Appl. Math. J. Chinese Univ.
2024, 39(1): 140-158

Asymptotic normality of error density estimator in stationary and explosive autoregressive models

WU Shi-peng[1,2] YANG Wen-zhi[1] GAO Min[1] HU Shu-he[1]

Abstract. In this paper, we consider the limit distribution of the error density function estimator in the first-order autoregressive models with negatively associated and positively associated random errors. Under mild regularity assumptions, some asymptotic normality results of the residual density estimator are obtained when the autoregressive models are stationary process and explosive process. In order to illustrate these results, some simulations such as confidence intervals and mean integrated square errors are provided in this paper. It shows that the residual density estimator can replace the density "estimator" which contains errors.

§1 Introduction

Consider the following first-order autoregressive process

$$X_i = \rho X_{i-1} + \varepsilon_i, \quad 1 \le i \le n, \tag{1}$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are real valued random errors with the unknown probability density function (p.d.f.) $f(x), x \in \mathbb{R}$. Note that the process (1) can be expressed as

$$X_i = \sum_{j=0}^{i-1} \rho^j \varepsilon_{i-j} + \rho^i X_0, \quad 1 \le i \le n. \tag{2}$$

Without loss of generality, we take $X_0 = 0$ in this paper. It is known that the autoregressive process $\{X_i\}$ may be different process if ρ takes different values. For instance,

(i) If $|\rho| < 1$, then the process $\{X_i\}$ is a stationary process which can be presented as

$$X_i = \sum_{j=0}^{\infty} \rho^j \varepsilon_{i-j}, \quad i \ge 1. \tag{3}$$

Received: 2021-09-07. Revised: 2021-11-01.
MR Subject Classification: 62G07, 62G20.
Keywords: explosive autoregressive models, residual density estimator, asymptotic distribution, association sequence.
Digital Object Identifier(DOI): https://doi.org/10.1007/s11766-024-4558-x.
Supported by the National Natural Science Foundation of China (11701004), National Natural Science Foundation of Anhui Province (2008085MA14) and Science and Technology Innovation Project of Shanxi Province Universities (2023L165).

(ii) If $|\rho| = 1$, then the process $\{X_i\}$ is a random walk.

(iii) Let $\rho = \rho_n = 1 + c/b_n$, where constant $c \in \mathbb{R}$ and b_n is a sequence increasing to ∞ such that $b_n = o(n)$ as $n \to \infty$. If $c > 0$, then the process $\{X_i\}$ is a near-explosive process; if $c < 0$, then the process $\{X_i\}$ is a near-stationary process.

(iv) If $|\rho| > 1$, then the process $\{X_i\}$ is an explosive process.

For further details of model (1) and its applications, one can refer to Aue and Horváth (2007), Brockwell and Davis (1991), Phillips and Magdalinos (2007), Magdalinos (2012), Yang et al. (2018), Oh et al. (2018), Shen and Pang (2020) and the references therein.

In order to make statistical inferences, one might require some knowledge of the density function $f(\cdot)$ of error sequence $\{\varepsilon_i, 1 \le i \le n\}$, which is assumed to exist in this paper. If the errors $\varepsilon_1, \dots, \varepsilon_n$ were observed, then the error density function "estimator"

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\Big(\frac{x - \varepsilon_i}{h_n}\Big), \quad x \in \mathbb{R}, \tag{4}$$

could be used to estimate p.d.f. $f(x)$ with an appropriately chosen bandwidth h_n and a kernel density function $K(\cdot)$. However, only the random variables $X_0, X_1, \dots, X_n$ are observed in the autoregressive model (1). Thus, in order to obtain an estimator of p.d.f $f(x)$, we modify definition (4) of $f_n(x)$ by plugging in the random variables

$$\hat{\varepsilon}_i = X_i - \hat{\rho}_n X_{i-1}, 1 \le i \le n, \tag{5}$$

instead of $\varepsilon_1, \dots, \varepsilon_n$, where $\hat{\rho}_n = \hat{\rho}_n(X_0, X_1, \dots, X_n)$ is an estimator of ρ computed by the observations $X_0, X_1, \dots, X_n$. For example, the least squares estimator

$$\hat{\rho}_n = \frac{\sum_{i=1}^{n} X_{i-1} X_i}{\sum_{i=1}^{n} X_{i-1}^2} \tag{6}$$

can be used to estimate ρ. Consequently, the residual kernel-type estimator $\hat{f}_n(x)$ of $f(x)$ is given as

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\Big(\frac{x - \hat{\varepsilon}_i}{h_n}\Big), \quad x \in \mathbb{R}. \tag{7}$$

The properties of kernel density estimator $f_n(x)$ have been well investigated in the literature. For example, based on the independent and identically distributed (*i.i.d.*) sample, Bickel and Rosenblatt (1973) provided the limit distribution of the L_2-norm of density estimator $f_n(x)$; Masry (1986) obtained the asymptotic normality for the recursive probability density estimator under α-mixing sample; Castellana and Leadbetter (1986) considered the consistency and asymptotic distribution of $f_n(x)$ under stationary processes; Roussas (1999), Lu (2001), and Yang and Hu (2012) investigated the convergence rate, asymptotic normality and Berry-Esseen bounds for density estimator $f_n(x)$ under mixing sample; For more details of density estimator, one can refer to Fan and Yao (2003), Li and Rcine (2007), etc.

There are many authors that studied the asymptotic properties of residual density $\hat{f}_n(x)$ based on the *i.i.d.* errors. For example, in the model (1) with $|\rho| > 1$, Koul and Levental (1989) proved the weak convergence of the residual empirical process to the Brownian bridge; Lee and Na (2002) considered the first order autoregressive models (1) with $|\rho| < 1$ and $|\rho| > 1$ processes, and showed that the difference between the L_2-norm between $f_n(x)$ and $\hat{f}_n(x)$ is so small that the Bickel-Rosenblatt asymptotic normality for $f_n(x)$ also holds for $\hat{f}_n(x)$; Horváth and Zitikis (2004) extended the results of Lee and Na (2002) to L_p-norm; Cheng (2005) considered the autoregressive model (1) with $|\rho| < 1$ process and obtained the asymptotic normality of $\frac{\hat{f}_n(x)-Ef_n(x)}{\sqrt{\mathrm{Var}(f_n(x))}}$; Wang and Yu (2015) obtained that the coefficient based test and the t test have better power for testing the hypothesis of zero intercept in the explosive process than in the stationary process. In addition, Liebscher (1999), Cheng (2010) and Cheng and Sun (2008) considered the residual density estimator $\hat{f}_n$ in the nonlinear autoregressive models with *i.i.d.* errors.

Obviously, the condition of independence of $\{\varepsilon_i, 1 \le i \le n\}$ seems to be strong. Therefore, Kim et al. (2014) extended the result of Cheng and Sun (2008) in the nonlinear autoregressive model with *i.i.d.* error terms to the model with stationary α-mixing error terms. Gao et al. (2022) studied the asymptotic normality of residual density estimator $\hat{f}_n(x)$ in the autoregressive model with the α-mixing errors. In this paper, we go on studying the residual density estimator $\hat{f}_n(x)$ with the associated errors. Now, let us recall some conceptions of association such as negatively associated (NA) and positively associated (PA).

Definition 1.1 *A finite family $\{\varepsilon_1, \dots, \varepsilon_n\}$ is said to be NA if for any disjoint subsets A, B $\subset \{1, 2, \dots, n\}$, and any real coordinatewise nondecreasing functions f on $\mathbb{R}^A$, g on $\mathbb{R}^B$,*

$$Cov(f(\varepsilon_k, k \in A), g(\varepsilon_k, k \in B)) \le 0,$$

whenever this covariance is defined. An infinite sequence of random variables $\{\varepsilon_i, i \ge 1\}$ is said to be NA, if every finite subcollection is NA.

Definition 1.2 *A finite collection of random variables $\{\varepsilon_1, \dots, \varepsilon_n\}$ is said to be PA if for any two coordinatewise nondecreasing functions f, g: $\mathbb{R}^n \to \mathbb{R}$,*

$$Cov(f(\varepsilon_1, \dots, \varepsilon_n), g(\varepsilon_1, \dots, \varepsilon_n)) \ge 0,$$

whenever this covariance is defined. An infinite sequence $\{\varepsilon_i, i \ge 1\}$ is said to be PA, if every finite subcollection is PA.

The concepts of NA and PA were introduced by Joag-Dev and Proschan (1983) and Esary et al. (1967), respectively. They can be used in many fields such as Gaussian system, survival analysis system, etc. For more examples and applications, one can refer to Bulinski and Shaskin (2007), Tan and Yang (2008), Prakasa (2012), Oliveira (2012) and Zhang et al.(2022). Based on the NA or PA random variables, Roussas (2000) considered the kernel-type density estimator $f_n(x)$ in (4), and obtained asymptotic distribution for density estimator $f_n(x)$ of $f(x)$. Inspired by Roussas (2000) and Cheng (2005), we study the asymptotic distribution of the residual kernel-type density estimator $\hat{f}_n(x)$ in stationary (i.e. $|\rho| < 1$) and explosive (i.e. $|\rho| > 1$) models with association errors (see our main results in Section 3).

The rest of this paper is organized as follows. In Section 2, we list some basic assumptions for the first-order autoregressive model (1). The main results of asymptotic distributions are given in Section 3. Some simulations are performed to evaluate the performance of the estimator $\hat{f}_n(t)$ in Section 4. We offer some conclusions in Section 5. The proofs of the main results are given in Section 6. Finally, some lemmas are presented in the Appendix.

Throughout the paper, we assume that all limits are taken as $n \to \infty$ unless otherwise specified. Let $\xrightarrow{d}$ stand the convergence in distribution. Let $O_P(\cdot)$ stand the bounded in probability and $o_P(1)$ stand the tend to 0 in probability. The C, C_1, C_2 denote some positive constants not depending on n, which may be different in different places.

§2 Basic assumptions

In this section, we shall list some basic assumptions such as for the density function $f(x)$ and moment of error, kernel function $K(\cdot)$ and estimator $\hat{\rho}_n$ of ρ, which are required for the limit distribution of residual density estimator $\hat{f}_n(x)$ in the model (1).

Assumptions:

(A1) (i) The sequence $\{\varepsilon_n\}$ is a stationary sequence of identically distributed association random variables with unknown marginal bounded probability density function $f(\cdot)$.

(ii) The second-order derivative $f''(\cdot)$ exists and is bounded in $\mathbb{R}$.

(iii) Assume that $f_{1,j}(x, y)$ is the joint probability distribution function of random variables ε_1 and ε_{1+j}, which satisfies

$$|f_{1,j}(x, y) - f(x)f(y)| \leq C \ for \ all \ x, y \in \mathbb{R} \ and \ j \geq 1.$$

(A2) (i) The $K(\cdot)$ is a known kernel density function such that:

$$K(u) \leq C, u \in \mathbb{R}; \ \lim(|u|K(u)) = 0 \ as \ |u| \to \infty.$$

(ii) The derivative $k(u) = K'(u)$ exists and is bounded $|k(u)| \leq C_1, u \in \mathbb{R}$.

(iii) Let function $K(\cdot)$ be satisfied

$$\int_{-\infty}^{\infty} uK(u)du = 0, \quad \int_{-\infty}^{\infty} u^2 K(u)du = C_2 \neq 0.$$

(A3) Let the bandwidth sequence h_n tend to 0 and satisfy the requirements:

(i) $nh_n^3 \to \infty$ as $n \to \infty$.

(ii) $nh_n^5 \to 0$ as $n \to \infty$.

(A4) (i) Let $E\varepsilon_1 = 0$ and $E\varepsilon_1^2 = \sigma_1^2 < \infty$.

(ii) Denote $u(n) = \sum_{j=n}^{\infty} |\text{Cov}(\varepsilon_1, \varepsilon_{j+1})|$ and suppose that $u(1) < \infty$.

(A5) Let $1 < p_n < n$, $0 < q_n < n$, be integers tending to ∞ along with n and let $k_n \to \infty$ be defined by $k_n = \lfloor n/(p_n + q_n) \rfloor$, where $\lfloor x \rfloor$ stands for the integral parts of x. Then there is a determination of them for which:

(i) $p_n k_n / n \to 1$ as $n \to \infty$.

(ii) $p_n h_n \to 0$ and $p_n^2/(nh_n) \to 0$ as $n \to \infty$.

(iii) $(1/h_n^3) \sum_{j=q_n}^{\infty} |\text{Cov}(\varepsilon_1, \varepsilon_{1+j})| \to 0$ as $n \to \infty$.

(A6) For the estimator $\hat{\rho}_n$ of ρ, we assume that

(i) $n^{1/2}(\hat{\rho}_n - \rho) = O_P(1)$, if $|\rho| < 1$.

$(i)^*$ $|\rho|^n(\hat{\rho}_n - \rho) = O_P(1)$, if $|\rho| > 1$.

Remark 2.1. We list some comments on the assumptions.

(R1) The Assumptions (A1)(i)-(iii) are used commonly in the kernel density estimate for the error sequence $\{\varepsilon_n, n \geq 1\}$. For more details, see Masry (1986), Roussas (2000), Fan and Yao (2003) and Li and Rcine (2007), etc.

(R2) The Assumptions (A2)(i)-(iii) are the conditions of kernel function $K(\cdot)$. The Assumption (A2)(ii) may be weakened to read as follows: The kernel density function $K(\cdot)$ is continuous, its left-hand side and right-hand side derivatives exist and are equal except for a finite number of points, and they are bounded in $\mathbb{R}$. The Assumption (A2)(iii) is the integrability condition of K. The Assumptions (A2)(i)(ii) are used in many literature such as Fan and Yao (2003) and Li and Rcine (2007), etc.

(R3) The Assumptions (A3) is the condition for bandwidth $\{h_n, n \geq 1\}$. The h_n needs to satisfy the requirements $h_n \gg n^{-1/3}$ and $h_n \ll n^{-1/5}$. Here, $a_n \ll b_n$ denotes that there exists a constant $c > 0$ such that $a_n \leq cb_n$ for n sufficiently large, $a_n \gg b_n$ is similar defined. It is easily seen that it is a common condition to study the properties of kernel-type density estimator.

(R4) The Assumptions (A4)(i)(ii) are the conditions of the moment and covariance for the error sequence $\{\varepsilon_n, n \geq 1\}$. The similar conditions were used by Roussas (2000) and Qin et al. (2011) etc.

(R5) The Assumptions (A5)(i)(ii) are easily satisfied, if p_n and q_n are chosen as follows: with $h_n \to 0$, let $p_n \sim h_n^{-\delta_1}$, $q_n \sim h_n^{-\delta_2}(0 < \delta_2 < \delta_1 < 1)$, where $x_n \sim y_n$ means $x_n/y_n \to 1$ as $n \to \infty$. Obviously, $k_n \sim nh_n^{\delta_1}$, so that Assumption (A5)(i)(ii) are fulfilled, provided $nh_n^{1+2\delta_1} \to \infty$ as $n \to \infty$. Also, the Assumption (A5)(iii) is satisfied for the following forms of the covariance function. Set $C_j = |\text{Cov}(\varepsilon_1, \varepsilon_{1+j})|$ and let $C_j = r_0 r^j$ $(0 < r < 1, r_0 > 0)$. Then,

$$h_n^{-3} \sum_{j \geq q_n} C_j \sim h_n^{-3} r^{q_n} \sim h_n^{-3} / \exp[(-\log r) h_n^{-\delta_2}],$$

and this tends to 0, as $n \to \infty$. Thus, Assumption (A5) is satisfied under the condition that $nh_n^{1+2\delta_1}$ as $n \to \infty$. Next, let $C_j = j_0 j^{-\theta}$ $(\theta > 1, j_0 > 0)$. Then

$$h_n^{-3} \sum_{j \geq q_n} C_j \sim h_n^{-3+(\theta-1)\delta_2} \to 0, \text{ as } n \to \infty$$

provided, $\theta > 1 + 3/\delta_2$. Thus, for this choice of the covariance function, the Assumption (A5) is satisfied for $\theta > 1 + 3/\delta_2$ and $nh_n^{1+2\delta_1} \to \infty$, $n \to \infty$ $(0 < \delta_2 < \delta_1 < 1)$; in fact, it suffices that $\theta > 4.1$, say. It is easily seen that the Assumption (A3) satisfies the requirement for h_n in Assumption (A5). For more details, see Roussas (2000).

(R6) The Assumptions (A6) is the condition of estimator $\hat{\rho}_n$ of ρ in the first order autoregressive model. For example, Assumption (A6)(i) is for the case $|\rho| < 1$, i.e. stationary process and (A6)(ii) is for the case $|\rho| > 1$, i.e. explosive process. The Assumption (A6) is used commonly by authors such as Koul and Levental (1989), Lee and Na (2002), Cheng (2005) and Horváth and Zitikis (2004), etc.

§3 Main results

First, we consider the autoregressive model (1) with stationary process (i.e. $|\rho| < 1$) and obtain the asymptotic normality of residuals density estimator $\hat{f}_n(x)$ in Theorem 3.1.

Theorem 3.1. *Consider the model* (1) *with* $|\rho| < 1$, *where* $\{\varepsilon_n, n \geq 1\}$ *is either NA sequence or PA sequence. Let the Assumptions (A1)-(A5) and (A6)(i) hold. Then, for* $x \in \mathbb{R}$,

$$\sqrt{nh_n}(\hat{f}_n(x) - f(x)) \xrightarrow{d} N(0, \sigma^2(x)). \tag{8}$$

where $\sigma^2(x) = f(x) \int_{-\infty}^{\infty} K^2(v) dv$ *and assume that* x *is a point of continuity of* $f(x)$ *with* $f(x) > 0$.

Similar to Theorem 3.1, we consider the autoregressive model with explosive process (i.e. $|\rho| > 1$) and obtain the following Theorem 3.2.

Theorem 3.2. *Consider the model* (1) *with* $|\rho| > 1$, *where* $\{\varepsilon_n, n \geq 1\}$ *is either NA sequence or PA sequence. Let the Assumptions (A1)-(A5) and (A6)(i)* hold. Then, for* $x \in \mathbb{R}$,

$$\sqrt{nh_n}(\hat{f}_n(x) - f(x)) \xrightarrow{d} N(0, \sigma^2(x)), \tag{9}$$

where $\sigma^2(x)$ *is defined as that in* (8).

Remark 3.1. In the first order autoregressive model (1) with $|\rho| < 1$ and *i.i.d.* errors, Cheng (2005) obtained asymptotic distribution of the residual kernel-type density estimator $\hat{f}_n(x)$ such that $\frac{\hat{f}_n(x) - Ef_n(x)}{\sqrt{Var(f_n(x))}} \xrightarrow{d} N(0, 1)$. Based on the association random variables, Roussas (2000) obtained the asymptotic normality for the kernel-type density estimator $f_n(x)$. Inspired by Cheng (2005) and Roussas (2000), we, under the identically distributed dependent errors such as NA and PA random variables, consider the autoregressive models with $|\rho| < 1$ and $|\rho| > 1$, and obtain the asymptotic normality results in Theorems 3.1-3.2.

Remark 3.2. The Assumption (A3)(i)(ii) is satisfied for $h_n \gg n^{-1/3}$ and $h_n \ll n^{-1/5}$. Taking $h_n = n^{-1/5} \log^{-1} n$, we obtain the absolute convergence rate between $\hat{f}_n(x)$ and $f(x)$ in Theorems 3.1 and 3.2.

$$|\hat{f}_n(x) - f(x)| = O_P(\frac{1}{\sqrt{nh_n}}) = O_P(n^{-2/5} \log^{1/2} n).$$

Theorem 3.1 with $|\rho| < 1$ implies that a straight-forward approach to construct a confidence regions is to use asymptotic normality with a variance estimator. Since $\hat{f}_n(x)$ is a consistent estimator of $f(x)$, so a "plug-in method" is replacing $f(x)$ in the asymptotic variance by its estimator $\hat{f}_n(x)$, leading to the following $1 - \alpha$ confidence interval of $f(x)$:

$$\Big[\hat{f}_n(x) - z_{1-\alpha/2}\sqrt{\frac{\hat{\sigma}_n^2(x)}{nh_n}}, \hat{f}_n(x) + z_{1-\alpha/2}\sqrt{\frac{\hat{\sigma}_n^2(x)}{nh_n}}\Big], \tag{10}$$

where $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard normal distribution and

$$\hat{\sigma}_n^2(x) = \hat{f}_n(x)\int_{-\infty}^{\infty} K^2(v)dv > 0.$$

Similarly, we can construct a confidence region for $f(t)$ based on Theorem 3.2 with $|\rho| > 1$. Therefore, we provided a unified approach for interval estimation of density function of error in AR(1) model with $|\rho| < 1$ and $|\rho| > 1$. In future research, it is also interested to study a unified approach that confidence interval estimation of density function of error in AR(1) model to encompass all situations (ρ take any value).

§4 Simulation

In this section, we conduct some simulation experiments to compare the kernel density "estimator" $f_n(x)$ defined by (4) and estimator $\hat{f}_n(x)$ defined by (7) based on errors $\{\varepsilon_i, i = 1, 2, \ldots, n\}$ and residuals $\{\hat{\varepsilon}_i, i = 1, 2, \ldots, n\}$, respectively. Since the error sequence $\{\varepsilon_i, 1 \le i \le n\}$ is generated by simulation, similar to (10), the $(1-\alpha)\%$ confidence region of $f(x)$ by $f_n(x)$ is given by

$$\Big[f_n(x) - z_{1-\alpha/2}\sqrt{\frac{\sigma_n^2(x)}{nh_n}}, f_n(x) + z_{1-\alpha/2}\sqrt{\frac{\sigma_n^2(x)}{nh_n}}\Big], \tag{11}$$

where $\sigma_n^2(x) = f_n(x)\int_{-\infty}^{\infty} K^2(v)dv$ and $z_{1-\alpha/2}$ is defined as that in (10).

The experimental data are generated by the following first order autoregressive time series models:

$$X_i = \rho X_{i-1} + \varepsilon_i,\ \ i = 1, 2, \ldots, n. \tag{12}$$

where $|\rho| < 1$ is for stationary process and $|\rho| > 1$ is for explosive process. Let $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote n-dimensional multivariate normal random vector, $\phi_{\boldsymbol{\theta}}(\boldsymbol{x})$ denote the multivariate normal density with parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{x} =_d \boldsymbol{y}$ mean that $\boldsymbol{x}$ and $\boldsymbol{y}$ have the same distribution function. Then random error vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$ are drawn from mixture Gaussian distribution (see Hastie and Tibshirani (1996)), i.e.

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)' =_d (1-\Delta)N_n(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_n) + \Delta N_n(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_n)$$

where Δ, $N_n(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_n)$ and $N_n(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_n)$ are independent, $\Delta \in \{0, 1\}$, $P(\Delta = 1) = 1/2$. Thus the density of random vector $\boldsymbol{\varepsilon}$ is

$$g_{\boldsymbol{\varepsilon}}(\boldsymbol{x}) = \frac{1}{2}\phi_{\boldsymbol{\theta}_1}(\boldsymbol{x}) + \frac{1}{2}\phi_{\boldsymbol{\theta}_2}(\boldsymbol{x}), \tag{13}$$

where $\boldsymbol{\theta}_1 = (\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_n)$ and $\boldsymbol{\theta}_2 = (\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_n)$. It is easy to check that $E(\boldsymbol{\varepsilon}) = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$ and $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}_n$. Let

$$\boldsymbol{\Sigma}_n = \boldsymbol{I}_n - \boldsymbol{\Lambda}, \tag{14}$$

where $\boldsymbol{I}_n$ is $n \times n$ unit matrix and $\boldsymbol{\Lambda} = (\beta^{|i-j|})_{1 \le i,j \le n, i \ne j}$ with some $\beta \in (0, 1)$. Then we have that $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)'$ is a NA random vector. Similarly, $(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)'$ is a PA random vector if covariance matrix $\boldsymbol{\Sigma}_n = \boldsymbol{I}_n - \boldsymbol{\Lambda}$ is replace by $\boldsymbol{\Sigma}_n = \boldsymbol{I}_n + \boldsymbol{\Lambda}$ for some $\beta \in (0, 1)$. For more details, see for example Joag-Dev and Proschan (1983) and Bulinski and Shaskin (2007). In addition, let $\boldsymbol{\mu}_1 = (\overbrace{\mu_1, \cdots, \mu_1}^{n})'$, $\boldsymbol{\mu}_2 = (\overbrace{\mu_2, \cdots, \mu_2}^{n})'$ and $\boldsymbol{\Sigma}_n = \boldsymbol{I}_n - \boldsymbol{\Lambda}$ or $\boldsymbol{\Sigma}_n = \boldsymbol{I}_n + \boldsymbol{\Lambda}$. So by (13), the random variables $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ have the same density function such as

$$f(x) = \frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{(x-\mu_1)^2}{2}\Big) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{(x-\mu_2)^2}{2}\Big), \quad x \in \mathbb{R}. \tag{15}$$

Meanwhile, the Gaussian kernel density function is taken by

$$K(x) = \frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{1}{2}x^2\Big), \quad x \in \mathbb{R},$$

where the smoothing parameter h_n is taken by $h_n = n^{-1/4}$ which satisfies Assumption (A3). For the plug-in and bootstrap methods of selection h_n, one can refer to Chen (2017) and the references therein. It is easy to calculate $\int_R K^2(x)dx = \frac{1}{2\sqrt{\pi}}$. Finally, we can construct error density "estimator" $f_n(x)$ by (4) and obtain the residuals density estimator $\hat{f}_n(x)$ using the least squares estimator $\hat{\rho}_n$ defined by (6) of ρ and residuals $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$.

Now, we consider the stationary model (12) with $\rho = 0.8$, take $n = [50, 100, 150, 200]$, $\mu_1 = -2$, $\mu_2 = 2$ in (15) and choose $\beta = 0.3$ in $\boldsymbol{\Sigma}_n = \boldsymbol{I}_n - \boldsymbol{\Lambda}$ for NA structure errors defined by (14). By 1000 replications, we obtain the confidence intervals for kernel-type density estimators $f_n(x)$ and $\hat{f}_n(x)$ at the confidence level $1-\alpha = 0.95$, which are shown in Figs 1-4.

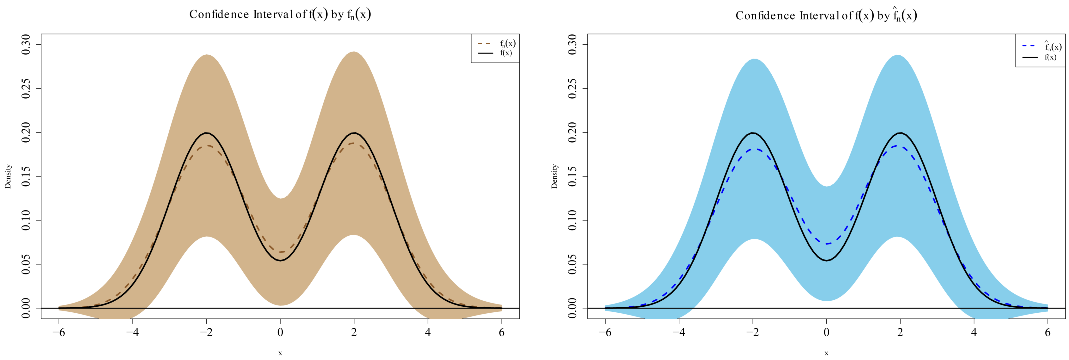

Figure 1. 95% confidence intervals and fitted curves for $f(x)$ by $f_n(x)$ and $\hat{f}_n(x)$ in the stationary model with $\rho = 0.8$, NA errors and $n = 50$.

From the Figs 1-4, it can be seen that the confidence intervals of $f(x)$ by $f_n(x)$ and $\hat{f}_n(x)$ will narrow as sample n increases. Meanwhile, the fitted curves of $f_n(x)$ and $\hat{f}_n(x)$ will approach to the true function $f(x)$ as sample n increases. In addition, the fitted curves of $f_n(x)$ and $\hat{f}_n(x)$ are close to the true curve $f(x)$. Thus, in stationary model, the residual density estimator $\hat{f}_n(x)$ can replace density "estimator" $f_n(x)$. Similarly, in the explosive model (12) with $\rho = 1.1$, Figs 5-8 show the confidence intervals of $f(x)$ by $f_n(x)$ and $\hat{f}_n(x)$ at the confidence level $1-\alpha = 0.95$.

By Figs 5-8, it can be also seen that the confidence intervals of $f(x)$ by $f_n(x)$ and $\hat{f}_n(x)$ will narrow as sample n increases. Meanwhile, the fitted curves of $f_n(x)$ and $\hat{f}_n(x)$ will approach to the true function $f(x)$ as sample n increases. Thus, in explosive model, the residual density estimator $\hat{f}_n(x)$ is a effective estimator and can replace density "estimator" $f_n(x)$. In addition, we take different $\beta = [0.1, 0.2]$ in $\boldsymbol{\Sigma}_n = \boldsymbol{I}_n - \boldsymbol{\Lambda}$ for NA structure errors and $\beta = [0.1, 0.2, 0.3]$ in

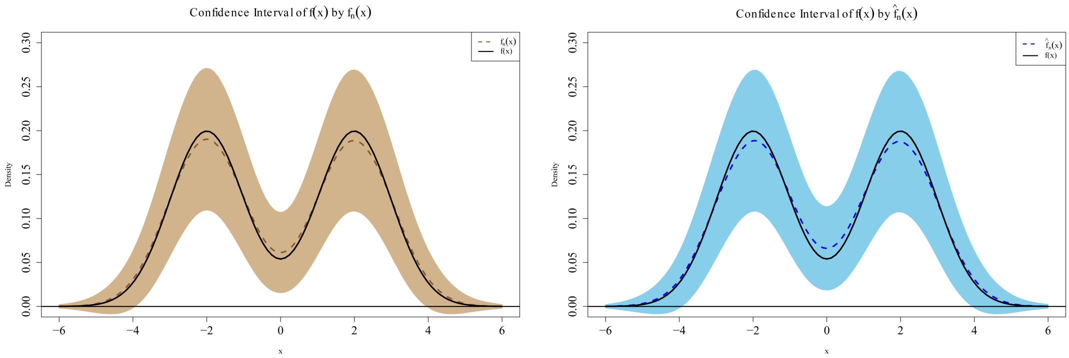

Figure 2. 95% confidence intervals and fitted curves for $f(x)$ by $f_n(x)$ and $\hat{f}_n(x)$ in the stationary model with $\rho = 0.8$, NA errors and $n = 100$.

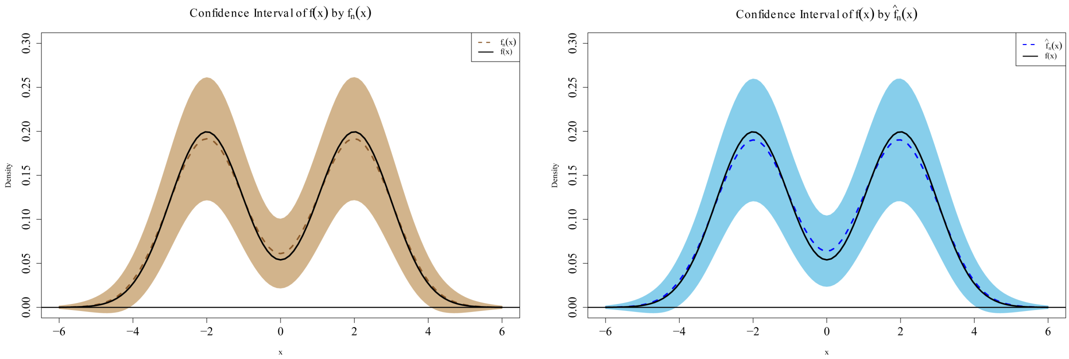

Figure 3. 95% confidence intervals and fitted curves for $f(x)$ by $f_n(x)$ and $\hat{f}_n(x)$ in the stationary with $\rho = 0.8$, with NA errors and $n = 150$.

$\boldsymbol{\Sigma}_n = \boldsymbol{I}_n + \boldsymbol{\Lambda}$ for PA structure errors, and have some similar results of Figs 1-8 in stationary and explosive models. Due to limitation of space, we didn't give these Figs in this paper.

To make an accurate comparison among the estimators of density function, we calculate the mean integrated square error (MISE) for the kernel-type density "estimator" $f_n(x)$ and residual kernel-type density estimator $\hat{f}_n(x)$, which are respectively defined by

$$MISE(f_n) = E\int_a^b \Big(f_n(x) - f(x)\Big)^2 dx \text{ and } MISE(\hat{f}_n) = E\int_a^b \Big(\hat{f}_n(x) - f(x)\Big)^2 dx.$$

For the different $\beta = [0.1, 0.2, 0.3]$, the errors were set NA structure with $\boldsymbol{\Sigma}_n = \boldsymbol{I}_n - \boldsymbol{\Lambda}$ and PA structure with $\boldsymbol{\Sigma}_n = \boldsymbol{I}_n + \boldsymbol{\Lambda}$. If $\rho = 0.8$, then $\{X_i\}$ is a stationary process, and if $\rho = 1.1$, $\{X_i\}$ is an explosive process. We perform 1000 replications and compute the mean of $\int_{-6}^{6}\Big(f_n(x) - f(x)\Big)^2 dx$ and $\int_{-6}^{6}\Big(\hat{f}_n(x) - f(x)\Big)^2 dx$, respectively. Thus, the $MISEs$ of $f_n(x)$ and $\hat{f}_n(x)$ is summarized in Table 1.

In Table 1 with NA errors and PA errors, we can see that both $MISE(f_n)$ and $MISE(\hat{f}_n)$

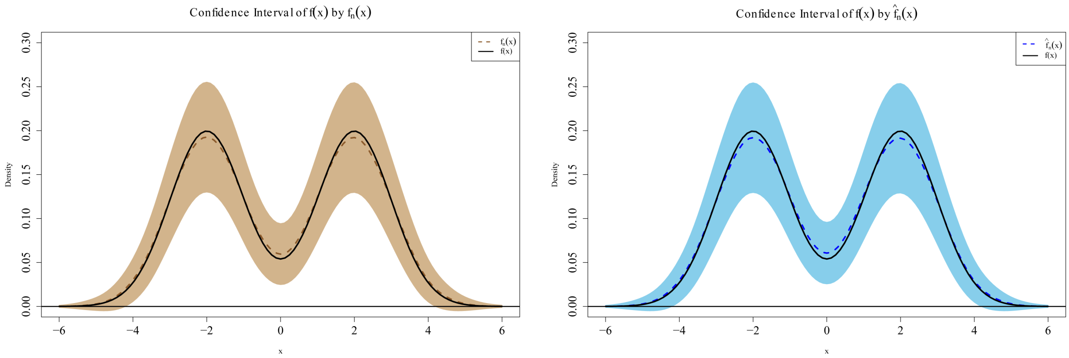

Figure 4. 95% confidence intervals and fitted curves for $f(x)$ by $f_n(x)$ and $\hat{f}_n(x)$ in the stationary with $\rho = 0.8$, with NA errors and $n = 200$.

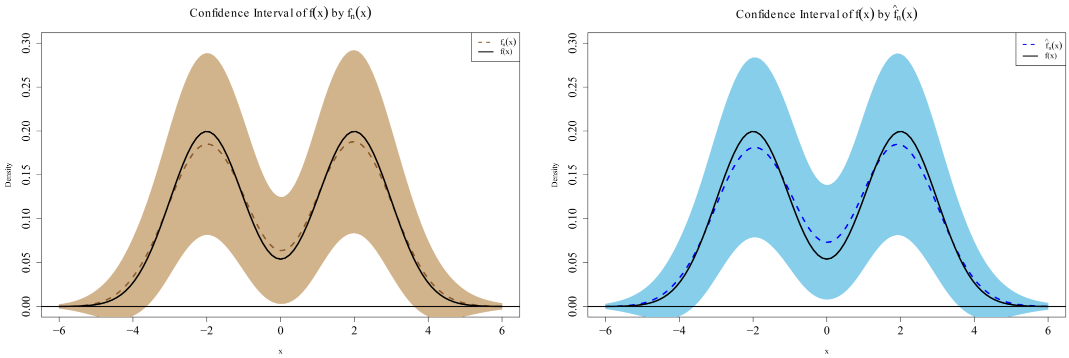

Figure 5. 95% confidence intervals and fitted curves for $f(x)$ by $f_n(x)$ and $\hat{f}_n(x)$ in the explosive model with $\rho = 1.1$, NA errors and $n = 50$.

will going to be smaller as the sample n increases in first order autoregressive model (12) with $\rho = 0.8$ and $\rho = 1.1$. Moreover, the $MISE(\hat{f}_n)$ is very close to $MISE(f_n)$. Therefore, when n is large enough, residual density estimator $\hat{f}_n(x)$ can replace density "estimator" $f_n(x)$ to make statistical inferences; in effect, if $n = 200$, then $\hat{f}_n(x)$ is very close to $f_n(x)$ in the simulation of this paper.

§5 Conclusions

In this paper, we consider the residual kernel density estimator $\hat{f}_n(x)$ in the autoregressive models (1) with $|\rho| < 1$ and $|\rho| > 1$. In practice, the errors always were not independent in the models. Thus, in our autoregressive model (1), the errors are association (NA or PA) random variables. This work extends the the range of applications of the first-order autoregressive model. We obtain the asymptotic normality results for the residual density estimator $\hat{f}_n(x)$ of error density $f(x)$ in Theorems 3.1.-3.2., which extend some results of Cheng (2005) and

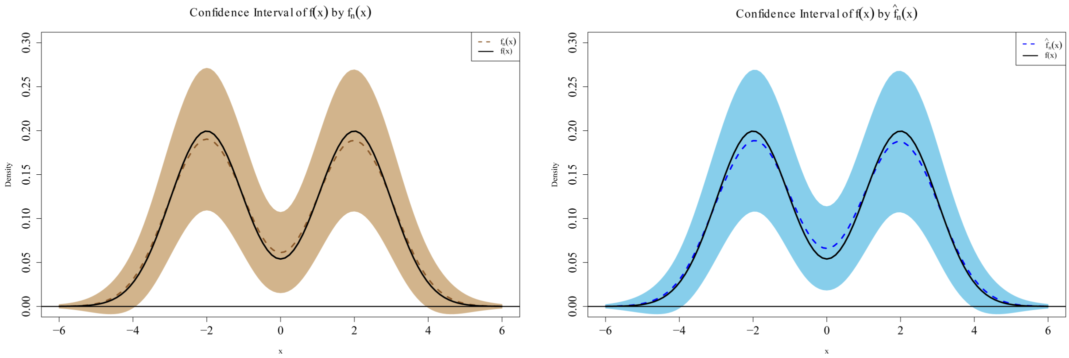

Figure 6. 95% confidence intervals and fitted curves for $f(x)$ by $f_n(x)$ and $\hat{f}_n(x)$ in the explosive model with $\rho = 1.1$, NA errors and $n = 100$.

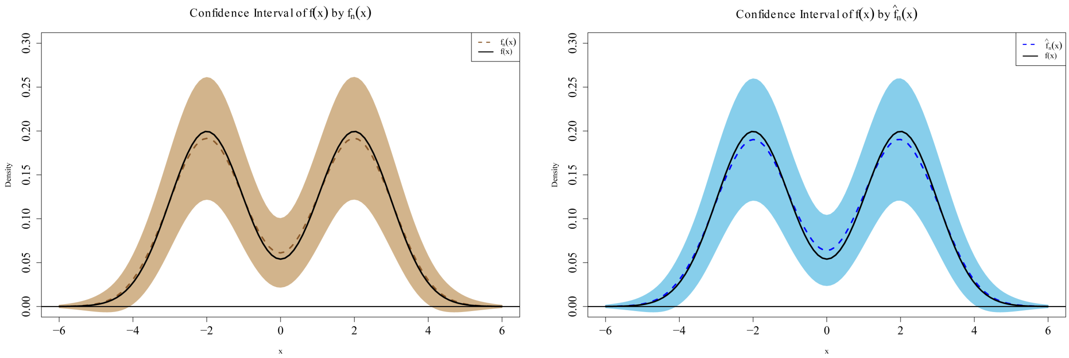

Figure 7. 95% confidence intervals and fitted curves for $f(x)$ by $f_n(x)$ and $\hat{f}_n(x)$ in the explosive model with $\rho = 1.1$, NA errors and $n = 150$.

Roussas (2000). In order to check our results, some simulations of fitted curves, confidence intervals and $MISE$ of $\hat{f}_n(x)$ and $f_n(x)$ are illustrated in Section 4. The results indicate that the residual estimator $\hat{f}_n(x)$ can replace the "estimator" $f_n(x)$. In addition, there are at least three directions to extend this work. First, Phillips and Magdalinos (2007) and Oh et al. (2018) studied the limit properties for $\hat{\rho}_n$ in the mildly explosive autoregressive model

$$X_i = \rho_n X_{i-1} + \varepsilon_i, i = 1, 2, \ldots, n,$$

where $\rho_n = c + c/b_n$, $b_n \to \infty$, $X_0 = o_P(\sqrt{b_n})$, $c \in \mathbb{R}$ and $\varepsilon_1, \ldots, \varepsilon_n$ are *i.i.d.* random errors or strictly stationary and ergodic α-mixing errors, and obtained some limit distribution for least squares estimator $\hat{\rho}_n$ defined by (6). Thus, it is interesting to study the asymptotic properties of residual density estimator $\hat{f}_n(x)$ in the mildly explosive autoregressive model with dependent errors. Second, the paper considers the first-order autoregressive models, which are fundamental model in time series analysis. Cheng and Sun (2008), Cheng (2010) and Liebscher

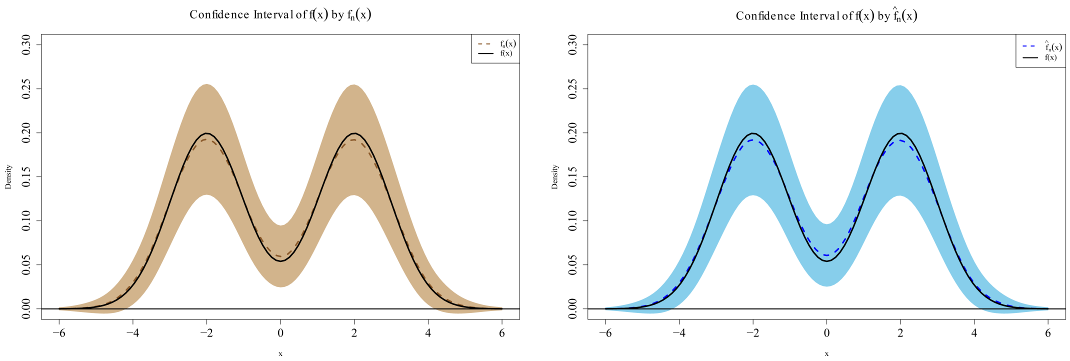

Figure 8. 95% confidence intervals and fitted curves for $f(x)$ by $f_n(x)$ and $\hat{f}_n(x)$ in the explosive model with $\rho = 1.1$, NA errors and $n = 200$.

(1999) investigated the nonlinear autoregressive time series model with *i.i.d.* errors

$$X_i = g(X_{i-1}, X_{i-2}, \ldots, X_{i-p}|\theta) + \varepsilon_i, i = p+1, \ldots, n,$$

where the autoregressive function $g : \mathbb{R}^p \to \mathbb{R}$ is a measurable function and $\theta = (\theta_1, \ldots, \theta_q)' \in \Theta \subset \mathbb{R}^q$ is the vector of parameters of the autoregressive model. Thus, the research of the residual density estimator for nonlinear autoregressive is also interesting. More challenging work remains on other time series models such as binomial AR(1) models (see Weiss and Kim (2013)), single-index varying-coefficient model (see Guo et al. (2018)) and Threshold autoregression model (see Caner and Hansen (2001)), etc. Third, the distribution function $F(x)$ of error sequence $\varepsilon_1, \ldots, \varepsilon_n$ is also fundamental function in statistics research, Cheng (2015) and Cheng (2018) studied the consistency for the residual empirical distribution function estimator $\bar{F}_n(x)$ defined by

$$\bar{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(\hat{\varepsilon}_i \le x), x \in \mathbb{R},$$

in nonlinear autoregressive model and first-order autoregressive model, where $I(\cdot)$ denotes the indicator function. On the other hand, the smooth residual distribution estimator of $F(x)$ can be defined by

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K\Big(\frac{x - \hat{\varepsilon}_i}{h_n}\Big), \quad x \in \mathbb{R},$$

where $K(\cdot)$ denotes the kernel distribution function. Thus, it is interested to study the properties of $\hat{F}_n(x)$ in various models in future research work.

§6 The proofs of main results

Lemma 6.1. *Consider the model* (1) *with* $|\rho| < 1$, *where* $\{\varepsilon_n, n \ge 1\}$ *is either NA sequence or PA sequence. Let assumptions (A1)(i), (A2)(i)(ii), (A4)(i) and (A6)(i) be satisfied. For*

Table 1. The MISE of stationary and explosive models.

	β	n	stationary model		explosive model	
			$MISE(f_n)$	$MISE(\hat{f}_n)$	$MISE(f_n)$	$MISE(\hat{f}_n)$
NA Case	0.1	50	0.0130	0.0140	0.0126	0.0143
		100	0.0076	0.0079	0.0077	0.0081
		150	0.0058	0.0060	0.0058	0.0059
		200	0.0046	0.0048	0.0048	0.0049
	0.2	50	0.0123	0.0136	0.0128	0.0142
		100	0.0077	0.0081	0.0078	0.0081
		150	0.0059	0.0062	0.0059	0.0060
		200	0.0046	0.0048	0.0046	0.0047
	0.3	50	0.0124	0.0134	0.0121	0.0141
		100	0.0073	0.0078	0.0073	0.0077
		150	0.0056	0.0058	0.0056	0.0058
		200	0.0046	0.0047	0.0046	0.0048
PA Case	0.1	50	0.0131	0.0140	0.0131	0.0146
		100	0.0080	0.0082	0.0078	0.0081
		150	0.0061	0.0063	0.0059	0.0061
		200	0.0047	0.0048	0.0049	0.0051
	0.2	50	0.0130	0.0137	0.0133	0.0143
		100	0.0080	0.0082	0.0081	0.0083
		150	0.0060	0.0060	0.0059	0.0061
		200	0.0049	0.0049	0.0050	0.0051
	0.3	50	0.0137	0.0140	0.0139	0.0152
		100	0.0085	0.0085	0.0084	0.0087
		150	0.0063	0.0063	0.0060	0.0061
		200	0.0050	0.0050	0.0050	0.0051

$x \in R$, if $nh_n^2 \to \infty$, then

$$|\hat{f}_n(x) - f_n(x)| = O_P(\frac{1}{nh_n^2}). \tag{16}$$

Proof. According to (1) and (5), it follow that

$$\hat{\varepsilon}_i - \varepsilon_i = -(\hat{\rho}_n - \rho)X_{i-1},\ 1 \le i \le n. \tag{17}$$

For $x \in \mathbb{R}$, by (4), (7) and Taylor's expansion to $K(\cdot)$, we obtain that

$$\begin{aligned} |\hat{f}_n(x) - f_n(x)| &= \Big|\frac{1}{nh_n}\sum_{i=1}^n \Big[K(\frac{x-\hat{\varepsilon}_i}{h_n}) - K(\frac{x-\varepsilon_i}{h_n})\Big]\Big| = \frac{1}{nh_n}\Big|\sum_{i=1}^n \Big[\frac{\varepsilon_i - \hat{\varepsilon}_i}{h_n}k(\eta_{ix})\Big]\Big| \\ &\le \frac{|\hat{\rho}_n - \rho|}{nh_n^2}\Big|\sum_{i=1}^n X_{i-1}k(\eta_{ix})\Big| := L_n(x), \end{aligned} \tag{18}$$

where $k(x)$ is the first-order derivative of $K(x)$ and η_{ix} is a random quantity between $(x-\hat{\varepsilon}_i)/h_n$ and $(x-\varepsilon_i)/h_n$. Now, we consider the term $\Big|\sum\limits_{i=1}^n X_{i-1}k(\eta_{ix})\Big|$ in (18). Let $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$ and denote $Y_i = k(\eta_{ix})$, $1 \le i \le n$. Since $Y_i = Y_i^+ - Y_i^-$, it has $\sum\limits_{i=1}^n X_{i-1}Y_i = \sum\limits_{i=1}^n X_{i-1}Y_i^+ - \sum\limits_{i=1}^n X_{i-1}Y_i^-$. In addition, we sort $Y_1^+, \dots, Y_n^+$ by $Y_{(1)}^+ \le \dots \le Y_{(n)}^+$ and denote the corresponding parts $X_0, \dots, X_{n-1}$ as $X_{(0)}^*, \dots, X_{(n-1)}^*$. Similarly, we sort $Y_1^-, \dots, Y_n^-$ by $Y_{(1)}^- \le \dots \le Y_n^-$ and denote the corresponding parts $X_0, \dots, X_{n-1}$ as $X_{(0)}^{**}, \dots, X_{(n-1)}^{**}$. Moreover, by

Lemma A.3. with the assumption (A2)(i), we obtain that

$$\max_{1\le k\le n}\Big|\sum_{i-1}^{k}X_{i-1}\Big|=O_P(n^{1/2}).$$

Thus, using Abel inequality(see Mitrinovic (1970)), it has

$$\begin{aligned}\Big|\sum_{i=1}^{n}X_{i-1}k\Big(\frac{x-\varepsilon_i}{h_n}\Big)\Big| &= \Big|\sum_{i=1}^{n}X_{i-1}Y_i\Big|\le\Big|\sum_{i=1}^{n}X^{*}_{(i-1)}Y^{+}_{(i)}\Big|+\Big|\sum_{i=1}^{n}X^{**}_{(i-1)}Y^{-}_{(i)}\Big|\\ &\le 5\max_{1\le k\le n}Y^{+}_{(k)}\max_{1\le k\le n}\Big|\sum_{i=1}^{k}X^{*}_{(i-1)}\Big|+5\max_{1\le k\le n}Y^{-}_{(k)}\max_{1\le k\le n}\Big|\sum_{i=1}^{k}X^{**}_{(i-1)}\Big|\\ &\le C\max_{1\le k\le n}\Big|\sum_{i=1}^{k}X_{i-1}\Big|=O_P(n^{1/2}). \qquad (19)\end{aligned}$$

Combining this and assumption (A.5)(i), we have

$$L_n(x)=O_P(\frac{1}{nh_n^2}). \qquad (20)$$

Therefore, by (18) and (20), we obtain

$$|\hat f_n(x)-f_n(x)|=O_P(\frac{1}{nh_n^2}). \qquad (21)$$

We complete the proof of Lemma 6.1. □

Lemma 6.2. *(Corollary 2.1. of Roussas (2000)) Let the assumptions (A1), (A2), (A3), (A4) and (A5) be satisfied, where $\{\varepsilon_i, i\ge 1\}$ is either NA sequence or PA sequence. Then, for $x\in\mathbb{R}$,*

$$\sqrt{nh_n}(f_n(x)-f(x))\stackrel{d}{\longrightarrow}N(0,\sigma^2(x)), \qquad (22)$$

where $\sigma^2(x)=f(x)\int_{-\infty}^{\infty}K^2(v)dv$ with $f(x)>0$.

Lemma 6.3. *Consider the model (1) with $|\rho|>1$, where $\{\varepsilon_i, i\ge 1\}$ is either NA sequence or PA sequence. Let the assumption (A1)(i), (A2)(i)(ii), (A4)(i) and (A6)(i)* be satisfied. If $\frac{\log n}{nh_n}\to 0$, then, for $x\in\mathbb{R}$,*

$$|\hat f_n(x)-f_n(x)|=O_P(\frac{\log n}{nh_n}). \qquad (23)$$

Proof. Let $a_n=\lfloor a\log n\rfloor$ with some large enough positive constant a, where $\lfloor x\rfloor$ denotes the largest integer not exceeding x. According to (4) and (7), we have

$$\begin{aligned}|\hat f_n(x)-f_n(x)|&\le\frac{1}{nh_n}\sum_{i=1}^{n-a_n}\Big|K(\frac{x-\hat\varepsilon_i}{h_n})-K(\frac{x-\varepsilon_i}{h_n})\Big|+\frac{1}{nh_n}\sum_{i=n-a_n+1}^{n}\Big|K(\frac{x-\hat\varepsilon_i}{h_n})-K(\frac{x-\varepsilon_i}{h_n})\Big|\\ &:=H_{n1}(x)+H_{n2}(x). \qquad (24)\end{aligned}$$

Now, we shall prove the convergence rates of two terms on the right-hand side of (24). By the boundness of $k(\cdot)$ in (A2)(ii), (17) and (2), it is easy to establish that

$$\begin{aligned}H_{n1}(x)&=\frac{1}{nh_n}\sum_{i=1}^{n-a_n}\Big|K(\frac{x-\hat\varepsilon_i}{h_n})-K(\frac{x-\varepsilon_i}{h_n})\Big|\le\frac{C}{nh_n^2}\sum_{i=1}^{n-a_n}|\hat\varepsilon_i-\varepsilon_i|\le\frac{C|\hat\rho-\rho|}{nh_n^2}\sum_{i=1}^{n-a_n}|X_{i-1}|\\ &\le\frac{C|\rho|^n|\hat\rho_n-\rho|}{nh_n^2|\rho|^{a_n}}\Big(\frac{1}{|\rho|^{n-a_n}}\sum_{i=1}^{n-a_n}\Big|\sum_{j=0}^{i-1}\rho^j\varepsilon_{i-1-j}\Big|\Big).\end{aligned}$$

By Assumption (A1)(i), (A4)(i) and $|\rho|>1$, we obtain that

$$E\Big(\frac{1}{|\rho|^{n-a_n}}\sum_{i=1}^{n-a_n}\Big|\sum_{j=0}^{i-1}\rho^j\varepsilon_{i-1-j}\Big|\Big)\le\frac{1}{|\rho|^{n-a_n}}\sum_{i=1}^{n-a_n}\sum_{j=0}^{i-1}|\rho^j|E|\varepsilon_{i-1-j}|\le\frac{C_1|\rho|^{n-a_n}}{|\rho|^{n-a_n}}=O(1)$$

which implies

$$\frac{1}{|\rho|^{n-a_n}}\sum_{i=1}^{n-a_n}|\sum_{j=0}^{i-1}\rho^j\varepsilon_{i-1-j}|=O_P(1).$$

Consequently, by (A6)(i)* and $|\rho|>1$, it follow that

$$H_{n1}(x)\quad=\quad O_P(\frac{1}{nh_n^2|\rho|^{a_n}}).\tag{25}$$

Next, we estimate the term of $H_{n2}(x)$ in (24). By the bounded of $K(\cdot)$ in (A2)(i) and $a_n=\lfloor a\log n\rfloor$, it follows that

$$H_{n2}(x)\le\frac{1}{nh_n}\sum_{i=n-a_n+1}^{n}\Big(\Big|K(\frac{x-\hat{\varepsilon}_i}{h_n})\Big|+\Big|K(\frac{x-\varepsilon_i}{h_n})\Big|\Big)=O(\frac{a_n}{nh_n}),\ a.s.\tag{26}$$

Finally, using the (24) - (26), $|\rho|>1$ and $a_n=\lfloor a\log n\rfloor$ with large enough positive constant a, we obtain that

$$|\hat{f}_n(x)-f_n(x)|=O_P(\frac{1}{nh_n^2|\rho|^{a_n}})+O(\frac{\log n}{nh_n})=O_P(\frac{\log n}{nh_n}).$$

This complete the proof of Lemma 6.3. □

Proof of Theorem 3.1. For $x\in\mathbb{R}$, we decompose the difference $\hat{f}_n(t)-f(t)$ into two parts,

$$\sqrt{nh_n}(\hat{f}_n(x)-f(x))=\sqrt{nh_n}(\hat{f}_n(x)-f_n(x))+\sqrt{nh_n}(f_n(x)-f(x)).\tag{27}$$

By Lemma 6.2, we known that the second term on the right-hand side of (27) converges in distribution to a normal random variable. Thus, we only need to prove that $\sqrt{nh_n}(\hat{f}_n(x)-f_n(x))$ converges to 0 in probability. Combining assumption (A3) and Lemma 6.1, it follow that

$$\sqrt{nh_n}|\hat{f}_n(x)-f_n(x)|=O_P(\frac{1}{\sqrt{nh_n^3}})=o_P(1),\tag{28}$$

since $nh_n^3\to\infty$. Therefore, by Slutsky's theorem and Lemma 6.2, we obtain that (8), which completes the the proof of Theorem 3.1. □

Proof of Theorem 3.2. Similar to the proof of Theorem 3.1. By Lemma 6.3 and $|\rho|>1$, we obtain

$$\sqrt{nh_n}|\hat{f}_n(x)-f_n(x)|=O_P(\frac{\log n}{\sqrt{nh_n}})=o_P(1).\tag{29}$$

Combining this with Lemma 6.2, we obtain that (9). □

§7 Auxiliary lemma

Lemma A.1 *(Shao (2000)). Let $p\ge1$ and $\{\varepsilon_i,i\ge1\}$ be a sequence of mean zero NA random variables with $E|\varepsilon_i|^p<\infty$ for every $1\le i\le n$. Then*

$$E\Big(\max_{1\le k\le n}\Big|\sum_{i=1}^{k}\varepsilon_i\Big|^p\Big)\quad\le\quad 2^{3-p}\sum_{i=1}^{n}E|\varepsilon_i|^p\quad for\ 1<p\le2;$$

(30)

$$E\Big(\max_{1\le k\le n}\Big|\sum_{i=1}^{k}\varepsilon_i\Big|^p\Big) \le 2(15p/\ln p)^p\Big(\sum_{i=1}^{n}E|\varepsilon_i|^p+\Big(\sum_{i=1}^{n}E\varepsilon_i^2\Big)^{p/2}\Big) \quad for\ p>2.$$

Lemma A.2 *(Newman and Wright (1981)). Let $\{\varepsilon_i, i\ge 1\}$ be centred, square-integrable and PA random variables. Then*

$$E\Big(\max_{1\le k\le n}\Big|\sum_{i=1}^{k}\varepsilon_i\Big|^2\Big)\le Var\Big(\sum_{i=1}^{n}\varepsilon_i\Big).$$

Remark A.2. Let $\{\varepsilon_i, i\ge 1\}$ be stationary, centred, square-integrable and PA random variables. Denote $u(n)=\sum_{j=n}^{\infty}\mathrm{Cov}(\varepsilon_1,\varepsilon_{j+1})$ and suppose that $u(1)<\infty$. Then

$$E\Big(\max_{1\le k\le n}\Big|\sum_{i=1}^{k}\varepsilon_i\Big|^2\Big)\le C\sum_{i=1}^{n}Var(\varepsilon_i),$$

where C is a positive constant not depending on n.

Lemma A.3. *Consider the model (1) with $|\rho|<1$. where $\{\varepsilon_i, i\ge 1\}$ is either identical distribution PA sequence or NA sequence. Let assumption (A1)(i), (A4)(i) and (A4)(ii) be satisfied. Then*

$$E\Big(\max_{1\le k\le n}\Big|\sum_{i=1}^{k}X_i\Big|^2\Big)\le Cn,$$

where C is a positive constant that does not depend on n.

Proof. The proof is inspired by Theorem 3.2 of Horváth and Zitikis (2004). By (A1)(i), (A4)(i) and (A4)(ii), it is easy to check that the conditions of maximal inequalities of NA in Lemma A.1. and PA in Remark A.2. are satisfied. Then, by (3) and Hölder inequality, $|\rho|<1$ and Lemma A.1. for NA case, we obtain that

$$\begin{aligned}E\Big(\max_{1\le k\le n}\Big|\sum_{i=1}^{k}X_i\Big|^2\Big) &= E\Big(\max_{1\le k\le n}\Big|\sum_{i=1}^{k}\sum_{j=0}^{\infty}\rho^j\varepsilon_{i-j}\Big|^2\Big)=E\Big(\max_{1\le k\le n}\Big|\sum_{j=0}^{\infty}\rho^j\sum_{i=1}^{k}\varepsilon_{i-j}\Big|^2\Big)\\ &\le E\Big(\Big(\sum_{j=0}^{\infty}|\rho|^{j/2}|\rho|^{j/2}\max_{1\le k\le n}|\sum_{i=1}^{k}\varepsilon_{i-j}|\Big)^2\Big)\\ &\le \Big(\sum_{j=0}^{\infty}|\rho|^j\Big)\Big(\sum_{j=0}^{\infty}|\rho|^jE\Big|\max_{1\le k\le n}\sum_{i=1}^{k}\varepsilon_{i-j}\Big|^2\Big)\le Cn.\end{aligned}$$

The proof of PA case is similar, so we omit its details here. □

Acknowledgement

The authors are deeply grateful to editor and anonymous referees for their careful reading and insightful comments.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

[1] A Aue, L Horváth. *A limit theorem for mildly explosive autoregressions with stable errors*, Econometric Theory, 2007, 23(2): 201-220.

[2] P Bickel, M Rosenblatt. *On some global measures of the deviations of density function estimates*, Ann Statist, 1973, 1(6): 1071-1095.

[3] P Brockwell, R Davis. *Time Series: Theory and Methods*, Spring-Verlag, New York, 1991.

[4] A Bulinski, A Shaskin. *Limit theorems for associated random fields and related systems*, World Scientific, Singapore, 2007.

[5] M Caner, B Hansen. *Threshold autoregression with a unit root*, Econometrica, 2001, 69(6): 1555-1596.

[6] J Castellana, M Leadbetter. *On smoothed probability density estimation for stationary processes*, Stochastic Process Appl, 1986, 21(2): 179-193.

[7] Y Chen. *A tutorial on kernel density estimation and recent advances*, Biostatistics & Epidemiology, 2017, 1: 161-187.

[8] F Cheng. *Asymptotic distributions of error density estimators in first-order autoregressive models*, Sankhyā, 2005, 67(3): 553-567.

[9] F Cheng, S Sun. *A goodness-of-fit test of the errors in nonlinear autoregressive time series models*, Statist Probab Lett, 2008, 78(1): 50-59.

[10] F Cheng. *Global property of error density estimation in nonlinear autoregressive time series models*, Stat Inference Stoch Process, 2010, 13(1): 43-53.

[11] F Cheng. *Strong consistency of the distribution estimator in the nonlinear autoregressive time series*, J Multivariate Anal, 2015, 142: 41-47.

[12] F Cheng. *Glivenko-Cantelli Theorem for the kernel error distribution estimator in the first-order autoregressive model*, Statist Probab Lett, 2018, 139: 95-102.

[13] J Esary, F Proschan, D Walkup. *Association of random variables with applications*, Ann Math Statist, 1967, 38(5): 1466-1474.

[14] J Fan, Q Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer Series in Statistics, New York, 2003.

[15] M Gao, W Yang, S Wu, W Yu. *Asymptotic normality of residual density estimator in stationary and explosive autoregressive models*, Comput Statist Data Anal, 2022, 175: 107549.

[16] C Guo, H Yang, J Lv. *Two step estimations for a single-index varying-coefficient model with longitudinal data*, Statist Papers, 2018, 59(3): 957-983.

[17] T Hastie, R Tibshirani. *Discriminant Analysis by Gaussian Mixtures*, J Roy Statist Soc Ser B, 1996, 58(1): 155-176.

[18] L Horváth, R Zitikis. *Asymptotics of L_p-norms of density estimators in first-order autoregressive models*, Statist Probab Lett, 2004, 66(2): 91-103.

[19] K Joag-Dev, F Proschan. *Negative association of random variables with application*, Ann Statist, 1983, 11(1): 286-295.

[20] H Koul, S Levental. *Weak convergence of the residual empirical process in explosive autoregression*, Ann. Statist, 1989, 17(4): 1784-1794.

[21] H Kim, Y Park. *A non-stationary integer-valued autoregressive model*, Statist Papers, 2008, 49(3): 485-502.

[22] H Kim, M Sin, O Kim. *A goodness-of-fit test of the errors in nonlinear autoregressive time series models with stationary α-mixing error terms*, ROMAI J, 2014, 10(1): 63-70.

[23] Q Li, J Rcine. *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, Princeton, 2007.

[24] E Liebscher. *Estimating the density of the residuals in autoregressive models*, Stat Inference Stoch Process, 1999, 2(2): 105-117.

[25] S Lee, S Na. *On the Bickel-Rosenblatt test for first-order autoregressive models*, Statist Probab Lett, 2002, 56(1): 23-35.

[26] Z Lu. *Asymptotic normality of kernel density estimators under dependence*, Ann Inst Statist. Math, 2001, 53(3): 447-468.

[27] E Masry. *Recursive probability density estimation for weakly dependent stationary processes*, IEEE Trans Inform Theory, 1986, 32(2): 254-267.

[28] T Magdalinos. *Mildly explosive autoregression under weak and strong dependence*, J Econometrics, 2012, 169(2): 179-187.

[29] D Mitrinovic. *Analytic Inequalities*, Springer, New York, 1970.

[30] C Newman, A Wright. *An invariance principle for certain dependent sequences*, Ann Probab, 1981, 9(4): 671-675.

[31] H Oh, S Lee, N Chan. *Mildly explosive autoregression with mixing innovations*, J Korean Statist Soc, 2018, 47(1): 41-53.

[32] P Oliveira. *Asymptotics for associated random variables*, Springer, Berlin, 2012.

[33] P Prakasa. *Associated sequences, demimartingales and nonparametric inference*, Springer, Basel, 2012.

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

[35] Y Qin, Y Li, Q Lei. *Empirical likelihood for probability density functions under negatively associated samples*, J Statist Plann Inference, 2011, 141(1): 373-381.

[36] M Rosenblatt. *A central limit theorem and a strong mixing condition*, Proc Nat Acad Sci USA, 1956, 42(1): 43-47.

[37] G G Roussas. *Positive and negative dependence with some statistical applications*, Asymptotics, Nonparametrics and Time Series, New York, 1999.

[38] G Roussas. *Asymptotic normality of the kernel estimate of a probability density function under association*, Statist Probab Lett, 2000, 50(1): 1-12.

[39] Q Shao. *A comparison theorem on moment inequalities between negatively associated and independent random variables*, J Theoret Probab, 2000, 13(2): 343-356.

[40] J Shen, T Pang. *Asymptotic inference for AR(1) panel data*, Appl Math J Chinese Univ Ser B, 2020, 35(3): 265-280.

[41] X Tan, X Yang. *A general result on precise asymptotics for linear processes of positively associated sequences*, Appl Math J Chinese Univ Ser B, 2008, 23(2): 190-196.

[42] X Wang, J Yu. *Limit theory for an explosive autoregressive process*, Econom Lett, 2015, 126: 176-180.

[43] C H Weiß, H-Y Kim. *Parameter estimation for binomial AR(1) models with applications in finance and industry*, Statist Papers, 2013, 54(3): 563-590.

[44] W Yang, S Hu. *The Berry-Esséen bounds for kernel density estimator under dependent sample*, J Inequal Appl, 2012, 2012: 287.

[45] Y Yang, D Wang, Z Zhao. *Empirical likelihood for first-order mixed integer-valued autoregressive model*, Appl Math J Chinese Univ Ser B, 2018, 33(3): 313-322.

[45] L Zhang, W Yu, X Wang. *Weak consistency for the estimators in a semiparametric regression model based on negatively associated random errors*, J Stat Comput Simul, 2022, 92(1): 19-42.

[1]School of Mathematical Sciences & School of Big Data and Statistics, Anhui University, Hefei 230601, China.

[2]School of Statistics, Shanxi University of Finance and Economics, Taiyuan, 030006, China

Email: wzyang@ahu.edu.cn