Submitted to the Annals of Statistics
arXiv: arXiv:0000.0000

ORACALLY EFFICIENT ESTIMATION OF AUTOREGRESSIVE ERROR DISTRIBUTION WITH SIMULTANEOUS CONFIDENCE BAND *

By Jiangyan Wang[1], Rong Liu[2], Fuxia Cheng[3] and Lijian Yang[1,4]

[1]*Soochow University,* [2]*University of Toledo,* [3]*Illinois State University,* [4]*Michigan State University*

Abstract We propose kernel estimator for the distribution function of unobserved errors in autoregressive time series, based on residuals computed by estimating the autoregressive coefficients with Yule-Walker method. Under mild assumptions, we establish oracle efficiency of the proposed estimator, i.e., it is asymptotically as efficient as the kernel estimator of the distribution function based on the unobserved error sequence itself. Applying result of [28], the proposed estimator is also asymptotically indistinguishable from the empirical distribution function based on the unobserved errors. A smooth simultaneous confidence band (SCB) is then constructed based on the proposed smooth distribution estimator and Kolmogorov distribution. Simulation examples support the asymptotic theory.

1. Introduction. Consider an AR(p) process $\{X_t\}_{t=-\infty}^{\infty}$ that satisfies

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t$$

in which $\{Z_t\}_{t=-\infty}^{\infty}$ are i.i.d. noises, called errors, $\mathrm{E}\, Z_t = 0, \mathrm{E}\, Z_t^2 = \sigma^2$, with probability density function (pdf) $f(z)$ and cumulative distribution function (cdf) $F(z) = \int_{-\infty}^{z} f(u)\, du$. For a positive integer k, the k-step ahead linear predictor $\hat{X}_{n+k}$ of X_{n+k}, based on a length $n+p$ realization $\{X_t\}_{t=1-p}^{n}$ up to time n, is well studied in Chapters 5 and 9 of [7]. While efficient methods are given to compute $\hat{X}_{n+k}$ and its mean squared error, prediction intervals is unavailable unless the process is Gaussian, see Section 5.4 of [7].

If $F(z)$ were known, all possible sample paths of the future observation X_{n+1} could be generated, and $P\left[F^{-1}(\alpha_1) \le X_{n+1} - \hat{X}_{n+1} \le F^{-1}(\alpha_2)\right] =$

*Supported in part by NSF award DMS 1007594, Jiangsu Specially-Appointed Professor Program SR10700111, Jiangsu Province Key-Discipline (Statistics) Program ZY107002, National Natural Science Foundation of China award 11371272, Research Fund for the Doctoral Program of Higher Education of China award 20133201110002, and Summer Fellowship of University of Toledo.

AMS 2000 subject classifications: Primary 62G15; secondary 62M10

Keywords and phrases: AR(p), bandwidth, error, kernel, oracle efficiency, residual

$\alpha_2 - \alpha_1$ for $0 < \alpha_1 < \alpha_2 < 1$. An efficient estimator $\hat{F}(z)$ of $F(z)$ can be used to construct a prediction interval $\left[\hat{X}_{n+1} + \hat{F}^{-1}(\alpha_1), \hat{X}_{n+1} + \hat{F}^{-1}(\alpha_2)\right]$ for X_{n+1}, with confidence level $\alpha_2 - \alpha_1$. It is also pointed out in [3] that knowledge of the cdf $F(z)$ can improve related bootstrapping procedures.

While asymptotically normal estimators of the error density $f(z)$ have been studied in [16], [1] and [9], consistent estimator for error distribution $F(z)$ does not exist for AR(p) model. On the other hand, such estimator has been proposed for nonparametric regression in [8], and uniformly $\sqrt{n}$-consistent estimator of error distribution for nonparametric AR(1)-ARCH(1) model in [24] and nonparametric regression model in [10] and [15]. It has been used for symmetry testing in parametric nonlinear time series by [3], and in nonparametric regression by [21], as well as test of parameter constancy in [2]. Other applications of error distribution estimation include functional estimation: [18]; testing parametric form of distribution and variance functions: [20] and [11]; testing for change-point in distribution: [22], and testing for additivity in regression: [23] and [19].

Assume for the sake of discussion that a sequence $\{Z_t\}_{t=1}^n$ of the errors were actually observed, [30], [13], [17], and more recently [28] propose a kernel distribution estimator (KDE) of $F(z)$ as

$$\tilde{F}(z) = \int_{-\infty}^{z} \tilde{f}(u)\, du = n^{-1} \sum\nolimits_{t=1}^{n} \int_{-\infty}^{z} K_h(u - Z_t)\, du, z \in \mathbb{R} \tag{1.1}$$

in which K is a kernel function, with $K_h(u) = h^{-1}K(u/h)$, and $h = h_n > 0$ is called bandwidth. It has been established in [13] for Lipschitz continuous F, and for Hölder continuous F in [28], that $\tilde{F}(z)$ is uniformly close to the empirical cdf $F_n(z)$ at a rate of $o_p\left(n^{-1/2}\right)$, thus inheriting all asymptotic properties of the latter. The general kernel smoothing results based on empirical process in [27] require that $F \in C^{(2)}(\mathbb{R})$, thus excluding distributions such as the double exponential distribution in our simulation study.

Unfortunately $\tilde{F}(z)$ is infeasible, as one observes only $\{X_t\}_{t=1-p}^n$, not $\{Z_t\}_{t=1}^n$. Denote by $\hat{\boldsymbol{\phi}}$ the Yule-Walker estimator of $\boldsymbol{\phi} = \left(\phi_1, ..., \phi_p\right)^T$, then

$$\begin{aligned} \hat{\boldsymbol{\phi}} &= \hat{\boldsymbol{\Gamma}}_p^{-1}\hat{\boldsymbol{\gamma}}_p, \hat{\boldsymbol{\Gamma}}_p = \{\hat{\gamma}(i-j)\}_{i,j=1}^p, \hat{\boldsymbol{\gamma}}_p = (\hat{\gamma}(1), ..., \hat{\gamma}(p))^T, \\ \hat{\gamma}(l) &= n^{-1} \sum\nolimits_{i=1-p}^{n-|l|} X_i X_{i+l}, l = 0, \pm 1, ..., \pm p. \end{aligned} \tag{1.2}$$

We propose to estimate $F(z)$ by a two-step plug-in estimator

$$\hat{F}(z) = \int_{-\infty}^{z} \hat{f}(u)\, du = n^{-1} \sum\nolimits_{t=1}^{n} \int_{-\infty}^{z} K_h\left(u - \hat{Z}_t\right) du, z \in \mathbb{R} \tag{1.3}$$

in which residuals $\hat{Z}_t = X_t - \hat{\phi}_1 X_{t-1} - \cdots - \hat{\phi}_p X_{t-p}, 1 \leq t \leq n$.

Denote the empirical cdf's based respectively on $\hat{Z}_t$ and Z_t as

$$\hat{F}_n(z) = n^{-1} \sum\nolimits_{t=1}^{n} I\left\{\hat{Z}_t \leq z\right\}, F_n(z) = n^{-1} \sum\nolimits_{t=1}^{n} I\left\{Z_t \leq z\right\}. \tag{1.4}$$

While $\hat{F}_n(z)$ is used for estimating $F(z)$, e.g., in [2], [3], [8], [18], [10], [20], [11], [21], [22], [23], [15], [19], [24], it is consistently shown to be less efficient than $F_n(z)$, as one Referee observes, see also subsection 4.1. Our unique innovation is proving that the smooth estimator $\hat{F}(z)$ based on residuals is asymptotically equivalent to, not less efficient than, the smooth estimator $\tilde{F}(z)$ based on errors. As the Associate Editor points out, this result depends crucially on the independence of Z_t with X_{t-r} for $r \geq 1$, ensured by the causal representation of the X_t (proof of Lemma A.4). We have also learned from a Referee that our result is related to the orthogonality between innovation density $f(z)$ and coefficient parameter $\boldsymbol{\phi}$, see e.g. [14].

Oracle efficiency of $\hat{F}(z)$ has powerful implications, as simultaneous confidence band (SCB) can be constructed for $F(z)$ over the entire real line, a natural tool for statistical inference on the global shape of $F(z)$, which does not exist in previous works. Working with smooth estimator based on residuals can be adopted to other settings such as nonparametric regression/autoregression, additive regression, functional autoregression (FAR), etc., the present paper thus serves as a first step in this direction.

Denote the distance between distribution functions as:

$$d(F_1, F_2) = \|F_1 - F_2\|_\infty = \sup_{z \in \mathbb{R}} |F_1(z) - F_2(z)|, \tag{1.5}$$

$$D_n(F_n) = d(F_n, F), \; D_n\left(\hat{F}\right) = d\left(\hat{F}, F\right), \; D_n\left(\tilde{F}\right) = d\left(\tilde{F}, F\right). \tag{1.6}$$

According to [28], $d\left(F_n, \tilde{F}\right) = o_p\left(n^{-1/2}\right)$, while it is well known that

$$P\left\{\sqrt{n} D_n(F_n) \leq Q\right\} \longrightarrow L(Q), \; as \; n \longrightarrow \infty, \tag{1.7}$$

where $L(Q)$ is the classic Kolmogorov distribution function, defined as:

$$L(Q) \equiv 1 - 2 \sum_{j=1}^{\infty} (-1)^{j-1} \exp\left(-2j^2 Q^2\right), Q > 0. \tag{1.8}$$

Table 1 displays the percentiles of $D_n(F_n)$ $(n \geq 50)$, $L^{-1}(1-\alpha)/\sqrt{n}$, critical values for the two-sided Kolmogorov-Smirnov test.

Theorem 2 entails that $d\left(\hat{F}, \tilde{F}\right) = o_p\left(n^{-1/2}\right)$, which, together with [28], lead to $\left|\sqrt{n}\left\{D_n\left(\hat{F}\right) - D_n(F_n)\right\}\right| \leq \sqrt{n}\left\{d\left(\hat{F}, \tilde{F}\right) + d\left(F_n, \tilde{F}\right)\right\} = o_p(1)$.

n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.2$
≥ 50	$1.63/\sqrt{n}$	$1.36/\sqrt{n}$	$1.22/\sqrt{n}$	$1.07/\sqrt{n}$

TABLE 1. *Critical Values of Kolmogorov-Smirnov Test*

Applying Slutzky's theorem produces a smooth asymptotic SCB by replacing $\sqrt{n}D_n(F_n)$ in (1.7) with $\sqrt{n}D_n\left(\hat{F}\right)$.

The rest of the paper is organized as follows. Main theoretical results on uniform asymptotics are given in Section 2. Data-driven implementation of procedure is described in Section 3, with simulation results presented in Section 4. Technical proofs are in the Appendix and the supplemental article [29].

2. Asymptotic results. In this section, we prove uniform closeness of estimators $\hat{F}(z)$ and $\tilde{F}(z)$ under Hölder continuity assumption on F. For integer $\nu \geq 0$ and $\beta \in (0,1]$, denote by $C^{(\nu,\beta)}(\mathbb{R})$ the space of functions whose ν-th derivative satisfies Hölder condition of order β,

$$C^{(\nu,\beta)}(\mathbb{R}) = \left\{ \phi : \mathbb{R} \to \mathbb{R} \,\middle|\, \sup_{x,y\in\mathbb{R}} \frac{\left|\phi^{(\nu)}(x) - \phi^{(\nu)}(y)\right|}{|x-y|^{\beta}} < +\infty \right\}. \tag{2.1}$$

We list some basic assumptions, where it is assumed that $\beta \in (1/3, 1]$.

(C1) *The cumulative distribution function* $F \in C^{(1,\beta)}(\mathbb{R})$, $0 < f(z) \leq C_f, \forall z \in \mathbb{R}$, *where* C_f *is a positive constant.*
(C2) *The process* $\{X_t\}_{t=-\infty}^{\infty}$ *is strictly stationary with* $\{Z_t\}_{t=-\infty}^{\infty} \sim \mathrm{IID}\left(0, \sigma^2\right)$. $\{X_t\}_{t=-\infty}^{\infty}$ *is causal, i.e.,* $\inf_{|z|\leq 1}\left|1 - \phi_1 z - \cdots - \phi_p z^p\right| > 0$.
(C3) *The univariate kernel function* $K(\cdot)$ *is a symmetric probability density, supported on* $[-1,1]$ *and* $K \in C^{(2)}(\mathbb{R})$.
(C4) *As* $n \to \infty$, $n^{-3/8} \ll h = h_n \ll n^{-\{2(1+\beta)\}^{-1}}$.
(C5) $\mathrm{E}\left|Z_t\right|^{6+3\eta} < \infty$, *for some* $\eta \in (6/5, +\infty)$.

Conditions (C2), (C5) are typical for time series. Conditions (C1), (C3), (C4) are similar to those in [28]. In particular, Condition (C4) on bandwidth h is rather different from those for constructing SCB in [4].

The infinite moving average expansion $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, t \in \mathbb{Z}$ and equation (3.3.6) of [7] ensure that there exist $C_\psi > 0, 0 < \rho_\psi < 1$, such that $\left|\psi_j\right| \leq C_\psi \rho_\psi^j, j \in \mathbb{N}$. In particular,

$$\left\{\mathrm{E}\left|X_t\right|^{6+3\eta}\right\}^{1/(6+3\eta)} \leq \sum\nolimits_{j=0}^{\infty} \left|\psi_j\right| \left\{\mathrm{E}\left|Z_{t-j}\right|^{6+3\eta}\right\}^{1/(6+3\eta)} < \infty.$$

In addition, the infinite moving average expansion and [25] ensure that there exist positive constants C_ρ and $\rho \in (0,1)$ such that $\alpha(k) \leq C_\rho \rho^k$ holds for all k, where the k-th order strong mixing coefficient of the strictly stationary process $\{X_s\}_{s=-\infty}^{\infty}$ is defined as

$$\alpha(k) = \sup_{B \in \sigma\{X_s, s \leq t\}, C \in \sigma\{X_s, s \geq t+k\}} |P(B \cap C) - P(B)P(C)|, k \geq 1.$$

Our first result concerns asymptotic uniform oracle efficiency of $\hat{F}$ given in (1.3) over intervals that grow to infinity with sample size.

THEOREM 1. *Under Conditions (C1)-(C5), the oracle estimator $\hat{F}(z)$ is asymptotically as efficient as the infeasible estimator $\tilde{F}(z)$ over $z \in [-a_n, a_n]$ where the sequence $a_n > 0$, $a_n \to \infty$, $a_n \leq C_1 n^{C_2}$ for some $C_1, C_2 > 0$, i.e., as $n \to \infty$, $\sup_{z \in [-a_n, a_n]} \left|\hat{F}(z) - \tilde{F}(z)\right| = o_p\left(n^{-1/2}\right)$.*

The above oracle efficiency of $\hat{F}$ extends to entire $\mathbb{R}$ provided that the extreme value of $\{|X_t|\}_{t=1}^{n}$ has a mild growth bound. Denote

$$M_n = \max(|X_1|, |X_2|, \cdots, |X_n|) \tag{2.2}$$

(C6) *There exists some $\gamma > 0$, such that $M_n = O_p(n^\gamma)$.*

Condition (C6) is satisfied, for instance, if the innovations $\{Z_t\}_{t=-\infty}^{\infty}$ have exponential tails. Denote by Z a random variable with the same distribution as the Z_t's, the following assumptions are as in A.1 and B.3 of [26].

(C5') *There exist constants $\sigma_z > 0$ and $\lambda > 0$, such that one of the following conditions holds: (1) $0 < \lambda \leq 1$ and for constants $C_z > 0, c_z \in \mathbb{R}$*

$$P(|\sigma_z Z| > z) \sim C_z z^{c_z} \exp\left(-z^\lambda\right), z \to +\infty; \tag{2.3}$$

(2) $\lambda > 1$ and for constants $C_{z,+}, C_{z,-} > 0, c_{z,+}, c_{z,-} \in \mathbb{R}$

$$\begin{aligned} \sigma_z^{-1} f\left(\sigma_z^{-1} z\right) &\sim C_{z,+} z^{c_{z,+}} \exp\left(-z^\lambda\right), z \to +\infty, \\ \sigma_z^{-1} f\left(\sigma_z^{-1} z\right) &\sim C_{z,-} (-z)^{c_{z,-}} \exp\left(-(-z)^\lambda\right), z \to -\infty. \end{aligned}$$

For $D(z) = \sigma_z^{-1} f\left(\sigma_z^{-1} z\right) \exp\left(|z|^\lambda\right)$ and its derivative $D'(z)$

$$D(z) \sim C_{z,+} z^{c_{z,+}}, z \to +\infty, D(z) \sim C_{z,-} (-z)^{c_{z,-}}, z \to -\infty,$$
$$\limsup_{|z| \to \infty} \left|zD'(z)/D(z)\right| < \infty.$$

Clearly the exponential tail Condition (C5') implies Condition (C5), while the next lemma establishes that it also entails Condition (C6).

LEMMA 1. *Conditions (C2), (C5') imply* $M_n = O_p\left((\log n)^{1/\lambda}\right)$.

The next Theorem 2 extends Theorem 1 with the additional Condition (C6) in general, or (C5') in particular. As pointed out by the Associate Editor, future works may lead to weaker conditions than (C5') that ensure (C6), aided by more powerful extreme value results than in [26]. We conjecture that Theorem 2 holds for functional autoregression model (FAR) as well.

THEOREM 2. *Under Conditions (C1)-(C6), the oracle estimator* $\hat{F}(z)$ *is asymptotically as efficient as the infeasible estimator* $\tilde{F}(z)$ *over* $z \in \mathbb{R}$ *i.e., as* $n \to \infty, d\left(\hat{F}, \tilde{F}\right) = \sup_{z\in\mathbb{R}} \left|\hat{F}(z) - \tilde{F}(z)\right| = o_p\left(n^{-1/2}\right)$. *Especially, the above holds under Conditions (C1)-(C4), (C5').*

By [5], as $n \to \infty$, $\sqrt{n}\left(F_n(z) - F(z)\right) \to_d B\left(F(z)\right)$, where B denotes the Brownian bridge. It is established in [28] that $d\left(F_n, \tilde{F}\right) = o_p\left(n^{-1/2}\right)$ and hence Theorem 2 provides that $d\left(\hat{F}, F_n\right) = o_p\left(n^{-1/2}\right)$, and the following

COROLLARY 1. *Under conditions of Theorem 2, as* $n \to \infty$,

$$\sqrt{n}\left(\hat{F}(z) - F(z)\right) \to_d B\left(F(z)\right).$$

For any $\alpha \in (0,1)$, $\lim_{n\to\infty} P\left\{F(z) \in \hat{F}(z) \pm L_{1-\alpha}/\sqrt{n}, z \in \mathbb{R}\right\} = 1 - \alpha$, *and a smooth SCB for* $F(z)$ *is*

$$\left[\max\left(0, \hat{F}(z) - L_{1-\alpha}/\sqrt{n}\right), \min\left(1, \hat{F}(z) + L_{1-\alpha}/\sqrt{n}\right)\right], z \in \mathbb{R}. \tag{2.4}$$

3. Implementation. We now describe steps to construct the smooth SCB in (2.4). For $n \geq 50$, the following critical values from Table 1 are used

$$L_{1-0.01} = 1.63,\ L_{1-0.05} = 1.36,\ L_{1-0.1} = 1.22,\ L_{1-0.2} = 1.07.$$

To compute $\hat{F}(z)$ in (1.3), we use the quartic kernel $K(u) = 15(1 - u^2)^2 I\{|u| \leq 1\}/16$ and a data-driven bandwidth $h = \text{IQR} \times n^{-1/3}$, with IQR denoting the Sample Inter-Quartile Range of $\left\{\hat{Z}_t\right\}_{t=1}^{n}$. This bandwidth satisfies Condition (C4) as long as the Hölder order $\beta > 1/2$. It is also similar to the robust and simple one in [28].

4. Simulation examples. In this section, we compare the performance of the estimator $\hat{F}$ with the benchmark infeasible estimator $\tilde{F}$.

For sample sizes $n = 50, 100, 500, 1000$, a total of 1000 samples $\{Z_t\}_{t=1}^n$ are generated, from the standard normal distribution and the standard double exponential distribution, both of which $\in C^{(1,1)}(\mathbb{R})$,

$$F(z) = \int_{-\infty}^{z} (2\pi)^{-1/2} e^{-u^2/2} du, \text{ or } F(z) = \begin{cases} 1 - 1/2 \exp(-z), & z \geq 0 \\ 1/2 \exp(z), & z < 0 \end{cases},$$

hence one would expect the data-driven bandwidth described in Section 3 to perform well. We present results only for Case 1: standard normal distributon with AR(1) model, and Case 2: standard double exponential distribution with AR(2) model. Other combinations of error distributions and AR models have yielded similar results which are omitted to save space.

4.1. *Global errors.* In this subsection, we examine the global errors of $\tilde{F}$ and $\hat{F}$, measured by the maximal deviations $D_n\left(\hat{F}\right), D_n\left(\tilde{F}\right)$ defined in (1.6), and the Mean Integrated Squared Error (MISE) defined as:

$$\begin{aligned} \text{MISE}\left(\hat{F}\right) &= \text{E} \int \left\{\hat{F}(z) - F(z)\right\}^2 dz, \\ \text{MISE}\left(\tilde{F}\right) &= \text{E} \int \left\{\tilde{F}(z) - F(z)\right\}^2 dz. \end{aligned} \tag{4.1}$$

Of interests are the means $\bar{D}_n(\hat{F})$ and $\bar{D}_n(\tilde{F})$ of $D_n(\hat{F})$ and $D_n(\tilde{F})$ over the 1000 replications, and similar means for MISE$(\hat{F})$ and MISE$(\tilde{F})$ for Case 1. Table 2 contains these values, while Figure 1 is created based on the ratios $D_n(\hat{F})/D_n(\tilde{F})$ with four sets of coefficients. Both show that as n increases, both deterministic ratios $\bar{D}_n(\hat{F})/\bar{D}_n(\tilde{F})$ and $\overline{\text{MISE}}(\hat{F})/\overline{\text{MISE}}(\tilde{F}) \to 1$, while the random ratio $D_n(\hat{F})/D_n(\tilde{F}) \to_p 1$, all consistent with Theorem 2.

Table 4 in Section 2 of the supplemental article [29] contains $\bar{D}_n(\hat{F})$, $\bar{D}_n(\hat{F}_n), \overline{\text{MISE}}(\hat{F}), \overline{\text{MISE}}(\hat{F}_n)$ with $\hat{F}_n$ defined in (1.4). Clearly $\hat{F}$ outperforms $\hat{F}_n$ as we have commented on page 3.

4.2. *Smooth SCBs.* In this subsection, we compare the SCBs based on smooth $\hat{F}$, and the infeasible $\tilde{F}$ and F_n for Case 2, and $1-\alpha = 0.99, 0.95, 0.90, 0.80$. Table 3 contains the coverage frequencies over 1000 replications of the SCBs. The smooth SCB is always conservative, the infeasible one more than the data-based one in all cases except a few. The nonsmooth SCB based on F_n has coverage frequencies closest to the nominal levels.

Figure 2 depicts the true F (thick), the infeasible $\tilde{F}$ (solid), the data-based $\hat{F}$ with its 90% SCB (solid), and F_n (dashed), for a data of size $n = 100$. The

three estimators are very close, with $\hat{F}$ practically distinguishable from $\tilde{F}$, consistent with our asymptotic theory. Similar patterns have been observed for larger n.

APPENDIX A: PROOFS

A.1. Preliminaries. In this Appendix, C (or c) denote any positive constants, U_p (or u_p) sequences of random variables uniformly O (or o) of certain order and by $O_{a.s.}$ (or $o_{a.s.}$) almost surely O (or o), etc.

The next two lemmas are used in the proof of Theorem 1.

LEMMA A.1. *[[6], Theorem 1.4] Let $\{\xi_t\}$ be a zero mean real valued process. Suppose that there exists $c > 0$ such that for $i = 1, \cdots, n$, $k \geq 3$, $E|\xi_i|^k \leq c^{k-2}k!E\xi_i^2 < +\infty$, $m_r = \max_{1\leq i\leq N} \|\xi_i\|_r, r \geq 2$. Then for each $n > 1$, integer $q \in [1, n/2]$, each $\varepsilon_n > 0$ and $k \geq 3$*

$$P\left\{\left|\sum\nolimits_{i=1}^{n} \xi_i\right| > n\varepsilon_n\right\} \leq a_1 \exp\left(-\frac{q\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n}\right) + a_2(k)\,\alpha\left(\left[\frac{n}{q+1}\right]\right)^{\frac{2k}{2k+1}}$$

where $a_1 = 2\dfrac{n}{q} + 2\left(1 + \dfrac{\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n}\right), a_2(k) = 11n\left(1 + \dfrac{5m_k^{2k/(2k+1)}}{\varepsilon_n}\right).$

LEMMA A.2. *[[7], Theorem 8.1.1] The Yule-Walker estimator $\hat{\boldsymbol{\phi}} = \left(\hat{\phi}_1, ..., \hat{\phi}_p\right)^T$ of $\boldsymbol{\phi} = \left(\phi_1, ..., \phi_p\right)^T$ satisfies $n^{1/2}\left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right) \to N\left(0, \sigma^2\Gamma_p^{-1}\right)$, where Γ_p is the covariance matrix $[\gamma(i-j)]_{i,j=1}^p$ with $\gamma(h) = \operatorname{cov}(X_t, X_{t+h})$ for the causal AR(p) process $\{X_t\}$.*

A.2. Proof of Theorem 1.

LEMMA A.3. *Under Conditions (C4) and (C5), there exists an $a > 0$, such that the following are fulfilled for the sequence $\{D_n\} = \{n^a\}$*

(A.1)
$$\sum\nolimits_{n=1}^{\infty} D_n^{-(2+\eta)} < \infty, D_n^{-(1+\eta)} n^{1/2} h^{1/2} \to 0, D_n n^{-1/2} h^{-1/2} (\log n) \to 0.$$

LEMMA A.4. *Under Conditions (C1) - (C5), for any $1 \leq r$, s, $v \leq p$*

$$\sup\nolimits_{|z|\leq a_n} \left|n^{-1} \sum\nolimits_{t=1}^{n} K_h(z - Z_t) X_{t-r}\right| = O_{a.s.}\left(n^{-1/2} h^{-1/2} \log n\right). \tag{A.2}$$

LEMMA A.5. *Under Conditions (C1) - (C5), for any $1 \leq r, s, v \leq p$*

$$\sup\nolimits_{|z|\leq a_n} \left|n^{-1} \sum\nolimits_{t=1}^{n} K_h'(z - Z_t) X_{t-r} X_{t-s}\right| = O_p(1), \tag{A.3}$$

$$\sup\nolimits_{|z|\leq a_n} \left|n^{-1} \sum\nolimits_{t=1}^{n} K_h''(z - Z_t) X_{t-r} X_{t-s} X_{t-v}\right| = O_p(1). \tag{A.4}$$

LEMMA A.6. *Under Conditions (C1) - (C5), as $n \to \infty$,*

$$n^{-1}\sum\nolimits_{t=1}^{n}\left|X_{t-r}X_{t-s}X_{t-v}X_{t-w}\right| = O_p\left(1\right), 1 \le r,s,v,w \le p.$$

The proofs of Lemmas A.3-A.6 are in the supplemental article [29].

Proof of Theorem 1.

Recall the definition of $\hat{Z}_t$ and Z_t in Section 1, one has

(A.5)

$$\hat{F}\left(z\right)-\tilde{F}\left(z\right) = n^{-1}\sum\nolimits_{t=1}^{n}\int_{(z-Z_t)/h}^{(z-\hat{Z}_t)/h} K\left(v\right)dv = n^{-1}\sum\nolimits_{t=1}^{n}\left\{G\left(\tfrac{z-\hat{Z}_t}{h}\right) - G\left(\tfrac{z-Z_t}{h}\right)\right\}$$

where $G\left(z\right) = \int_{-\infty}^{z} K\left(u\right)du$. The right hand side of equation (A.5) is

$$\frac{1}{n}\sum\nolimits_{t=1}^{n}\left\{G'\left(\tfrac{z-Z_t}{h}\right)\tfrac{\hat{Z}_t-Z_t}{h} + \frac{1}{2}G''\left(\tfrac{z-Z_t}{h}\right)\left(\tfrac{\hat{Z}_t-Z_t}{h}\right)^2 + \frac{1}{6}G^{(3)}\left(\tfrac{z-Z_t}{h}\right)\left(\tfrac{\hat{Z}_t-Z_t}{h}\right)^3 + R_t\right\}.$$

Therefore $\hat{F}\left(z\right) - \tilde{F}\left(z\right) = I_1 + I_2 + I_3 + I_4$, where

$$\begin{aligned}
\text{(A.6)}\quad I_1 &= n^{-1}\sum\nolimits_{t=1}^{n} K\left(\tfrac{z-Z_t}{h}\right)\tfrac{\hat{Z}_t-Z_t}{h},\\
I_2 &= (2n)^{-1}\sum\nolimits_{t=1}^{n} K'\left(\tfrac{z-Z_t}{h}\right)\left(\tfrac{\hat{Z}_t-Z_t}{h}\right)^2,\\
I_3 &= (6n)^{-1}\sum\nolimits_{t=1}^{n} K''\left(\tfrac{z-Z_t}{h}\right)\left(\tfrac{\hat{Z}_t-Z_t}{h}\right)^3, I_4 = n^{-1}\sum\nolimits_{t=1}^{n} R_t.
\end{aligned}$$

We now bound the four parts in (A.6).

Combining (A.6), Lemmas A.2 and A.4, for $1 \le r \le p$

$$\begin{aligned}
\text{(A.7)}\qquad & \sup\nolimits_{|z|\le a_n}\left|I_1\right|\\
&= \sup\nolimits_{|z|\le a_n} n^{-1}\left|\sum\nolimits_{t=1}^{n} K\left\{\left(z-Z_t\right)/h\right\}\left\{\left(\hat{Z}_t - Z_t\right)/h\right\}\right|\\
&= O_p\left(n^{-1/2}\right)O_{a.s.}\left(n^{-1/2}h^{-1/2}\log n\right) = o_p\left(n^{-1/2}\right).
\end{aligned}$$

From (A.6), by applying Lemmas A.2 and A.5, for $1 \le r,s,v \le p$

$$\text{(A.8)}\qquad \sup\nolimits_{|z|\le a_n}\left|I_2\right| = o_p\left(n^{-1/2}\right),\ \sup\nolimits_{|z|\le a_n}\left|I_3\right| = o_p\left(n^{-1/2}\right).$$

According to (A.6), one has $\left|I_4\right| \le n^{-1}\sum_{t=1}^{n} C\left|\left(\hat{Z}_t - Z_t\right)/h\right|^4$. Thus

$$\begin{aligned}
\text{(A.9)}\qquad & \sup_{z\in\mathbb{R}}\left|I_4\right| \le C\sup_{1\le t\le n}\left|\left(\hat{Z}_t - Z_t\right)/h\right|^4\\
&\le h^{-4}p^4\left(\max\left|\phi_r - \hat{\phi}_r\right|\right)^4 \sup n^{-1}\sum\nolimits_{t=1}^{n}\left|X_{t-r}X_{t-s}X_{t-v}X_{t-w}\right|\\
&= O_p\left(n^{-2}h^{-4}\right)\times O_p\left(1\right) = o_p\left(n^{-1/2}\right), 1 \le r,s,v,w \le p
\end{aligned}$$

which holds by using Lemmas A.2 and A.6 simultaneously.

Since $\sup_{|z|\leq a_n}\left|\hat{F}(z)-\tilde{F}(z)\right|\leq\sup_{|z|\leq a_n}(|I_1|+|I_2|+|I_3|+|I_4|)$, Theorem 1 follows by (A.7), (A.8) and (A.9) automatically.

A.3. Proof of Lemma 1. Condition (C2) provides the infinite moving average expansion $X_t=\sum_{j=0}^{\infty}\psi_j Z_{t-j}, t\in\mathbb{Z}$. Define $\tilde{X}_t=\sum_{j=0}^{\infty}\left|\psi_j\right|\left|Z_{t-j}\right|$, so that $|X_t|\leq\tilde{X}_t, t\in\mathbb{Z}$. It is obvious that

$$M_n\leq\max\left(\tilde{X}_1,\tilde{X}_2,\cdots,\tilde{X}_n\right). \tag{A.10}$$

If $0<\lambda\leq 1$, then $C_\psi\rho_\psi^j=O\left(j^{-\theta}\right)$, for some $\theta>1$, $j\in\mathbb{N}$ and $\left|\psi_j\right|\geq 0$ according to Condition (C5'), thus condition A.1 of [26] is fulfilled, so Theorem 7.4 and Theorem 8.5 of [26] imply that $\max\left(\tilde{X}_1,\tilde{X}_2,\cdots,\tilde{X}_n\right)=O_p\left((\log n)^{1/\lambda}\right)$. Thus $M_n=O_p\left((\log n)^{1/\lambda}\right)$ by (A.10).

If $\lambda>1$, then $C_\psi\rho_\psi^j=O\left(j^{-\theta}\right)$, for some $\theta>\max\left\{1,2\left(1-1/\lambda\right)^{-1}\right\}$, $j\in\mathbb{N}$ according to Condition (C5'), thus condition B.3 of [26] is fulfilled. Theorem 6.1 in [26] implies that $\max\left(X_1,\cdots,X_n,-X_1,\cdots,-X_n\right)=O_p\left((\log n)^{1/\lambda}\right)$, hence $M_n=O_p\left((\log n)^{1/\lambda}\right)$.

Summarizing both scenarios, one concludes that under Conditions (C2), (C5'), $M_n=O_p\left((\log n)^{1/\lambda}\right)$, which completes the proof.

A.4. Proof of Theorem 2. As in Theorem 1, equation (A.5) implies

$$\sup_{z\in\mathbb{R}}\left|\hat{F}(z)-\tilde{F}(z)\right|\leq\sup_{z\in\mathbb{R}}(|I_1|+|I_2|+|I_3|+|I_4|)$$

where the four parts at the right hand side are different from Theorem 1 except $\sup_{z\in\mathbb{R}}|I_4|$. So it remains to give the proof of parts I_1, I_2 and I_3 under Conditions (C1)-(C6). The proof of next lemma is in the supplemental article [29], where constants η and γ are given in Conditions (C5) and (C6).

LEMMA A.7. *Under Conditions (C1)-(C6), for any* $1\leq r,s,v\leq p, a_n=h+n^\delta$, *where* $\delta>(7/4+6\gamma)(6+3\eta)^{-1}$

$$\sup_{|z|>a_n}\left|n^{-1}\sum_{t=1}^n K_h(z-Z_t)X_{t-r}\right|=O_p\left(n^{-1}\right), \tag{A.11}$$

$$\sup_{|z|>a_n}n^{-1}\left|\sum_{t=1}^n K_h'(z-Z_t)X_{t-r}X_{t-s}\right|=O_p\left(n^{-1}\right), \tag{A.12}$$

$$\sup_{|z|>a_n}n^{-1}\left|\sum_{t=1}^n K_h''(z-Z_t)X_{t-r}X_{t-s}X_{t-v}\right|=O_p\left(n^{-1}\right). \tag{A.13}$$

Theorem 2 is proved by combining Lemmas A.4, A.5, A.6 and A.7.

ACKNOWLEDGEMENTS

This work is part of the first author's dissertation under the supervision of the last author. The helpful comments from the Co Editor, the Associate Editor and two anonymous Referees are gratefully acknowledged.

SUPPLEMENTARY MATERIAL

Supplement to "Oracally efficient estimation of autoregressive error distribution with simultaneous confidence band." (doi: TBA; .pdf). This supplement contains additional technical proofs and some supporting numerical results.

REFERENCES

[1] Bachmann, D. and Dette, H. (2005). A note on the Bickel-Rosenblatt test in autoregressive time series. *Statistics & Probability Letters* **74**, 221-234.

[2] Bai, J. (1996). Testing for parametric constancy in linear regressions: An emipirical distribution function approach. *Econometrica* **64**, 597-622.

[3] Bai, J. and Ng, S. (2001). A consistent test for conditional symmetry in time series models. *Journal of Econometrics* **103**, 225-258.

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

[5] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[6] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes.* Springer-Verlag, New York.

[7] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods, 2nd ed.* Springer-Verlag, New York.

[8] Cheng, F. (2002). Consistency of error density and distribution function estimators in nonparametric regression. *Statistics & Probability Letters* **59**, 257-270.

[9] Cheng, F. (2005a). Asymptotic distributions of error density estimators in first-order autoregressive models. *Sankhyā: the Indian Journal Statistics* **67**, 553-567.

[10] Cheng, F. (2005b). Asymptotic distributions of error density and distribution function estimators in nonparametric regression. *J. Statist. Plann. Inference* **128**, 327-349.

[11] Dette, H., Neumeyer, N. and Van Keilegom, I. (2007). A new test for the parametic form of the variance function in non-parametirc regression. *J.R.Statist.Soc.B* **69**, 903-917.

[12] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer-Verlag, New York.

[13] Fernholz, L. (1991). Almost sure convergence of smoothed empirical distribution functions. *Scand J Statist.* **18**, 255-262.

[14] Huang, W. (1986). A characterization of limiting distributions of estimators in an autoregressive process. *Ann. Inst. Statist. Math.* **38**, 137-144.

[15] Kiwitt, S. and Neumeyer, N. (2012). Estimating the conditional error distribution in non-parametric regression. *Scand J Statist.* **39**, 259-281.

[16] Lee, S. and Na, S. (2002). On the Bickel-Rosenblatt test for first-order autoregressive models. *Statistics & Probability Letters* **56**, 23-35.

[17] Liu, R. and Yang, L. (2008). Kernel estimation of multivariate cumulative distribution function. *J. Nonparametr. Stat.* **20**, 661-677.

[18] Müller, U.U., Schick, A. and Wefelmeyer, W. (2004). Estimating functionals of the error distribution in parametric and nonparametric regression. *J. Nonparametr. Stat.* **16**, 525-548.

[19] Müller, U.U., Schick, A. and Wefelmeyer, W. (2012). Estimating the error distribution function in semiparametric additive regression models. *J. Statist. Plann. Inference* **142**, 552-566.

[20] Neumeyer, N., Dette, H. and Nagel, E. (2006). Bootstrap tests for the error distribution in linear and nonparametirc regression models. *Aust.N.Z.J.Stat.* **48**, 129-156.

[21] Neumeyer, N. and Dette, H. (2007). Testing for symmetric error distribution in nonparametirc regression. *Statist. Sinica* **17**, 775-795.

[22] Neumeyer, N. and Van Keilegom, I. (2009). Change-point tests for the error distribution in non-parametric regression. *Scand J Statist.* **36**, 518-541.

[23] Neumeyer, N. and Van Keilegom, I. (2012). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multivariate Anal.* **101**, 1067-1078.

[24] Neumeyer, N. and Selk, L. (2013). A note on non-parametric testing for Gaussian innovations in AR-ARCH models. *J. Time Ser. Anal.* **34**, 362-367.

[25] Pham, T. and Tran, L. (1985). Some mixing properties of time series models. *Stochastic Processes and Their Applications* **19**, 279-303.

[26] Rootzén, H. (1986). Extreme value theory for moving average processes. *The Annals of Probability* **14**, 612-652.

[27] Van der Vaart, A. (1994). Weak convergence of smoothed empirical processes. *Scand J Statist.* **21**, 501-504.

[28] Wang, J., Cheng, F. and Yang, L. (2013). Smooth simultaneous confidence bands for cumulative distribution functions. *J. Nonparametr. Stat.* **25**, 395-407.

[29] Wang, J., Liu, R., Cheng, F. and Yang, L. (2013). Supplement to "Oracally efficient estimation of autoregressive error distribution with simultaneous confidence band." **DOI: TBA**

[30] Yamato, H. (1973). Uniform convergence of an estimator of a distribution function. *Bull.Math.Statist.* **15**, 69-78.

Jiangyan Wang
Center for Advanced Statistics and Econometrics Research
Soochow University
Suzhou 215006, CHINA
E-mail: wangjiangyan2007@126.com

Rong Liu
Department of Mathematics and Statistics
University of Toledo
Toledo, OH 43606, USA
E-mail: rong.liu@utoledo.edu

Fuxia Cheng
Department of Mathematics
Illinois State University
Normal, IL 61790, USA
E-mail: fcheng@ilstu.edu

Lijian Yang
Center for Advanced Statistics and Econometrics Research
Soochow University
Suzhou 215006, CHINA
and
Department of Statistics and Probability
Michigan State University
East Lansing, MI 48824, USA
E-mail: yanglijian@suda.edu.cn

ϕ	n	$\overline{D}_n(\hat{F})$	$\overline{D}_n(\hat{F})/\overline{D}_n(\tilde{F})$	$\overline{\text{MISE}}(\hat{F})$	$\overline{\text{MISE}}(\hat{F})/\overline{\text{MISE}}(\tilde{F})$
-0.8	50	0.0857	1.0012	0.0028	0.9881
	100	0.0649	1.0034	0.0015	0.9927
	500	0.0306	0.9966	0.0003	0.9983
	1000	0.0228	1.0022	0.0002	1.0012
-0.2	50	0.0865	1.0105	0.0029	1.0213
	100	0.0646	0.9985	0.0015	0.9927
	500	0.0307	0.9997	0.0003	1.0020
	1000	0.0228	1.0027	0.0002	1.0048
0.2	50	0.0879	1.0266	0.0030	1.0769
	100	0.0648	1.0016	0.0015	1.0112
	500	0.0308	1.0001	0.0003	1.0052
	1000	0.0228	1.0016	0.0002	1.0071
0.8	50	0.0977	1.1418	0.0046	1.6205
	100	0.0688	1.0631	0.0019	1.2780
	500	0.0311	1.0122	0.0003	1.0615
	1000	0.0230	1.0094	0.0002	1.0333

TABLE 2. *Comparing $\hat{F}$ and $\tilde{F}$: standard normal distribution errors in $AR(1)$.*

ϕ	n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.2$
	50	0.998(0.995)1.000	0.992(0.973)0.991	0.976(0.929)0.980	0.944(0.858)0.950
$(-0.8,$	100	0.998(0.992)0.997	0.987(0.963)0.990	0.972(0.924)0.977	0.928(0.858)0.936
$-0.4)$	500	1.000(0.997)1.000	0.987(0.965)0.984	0.969(0.927)0.965	0.917(0.830)0.923
	1000	0.995(0.992)0.995	0.985(0.951)0.982	0.954(0.904)0.949	0.889(0.814)0.901
	50	0.994(0.995)1.000	0.981(0.973)0.991	0.956(0.929)0.980	0.928(0.858)0.950
$(0.8,$	100	0.997(0.992)0.997	0.983(0.963)0.990	0.961(0.924)0.977	0.923(0.858)0.936
$-0.4)$	500	1.000(0.997)1.000	0.982(0.965)0.984	0.966(0.927)0.965	0.914(0.830)0.923
	1000	0.995(0.992)0.995	0.981(0.951)0.982	0.950(0.904)0.949	0.888(0.814)0.901
	50	0.993(0.995)1.000	0.984(0.973)0.991	0.965(0.929)0.980	0.930(0.858)0.950
$(0.2,$	100	0.998(0.992)0.997	0.983(0.963)0.990	0.961(0.924)0.977	0.915(0.858)0.936
$-0.1)$	500	0.999(0.997)1.000	0.986(0.965)0.984	0.966(0.927)0.965	0.914(0.830)0.923
	1000	0.995(0.992)0.995	0.982(0.951)0.982	0.952(0.904)0.949	0.896(0.814)0.901
	50	0.991(0.995)1.000	0.979(0.973)0.991	0.958(0.929)0.980	0.919(0.858)0.950
$(0.2,$	100	0.997(0.992)0.997	0.978(0.963)0.990	0.951(0.924)0.977	0.915(0.858)0.936
$0.1)$	500	0.999(0.997)1.000	0.985(0.965)0.984	0.964(0.927)0.965	0.910(0.830)0.923
	1000	0.995(0.992)0.995	0.983(0.951)0.982	0.951(0.904)0.949	0.894(0.814)0.901

TABLE 3. *Coverage frequencies for $AR(2)$ model with double exponential errors: left of parentheses-$\hat{F}$; right of parentheses-$\tilde{F}$; inside the parentheses-F_n.*

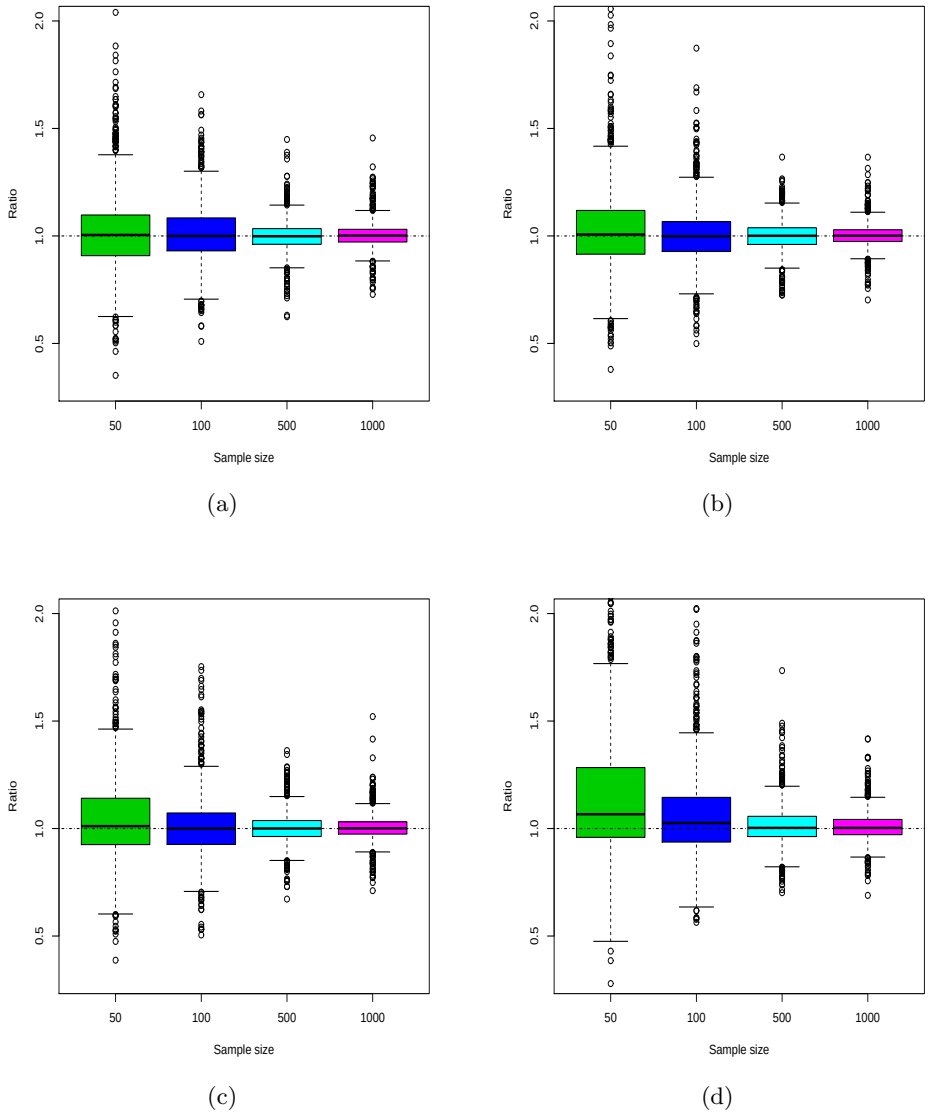

FIGURE 1. *Boxplot of the ratios* $D_n\left(\hat{F}\right)/D_n\left(\tilde{F}\right)$ *for* $AR(1)$ *model with standard normal errors. The AR coefficients of (a)-(d) are* -0.8, -0.2, 0.2, 0.8, *respectively.*

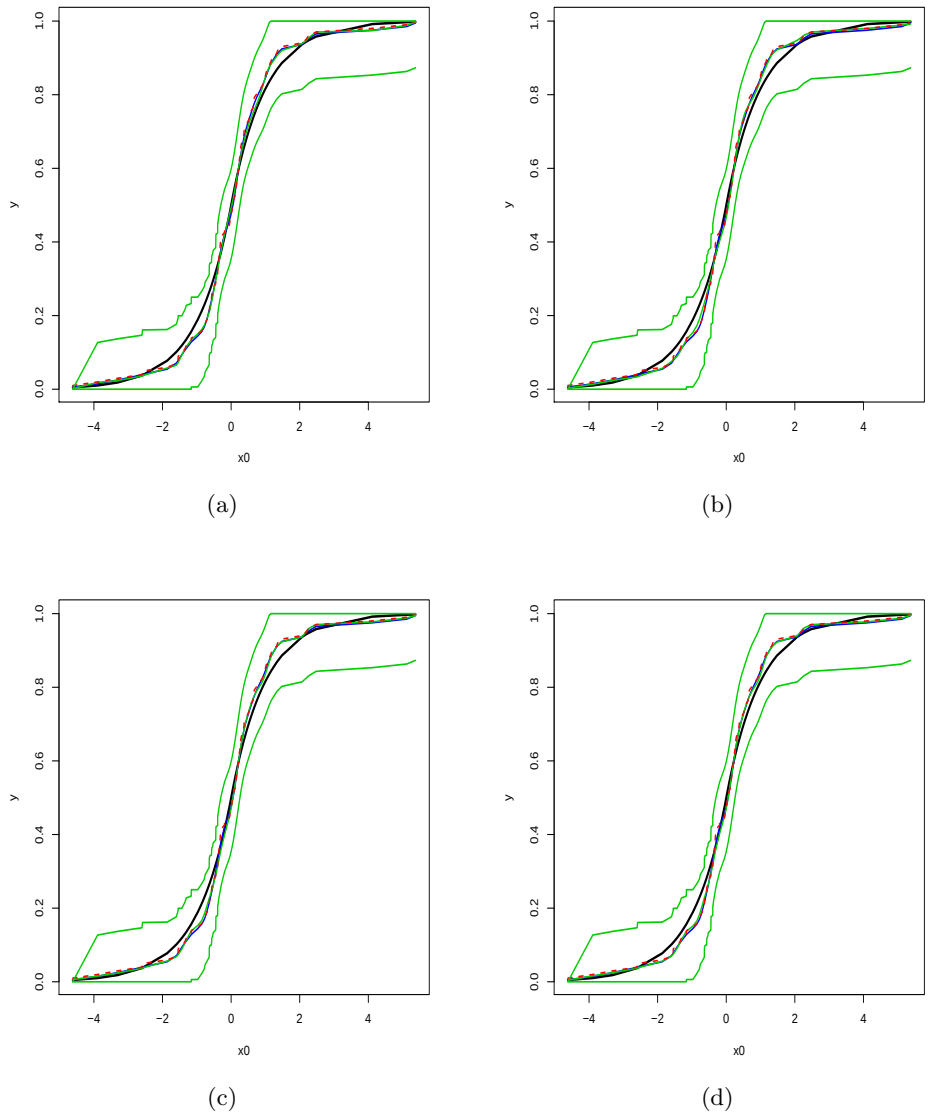

FIGURE 2. *Plots of the true cdf F (thick), the infeasible estimator $\tilde{F}$ (solid), the data-based estimator $\hat{F}$ together with its smooth* 90% *SCB (solid), and F_n (dashed) for $AR(2)$ model with $n = 100$ standard double exponential errors. The AR coefficients of (a)-(d) are* $(-0.8, -0.4)$, $(0.8, -0.4)$, $(0.2, -0.1)$, $(0.2, 0.1)$, *respectively.*