JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **39**: 690–708 (2018)
Published online 9 March 2018 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12293

ORIGINAL ARTICLE

PREDICTION INTERVAL FOR AUTOREGRESSIVE TIME SERIES VIA ORACALLY EFFICIENT ESTIMATION OF MULTI-STEP-AHEAD INNOVATION DISTRIBUTION FUNCTION

JUANJUAN KONG,[a] LIJIE GU[b†] AND LIJIAN YANG[c*]

[a] *Haimen Branch of the People's Bank of China, Haimen, China*
[b] *School of Mathematical Sciences, Soochow University, Suzhou, China*
[c] *Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing, China*

A kernel distribution estimator (KDE) is proposed for multi-step-ahead prediction error distribution of autoregressive time series, based on prediction residuals. Under general assumptions, the KDE is proved to be oracally efficient as the infeasible KDE and the empirical cumulative distribution function (cdf) based on unobserved prediction errors. Quantile estimator is obtained from the oracally efficient KDE, and prediction interval for multi-step-ahead future observation is constructed using the estimated quantiles and shown to achieve asymptotically the nominal confidence levels. Simulation examples corroborate the asymptotic theory.

Received 26 January 2017; Accepted 7 February 2018

Keywords: AR(p); non-Gaussian distribution; prediction interval; residuals; Yule–Walker estimator

JEL codes: C14; C53.

MOS subject classification: 62M20; 62G15; 62G20.

1. INTRODUCTION

Forecasting occupies a central place in the study of time series, with wide applications to economics, finance, and other disciplines, see data examples from Brockwell and Davis (1991). Consider a causal AR(p) time series $\left\{X_t\right\}_{t=-\infty}^{+\infty}$, which is a stochastic process that satisfies

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

in which the unobserved $\left\{Z_t\right\}_{t=-\infty}^{+\infty}$ are i.i.d. innovations, $\mathbb{E}Z_t = 0$, and $\mathbb{E}Z_t^2 = \sigma^2$, with pdf $f(z)$ and cdf $F(z) = \int_{-\infty}^{z} f(u)\,\mathrm{d}u$. A time series data $\left\{X_t\right\}_{t=1-p}^{n}$ consists of a length $n+p$ realization of $\left\{X_t\right\}_{t=-\infty}^{+\infty}$, and the k-step-ahead linear predictor $\tilde{X}_{n+k}^{[k]}$ for $X_{n+k}, k \geq 1$ based on $\left\{X_t\right\}_{t=1-p}^{n}$ is defined recursively by

$$\tilde{X}_{n+k}^{[k]} = \phi_1 \tilde{X}_{n+k-1}^{[k-1]} + \cdots + \phi_p \tilde{X}_{n+k-p}^{[k-p]}, \tag{1.1}$$

and satisfies

$$\tilde{X}_{n+k}^{[k]} = \phi_1^{[k]} X_n + \cdots + \phi_p^{[k]} X_{n-p+1}, \tag{1.2}$$

* Correspondence to: Lijian Yang, Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing 100084, China. E-mail: yanglijian@mail.tsinghua.edu.cn

†Co-first author.

in which the coefficient vector $\boldsymbol{\phi}^{[k]} = \left(\phi_1^{[k]}, \dots, \phi_p^{[k]}\right)^{\mathrm{T}}$ is a polynomial function g_k of $\boldsymbol{\phi} = \left(\phi_1, \dots, \phi_p\right)^{\mathrm{T}}$: $\boldsymbol{\phi}^{[k]} = g_k(\boldsymbol{\phi})$, with g_k defined recursively by repeated applications of (1.1). One would naturally ask if there exists a data-driven prediction interval (PI) for X_{n+k} based on $\left\{X_t\right\}_{t=1-p}^{n}$.

A $100(1-\alpha)\%$ k-step-ahead *normal PI* for X_{n+k} is given in Section 5.4 of Brockwell and Davis (1991) as $\tilde{X}_{n+k}^{[k]} \pm \Phi^{-1}(1-\alpha/2)\,\sigma_{[k]}$, where $\sigma_{[k]}$ is the standard deviation of the k-step-ahead prediction errors $Z_{n+k}^{[k]} = X_{n+k} - \tilde{X}_{n+k}^{[k]}$ and $\Phi^{-1}(1-\alpha/2)$ is the $(1-\alpha/2)$-quantile of the standard normal distribution. Clearly, the validity of this PI presumes the conditional distribution of $Z_{n+k}^{[k]}$ as $\mathrm{N}\left(0, \sigma_{[k]}^2\right)$, so the question remains unaddressed as to how this normal PI would perform if prediction errors $Z_{n+k}^{[k]}$ are significantly non-Gaussian.

If one had the knowledge of the k-step-ahead prediction error distribution $F^{[k]}(z)$ and obtained its $\alpha/2$th and $(1-\alpha/2)$th quantiles $q_{\alpha/2}^{[k]}$ and $q_{1-\alpha/2}^{[k]}$, then

$$\mathrm{P}\left(X_{n+k} \in \left[\tilde{X}_{n+k}^{[k]} + q_{\alpha/2}^{[k]}, \tilde{X}_{n+k}^{[k]} + q_{1-\alpha/2}^{[k]}\right]\right) = \mathrm{P}\left(Z_{n+k}^{[k]} \in \left[q_{\alpha/2}^{[k]}, q_{1-\alpha/2}^{[k]}\right]\right) = 1-\alpha$$

would lead to the following $100(1-\alpha)\%$ *infeasible PI* for X_{n+k}, which achieves the nominal confidence level

$$\left[\tilde{X}_{n+k}^{[k]} + q_{\alpha/2}^{[k]}, \tilde{X}_{n+k}^{[k]} + q_{1-\alpha/2}^{[k]}\right]. \tag{1.3}$$

This PI is termed 'infeasible' as it contains unknown quantities $q_{\alpha/2}^{[k]}, q_{1-\alpha/2}^{[k]}$ and unknown coefficients $\phi_1^{[k]}, \dots, \phi_p^{[k]}$. It does make one ask the question: What if one substitutes $q_{\alpha/2}^{[k]}, q_{1-\alpha/2}^{[k]}, \phi_1^{[k]}, \dots, \phi_p^{[k]}$ by some 'good' estimates and reconstructs a different PI?

Chapter 8 of Brockwell and Davis (1991) provides the Yule–Walker estimator $\hat{\boldsymbol{\phi}} = \left(\hat{\phi}_1, \dots, \hat{\phi}_p\right)^{\mathrm{T}}$ of $\boldsymbol{\phi}$. One then obtains a plug-in estimate $\hat{\boldsymbol{\phi}}^{[k]} = \left(\hat{\phi}_1^{[k]}, \dots, \hat{\phi}_p^{[k]}\right)^{\mathrm{T}} = g_k\left(\hat{\boldsymbol{\phi}}\right)$ of $\boldsymbol{\phi}^{[k]} = g_k(\boldsymbol{\phi})$. Define $\hat{X}_{n+k}^{[k]}$ as the data version of the linear predictor $\tilde{X}_{n+k}^{[k]}$:

$$\hat{X}_{n+k}^{[k]} = \hat{\phi}_1^{[k]} X_n + \cdots + \hat{\phi}_p^{[k]} X_{n-p+1}, \tag{1.4}$$

and $\hat{Z}_{n+k}^{[k]} = X_{n+k} - \hat{X}_{n+k}^{[k]}$ as the k-step-ahead prediction residuals. We further propose an oracle estimator $\hat{q}_{n,\alpha}^{[k]} = \left(\hat{F}^{[k]}\right)^{-1}(\alpha) = \inf\left\{z : \hat{F}^{[k]}(z) \geq \alpha\right\}$ of $q_\alpha^{[k]}$ based on a two-step plug-in kernel distribution estimator (KDE) $\hat{F}^{[k]}(z)$ of $F^{[k]}(z)$

$$\hat{F}^{[k]}(z) = \int_{-\infty}^{z} \hat{f}^{[k]}(u)\,\mathrm{d}u = n^{-1} \sum_{t=k}^{n} \int_{-\infty}^{z} K_h\left(u - \hat{Z}_t^{[k]}\right) \mathrm{d}u, z \in \mathbb{R}, \tag{1.5}$$

in which K is a kernel function, with $K_h(u) = h^{-1} K(u/h)$, $h = h_n > 0$ is called the bandwidth, and $\hat{Z}_t^{[k]} = X_t - \hat{X}_t^{[k]} = X_t - \hat{\phi}_1^{[k]} X_{t-k} - \cdots - \hat{\phi}_p^{[k]} X_{t-k-p+1}$, $k \leq t \leq n$ are the prediction residuals. A $100(1-\alpha)\%$ *oracle PI* for X_{n+k} is then

$$\left[\hat{X}_{n+k}^{[k]} + \hat{q}_{n,\alpha/2}^{[k]}, \hat{X}_{n+k}^{[k]} + \hat{q}_{n,1-\alpha/2}^{[k]}\right]. \tag{1.6}$$

The adjective '*oracle*' means that asymptotically the PI achieves the same coverage probability as the infeasible PI.

One may wonder why we have not used the more natural empirical cumulative distribution function (cdf) $\hat{F}_n^{[k]}(z) = n^{-1} \sum_{t=k}^{n} I\left(\hat{Z}_t^{[k]} \leq z\right)$ of $\left\{\hat{Z}_t^{[k]}\right\}_{t=k}^{n}$ to estimate $F^{[k]}(z)$, and based on empirical quantile $\tilde{q}_{n,\alpha}^{[k]} =$

DOI: 10.1111/jtsa.12293

Table I. The 95% PIs' coverage frequencies of X_{n+2} for AR(1) in Case 1 (normal) over 1000 replications

ϕ	n	Infeasible	Oracle	Normal	Empirical	Bootstrap
−0.8	50	0.952	0.934	0.936	0.916	0.936
	100	0.938	0.931	0.933	0.931	0.938
	500	0.952	0.950	0.952	0.951	0.967
	1000	0.957	0.954	0.955	0.954	0.943
−0.4	50	0.947	0.933	0.938	0.915	0.945
	100	0.943	0.941	0.938	0.937	0.948
	500	0.952	0.952	0.951	0.952	0.963
	1000	0.957	0.957	0.959	0.958	0.943
−0.2	50	0.952	0.926	0.940	0.918	0.945
	100	0.945	0.943	0.940	0.938	0.945
	500	0.951	0.957	0.949	0.954	0.954
	1000	0.954	0.952	0.950	0.954	0.946
0.2	50	0.944	0.928	0.941	0.918	0.939
	100	0.943	0.927	0.937	0.930	0.936
	500	0.962	0.960	0.963	0.960	0.944
	1000	0.952	0.947	0.950	0.949	0.948
0.4	50	0.947	0.920	0.941	0.913	0.935
	100	0.942	0.929	0.938	0.927	0.938
	500	0.965	0.962	0.961	0.961	0.941
	1000	0.957	0.955	0.952	0.954	0.955
0.8	50	0.954	0.924	0.928	0.913	0.912
	100	0.942	0.929	0.937	0.927	0.925
	500	0.958	0.952	0.953	0.948	0.940
	1000	0.965	0.963	0.963	0.964	0.954

$\inf\left\{z : \hat{F}_n^{[k]}(z) \geq \alpha\right\}$ to construct a $100(1-\alpha)\%$ *empirical PI* for X_{n+k} as

$$\left[\hat{X}_{n+k}^{[k]} + \tilde{q}_{n,\alpha/2}^{[k]}, \hat{X}_{n+k}^{[k]} + \tilde{q}_{n,1-\alpha/2}^{[k]}\right]. \tag{1.7}$$

In this article, we have decided against using the empirical cdf $\hat{F}_n^{[k]}(z)$ for two reasons. The aesthetic one is that $F(z)$ is smooth, while $\hat{F}_n^{[k]}(z)$ is not. More importantly, the definition of KDE $\hat{F}^{[k]}(z)$ in (1.5) allows for Taylor expansion and leads to a simpler proof of oracle efficiency, as in Wang *et al.* (2014) for the special case of $k=1$, when 1-step-ahead prediction errors $\left\{Z_t^{[1]}\right\}_{t=1}^n$ are exactly i.i.d. innovations $\left\{Z_t\right\}_{t=1}^n$. In contrast, the same oracle efficiency was established in Bai (1994) between empirical cdf of residuals and of innovations, for the same case of $k=1$, with much more tedious and longer proof. It is unclear to us whether the techniques of Bai (1994) extend to the case of $k>1$ when the k-step-ahead prediction errors $\left\{Z_t^{[k]}\right\}_{t=k}^n$ in (2.4) are no longer independent as the $\left\{Z_t^{[1]}\right\}_{t=1}^n$. We have carried out simulation studies on empirical PI as a benchmark for performance, without theoretical support. To secure an oracally efficient estimator for $F^{[k]}(z)$ leading to consistent estimates of its quantiles, the KDE $\hat{F}^{[k]}(z)$ is a better choice from both theoretical and numerical aspects.

An interesting simulation comparison can be made using normal, infeasible, oracle, and empirical PIs described above, together with the bootstrap PI proposed by Thombs and Schucany (1990). As a general principle, what one would like to have in an 'ideal' PI are the following: First of all, it needs to be accurate, i.e., the probability of the unknown quantity contained in the PI should be close to the prescribed nominal level $1-\alpha$. Second, it should be informative, that is, the interval being sufficiently narrow and therefore useful in locating the unknown quantity. With these in mind, consider an AR(2) time series with innovation distribution $F(z)$ being two normal mixtures studied in Marron and Wand (1992), for which we have coined the names 'bimodal' and 'kurtotic' because of their density shapes. Tables I–III list the empirical coverage frequencies from 1000 replications of the five PIs. Clearly, the normal PI's empirical coverage frequencies deviate considerably from the nominal coverage frequencies, which explains our motivation to propose the oracle PI as a robust device that works well regardless of the innovation distribution. A detailed comparison of the five PIs is given in Section 4.

The rest of the article is organized as follows. Section 2 gives the main asymptotically theoretical results, and Section 3 presents the steps of the implementation in detail. Section 4 reports the simulation results, some of which are relegated to the online Supporting Information. Section 5 describes the application to oil price data, while conclusions are given in Section 6. All technical proofs are in Appendix and the online Supporting Information.

DOI: 10.1111/jtsa.12293

Table II. The 95% PIs' coverage frequencies of X_{n+2} for AR(2) in Case 2 (kurtotic) over 1000 replications

(ϕ_1, ϕ_2)	n	Infeasible	Oracle	Normal	Empirical	Bootstrap
(−0.8, −0.4)	50	0.943	0.909	0.913	0.900	0.928
	100	0.958	0.943	0.939	0.942	0.949
	500	0.950	0.947	0.938	0.948	0.935
	1000	0.951	0.948	0.942	0.949	0.944
(0.8, −0.4)	50	0.944	0.900	0.904	0.894	0.925
	100	0.956	0.932	0.933	0.936	0.946
	500	0.962	0.958	0.950	0.957	0.940
	1000	0.949	0.944	0.933	0.947	0.965
(0.2, −0.1)	50	0.939	0.911	0.905	0.906	0.943
	100	0.950	0.941	0.934	0.942	0.946
	500	0.965	0.965	0.946	0.967	0.948
	1000	0.949	0.946	0.925	0.946	0.953
(−0.2, 0.1)	50	0.943	0.922	0.912	0.914	0.938
	100	0.949	0.947	0.933	0.947	0.941
	500	0.961	0.957	0.942	0.957	0.950
	1000	0.949	0.949	0.930	0.949	0.951
(0.1, −0.05)	50	0.943	0.916	0.905	0.907	0.943
	100	0.950	0.943	0.929	0.942	0.945
	500	0.964	0.965	0.945	0.965	0.954
	1000	0.947	0.950	0.922	0.950	0.956
(−0.1, 0.05)	50	0.943	0.923	0.909	0.919	0.937
	100	0.951	0.949	0.933	0.949	0.938
	500	0.963	0.958	0.943	0.958	0.952
	1000	0.948	0.943	0.927	0.945	0.953

Table III. The 95% PIs' coverage frequencies of X_{n+2} for AR(2) in Case 3 (bimodal) over 1000 replications

(ϕ_1, ϕ_2)	n	Infeasible	Oracle	Normal	Empirical	Bootstrap
(−0.8, −0.4)	50	0.948	0.928	0.963	0.899	0.937
	100	0.957	0.945	0.975	0.936	0.947
	500	0.933	0.930	0.983	0.926	0.954
	1000	0.962	0.961	0.988	0.959	0.945
(0.8, −0.4)	50	0.951	0.912	0.947	0.883	0.930
	100	0.954	0.935	0.975	0.927	0.937
	500	0.960	0.958	0.992	0.957	0.940
	1000	0.950	0.948	0.987	0.947	0.957
(0.2, −0.1)	50	0.957	0.934	0.994	0.901	0.962
	100	0.956	0.951	0.992	0.925	0.953
	500	0.961	0.962	1.000	0.957	0.941
	1000	0.954	0.957	0.996	0.953	0.951
(−0.2, 0.1)	50	0.947	0.944	0.992	0.910	0.964
	100	0.952	0.950	0.995	0.943	0.962
	500	0.937	0.946	0.997	0.939	0.956
	1000	0.965	0.966	0.997	0.965	0.947
(0.1, −0.05)	50	0.957	0.939	0.995	0.900	0.970
	100	0.959	0.957	0.997	0.939	0.966
	500	0.954	0.961	1.000	0.955	0.946
	1000	0.951	0.957	0.997	0.954	0.951
(−0.1, 0.05)	50	0.949	0.948	0.996	0.908	0.969
	100	0.952	0.955	0.997	0.939	0.968
	500	0.942	0.952	0.997	0.947	0.957
	1000	0.962	0.963	0.998	0.962	0.950

2. MAIN RESULTS

Asymptotic properties of $\hat{q}_{n,\alpha}^{[k]}$ and the oracle prediction bounds are stated. Denote by $C^{(\nu,\beta)}(\mathbb{R})$ the space of functions whose νth derivative satisfies the Hölder condition of order β where the integer $\nu \geq 0$ and $\beta \in (0, 1]$,

$$C^{(\nu,\beta)}(\mathbb{R}) = \left\{ \phi : \mathbb{R} \to \mathbb{R} \,\middle|\, \|\phi\|_{\nu,\beta} = \sup_{x_1, x_2 \in \mathbb{R}, x_1 \neq x_2} \frac{\left|\phi^{(\nu)}(x_1) - \phi^{(\nu)}(x_2)\right|}{\left|x_1 - x_2\right|^{\beta}} < +\infty \right\}.$$

The following assumptions are needed for the main Theorems 2.1, 2.2, and 2.3.

DOI: 10.1111/jtsa.12293

(A1) The cdf $F \in C^{(1,\beta)}(\mathbb{R}), \beta \in (1/3, 1], f(z) > 0, \forall z \in \mathbb{R}$, and $\sup_{z\in\mathbb{R}} f(z) = \|f\|_\infty < +\infty$.
(A2) The innovations $\left\{Z_t\right\}_{t=-\infty}^{\infty} \sim \text{IID}\left(0, \sigma^2\right)$, with $\mathbb{E}\left|Z_t\right|^{6+3\eta} = M_\eta < +\infty$, for some $\eta \in (6/5, +\infty)$.
(A3) The AR(p) process $\left\{X_t\right\}_{t=-\infty}^{\infty}$ is strictly stationary and causal, i.e., $\inf_{|z|\le 1}\left|1 - \phi_1 z - \cdots - \phi_p z^p\right| > 0$.
(A4) As $n \to \infty$, $n^{-3/8} \ll h = h_n \ll n^{-\{2(1+\beta)\}^{-1}}$.
(A5) The kernel function $K(\cdot)$ is a symmetric probability density, supported on $[-1, 1]$, and $K \in C^{(2,1)}(\mathbb{R})$, i.e., $\|K\|_{2,1} < +\infty$.

Remark 2.1. Assumptions (A2) and (A3) are standard for an AR(p) time series, see Wang *et al.* (2014). Note that Assumption (A1) and properties of convolution imply that $F^{[k]} \in C^{(1,\beta)}(\mathbb{R})$ and that the probability density function $f^{[k]}(z) = dF^{[k]}(z)/dz$ also satisfies $f^{[k]}(z) > 0, \forall z \in \mathbb{R}$, and $\sup_{z\in\mathbb{R}} f^{[k]}(z) = \left\|f^{[k]}\right\|_\infty < +\infty$.

Remark 2.2. Assumption (A3) ensures that X_t is an infinite moving average

$$X_t = \sum\nolimits_{j=0}^{\infty} \psi_j Z_{t-j}, a.s., t \in \mathbb{Z},$$
$$\sum_{j=0}^{\infty} \psi_j z^j = 1/\phi(z) = \left(1 - \phi_1 z - \cdots - \phi_p z^p\right)^{-1}, \tag{2.1}$$

while equation (3.3.6) of Brockwell and Davis (1991) ensures that

$$\left|\psi_j\right| \le C_\psi \rho_\psi^j, C_\psi > 0, 0 < \rho_\psi < 1, j \in \mathbb{N}. \tag{2.2}$$

Furthermore, the above imply that

$$\left\{\mathrm{E}\left|X_t\right|^{6+3\eta}\right\}^{1/(6+3\eta)} \le \sum\nolimits_{j=0}^{\infty} C_\psi \rho_\psi^j \left\{\mathrm{E}\left|Z_{t-j}\right|^{6+3\eta}\right\}^{1/(6+3\eta)} < \infty. \tag{2.3}$$

Remark 2.3. Theorem 3, p. 91, Doukhan (1994) and Assumptions (A1) and (A3) ensure that the induced p-dimensional Markov chain $\left\{\mathbf{Y}_t = \left(X_t, \ldots, X_{t-p+1}\right)\right\}_{t=-\infty}^{\infty}$ is geometrically β-mixing, thus the AR(p) process $\{X_t\}_{t=-\infty}^{\infty}$ itself is geometrically β-mixing as well. Since the α-mixing coefficient is no greater than the β-mixing coefficient according to Proposition 1, p. 4, Doukhan (1994), there exist constants $C_\rho > 0$ and $\rho \in (0, 1)$ such that the α-mixing coefficient $\alpha(k) \le C_\rho \rho^k$ holds for all k, where

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

According to equation (5.5.4), p.183, Brockwell and Davis (1991), the $\tilde{X}_{n+k}^{[k]}$ defined in Section 1, equation (1.2), satisfies (with $\left\{\psi_j\right\}$ defined in (2.1)):

$$\tilde{X}_{n+k}^{[k]} = \sum_{j=k}^{\infty} \psi_j Z_{n+k-j}, a.s.,$$

whereas the prediction error $Z_{n+k}^{[k]}$ is a linear combination of $\left\{Z_t\right\}_{t=n+1}^{n+k}$, i.e.

$$Z_{n+k}^{[k]} = \sum\nolimits_{j=0}^{k-1} \psi_j Z_{n+k-j} \tag{2.4}$$

DOI: 10.1111/jtsa.12293

and hence the linear predictor $\tilde{X}_{n+k}$ and the prediction noise $Z_{n+k}^{[k]}$ are independent. In particular, $\mathcal{L}\left(X_{n+k}-\tilde{X}_{n+k}^{[k]}\middle|X_n,X_{n-1},X_{n-2},...\right)=\mathcal{L}\left(Z_{n+k}^{[k]}\right)$. By the delta method, $\boldsymbol{\phi}^{[k]}$ is $\sqrt{n}$-consistent with $\hat{\boldsymbol{\phi}}^{[k]}$. Hence, $\hat{X}_{n+k}^{[k]}$ defined in (1.4) approximates $\tilde{X}_{n+k}^{[k]}$ at the rate of $\mathcal{O}_p\left(n^{-1/2}\right)$.

Assume for the sake of discussion that the error sequence $\left\{Z_t^{[k]}\right\}_{t=k}^{n}$ were actually observed, the two would-be estimators of $F^{[k]}(z)$ are the empirical cdf $F_n^{[k]}(z)$, and a KDE $\tilde{F}^{[k]}(z)$ given by

$$F_n^{[k]}(z) = n^{-1}\sum_{t=k}^{n} I\left(Z_t^{[k]}\le z\right), z\in\mathbb{R} \tag{2.5}$$

$$\tilde{F}^{[k]}(z) = \int_{-\infty}^{z}\tilde{f}^{[k]}(z)\,\mathrm{d}u = n^{-1}\sum_{t=k}^{n}\int_{-\infty}^{z}K_h\left(u-Z_t^{[k]}\right)\mathrm{d}u, z\in\mathbb{R}. \tag{2.6}$$

The integral form of the distribution estimator such as in (1.5) and (2.6) appeared in Reiss (1981) and more recently in Liu and Yang (2008) and Wang *et al.* (2013) for i.i.d. and stationary sequences. To use residuals instead of unobserved errors in KDE is the innovation of Wang *et al.* (2014), yielding efficient and smooth distribution estimators that previously did not exist.

The following Theorem 2.1 establishes the asymptotic equivalence of $\tilde{F}^{[k]}(z)$ and $F_n^{[k]}(z)$ at order $o_p\left(n^{-1/2}\right)$, similar to Wang *et al.* (2013). Although these 'estimators' of $F^{[k]}(z)$ use unobservable $\left\{Z_t^{[k]}\right\}_{t=k}^{n}$ and are therefore infeasible, they nonetheless serve as a performance benchmark for any data-driven estimator of $F^{[k]}(z)$.

Theorem 2.1. Under Assumptions (A1), (A4), and (A5), the infeasible estimator $\tilde{F}^{[k]}(z)$ given in (2.6) is as efficient as the empirical cdf $F_n^{[k]}(z)$ in (2.5) over $z\in\mathbb{R}$: that is, as $n\to\infty$, $\sup_{z\in\mathbb{R}}\left|\tilde{F}^{[k]}(z)-F_n^{[k]}(z)\right| = o_p\left(n^{-1/2}\right)$.

The next Theorem 2.2 establishes the asymptotic equivalence of $\hat{F}^{[k]}$ and $\tilde{F}^{[k]}$ up to order $o_p\left(n^{-1/2}\right)$, similar to Wang *et al.* (2014).

Theorem 2.2. Under Assumptions (A1)–(A5), the oracle estimator $\hat{F}^{[k]}(z)$ given in (1.5) is asymptotically as efficient as the infeasible estimator $\tilde{F}^{[k]}(z)$ over $z\in\mathbb{R}$: that is, as $n\to\infty$, $\sup_{z\in\mathbb{R}}\left|\hat{F}^{[k]}(z)-\tilde{F}^{[k]}(z)\right| = o_p\left(n^{-1/2}\right)$.

For any $\alpha\in(0,1)$, the next theorem (Theorem 2.3) provides the asymptotic $100(1-\alpha)\%$ oracle PI for X_{n+k}, whose implementation is described in Section 3, with data-driven predictor $\hat{X}_{n+k}^{[k]}$ is defined in (1.4).

Theorem 2.3. Under Assumptions (A1)–(A5), for any $\alpha\in(0,1)$, as $n\to\infty$

$$\mathrm{P}\left(X_{n+k}\in\left[\hat{X}_{n+k}^{[k]}+\hat{q}_{n,\alpha/2}^{[k]},\hat{X}_{n+k}^{[k]}+\hat{q}_{n,1-\alpha/2}^{[k]}\right]\right)\to 1-\alpha.$$

Proofs of these Theorems are given in Appendix.

3. IMPLEMENTATION

Based on a length $n+p$ realization $\left\{X_t\right\}_{t=1-p}^{n}$ of an AR(p) time series, one obtains first the Yule–Walker estimator $\hat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi}=\left(\phi_1,...,\phi_p\right)^{\mathrm{T}}$

$$\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\Gamma}}_p^{-1}\hat{\boldsymbol{\gamma}}_p, \hat{\boldsymbol{\Gamma}}_p = \{\hat{\gamma}(i-j)\}_{i,j=1}^{p}, \hat{\boldsymbol{\gamma}}_p = (\hat{\gamma}(1),...,\hat{\gamma}(p))^{\mathrm{T}},$$
$$\hat{\gamma}(l) = n^{-1}\sum\nolimits_{i=1-p}^{n-|l|}X_iX_{i+l}, l=0,\pm1,...,\pm p.$$

DOI: 10.1111/jtsa.12293

Then, the k-step-ahead innovations are estimated by $\hat{Z}_t^{[k]} = X_t - \hat{X}_t^{[k]} = X_t - \hat{\phi}_1^{[k]} X_{t-k} - \cdots - \hat{\phi}_p^{[k]} X_{t-k-p+1}$, $k \le t \le n$, where $\hat{\boldsymbol{\phi}}^{[k]} = \left(\hat{\phi}_1^{[k]}, ..., \hat{\phi}_p^{[k]}\right)^{\mathrm{T}} = g_k\left(\hat{\boldsymbol{\phi}}\right)$ is the plug-in Yule–Walker estimator of $g_k(\boldsymbol{\phi})$. For example, if $p = 1, k = 2$, then $\hat{Z}_t^{[2]} = X_t - \hat{\phi}_1^{[2]} X_{t-2} = X_t - \hat{\phi}_1^2 X_{t-2}$, $2 \le t \le n$, where $\hat{\phi}_1$ is the Yule–Walker estimator of ϕ_1.

The quantile estimator $\hat{q}_{n,\alpha}^{[k]} = \left(\hat{F}^{[k]}\right)^{-1}(\alpha)$ is the value of z from the 1001 equally spaced grid points from $\min\left(\left\{\hat{Z}_t^{[k]}\right\}_{t=k}^n\right) - h$ to $\max\left(\left\{\hat{Z}_t^{[k]}\right\}_{t=k}^n\right) + h$ with step size $\left\{\max\left(\left\{\hat{Z}_t^{[k]}\right\}_{t=k}^n\right) - \min\left(\left\{\hat{Z}_t^{[k]}\right\}_{t=k}^n\right) + 2h\right\}/1000$, such that $\hat{F}^{[k]}(z)$ defined by (1.5) is closest to α. The estimated cdf $\hat{F}^{[k]}(z)$ is computed with the triweight kernel $K(u) = 35\left(1 - u^2\right)^3 \times I\{|u| \le 1\}/32$, which satisfies Assumption (A5), and a data-driven bandwidth $h = \mathrm{IQR} \times n^{-1/3}$, where IQR denotes the sample interquartile range of $\left\{\hat{Z}_t^{[k]}\right\}_{t=k}^n$. As pointed out in Wang *et al.* (2014), this robust bandwidth satisfies Assumption (A4) as long as the Hölder order $\beta > 1/2$.

4. SIMULATION

4.1. Simulation Setup

In the following simulation study, three combinations of AR models and candidate error distributions $F(z)$ are involved:

Case 1 (normal). AR(1) with standard normal distribution N(0, 1);

Case 2 (kurtotic). AR(2) with kurtotic unimodal normal mixture distribution $\frac{2}{3}\mathrm{N}(0,1) + \frac{1}{3}\mathrm{N}\left(0, (\frac{1}{10})^2\right)$;

Case 3 (bimodal). AR(2) with separated bimodal normal mixture distribution $\frac{1}{2}\mathrm{N}\left(-\frac{3}{2}, (\frac{1}{2})^2\right) + \frac{1}{2}\mathrm{N}\left(\frac{3}{2}, (\frac{1}{2})^2\right)$.

Obviously, the above three candidate distributions satisfy Assumptions (A1) and (A2) on $F(z)$, and represent rather diverse density shapes as well, see Marron and Wand (1992), which introduced more details on the richness of normal mixture distributions. A number of other combinations of AR models and error distributions have been examined with similar results, and so they have been omitted.

For all examples, the sample size n is taken to be 50, 100, 500, and 1000, and realizations $\left\{X_t\right\}_{t=-999}^n$ of size $1000 + n$ are generated, with the first 1000 values thrown out to ensure strict stationarity of $\left\{X_t\right\}_{t=1}^n$.

4.2. Prediction Bounds for $X_{n+k}, k = 2, 3$

In the following, we construct 95% infeasible, oracle, normal, empirical, and bootstrap PIs mentioned in Section 1 and compare their performance over 1000 replications.

The infeasible PI: $\left[\tilde{X}_{n+k}^{[k]} + q_{0.025}^{[k]}, \tilde{X}_{n+k}^{[k]} + q_{0.975}^{[k]}\right]$ given in (1.3), based on presumed knowledge of the true error distribution $F^{[k]}(z)$;

The oracle PI: $\left[\hat{X}_{n+k}^{[k]} + \hat{q}_{n,0.025}^{[k]}, \hat{X}_{n+k}^{[k]} + \hat{q}_{n,0.975}^{[k]}\right]$ given in (1.6), by substituting $\hat{F}^{[k]}(z)$ for $F^{[k]}(z)$;

The normal PI: $\left[\hat{X}_{n+k}^{[k]} - 1.96\hat{\sigma}(k), \hat{X}_{n+k}^{[k]} + 1.96\hat{\sigma}(k)\right]$, where $\hat{\sigma}(k)$ is standard deviation of $\{\hat{Z}_t^{[k]}\}_{t=k}^n$, based on the naive presumption that $F^{[k]}(z)$ is normal.

The empirical PI: $\left[\hat{X}_{n+k}^{[k]} + \tilde{q}_{n,\alpha/2}^{[k]}, \hat{X}_{n+k}^{[k]} + \tilde{q}_{n,1-\alpha/2}^{[k]}\right]$ given in (1.7), based on empirical cdf of $\{\hat{Z}_t^{[k]}\}_{t=k}^n$.

The bootstrap PI: $\left[Q_B^{[k]}(0.025), Q_B^{[k]}(0.975)\right]$, where $Q_B^{[k]}(\alpha)$ is the αth quantile of B replicates of future values $X_{n+k}^{*1}, \ldots, X_{n+k}^{*B}$ generated by the bootstrap algorithm proposed by Thombs and Schucany (1990). The default value of B is set to 1000.

Tables I–III show the above PIs' coverage frequencies of X_{n+2}. The coverage frequencies of oracle and empirical PIs are nearly the same for large n. Although there is no significant difference of coverage frequencies between oracle and bootstrap PIs as n increases, the computing time of bootstrap PI is much longer than that of oracle PI, see Table IV. Comparison between oracle and normal PIs is another major objective. Table I is for Case 1 (normal), where the coverage frequencies of normal and oracle PIs are similar, and both become closer to that of

  J. Time Ser. Anal. **39**: 690–708 (2018)
DOI: 10.1111/jtsa.12293

Table IV. Computing time (minutes) of the 95% oracle and bootstrap PIs' coverage frequencies of X_{n+2} for AR(1) with coefficient −0.8 in Case 1 (normal) over 1000 replications

	n			
Computing time	50	100	500	1000
Oracle	0.250	0.393	1.830	3.450
Bootstrap	28.416	36.290	106.458	190.390
Bootstrap/oracle	113.664	92.341	58.174	55.186

the infeasible PI with increasing n. From Table II for Case 2 (kurtotic) and Table III for Case 3 (bimodal), one can see the coverage frequencies of normal PI are systematically lower than the nominal value of 0.95 for Case 2 and higher for Case 3, even for large sample sizes. The coverage frequencies for oracle PI, however, are consistently close to the nominal level 0.95, and become closer to those of the infeasible PI as the sample size n increases. Similar results of three cases for X_{n+3} are given in Tables S1–S3 of Supporting Information. All these observations are consistent with Theorem 2.3.

We have also created boxplots for the random ratio $\left(\hat{q}^{[k]}_{n,0.975} - \hat{q}^{[k]}_{n,0.025}\right) / \left(2 \times 1.96\hat{\sigma}_{[k]}\right)$ of the lengths of oracle and normal PIs where $k = 2$ or $k = 3$ in Figures 1–3. One observes that, as the sample size n increases, $\left(\hat{q}^{[k]}_{n,0.975} - \hat{q}^{[k]}_{n,0.025}\right) / \left(2 \times 1.96\hat{\sigma}_{[k]}\right) \overset{p}{\to} 1$ in Case 1 (normal), while for Case 2 (kurtotic) and Case 3 (bimodal) the random ratio converges to values much greater and smaller than 1 respectively. It indicates that the length of normal PI may be too narrow in Case 2 and too wide in Case 3, which partly demonstrates why the coverage frequencies of normal PI are lower or higher than the nominal level. Such significant difference in lengths provides another reason to use the smart oracle PI instead of the naive normal PI, in addition to the coverage frequencies discussed above.

Some may wonder why one would not prefer a PI that is better at catching the future true value, such as the normal PI in Case 3 (bimodal) with coverage frequencies higher than nominal. The short answer is: No, not at the price of precision. As shown in Figure 3, the normal PI for Case 3 (bimodal) is on average much wider than the oracle PI, and therefore substantially less useful in locating the whereabouts of the future value. The oracle PI is adaptive regardless of the distribution $F^{[k]}(z)$, and strikes an intelligent balance between coverage probability and precision.

5. APPLICATION

We analyze the monthly spot price series for crude oil (January 1986–December 2016), which can be downloaded from http://www.eia.gov/dnav/pet/pet_pri_spt_s1_m.htm. Based on the January 1986–January 2006 monthly oil price series, Cryer and Chan (2008) took the first difference of logarithms of the prices and further removed its mean to make the series look more stationary than original price series in Exhibit 5.4 (pp. 91), and then suggested that specifying an AR(1) model was reasonable in Exhibit 6.31 (pp. 139). Table V shows the coverage frequencies of oracle and normal prediction intervals for the last 50, 70, 100, 131 different points under two-step and three-step rolling forecasts. By comparison, the coverage frequencies of oracle PIs are much closer to the theoretical confidence level of 95%. Figure 4 shows 95% oracle and normal prediction intervals under the two-step rolling forecast for 131 difference points from February 2006 to December 2016.

6. CONCLUSIONS

In this article we have examined the infeasible, oracle, empirical, bootstrap, and normal k-step-ahead PIs for future values of the AR(p) time series. Oracle PI is shown to be theoretically optimal in the sense that it performs as well as the infeasible PI with true underlying distribution of independent innovation known and thus is better than the normal PI unless the true distribution is exactly Gaussian, a mathematical fact amply borne out in numerical

DOI: 10.1111/jtsa.12293

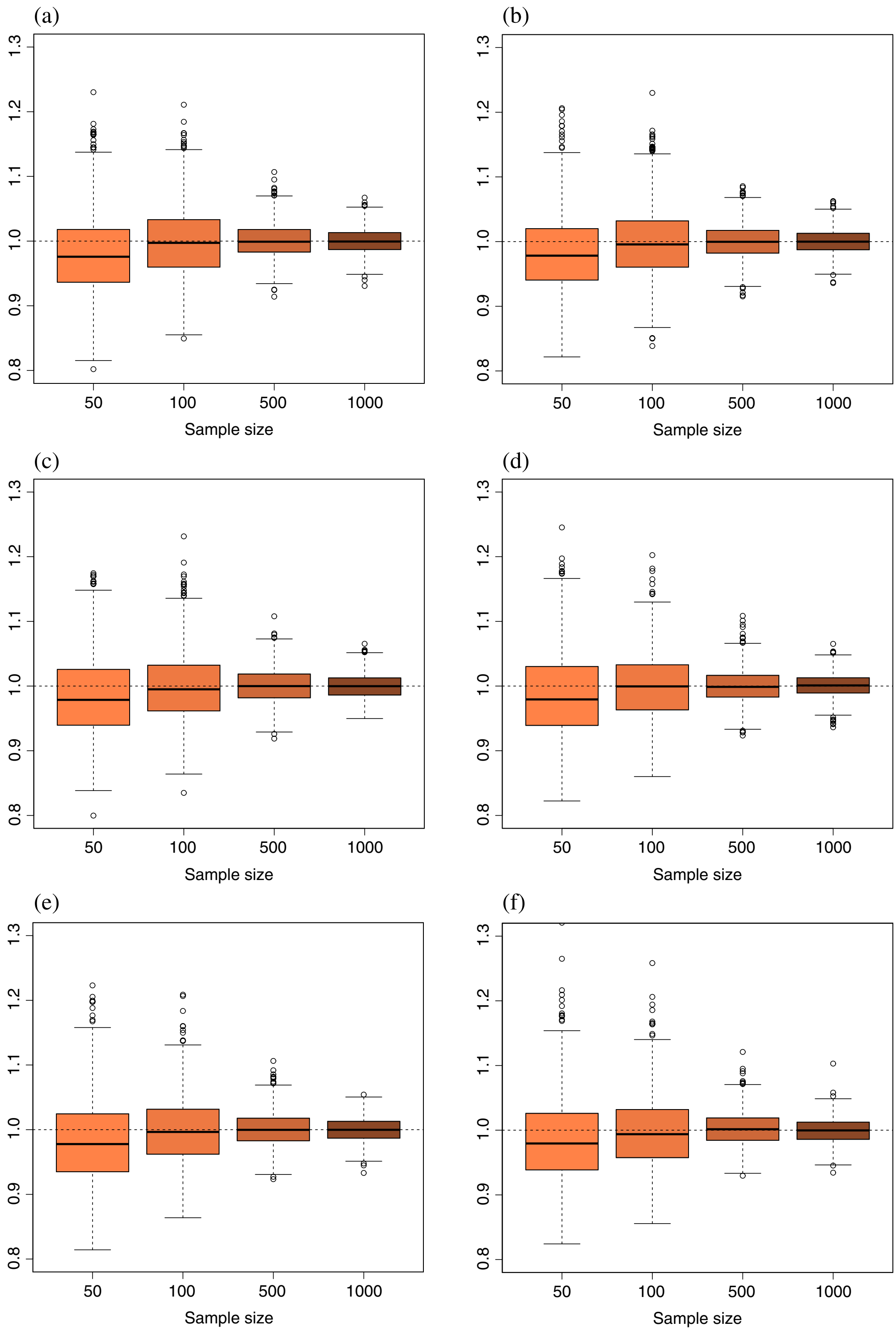

Figure 1. Boxplot of ratios $\left(\hat{q}^{[3]}_{n,0.975} - \hat{q}^{[3]}_{n,0.025}\right) / \left(2 \times 1.96\hat{\sigma}_{[3]}\right)$ for AR(1) in Case 1 (normal). The AR coefficients of (a)–(f) are −0.8, −0.4, −0.2, 0.2, 0.4, and 0.8 respectively [Color figure can be viewed at wileyonlinelibrary.com]

  J. Time Ser. Anal. **39**: 690–708 (2018)
DOI: 10.1111/jtsa.12293

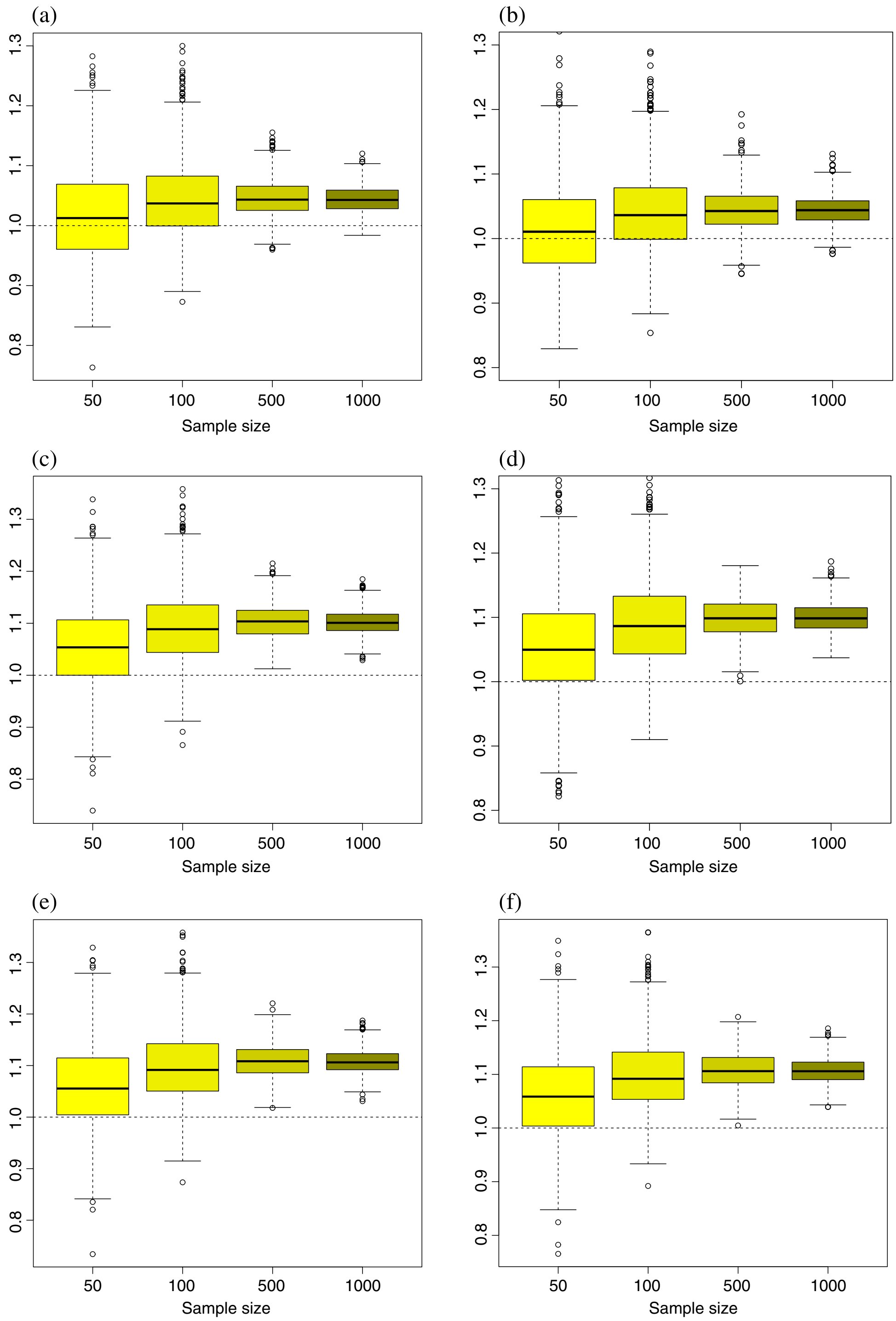

Figure 2. Boxplot of ratios $\left(\hat{q}^{[3]}_{n,0.975} - \hat{q}^{[3]}_{n,0.025}\right) / \left(2 \times 1.96\hat{\sigma}_{[3]}\right)$ for AR(2) in Case 2 (kurtotic). The AR coefficients of (a)–(f) are $(-0.8, -0.4)$, $(0.8, -0.4)$, $(0.2, -0.1)$, $(-0.2, 0.1)$, $(0.1, -0.05)$, and $(-0.1, 0.05)$ respectively [Color figure can be viewed at wileyonlinelibrary.com]

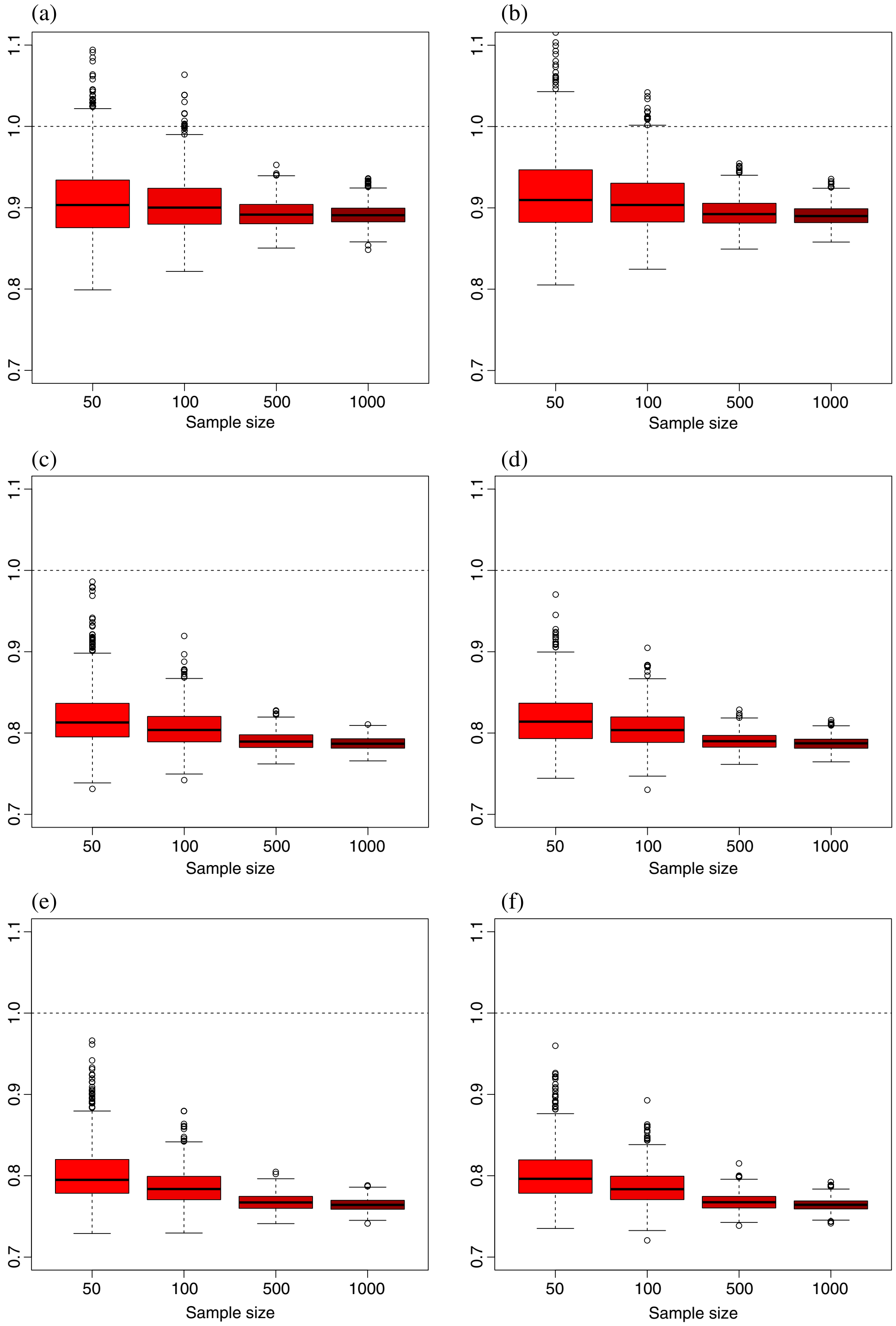

Figure 3. Boxplot of ratios $\left(\hat{q}^{[2]}_{n,0.975} - \hat{q}^{[2]}_{n,0.025}\right) / \left(2 \times 1.96\hat{\sigma}_{[2]}\right)$ for AR(2) in Case 3 (bimodal). The AR coefficients of (a)–(f) are $(-0.8, -0.4)$, $(0.8, -0.4)$, $(0.2, -0.1)$, $(-0.2, 0.1)$, $(0.1, -0.05)$, and $(-0.1, 0.05)$ respectively [Color figure can be viewed at wileyonlinelibrary.com]

  J. Time Ser. Anal. **39**: 690–708 (2018)
DOI: 10.1111/jtsa.12293

Table V. Coverage frequencies of rolling forecasts by 95% oracle and normal prediction intervals for the difference series of logged oil price

k-Step ahead	Number of forecasts	Oracle	Normal
$k = 2$	50	0.940	0.900
	70	0.957	0.929
	100	0.930	0.910
	131	0.947	0.931
$k = 3$	50	0.920	0.900
	70	0.943	0.929
	100	0.910	0.900
	131	0.931	0.924

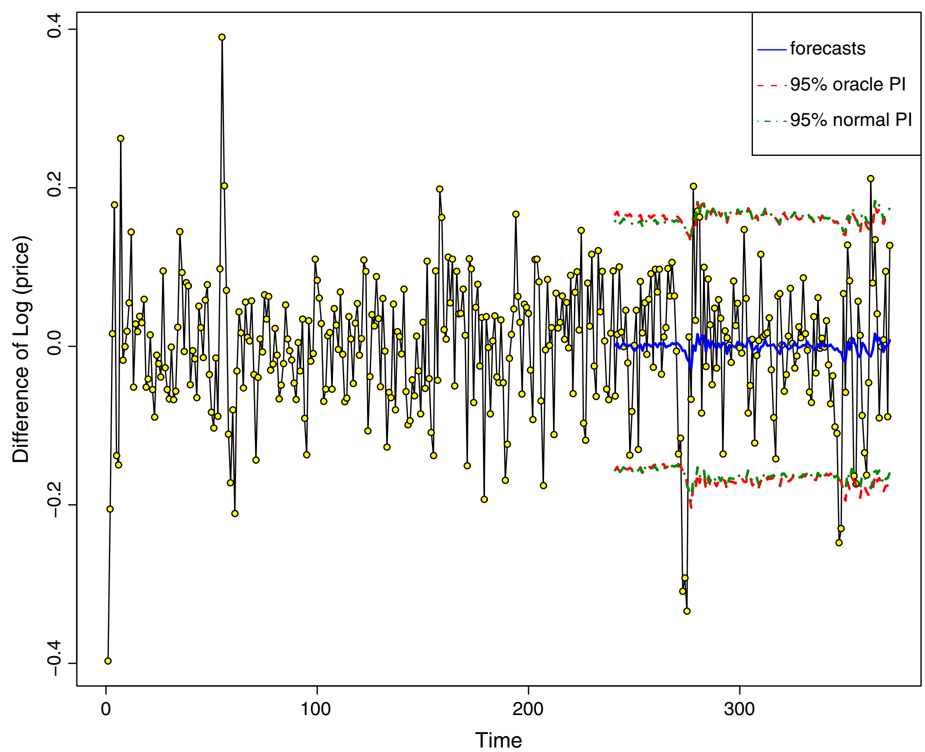

Figure 4. AR(1) model filling and two-step ahead rolling forecasts of 131 points for the difference series of logged oil price [Color figure can be viewed at wileyonlinelibrary.com]

studies. Although as good as the oracle PI in simulation, the empirical PI does not have any theoretical justification, while the bootstrap PI of Thombs and Schucany (1990) takes much longer computing time. Considering all of the above findings, we recommend with confidence the oracle PI as a robust multi-step-ahead prediction tool for the AR model.

Although in this article our focus was on the AR(p) model, the basic idea is applicable to other time series models such as ARMA, SARIMA, and VAR. In all cases, the KDE of multi-step-ahead innovation distribution based on residuals can be shown oracally efficient, and thus the corresponding quantiles are efficient estimators. Hence, much future work in this direction is expected, namely prediction intervals of various types developed for other time series models such as ARMA, PARMA, and VAR, by establishing uniform oracle efficiency of KDE based on residuals, as in Wang *et al.* (2014).

ACKNOWLEDGEMENTS

This work has been supported in part by the National Natural Science Foundation of China (award 11371272, 11771240, 11701403), the Research Fund for the Doctoral Program of Higher Education of China (award 20133201110002), the Jiangsu Key-Discipline Program (Statistics) (ZY107992), and the 2017 Jiangsu Overseas Visiting Scholar Program for University Prominent Young and Middle-aged Teachers and Presidents. Helpful comments from the co-editor and two reviewers are greatly appreciated.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

REFERENCES

Bai J. 1994. Weak convergence of the sequential empirical processes of residuals in ARMA models. *Annals of Statistics* **22**: 2051–2061.

Bosq D. 1998. *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, 2nd ed. New York: Springer-Verlag.

Brockwell PJ, Davis RA. 1991. *Time Series: Theory and Methods*, 2nd ed. New York: Springer-Verlag.

Cryer JD, Chan KS. 2008. *Time Series Analysis with Applications in R*, 2nd ed. New York: Springer-Verlag.

Dehling H, Mikosch T, Sørensen M. 2002. *Empirical Process Techniques for Dependent Data*. New York: Springer-Verlag.

Doukhan P. 1994. *Mixing: Properties and Examples*. New York: Springer-Verlag.

Liu R, Yang L. 2008. Kernel estimation of multivariate cumulative distribution function. *Journal of Nonparametric Statistics* **20**: 661–677.

Marron JS, Wand MP. 1992. Exact mean integrated squared error. *Annals of Statistics* **20**: 712–736.

Reiss RD. 1981. Nonparametric estimation of smooth distribution functions. *Scandinavian Journal of Statistics* **8**: 116–119.

Thombs L, Schucany W. 1990. Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association* **85**: 486–492.

Wang J, Cheng F, Yang L. 2013. Smooth simultaneous confidence bands for cumulative distribution functions. *Journal of Nonparametric Statistics* **25**: 395–407.

Wang J, Liu R, Cheng F, Yang L. 2014. Oracally efficient estimation of autoregressive error distribution with simultaneous confidence band. *Annals of Statistics* **42**: 654–668.

APPENDIX

A.1 PRELIMINARIES

In what follows, we take $\|\cdot\|_2$ and $\|\cdot\|_\infty$ as L^2-norm and supremum norm respectively. Denote by $\mathcal{O}_p$ (or o_p) sequences of random variables $\mathcal{O}$ (or o) of certain order in probability, by $\mathcal{O}_{a.s.}$ (or $o_{a.s.}$) almost surely $\mathcal{O}$ (or o), and by U (or u) uniformly $\mathcal{O}$ (or o), etc. For any real number a, one denotes by $[a]$ its integer part.

Lemma A.1 (Dehling *et al.* (2002), Theorem 4.3)**.** Let $\left\{\xi_j\right\}_{j=-\infty}^{\infty}$ be a strictly stationary sequence of d-dimensional random vectors defined on probability space $(\Omega, \mathcal{F}, P)$, with strong mixing coefficient $\alpha(n) \ll n^{-4-2d}$ and marginal distribution function F. For $s \in \mathbb{R}^d$, write $g_n(s) = I_{\{\xi_n \le s\}} - F(s)$. Then the series $\Gamma(s, s') = \mathbb{E}g_1(s)\,g_1(s') + \sum_{n\ge 2}\left\{\mathbb{E}g_1(s)\,g_n(s') + \mathbb{E}g_n(s)\,g_1(s')\right\}$ converges absolutely for $s, s' \in \mathbb{R}^d$ and defines a covariance function. There exists a sequence $\left\{\zeta_n(s)\right\}_{n=1}^{\infty}$ of i.i.d. Gaussian processes defined on $(\Omega, \mathcal{F}, P)$, indexed by $s \in \mathbb{R}^d$, with $\mathbb{E}\zeta_1(s) = 0, \mathbb{E}\zeta_1(s)\,\zeta_1(s') = \Gamma(s, s'), s, s' \in \mathbb{R}^d$, and $\lambda > 0$ that depends only on d such that with probability 1

$$\sup_{s\in\mathbb{R}^d}\left| n^{-1}\sum\nolimits_{j=1}^{n} I_{\{\xi_j \le s\}} - F(s) - n^{-1}\sum\nolimits_{j=1}^{n}\zeta_j(s)\right| \ll n^{-1/2}(\log n)^{-\lambda}.$$

DOI: 10.1111/jtsa.12293

Corollary A.1. Under Assumption (A3), there is a sequence $\left\{\zeta_n^{[k]}(z), z \in \mathbb{R}\right\}_{n=1}^{\infty}$ of Gaussian processes such that for $\lambda > 0$ in Lemma A.1 with $d = 1$, as $n \to \infty$

$$\sup_{z\in\mathbb{R}} \left| F_n^{[k]}(z) - F^{[k]}(z) - n^{-1/2}\zeta_n^{[k]}(z) \right| = \mathcal{O}_{a.s}\left(n^{-1/2}(\log n)^{-\lambda}\right),$$
$$\mathbb{E}\zeta_n^{[k]}(z) \equiv 0, \mathbb{E}\zeta_n^{[k]}(z)\zeta_n^{[k]}(z') \equiv \Gamma^{[k]}(z, z'),$$
$$\Gamma^{[k]}(z, z') \equiv \sum_{|j|\le k} \left\{ \mathbb{E}I\left(Z_t^{[k]} \le z\right) I\left(Z_{t+j}^{[k]} \le z'\right) - F^{[k]}(z) F^{[k]}(z') \right\}, z, z' \in \mathbb{R}.$$

Proof Assumption (A3) and the definition of $Z_j^{[k]}$ in (2.4) ensure that the $Z_j^{[k]}$'s are strictly stationary and geometrically strong mixing, hence the conclusion of Lemma A.1 applies to $\xi_j = Z_j^{[k]}, \zeta_n^{[k]}(z) = n^{-1/2}\sum_{j=1}^n \zeta_j(z)$. Furthermore, since the sequence $\left\{Z_j^{[k]}\right\}_{j=-\infty}^{\infty}$ is k-dependent, the sum in $\mathbb{E}\zeta_n^{[k]}(z)\zeta_n^{[k]}(z')$ is finite. The corollary is proved. □

Lemma A.2 (Bosq (1998), Theorem 1.4)**.** Let $\{\xi_i\}_{i=-\infty}^{\infty}$ be a mean-zero real-valued process with strong mixing coefficients $\alpha(n)$. Suppose that there exists a $c > 0$ such that for $i = 1, \ldots, n, \mu \ge 3$, $\mathbb{E}\left|\xi_i\right|^{\mu} \le c^{\mu-2}\mu!\mathbb{E}\xi_i^2 < +\infty$, and denote $m_r = \max_{1\le i\le n}\left\|\xi_i\right\|_r, r \ge 2$. Then for each $n > 1$, integer $q \in [1, n/2]$, each $\varepsilon_n > 0$ and $\mu \ge 3$

$$\mathrm{P}\left\{\left|\sum_{i=1}^n \xi_i\right| > n\varepsilon_n\right\} \le a_1 \exp\left(-\frac{q\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n}\right) + a_2(\mu)\,\alpha\left(\left[\frac{n}{q+1}\right]\right)^{2\mu/(2\mu+1)},$$

where

$$a_1 = 2\frac{n}{q} + 2\left(1 + \frac{\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n}\right), a_2(\mu) = 11n\left(1 + \frac{5m_\mu^{2\mu/(2\mu+1)}}{\varepsilon_n}\right).$$

Lemma A.3. Under Assumptions (A2) and (A3), there exists a $\gamma > 0$ such that as $n \to \infty$, $\max\left(\left|X_1\right|, \left|X_2\right|, \cdots, \left|X_n\right|\right) = \mathcal{O}_{a.s.}(n^{\gamma})$.

Proof By (2.1) and (2.2), $\left|X_t\right| \le C_\psi \sum_{j=0}^{\infty}\rho_\psi^j\left|Z_{t-j}\right|, t \in \mathbb{Z}$. Define a nonnegative random variable $W = C_\psi\sum_{j=0}^{\infty}\rho_\psi^j\left|Z_{1-j}\right|$, and Assumption (A2) entails that

$$\mathbb{E}W = C_\psi\sum_{j=0}^{\infty}\rho_\psi^j\mathbb{E}\left|Z_{1-j}\right| \le C_\psi\left(1-\rho_\psi\right)^{-1}M_\eta^{1/(6+3\eta)} < +\infty,$$

thus $W < +\infty$ almost surely. Note that

$$\max\left\{\left|X_1\right|, \ldots, \left|X_n\right|\right\} \le \max\left\{\left|Z_2\right|, \ldots, \left|Z_n\right|\right\} C_\psi\left(1-\rho_\psi\right)^{-1} + W \tag{A.1}$$

because $\left|X_1\right| \le W$ and

$$\begin{aligned}
\left|X_2\right| &\le C_\psi\left|Z_2\right| + \rho_\psi W,\\
&\cdots\\
\left|X_n\right| &\le C_\psi\left(\left|Z_n\right| + \rho_\psi\left|Z_{n-1}\right| + \cdots + \rho_\psi^{n-2}\left|Z_2\right|\right) + \rho_\psi^{n-1}W.
\end{aligned}$$

DOI: 10.1111/jtsa.12293

Under Assumption (A2),

$$\begin{aligned}
\mathrm{P}\left\{\max_{2\le t\le n}\left|Z_t\right| > n^{\gamma}\right\} &\le \sum_{t=2}^{n}\mathrm{P}\left\{\left|Z_t\right| > n^{\gamma}\right\} \\
&\le \sum_{t=2}^{n} n^{-\gamma(6+3\eta)}\mathbb{E}\left|Z_t\right|^{6+3\eta} \le n^{-\gamma(6+3\eta)+1}M_\eta,
\end{aligned}$$

hence for $\gamma > 2\,(6+3\eta)^{-1}$

$$\sum_{n=1}^{\infty}\mathrm{P}\left\{\max_{2\le t\le n}\left|Z_t\right| > n^{\gamma}\right\} \le \sum_{n=1}^{\infty} n^{-\gamma(6+3\eta)+1}M_\eta < +\infty,$$

so $\max_{2\le t\le n}\left|Z_t\right| = \mathcal{O}_{a.s.}\,(n^{\gamma})$ by the Borel–Cantelli lemma. Lemma A.3 is proved by noting (A.1). □

The following Lemma A.4 follows by elementary algebra.

Lemma A.4. Under Assumptions (A2) and (A4), there exists an $a > 0$, such that the following are fulfilled for the sequence $\left\{D_n\right\} = \{n^a\}$, $\sum_{n=1}^{\infty} D_n^{-(2+\eta)} < \infty$, $D_n^{-(1+\eta)}n^{1/2}h^{1/2} \to 0$, $D_n n^{-1/2}h^{-1/2}\,(\log n) \to 0$.

The next Lemma A.5 provides all the building blocks for proving Theorem 2.2. Its proof is given in Supporting Information.

Lemma A.5. Under Assumptions (A1)–(A5), for any $k \le r, s, v, w \le k+p-1$, as $n \to \infty$

$$\sup_{z\in\mathbb{R}}\left|n^{-1}\sum_{t=k}^{n} K_h\left(z - Z_t^{[k]}\right)X_{t-r}\right| = \mathcal{O}_p\left(n^{-1/2}h^{-1/2}\log n\right), \tag{A.2}$$

$$\sup_{z\in\mathbb{R}}\left|n^{-1}\sum_{t=k}^{n} K_h'\left(z - Z_t^{[k]}\right)X_{t-r}X_{t-s}\right| = \mathcal{O}_p\left(1\right), \tag{A.3}$$

$$\sup_{z\in\mathbb{R}}\left|n^{-1}\sum_{t=k}^{n} K_h''\left(z - Z_t^{[k]}\right)X_{t-r}X_{t-s}X_{t-v}\right| = \mathcal{O}_p\left(1\right), \tag{A.4}$$

$$n^{-1}\sum_{t=k}^{n}\left|X_{t-r}X_{t-s}X_{t-v}X_{t-w}\right| = \mathcal{O}_p\left(1\right). \tag{A.5}$$

In what follows, we take $\alpha \in (0,1)$. Recall the oracle quantile estimator $\hat{q}_{n,\alpha}^{[k]} = \left(\hat{F}^{[k]}\right)^{-1}(\alpha) = \inf\left\{x : \hat{F}^{[k]}(x) \ge \alpha\right\}$, and its basic property:

Lemma A.6. Under Assumptions (A1)–(A5), for any $\alpha \in (0,1)$, as $n \to \infty$, $\left|\hat{q}_{n,\alpha}^{[k]} - q_\alpha^{[k]}\right| = \mathcal{O}_p\left(n^{-1/2}\right)$.

Proof It follows from Theorems 2.1 and 2.2 and Corollary A.1 that

$$A_n = \sup_{z\in\mathbb{R}}\left|\hat{F}^{[k]}\,(z) - F^{[k]}\,(z)\right| = \mathcal{O}_p\left(n^{-1/2}\right). \tag{A.6}$$

As noted in Remark 2.1, $F^{[k]}$ has everywhere positive derivatives and there exists a unique $q_\alpha^{[k]}$ with $F^{[k]}\left(q_\alpha^{[k]}\right) = \alpha$. It is elementary to verify that $\hat{F}^{[k]}(z)$ is strictly increasing for $\hat{F}^{[k]}(z) \in (0,1)$. Therefore, $\hat{F}^{[k]}\left(\hat{q}_{n,\alpha}^{[k]}\right) = \alpha$.

DOI: 10.1111/jtsa.12293

According to the definition of A_n in (A.6), one has $\left|\hat{F}^{[k]}\left(\hat{q}^{[k]}_{n,\alpha}\right) - F^{[k]}\left(\hat{q}^{[k]}_{n,\alpha}\right)\right| \le A_n$, thus $\left|\alpha - F^{[k]}\left(\hat{q}^{[k]}_{n,\alpha}\right)\right| \le A_n$, so

$$\left|F^{[k]}\left(\hat{q}^{[k]}_{n,\alpha}\right) - F^{[k]}\left(q^{[k]}_{\alpha}\right)\right| \le A_n. \tag{A.7}$$

Since $A_n = \mathcal{O}_p\left(n^{-1/2}\right)$, $\forall \varepsilon > 0$, there $\exists M > 0$, s.t. $\mathrm{P}\left(\omega\,\middle|\,|A_n(\omega)| > n^{-1/2}M\right) < \varepsilon/2$, $n = 1, 2, \ldots$, letting $\Omega_n = \left\{\omega\,\middle|\,\left|F^{[k]}\left(\hat{q}^{[k]}_{n,\alpha}(\omega)\right) - F^{[k]}\left(q^{[k]}_{\alpha}\right)\right| > n^{-1/2}M\right\}$. Then (A.7) implies that $\mathrm{P}(\Omega_n) < \varepsilon/2, \forall n = 1, 2, \ldots$. Monotonicity of $F^{[k]}(z)$ (see Remark 2.1) entails that there $\exists \delta > 0$ such that

$$\left|F^{[k]}(z) - F^{[k]}\left(q^{[k]}_{\alpha}\right)\right| \ge \begin{cases} 1/2 f^{[k]}\left(q^{[k]}_{\alpha}\right)\left|z - q^{[k]}_{\alpha}\right|, \text{ if } \left|z - q^{[k]}_{\alpha}\right| \le \delta, \\ C_\delta, \text{ for some constant } C_\delta > 0, \text{ if } \left|z - q^{[k]}_{\alpha}\right| > \delta. \end{cases}$$

Let $\Delta_n = \left\{\omega\,\middle|\,\left|\hat{q}^{[k]}_{n,\alpha}(\omega) - q^{[k]}_{\alpha}\right| \le \delta\right\}$, then $\overline{\Delta}_n = \left\{\omega\,\middle|\,\left|\hat{q}^{[k]}_{n,\alpha}(\omega) - q^{[k]}_{\alpha}\right| > \delta\right\}$. For $\omega \in \Delta_n$, $\left|F^{[k]}\left(\hat{q}^{[k]}_{n,\alpha}(\omega)\right) - F^{[k]}\left(q^{[k]}_{\alpha}\right)\right| \ge 1/2 f^{[k]}\left(q^{[k]}_{\alpha}\right)\left|\hat{q}^{[k]}_{n,\alpha}(\omega) - q^{[k]}_{\alpha}\right|$, i.e.,

$$\left|\hat{q}^{[k]}_{n,\alpha} - q^{[k]}_{\alpha}\right| \le 2\left\{f^{[k]}\left(q^{[k]}_{\alpha}\right)\right\}^{-1}\left|F^{[k]}\left(\hat{q}^{[k]}_{n,\alpha}(\omega)\right) - F^{[k]}\left(q^{[k]}_{\alpha}\right)\right|.$$

For $\omega \in \overline{\Delta}_n$, $\left|F^{[k]}\left(\hat{q}^{[k]}_{n,\alpha}(\omega)\right) - F^{[k]}\left(q^{[k]}_{\alpha}\right)\right| \ge C_\delta > 0$. If $n > M/C_\delta^2$, then $n^{-1/2}M < C_\delta$, so $\omega \in \Omega_n$, and hence $\mathrm{P}\left(\overline{\Delta}_n\right) \le \mathrm{P}\left(\Omega_n\right) < \varepsilon/2$, for $n > n_0 = \left[\left(M/C_\delta\right)^2\right]$. Therefore

$$\mathrm{P}\left(\Delta_n \cap \overline{\Omega}_n\right) = 1 - \mathrm{P}\left(\overline{\Delta}_n \cup \Omega_n\right) \ge 1 - \varepsilon,$$

while $\omega \in \Delta_n \cap \overline{\Omega}_n$ implies that

$$\left|\hat{q}^{[k]}_{n,\alpha}(\omega) - q^{[k]}_{\alpha}\right| \le 2\left\{f^{[k]}\left(q^{[k]}_{\alpha}\right)\right\}^{-1}\left|F^{[k]}\left(\hat{q}^{[k]}_{n,\alpha}(\omega)\right) - F^{[k]}\left(q^{[k]}_{\alpha}\right)\right|$$

and $\left|F^{[k]}\left(\hat{q}^{[k]}_{n,\alpha}(\omega)\right) - F^{[k]}\left(q^{[k]}_{\alpha}\right)\right| \le n^{-1/2}M$, and thus

$$\left|\hat{q}^{[k]}_{n,\alpha}(\omega) - q^{[k]}_{\alpha}\right| \le 2\left\{f^{[k]}\left(q^{[k]}_{\alpha}\right)\right\}^{-1} n^{-1/2}M.$$

Consequently

$$\mathrm{P}\left(\sqrt{n}\left|\hat{q}^{[k]}_{n,\alpha} - q^{[k]}_{\alpha}\right| \le 2M\left\{f^{[k]}\left(q^{[k]}_{\alpha}\right)\right\}^{-1}\right) \ge \mathrm{P}(\Delta_n \cap \overline{\Omega}_n) \ge 1 - \varepsilon,$$

i.e.

$$\mathrm{P}\left(\sqrt{n}\left|\hat{q}^{[k]}_{n,\alpha} - q^{[k]}_{\alpha}\right| > 2M\left\{f^{[k]}\left(q^{[k]}_{\alpha}\right)\right\}^{-1}\right) \le \varepsilon,\ n \ge n_0 + 1,$$

which leads to the conclusion that $\left|\hat{q}^{[k]}_{n,\alpha} - q^{[k]}_{\alpha}\right| = \mathcal{O}_p\left(n^{-1/2}\right)$. □

DOI: 10.1111/jtsa.12293

A.2 PROOFS OF THEOREMS 2.1–2.3

Proof of Theorem 2.1 One notes that

$$\tilde{F}^{[k]}(z) = n^{-1}\sum_{t=k}^{n}\int_{-\infty}^{z} K_h\left(u - Z_t^{[k]}\right)\mathrm{d}u = \int_{-1}^{1} F_n^{[k]}(z - hv)\, K(v)\,\mathrm{d}v,$$

hence $\tilde{F}^{[k]}(z) - F_n^{[k]}(z) = \int_{-1}^{1}\left\{F_n^{[k]}(z-hv) - F_n^{[k]}(z)\right\}K(v)\,\mathrm{d}v$. Corollary A.1 implies that, as $n \to \infty$

$$\sup_{z\in\mathbb{R}}\left|F_n^{[k]}(z) - F^{[k]}(z) - n^{-1/2}\zeta_n^{[k]}(z)\right| = \mathcal{O}_{a.s.}\left(n^{-1/2}(\log n)^{-\lambda}\right).$$

The equicontinuity of $\zeta_n^{[k]}(z), z \in \mathbb{R}$ implies the following:

$$\sup_{v\in[-1,1]}\sup_{z\in\mathbb{R}}\left|\sqrt{n}\left\{F_n^{[k]}(z-hv) - F_n^{[k]}(z)\right\} - \sqrt{n}\left\{F^{[k]}(z-hv) - F^{[k]}(z)\right\}\right| = o_p(1).$$

Thus, one obtains

$$\int_{-1}^{1}\left\{F_n^{[k]}(z-hv) - F_n^{[k]}(z)\right\}K(v)\,\mathrm{d}v - \int_{-1}^{1}\left\{F^{[k]}(z-hv) - F^{[k]}(z)\right\}K(v)\,\mathrm{d}v$$
$$= u_p\left(n^{-1/2}\right).$$

Under Assumption (A1), applying Taylor expansion to $F^{[k]} \in C^{(1,\beta)}(\mathbb{R})$, there exists $c > 0$ such that for $z \in \mathbb{R}, v \in [-1,1]$,

$$\left|F^{[k]}(z-hv) - F^{[k]}(z) - f^{[k]}(z)(-hv)\right| \le ch^{1+\beta}|v|^{1+\beta} \le ch^{1+\beta},$$

which entails that

$$\left|\int_{-1}^{1}\left\{F^{[k]}(z-hv) - F^{[k]}(z) - f^{[k]}(z)(-hv)\right\}K(v)\,\mathrm{d}v\right|$$
$$\le ch^{1+\beta}\int_{-1}^{1}K(v)\,\mathrm{d}v \le ch^{1+\beta} = o\left(n^{-1/2}\right)$$

by the fact that $h^{1+\beta} = o\left(n^{-1/2}\right)$ according to Assumption (A4). Since $\int_{-1}^{1} vK(v)\,\mathrm{d}v = 0$, one obtains that

$$\left|\int_{-1}^{1}\left\{F^{[k]}(z-hv) - F^{[k]}(z)\right\}K(v)\,\mathrm{d}v\right| = o\left(n^{-1/2}\right),$$

and hence

$$\left|\int_{-1}^{1}\left\{F_n^{[k]}(z-hv) - F_n^{[k]}(z)\right\}K(v)\,\mathrm{d}v\right|$$
$$\le \left|\int_{-1}^{1}\left\{F^{[k]}(z-hv) - F^{[k]}(z)\right\}K(v)\right|\mathrm{d}v + u_p\left(n^{-1/2}\right) = u_p\left(n^{-1/2}\right).$$

Thus, it follows that

$$\sup_{z\in\mathbb{R}}\left|\tilde{F}^{[k]}(z) - F_n^{[k]}(z)\right|$$
$$= \sup_{z\in\mathbb{R}}\left|\int_{-1}^{1}\left\{F_n^{[k]}(z-hv) - F_n^{[k]}(z)\right\}K(v)\,\mathrm{d}v\right| = o_p\left(n^{-1/2}\right). \quad \square$$

DOI: 10.1111/jtsa.12293

Proof of Theorem 2.2 According to the definitions of $\hat{F}^{[k]}(z)$ and $\tilde{F}^{[k]}(z)$ given in (1.5) and (2.6), one has

$$\hat{F}^{[k]}(z) - \tilde{F}^{[k]}(z) = n^{-1} \sum_{t=k}^{n} \left\{ G\left(\frac{z - \hat{Z}_t^{[k]}}{h} \right) - G\left(\frac{z - Z_t^{[k]}}{h} \right) \right\}, \tag{A.8}$$

where $G(z) = \int_{-\infty}^{z} K(u)\,\mathrm{d}u$. The right-hand side of equation (A.8) is by third-order Taylor expansion $I_1+I_2+I_3+I_4$, where

$$\begin{aligned}
I_1 &= I_1(z) = n^{-1} \sum_{t=k}^{n} K\left(\frac{z - Z_t^{[k]}}{h} \right) \frac{\hat{Z}_t^{[k]} - Z_t^{[k]}}{h}, \\
I_2 &= I_2(z) = n^{-1} \sum_{t=k}^{n} K'\left(\frac{z - Z_t^{[k]}}{h} \right) \left(\frac{\hat{Z}_t^{[k]} - Z_t^{[k]}}{h} \right)^2, \\
I_3 &= I_3(z) = n^{-1} \sum_{t=k}^{n} K''\left(\frac{z - Z_t^{[k]}}{h} \right) \left(\frac{\hat{Z}_t^{[k]} - Z_t^{[k]}}{h} \right)^3, \\
I_4 &= I_4(z) = n^{-1} \sum_{t=k}^{n} R_t = n^{-1} \sum_{t=k}^{n} R_t(z),
\end{aligned}$$

in which $\left|R_t\right| \le \frac{\|K\|_{2,1}}{6} \left| \frac{\hat{Z}_t^{[k]} - Z_t^{[k]}}{h} \right|^4$. Straight algebra and (1.4) and (1.2) provide

$$\hat{Z}_t^{[k]} - Z_t^{[k]} = \tilde{X}_t^{[k]} - \hat{X}_t^{[k]} = \left(\phi_1^{[k]} - \hat{\phi}_1^{[k]} \right) X_{t-k} + \cdots + \left(\phi_p^{[k]} - \hat{\phi}_p^{[k]} \right) X_{t-k-p+1}.$$

According to Lemma A.5, one obtains next that

$$\begin{aligned}
\sup_{z\in\mathbb{R}} \left|I_1\right| &= \sup_{z\in\mathbb{R}} \left| n^{-1} \sum_{t=k}^{n} K\left(\frac{z - \hat{Z}_t^{[k]}}{h} \right) \frac{\hat{Z}_t^{[k]} - Z_t^{[k]}}{h} \right| \\
&= \sup_{z\in\mathbb{R}} \left| n^{-1} \sum_{t=k}^{n} K_h\left(z - \hat{Z}_t^{[k]} \right) \sum_{j=1}^{p} \left(\phi_j^{[k]} - \hat{\phi}_j^{[k]} \right) X_{t-k-j+1} \right| \\
&\le p \max_{1\le j\le p} \left| \phi_j^{[k]} - \hat{\phi}_j^{[k]} \right| \sup_{k\le r\le k+p-1} \sup_{z\in\mathbb{R}} \left| n^{-1} \sum_{t=k}^{n} K_h\left(z - \hat{Z}_t^{[k]} \right) X_{t-r} \right| \\
&= \mathcal{O}_p\left(n^{-1/2} \right) \mathcal{O}_{a.s.}\left(n^{-1/2} h^{-1/2} \log n \right) = o_p\left(n^{-1/2} \right).
\end{aligned} \tag{A.9}$$

similarly,

$$\sup_{z\in\mathbb{R}} \left|I_2\right| = \sup_{z\in\mathbb{R}} n^{-1} \left| \sum_{t=k}^{n} K'\left(\frac{z - Z_t^{[k]}}{h} \right) \left(\frac{\hat{Z}_t^{[k]} - Z_t^{[k]}}{h} \right)^2 \right| = o_p\left(n^{-1/2} \right) \tag{A.10}$$

$$\sup_{z\in\mathbb{R}} \left|I_3\right| = \sup_{z\in\mathbb{R}} n^{-1} \left| \sum_{t=k}^{n} K''\left(\frac{z - Z_t^{[k]}}{h} \right) \left(\frac{\hat{Z}_t^{[k]} - Z_t^{[k]}}{h} \right)^3 \right| = o_p\left(n^{-1/2} \right) \tag{A.11}$$

DOI: 10.1111/jtsa.12293

and together with Assumptions (A4) and (A5),

$$\sup_{z\in\mathbb{R}}\left|I_4\right| \le C \sup_{k\le t\le n}\left|\frac{\hat{Z}_t^{[k]} - Z_t^{[k]}}{h}\right|^4 \le h^{-4}p^4\left(\max_{1\le j\le p}\left|\phi_j^{[k]} - \hat{\phi}_j^{[k]}\right|\right)^4 \tag{A.12}$$

$$\times \sup_{k\le r,s,v,w\le k+p-1} n^{-1}\sum_{t=k}^{n}\left|X_{t-r}X_{t-s}X_{t-v}X_{t-w}\right|$$

$$= \mathcal{O}_p\left(n^{-2}h^{-4}\right)\mathcal{O}_p(1) = o_p\left(n^{-1/2}\right).$$

Since $\sup_{z\in\mathbb{R}}\left|\hat{F}^{[k]}(z) - \tilde{F}^{[k]}(z)\right| \le \sup_{z\in\mathbb{R}}\left(\left|I_1\right| + \left|I_2\right| + \left|I_3\right| + \left|I_4\right|\right)$, Theorem 2.2 can be easily proved by (A.9), (A.10), (A.11), and (A.12). □

Proof of Theorem 2.3 Let $Z^{[k]}$ be a random variable with distribution function $F^{[k]}(z)$. Recall (1.2) and that $X_{n+k} = \phi_1^{[k]}X_n + \cdots + \phi_p^{[k]}X_{n-p+1} + Z_{n+k}^{[k]}$. Thus as $n\to\infty$, $X_{n+k} - \tilde{X}_{n+k}^{[k]} = Z_{n+k}^{[k]} \xrightarrow{\mathcal{L}} Z^{[k]}$. Delta method and (1.2) and (1.4) imply that $\tilde{X}_{n+k}^{[k]} - \hat{X}_{n+k}^{[k]} \xrightarrow{p} 0$. Therefore $X_{n+k} - \tilde{X}_{n+k}^{[k]} + \tilde{X}_{n+k}^{[k]} - \hat{X}_{n+k}^{[k]} = X_{n+k} - \hat{X}_{n+k}^{[k]} \xrightarrow{\mathcal{L}} Z^{[k]}$ by applying Slutsky's Theorem. Next, Lemma A.6 provides that $\left|\hat{q}_{n,\alpha/2}^{[k]} - q_{\alpha/2}^{[k]}\right| \xrightarrow{p} 0$, so applying Slutsky's theorem again, one obtains that

$$X_{n+k} - \hat{X}_{n+k}^{[k]} - \left(\hat{q}_{n,\alpha/2}^{[k]} - q_{\alpha/2}^{[k]}\right) \xrightarrow{\mathcal{L}} Z^{[k]}.$$

Since the distribution function $F^{[k]}$ is continuous by Remark 2.1, one has

$$\lim_{n\to\infty} \mathrm{P}\left[X_{n+k} - \hat{X}_{n+k}^{[k]} - \left(\hat{q}_{n,\alpha/2}^{[k]} - q_{\alpha/2}^{[k]}\right) \le q_{\alpha/2}^{[k]}\right] = \mathrm{P}\left[Z^{[k]} \le q_{\alpha/2}^{[k]}\right] = \alpha/2,$$

i.e.

$$\mathrm{P}\left[X_{n+k} - \hat{X}_{n+k}^{[k]} \le \hat{q}_{n,\alpha/2}^{[k]}\right] \to \alpha/2.$$

Likewise,

$$\mathrm{P}\left[X_{n+k} - \hat{X}_{n+k}^{[k]} \le \hat{q}_{n,(1-\alpha/2)}^{[k]}\right] \to 1-\alpha/2,$$

and consequently

$$\mathrm{P}\left[X_{n+k} - \hat{X}_{n+k}^{[k]} \le \hat{q}_{n,(1-\alpha/2)}^{[k]}, X_{n+k} - \hat{X}_{n+k}^{[k]} > \hat{q}_{n,\alpha/2}^{[k]}\right] \to 1-\alpha/2-\alpha/2 = 1-\alpha,$$

or

$$\mathrm{P}\left(X_{n+k} \in \left[\hat{X}_{n+k}^{[k]} + \hat{q}_{n,\alpha/2}^{[k]}, \hat{X}_{n+k}^{[k]} + \hat{q}_{n,(1-\alpha/2)}^{[k]}\right]\right) \to 1-\alpha.$$ □

  J. Time Ser. Anal. **39**: 690–708 (2018)
DOI: 10.1111/jtsa.12293