the limiting distribution then derived, see Zhu and Ling (2011, 2013). Under mild conditions, the proposed estimator of (ϕ_0, α_0) are always strongly consistent and asymptotically normal whether the data mechanism is stationary or explosive. To measure the accuracy of the estimator, a random weighting method is proposed to estimate their asymptotic covariance matrix consistently in a unified framework. It is worthy noting that even if the robust QMLE of (ϕ_0, α_0) is consistent in every situation, the intercept term ω_0 is only consistent in the stationary case. For this reason, it is meaningful to test the sign of γ_0.

There are a lot of literature focusing on inference via the QMLE. Ling (2004, 2007) considered the QMLE of $(\phi_0, \alpha_0, \omega_0)$ and proved its asymptotics under the conditions $\gamma_0 < 0$ and $E\eta^4 < \infty$. On the other hand, in the explosive case, i.e., $\gamma_0 > 0$, Ling and Li (2008) investigated a constrained QMLE of (ϕ_0, α_0) in the sense that the value of the intercept term ω_0 is fixed and obtained the asymptotic normality when η is standard normal. Recently, Chen, Li and Ling (2014) studied an unconstrained QMLE of (ϕ_0, α_0) when η is symmetric with $E\eta^4 < \infty$. One disadvantage of the QMLE is that the assumption of $E\eta^4 < \infty$ is indispensable for valid inference. However, this may fail in practice and consequently the standard QMLE procedure may not be reliable. To tackle this difficulty, a robust procedure is provided, see Chan and Peng (2005) and Zhu and Ling (2013) for the stationary case. In presence of non-stationarity and heavy-tailed noises, the inference becomes more challenging and, to the best of our knowledge, no results are available so far in the literature. This is also one of our aims in this paper.

The remainder of the paper is organized as follows. Section 2 considers the robust QMLE with asymptotic properties and discusses how to estimate the asymptotic variance matrix via random weighting approach. In Section 3, we discuss the estimation of γ_0 and tests of stationarity and nonstationarity. The associate asymptotics are established. Section 4 reports numerical results on the performance of our proposed methodology in finite samples, supporting our theories in Sections 2 and 3. The stationarity tests for daily returns of 6 major stock indices in the world are analyzed in Section 4. A conclusion remark is in Section 5 and all the technical proofs are relegated to the Appendix.

2. Robust Quasi-Maximum Likelihood Estimation.

2.1. *Robust Estimation and its Asymptotic Properties.* Suppose that the observations $\{y_0, y_1, ..., y_n\}$ are from model (1.1). When η is double exponential, the log-likelihood function (ignoring a constant) can be written as

$$L_n(\theta) = \sum_{t=1}^{n} \left\{ \frac{1}{2} \log(\omega + \alpha y_{t-1}^2) + \frac{|y_t - \phi y_{t-1}|}{\sqrt{\omega + \alpha y_{t-1}^2}} \right\},$$

where $\theta = (\phi, \alpha, \omega)^T$ is the parameter. The proposed estimator is defined as

$$\hat{\theta}_n = (\hat{\phi}_n, \hat{\alpha}_n, \hat{\omega}_n)^T = \arg\min_{\theta \in \Theta} L_n(\theta), \tag{2.1}$$

where Θ is a compact subset of $R \times R_+^2$ containing the true value $\theta_0 = (\phi_0, \alpha_0, \omega_0)^T$. Here $R_+ = (0, \infty)$. Since we do not assume that η is double exponential, the estimator $\hat{\theta}_n$ is often called quasi-maximum exponential likelihood estimator as in Zhu and Ling (2013) or least absolute deviation estimator (LADE) as in Chan and Peng (2005). Since it shares nice robust properties as LADE, we refer to $\hat{\theta}_n$ the robust QMLE (or rQMLE) throughout the paper.

THEOREM 2.1. *Suppose that $\{\eta_t\}$ is i.i.d. and symmetric with $E|\eta_t| = 1$ and Θ is compact. Then, for the DAR(1) model (1.1), the rQMLE defined in (2.1) satisfies the following properties.*
(i). *If $\gamma_0 < 0$, then $\hat{\phi}_n \to \phi_0$, $\hat{\alpha}_n \to \alpha_0$, and $\hat{\omega}_n \to \omega_0$ a.s. as $n \to \infty$.*
(ii). *If $\gamma_0 > 0$, then $\hat{\phi}_n \to \phi_0$ and $\hat{\alpha}_n \to \alpha_0$ a.s. as $n \to \infty$.*

To further discuss asymptotic normality of $\hat{\theta}_n$, we need three assumptions.

ASSUMPTION 1. *$\{\eta_t\}$'s are i.i.d. with $E|\eta_1| = 1$ and $\kappa_\eta = E\eta_1^2 < \infty$.*

ASSUMPTION 2. *The density $f(x)$ of η_1 is symmetric and bounded continuous on R with $f(0) > 0$.*

ASSUMPTION 3. *Θ is compact and θ_0 is the interior point of Θ.*

THEOREM 2.2. *Suppose that Assumptions 1–3 hold.*
(i). *If $\gamma_0 < 0$, then*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1) \quad \textit{and} \quad \sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N\big(0,\ \mathcal{J}_S\big)$$

as $n \to \infty$, where '$\xrightarrow{\mathcal{L}}$' stands for the convergence in distribution,

$$\mathcal{J}_S = \operatorname{diag}\left\{\frac{1}{4\sigma_{11} f^2(0)},\ 4(\kappa_\eta - 1)\Sigma^{-1}\right\}, \tag{2.2}$$

and both Σ and σ_{11} are defined as

$$\Sigma = \begin{pmatrix} \sigma_{22} & \sigma_{12} \\ \sigma_{12} & \sigma_{02} \end{pmatrix} \quad \textit{with} \quad \sigma_{ij} = E\left\{\frac{y_t^{2i}}{(\omega_0 + \alpha_0 y_t^2)^j}\right\}, i, j = 0, 1, 2. \tag{2.3}$$

(ii). *If $\gamma_0 > 0$, then*

$$\sqrt{n}(\hat{\phi}_n - \phi_0, \hat{\alpha}_n - \alpha_0)^T = O_p(1) \textit{ and } \sqrt{n}(\hat{\phi}_n - \phi_0, \hat{\alpha}_n - \alpha_0)^T \xrightarrow{\mathcal{L}} N\Big(0, \mathcal{J}_N\Big),$$

as $n \to \infty$, where

$$\mathcal{J}_N = \operatorname{diag}\Big\{\frac{\alpha_0}{4f^2(0)},\ 4\alpha_0^2(\kappa_\eta - 1)\Big\}. \tag{2.4}$$

REMARK 1. *From Theorem 2.2, we can see that our rQMLE is global, i.e., the convergence rate is first obtained and the limiting distribution then derived, see Zhu and Ling (2011, 2013). This is totally different from the LADE of linear regression or time series models and other local LADE of nonlinear models. On the other hand, it is worth noting that* $\hat{\omega}_n$ *is inconsistent when* $\gamma_0 > 0$. *A similar phenomenon was observed by Chen, Li and Ling (2014), who studied an unconstrained QMLE of an explosive DAR(1) model, and by Francq and Zakoïan (2012) who studied the QMLE of nonstationary GARCH(1,1) models, see also Jensen and Rahbek (2004a,b).*

REMARK 2. *We remark here the hidden relationship between expressions (2.2) and (2.4), which reveals why several results in the stationary situation can still be applicable in the explosive one. Note that all* σ_{ij}*'s in (2.3) are finite constants when* $\gamma_0 < 0$ *since* $\{y_t\}$ *is strictly stationary and ergodic. In the explosive situation, i.e.* $\gamma_0 > 0$, *however,* $|y_t|$ *diverges to infinity at an exponential rate as* $t \to \infty$, *see Theorem 1 in Chen, Li and Ling (2014) or Theorem 2.1 in Liu, et al. (2016). Thus,* $\sigma_{11} \to 1/\alpha_0$, $\sigma_{22} \to 1/\alpha_0^2$ *and other* σ_{ij}*'s go to zero a.s. as* $t \to \infty$, *which implies that* $\mathcal{J}_S$ *can be reduced to* $\mathrm{diag}(\mathcal{J}_N, 0)$.

2.2. *Variance Estimation.* Theorem 2.2 demonstrates that the proposed global robust estimator shares nice asymptotic normality in both situations. To evaluate the accuracy of the estimator, we need a consistent estimator of their asymptotic covariance matrix. Unlike the QMLE, however, their asymptotic covariance matrix involves the density function $f(\cdot)$ and cannot be properly estimated using the plug-in rules. One choice is to use the nonparametric kernel method to estimate the density, see Zhu and Ling (2013). However, this method involves the choice of the bandwidth, which is another troublesome problem. To avoid density estimation, we propose a resampling method via random weighting to covariance matrix estimation. (Related works include, for example, Chen, et al. (2008, 2010) for regression models and Zhu (2015) for time series models). To be specific, let $\varpi_1, \cdots, \varpi_n$ be a sequence of i.i.d. nonnegative random variables, with both mean and variance equal to one. For example, the standard exponential distribution satisfies this requirement. Define

$$L_n^*(\theta) = \sum_{t=1}^{n} \varpi_t \left\{ \frac{1}{2} \log(\omega + \alpha y_{t-1}^2) + \frac{|y_t - \phi y_{t-1}|}{\sqrt{\omega + \alpha y_{t-1}^2}} \right\}$$

and $\hat{\theta}_n^* = \arg\min_{\theta \in \Theta} L_n^*(\theta)$. In the stationary case, the distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ can be approximated by the resampling distribution of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$. It turns out that, in the explosive case, the distribution of $\sqrt{n}(\hat{\phi}_n - \phi_0, \hat{\alpha}_n - \alpha_0)$ can still be approximated by that of $\sqrt{n}(\hat{\phi}_n^* - \hat{\phi}_n, \hat{\alpha}_n^* - \hat{\alpha}_n)$.

THEOREM 2.3. *Suppose Assumptions 1–3 hold.*
(i). *If $\gamma_0 < 0$, then, conditionally on the data $\{y_0, y_1, \cdots, y_n\}$, in probability,*

$$\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \xrightarrow{\mathcal{L}} N(0,\ \mathcal{J}_S), \quad \text{as } n \to \infty;$$

(ii). *If $\gamma_0 > 0$, then, given the data $\{y_0, y_1, \cdots, y_n\}$, in probability,*

$$\sqrt{n}(\hat{\phi}_n^* - \hat{\phi}_n,\ \hat{\alpha}_n^* - \hat{\alpha}_n)^T \xrightarrow{\mathcal{L}} N(0,\ \mathcal{J}_N), \quad \text{as } n \to \infty.$$

The proof of Theorem 2.3 is provided in Appendix. The inference procedure via random weighting is done as follows. First, i.i.d. $\{\varpi_1, \cdots, \varpi_n\}$ are generated B times from standard exponential distribution, where B is a large number. In simulation, B is taken to be 500. Each time, the minimizer $\hat{\theta}_n^*$ is computed. Denote them as $\hat{\theta}_n^{*1}, \cdots, \hat{\theta}_n^{*B}$. Then, the asymptotic variances of $\sqrt{n}(\hat{\phi}_n - \phi_0)$ and $\sqrt{n}(\hat{\alpha}_n - \alpha_0)$ can be approximated by, respectively,

$$\hat{\sigma}_\phi^2 = \frac{1}{B-1} \sum_{b=1}^{B} \big(\hat{\phi}_n^{*b} - \bar{\phi}_n^*\big)^2 \ \text{ and } \ \hat{\sigma}_\alpha^2 = \frac{1}{B-1} \sum_{b=1}^{B} \big(\hat{\alpha}_n^{*b} - \bar{\alpha}_n^*\big)^2,$$

with $\bar{\phi}_n^* = B^{-1} \sum_{b=1}^{B} \hat{\phi}_n^{*b}$ and $\bar{\alpha}_n^* = B^{-1} \sum_{b=1}^{B} \hat{\alpha}_n^{*b}$. In the stationary case, the asymptotic variance of $\sqrt{n}(\hat{\omega}_n - \omega_0)$ is estimated analogously.

3. Strict Stationarity Testing.

As described in the Introduction, the parameter γ_0 plays a key role in characterizing the strict stationarity of DAR (1) model. It also determines the consistency of $\hat{\omega}_n$. In this section, we propose a consistent estimator $\hat{\gamma}_n$ of γ_0 and then construct t-type tests based on $\hat{\gamma}_n$ for whether $\{y_t\}_{t=1}^n$ is strictly stationary or nonstationary. The asymptotic normality of $\hat{\gamma}_n$ is established and correspondingly, a random weighting approach is suggested to estimate asymptotic variance of $\hat{\gamma}_n$ in Section 3.1. Section 3.2 shows that the t-type stationarity tests are consistent.

3.1. *Consistent Estimation of γ_0.*

Define the rescaled residuals as

$$\hat{\eta}_t = \eta_t(\hat{\theta}_n), \quad \eta_t(\theta) = \frac{y_t - \phi y_{t-1}}{\sqrt{\omega + \alpha y_{t-1}^2}}. \tag{3.1}$$

By the definition of γ_0, a natural estimator of γ_0 is

$$\tilde{\gamma}_n = \frac{1}{2n} \sum_{t=1}^{n} \big(\log|\hat{\phi}_n + \hat{\eta}_t\sqrt{\hat{\alpha}_n}| + \log|\hat{\phi}_n - \hat{\eta}_t\sqrt{\hat{\alpha}_n}|\big). \tag{3.2}$$

To facilitate our proofs, we instead propose a truncated estimator for γ_0:

$$\hat{\gamma}_n = \frac{1}{2n}\Big(\sum_{t\in\mathcal{A}_1}\log|\hat{\phi}_n + \hat{\eta}_t\sqrt{\hat{\alpha}_n}| + \sum_{t\in\mathcal{A}_2}\log|\hat{\phi}_n - \hat{\eta}_t\sqrt{\hat{\alpha}_n}|\Big), \tag{3.3}$$

where $\mathcal{A}_1 = \{t : \hat{\phi}_n + \hat{\eta}_t\sqrt{\hat{\alpha}_n} \in \mathcal{I}_n, 1 \le t \le n\}$ and $\mathcal{A}_2 = \{t : \hat{\phi}_n - \hat{\eta}_t\sqrt{\hat{\alpha}_n} \in \mathcal{I}_n, 1 \le t \le n\}$ with $\mathcal{I}_n = [-n^2, -n^{-2}] \cup [n^{-2}, n^2]$. Indeed, both estimators are almost identical unless $|\hat{\phi}_n + \hat{\eta}_t\sqrt{\hat{\alpha}_n}|$ or $|\hat{\phi}_n - \hat{\eta}_t\sqrt{\hat{\alpha}_n}|$ is extremely small for some t.

To prove the asymptotics of $\hat{\gamma}_n$, we impose the following assumption.

Assumption 4. *The density $f(\cdot)$ of η_1 is positive and differentiable a.s. on R with $\sup_{x\in R} f(x) < \infty$, and $\int \log|\phi_0 + x\sqrt{\alpha_0}|(f(x) + |xf'(x)|)dx$ exists and is finite.*

Remark 3. *The most commonly used distributions, like normal and Student's t distributions, satisfy this condition. The Laplace distribution is also allowed since it is differentiable except only one point.*

To state the asymptotic normality of $\hat{\gamma}_n$, we first introduce some notation. Denote

$$\mu_1 = -\alpha_0^{-1/2}\int \log|\phi_0 + x\sqrt{\alpha_0}|f'(x)dx,$$
$$\mu_2 = -\alpha_0^{-1/2}\int \log|\phi_0 + x\sqrt{\alpha_0}|\{f(x) + xf'(x)\}dx.$$

For each $t = 1, \cdots, n$, define

$$\zeta_t = (\zeta_{1t} + \zeta_{2t})/2, \quad \zeta_{1t} = \log|\phi_0 + \eta_t\sqrt{\alpha_0}| - E(\log|\phi_0 + \eta_t\sqrt{\alpha_0}|),$$
$$\zeta_{2t} = \log|\phi_0 - \eta_t\sqrt{\alpha_0}| - E(\log|\phi_0 - \eta_t\sqrt{\alpha_0}|).$$

For $\gamma_0 < 0$, denote $\xi_t = \left(\xi_{1t}, \xi_{2t}^T\right)^T$, where

$$\xi_{1t} = \frac{1}{2\sigma_{11}f(0)}\frac{y_{t-1}\mathrm{sign}(\eta_t)}{\sqrt{\omega_0 + \alpha_0 y_{t-1}^2}}, \quad \xi_{2t} = \Sigma^{-1}\Big(\frac{2y_{t-1}^2}{\omega_0 + \alpha_0 y_{t-1}^2}, 2\Big)^T (|\eta_t| - 1),$$

and $\nu = (\nu_1, \nu_2)^T$, where

$$\nu_1 = E\frac{\omega_0}{2\sqrt{\alpha_0}\left(\omega_0 + \alpha_0 y_{t-1}^2\right)}, \quad \nu_2 = -E\frac{\sqrt{\alpha_0}}{2\left(\omega_0 + \alpha_0 y_{t-1}^2\right)}.$$

In the case of $\gamma_0 > 0$, we denote

$$\tilde{\xi}_{1t} = \frac{\sqrt{\alpha_0}}{2f(0)}\text{sign}(\eta_t), \quad t = 1, \cdots, n.$$

The following theorem gives the asymptotic properties of $\hat{\gamma}_n$.

THEOREM 3.1. *Suppose that Assumptions 1–4 hold.*
(i). *If* $\gamma_0 < 0$, *then*

$$\sqrt{n}\,(\hat{\gamma}_n - \gamma_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big(\zeta_t + \mu_1\xi_{1t} + \mu_2\nu^T\xi_{2t}\big) + o_p(1)$$
$$\xrightarrow{\mathcal{L}} N\big(0,\ E\big(\zeta_t + \mu_1\xi_{1t} + \mu_2\nu^T\xi_{2t}\big)^2\big).$$

(ii). *If* $\gamma_0 > 0$, *then*

$$\sqrt{n}\,(\hat{\gamma}_n - \gamma_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big(\zeta_t + \mu_1\tilde{\xi}_{1t}\big) + o_p(1)$$
$$\xrightarrow{\mathcal{L}} N\big(0,\ E\big(\zeta_t + \mu_1\tilde{\xi}_{1t}\big)^2\big).$$

Here we provide a heuristic analysis so as to understand why its associate theory involves $f(\cdot)$ and $f'(\cdot)$ and is different from that in Francq and Zakoïan (2012). In Appendix, we will give detailed proofs. To illustrate it, we consider the following toy problems. Suppose that we observe i.i.d. data $\{(X_i, Z_i)\}_{i=1}^n$ sampled from a bivariate normal variable (X, Z) with mean $(0, a)^T$, where a is an unknown parameter and $a \neq 0$. We wish to estimate two quantities: $\gamma_1 = E\log|X + a|$ and $\gamma_2 = E\log(X^2 + a^2)$. It is easy to see that γ_1 is similar to γ_0 in a DAR(1) model. Recall that the resulting top Lyapunov exponent is $\gamma_0 = E\log(\beta_1 + \alpha_1\epsilon_1^2)$ in a GARCH(1,1) model:

$$y_t = \sigma_t\epsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1\sigma_{t-1}^2,$$

where $\{\epsilon_t\}_{t=1}^n$ is an i.i.d. sequence of random variables with mean zero and unit variance, and the coefficients satisfy $\alpha_0 > 0$ and $\alpha_1, \beta_1 \geq 0$. The quantity γ_2 is closely related to γ_0 in GARCH(1,1) model. Our questions are: (a) *How to estimate* γ_1 *and* γ_2*?* (b) *How to derive their asymptotic normality of the proposed estimators?*

Obviously, natural estimators of both γ_1 and γ_2 are

$$\hat{\gamma}_1 = \frac{1}{n}\sum_{i=1}^{n}\log|X_i + \hat{a}_n| \quad \text{and} \quad \hat{\gamma}_2 = \frac{1}{n}\sum_{i=1}^{n}\log(X_i^2 + \hat{a}_n^2),$$

respectively, where $\hat{a}_n = n^{-1}\sum_{i=1}^n Z_i$. Let us look at $\hat{\gamma}_2$ first. To derive asymptotic normality of $\hat{\gamma}_2$, we rewrite $\hat{\gamma}_2 - \gamma_2$ as

$$\begin{aligned}\hat{\gamma}_2 - \gamma_2 &= \frac{1}{n}\sum_{i=1}^n \Big\{ \log(X_i^2 + \hat{a}_n^2) - \log(X_i^2 + a^2)\Big\} + \frac{1}{n}\sum_{i=1}^n \Big\{ \log(X_i^2 + a^2) - \gamma_2\Big\} \\ &:= I_1 + I_2.\end{aligned}$$

The analysis of I_2 is trivial. The i.i.d. representation of I_1 can be derived by Taylor's expansion. To be specific,

$$I_2 = \frac{1}{n}\sum_{i=1}^n \frac{2a}{X_i^2 + a^2}\big(\hat{a}_n - a\big) + O_p((\hat{a}_n - a)^2).$$

Since $n^{-1}\sum_{i=1}^n \big(X_i^2 + a^2\big)^{-1} \to b = E\big(X_i^2 + a^2\big)^{-1}$ a.s., $\sqrt{n}(\hat{\gamma}_2 - \gamma_2)$ is further expressed as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \Big\{2ab(Z_i - a) + \log(X_i^2 + a^2) - \gamma_2\Big\} + o_p(1),$$

and hence the asymptotic normality follows.

Consider the estimator $\hat{\gamma}_1$ now. Similarly, decompose $\hat{\gamma}_1 - \gamma_1$ as

$$\begin{aligned}\hat{\gamma}_1 - \gamma_1 &= \frac{1}{n}\sum_{i=1}^n (\log|X_i + \hat{a}_n| - \log|X_i + a|) + \frac{1}{n}\sum_{i=1}^n (\log|X_i + a| - \gamma_1) \\ &:= I_1 + I_2.\end{aligned}$$

The key lies in I_1. Since $\log|x + a| = \frac{1}{2}\log(x + a)^2$, by Taylor's expansion, the term I_1 becomes

$$I_1 = \frac{1}{n}\sum_{i=1}^n \frac{1}{X_i + a}(\hat{a}_n - a) + o_p(1) = E\big(X + a\big)^{-1}(\hat{a}_n - a) + o_p(1).$$

Unfortunately, $E(X + a)^{-1}$ does not exist even for normal distribution. In other words, the limit of $n^{-1}\sum_{i=1}^n (X_i + a)^{-1}$ may be infinite and, as a result, this Taylor's expansion approach fails.

A possible remedy is to apply modern empirical process theory to I_1. Denote empirical and true distributions of X by $P_n(x) = n^{-1}\sum_{i=1}^n I(X_i \le x)$ and $P(x) = P(X \le x)$, where $I(A)$ is an indicator function of a set A. Then $I_1 = \int_{x\in R} \log|x| d(P_n(x - \hat{a}_n) - P_n(x - a))$. Empirical process approximations enable us to obtain that

$$\sup_{x\in R, |a_1 - a| \le \delta_n} |P_n(x - a_1) - P(x - a_1) - P_n(x - a) + P(x - a)| = o_p(n^{-1/2-\epsilon}),$$

provided by $\delta_n \downarrow 0$ at some rate with some $\epsilon > 0$, and hence, intuitively,

$$I_1 = \int_{x\in R} \log|x| d(P(x-\hat{a}_n) - P(x-a)) + o_p(n^{-1/2}). \tag{3.4}$$

(Of course, this assertion is not easy to prove and needs to be analyzed carefully since $\log|x|$ is unbounded on R. For instance, we added a truncation into our estimator. One reason is to prove (3.4) in a much easier way.) Through this approximation, I_1 reduces to

$$\begin{aligned} I_1 &= \int \log|x| \big\{ f(x-\hat{a}_n) - f(x-a) \big\} dx + o_p(n^{-1/2}) \\ &= -\int \log|x| f'(x-a) dx (\hat{a}_n - a) + o_p(n^{-1/2}), \end{aligned}$$

where $f(\cdot)$ is the density of X. Define $c = -\int \log|x+a| f'(x)dx$. Then,

$$\sqrt{n}(\hat{\gamma}_1 - \gamma_1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ c(Z_i - a) + \log|X_i + a| - \gamma_1 \} + o_p(1). \tag{3.5}$$

This explains how different estimator of γ_1 is from that of γ_2. It also illustrates why asymptotic variance of $\hat{\gamma}_1$ involves the derivative function $f'(\cdot)$.

In the following, we discuss how to estimate the asymptotic variance of $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$. Since it involves the density function $f(\cdot)$ and its derivative $f'(\cdot)$, the asymptotic expression would not be applicable directly. With similar spirits in Section 2, we propose a random weighting approach for variance estimation. In particular, let $\varpi_1, \cdots, \varpi_n$ be a sequence of i.i.d. nonnegative random variables, with both mean and variance equal to one. Define

$$L_n^*(\theta) = \sum_{t=1}^{n} \varpi_t \left\{ \frac{1}{2} \log(\omega + \alpha y_{t-1}^2) + \frac{|y_t - \phi y_{t-1}|}{\sqrt{\omega + \alpha y_{t-1}^2}} \right\}$$

and $\hat{\theta}_n^* = \arg\min_{\theta\in\Theta} L_n^*(\theta)$. For $t = 1, \cdots, n$ and $k = 1, 2$, denote

$$\begin{aligned} &\hat{\eta}_t^* = \eta_t(\hat{\theta}_n^*), \quad \mathcal{A}_k^* = \{t : \hat{\eta}_t^* \sqrt{\hat{\alpha}_n^*} - (-1)^k \hat{\phi}_n^* \in \mathcal{I}_n\}, \\ &\hat{\gamma}_{nk}^* = \frac{1}{\sum_{t\in\mathcal{A}_k^*} \varpi_t} \sum_{t\in\mathcal{A}_k^*} \varpi_t \log|\hat{\eta}_t^* \sqrt{\hat{\alpha}_n^*} - (-1)^k \hat{\phi}_n^*|. \end{aligned} \tag{3.6}$$

The final resampling estimator of $\hat{\gamma}_n$ is defined as

$$\hat{\gamma}_n^* = (\hat{\gamma}_{n1}^* + \hat{\gamma}_{n2}^*)/2.$$

To construct the resampling estimator, we put the random weights into two locations in (3.6). They exactly correspond to two different parts in (3.5). The following theorem shows that the distribution of $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$ can be approximated by the resampling distribution of $\sqrt{n}(\hat{\gamma}_n^* - \hat{\gamma}_n)$.

THEOREM 3.2. *Suppose that Assumptions 1–4 hold and $\gamma_0 \neq 0$. Then, conditional on the data $\{y_0, y_1, \cdots, y_n\}$, in probability, $\sqrt{n}(\hat{\gamma}_n^* - \hat{\gamma}_n)$ is asymptotically normal as $n \to \infty$, and its asymptotic variance is the same as that of $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$.*

In practice, the asymptotic variance of $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$ can be estimated by using random weighting as follows. First, i.i.d. $\{\varpi_1, \cdots, \varpi_n\}$ are generated B times from standard exponential distribution, where B is a large number. Each time, minimize $L_n^*(\theta)$ to obtain $\hat{\theta}_n^*$ and then $\hat{\gamma}_n^*$ is computed. Denote them as $\hat{\gamma}_n^{*1}, \cdots, \hat{\gamma}_n^{*B}$. Then, the asymptotic variance σ_γ^2 of $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$ is approximated by

$$\hat{\sigma}_\gamma^2 = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\gamma}_n^{*b} - \bar{\gamma}_n^*\right)^2 \quad \text{with} \quad \bar{\gamma}_n^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\gamma}_n^{*b}.$$

Once variance estimation is given, a confidence interval is constructed, as stated in the following corollary.

COROLLARY 3.1. *Under the conditions in Theorem 3.1 and $B = O(n)$, $\hat{\sigma}_\gamma^2 \to \sigma_\gamma^2$ in probability as $n \to \infty$. Therefore, at the significance level $\underline{\alpha} \in (0,1)$, a confidence interval for γ_0 is*

$$\left[\hat{\gamma}_n - \frac{\hat{\sigma}_\gamma}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right),\ \hat{\gamma}_n + \frac{\hat{\sigma}_\gamma}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right],$$

where $\Phi(\cdot)$ is the standard normal distribution.

3.2. *Strict Stationarity Testing.* Consider the strict stationarity testing

$$H_0: \ \gamma_0 < 0 \quad \text{v.s.} \quad H_1: \ \gamma_0 \geq 0, \tag{3.7}$$

and

$$H_0: \ \gamma_0 > 0 \quad \text{v.s.} \quad H_1: \ \gamma_0 \leq 0. \tag{3.8}$$

With the aid of estimation of γ_0 and its asymptotics, the above tests are tractable. We give asymptotic critical regions for both testing problems as follows.

THEOREM 3.3. *Suppose that Assumptions 1–4 hold. Let*

$$T_n = \sqrt{n}\frac{\hat{\gamma}_n}{\hat{\sigma}_\gamma} \tag{3.9}$$

be the test statistics for (3.7) and (3.8), where $\hat{\sigma}_\gamma$ *is estimated with the resampling size* $B = O(n)$.

(i). *For the test (3.7), the test defined by the stationary (ST) critical region*

$$\mathrm{C}^{\mathrm{ST}} = \left\{T_n > \Phi^{-1}(1-\underline{\alpha})\right\}$$

has its asymptotic significance level $\underline{\alpha}$ *and is consistent for all* $\gamma_0 > 0$.

(ii). *For the test (3.8), the test defined by the nonstationary (NS) critical region*

$$\mathrm{C}^{\mathrm{NS}} = \{T_n < \Phi^{-1}(\underline{\alpha})\}$$

has its asymptotic significance level $\underline{\alpha}$ *and is consistent for all* $\gamma_0 < 0$.

Here we use the normal approximation to construct critical regions for (3.7) and (3.8). An alternative approach is to use the resampling distribution of $\hat{\gamma}_n$. If such approach is applied, the resampling size would be taken to be larger than here.

4. Numerical Studies.

4.1. *Simulation Studies.* In this subsection, we first conduct numerical studies to evaluate the finite sample performance of our global robust QMLE and strict stationarity tests. In particular, we compare our proposed estimation method with the QMLE studied in Ling (2004) and Chen, Li and Ling (2014) in finite samples.

First, we are interested in the performance of the estimators $(\hat{\phi}_n, \hat{\alpha}_n)$ and $\hat{\gamma}_n$ in finite samples. We generate the innovations $\{\eta_t\}$ in the following three scenarios: (a) *the normal distribution* $N(0, \pi/2)$; (b) *the Laplace distribution with density* $f(x) = 0.5\exp(-|x|)$; and (c) *the standardized Student's* t_3 (st_3) *distribution with density* $f(x) = 4\pi^2/(\pi^2 + 4x^2)^2$. The true parameters are set to be $(\phi_0, \alpha_0, \omega_0)^T = (0.7, 0.4, 0.5)^T$ and $(1.0, 3.0, 0.5)^T$, corresponding to the stationary and explosive cases, respectively. The sample size n is 200 and 400 and 1000 replications are used for each configuration. We take the resampling size $B = 500$ for variance estimation for each configuration. For all cases, the true top Lyapunov exponents have been calculated in Table 1.

TABLE 1

Summary statistics for our proposed estimation procedures under various scenarios based on 1000 replications

		$n = 200$				$n = 400$			
η	Parameters	Bias	SE	SEE	CP	Bias	SE	SEE	CP
Normal	$\phi_0 = 0.700$	-0.003	0.102	0.106	0.944	-0.004	0.072	0.074	0.948
	$\alpha_0 = 0.400$	-0.014	0.085	0.082	0.925	-0.007	0.056	0.058	0.951
	$\gamma_0 = -0.523$	-0.005	0.109	0.113	0.958	-0.005	0.074	0.076	0.950
Laplace	$\phi_0 = 0.700$	-0.002	0.068	0.073	0.963	-0.001	0.048	0.049	0.954
	$\alpha_0 = 0.400$	-0.009	0.107	0.102	0.915	-0.005	0.072	0.072	0.941
	$\gamma_0 = -0.440$	-0.005	0.086	0.093	0.969	-0.002	0.060	0.062	0.960
st_3	$\phi_0 = 0.700$	-0.002	0.082	0.085	0.943	0.001	0.057	0.058	0.940
	$\alpha_0 = 0.400$	-0.008	0.140	0.124	0.895	-0.009	0.087	0.084	0.903
	$\gamma_0 = -0.473$	-0.003	0.106	0.110	0.952	-0.001	0.071	0.075	0.954
Normal	$\phi_0 = 1.000$	0.011	0.193	0.200	0.939	0.003	0.139	0.139	0.942
	$\alpha_0 = 3.000$	-0.009	0.338	0.326	0.932	-0.006	0.223	0.228	0.952
	$\gamma_0 = 0.242$	0.004	0.068	0.070	0.951	0.003	0.045	0.048	0.953
Laplace	$\phi_0 = 1.000$	0.002	0.128	0.137	0.953	-0.001	0.090	0.094	0.959
	$\alpha_0 = 3.000$	-0.001	0.438	0.430	0.923	-0.002	0.292	0.302	0.946
	$\gamma_0 = 0.227$	0.001	0.073	0.075	0.948	-0.001	0.049	0.052	0.958
st_3	$\phi_0 = 1.000$	-0.006	0.158	0.159	0.942	0.001	0.109	0.110	0.949
	$\alpha_0 = 3.000$	0.022	0.536	0.521	0.921	0.001	0.375	0.356	0.926
	$\gamma_0 = 0.183$	0.005	0.075	0.079	0.954	0.001	0.054	0.055	0.947

Note: Bias and SE are the finite sample bias and standard error of the parameter estimator, SEE means the mean of standard error estimator, and CP is the empirical coverage probabilities of the 95% confidence intervals.

In Table 1, we report the finite sample biases (Bias), the standard errors (SE), the sample mean of the standard error estimators (SEE) and the empirical coverage probabilities (CP) of the 95% confidence intervals via normal approximation. We observe that all estimators are virtually unbiased, the variance estimators accurately reflect the true variations and the coverage probabilities agree well with the nominal level 95% for almost all cases. These findings confirm that the stationarity assumption is not necessary for estimation of these parameters. They also illustrate that the resampling approach works well for variance estimation. We comment that in some cases, the coverage rates of α_0 is slightly lower than the nominal level 95%. One possible reason is that $\sqrt{n}(\hat{\alpha}_n - \alpha_0)$ is skewed relative to the normal distribution, especially for small positive α_0, since all $\hat{\alpha}_n$ are positive. The basic idea for improvement is to incorporate the skewness information into confidence interval construction. This may be done, for example, by better confidence intervals proposed by Efron (1987).

Next, we evaluate the finite sample performance of our estimators $(\hat{\phi}_n, \hat{\alpha}_n)$ by comparing with the QMLE. We consider $n = 200$ and generate 1000 replications for each configuration. Since both estimation procedures require different conditions on η for identification in model (1.1), it is not a good idea to compare them directly. To make feasible comparison, we use the average absolute errors (AAE), defined as

$$\mathrm{AAE} = \frac{1}{2}\Big(\Big|\hat{\phi} - \phi_0\Big| + \Big|\frac{\hat{\alpha}}{\alpha_0} - 1\Big|\Big),$$

where $(\hat{\phi}, \hat{\alpha})$ are the rQMLE or QMLE of (ϕ_0, α_0), respectively. Unlike Zhu and Ling (2013), the advantage of this AAE is that it is not necessary to adjust the estimators for both procedures.

Fig. 2 depicts box-plots of AAE for the rQMLE and QMLE. It is observed from Fig. 2 that the rQMLE outperforms the QMLE when the innovation η follows from st_3 and Laplace distribution, supporting the robust properties of our rQMLE, either stationary or explosive. In particular, the QMLE performs much worse when errors have infinite fourth moments. We also observe that the QMLE with normal errors behaves better. It is not surprising since in this case, the QMLE is efficient, while the rQMLE is not. Nevertheless, their difference is small.

Finally, we illustrate the behavior of strict stationarity tests developed in Section 3 through simulations. To access the performance, we keep the standardized Student's t_3 distribution for η_t, $\omega_0 = 0.5$ but ϕ_0 varies from 0.6 to 1.3 and $\alpha_0 = 2\phi_0$. In this scenario, we have $\gamma_0 = 0$ for $(\phi_0, \alpha_0) = (0.922, 1.844)$ and hence, $\gamma_0 > 0, < 0$ if $\phi_0 > 0.922$ or $\phi_0 < 0.922$, respectively. We take the sample size $n = 200, 400, 800$ and 1000 replications for each configuration.

Tables 2 and 3 summarized empirical frequencies of rejection for (3.7) and (3.8) for various values of ϕ_0, respectively. It is observed that when $\gamma_0 = 0$, the rejection frequencies of the two tests agree with the nominal level 5% as n increases. As expected, the frequency of rejection of the C^{ST} test increases with γ_0, while that of the C^{NS} test decreases. Overall, the power of the two tests is significant, as shown in theory.

TABLE 2
Relative frequency of rejection of the test (3.7): $H_0 : \gamma_0 < 0$ with $(\omega_0, \alpha_0) = (0.5, 2\phi_0)$ based on 1000 replications

	ϕ_0						
	0.600	0.700	0.800	0.922	1.000	1.100	1.300
$n = 200$	0.000	0.000	0.001	0.053	0.173	0.528	0.943
$n = 400$	0.000	0.000	0.000	0.062	0.252	0.754	1.000
$n = 800$	0.000	0.000	0.000	0.057	0.452	0.957	1.000

4.2. *Empirical Study.* We applied the proposed strict stationarity tests to the daily returns of 6 major stock market indices. We collected the dai-

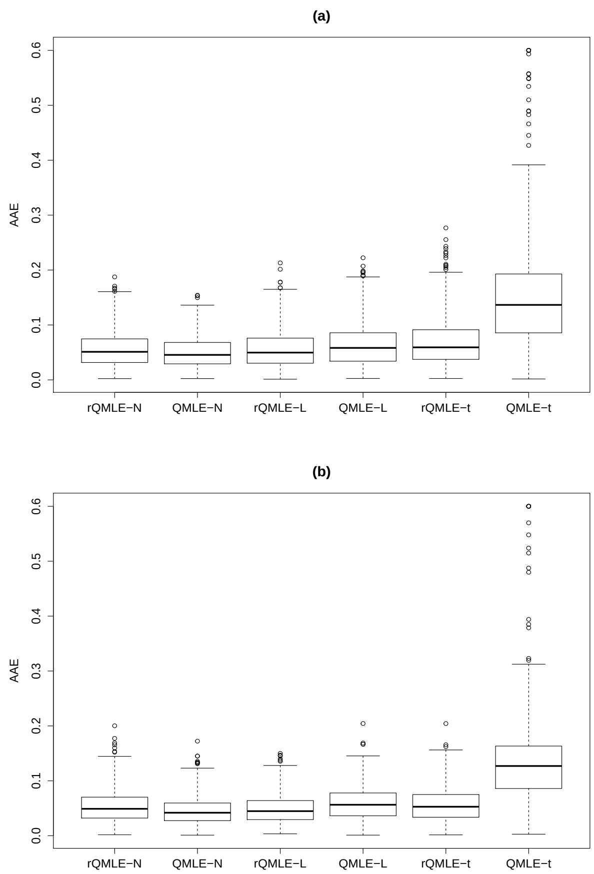

FIG 2. *Boxplots of the average absolute errors for the rQMLE and the QMLE based on 1000 replications. (a)* $(\phi_0, \alpha_0, \omega_0)^T = (0.7, 0.4, 0.5)^T$ *and (b)* $(\phi_0, \alpha_0, \omega_0)^T = (1.0, 3.0, 0.5)^T$. *'rQMLE-N','rQMLE-L' and 'rQMLE-t' mean 'rQMLE's when errors are normal, Laplace and* st_3*, respectively. 'QMLE-N',QMLE-L' and 'QMLE-t' are defined similarly.*

TABLE 3
Relative frequency of rejection of the test (3.8): $H_0 : \gamma_0 > 0$ with $(\omega_0, \alpha_0) = (0.5, 2\phi_0)$ based on 1000 replications

	ϕ_0						
	0.600	0.700	0.800	0.922	1.000	1.100	1.300
$n = 200$	0.954	0.713	0.313	0.038	0.009	0.000	0.000
$n = 400$	0.996	0.933	0.501	0.048	0.003	0.000	0.000
$n = 800$	1.000	1.000	0.796	0.054	0.002	0.000	0.000

ly returns for CAC, JAX, NASDAQ, NIKKEI, SMI and SP500 from the period January 2011 to May 2015. The focus here is to check whether the daily returns of each index over this period is strictly stationary by fitting a DAR(1) model. If the significance level is set to be 5%, then this stationarity hypothesis can be tested by simply comparing the t-type statistic T_n to 1.64 and rejecting the stationary hypothesis if $|T_n| > 1.64$. To calculate T_n, the resampling size is taken to be 1000 for variance estimation of $\hat{\gamma}_n$. Table 4 displays the test statistic T_n defined in (3.9) for each time series. Since all values of T_n shown in Table 4 are negative and very small, the stationarity assumption of a DAR(1) model cannot be rejected for any of these series.

TABLE 4
Test Statistic T_n of the strict stationarity tests (3.7)

Index	CAC	DAX	NASDAQ	NIKKEI	SMI	SP500
T_n	-11.11	-11.18	-9.9	-10.07	-10.16	-8.71

5. Conclusion. Testing for strict stationarity is an important issue in the context of nonlinear time series. This paper concerns strict stationarity testing of DAR models and develops a unified framework for the inference of both stationary and explosive DAR(1) models. We proposed a global robust QMLE of (ϕ_0, α_0) in the DAR(1) model and established their asymptotic theory without strict stationarity assumption. If one is interested in inference on ϕ_0 or α_0, stationarity testing is unnecessary.

When we focus on applications of a DAR model in which ω_0 is used (e.g. estimating conditional variance $\text{var}(y_t|y_{t-1})$), testing stationarity is needed since the rQMLE or QMLE of ω_0 is consistent if $\gamma_0 < 0$.

In this paper, one major assumption is that η is symmetric. This is because the dynamic behavior of y_t is unknown if the symmetric assumption is violated. To relax this condition is an interesting problem and we leave it as future study.

APPENDIX A: PROOFS OF THEOREMS

Denote by M any positive constant whose value is unimportant and can be different throughout the proofs. Recall the fact that

$$|y_t|/\rho^t \to \infty \quad \text{a.s. as } t \to \infty \tag{A.1}$$

for any $\rho \in (1, e^{\gamma_0})$ by Theorem 1 in Chen, Li and Ling (2014).

A.1. Proof of Theorem 2.1. The result stated in (i) is standard, see Zhu and Ling (2013). Consider the case (ii). Clearly, $(\hat{\phi}_n, \hat{\alpha}_n, \hat{\omega}_n)^T = \arg\min_{\theta\in\Theta} Q_n(\theta)$, where $Q_n(\theta) = \{L_n(\theta) - L_n(\theta_0)\}/n$. We have

$$Q_n(\theta) = O_n(\phi, \alpha) + R_{1n}(\theta) + R_{2n}(\theta),$$

where

$$O_n(\phi, \alpha) = \frac{1}{n}\sum_{t=1}^{n}\left\{\frac{1}{2}\log\frac{\alpha}{\alpha_0} + \frac{|\eta_t \text{sign}(y_{t-1})\sqrt{\alpha_0} - (\phi - \phi_0)|}{\sqrt{\alpha}} - |\eta_t|\right\}$$

and

$$R_{1n}(\theta) = \frac{1}{2n}\sum_{t=1}^{n}\log\frac{\alpha_0(\omega + \alpha y_{t-1}^2)}{\alpha(\omega_0 + \alpha_0 y_{t-1}^2)},$$

$$R_{2n}(\theta) = \frac{1}{2n}\sum_{t=1}^{n}\left\{\frac{|\eta_t\sqrt{\omega_0 + \alpha_0 y_{t-1}^2} - (\phi - \phi_0)y_{t-1}|}{\sqrt{\omega + \alpha y_{t-1}^2}} - \frac{|\eta_t \text{sign}(y_{t-1})\sqrt{\alpha_0} - (\phi - \phi_0)|}{\sqrt{\alpha}}\right\}.$$

Note that $(\eta_1\text{sign}(y_0), ..., \eta_n\text{sign}(y_{n-1}))$ and $(\eta_1, ..., \eta_n)$ have the same distribution. Thus,

$$\begin{aligned}\lim_{n\to\infty} O_n(\phi, \alpha) &= \tfrac{1}{2}\log\frac{\alpha}{\alpha_0} + \frac{E|\eta_t\sqrt{\alpha_0} - (\phi - \phi_0)|}{\sqrt{\alpha}} - 1 \\ &\geq \tfrac{1}{2}\log(\alpha/\alpha_0) + \sqrt{\alpha_0/\alpha} - 1 \geq 0\end{aligned}$$

because $E|\eta_t - c| \geq E|\eta_t|$ for any $c \in R$ and the inequality $\log x \leq x - 1$ for $x > 0$. The equality holds if and only if $\phi = \phi_0$ and $\alpha = \alpha_0$.

For $R_{1n}(\theta)$, by the mean value theorem, the compactness of Θ and (A.1), we have

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}|R_{1n}(\theta)| \leq M\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{1 + y_{t-1}^2} \to 0 \quad \text{a.s.}$$

For $R_{2n}(\theta)$, by the compactness of Θ and (A.1), a simple calculus yields

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}|R_{2n}(\theta)| \le M \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\frac{|\eta_t|}{|y_{t-1}|}\to 0 \quad \text{a.s.}$$

Thus, the proof can be completed by standard arguments, invoking the compactness of Θ.

A.2. Proof of Theorem 2.2. (i). When $\gamma < 0$, $\{y_t\}$ is strictly stationary and ergodic, and there exists some $\iota > 0$ such that $E|y_t|^\iota < \infty$, see Borkovec and Klüppelberg (2001). Thus, Assumption 2 in Zhu and Ling (2013) is satisfied and in turn the result holds.

(ii). We first reparameterize the log-likelihood function as

$$H_n(u) = L_n(\theta_0 + u) - L_n(\theta_0),$$

where $u \in \Lambda := \{u = (u_1, u_2, u_3)^T : u + \theta_0 \in \Theta\}$.

Let $\hat{u}_n = \hat{\theta}_n - \theta_0 := (\hat{u}_{1n}, \hat{u}_{2n}, \hat{u}_{3n})^T = (\hat{v}_n^T, \hat{u}_{3n})^T$. We can see that $\hat{u}_n$ is the minimizer of $H_n(u)$ in Λ. By Theorem 2.1(ii), $\hat{v}_n = o_p(1)$. Using the similar arguments as in Zhu and Ling (2011, 2013), by the fact (A.1), we have

$$H_n(\hat{u}_n) = (\sqrt{n}\hat{v}_n)^T T_n + (\sqrt{n}\hat{v}_n)^T \Omega(\sqrt{n}\hat{v}_n) + o_p(\sqrt{n}\|\hat{v}_n\| + n\|\hat{v}_n\|^2),$$

where

$$T_n = -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big(\frac{\mathrm{sign}(\eta_t)}{\sqrt{\alpha_0}}, \frac{|\eta_t|-1}{2\alpha_0}\Big)^T \text{ and } \Omega = \mathrm{diag}(f(0)/\alpha_0,\ 1/(8\alpha_0^2)).$$

By the central limit theorem, we have $T_n \xrightarrow{\mathcal{L}} N(0, \mathrm{diag}(1/\alpha_0,\ (\kappa_\eta - 1)/(4\alpha_0^2)))$. Let $\lambda_{\min} = \min\{f(0)/\alpha_0,\ 1/(8\alpha_0^2)\} > 0$. Then

$$H_n(\hat{u}_n) \ge -\|\sqrt{n}\hat{v}_n\|\{\|T_n\| + o_p(1)\} + \|\sqrt{n}\hat{v}_n\|^2\{\lambda_{\min} + o_p(1)\}.$$

Note that $H_n(\hat{u}_n) \le 0$ by the definition of $\hat{\theta}_n$. Thus,

$$\|\sqrt{n}\hat{v}_n\| \le \{\lambda_{\min} + o_p(1)\}^{-1}\{\|T_n\| + o_p(1)\} = O_p(1).$$

Next, let $v_n^* = -\Omega^{-1}T_n/(2\sqrt{n})$. Then,

$$\sqrt{n}v_n^* = -\Omega^{-1}T_n/2 \xrightarrow{\mathcal{L}} N(0, \mathcal{J}_N).$$

Using the previous facts, a simple calculus gives that

$$\begin{aligned} H_n(\hat{u}_n) - H_n(u_n^*) &= (\sqrt{n}\hat{v}_n - \sqrt{n}v_n^*)^T \Omega (\sqrt{n}\hat{v}_n - \sqrt{n}v_n^*) + o_p(1) \\ &\geq \lambda_{\min} \|\sqrt{n}\hat{v}_n - \sqrt{n}v_n^*\|^2 + o_p(1), \end{aligned}$$

where $u_n^* = (v_n^{*T}, 0)^T$. Note that $H_n(\hat{u}_n) - H_n(u_n^*) = L_n(\theta_0 + \hat{u}_n) - L_n(\theta_0 + u_n^*) \leq 0$ a.s. Thus, we have $\|\sqrt{n}\hat{v}_n - \sqrt{n}v_n^*\| = o_p(1)$.

Finally, we have $\sqrt{n}(\hat{\phi}_n - \phi_0,\ \hat{\alpha}_n - \alpha_0)^T = \sqrt{n}\hat{v}_n = \sqrt{n}v_n^* + o_p(1)$ and then (ii) holds.

A.3. Proof of Theorem 2.3. The proof is similar to that of Theorem 2.2. We now prove the results for $\gamma_0 > 0$. Let P^* be the joint probability of $(\varpi_1, \cdots, \varpi_n)$ and $(y_1, \cdots, y_n)$. Similar to Theorem 2.2, it can be shown that

$$\begin{aligned} \sqrt{n}(\hat{\phi}_n^* - \phi_0) &= \frac{\sqrt{\alpha_0}}{2f(0)} \frac{1}{\sqrt{n}} \sum_{t=1}^n \varpi_t \mathrm{sign}(\eta_t) + o_{P^*}(1), \\ \sqrt{n}(\hat{\alpha}_n^* - \alpha_0) &= \frac{2\alpha_0}{\sqrt{n}} \sum_{t=1}^n \varpi_t \{|\eta_t| - 1\} + o_{P^*}(1). \end{aligned}$$

Note that $(\hat{\phi}_n^* - \hat{\phi}_n) = (\hat{\phi}_n^* - \phi_0) - (\hat{\phi}_n - \phi_0)$ and so does for $(\hat{\alpha}_n^* - \hat{\alpha}_n)$. Hence, together with asymptotic representations of $\sqrt{n}(\hat{\phi}_n - \phi_0)$ and $\sqrt{n}(\hat{\alpha}_n - \alpha_0)$ in the proof of Theorem 2.2, we immediately obtain that

$$\begin{aligned} \sqrt{n}(\hat{\phi}_n^* - \hat{\phi}_n) &= \frac{\sqrt{\alpha_0}}{2f(0)} \frac{1}{\sqrt{n}} \sum_{t=1}^n (\varpi_t - 1) \mathrm{sign}(\eta_t) + o_{P^*}(1), \\ \sqrt{n}(\hat{\alpha}_n^* - \hat{\alpha}_n) &= \frac{2\alpha_0}{\sqrt{n}} \sum_{t=1}^n (\varpi_t - 1) \{|\eta_t| - 1\} + o_{P^*}(1), \end{aligned}$$

and therefore, conditional on the data $\{y_0, y_1, \cdots, y_n\}$, the part (ii) follows. The proof of (i) is proved analogously.

A.4. Proof of Theorem 3.1. First, we decompose $\hat{\phi}_n + \hat{\eta}_t\sqrt{\hat{\alpha}_n}$ for each $t = 1, \cdots, n$. In fact, $\hat{\phi}_n + \hat{\eta}_t\sqrt{\hat{\alpha}_n}$ is rewritten as

$$\begin{aligned} \hat{\phi}_n + \hat{\eta}_t\sqrt{\hat{\alpha}_n} = \hat{\phi}_n &+ \frac{y_t - \hat{\phi}_n y_{t-1}}{\sqrt{\hat{\omega}_n + \hat{\alpha}_n y_{t-1}^2}} \sqrt{\hat{\alpha}_n} \\ &= \hat{\phi}_n + \frac{(\phi_0 - \hat{\phi}_n) y_{t-1}}{\sqrt{\hat{\omega}_n + \hat{\alpha}_n y_{t-1}^2}} \sqrt{\hat{\alpha}_n} + \sqrt{\frac{\omega_0 + \alpha_0 y_{t-1}^2}{\hat{\omega}_n + \hat{\alpha}_n y_{t-1}^2}} \eta_t \sqrt{\hat{\alpha}_n} \\ &= \phi_0 + \eta_t\sqrt{\alpha_0} + v_t^{(1)}(\hat{\theta}_n) + \eta_t u_t(\hat{\theta}_n). \end{aligned}$$

where $\hat{\theta}_n = (\hat{\phi}_n, \hat{\alpha}_n, \hat{\omega}_n)^T$,

$$v_t^{(1)}(\theta) = (\phi - \phi_0)\Big(1 - \frac{y_{t-1}\sqrt{\alpha}}{\sqrt{\omega + \alpha y_{t-1}^2}}\Big), \quad u_t(\theta) = \sqrt{\frac{\omega_0 + \alpha_0 y_{t-1}^2}{\omega + \alpha y_{t-1}^2}}\sqrt{\alpha} - \sqrt{\alpha_0}.$$

Similarly,

$$\hat{\eta}_t\sqrt{\hat{\alpha}_n} - \hat{\phi}_n = \eta_t\sqrt{\alpha_0} - \phi_0 + v_t^{(2)}(\hat{\theta}_n) + \eta_t u_t(\hat{\theta}_n).$$

where $v_t^{(2)}(\theta) = -(\phi - \phi_0)\Big(1 + \frac{y_{t-1}\sqrt{\alpha}}{\sqrt{\omega+\alpha y_{t-1}^2}}\Big)$.

(I) Consider the stationary case, i.e. $\gamma_0 < 0$. Let $v_{nt}^{(k)}(\vartheta) = n^{1/2} v_t^{(1)}(\theta_0 + n^{-1/2}\vartheta)$, $k = 1, 2$, and $u_{nt}(\vartheta) = n^{1/2} u_t(\theta_0 + n^{-1/2}\vartheta)$. For $k = 1, 2$, define

$$\begin{aligned} H_{nk}(x;\vartheta) &= \frac{1}{n}\sum_{t=1}^n I\Big(\eta_t \le \frac{x + (-1)^k\phi_0 - n^{-1/2}v_{nt}^{(k)}(\vartheta)}{\sqrt{\alpha_0} + n^{-1/2}u_{nt}(\vartheta)}\Big), \\ H_k(x;\vartheta) &= \frac{1}{n}\sum_{t=1}^n F\Big(\frac{x + (-1)^k\phi_0 - n^{-1/2}v_{nt}^{(k)}(\vartheta)}{\sqrt{\alpha_0} + n^{-1/2}u_{nt}(\vartheta)}\Big). \end{aligned}$$

If we denote $\hat{\vartheta}_n = n^{1/2}(\hat{\theta}_n - \theta_0)$ and $\mathcal{I}_n = [-n^2, -n^{-2}] \cup [n^{-2}, n^2]$, then

$$2\hat{\gamma}_n = \int_{\mathcal{I}_n} \log|x| dH_{n1}(x;\hat{\vartheta}_n) + \int_{\mathcal{I}_n} \log|x| dH_{n2}(x;\hat{\vartheta}_n) = \hat{\gamma}_{n1} + \hat{\gamma}_{n2}.$$

We decompose the $\hat{\gamma}_{nk} - \gamma_0$ $(k = 1, 2)$, respectively, as

$$\begin{aligned} \hat{\gamma}_{nk} - \gamma_0 &= \int_{\mathcal{I}_n} \log|x| d\big(H_{nk}(x;\hat{\vartheta}_n) - H_{nk}(x;0) - H_k(x;\hat{\vartheta}_n) + H_k(x;0)\big) \\ &\quad + \int_{\mathcal{I}_n} \log|x| d(H_{nk}(x;0) - H_k(x;0)) + \int_{R/\mathcal{I}_n} \log|x| dH_k(x;0) \\ &\quad + \int_{\mathcal{I}_n} \log|x| d(H_k(x;\hat{\vartheta}_n) - H_k(x;0)) \\ &= I_{n1}^{(k)} + I_{n2}^{(k)} + I_{n3}^{(k)} + I_{n4}^{(k)}. \end{aligned}$$

In the following, we will show that

(a) $\sqrt{n} I_{n1}^{(k)} = o_p(1), \quad k = 1, 2;$

(b) $\sqrt{n}\big(I_{n2}^{(1)} + I_{n2}^{(2)}\big) = \dfrac{1}{\sqrt{n}}\displaystyle\sum_{t=1}^n \big(\zeta_{1t} + \zeta_{2t}\big) + o_p(1);$

(c) $\sqrt{n} I_{n3}^{(k)} = o_p(1), \quad k = 1, 2;$

(d) $\sqrt{n}\big(I_{n4}^{(1)} + I_{n4}^{(2)}\big) = \dfrac{1}{\sqrt{n}}\displaystyle\sum_{t=1}^n \big(2\mu_1\xi_{1t} + 2\mu_2\nu^T\xi_{2t}\big) + o_p(1).$

Suppose that (a)-(d) are proved. It immediately follows that

$$\sqrt{n}\,(\hat{\gamma}_n - \gamma_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \big(\zeta_t + \mu_1 \xi_{1t} + \mu_2 \nu^T \xi_{2t}\big) + o_p(1),$$

and hence $\sqrt{n}\,(\hat{\gamma}_n - \gamma_0)$ is asymptotically normal.

(a) Consider the term $I_{n1}^{(k)}(k = 1, 2)$. If follows from Lemma B.1 below that, for any fixed $0 < b < \infty$, there exists a positive constant $\epsilon > 0$ such that, for $k = 1, 2$,

$$\sup_{x \in R,\, \|\vartheta\| \le b} \sqrt{n}\big|H_{nk}(x;\vartheta) - H_{nk}(x;0) - H_k(x;\vartheta) + H_k(x;0)\big| = O_p(n^{-\epsilon}),$$

and therefore

$$\sup_{x \in R} |H_{nk}(x;\hat{\vartheta}_n) - H_{nk}(x;0) - H_k(x;\hat{\vartheta}_n) + H_k(x;0)| = O_p\big(n^{-1/2-\epsilon}\big).$$

Note that, through integration by parts, for a function $G(x)$,

$$\int_{n^{-2}}^{n^2} \log|x| dG(x) = \log(x)G(x)|_{n^{-2}}^{n^2} - \int_{n^{-2}}^{n^2} G(x) d\log(x).$$

Hence,

$$\Big|\int_{n^{-2}}^{n^2} \log|x| dG(x)\Big| \le 8\log(n) \sup_{x \in [n^{-2}, n^2]} |G(x)|.$$

Similarly, $\big|\int_{-n^2}^{-n^{-2}} \log|x| dG(x)\big| \le 8\log(n) \sup_{x \in [-n^2, -n^{-2}]} |G(x)|$. Applying them to $|I_{n1}^{(k)}|$, we immediately have that, for $k = 1, 2$,

$$\begin{aligned}|I_{n1}^{(k)}| &\le 16\log(n) \sup_{x \in R} \big|H_{nk}(x;\hat{\vartheta}_n) - H_{nk}(x;0) - H_k(x;\hat{\vartheta}_n) + H_k(x;0)\big| \\ &= o_p(n^{-1/2}).\end{aligned}$$

(b) Consider the term $I_{n2}^{(k)}(k = 1, 2)$. Observe that $\sqrt{n}\big(1 - P(|\eta_t\sqrt{\alpha_0} + \phi_0 \in \mathcal{I}_n)\big) = o(1)$. Thus,

$$\begin{aligned}\sqrt{n} I_{n1}^{(k)} &= \frac{1}{n} \sum_{t=1}^{n} \Big(I(|\eta_t\sqrt{\alpha_0} - (-1)^k \phi_0| \in \mathcal{I}_n) \log|\eta_t\sqrt{\alpha_0} - (-1)^k \phi_0| \\ &\qquad - \int_{\mathcal{I}_n} \log|x| dH_k(x;0)\Big) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \zeta_{kt} + o_p(1),\end{aligned}$$

and consequently,

$$\sqrt{n}\big(I_{n1}^{(1)} + I_{n2}^{(2)}\big) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\zeta_{1t} + \zeta_{2t}) + o_p(1).$$

(c) Consider the term $I_{n3}^{(k)}$ $(k = 1, 2)$. Since $E\eta^2 < \infty$ and $f(\cdot)$ is bounded,

$$|I_{n3}^{(k)}|^2 \le E\big(\log|(-1)^{k-1}\phi_0 + \eta\sqrt{\alpha_0}|\big)^2 P(|(-1)^{k-1}\phi_0 + \eta\sqrt{\alpha_0}| \notin \mathcal{I}_n) = o(n^{-1}).$$

(d) Consider the term $I_{n4}^{(k)}$ $(k = 1, 2)$. Define

$$\tilde{v}_t^{(k)} = (-1)^{k-1}\Big(1 + \frac{(-1)^k y_{t-1}\sqrt{\alpha_0}}{\sqrt{\omega_0 + \alpha_0 y_{t-1}^2}}\Big), \tilde{u}_{1t} = \frac{1}{\sqrt{\alpha_0}}\tilde{u}_{2t}, \tilde{u}_{2t} = \frac{-\sqrt{\alpha_0}}{2(\omega_0 + \alpha_0 y_{t-1}^2)}.$$

By Taylor's expansion, $v_t^{(k)}(\hat{\theta}_n)(k = 1, 2)$ and $u_t(\hat{\theta}_n)$ can be approximated by

$$\begin{aligned} v_t^{(k)}(\hat{\theta}_n) &= \tilde{v}_t^{(k)}(\hat{\phi}_n - \phi_0) + o_p(n^{-1/2}), \\ u_t(\hat{\theta}_n) &= \tilde{u}_{1t}(\hat{\alpha}_n - \alpha_0) + \tilde{u}_{2t}(\hat{\omega}_n - \omega_0) + o_p(n^{-1/2}), \end{aligned}$$

respectively. It follows from the laws of large numbers that

$$\frac{1}{n}\sum_{t=1}^{n}\tilde{u}_{kt} \to \nu_k, a.s., k = 1, 2.$$

Hence, it follows from Lemma B.2 and Theorem 2.2 that

$$\begin{aligned} \sqrt{n}\big(I_{n4}^{(1)} + I_{n4}^{(2)}\big) &= 2\mu_1\sqrt{n}(\hat{\phi}_n - \phi_0) + 2\mu_2\nu_1\sqrt{n}(\hat{\alpha}_n - \alpha_0) \\ &\quad + 2\mu_2\nu_2\sqrt{n}(\hat{\omega}_n - \omega_0) + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big(2\mu_1\xi_{1t} + 2\mu_2\nu^T\xi_{2t}\big) + o_p(1). \end{aligned}$$

The part (d) follows.

(II) Consider the nonstationary case, i.e., $\gamma_0 > 0$. Let $v_{nt}^{(k)}(\vartheta) = n^{1/2}v_t^{(k)}\big((\phi_0 + n^{-1/2}\vartheta_1, \alpha_0 + n^{-1/2}\vartheta_2, \omega_0 + \vartheta_3)^T\big)$ and $u_{nt}(\vartheta) = n^{1/2}u_t\big((\phi_0 + n^{-1/2}\vartheta_1, \alpha_0 + n^{-1/2}\vartheta_2, \omega_0 + \vartheta_3)^T\big)$. Define

$$\begin{aligned} \tilde{H}_{nk}(x;\vartheta) &= \frac{1}{n}\sum_{t=1}^{n} I\Big(\eta_t \le \frac{x + (-1)^k\phi_0 - n^{-1/2}v_{nt}^{(k)}(\vartheta)}{\sqrt{\alpha_0} + n^{-1/2}u_{nt}(\vartheta)}\Big), \\ \tilde{H}_k(x;\vartheta) &= \frac{1}{n}\sum_{t=1}^{n} F\Big(\frac{x + (-1)^k\phi_0 - n^{-1/2}v_{nt}^{(k)}(\vartheta)}{\sqrt{\alpha_0} + n^{-1/2}u_{nt}(\vartheta)}\Big). \end{aligned}$$

Denote $\hat{\vartheta}_n = \big(n^{1/2}(\hat{\phi}_n - \phi_0), n^{1/2}(\hat{\alpha}_n - \alpha_0), \hat{\omega}_n - \omega_0\big)^T$, then

$$2\hat{\gamma}_n = \int_{\mathcal{I}_n} \log|x| d\tilde{H}_{n1}(x;\hat{\vartheta}_n) + \int_{\mathcal{I}_n} \log|x| d\tilde{H}_{n2}(x;\hat{\vartheta}_n) = \hat{\gamma}_{n1} + \hat{\gamma}_{n2}.$$

We decompose the $\hat{\gamma}_{nk} - \gamma_0$ $(k = 1, 2)$, respectively, as

$$\begin{aligned}
\hat{\gamma}_{nk} - \gamma_0 = & \int_{\mathcal{I}_n} \log|x| d\left(\tilde{H}_{nk}(x;\hat{\vartheta}_n) - \tilde{H}_{nk}(x;0) - \tilde{H}_k(x;\hat{\vartheta}_n) + \tilde{H}_k(x;0)\right) \\
& + \int_{\mathcal{I}_n} \log|x| d(\tilde{H}_{nk}(x;0) - \tilde{H}_k(x;0)) + \int_{R/\mathcal{I}_n} \log|x| d\tilde{H}_k(x;0) \\
& + \int_{\mathcal{I}_n} \log|x| d(\tilde{H}_k(x;\hat{\vartheta}_n) - \tilde{H}_k(x;0)) \\
:= & I_{n1}^{(k)} + I_{n2}^{(k)} + I_{n3}^{(k)} + I_{n4}^{(k)}.
\end{aligned}$$

To complete the proof, it suffices to show that

(a) $\quad \sqrt{n} I_{n1}^{(k)} = o_p(1), \quad k = 1, 2;$

(b) $\quad \sqrt{n}\big(I_{n2}^{(1)} + I_{n2}^{(2)}\big) = \dfrac{1}{\sqrt{n}} \sum_{t=1}^{n} \big(\zeta_{1t} + \zeta_{2t}\big) + o_p(1);$

(c) $\quad \sqrt{n} I_{n3}^{(k)} = o_p(1), \quad k = 1, 2;$

(d) $\quad \sqrt{n}\big(I_{n4}^{(1)} + I_{n3}^{(2)}\big) = \dfrac{1}{\sqrt{n}} \sum_{t=1}^{n} \big(2\mu_1 \xi_{1t}\big) + o_p(1).$

Combining (a)-(d), we immediately obtain that

$$\sqrt{n}\,(\hat{\gamma}_n - \gamma_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \big(\zeta_t + \mu_1 \tilde{\xi}_{1t}\big) + o_p(1).$$

In fact, (a), (b) and (c) can be obtained in a similar fashion as those in the stationary case. Now we only need to prove (d). First, it follows from Lemma B.1 below and Taylor's expansion that

$$\begin{aligned}
\sqrt{n}\big(I_{n4}^{(1)} + I_{n4}^{(2)}\big) &= \frac{\mu_1}{\sqrt{n}} \sum_{t=1}^{n} \big\{v_t^{(1)}(\hat{\theta}_n) + v_t^{(2)}(\hat{\theta}_n)\big\} + \frac{2\mu_2}{\sqrt{n}} \sum_{t=1}^{n} u_t(\hat{\theta}_n) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \big(2\mu_1 \tilde{\xi}_{1t}\big) + \frac{2\mu_2}{\sqrt{n}} \sum_{t=1}^{n} u_t(\hat{\theta}_n) + o_p(1).
\end{aligned}$$

By the compactness of Θ and the fact (A.1), we have

$$\sum_{t=1}^{n} |u_t(\hat{\theta}_n)| \le M \sum_{t=1}^{n} \frac{1}{1+y_{t-1}^2} < \infty.$$

Thus, $\sum_{t=1}^{n} |u_t(\hat{\theta}_n) = O_p(1)$ and

$$\sqrt{n}\big(I_{n4}^{(1)} + I_{n4}^{(2)}\big) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \big(2\mu_1 \tilde{\xi}_{1t}\big) + o_p(1).$$

The proof is complete.

A.5. Proof of Theorem 3.2. Consider the case of $\gamma_0 < 0$. Using similar arguments in the proof of Theorem 3.1, we obtain that

$$\sqrt{n}\,(\hat{\gamma}_n^* - \gamma_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varpi_t \big(\zeta_t + \mu_1 \xi_{1t} + \mu_2 \nu^T \xi_{2t}\big) + o_{P^*}(1).$$

Then, with the asymptotic representation of $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$, it follows that

$$\sqrt{n}\,(\hat{\gamma}_n^* - \hat{\gamma}_n) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\varpi_t - 1)\big(\zeta_t + \mu_1 \xi_{1t} + \mu_2 \nu^T \xi_{2t}\big) + o_{P^*}(1).$$

Thus, conditionally on $\{y_0, y_1, \cdots, y_n\}$, in probability, as $n \to \infty$,

$$\sqrt{n}\big(\hat{\gamma}_n^* - \hat{\gamma}_n\big) \xrightarrow{\mathcal{L}} N\Big(0,\ E(\zeta_t + \mu_1 \xi_{1t} + \mu_2 \nu^T \xi_{2t})^2\Big).$$

The results for $\gamma_0 > 0$ can be proved analogously. The proof is complete.

A.6. Proof of Theorem 3.3. The proof is immediate from Theorem 3.2, and hence it is omitted.

APPENDIX B: LEMMAS WITH PROOFS

In this subsection, two lemmas are given, which are used in the proofs of Theorems 3.1 and 3.2.

LEMMA B.1. *Let $\{\eta_t, t = 1, \cdots, n\}$ are i.i.d. with the distribution $F(\cdot)$, which has an a.e. positive bounded density $f(\cdot)$ with $\int_{x\in R} |x| f(x)dx < \infty$. Let $\mathcal{A}_{ni}$ be an array of sub σ-fields such that $\mathcal{A}_{nt} \subset \mathcal{A}_{n(t+1)}, 1 \le t \le n, n \ge 1$; for each θ, $(v_{n1}(\theta), u_{n1}(\theta))$ is $\mathcal{A}_{n1}$-measurable and $(\eta_1, \cdots, \eta_{t-1}, v_{nt}(\theta), u_{nt}(\theta))$,*

$t \le j$ are $\mathcal{A}_{nj}$ measurable, $2 \le j \le n$; η_t is independent of $\mathcal{A}_{nt}, 1 \le t \le n$. For $v_{nt}(\theta)$ and $u_{nt}(\theta)$, assume that

$$\frac{1}{n}\sum_{t=1}^{n} \sup_{||\theta_1-\theta_2||\le\delta} \Big|v_{nt}(\theta_1) - v_{nt}(\theta_2)\Big| \le C_1\delta, \quad a.s.,$$

$$\text{and} \quad \frac{1}{n}\sum_{t=1}^{n} \sup_{||\theta_1-\theta_2||\le\delta} \Big|u_{nt}(\theta_1) - u_{nt}(\theta_2)\Big| \le C_2\delta, \quad a.s.$$

for some universal positive constants C_1 and C_2. For each $x \in R$, define

$$H_n(x;\theta) = \frac{1}{n}\sum_{t=1}^{n} I\Big(\eta_t \le \frac{x - n^{-1/2}v_{nt}(\theta)}{\sqrt{\alpha_0} + n^{-1/2}u_{nt}(\theta)}\Big) \quad \text{and}$$

$$H(x;\theta) = \frac{1}{n}\sum_{t=1}^{n} F\Big(\frac{x - n^{-1/2}v_{nt}(\theta)}{\sqrt{\alpha_0} + n^{-1/2}u_{nt}(\theta)}\Big),$$

where α_0 are a positive constant. Then, for any constant b with $0 < b < \infty$, there exists a positive constant ϵ such that

$$\sup_{x\in R,\, \|\theta\|\le b} \sqrt{n}\,|H_n(x;\theta) - H_n(x;0) - H(x;\theta) + H(x;0)| = O_p(n^{-\epsilon}).$$

PROOF. This lemma is similar to that of Lemma 8.3.2 in Koul (2002) but the rate of convergence is strengthened. It should be noted that we use the pseudometric

$$d_b(x,y) = \sup_{|z|\le b}\Big|F\Big(\frac{x - \phi_0 - n^{-1/2}z_1}{\sqrt{\alpha_0} + n^{-1/2}z_2}\Big) - F\Big(\frac{y - \phi_0 - n^{-1/2}z_1}{\sqrt{\alpha_0} + n^{-1/2}z_2}\Big)\Big|$$

for this local and scale setup, where $z = (z_1, z_2) \in R^2$, $|z| = |z_1| \vee |z_2|$, $x, y \in R$, $b > 0$. If we let $\mathfrak{N}(\delta, b)$ be the cardinality of the minimal δ-net of (R, d_b), we can show that $\mathfrak{N}(\delta, b) \le C_b\delta^{-4}$ with $C_b < \infty$ for any $0 < n^{-1/2}b < 1$. We can repeat the similar arguments as in Lemma 8.3.2 in Koul (2002) using the metric d_b. The convergence rate is strengthened since we make stronger conditions of the terms $u_{nt}(\theta)$ and $v_{nt}(\theta)$. The details are omitted here.

LEMMA B.2. *Suppose that η is a random variable with the density $f(\cdot)$ and the derivative $f'(x)$ of $f(x)$ exists for each x. The constants ϕ_0 and α_0 are finite and $\alpha_0 > 0$. Assume $b_1 = -\alpha_0^{-1/2}\int \log|\phi_0 + \eta\sqrt{\alpha_0}| f'(\eta)d\eta < \infty$ and $b_2 = -\alpha_0^{-1/2}\int \log|\phi_0 + \eta\sqrt{\alpha_0}|(f(\eta) + \eta f'(\eta))d\eta < \infty$. Denote $\delta =$*

$(\delta_1, \delta_2)^T$ *and* $h(\eta;\delta) = \log|\phi_0 + \eta\sqrt{\alpha_0} + \delta_1 + \eta\delta_2|$. *Then, for small constants* δ_1 *and* δ_2 *with* $\delta_1 \to 0$ *and* $\delta_2 \to 0$, *the following approximation holds:*

$$E\big(h(\eta;\delta) - h(\eta;0)\big) = b_1\delta_1 + b_2\delta_2 + o(|\delta_1| + |\delta_2|).$$

PROOF. $Eh(\eta;\delta)$ is expressed as

$$Eh(\eta;\delta) = \int f\big(\frac{\eta - \phi_0 - \delta_1}{\sqrt{\alpha_0} + \delta_2}\big)\frac{\log|\eta|}{\sqrt{\alpha_0} + \delta_2}d\eta.$$

Hence,

$$\begin{aligned} E\big(h(\eta;\delta) - h(\eta;0)\big) &= \int \log|\eta|\Big\{f\big(\frac{\eta - \phi_0 - \delta_1}{\sqrt{\alpha_0} + \delta_2}\big) - f\big(\frac{\eta - \phi_0}{\sqrt{\alpha_0}}\big)\Big\}d\eta\frac{1}{\sqrt{\alpha_0} + \delta_2} \\ &\quad + \Big(\frac{1}{\sqrt{\alpha_0} + \delta_2} - \frac{1}{\sqrt{\alpha_0}}\Big)\int \log|\eta| f\big(\frac{\eta - \phi_0}{\sqrt{\alpha_0}}\big)d\eta \\ &= I_1 + I_2. \end{aligned}$$

Let us handle the first term I_1. By Taylor 's expansion for $f(\cdot)$ at the point $(\eta - \phi_0)/\sqrt{\alpha_0}$, we obtain that

$$\begin{aligned} & f\big(\frac{\eta - \phi_0 - \delta_1}{\sqrt{\alpha_0} + \delta_2}\big) - f\big(\frac{\eta - \phi_0}{\sqrt{\alpha_0}}\big) \\ &= f'\big(\frac{\eta - \phi_0}{\sqrt{\alpha_0}}\big)\Big(\frac{\eta - \phi_0 - \delta_1}{\sqrt{\alpha_0} + \delta_2} - \frac{\eta - \phi_0}{\sqrt{\alpha_0}}\Big) + o(|\delta_1| + |\delta_2|) \\ &= f'\big(\frac{\eta - \phi_0}{\sqrt{\alpha_0}}\big)\Big(\frac{-\delta_1\sqrt{\alpha_0} - \delta_2\eta + \phi_0\delta_2}{\alpha_0}\Big) + o(|\delta_1| + |\delta_2|). \end{aligned}$$

Thus,

$$\begin{aligned} I_1 &= \int \log|\eta| f'\big(\frac{\eta - \phi_0}{\sqrt{\alpha_0}}\big)\Big(\frac{-\delta_1\sqrt{\alpha_0} - \delta_2\eta + \phi_0\delta_2}{\alpha_0}\Big)d\eta\frac{1}{\sqrt{\alpha_0}} + o(|\delta_1| + |\delta_2|) \\ &= -\alpha_0^{-1/2}\delta_1\int \log|\phi_0 + \eta\sqrt{\alpha_0}| f'(\eta)d\eta \\ &\quad - \alpha_0^{-1/2}\delta_2\int \log|\phi_0 + \eta\sqrt{\alpha_0}| f'(\eta)\eta d\eta + o(|\delta_1| + |\delta_2|). \end{aligned}$$

Similarly, we have that

$$I_2 = -\alpha_0^{-1/2}\delta_2\int \log|\phi_0 + \eta\sqrt{\alpha_0}| f\,(\eta)\,d\eta + o(\delta_2).$$

Combining the above results, the proof follows.

REFERENCES

Borkovec, M. and Klüppelberg, C. (2001). The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. *Ann. Appl. Prob.*, **11**, 1220–1241.

Chan, N. H. and Peng, L. (2005). Weighted least absolute deviation estimation for an AR(1) process with ARCH(1) errors. *Biometrika*, **92**, 477–484.

Chen, K., Ying, Z., Zhang, H., and Zhao, L. (2008). Analysis of least absolute deviation. *Biometrika*, **95(1)**, 107-122.

Chen, K., Guo, S., Lin, Y., and Ying, Z. (2010). Least absolute relative error estimation. *J. Amer. Statist, Assoc.*, **105**, 1104-1112.

Chen, M., Li, D. and Ling, S. (2014). Non-stationarity and quasi-maximum likelihood estimation on a double autoregressive model. *J. Time Series Anal.*, **35**, 189-202.

Cline, D. B. H and Pu, H. H. (2004). Stability and the Lyapounov exponent of threshold AR-ARCH models. *Ann. Appl. Prob.*, **14**, 1920–1949.

Douc, R., Moulines, E. and Stoffer, D. (2014). *Nonlinear time series: Theory, Methods and Applications with R Examples.* Chapman & Hall.

Efron, B. (1987). Better bootstrap confidence intervals. *J. Amer. Statist. Assoc.*, **82**, 171-185.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.

Francq, C. and Zakoïan, J.-M. (2012). Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models. *Econometrica*, **80**, 821–861.

Francq, C. and Zakoïan, J.-M. (2013). Inference in nonstationary asymmetric GARCH models. *Ann. Stat.* **41**, 1970-1998.

Jensen, S. T. and Rahbek, A. (2004a). Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case. *Econometrica*, **72**, 641–646.

Jensen, S. T. and Rahbek, A. (2004b). Asymptotic inference for nonstationary GARCH. *Econometric Theory*, **20**, 1203–1226.

Koul, H. L. (2002). *Weighted empirical processes in dynamic nonlinear models.* Springer-Verlag, New York.

Li, D., Ling, S. and Zakoïan, J.-M. (2015). Asymptotic inference in multiple-threshold double autoregressive models. *J. Econometrics*, **189**, 415–427.

Li, D., Ling, S. and Zhang, R.M. (2016). On a threshold double autoregressive model. *J. Bus. Econom. Statist.*, **34**,

Li, G., Zhu, Q., Liu, Z. and Li, W.K. (2015). On mixturedouble autoregressive time series models. *J. Bus. Econom. Statist.* (In press)

Ling, S. (2004). Estimation and testing stationarity for double autoregressive models. *J. Roy. Statist. Soc. Ser. B*, **66**, 63–78.

Ling, S. (2007). A double AR(p) model: structure and estimation. *Statist. Sinica.*, **17**, 161–175.

Ling, S. and Li, D. (2008). Asymptotic inference for a nonstationary double AR(1) model. *Biometrika*, **95**, 257–263.

Liu, F., Li, D. and Kang, X. M. (2016). Sample path properties of an explosive double autoregressive model. *Econometric Rev.* (In press)

Lu, Z. (1998). On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term. *Statist. Sinica.*, **8**, 1205–1217.

Tsay, R. S. (1987). Conditional heteroscedastic time series models, *J. Amer. Statist. Assoc.*, **82**, 590–604.

Weiss, A. A. (1984). ARMA models with ARCH errors. *J. Time Ser. Anal.*, **5**, 129–143.

Zhu, K. (2015). Bootstrapping the portmanteau tests in weak auto-regressive moving average models. *J. Roy. Statist. Soc. Ser. B.*(In press)

Zhu, K. and Ling, S. (2011). Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. *Ann. Statist.*, **39**, 2131–2163.

Zhu, K. and Ling, S. (2013). Quasi-maximum exponential likelihood estimators for a double AR (p) model. *Statist. Sinica.*, **23**, 251–270.

Academy of Mathematics and Systems Science,
Chinese Academy of Sciences,
Beijing 100190, China;
& Department of Statistics,
London School of Economics, London,
WC2A 2AE, United Kingdom;
E-mail: guoshaoj@amss.ac.cn

Center for Statistical Science,
Tsinghua University,
Beijing 100084, China
E-mail: malidong@tsinghua.edu.cn

Wang Ya'nan Institute for Studies in Economics (WISE)
and School of Economics,
Xiamen Unviersity,
Xiamen 361005, China
E-mail: limuyi@xmu.edu.cn