



Least absolute deviation estimation for AR(1) processes with roots close to unity

Nannan Ma¹ · Hailin Sang² · Guangyu Yang³

Received: 7 July 2022 / Revised: 27 November 2022 / Accepted: 16 December 2022 /

Published online: 23 January 2023

© The Institute of Statistical Mathematics, Tokyo 2023

Abstract

We establish the asymptotic theory of least absolute deviation estimators for AR(1) processes with autoregressive parameter satisfying $n(\rho_n - 1) \rightarrow \gamma$ for some fixed γ as $n \rightarrow \infty$, which is parallel to the results of ordinary least squares estimators developed by Andrews and Guggenberger (Journal of Time Series Analysis, 29, 203–212, 2008) in the case $\gamma = 0$ or Chan and Wei (Annals of Statistics, 15, 1050–1063, 1987) and Phillips (Biometrika, 74, 535–574, 1987) in the case $\gamma \neq 0$. Simulation experiments are conducted to confirm the theoretical results and to demonstrate the robustness of the least absolute deviation estimation.

Keywords Asymptotic distribution · Autoregressive processes · Least absolute deviation estimation · Local to unity · Unit root test

1 Introduction

Consider the following AR(1) process

$$y_i = \rho y_{i-1} + \epsilon_i, \quad 1 \leq i \leq n,$$

✉ Guangyu Yang
guangyu@zzu.edu.cn

Nannan Ma
1055165454@qq.com

Hailin Sang
sang@olemiss.edu

¹ Zhengzhou Zhongyuan Sub-branch, Agricultural Bank of China, 101 Funiu Road, Zhongyuan District, Zhengzhou 450007, Henan, China

² Department of Mathematics, University of Mississippi, University, MS 38677, USA

³ School of Mathematics and Statistics, Zhengzhou University, 100 Kexue Blvd, High-Tech District, Zhengzhou 450001, Henan, China

where ρ is a deterministic parameter and $\{\epsilon_i\}_{i \in \mathbb{Z}}$ is a sequence of independent and identically distributed (i.i.d.) random variables with mean zero and finite variance. The asymptotic properties of the ordinary least squares (OLS) estimator of ρ have been extensively studied in the literature; we refer to Anderson (1959), White (1958), Dickey and Fuller (1979), Phillips (1987), Chan (2009), Miao and Shen (2009), and references therein. To further handle the data that allows for large shocks in dynamic structure of the process, e.g. modeling the asset-price bubbles, it is usually to require the parameter ρ to depend on the sample size n . On the other hand, to understand the phenomena that the unit root test generally has a low discriminatory power against the alternative of root close to but not equal to unity, Bobkoski (1983) and Cavanagh (1985) introduced a local unit root model with the parameter ρ depending on the sample size n and tending to unity as $n \rightarrow \infty$. Since then, many researchers systematically established the limit theory for various near unit root processes. See Chan and Wei (1987), Phillips (1987, 1988), Phillips and Magdalinos (2007), Aue and Horváth (2007), Andrews and Guggenberger (2008), Buchmann and Chan (2013), Miao et al. (2015), Jiang et al. (2022), Tanaka (2017), and references therein. In particular, we refer to Stock (1991) for the empirical research or Phillips (2021) for the recent theoretical progress and empirical research on the processes with near unit roots.

Recently, Zhou and Lin (2014) and Wang et al. (2020) studied the statistical inference for the autoregressive parameter under the framework of the least absolute deviation (LAD) estimation. They proved that, if $\rho_n \rightarrow 1$ and $n|\rho_n - 1| \rightarrow \infty$ as $n \rightarrow \infty$, then the LAD estimators have normal and Cauchy asymptotic distributions under the mildly-stationary case and the mildly-explosive case, respectively, which are complementary to the results of the OLS estimators established by Giraitis and Phillips (2006) and Phillips and Magdalinos (2007). In fact, the LAD estimator was first considered in the study of autoregressive processes in the case that the regressive parameters are constants and the innovations have infinite variance due to the robust property of LAD estimation. For example, see the papers (Pollard 1991; Phillips 1991; Davis et al. 1992; and Li and Li 2009). Specially, Herce (1996) studied the asymptotic property of the LAD estimator in a unit root process with finite variance innovations and correspondingly developed the unit root test in this case which complement similar ones obtained by Knight (1989).

Motivated by the above work, the goal of the present article is to establish the asymptotic theory of LAD estimators for AR(1) processes when the regressive parameter satisfies $n(\rho_n - 1) \rightarrow 0$ or $n(\rho_n - 1) \rightarrow \gamma$ for some fixed $\gamma \neq 0$ as $n \rightarrow \infty$. To the best of our knowledge, this part of research is still missing although we already have a rich literature in LAD estimations and unit root test for AR(1) processes. It is shown that, if $n(\rho_n - 1) \rightarrow 0$ as $n \rightarrow \infty$, the limiting distributions are dominated by the initial conditions both in the near-stationary case and the near-explosive case. This phenomena have been studied by Müller and Elliott (2003), Phillips and Magdalinos (2009), and references therein. Our results in the near-stationary case correspond exactly to the theory on the OLS estimators for the same model developed by Andrews and Guggenberger (2008). With the condition $n(\rho_n - 1) \rightarrow \gamma$ for some fixed $\gamma \neq 0$ as $n \rightarrow \infty$, we study the asymptotic theory of the LAD estimator under the assumption $y_0 = 0$. The work in this case is

complementary to the theory of the OLS estimation developed by Chan and Wei (1987) and Phillips (1987).

This paper is organized as follows. After introducing the model, we state the main results in Sect. 2. Section 3 reports some simulation studies to illustrate the finite sample performance and to confirm the asymptotic results. Section 4 consists of concluding remarks and Sect. 5 provides the proofs of main results. All the other technical proofs are given in Appendix.

Throughout this paper, the symbols ' \rightarrow_d ', ' \rightarrow_p ' and ' \equiv_d ' denote the weak convergence, convergence in probability and equality in distribution, respectively; $o_p(1)$ means tending to zero in probability; \mathcal{C} denotes the standard Cauchy random variable and $\mathcal{N}(\mu, \sigma^2)$ denotes the normal random variable with mean μ and variance σ^2 ; $[x]$ represents the integral part of x ; $\text{sign}(\cdot)$ is the signum function; \mathbb{I}_A is the indicator function of set A and $\det(M)$ is the determinant of matrix M .

2 Main results

Suppose that the data $\{y_i\}$ are generated by the following autoregressive model with order one

$$y_i = \rho_n y_{i-1} + \epsilon_i, \quad 1 \leq i \leq n, \quad (1)$$

where the parameter ρ_n satisfies $\rho_n \rightarrow 1$ as $n \rightarrow \infty$, and the noises $\{\epsilon_i\}_{i \in \mathbb{Z}}$ satisfy the following assumptions:

- (i). $\{\epsilon_i\}_{i \in \mathbb{Z}}$ are i.i.d. random variables with mean zero and finite variance σ^2 ;
- (ii). ϵ_1 has zero median and a differentiable density function $f(x)$ in \mathbb{R} with $f(0) > 0$ and $\sup_{x \in \mathbb{R}} |f'(x)| < \infty$.

Note that, formally the data $\{y_i\}$ is a triangular array but here n is omitted for notational simplicity. With the above assumptions on the noises, one can also study the estimation of the parameter ρ_n with $\rho_n \rightarrow -1$ as $n \rightarrow \infty$. For the purpose of simplicity, we only state the results for the positive ρ_n case in this paper.

The LAD estimator of ρ_n is defined as a solution of the following extremum problem

$$\hat{\rho}_{\text{LAD}} = \arg \min_{\rho \in \mathbb{R}} \left\{ \frac{1}{n} \sum_{i=1}^n |y_i - \rho y_{i-1}| \right\}. \quad (2)$$

Notice that, the LAD estimators usually do not have closed forms and are not unique if the object function has a flat segment; see, e.g. Hecce (1996). Moreover, the LAD estimators are robust, especially they are not significantly affected by the presence of outliers. This is confirmed by the simulation study in Sect. 3.

2.1 The case $n(\rho_n - 1) \rightarrow 0$ as $n \rightarrow \infty$

We first consider the case that the parameter ρ_n satisfies $n(\rho_n - 1) \rightarrow 0$ as $n \rightarrow \infty$. That is, the autoregressive parameter ρ_n is very nearly unity in the sense that ρ_n is away from unity by $o(n^{-1})$. The following two theorems are the main results in this case. One is for the near-stationary case $0 < \rho_n < 1$ and the other is for the near-explosive case $\rho_n > 1$.

Theorem 1 (The near-stationary case) *For model (1) with $0 < \rho_n < 1$ and $n(\rho_n - 1) \rightarrow 0$ as $n \rightarrow \infty$, assume that y_0 depends on the full past of the noise, i.e. $y_0 = \sum_{j=0}^{\infty} \rho_n^j \epsilon_{-j}$, then, as $n \rightarrow \infty$, we have*

$$\sqrt{\frac{n}{1 - \rho_n^2}} (\hat{\rho}_{\text{LAD}} - \rho_n) \rightarrow_d \frac{\mathcal{C}}{2\sigma f(0)} \quad (3)$$

and

$$\sqrt{\sum_{i=1}^n y_{i-1}^2} (\hat{\rho}_{\text{LAD}} - \rho_n) \rightarrow_d \mathcal{N}\left(0, \frac{1}{4f^2(0)}\right). \quad (4)$$

Theorem 2 (The near-explosive case) *For model (1) with $\rho_n > 1$ and $n(\rho_n - 1) \rightarrow 0$ as $n \rightarrow \infty$, assume that y_0 is an infinitely distant initialization, i.e. $y_0 = \sum_{j=0}^{\kappa_n} \rho_n^j \epsilon_{-j}$, with $\kappa_n/n \rightarrow \infty$ and $\kappa_n(\rho_n - 1) \rightarrow 0$ as $n \rightarrow \infty$, then, as $n \rightarrow \infty$, we have*

$$\sqrt{n\kappa_n} (\hat{\rho}_{\text{LAD}} - \rho_n) \rightarrow_d \frac{\mathcal{C}}{2\sigma f(0)} \quad (5)$$

and

$$\sqrt{\sum_{i=1}^n y_{i-1}^2} (\hat{\rho}_{\text{LAD}} - \rho_n) \rightarrow_d \mathcal{N}\left(0, \frac{1}{4f^2(0)}\right). \quad (6)$$

We give some comments on Theorems 1 and 2.

Remark 1 Since the AR(1) model (1) is causal when $0 < \rho_n < 1$, the initial value y_0 can be written as a linear combination over the past information in the linear process form $y_0 = \sum_{j=0}^{\infty} \rho_n^j \epsilon_{-j}$. In the same case, $\rho_n \in (-1, 1)$, $n(\rho_n - 1) \rightarrow 0$ and y_0 depends on the full past of the noise, Andrews and Guggenberger (2008) obtained the limit theorems for the OLS estimators which are similar to Theorem 1 except for the appearance of $f(0)$ here. Furthermore, because the limiting distributions of the t -type estimator include $f(0)$ in LAD estimations, in practice we use the following density estimator (Silverman 1986) for statistical inference

$$\hat{f}_n(0) := \frac{1}{nb_n} \sum_{i=1}^n K\left(\frac{y_i - \hat{\rho}_{\text{LAD}} y_{i-1}}{b_n}\right),$$

where b_n is the bandwidth and $K(\cdot)$ is a kernel function, e.g. the Gaussian or logistic kernel.

Remark 2 As we mentioned previously, Zhou and Lin (2014) and Wang et al. (2020) studied the LAD estimations for ρ_n satisfying $n|\rho_n - 1| \rightarrow \infty$ where the initial value $y_0 = o_p(|\rho_n - 1|^{-1/2})$ independent of $\{\epsilon_i, i \geq 1\}$. To be explicit, they obtained that

$$\begin{cases} \sqrt{\frac{n}{1-\rho_n^2}}(\hat{\rho}_{\text{LAD}} - \rho_n) \rightarrow_d \mathcal{N}\left(0, \frac{1}{4f^2(0)}\right) \\ \sqrt{\sum_{i=1}^n y_{i-1}^2}(\hat{\rho}_{\text{LAD}} - \rho_n) \rightarrow_d \mathcal{N}\left(0, \frac{1}{4f^2(0)}\right) \end{cases} \quad (7)$$

and

$$\frac{\rho_n^n}{\rho_n^2 - 1}(\hat{\rho}_{\text{LAD}} - \rho_n) \rightarrow_d \frac{\mathcal{C}}{2\sigma f(0)}, \quad (8)$$

for the mildly-stationary case and the mildly-explosive case, respectively. Their results match the classic results in Phillips and Magdalinos (2007) except for the appearance of $f(0)$, just like the usual LAD estimations.

It is worth pointing out that, here for the very nearly unit root processes, Theorems 1 and 2 show that, in both near-stationary and near-explosive cases, the initial value y_0 dominates the asymptotic distributions of the LAD estimator and the t -type estimator which are Cauchy and normal, respectively. Moreover, since $\rho_n \rightarrow 1$ at a much faster rate than $1/n$, for the near-stationary case, by Lemma 1, the assumption $y_0 = o_p(|\rho_n - 1|^{-1/2})$ can not hold and the convergence rate $\sqrt{n/(1-\rho_n^2)}$ has a larger order than n which enlarges the convergence rate spectrum, while, for the near-explosive case, the convergence rate $\sqrt{n\kappa_n}$ also has a large order than n which is different from that in Eq. (8).

2.2 The case $n(\rho_n - 1) \rightarrow \gamma$ as $n \rightarrow \infty$

Now we consider the local unit root case that the parameter ρ_n satisfies $n(\rho_n - 1) \rightarrow \gamma$ for some fixed $\gamma \neq 0$ as $n \rightarrow \infty$. We remark that, unlike the case, $\rho_n \neq 1$ and $n(\rho_n - 1) \rightarrow 0$ as $n \rightarrow \infty$, where the initial condition entirely dominates the asymptotic distribution of the sample variance $\sum_{i=1}^n y_{i-1}^2$, in this case, the initial value depending on the past of the noise does not affect the analysis methods essentially so we can assume that $y_0 = 0$ for simplicity.

To state the main results, we first introduce some notations. For $1 \leq i \leq n$, let

$$K_{n,i} := \frac{1}{\sqrt{n}}\rho_n^{i-1-n}\text{sign}(\epsilon_i) \quad \text{and} \quad L_{n,i} := \frac{1}{\sqrt{n}}\rho_n^{n-i}\epsilon_i,$$

and define

$$K_n(t) := \sum_{i=1}^{[nt]} K_{n,i} \quad \text{and} \quad L_n(t) := \sum_{i=1}^{[nt]} L_{n,i}, \quad (9)$$

for $0 \leq t \leq 1$. If we further assume that $\sigma^2 = 1$ and $\mathbb{E}|\epsilon_1|^{2+\delta} < \infty$ for some $\delta > 0$, by Lemma 3, the process $\{(K_n(t), L_n(t)), 0 \leq t \leq 1\}$ converges weakly to a continuous process

$$\mathbf{X}(t) = (K(t), L(t)), \quad 0 \leq t \leq 1 \quad (10)$$

with independent Gaussian increments, mean vector zero and covariance matrix

$$\Gamma(t) = (\gamma_{ij}(t)) := \begin{pmatrix} \frac{e^{-2\gamma}}{2\gamma}(e^{2\gamma t} - 1) & t\mathbb{E}|\epsilon_1| \\ t\mathbb{E}|\epsilon_1| & \frac{e^{2\gamma}}{2\gamma}(1 - e^{-2\gamma t}) \end{pmatrix}, \quad 0 \leq t \leq 1. \quad (11)$$

Theorem 3 For model (1), assume that $y_0 = 0$, $\sigma^2 = 1$, $n(\rho_n - 1) \rightarrow \gamma$ for some fixed $\gamma \neq 0$ as $n \rightarrow \infty$, and $\mathbb{E}|\epsilon_1|^{2+\delta} < \infty$ for some $\delta > 0$, then, as $n \rightarrow \infty$, we have

$$n(\hat{\rho}_{\text{LAD}} - \rho_n) \rightarrow_d \frac{1}{2f(0)} \cdot \frac{\int_0^1 L(t) dK(t)}{\int_0^1 e^{-2\gamma(1-t)} L^2(t) dt} \quad (12)$$

and

$$\sqrt{\sum_{i=1}^n y_{i-1}^2} (\hat{\rho}_{\text{LAD}} - \rho_n) \rightarrow_d \frac{1}{2f(0)} \cdot \frac{\int_0^1 L(t) dK(t)}{\sqrt{\int_0^1 e^{-2\gamma(1-t)} L^2(t) dt}}. \quad (13)$$

Remark 3 Zhou and Lin (2014) and Wang et al. (2020) worked on the case that $\rho_n \rightarrow 1$ and $n|\rho_n - 1| \rightarrow \infty$ as $n \rightarrow \infty$, where the convergence rate for the asymptotic distributions of the LAD estimators has an order in the range $(n^{1/2}, n)$ for the mildly-stationary case. Together with the results in Theorems 1 and 2, we enlarge the convergence rate spectrum in the asymptotic distributions of the LAD estimators. The similar relationship between the rate of ρ_n and the rate in the asymptotic distributions was observed for OLS estimators in the literature, see, e.g. Chan and Wei (1987), Phillips (1987), Phillips and Magdalinos (2007), and Andrews and Guggenberger (2008).

Remark 4 Let

$$\mathcal{D}(\gamma) := \frac{1}{2f(0)} \cdot \frac{\int_0^1 L(t) dK(t)}{\int_0^1 e^{-2\gamma(1-t)} L^2(t) dt} \quad (14)$$

and

$$\mathcal{L}(\gamma) := \frac{1}{2f(0)} \cdot \frac{\int_0^1 L(t) dK(t)}{\sqrt{\int_0^1 e^{-2\gamma(1-t)} L^2(t) dt}}. \quad (15)$$

Notice that $\mathcal{D}(\gamma)$ and $\mathcal{L}(\gamma)$ are continuous families of distributions indexed by the parameter γ . Thus, for the case, $\rho_n = 1 + \gamma/n$, by Lemma 3, when $\gamma = 0$, i.e. $\rho_n \equiv 1$, we have

$$\mathcal{D}(0) = \frac{1}{2f(0)} \cdot \frac{\int_0^1 W_2(t) dW_1(t)}{\int_0^1 W_2^2(t) dt} \quad \text{and} \quad \mathcal{L}(0) = \frac{1}{2f(0)} \cdot \frac{\int_0^1 W_2(t) dW_1(t)}{\sqrt{\int_0^1 W_2^2(t) dt}},$$

here $\mathbf{W} = (W_1, W_2)$ is a bivariate Brownian motion with covariance matrix

$$\Xi = \begin{pmatrix} 1 & \mathbb{E}|e_1| \\ \mathbb{E}|e_1| & 1 \end{pmatrix}.$$

These exactly correspond to the main results in Herce (1996), where the author discussed the structure of $\mathcal{D}(0)$, a combination of a “unit root” distribution and a scale mixture of normal distributions, and used it to construct the LAD-based unit root tests. In addition, since $\sigma^2 = 1$, we have $\mathbb{E}|e_1| \leq 1$, hence the matrix Ξ is non-negative definite. However, here $\mathcal{D}(\gamma)$ and $\mathcal{L}(\gamma)$ are too complicate to be analyzed effectively so we do some simulations to give an overall view on the distributions of $\mathcal{D}(\gamma)$ and $\mathcal{L}(\gamma)$ in Sect. 3 which illustrate how they depend on the parameter γ . Finally, we also remark that, from Theorem 1 in Herce (1996), the local unit root case has the optimal rate of convergence for the alternative hypothesis, that is, it has the same rate as in the unit root case.

Followed the above remark, there is a natural question, what can we say about the distributions of $\mathcal{L}(\gamma)$ as $|\gamma| \rightarrow \infty$? In fact, like the case for the OLS estimators in Chan and Wei (1987) and Phillips (1987), we have the following result.

Theorem 4 *For model (1), assume that $y_0 = 0$, $\sigma^2 = 1$, $n(\rho_n - 1) \rightarrow \gamma$ for some fixed $\gamma \neq 0$ as $n \rightarrow \infty$, and $\mathbb{E}|e_1|^{2+\delta} < \infty$ for some $\delta > 0$, then we have*

$$2f(0) \cdot \mathcal{L}(\gamma) \rightarrow_d \mathcal{N}(0, 1), \quad \text{as } |\gamma| \rightarrow \infty. \quad (16)$$

Remark 5 Recall from Remark 2, Zhou and Lin (2014) and Wang et al. (2020) proved that

$$2f(0) \cdot \sqrt{\sum_{i=1}^n y_{i-1}^2} (\hat{\rho}_{\text{LAD}} - \rho_n) \rightarrow_d \mathcal{N}(0, 1), \quad (17)$$

for the mildly-stationary case, $0 < \rho_n < 1$ and $n(1 - \rho_n) \rightarrow \infty$ as $n \rightarrow \infty$. Theorem 4 together with Theorem 3 yields that, if let $n(\rho_n - 1) \rightarrow \gamma$ first, and then let $|\gamma| \rightarrow \infty$, we can also get the asymptotic normal distribution. Moreover, it is worth noting

that, although Theorem 4 holds for $\gamma \rightarrow \infty$ and $\gamma \rightarrow -\infty$, the underlying reasoning is quite different between these two cases. In fact, for the large negative γ 's, ρ_n can be thought of much less than one so the model (1) can be regarded as stationary and then the asymptotic normality holds; for the positive γ , ρ_n is larger than one and the model (1) is explosive, however Theorem 4 shows that the asymptotic normality is still valid which is different from the mildly-explosive case discussed in Zhou and Lin (2014) and Wang et al. (2020), i.e. Eq. (8) in Remark 2.

3 Simulations

In this section, we work on Monte Carlo simulation to examine the finite sample performance of the estimators in our main results. For convenience, in all experiments, we always suppose that the data are generated by the AR(1) model, $y_i = \rho_n y_{i-1} + \epsilon_i$ with $\rho_n = 1 + \gamma n^{-\beta}$ for some $\gamma \in \mathbb{R}$ and $\beta \geq 1$. Furthermore, we also assume that the innovations are i.i.d. and have $\mathcal{N}(0, 1)$ or $U(-1, 1)$ distributions. Here $U(-1, 1)$ denotes the uniform random variable on the interval $(-1, 1)$. To estimate the parameter $f(0)$, we apply the following density estimator (Silverman 1986)

$$\hat{f}_n(0) = \frac{1}{nb_n} \sum_{i=1}^n K\left(\frac{y_i - \hat{\rho}_{\text{LAD}} y_{i-1}}{b_n}\right),$$

where $K(\cdot)$ is the Gaussian kernel and the optimal bandwidth b_n associated with the MATLAB function `ksdensity`($\cdot, 0$) is automatically selected from the data.

3.1 Density curves of estimates

We first consider the simulations of Theorems 1 and 2. For the near-stationary case, $y_0 = \sum_{j=0}^{\infty} \rho_n^j \epsilon_{-j}$, the true parameters are taken to be $(\gamma, \beta) = (-10, 2)$ and we denote the normalized estimator by $\hat{\rho}_s = 2\sigma \hat{f}_n(0) \sqrt{n/(1 - \rho_n^2)} (\hat{\rho}_{\text{LAD}} - \rho_n)$. For the near-explosive case, take $y_0 = \sum_{j=0}^{\kappa_n} \rho_n^j \epsilon_{-j}$, $(\gamma, \beta) = (10, 2)$, $\kappa_n = \lceil n^{1.3} \rceil$, and denote the normalized estimator by $\hat{\rho}_e = 2\sigma \hat{f}_n(0) \sqrt{n\kappa_n} (\hat{\rho}_{\text{LAD}} - \rho_n)$. For both cases, denote the t -type estimators by $T_n := 2\hat{f}_n(0) \sqrt{\sum_{i=1}^n y_{i-1}^2} (\hat{\rho}_{\text{LAD}} - \rho_n)$. We simulate 3000 replications with sample size $n = 1000$ for each case. Figure 1 shows that the density curves of $\hat{\rho}_s$ and $\hat{\rho}_e$ are close to that of the standard Cauchy random variable. Figure 2 confirms the asymptotic normality of T_n in the near-stationary and near-explosive cases by using the Q-Q graphs.

Next we simulate the limiting distributions, $\mathcal{D}(\gamma)$ and $\mathcal{L}(\gamma)$, defined as in Remark 4. Let $\beta = 1$ and suppose that the innovations are standard normal random variables so $f(0) = 1/\sqrt{2\pi}$. For $\mathcal{D}(\gamma)$, simulate 1000 replications with sample size $n = 200$ for each case; Figure 3 illustrates the shape of the asymptotic density curves of $2f(0)\mathcal{D}(\gamma)$ for different γ . For $\mathcal{L}(\gamma)$, simulate 1000 replications with sample size

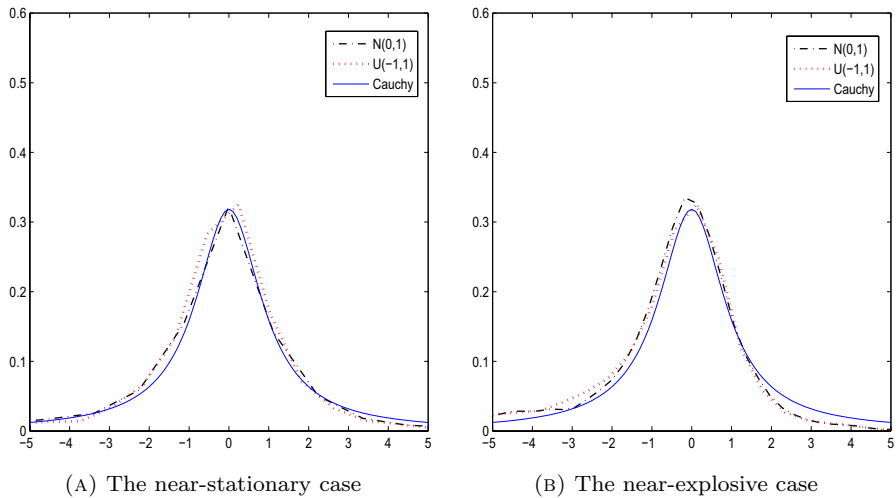


Fig. 1 Density curves of $\hat{\rho}_s$ and $\hat{\rho}_e$ with $\mathcal{N}(0, 1)$ and $U(-1, 1)$ innovations

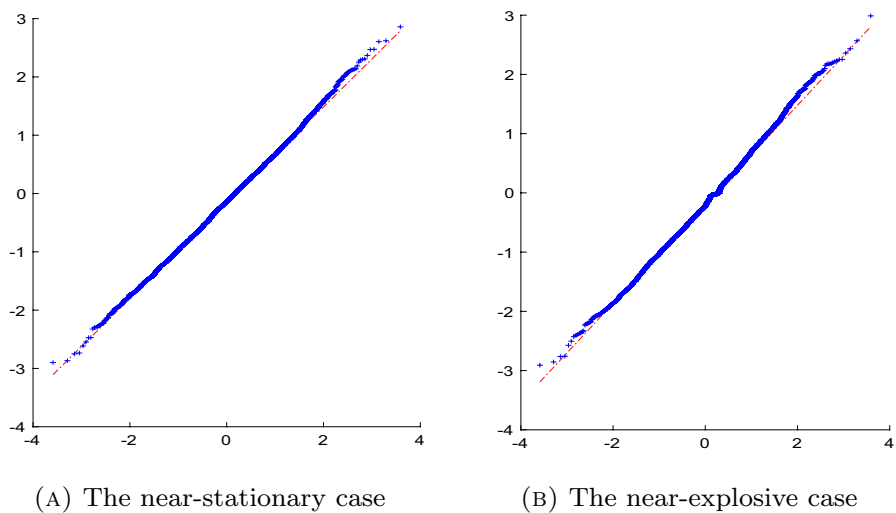


Fig. 2 Q-Q graphs of statistics T_n with $\mathcal{N}(0, 1)$ innovation

$n = 2000$ ($\gamma < 0$) and $n = 1000$ ($\gamma > 0$) for each case; Figure 4 confirms Theorem 4, i.e. the asymptotic normality of $2f(0)\mathcal{L}(\gamma)$ as $|\gamma| \rightarrow \infty$, which also demonstrates the difference of the convergence rate between $\gamma \rightarrow \infty$ and $\gamma \rightarrow -\infty$.

3.2 Accuracy of estimators

Andrews and Guggenberger (2008) achieved the asymptotic theory of OLS estimators for model (1) with $n(\rho_n - 1) \rightarrow 0$ and $0 < \rho_n < 1$. Here we will compare

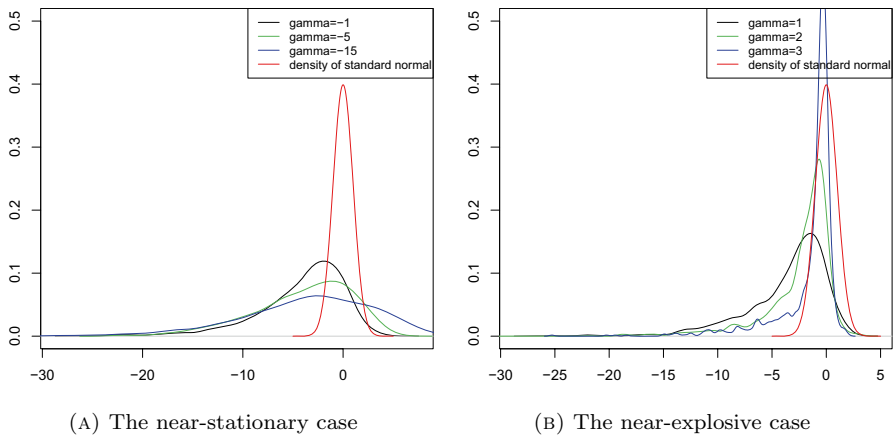


Fig. 3 Asymptotic density curves of $2f(0)\mathcal{Z}(\gamma)$ with $\mathcal{N}(0, 1)$ innovation

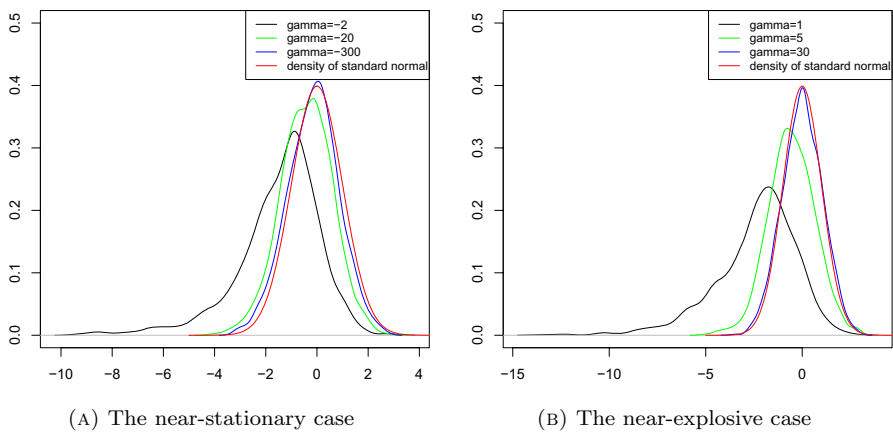


Fig. 4 Asymptotic density curves of $2f(0)\mathcal{Z}(\gamma)$ with $\mathcal{N}(0, 1)$ innovation

the accuracy between the estimators $\hat{\rho}_{\text{LAD}}$ and $\hat{\rho}_{\text{OLS}}$, where $\hat{\rho}_{\text{OLS}}$ denotes the OLS estimator of ρ_n . For the near-stationary case, let $y_0 = \sum_{j=0}^{\infty} \rho_n^j \epsilon_{-j}$, $\beta = 1.1, 1.3$ and $\gamma = -5, -50$. Here we add 10 for each randomly selected 5% sample points to construct outliers when generating data. In these experiments, we simulate 1000 replications each with sample size $n = 200$ or 500 to compare the empirical means (EM), absolute errors (AE) and mean squared errors (MSE) of $\hat{\rho}_{\text{LAD}}$ and $\hat{\rho}_{\text{OLS}}$ in two cases, with and without outliers. In the case without outliers, Table 1 shows that EM is very close to the true value ρ_n and AE and MSE are very small, which verifies the accuracy of LAD estimator for very nearly unit root model. The errors (AE and MSE) decrease as the sample size n increases. If there are outliers in the data, Table 2 shows that the errors (AE and MSE) of OLS estimator are much larger than those of LAD estimator. Furthermore, the OLS estimator occasionally produces estimates greater than one. Hence it is concluded that $\hat{\rho}_{\text{LAD}}$ is more robust than $\hat{\rho}_{\text{OLS}}$.

Table 1 AE/EM and MSE of $\hat{\rho}_{\text{LAD}}(\hat{\rho}_{\text{OLS}})$ for near-stationary AR(1) processes with $\mathcal{N}(0, 1)$, $U(-1, 1)$ noises: no outliers case

β	γ	n	ρ_n	ϵ_i			
				$N(0, 1)$		$U(-1, 1)$	
				AE/EM	MSE	AE/EM	MSE
1.1	-50	200	0.8528	0.0097/0.8431	0.0127	0.0112/0.8416	0.0207
				(0.0083/0.8445)	(0.0083)	(0.0084/0.8445)	(0.0077)
		500	0.9463	0.0037/0.9426	0.0017	0.0049/0.9414	0.0030
				(0.0034/0.9428)	(0.0012)	(0.0036/0.9427)	(0.0012)
	-5	200	0.9853	0.0046/0.9807	0.0010	0.0037/0.9816	0.0011
				(0.0083/0.9770)	(0.0019)	(0.0073/0.9780)	(0.0016)
1.3	-50	200	0.9490	0.0072/0.9418	0.0038	0.0101/0.9389	0.0059
				(0.0083/0.9407)	(0.0035)	(0.0096/0.9394)	(0.0035)
		500	0.9845	0.0028/0.9817	0.0005	0.0038/0.9807	0.0008
				(0.0037/0.9808)	(0.0005)	(0.0034/0.9811)	(0.0005)
	-5	200	0.9949	0.0029/0.9920	0.0003	0.0025/0.9924	0.0004
				(0.0065/0.9884)	(0.0011)	(0.0047/0.9890)	(0.0006)
	-5	500	0.9985	0.0018/0.9966	0.0001	0.0018/0.9967	0.0001
				(0.0025/0.9960)	(0.0001)	(0.0019/0.9966)	(0.0001)

For the near-explosive case, we take $\kappa_n = [n^{1.3}]$ in $y_0 = \sum_{j=0}^{\kappa_n} \rho_n^j e_{-j}$, and let $\beta = 1.5, 1.8, \gamma = 5, 50$. The selection of the parameter values here guarantees that $n(\rho_n - 1) \rightarrow 0$, $\kappa_n(1 - \rho_n) \rightarrow 0$ and $\kappa_n/n \rightarrow \infty$, as $n \rightarrow \infty$. Since the y values from different cases have quite different scales, here we add 10 times of the absolute value of the maximum for each randomly selected 5% sample points to construct outliers when generating data. In these experiments, we simulate 1000 replications each with sample size $n = 200$ or 500 and study the empirical means (EM), absolute errors (AE) and mean squared errors (MSE) of $\hat{\rho}_{\text{LAD}}$ and $\hat{\rho}_{\text{OLS}}$ in two cases, with and without outliers. Table 3 shows that, if there are no outliers in the data, EM is very close to the true value ρ_n and AE and MSE are very small, which verifies the accuracy of the LAD estimator and OLS estimator, while, if there are outliers in the data, it is clear from Table 4 that $\hat{\rho}_{\text{LAD}}$ is more robust than $\hat{\rho}_{\text{OLS}}$.

4 Concluding remarks

In this article we develop the limit theory of the LAD estimator for the AR(1) process with a root close to unity. The parameter ρ_n satisfies $n(\rho_n - 1) \rightarrow 0$ or $n(\rho_n - 1) \rightarrow \gamma$ for some fixed number $\gamma \neq 0$, as $n \rightarrow \infty$. It is shown that, in the first case, for both near-stationary and near-explosive processes, the LAD

Table 2 AE/EM and MSE of $\hat{\rho}_{\text{LAD}}(\hat{\rho}_{\text{OLS}})$ for near-stationary AR(1) processes with $\mathcal{N}(0, 1)$, $U(-1, 1)$ noises: outliers case

β	γ	n	ρ_n	ϵ_i			
				$N(0, 1)$		$U(-1, 1)$	
				AE/EM	MSE	AE/EM	MSE
1.1	-50	200	0.8528	0.0046/0.8574	0.0015	0.0046/0.8574	0.00009
				(0.0451/0.8979)	(0.0121)	(0.0500/0.9028)	(0.0138)
		500	0.9463	0.0039/0.9501	0.0002	0.0033/0.9495	0.0001
				(0.0320/0.9783)	(0.0052)	(0.0338/0.9801)	(0.0058)
	-5	200	0.9853	0.0015/0.9868	0.0001	0.0013/0.9866	0.0000
				(0.0145/0.9998)	(0.0011)	(0.0150/1.0003)	(0.0012)
		500	0.9946	0.0007/0.9954	0.0000	0.0006/0.9953	0.0000
				(0.0065/1.0011)	(0.0002)	(0.0066/1.0012)	(0.0002)
1.3	-50	200	0.9490	0.0024/0.9514	0.0004	0.0029/0.9519	0.0002
				(0.0297/0.9787)	(0.0046)	(0.0320/0.9810)	(0.0052)
		500	0.9845	0.0016/0.9861	0.0000	0.0013/0.9858	0.0000
				(0.0137/0.9982)	(0.0009)	(0.0140/0.9985)	(0.0010)
	-5	200	0.9949	0.0009/0.9958	0.0000	0.0011/0.9960	0.0001
				(0.0091/1.0040)	(0.0005)	(0.0095/1.0044)	(0.0006)
		500	0.9985	0.0006/0.9990	0.0001	0.0004/0.9989	0.0001
				(0.0040/1.0024)	(0.0001)	(0.0040/1.0025)	(0.0001)

estimator and the t -type statistic have Cauchy and normal asymptotic distribution, respectively. The simulation study in this case confirms the theoretical results and it illustrates that the theory should be useful in statistical inference and the LAD estimator is robust if there are outliers in the data. In the case that $n(\rho_n - 1) \rightarrow \gamma$ for some fixed number $\gamma \neq 0$, as $n \rightarrow \infty$, we also develop asymptotic theory for the LAD estimator and the t -type statistic under the assumption that $y_0 = 0$ which gives us the connection with the existing results in literature, e.g. Chan and Wei (1987); Phillips (1987), and Herce (1996).

In summary, we establish the asymptotic theory of the LAD estimators for the nearly unit root processes which correspond to the results of the OLS estimators developed by Chan and Wei (1987), Phillips (1987, 1988, 2021), Phillips and Magdalinos (2007), Andrews and Guggenberger (2008), Chan (2009) and so on. The results in this article are also the completion of the LAD estimators studied in Zhou and Lin (2014) and Wang et al. (2020). These authors worked on the case that $\rho_n \rightarrow 1$ and $n|\rho_n - 1| \rightarrow \infty$ as $n \rightarrow \infty$. The normalizer of the asymptotic distributions for the LAD estimators there has a rate in the range $(n^{1/2}, n)$ for the near-stationary case. We enlarge the rate spectrum of the normalizer in the asymptotic distributions of the LAD estimators.

Table 3 AE/EM and MSE of $\hat{\rho}_{\text{LAD}}(\hat{\rho}_{\text{OLS}})$ for near-explosive AR(1) processes with $\mathcal{N}(0, 1)$, $U(-1, 1)$ noises: no outliers case

β	γ	n	ρ_n	ϵ_i			
				$N(0, 1)$		$U(-1, 1)$	
				AE/EM	MSE	AE/EM	MSE
1.5	50	200	1.0177	2.2e-5/1.0177	2.5e-9	2.2e-5/1.0177	2.5e-9
				(2.2e-5/1.0177)	(2.5e-9)	(2.2e-5/1.0177)	(2.5e-9)
		500	1.0045	2.8e-5/1.0045	3.9e-9	2.8e-5/1.0045	3.9e-9
				(2.8e-5/1.0045)	(3.9e-9)	(2.8e-5/1.0045)	(3.9e-9)
	5	200	1.0018	-6e-4/1.0012	4.5e-5	-8e-4/1.0009	6.7e-5
				(-e-6/1.0018)	(e-5)	(1.8e-4/1.0019)	(e-5)
		500	1.0004	-3.8e-4/1.00007	1.3e-5	-5.7e-4/0.9999	2.8e-5
				(1.2e-4/1.0006)	(1.5e-6)	(0.0002/1.0007)	(e-6)
1.8	50	200	1.0036	-1.6e-4/1.0034	1.4e-5	-2e-4/1.0034	1.5e-5
				(-5e-5/1.0036)	(3e-6)	(-e-4/1.0035)	(3.7e-6)
		500	1.0007	-2.6e-4/1.0004	9e-6	3.7e-4/1.0003	1.8e-5
				(4e-5/1.0007)	(e-6)	(9e-5/1.0008)	(9e-7)
	5	200	1.0004	-0.0017/0.9986	e-4	-0.0020/0.9984	1.7e-4
				(4e-4/1.0007)	(1.9e-5)	(6e-4/1.0010)	(e-5)
		500	1.00007	-9.6e-4/0.9991	3.9e-5	-0.0011/0.9989	5.7e-5
				(3e-4/1.00036)	(1.6e-6)	(2.7e-4/1.00035)	(9e-7)

5 Proof of main results

We first give some comments on the proofs. It is known that the methods of the LAD estimation are classic which were developed by Pollard (1991), Davis et al. (1992), Knight (1989, 1998) and Ling (2005). The limiting behaviors of the quadratic functionals, $\sum_{i=1}^n y_{i-1}^2$ and $\sum_{i=1}^n y_{i-1} \text{sign}(\epsilon_i)$, play the crucial role in the analysis. Concretely, for the very nearly case, $n(\rho_n - 1) \rightarrow 0$, we mainly follow the strategies of Zhou and Lin (2014) and Wang et al. (2020), while, under our framework, the initial values y_0 dominate the asymptotic behavior of $\sum_{i=1}^n y_{i-1}^2$; see Lemmas 1 and 2. More work is required for the local unit root case, $n(\rho_n - 1) \rightarrow \gamma$ for some $\gamma \neq 0$. To treat the asymptotic joint distributions of $\left(\frac{1}{n} \sum_{i=1}^n y_{i-1} \text{sign}(\epsilon_i), \frac{1}{n^2} \sum_{i=1}^n y_{i-1}^2\right)$, we need to develop a functional central limit theorem for the process $\{(K_n(t), L_n(t)), 0 \leq t \leq 1\}$ defined as in (9), i.e. Lemma 3, by the vector-value martingale invariance principle. We remark that, although Phillips and Durlauf (1986) (Lemma 3.1) established the asymptotic theory for sample moments of vector-value integrated processes, their results can not be used here because the regression parameter ρ_n depends on the sample size n which causes the covariance matrix defined as in (11) to be time-dependent. Finally, by using the tools from stochastic calculus, we achieve the estimates, Lemmas 4 and 5, which is the key of the proof in Theorem 4.

Table 4 AE/EM and MSE of $\hat{\rho}_{\text{LAD}}(\hat{\rho}_{\text{OLS}})$ for near-explosive AR(1) processes with $\mathcal{N}(0, 1)$, $U(-1, 1)$ noises: outliers case

β	γ	n	ρ_n	ϵ_i			
				$N(0, 1)$		$U(-1, 1)$	
				AE/EM	MSE	AE/EM	MSE
1.5	50	200	1.0177	− 0.0144/1.0033	0.0020	− 0.0140/1.0037	0.0019
				(− 0.4835/0.5341)	(2.1195)	(− 0.4703/0.5474)	(2.0560)
		500	1.0045	− 7.1e-3/0.9974	5.1e-4	− 0.0069/0.9976	5e-4
	5	200	1.0018	(− 0.5434/0.4610)	(1.9600)	(− 0.5345/0.4700)	(1.9105)
				− 0.0080/0.9939	3.8e-4	− 0.0090/0.9928	4.6e-4
		500	1.0004	(− 0.7847/0.2171)	(3.1892)	(− 0.7748/0.2270)	(3.1400)
1.8	50	200	1.0036	− 0.0064/0.9941	2.5e-4	− 0.0074/0.9930	3.2e-4
				(− 0.7823/0.2182)	(3.1841)	(− 0.7731/0.2273)	(3.1062)
		500	1.0007	(− 0.7888/0.2148)	(3.1625)	(− 0.7929/0.2108)	(3.1825)
	5	200	1.0004	− 0.0033/1.0003	1.2e-4	− 0.0038/0.9998	1.5e-4
				(− 0.7888/0.2148)	(3.1625)	(− 0.7929/0.2108)	(3.1825)
		500	1.0007	− 0.0045/0.9962	1.5e-4	− 0.0055/0.9952	2.1e-4
	50	200	1.0036	(− 0.7930/0.2077)	(3.2306)	(− 0.7952/0.2055)	(3.2533)
				− 9.6e-3/0.9907	4.8e-4	− 9.8e-3/0.9905	5e-4
		500	1.0007	(− 0.7404/0.2600)	(2.9465)	(− 0.7364/0.2640)	(2.9257)
	5	200	1.0004	− 8.4e-3/0.9917	3.8e-4	− 9.0e-3/0.9911	4.3e-4
				(− 0.7375/0.2625)	(2.9256)	(− 0.7394/0.2607)	(2.9476)
		500	1.00007				

Throughout the proof, we use the following identity of Knight (1998),

$$\text{for } x \neq 0, \quad |x - y| - |x| = -\text{ysign}(x) + 2 \int_0^y (\mathbb{I}_{\{x \leq s\}} - \mathbb{I}_{\{x \leq 0\}}) ds. \quad (18)$$

From the model (1), we also observe that

$$y_i = \rho_n^i y_0 + \sum_{j=1}^i \rho_n^{i-j} \epsilon_j, \quad 1 \leq i \leq n. \quad (19)$$

5.1 The case $n(\rho_n - 1) \rightarrow 0$ as $n \rightarrow \infty$

In this subsection, we will prove Theorems 1 and 2. Let us start by presenting two lemmas proved in Appendix.

Lemma 1 *Under the assumptions of Theorem 1, we have*

$$(1). \quad \sqrt{1 - \rho_n^2} y_0 \rightarrow_d \sigma X;$$

- (2). $\frac{1}{\sqrt{n}} \sum_{i=1}^n \rho_n^{i-1} \text{sign}(\epsilon_i) \rightarrow_d Y$;
 (3). $\sqrt{(1 - \rho_n^2)/n} \max_{1 \leq i \leq n} |y_{i-1}| \rightarrow_p 0$;

here X and Y are independent standard normal random variables.

Lemma 2 Under the assumptions of Theorem 2, we have

- (1). $\frac{1}{\sqrt{\kappa_n}} y_0 \rightarrow_d \sigma U$;
 (2). $\frac{1}{\sqrt{n}} \sum_{i=1}^n \rho_n^{i-1} \text{sign}(\epsilon_i) \rightarrow_d V$;
 (3). $\frac{1}{\sqrt{n\kappa_n}} \max_{1 \leq i \leq n} |y_{i-1}| \rightarrow_p 0$;

here U and V are independent standard normal random variables.

Proof of Theorem 1 Denote $\hat{u}_n = \sqrt{n/(1 - \rho_n^2)}(\hat{\rho}_{\text{LAD}} - \rho_n)$. Then \hat{u}_n is the minimizer of the following convex objective function,

$$Z_n(u) = \sum_{i=1}^n \left(|\epsilon_i - \sqrt{(1 - \rho_n^2)/n} y_{i-1} u| - |\epsilon_i| \right).$$

Based on the ideas of Davis et al. (1992) and Ling (2005), if we can prove that, for each u , $Z_n(u)$ converges weakly to a random variable which has a unique minimizer u_{\min} , then \hat{u}_n must converge weakly to u_{\min} .

By (18), we rewrite $Z_n(u)$ as follows

$$Z_n(u) = -uA_n + 2 \sum_{i=1}^n \xi_i(u), \quad (20)$$

here

$$A_n = \sqrt{(1 - \rho_n^2)/n} \sum_{i=1}^n y_{i-1} \text{sign}(\epsilon_i)$$

and

$$\xi_i(u) = \int_0^{\sqrt{(1 - \rho_n^2)/n} y_{i-1} u} (\mathbb{I}_{\{\epsilon_i \leq s\}} - \mathbb{I}_{\{\epsilon_i \leq 0\}}) ds.$$

Next we analyze the asymptotic properties of A_n and $\xi_i(u)$, respectively. By (19), we can decompose A_n as follows

$$\begin{aligned} A_n &= \sqrt{(1 - \rho_n^2)/n} \sum_{i=1}^n \left(\rho_n^{i-1} y_0 + \sum_{j=1}^{i-1} \rho_n^{i-1-j} \epsilon_j \right) \text{sign}(\epsilon_i) \\ &= \sqrt{(1 - \rho_n^2)} y_0 \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n \rho_n^{i-1} \text{sign}(\epsilon_i) + R_{n,1}, \end{aligned} \quad (21)$$

where the remainder

$$R_{n,1} = \sqrt{(1 - \rho_n^2)/n} \sum_{i=1}^n \sum_{j=1}^{i-1} \rho_n^{i-1-j} \epsilon_j \operatorname{sign}(\epsilon_i).$$

The parts (1) and (2) of Lemma 1, together with the continuous mapping theorem, yield that the first term in Eq. (21) converges weakly to the random variable σXY , where X and Y are independent standard normal random variables. Now, we show that the remainder converges to zero in probability. Obviously, by simple calculations, $\mathbb{E}(R_{n,1}) = 0$ and

$$\begin{aligned} \mathbb{E}(R_{n,1}^2) &= \frac{1 - \rho_n^2}{n} \mathbb{E} \left[\left(\sum_{i=1}^n \sum_{j=1}^{i-1} \rho_n^{i-1-j} \epsilon_j \operatorname{sign}(\epsilon_i) \right)^2 \right] \\ &= \frac{1 - \rho_n^2}{n} \sum_{i=1}^n \mathbb{E} \left[\left(\sum_{j=1}^{i-1} \rho_n^{i-1-j} \epsilon_j \right)^2 \right] = \frac{n-1}{n} \left(1 - \frac{\rho_n^2 - \rho_n^{2n}}{(n-1)(1 - \rho_n^2)} \right) \sigma^2 \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$, where the last step is due to Proposition 2. Consequently, A_n converges weakly to the random variable σXY .

We now analyze the second term $\sum_{i=1}^n \xi_i(u)$ in Eq. (20) by the martingale method. Denote the filtration by $\mathcal{F}_{n,i} = \sigma(y_0, \epsilon_1, \dots, \epsilon_i)$ for $1 \leq i \leq n$ and $n \geq 1$, then we can write

$$\sum_{i=1}^n \xi_i(u) = \sum_{i=1}^n \mathbb{E}(\xi_i(u) | \mathcal{F}_{n,i-1}) + R_{n,2},$$

where the remainder

$$R_{n,2} = \sum_{i=1}^n (\xi_i(u) - \mathbb{E}(\xi_i(u) | \mathcal{F}_{n,i-1})).$$

is a martingale with respect to the filtration, $\{\mathcal{F}_{n,i}, 1 \leq i \leq n\}$.

By the Taylor's formula,

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}(\xi_i(u) | \mathcal{F}_{n,i-1}) &= \sum_{i=1}^n \int_0^{\sqrt{(1-\rho_n^2)/ny_{i-1}}u} \mathbb{E}(\mathbb{I}_{\{\epsilon_i \leq s\}} - \mathbb{I}_{\{\epsilon_i \leq 0\}}) ds \\ &= \sum_{i=1}^n \int_0^{\sqrt{(1-\rho_n^2)/ny_{i-1}}u} (F(s) - F(0)) ds \\ &= \sum_{i=1}^n \int_0^{\sqrt{(1-\rho_n^2)/ny_{i-1}}u} \left(f(0)s + \frac{1}{2} f'(s^*) s^2 \right) ds \\ &= \frac{u^2 f(0)}{2} \cdot \frac{1 - \rho_n^2}{n} \sum_{i=1}^n y_{i-1}^2 + R_{n,3}, \end{aligned} \tag{22}$$

where $F(\cdot)$ is the distribution function of ϵ_1 , $s^* \in (0, s)$ and the remainder

$$R_{n,3} = \frac{1}{2} \sum_{i=1}^n \int_0^{\sqrt{(1-\rho_n^2)/ny_{i-1}u}} f'(s^*)s^2 ds.$$

We first consider the sample variance, $\sum_{i=1}^n y_{i-1}^2$. From Eq. (19),

$$\sum_{i=1}^n y_{i-1}^2 = \sum_{i=1}^n \rho_n^{2(i-1)} y_0^2 + \sum_{i=1}^n \rho_n^{i-1} \sum_{j=1}^{i-1} \rho_n^{i-j-1} \epsilon_j y_0 + \sum_{i=1}^n \left(\sum_{j=1}^{i-1} \rho_n^{i-j-1} \epsilon_j \right)^2. \quad (23)$$

Simple calculations yield that

$$\mathbb{E} \left[\sum_{i=1}^n \left(\sum_{j=1}^{i-1} \rho_n^{i-j-1} \epsilon_j \right)^2 \right] = \sum_{i=1}^n \sum_{j=1}^{i-1} \rho_n^{2(i-1-j)} \sigma^2 = \frac{(n-1)\sigma^2}{1-\rho_n^2} \left(1 - \frac{\rho_n^2 - \rho_n^{2n}}{(n-1)(1-\rho_n^2)} \right).$$

Note that, by the Cauchy-Schwarz's inequality,

$$\left| \sum_{i=1}^n \rho_n^{i-1} \sum_{j=1}^{i-1} \rho_n^{i-j-1} \epsilon_j \right|^2 \leq \sum_{i=1}^n \rho_n^{2(i-1)} \sum_{i=1}^n \left(\sum_{j=1}^{i-1} \rho_n^{i-j-1} \epsilon_j \right)^2,$$

so, applying Proposition 2 for $k_n = n$ and part (1) of Lemma 1, for the second term in (23), we can obtain

$$\frac{1-\rho_n^2}{n} \sum_{i=1}^n \rho_n^{i-1} \sum_{j=1}^{i-1} \rho_n^{i-j-1} \epsilon_j y_0 \rightarrow_p 0.$$

It also follows, for the third term in (23),

$$\frac{1-\rho_n^2}{n} \sum_{i=1}^n \left(\sum_{j=1}^{i-1} \rho_n^{i-j-1} \epsilon_j \right)^2 \rightarrow_p 0.$$

Moreover, Proposition 2 and part (1) of Lemma 1, combined with the continuous mapping theorem, imply that

$$\frac{1-\rho_n^2}{n} \sum_{i=1}^n \rho_n^{2(i-1)} y_0^2 \rightarrow_d \sigma^2 X^2.$$

Therefore, we have shown that

$$\frac{1-\rho_n^2}{n} \sum_{i=1}^n y_{i-1}^2 \rightarrow_d \sigma^2 X^2. \quad (24)$$

It remains to deal with the remainders, $R_{n,2}$ and $R_{n,3}$. Notice that

$$\begin{aligned}
|R_{n,3}| &\leq \frac{|u|^3 \sup_{x \in \mathbb{R}} |f'(x)|}{3} \left(\frac{1 - \rho_n^2}{n} \right)^{3/2} \sum_{i=1}^n |y_{i-1}|^3 \\
&\leq \frac{|u|^3 \sup_{x \in \mathbb{R}} |f'(x)|}{3} \left(\frac{1 - \rho_n^2}{n} \right)^{3/2} \max_{1 \leq i \leq n} |y_{i-1}| \sum_{i=1}^n y_{i-1}^2,
\end{aligned}$$

by Eq. (24) and part (3) of Lemma 1, $R_{n,3}$ converges to zero in probability. Since $R_{n,2}$ is a martingale, with the methods of Wang et al. (2020) or Zhou and Lin (2014) and the help of Lemma 1, we can prove that $R_{n,2}$ converges to zero in probability.

Summarizing above discussions, we get that

$$\begin{aligned}
Z_n(u) &= -uA_n + 2 \sum_{i=1}^n \xi_i(u) \\
&= -u \sqrt{(1 - \rho_n^2)y_0} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n \rho_n^{i-1} \text{sign}(\epsilon_i) + u^2 f(0) \frac{1 - \rho_n^2}{n} \sum_{i=1}^n \rho_n^{2(i-1)} y_0^2 + o_p(1) \\
&\rightarrow_d -u\sigma XY + u^2 \sigma^2 f(0) X^2 =: Z(u).
\end{aligned}$$

Because $Z(u)$ has a unique minimum at

$$u_{\min} = Y/(2\sigma f(0)X),$$

by Lemma 2.2 of Davis et al. (1992), we derive that

$$\hat{u}_n = \sqrt{n/(1 - \rho_n^2)}(\hat{\rho}_{\text{LAD}} - \rho_n) \rightarrow_d \frac{1}{2\sigma f(0)} C.$$

Combining this with Eq. (24) and using the continuous mapping theorem, we complete the proof of Theorem 1. \square

Proof of Theorem 2 The proof is similar to that of Theorem 1. In this case, $\hat{u}_n = \sqrt{n\kappa_n}(\hat{\rho}_{\text{LAD}} - \rho_n)$, $Z_n(u) = \sum_{i=1}^n (|\epsilon_i - (n\kappa_n)^{-1/2} y_{i-1} u| - |\epsilon_i|)$, and in (20),

$$A_n = \frac{1}{\sqrt{n\kappa_n}} \sum_{i=1}^n y_{i-1} \text{sign}(\epsilon_i)$$

and

$$\xi_i(u) = \int_0^{y_{i-1} u / \sqrt{n\kappa_n}} (\mathbb{I}_{\{\epsilon_i \leq s\}} - \mathbb{I}_{\{\epsilon_i \leq 0\}}) ds.$$

Lemma 2 is applied to prove the corresponding terms as in the proof of Theorem 1. Here we only partially demonstrate the role of the assumptions, $\kappa_n/n \rightarrow \infty$ and $\kappa_n(\rho_n - 1) \rightarrow 0$ as $n \rightarrow \infty$. Let us consider the normalized sample variance. For the third term of (23), we want to show that

$$\frac{1}{n\kappa_n} \sum_{i=1}^n \left(\sum_{j=1}^{i-1} \rho_n^{i-j-1} \epsilon_j \right)^2 \rightarrow_p 0.$$

By a simple calculation, we get

$$\mathbb{E} \left[\sum_{i=1}^n \left(\sum_{j=1}^{i-1} \rho_n^{i-j-1} \epsilon_j \right)^2 \right] = \sum_{i=1}^n \sum_{j=1}^{i-1} \rho_n^{2(i-1-j)} \sigma^2 = \frac{(n-1)\sigma^2}{1-\rho_n^2} \left(1 - \frac{\rho_n^2 - \rho_n^{2n}}{(n-1)(1-\rho_n^2)} \right).$$

Then, Proposition 2 yields that

$$\frac{1}{n\kappa_n} \cdot \frac{(n-1)\sigma^2}{1-\rho_n^2} \left(1 - \frac{\rho_n^2 - \rho_n^{2n}}{(n-1)(1-\rho_n^2)} \right) = O\left(\frac{n}{\kappa_n}\right),$$

which tends to zero if $\kappa_n/n \rightarrow \infty$ as $n \rightarrow \infty$. The proof is complete. \square

5.2 The case $n(\rho_n - 1) \rightarrow \gamma$ as $n \rightarrow \infty$

In this subsection, we will prove Theorems 3 and 4.

Firstly, we prove Theorem 3 following the line as in the proofs of the case $n(\rho_n - 1) \rightarrow 0$. In this case, $\hat{u}_n = n(\hat{\rho}_{\text{LAD}} - \rho_n)$ and it is the minimizer of the following convex function,

$$Z_n(u) = \sum_{i=1}^n (|\epsilon_i - n^{-1}y_{i-1}u| - |\epsilon_i|). \quad (25)$$

Again by the Knight's identity (18), we can rewrite the quantity $Z_n(u)$ in (25) as identity (20) with

$$A_n = \frac{1}{n} \sum_{i=1}^n y_{i-1} \text{sign}(\epsilon_i) \quad (26)$$

and

$$\xi_i(u) = \int_0^{n^{-1}y_{i-1}u} (\mathbb{1}_{\{\epsilon_i \leq s\}} - \mathbb{1}_{\{\epsilon_i \leq 0\}}) ds. \quad (27)$$

Again we use the filtration $\mathcal{F}_{n,i} = \sigma(y_0, \epsilon_1, \dots, \epsilon_i)$ for $1 \leq i \leq n$ and $n \geq 1$, and write

$$\sum_{i=1}^n \xi_i(u) = \sum_{i=1}^n \mathbb{E}(\xi_i(u) | \mathcal{F}_{n,i-1}) + R_{n,1},$$

where

$$R_{n,1} = \sum_{i=1}^n (\xi_i(u) - \mathbb{E}(\xi_i(u) | \mathcal{F}_{n,i-1})) \quad (28)$$

is a martingale with respect to the filtration, $\{\mathcal{F}_{n,i}, 1 \leq i \leq n\}$. By the same argument as in (22),

$$\sum_{i=1}^n \mathbb{E}(\xi_i(u) | \mathcal{F}_{n,i-1}) = \frac{u^2 f(0)}{2} \cdot \frac{1}{n^2} \sum_{i=1}^n y_{i-1}^2 + R_{n,2},$$

where

$$R_{n,2} = \frac{1}{2} \sum_{i=1}^n \int_0^{n^{-1}y_{i-1}u} f'(s^*) s^2 ds \quad (29)$$

and s^* is between 0 and s . As in the proofs of Theorems 1 and 2, to obtain the asymptotic distribution of $Z_n(u)$, the key point is to analyze the joint distribution of $\left(\frac{1}{n} \sum_{i=1}^n y_{i-1} \text{sign}(\epsilon_i), \frac{1}{n^2} \sum_{i=1}^n y_{i-1}^2\right)$.

Lemma 3 *Under the assumptions of Theorem 3, the process $(K_n(t), L_n(t), 0 \leq t \leq 1)$ defined as in (9) converges weakly to the continuous process $(\mathbf{X}(t), 0 \leq t \leq 1)$ defined as in (10) with independent Gaussian increments, mean vector zero and covariance matrix $\Gamma(t)$ defined as in (11). Moreover, we also have*

$$\mathbf{X}^T(t) = \int_0^t \Lambda^{1/2}(s) d\mathbf{B}^T(h(s)), \quad 0 \leq t \leq 1, \quad (30)$$

where $\Lambda(\cdot)$ is a non-negative definite matrix-valued function given by

$$\Lambda(t) = (\lambda_{ij}(t)) := \frac{1}{2} \begin{pmatrix} 1 - \tanh(2\gamma(1-t)) & \mathbb{E}|\epsilon_1| \text{sech}(2\gamma(1-t)) \\ \mathbb{E}|\epsilon_1| \text{sech}(2\gamma(1-t)) & 1 + \tanh(2\gamma(1-t)) \end{pmatrix}, \quad 0 \leq t \leq 1,$$

$\mathbf{B} = (B_1, B_2)$ is a two-dimensional standard Brownian motion and

$$h(t) = \frac{1}{\gamma} (\sinh(2\gamma) - \sinh(2\gamma(1-t))), \quad 0 \leq t \leq 1.$$

Proposition 1 *Under the assumptions of Theorem 3, as $n \rightarrow \infty$,*

$$\left(\frac{1}{n} \sum_{i=1}^n y_{i-1} \text{sign}(\epsilon_i), \frac{1}{n^2} \sum_{i=1}^n y_{i-1}^2\right) \rightarrow_d \left(\int_0^1 L(t) dK(t), \int_0^1 e^{-2\gamma(1-t)} L^2(t) dt\right).$$

Proof Observe that

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n y_{i-1} \text{sign}(\epsilon_i) &= \sum_{i=1}^n \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} \rho_n^{n-j} \epsilon_j \right) \frac{1}{\sqrt{n}} \rho_n^{i-1-n} \text{sign}(\epsilon_i) \\ &= \int_0^1 L_n(t) dK_n(t)\end{aligned}$$

and

$$\begin{aligned}\frac{1}{n^2} \sum_{i=1}^n y_{i-1}^2 &= \sum_{i=1}^n \frac{1}{n} \left(\frac{1}{\sqrt{n}} \sum_{j=1}^i \rho_n^{i-j} \epsilon_j \right)^2 \\ &= \int_0^1 e^{-2\gamma(1-t)} L_n^2(t) dt + R_{n,3},\end{aligned}$$

where the remainder

$$R_{n,3} = \frac{1}{n} \sum_{i=1}^n \rho_n^{2(i-n)} L_n^2\left(\frac{i}{n}\right) - \int_0^1 e^{-2\gamma(1-t)} L_n^2(t) dt.$$

According to the same argument in the proof of Lemma 2.2 in Chan and Wei (1987), we can show that $R_{n,3} \rightarrow_p 0$ as $n \rightarrow \infty$. Hence Theorem 2.1 in Hansen (1992) and Lemma 3 immediately yield the desired result. \square

Proof of Theorem 3 Recall the analysis in the beginning of this subsection. Notice that $R_{n,2}$ in (29) satisfies

$$\begin{aligned}2|R_{n,2}| &\leq \frac{|u|^3 \sup_{x \in \mathbb{R}} |f'(x)|}{3} \frac{1}{n^3} \sum_{i=1}^n |y_{i-1}|^3 \\ &\leq \frac{|u|^3 \sup_{x \in \mathbb{R}} |f'(x)|}{3} \frac{1}{n} \max_{1 \leq i \leq n} |y_{i-1}| \cdot \frac{1}{n^2} \sum_{i=1}^n y_{i-1}^2.\end{aligned}\tag{31}$$

By Propositions 1 and 3, $R_{n,2}$ converges to zero in probability. In addition, following the same line of Wang et al. (2020) or Zhou and Lin (2014), we can also show that $R_{n,1}$ in (28) converges to zero in probability. Combining with Proposition 1, we have

$$\begin{aligned}Z_n(u) &= -uA_n + 2 \sum_{i=1}^n \xi_i(u) \\ &\rightarrow_d -u \int_0^1 L(t) dK(t) + u^2 f(0) \int_0^1 e^{-2\gamma(1-t)} L^2(t) dt =: Z(u).\end{aligned}$$

Because $Z(u)$ has a unique minimum at

$$u_{\min} = \frac{1}{2f(0)} \cdot \frac{\int_0^1 L(t) dK(t)}{\int_0^1 e^{-2\gamma(1-t)} L^2(t) dt},$$

by Lemma 2.2 of Davis et al. (1992),

$$\hat{u}_n = n(\hat{\rho}_{\text{LAD}} - \rho_n) \rightarrow_d \frac{1}{2f(0)} \cdot \frac{\int_0^1 L(t) dK(t)}{\int_0^1 e^{-2\gamma(1-t)} L^2(t) dt},$$

as $n \rightarrow \infty$. Applying the continuous mapping theorem and Proposition 1 again, we complete the proof of Theorem 3. \square

Proof of Theorem 4 Denote the square root of the matrix-valued function $\Lambda(t)$ defined as in Lemma 3 by

$$\Lambda^{1/2}(t) = \begin{pmatrix} \tilde{\lambda}_{11}(t) & \tilde{\lambda}_{12}(t) \\ \tilde{\lambda}_{12}(t) & \tilde{\lambda}_{22}(t) \end{pmatrix}, \quad 0 \leq t \leq 1.$$

Then, from Lemma 3, we know that

$$\begin{cases} dK(t) = \tilde{\lambda}_{11}(t) dB_1(h(t)) + \tilde{\lambda}_{12}(t) dB_2(h(t)) \\ dL(t) = \tilde{\lambda}_{12}(t) dB_1(h(t)) + \tilde{\lambda}_{22}(t) dB_2(h(t)) \end{cases}. \quad (32)$$

Moreover, by the time change for Itô integrals, e.g. Theorem 8.5.7 in Øksendal (2005), we can rewrite Eqs. (32) as

$$\begin{cases} dK(t) = \tilde{\lambda}_{11}(t) \sqrt{h'(t)} d\tilde{B}_1(t) + \tilde{\lambda}_{12}(t) \sqrt{h'(t)} d\tilde{B}_2(t) \\ dL(t) = \tilde{\lambda}_{12}(t) \sqrt{h'(t)} d\tilde{B}_1(t) + \tilde{\lambda}_{22}(t) \sqrt{h'(t)} d\tilde{B}_2(t) \end{cases}, \quad (33)$$

here $\tilde{B} = (\tilde{B}_1, \tilde{B}_2)$ is also a two-dimensional standard Brownian motion. If denote

$$d\hat{B}(t) = \frac{\tilde{\lambda}_{11}(t) \sqrt{h'(t)} d\tilde{B}_1(t) + \tilde{\lambda}_{12}(t) \sqrt{h'(t)} d\tilde{B}_2(t)}{\sqrt{\tilde{\lambda}_{11}^2(t) h'(t) + \tilde{\lambda}_{12}^2(t) h'(t)}},$$

then, by the Lévy characterization of Brownian motion, e.g. Theorem 8.6.1 in Øksendal (2005), we know that \hat{B} is a two-dimensional standard Brownian motion. Notice that

$$\tilde{\lambda}_{11}^2(t) h'(t) + \tilde{\lambda}_{12}^2(t) h'(t) = e^{-2\gamma(1-t)}, \quad 0 \leq t \leq 1,$$

hence the random variable $\mathcal{L}(\gamma)$ can be represented as

$$2f(0) \cdot \mathcal{L}(\gamma) = \frac{\int_0^1 e^{-\gamma(1-t)} L(t) d\hat{B}(t)}{\sqrt{\int_0^1 e^{-2\gamma(1-t)} L^2(t) dt}}.$$

Now, the subsequent proof will base on Theorem 1 of Rootzén (1980) which is restated as Theorem 6 in Appendix. We need verify the assumptions in Theorem 6. Let

$$\varphi_\gamma(t) = 2\gamma e^{-\gamma} e^{-\gamma(1-t)} L(t), \quad 0 \leq t \leq 1, \gamma > 0,$$

and

$$\psi_\gamma(t) = \sqrt{-2\gamma} e^{-\gamma(1-t)} L(t), \quad 0 \leq t \leq 1, \gamma < 0,$$

then we have the following lemmas whose proofs are postponed to Appendix. \square

Lemma 4 As $\gamma \rightarrow \infty$,

$$\sup_{0 \leq t \leq 1} \left| \int_0^t \varphi_\gamma(s) ds \right| \rightarrow_p 0 \quad (34)$$

and

$$\int_0^1 \varphi_\gamma^2(t) dt \rightarrow_p \mathcal{N}^2(0, 1). \quad (35)$$

Lemma 5 As $\gamma \rightarrow -\infty$,

$$\sup_{0 \leq t \leq 1} \left| \int_0^t \psi_\gamma(s) ds \right| \rightarrow_p 0 \quad (36)$$

and

$$\int_0^1 \psi_\gamma^2(t) dt \rightarrow_p 1. \quad (37)$$

With the help of Lemmas 4, 5, and Theorem 6, we complete the proof of Theorem 4. \square

6 Technical appendix and proofs

Proposition 2 Let $\{k_n, n \geq 1\}$ be a sequence of positive numbers such that $k_n \rightarrow \infty$ and $k_n |1 - \rho_n^2| \rightarrow 0$ as $n \rightarrow \infty$, then we have $\rho_n^{2k_n} \rightarrow 1$ and

$$\frac{1 - \rho_n^{2k_n}}{k_n(1 - \rho_n^2)} = 1 + O(k_n(1 - \rho_n^2)) \quad (38)$$

as $n \rightarrow \infty$.

Proof Applying the Taylor's formula, we have

$$\begin{aligned}\log(\rho_n^{2k_n}) &= k_n \log(1 + \rho_n^2 - 1) \\ &= k_n \left(\rho_n^2 - 1 - \frac{(\rho_n^2 - 1)^2}{2} + \dots \right).\end{aligned}$$

So $\rho_n^{2k_n} \rightarrow 1$. Moreover, using the Taylor's formula again, we obtain, as $n \rightarrow \infty$

$$\begin{aligned}\rho_n^{2k_n} &= \exp\{k_n(\rho_n^2 - 1) + O(k_n(\rho_n^2 - 1)^2)\} \\ &= 1 + k_n(\rho_n^2 - 1) + O(k_n^2(\rho_n^2 - 1)^2).\end{aligned}$$

Thus the estimation (38) follows. \square

Proposition 3 For model (1), assume that $y_0 = 0$ and $n(\rho_n - 1) \rightarrow \gamma$ for some fixed $\gamma \neq 0$ as $n \rightarrow \infty$, then we have

$$\frac{1}{n} \max_{1 \leq i \leq n} |y_{i-1}| \rightarrow_p 0, \text{ as } n \rightarrow \infty.$$

Proof For any $\varepsilon > 0$, by the Kolmogorov's maximal inequality,

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n} \max_{1 \leq i \leq n} |y_{i-1}| > \varepsilon\right) &= \mathbb{P}\left(\max_{1 \leq i \leq n} \left| \sum_{j=1}^{i-1} \rho_n^{i-1-j} \varepsilon_j \right| > n\varepsilon\right) \\ &\leq \frac{\sigma^2}{n^2 \varepsilon^2} \cdot \sum_{j=1}^{n-1} \rho_n^{2(n-1-j)} = \frac{\sigma^2(\rho_n^{2(n-1)} - 1)}{n^2 \varepsilon^2(\rho_n^2 - 1)} \rightarrow 0,\end{aligned}$$

as $n \rightarrow \infty$, which implies the desired result immediately. Here we used the facts, $n(\rho_n - 1) \rightarrow \gamma$ and $\rho_n^n \rightarrow e^\gamma$ as $n \rightarrow \infty$. \square

Proof of Lemma 1 (1). Note that there exists a sequence m_n such that $m_n(1 - \rho_n) \rightarrow \infty$ as $n \rightarrow \infty$, which implies that $\rho_n^{m_n} = o(1)$. So we can rewrite $\sqrt{1 - \rho_n^2} y_0$ as follows

$$\begin{aligned}\sqrt{1 - \rho_n^2} y_0 &= \sqrt{1 - \rho_n^2} \sum_{j=0}^{\infty} \rho_n^j \varepsilon_{-j} \\ &= \sqrt{1 - \rho_n^2} \sum_{j=0}^{m_n} \rho_n^j \varepsilon_{-j} + \sqrt{1 - \rho_n^2} \sum_{j=m_n+1}^{\infty} \rho_n^j \varepsilon_{-j} \\ &=: \Gamma_{n,1} + \Gamma_{n,2}.\end{aligned}$$

Obviously, $\mathbb{E}(\Gamma_{n,2}) = 0$ and

$$\mathbb{E}(\Gamma_{n,2}^2) = \rho_n^{2(m_n+1)} \sigma^2 = o(1),$$

which immediately yield that $\Gamma_{n,2} \rightarrow_p 0$ as $n \rightarrow \infty$. We apply the Lindeberg-Feller central limit theorem for the first term $\Gamma_{n,1}$. Denote $\zeta_{n,j} = \sqrt{1 - \rho_n^2} \rho_n^j \varepsilon_{-j}$, then $\mathbb{E}(\Gamma_{n,1}) = 0$ and

$$\mathbb{E}(\Gamma_{n,1}^2) = (1 - \rho_n^{2(m_n+1)})\sigma^2 \rightarrow \sigma^2$$

as $n \rightarrow \infty$. Hence, to complete the proof, we verify the Lindeberg condition,

$$\text{for any } \varepsilon > 0, \quad \sum_{j=0}^{m_n} \mathbb{E}(\zeta_{n,j}^2 \mathbb{I}_{\{|\zeta_{n,j}| > \varepsilon\}}) = o(1).$$

In fact, since $0 < \rho_n < 1$, by dominated convergence theorem, we have

$$\begin{aligned} \sum_{j=0}^{m_n} \mathbb{E}(\zeta_{n,j}^2 \mathbb{I}_{\{|\zeta_{n,j}| > \varepsilon\}}) &= (1 - \rho_n^2) \sum_{j=0}^{m_n} \rho_n^{2j} \mathbb{E}(\epsilon_{-j}^2 \mathbb{I}_{\{|\zeta_{n,j}| > \varepsilon\}}) \\ &\leq (1 - \rho_n^2) \sum_{j=0}^{m_n} \rho_n^{2j} \mathbb{E}(\epsilon_{-j}^2 \mathbb{I}_{\{\sqrt{1-\rho_n^2}|\epsilon_{-j}| > \varepsilon\}}) \\ &= (1 - \rho_n^{2(m_n+1)}) \mathbb{E}(\epsilon_1^2 \mathbb{I}_{\{\sqrt{1-\rho_n^2}|\epsilon_1| > \varepsilon\}}) = o(1). \end{aligned}$$

(2). Because the Lindeberg's condition holds obviously in this case, we only need to estimate the asymptotic variance. Noticing that ϵ_t have zero median, we get $\mathbb{E}(\text{sign}(\epsilon_t)) = 0$ and

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \rho_n^{i-1} \text{sign}(\epsilon_i)\right)^2\right] = \frac{1 - \rho_n^{2n}}{n(1 - \rho_n^2)} \rightarrow 1$$

as $n \rightarrow \infty$, here we use the facts, $(1 - \rho_n^n) \sim n(1 - \rho_n)$ and $\rho_n^n \rightarrow 1$ as $n \rightarrow \infty$.

(3). Since

$$\max_{1 \leq i \leq n} |y_{i-1}| \leq \max\{|y_0|, \max_{2 \leq i \leq n} |y_{i-1}|\}$$

and

$$\max_{2 \leq i \leq n} |y_{i-1}| \leq \max_{2 \leq i \leq n} |\rho_n^{i-1} y_0| + \max_{2 \leq i \leq n} \left| \sum_{j=1}^{i-1} \rho_n^{i-1-j} \epsilon_j \right|,$$

part (1) of this lemma implies that

$$\sqrt{(1 - \rho_n^2)/n} \max_{2 \leq i \leq n} |y_{i-1}| \leq \sqrt{(1 - \rho_n^2)/n} \max_{2 \leq i \leq n} \left| \sum_{j=1}^{i-1} \rho_n^{i-1-j} \epsilon_j \right| + o_p(1).$$

By the Kolmogorov's maximal inequality and Proposition 2, it follows that, for any $\varepsilon > 0$,

$$\begin{aligned}
& \mathbb{P}\left(\sqrt{(1-\rho_n^2)/n} \max_{2 \leq i \leq n} \left| \sum_{j=1}^{i-1} \rho_n^{i-1-j} \epsilon_j \right| > \varepsilon\right) \\
&= \mathbb{P}\left(\sqrt{(1-\rho_n^2)/n} \max_{2 \leq i \leq n} \left| \sum_{j=1}^{i-1} \rho_n^{j-1} \epsilon_{i-j} \right| > \varepsilon\right) \\
&\leq \frac{\sigma^2}{\varepsilon^2} \cdot \frac{1-\rho_n^{2n}}{n} = \frac{\sigma^2}{\varepsilon^2} \cdot \frac{1-\rho_n^{2n}}{n(1-\rho_n)} \cdot (1-\rho_n) = o(1),
\end{aligned}$$

which complete the proof of part (3). \square

Proof of Lemma 2 Noting that $\kappa_n(1-\rho_n) \rightarrow 0$ and applying Proposition 2, we can show that, $(1-\rho_n^{\kappa_n}) \sim \kappa_n(1-\rho_n)$ and $\rho_n^{\kappa_n} \rightarrow 1$ as $n \rightarrow \infty$. We omit the remainder of the argument since it is analogous to that in the proof of Lemma 1. \square

Proposition 1 are required to achieve Theorem 3 and 4. So, in the rest of this section, we mainly make the supplement for the proofs in Sect. 5. First of all, we need the following modified version of Theorem 1.4 in Chapter 7 of Ethier and Kurtz (1986).

Theorem 5 (Ethier and Kurtz 1986) *Let $\Gamma(t) = (\gamma_{ij}(t))_{i,j=1}^d$ be a continuous, symmetric, $d \times d$ matrix-valued function, defined on $[0, 1]$, satisfying $\Gamma(0) = 0$ and for any $0 \leq s < t \leq 1$,*

$$\sum_{i,j=1}^d (\gamma_{ij}(t) - \gamma_{ij}(s)) \alpha_i \alpha_j \geq 0, \quad \forall \alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{R}^d.$$

Let $\{\xi_k^n, \mathcal{F}_k^n, n \geq 1, 1 \leq k \leq n\}$ be an \mathbb{R}^d -valued square-integrable martingale difference array on a completed probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Denote $S_n(t) = \sum_{k=1}^{[nt]} \xi_k^n$ and

$$\Theta_n(t) = (\theta_n^{ij}(t))_{i,j=1}^d = \sum_{k=1}^{[nt]} \mathbb{E}((\xi_k^n)^T \xi_k^n | \mathcal{F}_{k-1}^n),$$

for $0 \leq t \leq 1$. Suppose that,

$$\mathbb{E}\left(\sup_{0 \leq t \leq 1} |S_n(t) - S_n(t-)|^2\right) \rightarrow 0, \quad \text{as } n \rightarrow \infty, \quad (39)$$

and for $i, j = 1, 2, \dots, d$,

$$\mathbb{E}\left(\sup_{0 \leq t \leq 1} |\theta_n^{ij}(t) - \theta_n^{ij}(t-)|\right) \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (40)$$

In addition, if, for each $0 \leq t \leq 1$ and $i, j = 1, 2, \dots, d$,

$$\theta_n^{ij}(t) \rightarrow_p \gamma_{ij}(t), \quad \text{as } n \rightarrow \infty, \quad (41)$$

then the process S_n converges weakly to a continuous process G with independent Gaussian increments, mean vector zero and covariance matrix $\Gamma(t)$, on the Skorohod space $D_{\mathbb{R}^d}([0, 1])$ of \mathbb{R}^d -valued cadlag paths on $[0, 1]$.

Proof of Lemma 3 Recall that, $\mathcal{F}_i = \mathcal{F}_{n,i} = \sigma(\epsilon_1, \dots, \epsilon_i)$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. For convenience, denote $\mathbf{X}_{n,i} = (K_{n,i}, L_{n,i})$ and $\mathbf{X}_n(t) = \sum_{i=1}^{[nt]} \mathbf{X}_{n,i}$ for $0 \leq t \leq 1$, then it is clear that $\{\mathbf{X}_{n,i}, \mathcal{F}_i\}$ is an \mathbb{R}^2 -valued square-integral martingale difference array. Note that, under these circumstances,

$$\Theta_n(t) = \sum_{i=1}^n \mathbb{E}(\mathbf{X}_{n,i}^T \mathbf{X}_{n,i} | \mathcal{F}_{i-1}) = \sum_{i=1}^n \mathbb{E}(\mathbf{X}_{n,i}^T \mathbf{X}_{n,i}).$$

To apply Theorem 5, we first verify the non-negative definiteness of the matrix-valued function $\Gamma(t)$ defined as in (11). Clearly, for $0 \leq s < t \leq 1$, the first principle minor of $\Gamma(t) - \Gamma(s)$ is positive for any $\gamma \neq 0$; moreover, since $\sigma^2 = 1$, we know that $\mathbb{E}|\epsilon_1| \leq 1$ and it follows that

$$\det(\Gamma(t) - \Gamma(s)) = \frac{1}{4\gamma^2} (e^{\gamma(t-s)} - e^{-\gamma(t-s)})^2 - ((t-s)\mathbb{E}|\epsilon_1|)^2 \geq 0.$$

So the matrix-valued function $\Gamma(t)$ is non-negative definite.

Now, we check the condition (39), i.e.

$$\mathbb{E} \left(\max_{1 \leq i \leq n} |\mathbf{X}_{n,i}|^2 \right) \rightarrow 0, \quad (42)$$

as $n \rightarrow \infty$. Obviously, since $\rho_n^n \rightarrow e^\gamma$ as $n \rightarrow \infty$, we know that

$$\max_{1 \leq i \leq n} |\mathbf{X}_{n,i}|^2 \leq \frac{C_1}{n} \left(1 \vee \max_{1 \leq i \leq n} \epsilon_i^2 \right),$$

for some positive constant C_1 independent of n . For any $\varepsilon > 0$, by the Chebyshev's inequality, we can get

$$\begin{aligned} \mathbb{P}(\max_{1 \leq i \leq n} \epsilon_i^2 > n\varepsilon) &\leq n\mathbb{P}(\epsilon_1^2 > n\varepsilon) \\ &\leq \varepsilon^{-(2+\delta)/2} n^{-\delta/2} \cdot \mathbb{E}|\epsilon_1|^{2+\delta} \rightarrow 0, \end{aligned}$$

as $n \rightarrow \infty$. So

$$\max_{1 \leq i \leq n} |\mathbf{X}_{n,i}|^2 \rightarrow_p 0. \quad (43)$$

On the other hand, notice that, for each n ,

$$\mathbb{E} \left[\left(\frac{\max_{1 \leq i \leq n} \epsilon_i^2}{n} \right)^{(2+\delta)/2} \right] \leq \frac{\mathbb{E}|\epsilon_1|^{2+\delta}}{n^{\delta/2}}.$$

Hence the family of random variables, $\{\max_{1 \leq i \leq n} |\mathbf{X}_{n,i}|^2, n \geq 1\}$, is uniformly integrable, which, combined with (43), suggests that the condition (42) holds.

We next calculate the covariance matrix. Observe that

$$\sum_{i=1}^{[nt]} \mathbb{E}(\mathbf{X}_{n,i}^T \mathbf{X}_{n,i}) = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^{[nt]} \rho_n^{2(i-1-n)} & t \rho_n^{-1} \mathbb{E}|\epsilon_1| \\ t \rho_n^{-1} \mathbb{E}|\epsilon_1| & \frac{1}{n} \sum_{i=1}^{[nt]} \rho_n^{2(n-i)} \end{pmatrix}.$$

Therefore it is straightforward to show that, as $n \rightarrow \infty$,

$$\begin{aligned} \sup_{0 \leq t \leq 1} |\theta_n^{11}(t) - \theta_n^{11}(t-)| &= \frac{1}{n} \max_{1 \leq i \leq n} \rho_n^{2(i-1-n)} \rightarrow 0, \\ \sup_{0 \leq t \leq 1} |\theta_n^{22}(t) - \theta_n^{22}(t-)| &= \frac{1}{n} \max_{1 \leq i \leq n} \rho_n^{2(n-i)} \rightarrow 0, \\ \sup_{0 \leq t \leq 1} |\theta_n^{12}(t) - \theta_n^{12}(t-)| &= \sup_{0 \leq t \leq 1} |\theta_n^{21}(t) - \theta_n^{21}(t-)| = 0, \end{aligned}$$

and

$$\Theta_n(t) \rightarrow \Gamma(t) = \begin{pmatrix} \frac{e^{-2\gamma}}{2\gamma}(e^{2\gamma t} - 1) & t \mathbb{E}|\epsilon_1| \\ t \mathbb{E}|\epsilon_1| & \frac{e^{2\gamma}}{2\gamma}(1 - e^{-2\gamma t}) \end{pmatrix}.$$

Hence, by Theorem 5, the process \mathbf{X}_n converges weakly to a continuous process $\mathbf{X} = (\mathbf{X}(t), 0 \leq t \leq 1)$ which has independent Gaussian increments, zero mean vector and covariance matrix $\Gamma(t)$.

We now turn to the proof of Eq. (30). From the proof of Theorem 1.1 in Chapter 7 of Ethier and Kurtz (1986), we know that the limit process \mathbf{X} can be represented by

$$\mathbf{X}^T(t) = \int_0^t \Lambda^{1/2}(s) d\mathbf{B}^T(h(s)), \quad 0 \leq t \leq 1,$$

where $\Lambda(\cdot)$ is a non-negative definite matrix-valued function, $\mathbf{B} = (B_1, B_2)$ is a two-dimensional standard Brownian motion, and

$$h(t) = \gamma_{11}(t) + \gamma_{22}(t), \quad 0 \leq t \leq 1.$$

In fact, the entries $\lambda_{ij}(\cdot)$ of $\Lambda(\cdot)$ are the solutions of the following equations,

$$\begin{cases} \gamma_{11}(t) = \int_0^t \lambda_{11}(s) dh(s), \\ \gamma_{22}(t) = \int_0^t \lambda_{22}(s) dh(s), \\ \gamma_{12}(t) = \int_0^t \lambda_{12}(s) dh(s), \\ \gamma_{21}(t) = \int_0^t \lambda_{21}(s) dh(s). \end{cases} \quad (44)$$

Solving (44), we can get

$$\begin{aligned}\lambda_{11}(t) &= \frac{e^{-2\gamma(1-t)}}{e^{2\gamma(1-t)} + e^{-2\gamma(1-t)}} = \frac{1 - \tanh(2\gamma(1-t))}{2}, \\ \lambda_{22}(t) &= \frac{e^{2\gamma(1-t)}}{e^{2\gamma(1-t)} + e^{-2\gamma(1-t)}} = \frac{1 + \tanh(2\gamma(1-t))}{2}, \\ \lambda_{12}(t) = \lambda_{21}(t) &= \frac{\mathbb{E}|\epsilon_1|}{e^{2\gamma(1-t)} + e^{-2\gamma(1-t)}} = \frac{\mathbb{E}|\epsilon_1|}{2 \cosh(2\gamma(1-t))}.\end{aligned}$$

Hence we complete the proof of Lemma 3. \square

Theorem 6 (Rootzén 1980) Let $(B(t), 0 \leq t \leq 1)$ be a standard Brownian motion with respect to the filtration $(\mathcal{F}_t, 0 \leq t \leq 1)$. Suppose $(\phi_n(t), 0 \leq t \leq 1)_{n \geq 1}$ is a sequence of random functions which is adapted to the filtration $(\mathcal{F}_t, 0 \leq t \leq 1)$. If

$$\sup_{0 \leq t \leq 1} \left| \int_0^t \phi_n(s) ds \right| \rightarrow_p 0, \quad \text{as } n \rightarrow \infty,$$

and

$$\int_0^1 \phi_n^2(t) dt \rightarrow_p \tau, \quad \text{as } n \rightarrow \infty,$$

for some random variable τ such that $\tau > 0$ a.s., then

$$\frac{\int_0^1 \phi_n(t) dB(t)}{\sqrt{\int_0^1 \phi_n^2(t) dt}} \rightarrow_d \mathcal{N}(0, 1), \quad \text{as } n \rightarrow \infty.$$

Theorem 7 (Graversen and Peskir 2000) Let $(V(t), t \geq 0)$ be the Ornstein-Uhlenbeck process solving

$$dV(t) = -\beta V(t)dt + dB(t)$$

with $V(0) = 0$, where $\beta > 0$ and $(B(t), t \geq 0)$ is a standard Brownian motion, then there exist universal positive constants c_0, c_1 such that, for all stopping times τ of $(V(t), t \geq 0)$,

$$\frac{c_0}{\sqrt{\beta}} \mathbb{E} \sqrt{\log(1 + \beta\tau)} \leq \mathbb{E} \left(\max_{0 \leq t \leq \tau} |V(t)| \right) \leq \frac{c_1}{\sqrt{\beta}} \mathbb{E} \sqrt{\log(1 + \beta\tau)}.$$

Proof of Lemma 4 Since

$$\sup_{0 \leq t \leq 1} \left| \int_0^t \varphi_\gamma(s) ds \right| \leq \int_0^1 |\varphi_\gamma(t)| dt,$$

it is enough to show $\int_0^1 |\varphi_\gamma(t)| dt \rightarrow_p 0$ as $\gamma \rightarrow \infty$. Notice that, from Lemma 3,

$$L(t) \sim \mathcal{N}\left(0, \frac{e^{2\gamma}(1 - e^{-2\gamma t})}{2\gamma}\right),$$

so

$$\begin{aligned} \mathbb{E} \int_0^1 |\varphi_\gamma(t)| \, dt &= 2\gamma e^{-\gamma} \int_0^1 e^{-\gamma(1-t)} \mathbb{E}|L(t)| \, dt \\ &= \sqrt{4\gamma/\pi} e^{-\gamma} \int_0^1 \sqrt{e^{2\gamma t} - 1} \, dt \\ &\leq \sqrt{4/(\pi\gamma)}(1 - e^{-\gamma}) \rightarrow 0, \quad \text{as } \gamma \rightarrow \infty. \end{aligned}$$

By the Markov's inequality, this establishes Eq. (34).

We next prove Eq. (35). Firstly, using integration by parts, we can get

$$\begin{aligned} \int_0^1 \varphi_\gamma^2(t) \, dt &= 4\gamma^2 e^{-4\gamma} \int_0^1 e^{2\gamma t} L^2(t) \, dt \\ &= 2\gamma e^{-4\gamma} \int_0^1 L^2(t) \, de^{2\gamma t} \\ &= 2\gamma e^{-4\gamma} \left(L^2(1)e^{2\gamma} - \int_0^1 e^{2\gamma t} \, dL^2(t) \right). \end{aligned}$$

Furthermore, since

$$\tilde{\lambda}_{12}^2(t)h'(t) + \tilde{\lambda}_{22}^2(t)h'(t) = e^{2\gamma(1-t)}, \quad 0 \leq t \leq 1,$$

the Itô's formula, together with Eq. (33), yields that

$$dL^2(t) = 2L(t) \, dL(t) + e^{2\gamma(1-t)} \, dt.$$

Therefore

$$\int_0^1 \varphi_\gamma^2(t) \, dt = 2\gamma e^{-2\gamma} L^2(1) - 2\gamma e^{-2\gamma} - 4\gamma e^{-4\gamma} \int_0^1 e^{2\gamma t} L(t) \, dL(t).$$

Noticing that

$$L(1) \sim \mathcal{N}\left(0, \frac{e^{2\gamma} - 1}{2\gamma}\right),$$

we have

$$2\gamma e^{-2\gamma} L^2(1) =_d \left(\sqrt{1 - e^{-2\gamma}} \mathcal{N}(0, 1) \right)^2 \rightarrow_p \mathcal{N}^2(0, 1),$$

as $\gamma \rightarrow \infty$. In addition,

$$\begin{aligned}\mathbb{E}\left(4\gamma e^{-4\gamma} \int_0^1 e^{2\gamma t} L(t) \, dL(t)\right)^2 &= 16\gamma^2 e^{-6\gamma} \int_0^1 e^{2\gamma t} \mathbb{E}L^2(t) \, dt \\ &= 8\gamma e^{-4\gamma} \int_0^1 (e^{2\gamma t} - 1) \, dt \leq 4e^{-2\gamma} \rightarrow 0,\end{aligned}$$

as $\gamma \rightarrow \infty$. It follows from the Chebyshev's inequality that

$$4\gamma e^{-4\gamma} \int_0^1 e^{2\gamma t} L(t) \, dL(t) \rightarrow_p 0, \quad \text{as } \gamma \rightarrow \infty.$$

Hence we achieve Eq. (35). The proof is complete. \square

Proof of Lemma 5 We first show that Eq. (36) holds. Let

$$dB^*(t) = \frac{\tilde{\lambda}_{12}(t)\sqrt{h'(t)}d\tilde{B}_1(t) + \tilde{\lambda}_{22}(t)\sqrt{h'(t)}d\tilde{B}_2(t)}{\sqrt{\tilde{\lambda}_{12}^2(t)h'(t) + \tilde{\lambda}_{22}^2(t)h'(t)}},$$

then, by the Lévy characterization of Brownian motion, B^* is a one-dimensional standard Brownian motion. Since

$$\tilde{\lambda}_{12}^2(t)h'(t) + \tilde{\lambda}_{22}^2(t)h'(t) = e^{2\gamma(1-t)}, \quad 0 \leq t \leq 1,$$

Eq. (33) implies that

$$dL(t) = e^{\gamma(1-t)} dB^*(t).$$

Hence, by the Itô's formula, we have

$$\begin{aligned}\int_0^t \psi_\gamma(s) \, ds &= \sqrt{-2\gamma} e^{-\gamma} \int_0^t e^{\gamma s} L(s) \, ds \\ &= \frac{\sqrt{-2\gamma} e^{-\gamma}}{\gamma} \left(e^{\gamma t} L(t) - \int_0^t e^{\gamma s} \, dL(s) \right) \\ &= \frac{\sqrt{-2\gamma}}{\gamma} \left(\int_0^t e^{\gamma(t-s)} \, dB^*(s) - B^*(t) \right).\end{aligned}$$

Therefore

$$\sup_{0 \leq t \leq 1} \left| \int_0^t \psi_\gamma(s) \, ds \right| \leq \sqrt{\frac{2}{-\gamma}} \left(\sup_{0 \leq t \leq 1} \left| \int_0^t e^{\gamma(t-s)} \, dB^*(s) \right| + \sup_{0 \leq t \leq 1} |B^*(t)| \right).$$

Obviously,

$$\sqrt{\frac{2}{-\gamma}} \left(\sup_{0 \leq t \leq 1} |B^*(t)| \right) \rightarrow_p 0,$$

as $\gamma \rightarrow -\infty$ and Theorem 7 shows that

$$\mathbb{E} \left(\sqrt{\frac{2}{-\gamma}} \sup_{0 \leq t \leq 1} \left| \int_0^t e^{\gamma(t-s)} dB^*(t) \right| \right) \leq \frac{c_1 \sqrt{2 \log(1-\gamma)}}{-\gamma} \rightarrow 0,$$

as $\gamma \rightarrow -\infty$. By the Markov's inequality again, we achieve Eq. (36).

Finally, we verify Eq. (37). Notice that

$$dL^2(t) = 2L(t) dL(t) + e^{2\gamma(1-t)} dt.$$

It follows that

$$\int_0^1 \psi_\gamma^2(t) dt = -L^2(1) + 1 + 2e^{-2\gamma} \int_0^1 e^{2\gamma t} L(t) dL(t).$$

Since

$$L(t) \sim \mathcal{N} \left(0, \frac{e^{2\gamma}(1 - e^{-2\gamma t})}{2\gamma} \right),$$

we have

$$L^2(1) \rightarrow_p 0, \quad \text{as } \gamma \rightarrow -\infty.$$

Moreover,

$$\begin{aligned} \mathbb{E} \left(2e^{-2\gamma} \int_0^1 e^{2\gamma t} L(t) dL(t) \right)^2 &= 4e^{-4\gamma} \int_0^1 e^{4\gamma t} e^{2\gamma(1-t)} \mathbb{E} L^2(t) dt \\ &= \frac{2}{\gamma} \int_0^1 (e^{2\gamma t} - 1) dt \rightarrow 0, \end{aligned}$$

as $\gamma \rightarrow -\infty$. It concludes from the Chebyshev's inequality that

$$2e^{-2\gamma} \int_0^1 e^{2\gamma t} L(t) dL(t) \rightarrow_p 0.$$

Hence we obtain Eq. (37). The proof is complete. \square

Acknowledgements The authors would like to express their sincere gratitude to the anonymous referees and AE for helpful comments which surely lead to an improved presentation of this paper. The authors are very grateful to Huarui He, Hui Jiang, Feng Li, Yu Miao, Shaochen Wang and Qingshan Yang for the helpful discussions. Hailin Sang's work was partially supported by the Simons Foundation Grant 586789, USA. Guangyu Yang's work was partially supported by the Foundation of Young Scholar of the Educational Department of Henan Province Grant 2019GGJS012, China.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- Anderson, T. W. (1959). On asymptotic distributions of estimators of parameters of stochastic difference equations. *Annals of Mathematical Statistics*, 30, 676–687.
- Andrews, D. W. K., Guggenberger, P. (2008). Asymptotics for stationary very nearly unit root processes. *Journal of Time Series Analysis*, 29, 203–212.
- Aue, A., Horváth, L. (2007). A limit theorem for mildly explosive autoregression with stable errors. *Econometric Theory*, 23, 201–220.
- Bobkoski M. J. (1983). *Hypothesis testing in nonstationary time series*. Ph.D. Thesis, Department of Statistics, University of Wisconsin, Madison, Wisconsin.
- Buchmann, B., Chan, N. H. (2013). Unified asymptotic theory for nearly unstable AR(p) processes. *Stochastic Processes and their Applications*, 123, 952–985.
- Cavanagh, C. (1985). *Roots local to unity*. Manuscript, Department of Economics, Harvard University, Cambridge, Massachusetts.
- Chan, N. H. (2009). Time series with roots on or near the unit circle. In T. G. Andersen, R. A. Davis, J. Kreiss, T. Mikosch (Eds.), *Springer handbook of financial time series* (pp. 696–707). Berlin, Germany: Springer-Verlag.
- Chan, N. H., Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics*, 15, 1050–1063.
- Davis, R. A., Knight, K., Liu, J. (1992). M-estimation for autoregressions with infinite variance. *Stochastic Processes and their Applications*, 40, 145–180.
- Dickey, D. A., Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74, 427–431.
- Ethier, S. N., Kurtz, T. G. (1986). *Markov processes: Characterization and convergence*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Giraitis, L., Phillips, P. C. B. (2006). Uniform limit theory for stationary autoregression. *Journal of Time Series Analysis*, 27, 51–60.
- Graversen, S. E., Peskir, G. (2000). Maximal inequalities for the Ornstein-Uhlenbeck process. *Proceedings of the American Mathematical Society*, 128, 3035–3041.
- Hansen, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory*, 8, 489–500.
- Herce, M. A. (1996). Asymptotic theory of LAD estimation in a unit root process with finite variance errors. *Econometric Theory*, 12, 129–153.
- Jiang, H., Wan, Y. L., Yang, G. Y. (2022). Deviation inequalities and Cramer-type moderate deviations for the explosive autoregressive process. *Bernoulli*, 28, 2634–2662.
- Knight, K. (1989). Limit theory for autoregressive parameter estimates in an infinite variance random walk. *Canadian Journal of Statistics*, 17, 261–278.
- Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *Annals of Statistics*, 26, 755–770.
- Li, G. D., Li, W. K. (2009). Least absolute deviation estimation for unit root processes with GARCH errors. *Econometric Theory*, 25, 1208–1227.
- Ling, S. Q. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society: Series B*, 67, 381–393.
- Miao, Y., Shen, S. (2009). Moderate deviation principle for autoregressive processes. *Journal of Multivariate Analysis*, 100, 1952–1961.
- Miao, Y., Wang, Y. L., Yang, G. Y. (2015). Moderate deviations principle for empirical covariance from a unit root. *Scandinavian Journal of Statistics*, 42, 234–255.
- Müller, U. K., Elliott, G. (2003). Tests for unit roots and the initial condition. *Econometrica*, 71, 1269–1286.
- Øksendal, B. (2005). *Stochastic differential equation: An introduction with applications* (6th ed.). Berlin Heidelberg: Springer-Verlag.
- Phillips, P. C. B. (1987). Toward a unified asymptotic theory of autoregression. *Biometrika*, 74, 535–574.
- Phillips, P. C. B. (1988). Regression theory for near-integrated time series. *Econometrica*, 56, 1021–1043.
- Phillips, P. C. B. (1991). A shortcut to LAD estimator asymptotics. *Econometric Theory*, 7, 450–463.
- Phillips P. C. B. (2021). Estimation and inference with near unit roots. *Cowles Foundation Discussion Papers*. 2654. <https://elischolar.library.yale.edu/cowles-discussion-paper-series/2654>.

- Phillips, P. C. B., Durlauf, S. N. (1986). Multiple time series regressions with integrated processes. *Review of Economic Studies*, 53, 473–495.
- Phillips, P. C. B., Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, 136, 115–130.
- Phillips, P. C. B., Magdalinos, T. (2009). Unit root and cointegrating limit theory when initialization is in the infinite past. *Econometric Theory*, 25, 1682–1715.
- Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, 7, 186–199.
- Rootzén, H. (1980). Limit distributions for the error in approximations of stochastic integrals. *Annals of Probability*, 8, 244–251.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. New York: Chapman and Hall.
- Stock, J. M. (1991). Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. *Journal of Monetary Economics*, 28, 435–459.
- Tanaka, K. (2017). *Time series analysis: nonstationary and noninvertible distribution theory* (2nd ed.). Hoboken, New Jersey: Wiley.
- Wang, X. H., Wang, H. L., Wang, H. R., Hu, S. H. (2020). Asymptotic inference of least absolute deviation estimation for AR(1) processes. *Communications in Statistics-Theory and Methods*, 49, 809–826.
- White, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics*, 29, 1188–1197.
- Zhou, Z. Y., Lin, Z. Y. (2014). Asymptotic theory for LAD estimation of moderate deviation from a unit root. *Statistics and Probability Letters*, 90, 25–32.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.