Appl. Math. J. Chinese Univ.
2020, 35(3): 265-280

Asymptotic inference for AR(1) panel data

SHEN Jian-fei PANG Tian-xiao

Abstract. A general asymptotic theory is given for the panel data AR(1) model with time series independent in different cross sections. The theory covers the cases of stationary process, local to unity process, unit root process, mildly integrated, mildly explosive and explosive processes. It is assumed that the cross-sectional dimension and time-series dimension are respectively N and T. The results in this paper illustrate that whichever the process is, with an appropriate regularization, the least squares estimator of the autoregressive coefficient converges in distribution to a normal distribution with rate at least $O(N^{-1/3})$. Since the variance is the key to characterize the normal distribution, it is important to discuss the variance of the least squares estimator. We will show that when the autoregressive coefficient ρ satisfies $|\rho| < 1$, the variance declines at the rate $O((NT)^{-1})$, while the rate changes to $O(N^{-1}T^{-2})$ when $\rho = 1$ and $O(N^{-1}\rho^{-2T+4})$ when $|\rho| > 1$. $\rho = 1$ is the critical point where the convergence rate changes radically. The transition process is studied by assuming ρ depending on T and going to 1. An interesting phenomenon discovered in this paper is that, in the explosive case, the least squares estimator of the autoregressive coefficient has a standard normal limiting distribution in the panel data case while it may not has a limiting distribution in the univariate time series case.

§1 Introduction

Dynamic models are useful in modeling time series data and have been well studied in the past few decades. One of the dynamic models is the AR(1) model which is given by

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad t = 1, 2, ..., T. \tag{1.1}$$

We assume that $\{\varepsilon_t, t \geq 1\}$ are independent and identically distributed (i.i.d.) random variables with $E[\varepsilon_1] = 0$ and $E[\varepsilon_1^2] = 1$.

Although the model (1.1) is simple, it is very useful and important in time series and econometrics literature since the model can be used to model some kinds of stationary or non-stationary time series data. The parameter ρ is the main concern in the model (1.1) since

Received: 2016-08-28. Revised: 2019-01-06.
MR Subject Classification: 62E20.
Keywords: AR(1) model, Least squares estimator, Limiting distribution, Non-stationray, Panel data.
Digital Object Identifier(DOI): https://doi.org/10.1007/s11766-020-3491-x.
Supported by the National Natural Science Foundation of China (11871425), Zhejiang Provincial Natural Science Foundation of China (LY19A010022) and the Department of Education of Zhejiang Province (N20140202).

whether the model is stationary is determined by the value of ρ. It is well-known that the necessary and sufficient condition for the stationarity of y_t in (1.1) is $|\rho| < 1$ when y_0 is an appropriate random variable. The least squares estimator (LSE) of ρ is given by

$$\hat{\rho} = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}. \tag{1.2}$$

For the stationary AR(1) model, Mann and Wald (1943) proved that, if $y_0 = O_P(1)$, then

$$\frac{\sqrt{T}}{\sqrt{1-\rho^2}}(\hat{\rho} - \rho) \xrightarrow{\text{d}} N(0,1).$$

When $|\rho| > 1$, model (1.1) is non-stationary and is called the explosive AR(1) model. For this model, Anderson (1959) showed that if $y_0 = 0$ and ε_t's are independent and normal distributed random variables, then

$$\frac{|\rho|^T}{\rho^2 - 1}(\hat{\rho} - \rho) \xrightarrow{\text{d}} C,$$

where C is a standard Cauchy variate. However, for general ε_t's, Anderson (1959) showed that the limiting distribution of $\hat{\rho}$ may not exist. The interesting case is $\rho = 1$, the corresponding AR(1) model is called the unit root model in econometrics. For this model, the central limit theorem is no longer applicable when exploring the limiting distribution of $\hat{\rho}$. Instead, by applying the functional central limit theorem, White (1958) and Rao (1978) showed that, if $y_0 = o_P(\sqrt{T})$, then

$$T(\hat{\rho} - \rho) \xrightarrow{\text{d}} \frac{\frac{1}{2}\left[W^2(1) - 1\right]}{\int_0^1 W^2(t)dt},$$

where $\{W(t), 0 \le t \le 1\}$ is a standard Wiener process. This limiting distribution is not standard. Noting that $P(W^2(1) \le 1) \approx 0.684$, the limiting distribution is not even symmetric.

In order to bridge the gaps of asymptotic theories between the stationary AR(1) model and the unit root model, Chan and Wei (1987) and Phillips (1987) independently studied the following model which is called nearly non-stationary AR(1) model:

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad y_0 = o_P(\sqrt{T}), \quad \rho = \rho_T = 1 - c/T, \quad t = 1, 2, ..., T, \tag{1.3}$$

where c is a fixed constant. Of late, in order to bridge the gaps of asymptotic theories between the unit root model and the explosive AR(1) model, Phillips and Magdalinos (2007) studied the following AR(1) model:

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad y_0 = o_P(\sqrt{k_T}), \quad \rho = \rho_T = 1 - c/k_T, \quad t = 1, 2, ..., T, \tag{1.4}$$

where c is a non-zero constant and k_T is a sequence of positive constants increasing to ∞ such that $k_T = o(T)$. Model (1.4) with $c > 0$ and with $c < 0$ is called mildly integrated AR(1) model and mildly explosive AR(1) model respectively according to Phillips and Magdalinos (2007).

In models (1.3) and (1.4), we denote $\hat{\rho}_T$ to be the LSE of ρ_T, and also suppose that $\{\varepsilon_t, t \ge 1\}$ are i.i.d. random variables with $E[\varepsilon_1] = 0$ and $E[\varepsilon_1^2] = 1$. It is worth noting that the limiting distributions of $\hat{\rho}_T$ are different from those in the stationary AR(1) model, unit root model and explosive model. Specifically, Chan and Wei (1987) proved that when $\rho = \rho_T = 1 - c/T$ with

$c \in R$,

$$T(\hat{\rho}_T - \rho_T) \xrightarrow{\mathrm{d}} \frac{2c}{b} \frac{\int_0^1 (1+bt)^{-1} W(t) dW(t)}{\int_0^1 (1+bt)^{-2} W^2(t) dt},$$

where $b = e^{2c} - 1$ ($\frac{2c}{b}$ in the above limiting distribution is replaced by 1 if $c = 0$), while Phillips and Magdalinos (2007) proved that when $\rho = \rho_T = 1 - c/k_T$ with $c > 0$,

$$\sqrt{Tk_T}(\hat{\rho}_T - \rho_T) \xrightarrow{\mathrm{d}} N(0, 2c),$$

and when $\rho = \rho_T = 1 - c/k_T$ with $c < 0$,

$$[k_T \rho_T^T/(-2c)](\hat{\rho}_T - \rho_T) \xrightarrow{\mathrm{d}} C,$$

where, as before, C stands for a standard Cauchy variate.

It is clear that the limiting distribution of the LSE of ρ varies in AR(1) models under different assumptions on ρ. Further, one can find that the limiting distribution is not standard in nearly non-stationary AR(1) model which includes the unit root model as a special case. This is harmful for making further statistical inferences, for example, confidence intervals of ρ.

However, with the panel data, the results may be extremely simple. A panel data set is the one that follows a given sample of individuals over time, and thus provides multiple observations on each individual in the sample. A panel data AR(1) model is formulated by

$$y_{it} = \rho y_{i,t-1} + \varepsilon_{it}, \quad t = 1, 2, ..., T, \quad i = 1, 2, ..., N, \tag{1.5}$$

where $\{\varepsilon_{it}, i \geq 1, t \geq 1\}$ are i.i.d. random variables with $E[\varepsilon_{11}] = 0$ and $E[\varepsilon_{11}^2] = 1$. The dimension of individual, N, is usually called cross-sectional dimension. There is no common effect on individuals in the model (1.5). Thus each individual generates an independent time series and the central limit theorem may be applied to cross-sectional dimension. There are many papers studying on dynamic panel data models with individual specific effects or/and time specific effects in the literature, for example, see Moon and Phillips (2000), Hahn and Moon (2006) and Lu and Su (2016). However, we focus on the most simple dynamic panel data model in this paper. For model (1.5), Levin and Lin (1992) proved that, when $\rho = 1$ (unit root case), $y_{i0} = 0$ for all $i \geq 1$ and an additional moment condition is fulfilled, that is, $E|\varepsilon_{11}|^{2+\lambda} < \infty$ for some $\lambda > 0$, then it is true that

$$\sqrt{N}T(\hat{\rho} - \rho) \xrightarrow{\mathrm{d}} N(0, 2), \quad N, T \to \infty. \tag{1.6}$$

Here and in what follows, $N, T \to \infty$ means $T \to \infty$ followed by $N \to \infty$; see Phillips and Moon (1999). Obviously, the limiting distribution of $\hat{\rho}$ in panel data unit root model is simpler than that in univariate time series unit root model. What is more important is the former is standard while the latter is not. This comparison motivates us to study other panel data AR(1) models.

Therefore, the aim of this paper is to study the limiting distribution of the LSE of ρ in various panel data AR(1) models. We are interested in the following question: whether, like the panel data unit root case, all the limiting distributions are normal in stationary, nearly non-stationary, mildly integrated, mildly explosive and explosive panel data.

The rest of the paper is organized as follows. We will extend the conclusion (1.6) to general cases for $\rho \in R$ in Section 2, and provide some applications in Section 3. Note that, in Section 3, all the limiting distributions have the form of normal distribution only with different rates of convergence. When $\rho = 1$, our result coincides with that in Levin and Lin (1992), but the moment condition $E[|\varepsilon_{11}|^{2+\lambda}] < \infty$ for some $\lambda > 0$ is replaced by a more weaker one, that is, $E[\varepsilon_{11}^2] < \infty$, in our paper.

§2 Asymptotics for the LSE of ρ

Consider the panel data AR(1) model:

$$y_{it} = \rho y_{i,t-1} + \varepsilon_{it}, \quad t = 1, 2, ..., T, \quad i = 1, 2, ..., N, \tag{2.1}$$

where the innovations $\{\varepsilon_{it}, i \geq 1, t \geq 1\}$ are i.i.d. random variables with $E[\varepsilon_{11}] = 0$ and $E[\varepsilon_{11}^2] = 1$. In this model, the LSE of ρ is

$$\hat{\rho} = \frac{\sum_{i=1}^N \sum_{t=1}^T y_{it} y_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T y_{i,t-1}^2}. \tag{2.2}$$

It is true that

$$\hat{\rho} - \rho = \frac{\sum_{i=1}^N \sum_{t=1}^T y_{i,t-1}\varepsilon_{it}}{\sum_{i=1}^N \sum_{t=1}^T y_{i,t-1}^2}. \tag{2.3}$$

To obtain a non-degenerated limiting distribution for (2.3), we can apply the central limit theorem to the numerator and the law of large numbers to the denominator, respectively. Before doing so, we need to put the normalizing constants on $\sum_{t=1}^T y_{i,t-1}\varepsilon_{it}$'s and $\sum_{t=1}^T y_{i,t-1}^2$'s such that they become bounded in probability. The following is our main result in this section.

Theorem 2.1. *In the model (2.1), we suppose the innovations $\{\varepsilon_{it}, i \geq 1, t \geq 1\}$ are i.i.d. random variables with $E[\varepsilon_{11}] = 0$ and $E[\varepsilon_{11}^2] = 1$. In addition, we assume there exist two positive functions of T, $Q(T)$ and $P(T)$, such that*

$$A_i^T := P(T) \sum_{t=1}^T y_{i,t-1}\varepsilon_{it} \xrightarrow{d} A_i, \quad T \to \infty,$$

and

$$B_i^T := Q(T) \sum_{t=1}^T y_{i,t-1}^2 \xrightarrow{d} B_i, \quad T \to \infty,$$

where A_i's and B_i's are random variables.
(1) If, as $T \to \infty$, $E[(A_i^T)^r] \to E[A_i^r]$ for $r = 1, 2$ and $E[B_i^T] \to E[B_i]$ with $0 < E[A_i^2] < \infty$ and $0 < E[B_i] < \infty$ for all $i \geq 1$, then we have

$$\sqrt{N}\frac{P(T)}{Q(T)}(\hat{\rho} - \rho) \xrightarrow{d} N\left(0, \frac{Var(A_1)}{(E[B_1])^2}\right), \quad N, T \to \infty. \tag{2.4}$$

(2) If the conditions in (1) are fulfilled, and in addition, as $T \to \infty$, $E[(B_i^T)^2] \to E[B_i^2] < \infty$ and $E[|A_i^T|^3] \to E[|A_i|^3] < \infty$ for all $i \geq 1$, then we have, as long as T is large enough, $\sqrt{N}\frac{P(T)}{Q(T)}(\hat{\rho} - \rho)$ converges to a normal random variable in distribution with the rate at least $O(N^{-\frac{1}{3}})$ as $N \to \infty$.

Remark 2.1. *We assume the cross section dimension N and the time series dimension T are independent in this paper. However, if N depends on T and is a monotonic function of T, one could extend the results in this paper via some limit theorems for triangular arrays (for example, central limit theorem for triangular arrays in Levin and Lin (1992) and the law of large numbers for triangular arrays in Sung (1999)).*

Remark 2.2. *In this paper, we assume that ε_{it}'s are independent across both i and t for ease of exposition. However, the results in Section 2 and Section 3 below can possibly be generalized when ε_{it}'s are weakly dependent over i or over t. For example, if $\varepsilon_{it} = \sum_{j=0}^{\infty} \phi_{ij} u_{i,t-j}$ with u_{ij}'s being i.i.d random variables with zero mean and finite variance, and $\sum_{j=0}^{\infty} j|\phi_{ij}| < \infty$ for all $i \geq 1$, then our results in this paper can be generalized easily by BN decomposition for linear processes. Moreover, If ε_{it}'s are weakly dependent over i for any fixed $t \geq 1$ such that $\{A_i^T, i \geq 1\}$ are also weakly dependent, then our results in this paper can also be generalized.*

Proof. (1) Apparently, $\{A_i^T, i \geq 1\}$ are i.i.d. random variables with $E[A_i^T] = 0$. Moreover, it follows from the conditions of moment convergence that there exists some $T_0 > 0$ such that when $T > T_0$, $0 < E[(A_i^T)^2] < \infty$. Denote

$$S_N^T = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{A_i^T}{\sqrt{Var(A_1^T)}}. \tag{2.5}$$

Note that $E\left[\frac{A_i^T}{\sqrt{Var(A_1^T)}}\right] = 0$ and $Var\left(\frac{A_i^T}{\sqrt{Var(A_1^T)}}\right) = 1$. Hence, when $T > T_0$, applying the central limit theorem for i.i.d. random variables with zero mean and finite second moment leads to

$$S_N^T \xrightarrow{\text{d}} N(0,1), \quad N \to \infty, \tag{2.6}$$

which, in view of the characteristic function arguments, further implies that

$$S_N^T \xrightarrow{\text{d}} N(0,1), \quad N, T \to \infty. \tag{2.7}$$

In addition, noting that $\{B_i^T, i \geq 1\}$ are also i.i.d. random variables and there exists some $T_1 > 0$ such that $E[B_i^T] < \infty$ when $T > T_1$ by the conditions of moment convergence, it follows from the law of large numbers that when $T > T_1$,

$$\frac{1}{N} \sum_{i=1}^{N} B_i^T \xrightarrow{\text{P}} E[B_1^T], \quad N \to \infty. \tag{2.8}$$

This easily yields

$$\frac{1}{N} \sum_{i=1}^{N} B_i^T \xrightarrow{\text{P}} E[B_1], \quad N, T \to \infty. \tag{2.9}$$

Combining (2.7) with (2.9) immediately leads to (2.4) by observing the following equality

$$\sqrt{N} \frac{P(T)}{Q(T)} (\hat{\rho} - \rho) = \sqrt{N} \frac{\sum_{i=1}^{N} A_i^T}{\sum_{i=1}^{N} B_i^T} = S_N^T \frac{\sqrt{Var(A_1^T)}}{\frac{1}{N} \sum_{i=1}^{N} B_i^T}.$$

(2) It follows from the conditions of moment convergence that there exists some $T_2 > 0$ such that, when $T > T_2$, $E[|A_i^T|]^3 < \infty$ and (2.6) is still true. Denote

$$\gamma_T = E[|A_i^T|^3], \quad \sigma_T^2 = E[(A_i^T)^2].$$

Then, according to the well-known Berry-Esseen bound for i.i.d. random variables with finite third moment, the rate of convergence for (2.6), when $T > T_2$, is characterized by the following inequality:

$$\sup_{x\in R} |P(S_N^T \le x) - \Phi(x)| \le \frac{c_0\gamma_T}{\sigma_T^3\sqrt{N}}, \tag{2.10}$$

where c_0 is some positive constant and $\Phi(x)$ is the distribution function of a standard normal random variable. In addition, by virtue of the conditions of moment convergence again, there esixts some $T_3 > 0$ such that $E[B_i^T] > 0$ and $E[(B_i^T)^2] < \infty$ when $T > T_3$. Denote

$$R_N^T = \frac{\frac{1}{N}\sum_{i=1}^N B_i^T}{E[B_1^T]}.$$

Note that R_N^T is a non-negative random variable and $E[R_N^T] = 1$. By applying Chebyshev's inequality, we have for any $0 < \delta < \frac{1}{2}$,

$$P(|R_N^T - 1| \ge \delta) \le \frac{Var(R_N^T)}{\delta^2} = \frac{1}{N\delta^2}\frac{Var(B_1^T)}{(E[B_1^T])^2}. \tag{2.11}$$

Next, we will explore the rate of convergence of S_N^T/R_N^T for $T > \max\{T_2, T_3\}$.

First, when $x \ge 0$, one has

$$\begin{aligned}
& \sup_{x\ge 0}\left|P\left(\frac{S_N^T}{R_N^T} < x\right) - \Phi(x)\right| \\
= \ & \sup_{x\ge 0}\left|P\left(\frac{S_N^T}{R_N^T} < x, |R_N^T - 1| < \delta\right) + P\left(\frac{S_N^T}{R_N^T} < x, |R_N^T - 1| \ge \delta\right) - \Phi(x)\right| \\
\le \ & \sup_{x\ge 0}\left|P\left(S_N^T < R_N^T x, |R_N^T - 1| < \delta\right) - \Phi(x)\right| + P\left(|R_N^T - 1| \ge \delta\right) \\
\le \ & \sup_{x\ge 0}\max\left\{P\left(S_N^T < (1+\delta)x\right) - \Phi(x), \Phi(x) - P\left(S_N^T < (1-\delta)x, |R_N^T - 1| < \delta\right)\right\} \\
& +P\left(|R_N^T - 1| \ge \delta\right) \\
\le \ & \sup_{x\ge 0}\max\left\{P\left(S_N^T < (1+\delta)x\right) - \Phi(x), \Phi(x) - P\left(S_N^T < (1-\delta)x\right) + P(|R_N^T - 1| \ge \delta)\right\} \\
& +P\left(|R_N^T - 1| \ge \delta\right) \\
\le \ & \max\left\{\sup_{x\ge 0}\left|P\left(S_N^T < (1+\delta)x\right) - \Phi((1+\delta)x)\right| + \sup_{x\ge 0}|\Phi((1+\delta)x) - \Phi(x)|\right. \\
& +P\left(|R_N^T - 1| \ge \delta\right), \sup_{x\ge 0}\left|P\left(S_N^T < (1-\delta)x\right) - \Phi((1-\delta)x)\right| + \sup_{x\ge 0}|\Phi((1-\delta)x) - \Phi(x)| \\
& \left. +2P\left(|R_N^T - 1| \ge \delta\right)\right\}.
\end{aligned} \tag{2.12}$$

Note that $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$ satisfies the following smooth conditions:

$$\sup_{x\ge 0}|\Phi((1+\delta)x) - \Phi(x)| = \sup_{x\ge 0}\int_x^{(1+\delta)x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt \le \sup_{x\ge 0}\frac{\delta}{\sqrt{2\pi}}xe^{-\frac{x^2}{2}} \le \frac{\delta}{\sqrt{2\pi e}}, \tag{2.13}$$

and similarly,

$$\sup_{x\ge 0}|\Phi(x) - \Phi((1-\delta)x)| \le \frac{1}{1-\delta}\frac{\delta}{\sqrt{2\pi e}} < \frac{2\delta}{\sqrt{2\pi e}}. \tag{2.14}$$

Substituting (2.10), (2.11), (2.13) and (2.14) into (2.12) and taking $\delta = N^{-\frac{1}{3}}$ $(N > 8)$, one has

$$\begin{aligned}\sup_{x\geq 0}\left|P\left(\frac{S_N^T}{R_N^T}<x\right)-\Phi(x)\right| &\leq \frac{c_0\gamma_T^3}{\sigma_T^3\sqrt{N}}+\frac{2}{N\delta^2}\frac{Var(B_1^T)}{(E[B_1^T])^2}+\frac{2\delta}{\sqrt{2\pi e}}\\ &=: C_1(T)N^{-\frac{1}{2}}+C_2(T)N^{-\frac{1}{3}}, \qquad (2.15)\end{aligned}$$

where $C_1(T) = \frac{c_0\gamma_T^3}{\sigma_T^3}$ and $C_2(T) = \frac{2Var(B_1^T)}{(E[B_1^T])^2} + \frac{2}{\sqrt{2\pi e}}$. Note that both $C_1(T)$ and $C_2(T)$ are bounded when $T > \max\{T_2, T_3\}$.

By the same arguments, when $x < 0$, one has

$$\begin{aligned}&\sup_{x<0}\left|P\left(\frac{S_N^T}{R_N^T}<x\right)-\Phi(x)\right|\\ \leq\ & \max\Big\{\sup_{x<0}\left|P\left(S_N^T<(1-\delta)x\right)-\Phi((1-\delta)x)\right|+\sup_{x<0}|\Phi((1-\delta)x)-\Phi(x)|\\ &+P\left(|R_N^T-1|\geq\delta\right),\sup_{x<0}\left|P\left(S_N^T<(1+\delta)x\right)-\Phi((1+\delta)x)\right|+\sup_{x<0}|\Phi((1+\delta)x)-\Phi(x)|\\ &+2P\left(|R_N^T-1|\geq\delta\right)\Big\}\\ \leq\ & C_1(T)N^{-\frac{1}{2}}+C_2(T)N^{-\frac{1}{3}}. \qquad (2.16)\end{aligned}$$

Thus we can unify (2.15) and (2.16) as

$$\sup_{x\in R}\left|P\left(\frac{S_N^T}{R_N^T}<x\right)-\Phi(x)\right|\leq C_1(T)N^{-\frac{1}{2}}+C_2(T)N^{-\frac{1}{3}}, \qquad (2.17)$$

where both $C_1(T)$ and $C_2(T)$ are bounded when $T > \max\{T_2, T_3\}$.

Noting that

$$\sqrt{N}\frac{P(T)}{Q(T)}(\hat{\rho}-\rho)=\frac{S_N^T}{R_N^T}\cdot\frac{\sqrt{Var(A_1^T)}}{E[B_1^T]}, \qquad (2.18)$$

it is true that $\sqrt{N}\frac{P(T)}{Q(T)}(\hat{\rho}-\rho)$ also converges to a standard normal random variable in distribution with rate at least $O(N^{-\frac{1}{3}})$ as long as T is large enough. □

Remark 2.3. *Generally, the requirements of* $0 < E[A_i^2] < \infty$, $0 < E[B_i] < \infty$ *and convergence of moments are not strong. They can be fulfilled in most of models we will discuss below.*

Theorem 2.1 illustrates that N determines the form of the limiting distribution while T portrays the speed of convergence (with $P(T)$ and $Q(T)$). Considering the limiting distribution is normal with zero mean, it can be totally depicted by its variance. So the rest of this paper focuses on studying the variance of the limiting distribution in various cases.

§3 Applications

In this section, the limiting distribution of the LSE of ρ in model (2.1) will be introduced one by one whenever ρ is a fixed constant or a constant depending on T.

The results in the following lemma are taken from Mann and Wald (1943) and Rao (1978), respectively.

Lemma 3.1. *In the model (1.1), we suppose that the innovations* $\{\varepsilon_t, t \geq 1\}$ *are i.i.d. random variables with* $E[\varepsilon_1] = 0$ *and* $E[\varepsilon_1^2] = 1$. *Then,* *(1) When $|\rho| < 1$ and $y_0 = O_P(1)$, one has*

$$\sqrt{\frac{1-\rho^2}{T}} \sum_{t=1}^{T} y_{t-1}\varepsilon_t \xrightarrow{d} N(0,1), \quad T \to \infty,$$

$$\frac{1-\rho^2}{T} \sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{P} 1, \quad T \to \infty;$$

(2) When $\rho = 1$ and $y_0 = o_P(\sqrt{T})$, one has

$$\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t, \frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2\right) \xrightarrow{d} \left(\frac{1}{2}(W(1)^2-1), \int_0^1 W^2(t)dt\right), \quad T \to \infty,$$

where $\{W(t), 0 \le t \le 1\}$ is a standard Wiener process.

Note that the LSE of ρ in model (2.1) is (2.2). With Theorem 2.1 and Lemma 3.1, the following results can be obtained.

Theorem 3.1. *In the model (2.1), we suppose the innovations $\{\varepsilon_{it}, i \ge 1, t \ge 1\}$ are i.i.d. random variables with $E[\varepsilon_{11}] = 0$ and $E[\varepsilon_{11}^2] = 1$. Then,*

(1) When $|\rho| < 1$ and $y_{i0} = O_P(1)$ for all $i \ge 1$, one has

$$\frac{\sqrt{NT}}{\sqrt{1-\rho^2}}(\hat{\rho} - \rho) \xrightarrow{d} N(0,1), \quad N, T \to \infty; \tag{3.1}$$

(2) When $\rho = 1$ and $y_{i0} = o_P(\sqrt{T})$ for all $i \ge 1$, one has

$$\sqrt{N}T(\hat{\rho} - \rho) \xrightarrow{d} N(0,2), \quad N, T \to \infty. \tag{3.2}$$

Proof. It is easy to see that the values of y_{i0} do not affect the limiting distribution of $\hat{\rho}$ as long as the assumptions on y_{i0} in (1) and (2) are satisfied. Hence, without loss of generality, we assume $y_{i0} = 0$ for all $i \ge 1$. The proof of (1) is easy and thus omitted. (2) is true because for any $i \ge 1$,

$$E\left[\frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_t\right] = 0, \quad E\left[\left(\frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_t\right)^2\right] = \frac{1}{T^2}\sum_{t=1}^{T}(t-1) \to \frac{1}{2}, \quad T \to \infty,$$

$$E\left[\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2\right] = \frac{1}{T^2}\sum_{t=1}^{T}(t-1) \to \frac{1}{2}, \quad T \to \infty,$$

$$E\left[\frac{1}{2}(W(1)^2-1)\right] = 0, \quad Var\Big(\frac{1}{2}(W(1)^2-1)\Big) = E\left[\Big(\frac{1}{2}(W(1)^2-1)\Big)^2\right] = \frac{1}{4} \times (3-1) = \frac{1}{2}$$

and

$$E\Big[\int_0^1 W^2(t)dt\Big] = \int_0^1 tdt = \frac{1}{2}.$$

□

Remark 3.1. *The result (2) in Theorem 3.1 is indeed one of the main results in Levin and Lin (1992), but the moment conditions in this paper are weaker than those in Levin and Lin (1992).*

Remark 3.2. *In Lemma 3.1, the case of $|\rho| > 1$ is excluded. Anderson (1959) proved that, if $y_0 = 0$ and $\{\varepsilon_t, t \geq 1\}$ are i.i.d. normal random variables with mean zeros and variance ones, then*

$$\rho^{-(T-2)} \sum_{t=1}^{T} y_{t-1}\varepsilon_t \xrightarrow{d} \xi\eta, \quad T \to \infty,$$

$$(\rho^2 - 1)\rho^{-2(T-1)} \sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{d} \xi^2, \quad T \to \infty,$$

$$\frac{\rho^T}{\rho^2 - 1}(\hat{\rho} - \rho) \xrightarrow{d} C, \quad T \to \infty,$$

where, ξ and η are independent and obey $N(0, \rho^2/(\rho^2 - 1))$, and C stands for a standard Cauchy variate. In general case, $\hat{\rho} - \rho$ may not has a limiting distribution. Consequently, the case of $|\rho| > 1$ is also excluded in Theorem 3.1.

Next, we will study the case of $|\rho| > 1$ in panel data AR(1) model without the help of Theorem 2.1.

Theorem 3.2. *In the model (2.1) with $|\rho| > 1$, we suppose that $y_{i0} = 0$ for all $i \geq 1$ and the innovations $\{\varepsilon_{it}, i \geq 1, t \geq 0\}$ are i.i.d. random variables with $E[\varepsilon_{11}] = 0$ and $E[\varepsilon_{11}^2] = 1$. Then*

$$\sqrt{N}\rho^{T-2}(\hat{\rho} - \rho) \xrightarrow{d} N(0, 1), \quad N, T \to \infty. \tag{3.3}$$

Proof. Denote $\beta = 1/\rho$, and

$$u_{iT} = \varepsilon_{i1} + \beta\varepsilon_{i2} + \cdots + \beta^{T-2}\varepsilon_{i,T-1}, \quad i \geq 1$$

$$v_{iT} = \varepsilon_{iT} + \beta\varepsilon_{i,T-1} \cdots + \beta^{T-2}\varepsilon_{i2} + \beta^{T-1}\varepsilon_{i1}, \quad i \geq 1.$$

Then, following the proofs of Theorem 2.1 and Theorem 2.2 in Anderson (1959), one has, as $N, T \to \infty$,

$$\frac{1}{N}\left|\beta^{T-2}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_{it} - \sum_{i=1}^{N} u_{iT}v_{iT}\right| \xrightarrow{\mathrm{P}} 0,$$

$$\frac{1}{N}\left|\beta^{2(T-2)}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}^2 - \sum_{i=1}^{N} u_{iT}^2\right| \xrightarrow{\mathrm{P}} 0.$$

As a result,

$$\begin{aligned}
\sqrt{N}\rho^{T-2}(\hat{\rho} - \rho) &= \sqrt{N}\rho^{T-2}\frac{\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}^2} \\
&= \sqrt{N}\frac{\beta^{T-2}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_{it}}{\beta^{2(T-2)}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}^2} \\
&= \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_{iT}v_{iT}}{\frac{1}{N}\sum_{i=1}^{N} u_{iT}^2}(1 + o_P(1)).
\end{aligned} \tag{3.4}$$

That is to say, we only need to derive the limiting distribution of $\frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_{iT}v_{iT}}{\frac{1}{N}\sum_{i=1}^{N} u_{iT}^2}$ in order to derive the limiting distribution of $\sqrt{N}\rho^{T-2}(\hat{\rho} - \rho)$.

First, for any fixed $T \geq 2$, it follows from the law of large numbers that

$$\frac{1}{N}\sum_{i=1}^{N} u_{iT}^2 \xrightarrow{\mathrm{P}} \frac{1-\beta^{2(T-1)}}{1-\beta^2}, \quad N \to \infty, \tag{3.5}$$

which yields

$$\frac{1}{N}\sum_{i=1}^{N} u_{iT}^2 \xrightarrow{\mathrm{P}} \frac{1}{1-\beta^2}, \quad N, T \to \infty. \tag{3.6}$$

Second, denote

$$u_{iT}^* = \sum_{t=1}^{[T/2]} \beta^{t-1}\varepsilon_{it}, \quad v_{iT}^* = \sum_{t=[T/2]+1}^{T} \beta^{T-t}\varepsilon_{it},$$

here the symbol $[x]$ denote the largest integer not greater than x. Then, by the proof of Theorem 2.3 in Anderson (1959), we have for any $i \geq 1$,

$$|u_{iT} - u_{iT}^*| \xrightarrow{\mathrm{P}} 0 \text{ and } |v_{iT} - v_{iT}^*| \xrightarrow{\mathrm{P}} 0, \quad T \to \infty.$$

It follows that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_{iT}v_{iT} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_{iT}^* v_{iT}^*(1+o_P(1)). \tag{3.7}$$

Note that the sequences $\{u_{iT}^*, i \geq 1\}$ and $\{v_{iT}^*, i \geq 1\}$ are independent for any fixed $T \geq 2$. Then, by virtue of the central limit theorem we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_{iT}^* v_{iT}^* \xrightarrow{\mathrm{d}} N(0, \frac{(1-\beta^{2[T/2]})(1-\beta^{2(T-[T/2])})}{(1-\beta^2)^2}), \quad N \to \infty, \tag{3.8}$$

which further implies that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} u_{iT}^* v_{iT}^* \xrightarrow{\mathrm{d}} N(0, \frac{1}{(1-\beta^2)^2}), \quad N, T \to \infty \tag{3.9}$$

by characteristic function arguments. Now, combining (3.4), (3.6), (3.7) with (3.9) yields (3.3). □

Remark 3.3. *It is interesting to see that $\hat{\rho}$ has a limiting distribution in panel data case while it may not has a limiting distribution in univariate time series case.*

Though in panel data, the form of limiting distribution is stable, noticing that the scale of $\hat{\rho} - \rho$ declines from $O(\frac{1}{\sqrt{NT}})$ when $|\rho| < 1$, to $O(\frac{1}{\sqrt{N}T})$ when $\rho = 1$ and to $O(\frac{1}{\sqrt{N}\rho^{T-2}})$ when $|\rho| > 1$, the rate of convergence changes radically at $\rho = 1$. Hence, it is necessary to discuss the case when ρ is near 1. In the rest of the paper, we suppose that ρ depends on T, so it is natural to use the notation $\hat{\rho}_T$ to denote the LSE of ρ, that is, (2.2).

We first follow the proposal of Chan and Wei (1987) and Phillips (1987) to study the case where $\rho = \rho_T = 1 - \frac{c}{T}$ with c a fixed constant. Consider the model

$$y_{it} = \rho_T y_{i,t-1} + \varepsilon_{it}, \quad \rho_T = 1 - \frac{c}{T}, \quad t = 1, 2, ..., T, \quad i = 1, 2, ..., N,$$

where $y_{i0} = o_P(\sqrt{T})$ for all $i \geq 1$ and $\{\varepsilon_{it}, i \geq 1, t \geq 1\}$ are i.i.d. random variables with $E[\varepsilon_{11}] = 0$ and $E[\varepsilon_{11}^2] = 1$.

The following lemma is not explicitly formulated in Chan and Wei (1987), but can be easily obtained by the proofs in Chan and Wei (1987), technique of change of variable and Itó formula. Thus, the details are omitted here.

Lemma 3.2. *Let $\rho_T = 1 - \frac{c}{T}$, where $c \neq 0$ is a fixed constant. Suppose that y_t comes from the following reparameterized AR(1) model,*

$$y_t = \rho_T y_{t-1} + \varepsilon_t, \quad t = 1, 2, ...T,$$

where $y_0 = o_P(\sqrt{T})$ and $\{\varepsilon_t, t \geq 1\}$ are i.i.d. random variables with $E[\varepsilon_1] = 0$ and $E[\varepsilon_1^2] = 1$. Then

$$T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \xrightarrow{d} \frac{b}{2c}\int_0^1 (1+bt)^{-1}W(t)dW(t), \quad T \to \infty, \tag{3.10}$$

$$T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{d} \left(\frac{b}{2c}\right)^2 \int_0^1 (1+bt)^{-2}W^2(t)dt, \quad T \to \infty, \tag{3.11}$$

where $b = e^{2c} - 1$ and $\{W(t), 0 \leq t \leq 1\}$ is a standard Wiener process.

With the help of the above lemma and Theorem 2.1, we have the following result.

Theorem 3.3. *Let $\rho = \rho_T = 1 - \frac{c}{T}$, where $c \neq 0$ is a fixed constant. For $t = 1, 2, ..., T$ and $i = 1, 2, ..., N$, we suppose y_{it} satisfies the following reparameterized AR(1) model,*

$$y_{it} = \rho_T y_{i,t-1} + \varepsilon_{it}, \quad i = 1, 2, ..., N, \quad t = 1, 2, ...T,$$

where $y_{i0} = o_P(\sqrt{T})$ for all $i \geq 1$ and $\{\varepsilon_{it}, i \geq 1, t \geq 1\}$ are i.i.d random variables with $E[\varepsilon_{11}] = 0$ and $E[\varepsilon_{11}^2] = 1$. Then

$$\sqrt{N}T(\hat{\rho}_T - \rho_T) \xrightarrow{d} N(0, \frac{4c^2}{2c - 1 + e^{-2c}}), \quad N, T \to \infty. \tag{3.12}$$

Proof. It follows from Theorem 2.1 and Lemma 3.2 that we only need to verify the corresponding conditions of moment convergence and calculate the variance of the random variable in the right hand side of (3.10) and the expectation of the random variable in the right hand side of (3.11). To verify the conditions of moment convergence. As before, without loss of generality, we assume $y_{i0} = 0$ for all $i \geq 1$. It is true that for every $i \geq 1$,

$$E\left[T^{-1}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_{it}\right] = 0,$$

$$\begin{aligned} E\left[\left(T^{-1}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_{it}\right)^2\right] &= \frac{1}{T^2}\sum_{t=1}^{T}\frac{1-\rho_T^{2(t-1)}}{1-\rho_T^2} \\ &= \frac{1}{T^2(1-\rho_T^2)}\left(T - \frac{1-\rho_T^{2T}}{1-\rho_T^2}\right) \\ &\to \frac{2c-1+e^{-2c}}{4c^2}, \quad T \to \infty, \end{aligned}$$

$$E\left[T^{-2}\sum_{t=1}^{T} y_{i,t-1}^2\right] = E\left[\left(T^{-1}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_{it}\right)^2\right] \to \frac{2c-1+e^{-2c}}{4c^2}, \quad T \to \infty,$$

$$E\left[\frac{b}{2c}\int_0^1 (1+bt)^{-1}W(t)dW(t)\right]=0,$$

$$\begin{aligned} E\left[\left(\frac{b}{2c}\int_0^1 (1+bt)^{-1}W(t)dW(t)\right)^2\right] &= \frac{b^2}{4c^2}E\left[\int_0^1 \left((1+bt)^{-1}W(t)\right)^2 dt\right] \\ &= \frac{b^2}{4c^2}\int_0^1 \frac{t}{(1+bt)^2}dt \\ &= \frac{1}{4c^2}\left[\ln(1+b)-b/(1+b)\right] \\ &= \frac{2c-1+e^{-2c}}{4c^2} \end{aligned}$$

according to Itó isometry theorem, and

$$\begin{aligned} E\left[\left(\frac{b}{2c}\right)^2\int_0^1 (1+bt)^{-2}W^2(t)dt\right] &= E\left[\left(\frac{b}{2c}\int_0^1 (1+bt)^{-1}W(t)dW(t)\right)^2\right] \\ &= \frac{2c-1+e^{-2c}}{4c^2}. \end{aligned}$$

To calculate the variance of the random variable in the right hand side of (3.10) and the expectation of the random variable in the right hand side of (3.11). Note that the latter has just been done. For the former, it is easy to see that

$$\begin{aligned} Var\left(\frac{b}{2c}\int_0^1 (1+bt)^{-1}W(t)dW(t)\right) &= E\left[\left(\frac{b}{2c}\int_0^1 (1+bt)^{-1}W(t)dW(t)\right)^2\right] \\ &= \frac{2c-1+e^{-2c}}{4c^2}. \end{aligned}$$

The proof is complete. □

Remark 3.4. *It is easy to see that*

$$\lim_{c\to 0}\frac{4c^2}{2c-1+e^{-2c}}=2.$$

Thus, the second part of Theorem 3.1 can be regarded as a complementary of Theorem 3.3.

Now we investigate the case where $\rho = 1-\frac{c}{k_T}$, where $c\neq 0$ and k_T is an increasing positive function of T diverging to infinity such that $k_T = o(T)$. First, we introduce a result about the limiting distribution of $\hat{\rho}_T$ in univariate time series AR(1) model which is taken from Phillips and Magdalinos (2007).

Lemma 3.3. *Let $\rho_T = 1-\frac{c}{k_T}$, where $c\neq 0$ is a fixed constant and k_T is an increasing positive function of T diverging to infinity such that $k_T = o(T)$. For $t = 1,2,...,T$, suppose y_t satisfied the following reparameterized AR(1) model,*

$$y_t = \rho_T y_{t-1} + \varepsilon_t, \quad t = 1,2,...T,$$

where $y_0 = o_P(\sqrt{k_T})$ and $\{\varepsilon_t, t\geq 1\}$ are i.i.d. random variables with $E[\varepsilon_1]=0$ and $E[\varepsilon_1^2]=1$. Then, for $c>0$, one has

$$\frac{1}{\sqrt{Tk_T}}\sum_{t=1}^T y_{t-1}\varepsilon_t \xrightarrow{d} N(0,\frac{1}{2c}), \quad T\to\infty, \tag{3.13}$$

$$\frac{1}{Tk_T}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{P} \frac{1}{2c}, \quad T\to\infty; \tag{3.14}$$

and for $c<0$, one has

$$\left(\frac{1}{\rho_T^T k_T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t, \frac{-2c}{(\rho_T^T k_T)^2}\sum_{t=1}^{T} y_{t-1}^2\right) \xrightarrow{d} (XY, Y^2), \quad T\to\infty, \tag{3.15}$$

where X and Y are two independent random variables obeying $N(0,1/(-2c))$.

We now study the panel data case. By employing Theorem 2.1 and Lemma 3.3, we immediately have the following result.

Theorem 3.4. *Let $\rho_T = 1-\frac{c}{k_T}$, where $c\neq 0$ is a fixed constant and k_T is an increasing positive function of T diverging to infinity such that $k_T = o(T)$. For $t=1,2,...,T$ and $i=1,2,..,N$, suppose y_{it} satisfies the following reparameterized AR(1) model,*

$$y_{it} = \rho_T y_{i,t-1} + \varepsilon_{it}, \quad t=1,2,...,T, \quad i=1,2,...,N,$$

where $y_{i0} = o_P(\sqrt{k_T})$ for all $i\geq 1$ and $\{\varepsilon_{it}, i\geq 1, t\geq 1\}$ are i.i.d. random variables with $E[\varepsilon_{11}]=0$ and $E[\varepsilon_{11}^2]=1$. Then, for $c>0$ we have

$$\sqrt{NTk_T}(\hat\rho_T-\rho_T) \xrightarrow{d} N(0,2c), \quad N,T\to\infty; \tag{3.16}$$

and for $c<0$ we have

$$\sqrt{N}k_T\rho_T^T(\hat\rho_T-\rho_T) \xrightarrow{d} N(0,4c^2), \quad N,T\to\infty. \tag{3.17}$$

Proof. The proofs of (3.16) and (3.17) are similar, so we only prove (3.17) here. To do so, it follows from Theorem 2.1 and Lemma 3.3 that we only need to verify the corresponding conditions of moment convergence and calculate the variance of XY and the expectation of Y^2 in the right hand side of (3.15). As before, without loss of generality, we assume $y_{i0}=0$ for all $i\geq 1$. Noting that $\rho_T^{-T} = o(k_T/T)$ by Proposition A.1 in Phillips and Magdalinos (2007), it is true that for every $i\geq 1$,

$$E\left[\frac{1}{\rho_T^T k_T}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_{it}\right] = 0,$$

$$\begin{aligned} E\left[\left(\frac{1}{\rho_T^T k_T}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_{it}\right)^2\right] &= \frac{1}{\rho_T^{2T}k_T^2}\sum_{t=1}^{T}\frac{1-\rho_T^{2(t-1)}}{1-\rho_T^2} \\ &= \frac{1}{\rho_T^{2T}k_T^2(1-\rho_T^2)}\left(T-\frac{1-\rho_T^{2T}}{1-\rho_T^2}\right) \\ &= o(1)+\frac{1}{k_T^2(1-\rho_T^2)^2} \\ &\to \frac{1}{4c^2}, \quad T\to\infty, \end{aligned}$$

$$E\left[\frac{-2c}{(\rho_T^T k_T)^2}\sum_{t=1}^{T} y_{i,t-1}^2\right] = \frac{-2c}{\rho_T^{2T}k_T^2}\sum_{t=1}^{T}\frac{1-\rho_T^{2(t-1)}}{1-\rho_T^2} = o(1)+\frac{-2c}{k_T^2(1-\rho_T^2)^2} \to -\frac{1}{2c}, \quad T\to\infty,$$

$$E[XY]=0, \quad Var(XY) = E\left[(XY)^2\right] = \frac{1}{4c^2}, \quad E[Y^2] = -\frac{1}{2c}.$$

The proof is complete. □

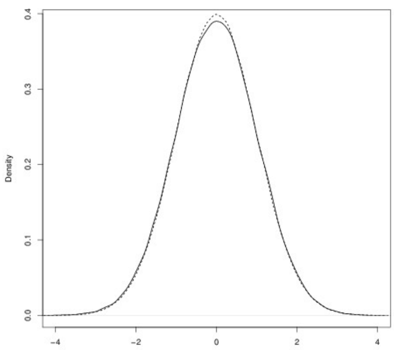
§4 Simulations

In this section, we perform some experiments to see how well the finite-sample properties of the LSE of the autoregressive coefficient ρ or ρ_T follow the asymptotic results in Section 3. We generate the observations y_{it}'s according to model (2.1). In the experiments, the cross-sectional dimension is set at $N = 300$, the time-series dimension is also set at $T = 300$, and the number of replications is set at $50,000$. In addition, all y_{i0}'s are set at zero for simplicity. We take $\rho = 0.8$ for the stationary case (corresponds to the result (3.1) in Theorem 3.1), take $\rho = 1$ for unit root case (corresponds to the result (3.2) in Theorem 3.1), take $\rho = 3$ for the explosive case (corresponds to the result (3.3) in Theorem 3.2), take $\rho_T = 1 - 1/T$ for the local to unity case (corresponds to the result (3.12) in Theorem 3.3), take $\rho_T = 1 - 1/T^{0.8}$ for the mildly integrated case (corresponds to the result (3.16) in Theorem 3.4), and take $\rho_T = 1 + 1/T^{0.8}$ for the mildly explosive case (corresponds to the result (3.17) in Theorem 3.4).

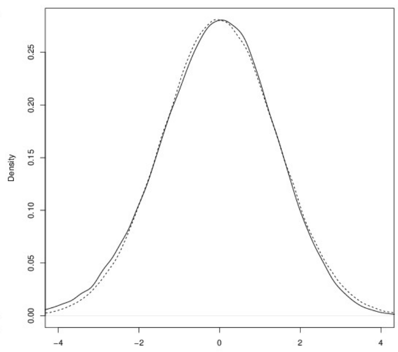
Parts (a) and (b) of Figure 1 show the distributions of $\hat{\rho}$ for Theorem 3.1, part (c) of Figure 1 shows the distribution of $\hat{\rho}$ for Theorem 3.2, part (d) of Figure 1 shows the distribution of $\hat{\rho}_T$ for Theorem 3.3, and parts (e) and (f) of Figure 1 show the distributions of $\hat{\rho}_T$ for Theorem 3.4. Obviously, Theorems 3.1-3.4 are all supported by Figure 1.

§5 Conclusions

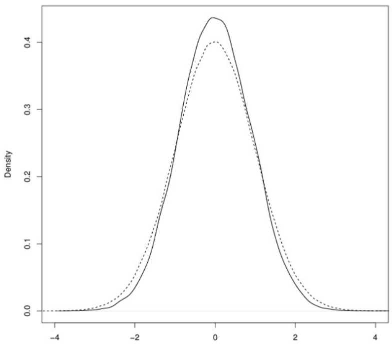
In this article, we study the asymptotic properties of the least squares estimator of the autoregressive coefficient ρ in the panel data AR(1) model. Unlike the univariate time series case, we prove in this paper that the limiting distribution of the least squares estimator of ρ is always normal for all $\rho \in R$, which is a standard distribution. Hence, this result is useful for conducting further statistical inferences. As some applications, we apply our result to some common used panel data AR(1) model, that is, stationary model, unit root model, explosive model, local to unity model, mildly integrated model and mildly explosive model. Monte Carlo simulations are then conducted to examine the finite-sample performance for the estimators. Our theoretical findings are supported by the Monte Carlo simulations.



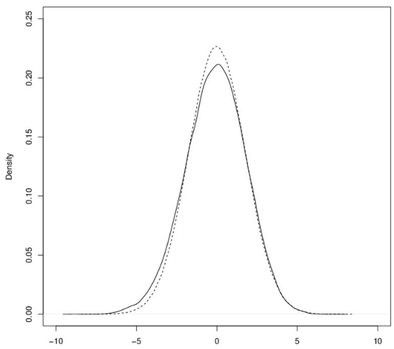
(a) stationary case



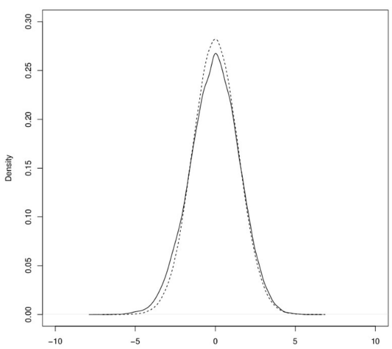
(b) unit root case



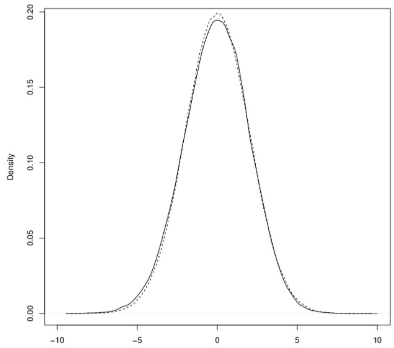
(c) explosive case



(d) local to unity case



(e) mildly integrated case



(f) mildly explosive case

Figure 1: The finite-sample distributions and the corresponding limiting distributions of $\hat{\rho}$ or $\hat{\rho}_T$. The solid lines represent the graphs when $N = T = 300$, and the dashed lines represent the graph when $N = T = \infty$.

References

[1] T W Anderson. *On asymptotic distributions of estimates of parameters of stochastic difference equations*, Ann Math Statist, 1959, 30(3): 676-687.

[2] N H Chan, C Z Wei. *Asymptotic inference for nearly nonstationary AR (1) processes*, Ann Statist, 1987, 15(3): 1050-1063.

[3] J Hahn, H R Moon. *Reducing bias of MLE in a dynamic panel model*, Econometric Theory, 2006, 22(3): 499-512.

[4] A T Levin, C F Lin. *Unit root tests in panel data: asymptotic and finite-sample properties*, Economics Working Paper Series, 1992.

[5] X Lu, L Su. *Shrinkage estimation of dynamic panel data models with interactive fixed effects*, Journal of Econometrics, 2016, 190(1): 148-175.

[6] H B Mann, A Wald. *On the statistical treatment of linear stochastic difference equations*, Econometrica, 1943, 11(3): 173-220.

[7] H R Moon, P C B Phillips. *Estimation of autoregressive roots near unity using panel data*, Econometric Theory, 2000, 16(6): 927-997.

[8] P C B Phillips. *Towards a unified asymptotic theory for autoregression*, Biometrika, 1987, 74(3): 535-547.

[9] P C B Phillips, T Magdalinos. *Limit theory for moderate deviations from a unit root*, Journal of Econometrics, 2007, 136(1): 115-130.

[10] P C B Phillips, H R Moon. *Linear regression limit theory for nonstationary panel data*, Econometrica, 1999, 67(5): 1057-1111.

[11] M M Rao. *Asymptotic distribution of an estimator of boundary parameter of an unstable process*, Ann Statist, 1978, 6(1): 185-190.

[12] S H Sung. *Weak law of large numbers for arrays of random variables*, Statistics & Probability Letters, 1999, 42(3): 293-298.

[13] J S White. *The limiting distribution of the serial correlation coefficient in the explosive case*, Ann Math Statist, 1958, 29(4): 1188-1197.

School of Mathematical Sciences, Zhejiang University, Hangzhou 310027, China.
Email: txpang@zju.edu.cn