Computational Statistics (2022) 37:1015–1039
https://doi.org/10.1007/s00180-021-01149-5

ORIGINAL PAPER

Kolmogorov–Smirnov simultaneous confidence bands for time series distribution function

Jie Li[1] · Jiangyan Wang[2] · Lijian Yang[1]

Received: 10 March 2021 / Accepted: 31 August 2021 / Published online: 18 September 2021

Abstract
Claims about distributions of time series are often unproven assertions instead of substantiated conclusions for lack of hypotheses testing tools. In this work, Kolmogorov–Smirnov type simultaneous confidence bands (SCBs) are constructed based on simple random samples (SRSs) drawn from realizations of time series, together with smooth SCBs using kernel distribution estimator (KDE) instead of empirical cumulative distribution function of the SRS. All SCBs are shown to enjoy the same limiting distribution as the standard Kolmogorov–Smirnov for i.i.d. sample, which is validated in simulation experiments on various time series. Computing these SCBs for the standardized S&P 500 daily returns data leads to some rather unexpected findings, i.e., student's t-distributions with degrees of freedom no less than 3 and the normal distribution are all acceptable versions of the standardized daily returns series' distribution, with proper rescaling. These findings present challenges to the long held belief that daily financial returns distribution is fat-tailed and leptokurtic.

Keywords Bandwidth · Brownian bridge · Kernel · Kolmogorov distribution · Simple random sample · Stationarity

---

Jie Li and Jiangyan Wang: Co-first author.

Supported by National Natural Science Foundation of China awards 11771240, 11801272, 12026242, Natural Science Foundation of Jiangsu Province of China BK20180820, Qinglan Project of Jiangsu Province of China, Philosophy and Social Science Research in Colleges and Universities of Jiangsu Province 2019SJA0353.

✉ Lijian Yang
yanglijian@tsinghua.edu.cn

[1] Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing 100084, China

[2] School of Statistics and Data Science, Nanjing Audit University, Nanjing 211815, China

1 Introduction

Probability distribution function contains complete information about a random quantity, such as financial returns. Claims about probability distribution are routinely made such as skewness vs. symmetry, normality, fat-tailedness, etc., often as baseless assertions. One such unsubstantiated claim is that distribution function of daily financial returns is fat-tailed and leptokurtic (Francq and Zakoian 2010, p 9, (v)), because its empirical distribution does not “look” normal, its sample kurtosis is much greater than 3, and classic normality tests lead to rejection.

Take, for example, the S&P 500 daily returns from January 3, 1950 to August 28, 2018, shown in Fig. 1(a). This realization of length 17,276 is visibly not stationary, with much wider range of variation towards the last 10 years, hence a standardized (and therefore stationary) returns series by adjusting for trend in variance is used for further analysis, shown in Fig. 1(b). Figure 2 shows the empirical distribution function together with 99% Kolmogorov–Smirnov simultaneous confidence band (SCB) based on the standardized returns and a normal distribution function with the mean and variance equal to the sample mean and sample variance. This figure is a nice illustration of “not looking normal ” as the normal curve falls outside of the SCB.

Classic normality tests such as the previous one based on Kolmogorov–Smirnov SCB, however, are suitable only for i.i.d. observations, hence it seems that a direct and reliable testing procedure on the shape of time series distribution is called for to decide on the validity of various presumptions. For the aforementioned S&P 500 daily returns data, 95% Kolmogorov–Smirnov type SCBs based on a simple random sample

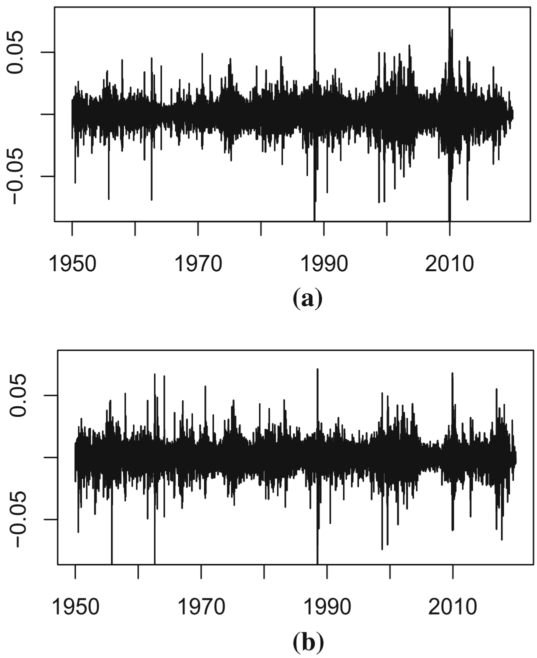

Fig. 1 (**a**) Scatter plot of $\{y_t\}_{t=1}^{17276}$, the S&P 500 daily returns, January 3, 1950–August 28, 2018; (**b**) scatter plot of the standardized returns $\{x_t\}_{t=1}^{17276}$

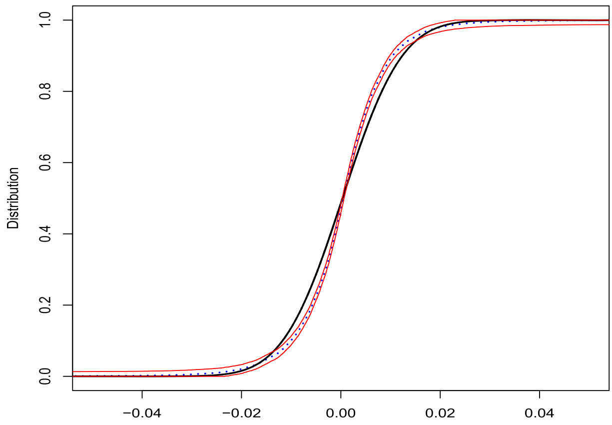

Fig. 2 The standardized S&P 500 daily returns $\{x_t\}_{t=1}^{17276}$. Dotted line—F_N, solid lines—99% Kolmogorov–Smirnov SCBs, thick line—normal cdf with mean and variance equal to the sample mean and sample variance of $\{x_t\}_{t=1}^{17276}$

(SRS) of much smaller size 200, defined in (7) and (9), contain rescaled normal and student's t-distributions of various degrees of freedom entirely, see Fig. 3. These SCBs are asymptotically correct according to Corollary 1 in the next section, thus any null hypothesis of normal or t-distribution for the S&P 500 daily returns is not rejectable at significance level $\alpha = 0.05$.

Given such example, consider now a stationary real valued time series $\{x_t\}_{t=1}^{N}$ where N denotes the sample size and the x_t's are continuous measurements, such as daily stock returns or monthly number of traffic accidents. In addition to the dependence structure, the stationary distribution function $F(\cdot)$ of x_t also provides useful information for making inference. One can estimate the unknown continuous function $F(\cdot)$ by the following empirical cumulative distribution function (cdf):

$$F_N(x) = N^{-1} \sum_{t=1}^{N} I(x_t \leq x), x \in \mathbb{R}.$$

This estimator $F_N(\cdot)$ converges to $F(\cdot)$ at the rate of $N^{-1/2}$ as $N \to \infty$ (see Lemma 4), but the limiting process distribution not only depends on $F(\cdot)$ but also the autocovariance structure of $\{I(x_t \leq x)\}_{t=1}^{N}$. Thus it is impossible to obtain a Kolmogorov–Smirnov type distribution free SCB for $F(\cdot)$ based on $F_N(\cdot)$.

To overcome this difficulty, one can draw a SRS $X_1, \ldots, X_n$ with a smaller size $n \ll N$ from the time series $\{x_t\}_{t=1}^{N}$, and define the empirical distribution function $F_n(x)$ as

$$F_n(x) = n^{-1} \sum_{i=1}^{n} I(X_i \leq x), x \in \mathbb{R}.$$

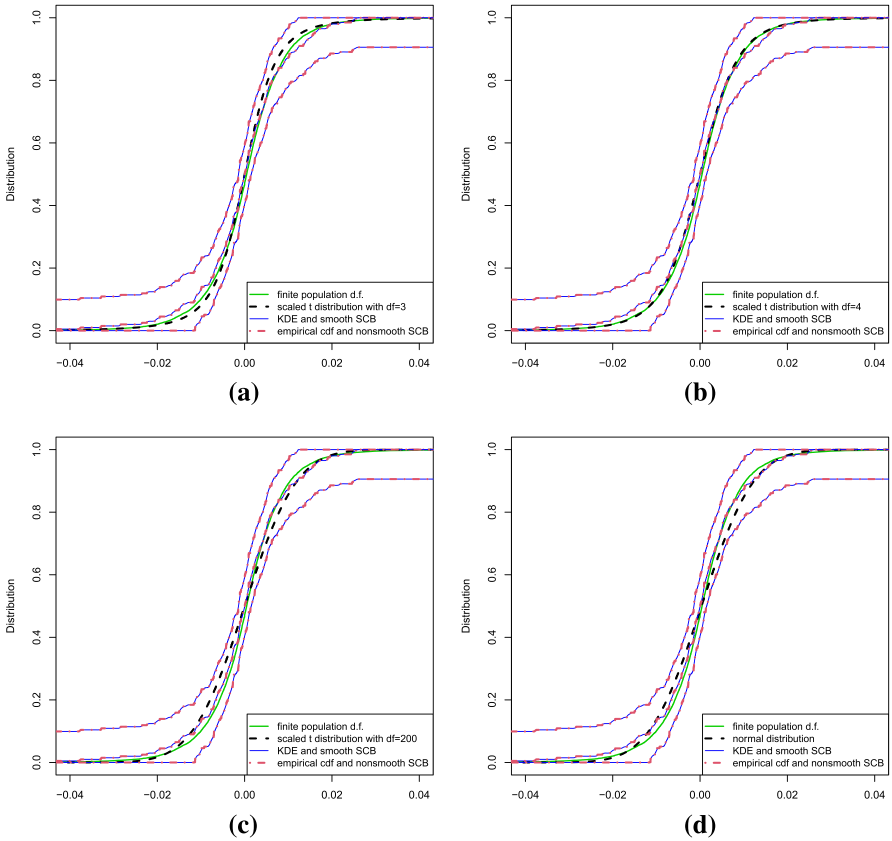

Fig. 3 The finite population distribution function F_N (thick line), smooth KDE and its accompanying corrected 95% SCBs (solid), empirical cdf and its accompanying corrected 95% SCBs (dashed) for the standardized S&P 500 daily returns $\{x_t\}_{t=1}^{17276}$ using SRS of size $n = 200$. The dotted line in four plots is: (**a**) rescaled t-distribution with degree of freedom 3; (**b**) rescaled t-distribution with degree of freedom 4; (**c**) rescaled t-distribution with degree of freedom 200; (**d**) normal distribution

Had the observations $X_1, \ldots, X_n$ been independent, the well-known Donsker's Theorem would entail that in uniform metric on the cadlag space $\mathcal{D}(-\infty, \infty)$

$$n^{1/2}\{F_n(\cdot) - F(\cdot)\} \xrightarrow{d} B\{F(\cdot)\}, \tag{1}$$

in which $B(t)$ denotes the Brownian bridge: $B(t) = W(t) - tW(1), t \in [0, 1]$, with $W(t), 0 \le t \le 1$ being the Wiener process. Therefore, the classic Kolmogorov–Smirnov asymptotic simultaneous confidence band (SCB) for $F(\cdot)$ of level $1 - \alpha$ would be:

$$\left[\max\left(F_n(x) - n^{-1/2}L_{1-\alpha}, 0\right), \min\left(F_n(x) + n^{-1/2}L_{1-\alpha}, 1\right)\right], \; x \in \mathbb{R},$$

Table 1 Quantiles $L_{1-\alpha}$ for the Kolmogorov distribution

$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.2$
1.63	1.36	1.22	1.07

where $L_{1-\alpha}$ is the $(1-\alpha)$-th quantile of the maximal absolute value of $B(t)$, i.e.,

$$\mathbb{P}\left[\sup_{t\in[0,1]} |B(t)| > L_{1-\alpha}\right] = \alpha, \forall \alpha \in (0,1). \tag{2}$$

Some commonly used values of $L_{1-\alpha}$ are listed in Table 1.

For a SRS $X_1, \ldots, X_n$ drawn from the time series $x_1, \ldots, x_N$ without replacement, a variant of (1) holds in spite of dependence, under Assumptions (A1) and (A3) in the next section. To be more precise, one considers the time series realization $x_1, \ldots, x_{N_k}$ together with the SRS $X_1, \ldots, X_{n_k}$ drawn from it, as being part of an infinite sequence of experiments $k = 1, 2, \ldots, \infty$, where for each k, N_k and n_k form one experiment and grow to infinity along the way in sync: $\lim_{k\to\infty} \min(n_k, N_k - n_k) = \infty$. Denote a sequence of time series realizations $\{\pi_k\}_{k=1}^{\infty}$, $\pi_k = \left\{x_1, x_2, \ldots, x_{N_k}\right\}$ together with its finite population distribution function

$$F_{N_k}(x) = N_k^{-1} \sum_{i=1}^{N_k} I(x_i \le x), \; x \in \mathbb{R}, \tag{3}$$

and the empirical cumulative distribution function (ECDF) based on $X_1, \ldots, X_{n_k}$

$$F_{n_k}(x) = n_k^{-1} \sum_{i=1}^{n_k} I(X_i \le x), \; x \in \mathbb{R}. \tag{4}$$

Making use of the finite population asymptotics in Rosén (1964), we will establish the asymptotics of $\sup_{x\in\mathbb{R}} \left|F_{N_k}(x) - F_{n_k}(x)\right|$ and $\sup_{x\in\mathbb{R}} \left|F(x) - F_{n_k}(x)\right|$, which are employed to construct a Kolmogorov–Smirnov asymptotic SCB for $F(\cdot)$ based on $F_{n_k}(\cdot)$.

For independent and identically distributed random sample, Yamato (1973), Reiss (1981), Falk (1985), Cheng and Peng (2002), and more recently Liu and Yang (2008), Xue and Wang (2010), Wang et al. (2013), Wang et al. (2016) had all argued that a smoothed version of ECDF is more preferable, since the smoothed estimator shares the smoothness feature with the true distribution function $F(\cdot)$, while the ECDF is a step function. The kernel distribution estimator (KDE) for $F(\cdot)$ is defined as:

$$\hat{F}_k(x) = \int_{-\infty}^{x} n_k^{-1} \sum_{i=1}^{n_k} K_h(u - X_i)\, du, x \in \mathbb{R}, \tag{5}$$

where $h = h_{n_k} > 0$ is the bandwidth and K is a kernel function, and $K_h(u) = K(u/h)/h$. A Kolmogorov–Smirnov asymptotic SCB for $F(\cdot)$ based on $\hat{F}_k(\cdot)$ will also be established, which is smooth.

In the statistics arsenal, SCB is a versatile tool for making inference on the entirety of a curve or function, and has been used in various contexts, which includes non-parametric regression: Song and Yang (2009), Wang and Yang (2009), Cai and Yang (2015), Zhang and Yang (2018); semiparametric dimension reduction: Gu and Yang (2015), Zheng et al. (2016); functional data analysis: Cardot and Josserand (2011), Degras (2011), Cao et al. (2012), Ma et al. (2012), Cardot et al. (2013), Song et al. (2014), Zheng et al. (2014), Gu et al. (2014), Cao et al. (2016); distribution function estimation for time series error: Wang et al. (2014), Kong et al. (2018).

Throughout this paper, one denotes the maximal deviation between two distribution functions $G_1(\cdot)$ and $G_2(\cdot)$ as

$$D(G_1, G_2) = \|G_1 - G_2\|_\infty = \sup_x |G_1(x) - G_2(x)|. \tag{6}$$

A finite population version of Donsker's Theorem, Theorem 1 states that

$$l_k \left\{F_{n_k}(\cdot) - F_{N_k}(\cdot)\right\} \xrightarrow{d} B\{F(\cdot)\},$$

in which $l_k = \left(n_k^{-1} - N_k^{-1}\right)^{-1/2} = n_k^{1/2}\left(\text{fpc}_k\right)^{-1/2}$ is a finite population corrected scale factor similar to $n_k^{1/2}$ for an i.i.d. sample of size n_k, with $\text{fpc}_k = 1 - n_k/N_k$ the *finite population correction* (fpc) factor. The Assumption (A4) that $n_k = o(N_k)$ and Lemma 4 then ensure that $D\left(F_{N_k}, F\right) = \mathcal{O}_p\left(N_k^{-1/2}\right) = o_p\left(n_k^{-1/2}\right)$ and $\lim_{k\to\infty} n_k^{-1/2}/l_k^{-1} = 1$, so both $l_k\left\{F_{n_k}(\cdot) - F(\cdot)\right\}$ and $n_k^{1/2}\left\{F_{n_k}(\cdot) - F(\cdot)\right\}$ converge to $B\{F(\cdot)\}$ in distribution. These facts then allow one to construct both "corrected" and "uncorrected" Kolmogorov–Smirnov SCB for $F(\cdot)$ based on $F_{n_k}(\cdot)$ with predetermined asymptotic coverage $1-\alpha$:

$$\left[\max\left(F_{n_k}(x) - l_k^{-1}L_{1-\alpha}, 0\right), \min\left(F_{n_k}(x) + l_k^{-1}L_{1-\alpha}, 1\right)\right],\ x \in \mathbb{R}. \tag{7}$$

$$\left[\max\left(F_{n_k}(x) - n_k^{-1/2}L_{1-\alpha}, 0\right), \min\left(F_{n_k}(x) + n_k^{-1/2}L_{1-\alpha}, 1\right)\right],\ x \in \mathbb{R}. \tag{8}$$

Theorem 2 shows in addition that $D\left(F_{n_k}, \hat{F}_k\right)$ is of order $o_p(l_k^{-1})$, which leads one to propose the "corrected" and "uncorrected" smooth SCB for $F(\cdot)$ based on $\hat{F}_k(\cdot)$ with predetermined asymptotic coverage $1-\alpha$:

$$\left[\max\left(\hat{F}_k(x) - l_k^{-1}L_{1-\alpha}, 0\right), \min\left(\hat{F}_k(x) + l_k^{-1}L_{1-\alpha}, 1\right)\right],\ x \in \mathbb{R}, \tag{9}$$

$$\left[\max\left(\hat{F}_k(x) - n_k^{-1/2}L_{1-\alpha}, 0\right), \min\left(\hat{F}_k(x) + n_k^{-1/2}L_{1-\alpha}, 1\right)\right],\ x \in \mathbb{R}. \tag{10}$$

Table 2 Coverage frequencies of **Model 1**: $x_t - \phi x_{t-1} = \varepsilon_t$, where $\varepsilon_t \sim N(0,1)$ and $\phi = 0.2$; left—smooth SCB based on KDE $\hat{F}_k$, right—nonsmooth SCB based on ECDF F_{n_k}; l_k^{-1} (corrected SCB), $n_k^{-1/2}$ (uncorrected SCB)

(n_k, N_k)	SCB	0.99	0.95	0.90	0.80
(200, 5000)	l_k^{-1}	0.994 0.993	0.946 0.939	0.901 0.894	0.817 0.797
	$n_k^{-1/2}$	0.995 0.995	0.952 0.951	0.907 0.906	0.832 0.823
(500, 15,000)	l_k^{-1}	0.989 0.989	0.953 0.951	0.897 0.896	0.815 0.810
	$n_k^{-1/2}$	0.991 0.991	0.961 0.959	0.904 0.903	0.827 0.824
(200, 20,000)	l_k^{-1}	0.993 0.992	0.956 0.951	0.905 0.901	0.807 0.792
	$n_k^{-1/2}$	0.993 0.992	0.959 0.953	0.910 0.903	0.813 0.797
(500, 20,000)	l_k^{-1}	0.992 0.992	0.962 0.963	0.927 0.927	0.839 0.833
	$n_k^{-1/2}$	0.995 0.994	0.967 0.967	0.929 0.929	0.845 0.842

In particular, the smooth SCB of (10) has exactly the same form as the smooth SCB of Wang et al. (2013).

The paper is organized as follows. Section 2 contains the main theoretical results on the four SCBs defined in (7), (8), (9) and (10), while Sect. 3 describes the steps to implement these SCBs. Simulation studies and analysis of the S&P 500 daily returns are reported in Sects. 4 and 5 with details. Section 6 discusses the contributions of the proposed SCBs in relation to the existing literature, while all technical proofs are in the Appendix.

2 Main results

The limiting distribution of stochastic process $l_k \left\{F_{n_k}(\cdot) - F_{N_k}(\cdot)\right\}$ is established in this section, together with the maximal deviation between $F_{n_k}(\cdot)$ and $\hat{F}_k(\cdot)$, and between $F_{N_k}(\cdot)$ and $F(\cdot)$, under rather mild assumptions. These asymptotics lead to Corollary 1 about the four SCBs in (7), (8), (9) and (10) for $F(\cdot)$, based on $F_{n_k}(\cdot)$ and $\hat{F}_k(\cdot)$.

For any $\mu \in (0,1]$ and nonnegative integer ν, denote by $C^{(\nu,\mu)}(\mathbb{R})$ the space of functions whose ν-th derivatives satisfy Hölder conditions of order μ

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

For any sequence $\{y_t, t = 0, \pm 1, \pm 2, \ldots\}$ of random variables of dimension $d > 0$, denote by $\mathcal{M}_a^b$ the σ -field generated by $y_a, \ldots, y_b$. The sequence is called α -mixing if

$$\alpha(n) := \sup\left\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{M}_1^k, B \in \mathcal{M}_{k+n}^{\infty}, k \geq 1\right\} \to 0.$$

The following general assumptions are needed:

(A1) The sequence $\{x_t, t = 0, \pm 1, \pm 2, \ldots\}$ is a stationary and ergodic time series satisfying the α- mixing condition with rate $\alpha(n) \ll n^{-6-\epsilon}$.
(A2) There exist an integer $\nu \geq 0$ and $\mu \in (1/2, 1]$ such that $F \in C^{(\nu,\mu)}(\mathbb{R})$, and $F(x)$ is uniformly continuous over $x \in \mathbb{R}$.
(A3) $\lim_{k\to\infty} \min(n_k, N_k - n_k) = \infty$.
(A4) $\lim_{k\to\infty} n_k/N_k = 0$, i.e., $\lim_{k\to\infty} \mathrm{fpc}_k = 1$.
(A5) The bandwidth $h = h_{n_k} > 0$ and $\lim_{k\to\infty} l_k h_{n_k}^{\nu+\mu} = 0$ (i.e., $\lim_{k\to\infty} n_k^{1/2} h_{n_k}^{\nu+\mu} = 0$).
(A6) The kernel $K \in C^{(0)}(\mathbb{R})$ and satisfies $K(u) = K(-u), \forall u \in \mathbb{R}$; $K(u) = 0$ if $|u| > 1$. It is an l-th order kernel for some even integer $l > \nu + \mu$, i.e., its moments $\mu_r(K) = \int K(w) w^r dw$ satisfy $\mu_0(K) \equiv 1$, $\mu_l(K) \neq 0$, $\mu_r(K) \equiv 0$ for any integer r, $0 < r < l$.

Assumption (A3) implies that n_k and $N_k - n_k$ both go to infinity as in Rosén (1964), while Assumptions (A2), (A5) and (A6) are similar to those in Wang et al. (2013). Assumption (A2) contains uniform continuity of $F(\cdot)$ in addition to $F \in C^{(\nu,\mu)}(\mathbb{R})$. Assumption (A6) allows the kernel K to have order higher than 2, so $\nu + \mu$ can be greater than 2. In contrast, K is restricted to be second order nonnegative kernel in Wang et al. (2013), thus a probability density, and $\nu = 0, 1$ so $\nu + \mu$ is always less than 2. Assumption (A4) gurantees that $\lim_{k\to\infty} n_k^{-1/2}/l_k^{-1} = 1$, and the difference $F_{N_k}(x) - F(x)$ is asymptotically $N_k^{-1/2} B\{F(x)\} = o_p\left(l_k^{-1}\right)$.

The following Theorem is an analog to Theorem 14.3 of Billingsley (1999) for the case of i.i.d. samples:

Theorem 1 *Under Assumptions (A1), (A3), there exist versions B_k^* of Brownian bridge such that as $k \to \infty$, $\sup_{x\in\mathbb{R}} \left| l_k \left\{F_{n_k}(x) - F_{N_k}(x)\right\} - B_k^*\{F(x)\}\right| \xrightarrow{a.s.} 0$ and consequently $l_k \left\{F_{n_k}(\cdot) - F_{N_k}(\cdot)\right\} \xrightarrow{d} B\{F(\cdot)\}$.*

The next Theorem extends Theorem 2.1 of Wang et al. (2013) to finite population:

Theorem 2 *Under Assumptions (A1)–(A6), as $k \to \infty$, the maximal deviation $D\left(F_{n_k}, \hat{F}_k\right)$ defined in (6) satisfies $l_k D\left(F_{n_k}, \hat{F}_k\right) = o_p(1)$. Consequently*

$$l_k \left\{\hat{F}_k(\cdot) - F_{N_k}(\cdot)\right\} \xrightarrow{d} B\{F(\cdot)\} \tag{11}$$

Moreover, $l_k D\left(F_{N_k}, F\right) \xrightarrow{d} 0$, and $n_k^{-1/2}/l_k^{-1} \to 1$, hence

$$\begin{aligned} &l_k \left\{F_{n_k}(\cdot) - F(\cdot)\right\} \xrightarrow{d} B\{F(\cdot)\}, n_k^{1/2} \left\{F_{n_k}(\cdot) - F(\cdot)\right\} \xrightarrow{d} B\{F(\cdot)\} \\ &l_k \left\{\hat{F}_k(\cdot) - F(\cdot)\right\} \xrightarrow{d} B\{F(\cdot)\}, n_k^{1/2} \left\{\hat{F}_k(\cdot) - F(\cdot)\right\} \xrightarrow{d} B\{F(\cdot)\} \end{aligned} \tag{12}$$

Theorems 1 and 2 imply the next Corollary, which provides theoretical justifications for all four asymptotic SCBs of $F(\cdot)$ in (7), (8), (9) and (10). The quantile $L_{1-\alpha}$ of

Table 3 Coverage frequencies of **Model 1**: $x_t - \phi x_{t-1} = \varepsilon_t$, where $\varepsilon_t \sim \mathrm{N}(0, 1)$ and $\phi = -0.4$; left—smooth SCB based on KDE $\hat{F}_k$, right—nonsmooth SCB based on ECDF F_{n_k}; l_k^{-1} (corrected SCB), $n_k^{-1/2}$ (uncorrected SCB)

(n_k, N_k)	SCB	0.99	0.95	0.90	0.80
(200, 5000)	l_k^{-1}	0.984 0.984	0.957 0.951	0.909 0.907	0.827 0.815
	$n_k^{-1/2}$	0.988 0.986	0.963 0.958	0.915 0.910	0.841 0.833
(500, 15,000)	l_k^{-1}	0.987 0.987	0.951 0.947	0.900 0.899	0.795 0.790
	$n_k^{-1/2}$	0.989 0.988	0.955 0.954	0.906 0.905	0.814 0.811
(200, 20,000)	l_k^{-1}	0.996 0.993	0.957 0.955	0.916 0.911	0.835 0.819
	$n_k^{-1/2}$	0.996 0.995	0.959 0.956	0.918 0.915	0.837 0.826
(500, 20,000)	l_k^{-1}	0.990 0.990	0.954 0.952	0.915 0.914	0.832 0.830
	$n_k^{-1/2}$	0.990 0.990	0.960 0.960	0.922 0.920	0.851 0.845

Table 4 Coverage frequencies of **Model 2**: $x_t = \left(\varepsilon_t + \varepsilon_{t-1}\right)/2$, $x_t \sim$ Cauchy (0, 1); left—smooth SCB based on KDE $\hat{F}_k$, right—nonsmooth SCB based on ECDF F_{n_k}; l_k^{-1} (corrected SCB), $n_k^{-1/2}$ (uncorrected SCB)

(n_k, N_k)	SCB	0.99	0.95	0.90	0.80
(200, 5000)	l_k^{-1}	0.987 0.986	0.958 0.953	0.911 0.897	0.800 0.786
	$n_k^{-1/2}$	0.991 0.989	0.964 0.959	0.927 0.912	0.825 0.799
(500, 15,000)	l_k^{-1}	0.994 0.991	0.954 0.948	0.900 0.892	0.780 0.767
	$n_k^{-1/2}$	0.995 0.995	0.962 0.957	0.912 0.904	0.799 0.784
(200, 20,000)	l_k^{-1}	0.992 0.991	0.953 0.946	0.896 0.884	0.810 0.796
	$n_k^{-1/2}$	0.993 0.991	0.954 0.948	0.900 0.888	0.811 0.802
(500, 20,000)	l_k^{-1}	0.987 0.986	0.941 0.936	0.893 0.884	0.789 0.781
	$n_k^{-1/2}$	0.988 0.987	0.943 0.942	0.900 0.894	0.802 0.788

the Kolmogorov distribution is defined in (2) and $1 - \alpha \in (0, 1)$ a predetermined confidence level.

Corollary 1 *Under Assumptions (A1)–(A6), for any* $\alpha \in (0, 1)$

$$\lim_{k\to\infty} \mathbb{P}\left[n_k^{1/2} D\left(F_{n_k}, F\right) \le L_{1-\alpha}\right] = 1 - \alpha\ , \ \lim_{k\to\infty} \mathbb{P}\left[l_k D\left(F_{n_k}, F\right) \le L_{1-\alpha}\right] = 1 - \alpha, ,$$
$$\lim_{k\to\infty} \mathbb{P}\left[n_k^{1/2} D\left(\hat{F}_k, F\right) \le L_{1-\alpha}\right] = 1 - \alpha\ , \ \lim_{k\to\infty} \mathbb{P}\left[l_k D\left(\hat{F}_k, F\right) \le L_{1-\alpha}\right] = 1 - \alpha, .$$

Consequently, the SCBs in (7)*,* (8)*,* (9) *and* (10) *are all asymptotically* $100(1-\alpha)\%$ *correct for* $F(\cdot)$*.*

Table 5 Coverage frequencies of **Model 3**: $x_t = \varepsilon_t + \theta\varepsilon_{t-1}, \varepsilon_t \sim \mathrm{E}(0, 1), \theta = 2$; left—smooth SCB based on KDE $\hat{F}_k$, right—nonsmooth SCB based on ECDF F_{n_k}; l_k^{-1} (corrected SCB), $n_k^{-1/2}$ (uncorrected SCB)

(n_k, N_k)	SCB	0.99	0.95	0.90	0.80
(200, 5000)	l_k^{-1}	0.986 0.982	0.941 0.936	0.898 0.881	0.805 0.784
	$n_k^{-1/2}$	0.991 0.986	0.951 0.943	0.908 0.898	0.816 0.799
(500, 15,000)	l_k^{-1}	0.990 0.987	0.939 0.937	0.899 0.889	0.814 0.805
	$n_k^{-1/2}$	0.994 0.993	0.963 0.960	0.918 0.910	0.835 0.827
(200, 20,000)	l_k^{-1}	0.990 0.988	0.957 0.951	0.904 0.893	0.810 0.784
	$n_k^{-1/2}$	0.990 0.988	0.961 0.951	0.907 0.895	0.816 0.788
(500, 20,000)	l_k^{-1}	0.988 0.988	0.956 0.953	0.917 0.908	0.824 0.809
	$n_k^{-1/2}$	0.988 0.988	0.960 0.957	0.922 0.916	0.835 0.825

3 Implementation

This section describes how SCBs are constructed based on estimators $F_{n_k}(\cdot)$ and $\hat{F}_k(\cdot)$ defined in (4) and (5), respectively. According to Corollary 1, for sample size $n_k > 50$, the corrected and uncorrected nonsmooth and smooth SCBs for the true cdf are computed and named as follows:

SCB in (7), "corrected, nonsmooth",
SCB in (9), "corrected, smooth",
SCB in (8), "uncorrected, nonsmooth",
SCB in (10), "uncorrected, smooth".

By using the quartic kernel $K(u) = 15\left(1-u^2\right)^2 I\{|u| \le 1\}/16$, the proposed function $\hat{F}_k(x)$ is computed as

$$\hat{F}_k(x) = n_k^{-1}\sum_{i=1}^{n_k}\int_{-\infty}^{x} h^{-1}K\left(\frac{u-X_i}{h}\right)du$$

in which $h = \mathrm{IQR} \times l_k^{-2}$, where IQR stands for the Inter-Quartile Range of $\left\{X_1, \ldots, X_{n_k}\right\}$. The bandwidth h automatically satisfies Assumption (A5) and is similar to that used in Wang et al. (2013).

4 Simulation

4.1 General simulation studies

In this section, we display the performance of the various SCBs on estimators $F_{n_k}(\cdot)$ and $\hat{F}_k(\cdot)$. Time series data $\{x_t\}_{t=1}^N$ is generated from three different models. The first one is causal Gaussian AR(1), the other two are 2-dependent, hence all three are geometrically ergodic and α -mixing.

Model 1. The data $\{x_t\}_{t=1}^N$ is a segment of $\{x_t\}_{t=-\infty}^{\infty}$ with

$$x_t - \phi x_{t-1} = \varepsilon_t,\ x_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \sim \mathrm{N}\left(0, \left(1-\phi^2\right)^{-1}\right)$$

where i.i.d. $\varepsilon_t \sim \mathrm{N}(0,1)$, $t = 0, \pm 1, \pm 2, \ldots, |\phi| < 1$ and the infinite series converges almost surely. The above entails that the stationary distribution function of x_t is

$$F(x) = \Phi\left\{\left(1-\phi^2\right)^{1/2} x\right\},$$

in which $\Phi(\cdot)$ is the standard normal distribution function. In our experiments, the parameter ϕ is taken to be 0.2, −0.4.

Model 2. The data $\{x_t\}_{t=1}^N$ is a segment of $\{x_t\}_{t=-\infty}^{\infty}$ with

$$x_t = (\varepsilon_t + \varepsilon_{t-1})/2,\ x_t \sim \text{Cauchy}(0,1)$$

where i.i.d. $\varepsilon_t \sim \text{Cauchy}(0,1)$, $t = 0, \pm 1, \pm 2, \ldots$ The above entails that the stationary distribution function of x_t is

$$F(x) = \pi^{-1} \arctan x + 1/2.$$

Model 3. The data $\{x_t\}_{t=1}^N$ is a segment of $\{x_t\}_{t=-\infty}^{\infty}$ with

$$x_t = \varepsilon_t + \theta \varepsilon_{t-1},$$

where i.i.d. $\varepsilon_t \sim \mathrm{E}(0,1)$, $t = 0, \pm 1, \pm 2, \ldots, \theta \in (0, \infty)$. The above entails that the stationary distribution function of x_t is

$$F(x) = \left\{1 + \left(e^{-x} - \theta e^{-x/\theta}\right)(\theta - 1)^{-1}\right\} I(x > 0).$$

In our experiments, the parameter θ is taken to be 2.

A realization π_k of size N_k is first generated from each of the above model, SRS $\left\{X_{n_1}, \ldots, X_{n_k}\right\}$ of size n_k are then drawn without replacement from π_k. The sample sizes and time series lengths are selected as $(n_k, N_k) = (200, 5000)$, (500, 15,000),

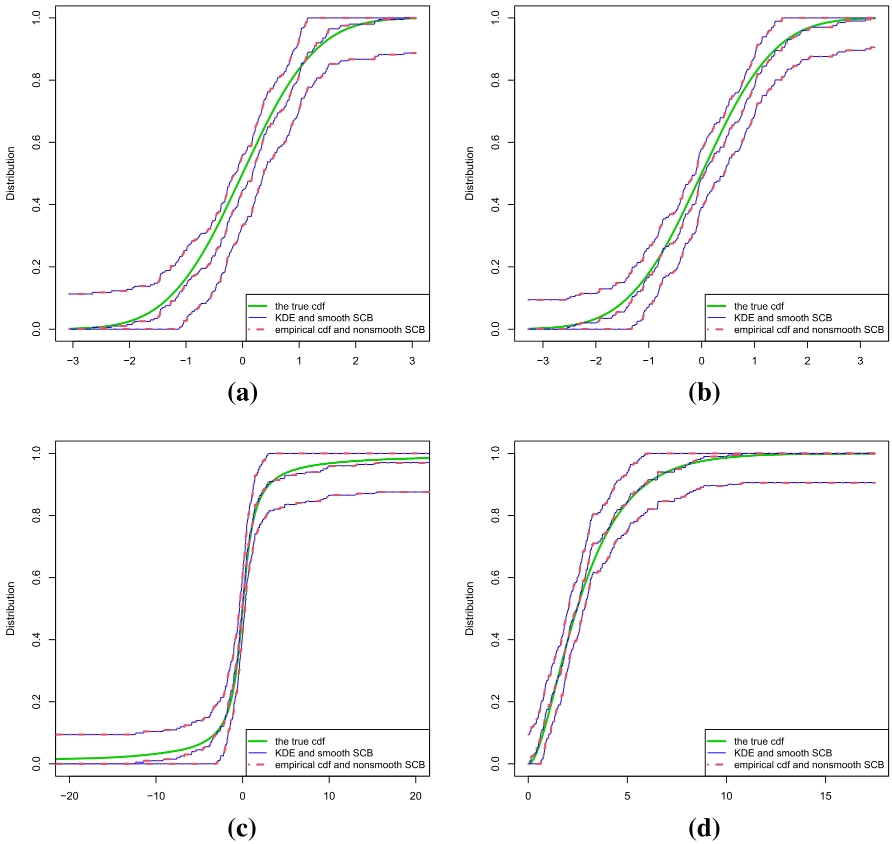

Fig. 4 The true cdf (thick line), smooth KDE and its accompanying corrected 95% SCBs (solid), empirical cdf and its accompanying corrected 95% SCBs (dashed) with $(n_k, N_k) = (200, 5000)$ for different models: (**a**, **b**) **Model 1** with $\phi = 0.2, -0.4$ respectively; (**c**) **Model 2**; (**d**) **Model 3** with $\theta = 2$. It shows negligible difference between KDE and empirical cdf estimator

(200, 20,000), (500, 20,000), with confidence levels $1-\alpha = 0.99, 0.95, 0.90, 0.80$ for constructing SCBs. Tables 2, 3, 4 and 5 display the frequencies out of 1000 replications of the true function $F(\cdot)$ being contained at all data points $\{x_1, x_2, \ldots, x_{N_k}\}$ by various SCBs. Main findings are summarized as follows:

1. In general, coverage frequencies of uncorrected SCBs are always slightly higher than the corrected ones. Smooth and nonsmooth SCBs have nearly the same coverage frequencies. Corrected SCBs have coverage frequencies closer to the nominal levels than uncorrected ones.
2. The performance of SCBs for all cdf $F(\cdot)$ of different models, with different sample sizes and confidence levels is satisfactory across the board (the coverage frequency is close to the predetermined nominal confidence level), which implies that the method is robust and widely applicable.
3. Since both n_k and N_k are quite large in the simulation, the ratio n_k/N_k makes little difference in the performance of the estimation and SCB coverage frequencies.

To visualize the SCBs, Fig. 4 depicts the true cdf $F(\cdot)$ (thick) in different models, the smooth KDE $\hat{F}_k(\cdot)$ together with its 95% SCB (solid), the empirical cdf $F_{n_k}(\cdot)$ together with its 95% SCB (dashed), for one specifically simulated time series and one sample in each simulation setting. In all plots we use the sample with the median confidence band width in the 1000 runs. To save space, only the combination of $(n_k, N_k) = (200, 5000)$ is shown. The figures for other combinations are similar. One clearly sees that the SCBs constructed based on $F_{n_k}(\cdot)$ and $\hat{F}_k(\cdot)$, and the estimators themselves are nearly indistinguishable from each other. For the same population size of 5000, the SCBs tend to be narrower as the sample size n_k increases and l_k^{-1} decreases.

4.2 Comparison with parametric SCB

In this subsection, simulations are conducted to compare the proposed Kolmogorov–Smirnov type SCBs with parametric ones. To the best of our knowledge, there are no other nonparametric SCBs for time series distribution function to make a comparison with ours.

Given a time series $\{x_t\}_{t=1}^N$, if one naively assumes that the data is generated from a causal Gaussian AR(1) model:

$$(x_t - \mu) - \phi(x_{t-1} - \mu) = \varepsilon_t, \quad \varepsilon_t \sim \text{IID } N(0, \sigma^2).$$

According to Lemma 1, a $100(1-\alpha)\%$ confidence interval for μ is

$$\left[\overline{x}_N - N^{-1/2}(1-\phi)^{-1}\sigma z_{1-\alpha/2}, \overline{x}_N - N^{-1/2}(1-\phi)^{-1}\sigma z_{\alpha/2}\right],$$

where $\overline{x}_N = N^{-1}\sum_{t=1}^N x_t$. Lemma 2 provides consistent estimators for the unknown ϕ and σ^2 as $\hat{\phi} = \hat{\gamma}(1)/\hat{\gamma}(0)$ and $\hat{\sigma}^2 = \left(\hat{\gamma}^2(0) - \hat{\gamma}^2(1)\right)/\hat{\gamma}(0)$ with

$$\hat{\gamma}(l) = N^{-1}\sum_{t=1}^{n-l}(x_t - \overline{x}_N)(x_{t+l} - \overline{x}_N), \quad l = 0, 1.$$

For notation simplicity, denote $\underline{\mu} = \overline{x}_N - N^{-1/2}(1-\hat{\phi})^{-1}\hat{\sigma} z_{1-\alpha/2}$ and $\overline{\mu} = \overline{x}_N - N^{-1/2}(1-\hat{\phi})^{-1}\hat{\sigma} z_{\alpha/2}$. Since $\Phi(x-\mu)$ has monotonicity in x and linearity in μ, a $100(1-\alpha)\%$ parametric confidence band for $F(x)$ is constructed as

$$\left[\Phi\left\{\left(1-\hat{\phi}^2\right)^{1/2}(x-\overline{\mu})\right\}, \Phi\left\{\left(1-\hat{\phi}^2\right)^{1/2}(x-\underline{\mu})\right\}\right], \ x \in \mathbb{R}. \qquad (13)$$

Time series data $\{x_t\}_{t=1}^N$ are generated by two different models. The first one is **Model 2**, the second one the following:

Table 6 Coverage frequencies of **Model 2**: $x_t = \left(\varepsilon_t + \varepsilon_{t-1}\right)/2$, $x_t \sim$ Cauchy (0, 1); left—smooth SCB based on KDE $\hat{F}_k$, right—nonsmooth SCB based on ECDF F_{n_k}; l_k^{-1} (corrected SCB), $n_k^{-1/2}$ (uncorrected SCB), parametric (parametric SCB)

(n_k, N_k)	SCB	0.99	0.95	0.90	0.80
(200, 5000)	l_k^{-1}	0.987 0.986	0.958 0.953	0.911 0.897	0.800 0.786
	$n_k^{-1/2}$	0.991 0.989	0.964 0.959	0.927 0.912	0.825 0.799
	Parametric	0	0	0	0
(500, 15,000)	l_k^{-1}	0.994 0.991	0.954 0.948	0.900 0.892	0.780 0.767
	$n_k^{-1/2}$	0.995 0.995	0.962 0.957	0.912 0.904	0.799 0.784
	Parametric	0	0	0	0
(200, 20,000)	l_k^{-1}	0.992 0.991	0.953 0.946	0.896 0.884	0.810 0.796
	$n_k^{-1/2}$	0.993 0.991	0.954 0.948	0.900 0.888	0.811 0.802
	Parametric	0	0	0	0
(500, 20,000)	l_k^{-1}	0.987 0.986	0.941 0.936	0.893 0.884	0.789 0.781
	$n_k^{-1/2}$	0.988 0.987	0.943 0.942	0.900 0.894	0.802 0.788
	Parametric	0	0	0	0

Model 4. The data $\{x_t\}_{t=1}^{N}$ is a segment of $\{x_t\}_{t=-\infty}^{\infty}$ with

$$(x_t - \mu) - \phi\,(x_{t-1} - \mu) = \varepsilon_t, \quad x_t = \mu + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \sim \mathrm{N}\left(\mu, \left(1 - \phi^2\right)^{-1}\right)$$

where i.i.d. $\varepsilon_t \sim \mathrm{N}\,(0, 1)$, $t = 0, \pm 1, \pm 2, \ldots$, $|\phi| < 1$ and the infinite series converges almost surely. The above entails that the stationary distribution function of x_t is

$$F(x) = \Phi\left\{\left(1 - \phi^2\right)^{1/2}(x - \mu)\right\},$$

in which $\Phi(\cdot)$ is the standard normal distribution function. In our experiments, the parameter ϕ is taken to be 0.2 and μ is 2.

Again, SRS $\left\{X_{n_1}, \ldots, X_{n_k}\right\}$ of size n_k are drawn without replacement after a realization π_k of size N_k is generated from the above two models. The combinations of (n_k, N_k) are the same as in the previous subsection. Four Kolmogorov–Smirnov type SCBs are constructed along with the parametric SCB with confidence levels $1 - \alpha = 0.99, 0.95, 0.90, 0.80$ for comparing performance.

Tables 6 and 7 display the coverage frequencies over 1000 replications of the various SCBs. It is clear that the parametric SCB severely suffers from the problem of model misspecification, leading to poor coverage frequency in Model 2, see Table 6. The KS type SCBs, on the other hand, always have coverage frequencies closer to the nominal level than the parametric SCB, even when the model is correctly specified as Model 4, see Table 7.

Table 7 Coverage frequencies of **Model 4**: $(x_t - \mu) - \phi(x_{t-1} - \mu) = \varepsilon_t$, where $\varepsilon_t \sim N(0, 1)$, $\mu = 2$ and $\phi = 0.2$; left—smooth SCB based on KDE $\hat{F}_k$, right—nonsmooth SCB based on ECDF F_{n_k}; l_k^{-1} (corrected SCB), $n_k^{-1/2}$ (uncorrected SCB), parametric (parametric SCB)

(n_k, N_k)	SCB	0.99	0.95	0.90	0.80
(200, 5000)	l_k^{-1}	0.987 0.987	0.951 0.946	0.900 0.897	0.811 0.802
	$n_k^{-1/2}$	0.989 0.987	0.959 0.952	0.912 0.907	0.829 0.819
	Parametric	0.972	0.920	0.885	0.776
(500, 15,000)	l_k^{-1}	0.992 0.992	0.957 0.957	0.907 0.904	0.828 0.822
	$n_k^{-1/2}$	0.992 0.992	0.959 0.959	0.921 0.916	0.842 0.839
	Parametric	0.970	0.931	0.878	0.763
(200, 20,000)	l_k^{-1}	0.995 0.994	0.961 0.957	0.911 0.902	0.828 0.818
	$n_k^{-1/2}$	0.995 0.995	0.961 0.958	0.911 0.906	0.834 0.825
	Parametric	0.979	0.932	0.880	0.784
(500, 20,000)	l_k^{-1}	0.989 0.989	0.949 0.949	0.905 0.904	0.818 0.813
	$n_k^{-1/2}$	0.991 0.991	0.953 0.950	0.909 0.910	0.830 0.825
	Parametric	0.976	0.935	0.877	0.788

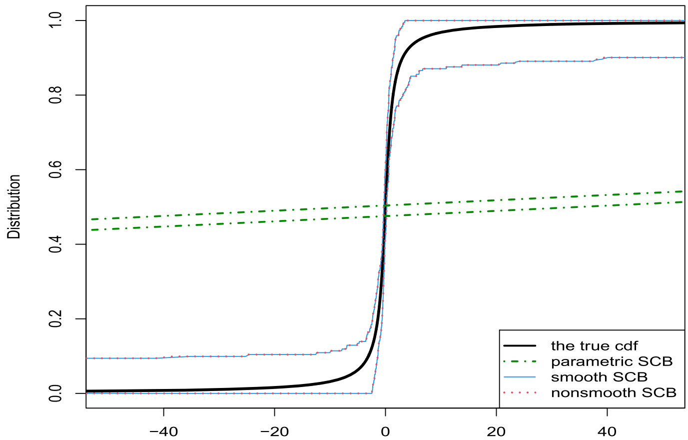

Fig. 5 Corrected smooth, corrected nonsmooth and parametric 95% SCBs with $(n_k, N_k) = (200, 5000)$ for **Model 2**

To visualize the SCBs, Figs. 5 and 6 show the true cdf $F(\cdot)$ (thick), the corrected smooth 95% SCB (solid), the corrected nonsmooth 95% SCB (dotted), the parametric 95% SCB (dashed) based on the combination of $(n_k, N_k) = (200, 5000)$. One clearly sees that the parametric SCB performs well in Fig. 6 but extremely poorly in Fig. 5, most likely due to the use of incorrect parametric form for the distribution function in Model 2. Our methods enjoy the advantage of stability and computational ease, thus reliable and efficient for practical applications.

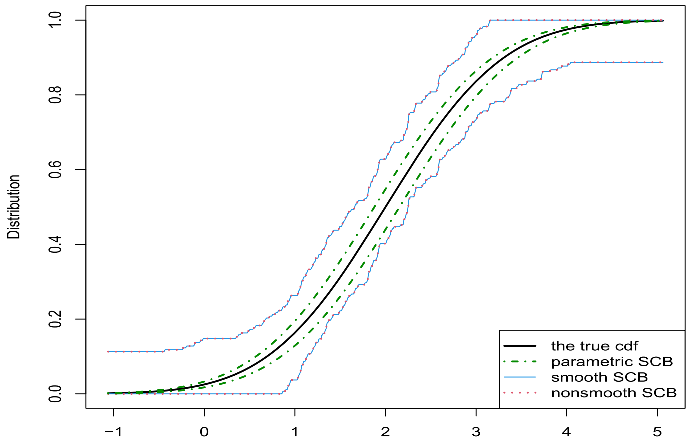

Fig. 6 Corrected smooth, corrected nonsmooth and parametric 95% SCBs with $(n_k, N_k) = (200, 5000)$ for **Model 4** with $\phi = 0.2$ and $\mu = 2$

5 Application

In this section, the proposed method is applied to the S&P 500 daily returns data, already discussed in the introduction. This data set includes observations from January 3, 1950 to August 28, 2018, a total of 17,277 closing prices SPI_t, $t = 0, \ldots, 17{,}276$, which were downloaded from https://finance.yahoo.com. The daily returns are calculated by $y_t = \log(\text{SPI}_t/\text{SPI}_{t-1})$, $t = 1, \ldots, 17{,}276$. The plot of $\{y_t\}_{t=1}^{17{,}276}$ is in Fig. 1a, which, as mentioned in the introduction, exhibits rather pronounced nonstationarity over the entire 68 years. It is thus meaningless to draw any conclusions on the distribution of this raw return series $\{y_t\}_{t=1}^{17{,}276}$. Instead, a cubic spline curve $\{g_t\}_{t=1}^{17{,}276}$ is fitted to the slowly varying trend of $\{y_t^2\}_{t=1}^{17{,}276}$, as in Shao and Yang (2017) and Zhang et al. (2020). Then the standardized returns $x_t = y_t g_t^{-1/2}$, $1 \le t \le 17{,}276$ are obtained as a stationary time series. The time plot of $\{x_t\}_{t=1}^{17{,}276}$ is in Fig. 1b, whose distributional properties are studied.

An ad hoc initial analysis is done by constructing an 99% Kolmogorov–Smirnov SCB based on $\{x_t\}_{t=1}^{17{,}276}$, namely

$$\left[\max\left(F_N(x) - 1.63N^{-1/2}, 0\right), \min\left(F_N(x) + 1.63N^{-1/2}, 1\right)\right], \ x \in \mathbb{R},$$

where $F_N(x) = N^{-1}\sum_{t=1}^{N} I(x_t \le x)$ with $N = 17{,}276$. The 99% Kolmogorov–Smirnov SCB is shown in Fig. 2 as the solid lines, with the empirical cumulative distribution function F_N the dotted line in the middle. The closest normal approximation to F_N is shown in Fig. 2 as the thick line with the distribution function $\Phi\left\{(x - \bar{x}_N)/\hat{s}_N\right\}$, where $\bar{x}_N$ and $\hat{s}_N^2$ are sample mean and sample variance of $\{x_t\}_{t=1}^{17{,}276}$ respectively. A naive statistician could conclude that at significance level 0.01, normality is rejected for F_N because the normal distribution function falls outside the 99% SCB.

The above analysis is incorrect because the Kolmogorov–Smirnov SCB is computed without regard to the dependence in $\{x_t\}_{t=1}^{17,276}$. Instead, one can draw a SRS from $\{x_t\}_{t=1}^{17,276}$ and compute corrected SCBs from (7) and (9), shown in Fig. 3. Clearly, all SCBs contain properly rescaled t-distribution functions with degrees of freedom 3, 4, 200 and normal distribution in entirety. Notice that all SCBs are reliable as the combination $n_k = 200$, $N_k = 17{,}276$ is in the range of sample sizes in simulation examples from previous section with satisfactory results. This well-founded analysis leads to the somewhat surprising conclusion that the distribution of $\{x_t\}_{t=1}^{17,276}$ could be either fat-tailed such as rescaled t-distribution with degree of freedom 3, 4, or, normal distribution.

6 Conclusions

In this work, Kolmogorov–Smirnov type SCBs are constructed by using SRS drawn from time series realization, and shown both theoretically and numerically to perform well despite dependence in the sample and in the time series. These SCBs are useful for testing against any hypothesis on the stationary distribution function of time series such as normality, fat-tailedness, etc., and are theoretically reliable, easy to implement alternatives to existing ad hoc approaches. Further research may yield multivariate and/or conditional extensions, and versions for other dependent data such as spatial-temporal data, functional data, etc. In addition, constructing prediction intervals for future observations of time series is a potentially interesting direction to pursue with the aid of KDE.

Appendix

Throughout this section, c denotes any positive constant and $\mathcal{O}_p$ (or o_p) a sequence of random variables of certain order in probability. In addition, u_p denotes a sequence of random functions which are o_p uniformly defined in the domain. For any continuous function ϕ defined on an interval $\mathcal{I}$, the modulus of continuity is defined as $\omega(\phi, \Delta) = \sup_{x,x' \in \mathcal{I}, |x-x'| \le \Delta} \left|\phi\left(x'\right) - \phi(x)\right|$

Lemma 1 (Theorem 7.1,2, Brockwell and Davis (1991)) *If $\{X_t\}_{t=1}^n$ is the stationary process,*

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \quad \varepsilon_t \sim IID(0, \sigma^2)$$

with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ and $\sum_{j=-\infty}^{\infty} |\psi_j| \neq 0$, then $\overline{X}_n$ is $AN\big(\mu, n^{-1}v\big)$, where $\overline{X}_n = n^{-1}\sum_{t=1}^{n} X_t$, $v = \sum_{h=-\infty}^{\infty} \gamma(h) = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j\right)^2$, and $\gamma(\cdot)$ is the auto-covariance function of $\{X_t\}_{t=1}^n$.

Lemma 2 (Theorem 8.1, Brockwell and Davis (1991)) *If* $\{X_t\}_{t=1}^n$ *is the zero-mean causal AR*(p) *process,*

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t, \quad Z_t \sim IID(0, \sigma^2)$$

and $\hat{\boldsymbol{\phi}}$ *is the Yule-Walker estimator of* $\boldsymbol{\phi}$*, that is* $\hat{\boldsymbol{\phi}} = \boldsymbol{\Gamma}_p^{-1}\hat{\boldsymbol{\gamma}}_p$ *with* $\hat{\boldsymbol{\Gamma}}_p = \left\{\hat{\gamma}(i-j)\right\}_{i,j=1}^p$ *and* $\hat{\boldsymbol{\gamma}}_p = \left(\hat{\gamma}(1), \ldots, \hat{\gamma}(p)\right)^\top$*, then*

$$\sqrt{n}\left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right) \xrightarrow{d} N\left(0, \sigma^2 \boldsymbol{\Gamma}_p^{-1}\right)$$

where $\boldsymbol{\Gamma}_p$ *is the covariance matrix with* $\boldsymbol{\Gamma}_p = \{\gamma(i-j)\}_{i,j=1}^p$*. Moreover,*

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2,$$

where $\hat{\sigma}^2 = \hat{\gamma}_0 - \hat{\boldsymbol{\phi}}^\top \hat{\boldsymbol{\gamma}}_p$*.*

A.1 Preliminary results on weak convergence

The next weak convergence result extends (1) of the Donsker's Theorem to strongly mixing time series.

Lemma 3 (Deo 1973) *Let* $\{\xi_n : -\infty < n < \infty\}$ *be a strictly stationary sequence of random variables,* $\{F_n(t) : 0 \le t \le 1\}$ *be the empirical process for* $\xi_1, \xi_2, \ldots, \xi_n$*, i.e.,* $F_n(t) = n^{-1}\sum_{i=1}^n I_{[0,t]}(\xi_i)$ *where* $I_{[0,t]}(\cdot)$ *is the indicator function of the interval* $[0, t]$*. Suppose that* $0 \le \xi_0 \le 1$ *and* ξ_0 *have continuous distribution function* F *with* $F(0) = 0$ *and* $F(1) = 1$*. Normalize* $F_n(t)$ *as*

$$Y_n(t) = n^{1/2}\left(F_n(t) - F(t)\right), \qquad 0 \le t \le 1.$$

For $0 \le t \le 1$*, define the function* g_t *by*

$$g_t(x) = I_{[0,t]}(x) - F(t),$$

and suppose further that $\{\xi_n\}$ *satisfies the mixing condition*

$$\sum_{n=1}^{\infty} n^2 \alpha(n)^{1/2-\tau} < \infty \qquad for\ some \quad \tau \in (0, 1/2).$$

Then the sequence $\{Y_n(t) : 0 \le t \le 1\}$ *of normalized empirical processes converges weakly in* $\mathcal{D}[0, 1]$ *to a Gaussian random function* $\{Y(t) : 0 \le t \le 1\}$ *specified by*

$\mathbb{E}(Y(t)) = 0$ and

$$\mathbb{E}\{Y(s)Y(t)\} = \mathbb{E}\{g_s(\xi_0)\, g_t(\xi_0)\} + \sum_{k=1}^{\infty} \mathbb{E}\{g_s(\xi_0)\, g_t(\xi_k)\} + \sum_{k=1}^{\infty} \mathbb{E}\{g_s(\xi_k)\, g_t(\xi_0)\} \tag{14}$$

Furthermore, the series in (14) converges absolutely and the sample paths of Y are continuous with probability one.

The following lemma yields a uniformly continuous Gaussian limiting process $\zeta(\cdot)$ on $\mathbb{R}$ for the empirical process $N_k^{1/2}\left\{F_{N_k}(\cdot) - F(\cdot)\right\}$, which is used in the proof of Theorem 2.

Lemma 4 *Under Assumptions (A1) and (A2), there exists a mean-zero Gaussian process $Y(\cdot)$ whose sample path is continuous on $[0, 1]$ with probability one such that as $k \to \infty$, $N_k^{1/2}\left\{F_{N_k}(\cdot) - F(\cdot)\right\} \xrightarrow{d} \zeta(\cdot) = Y(F(\cdot))$. Furthermore, the process $\zeta(\cdot)$ is uniformly continuous on $\mathbb{R}$ with modulus of continuity $\omega(\zeta, \Delta) \leq \omega(Y, \omega(F, \Delta)) \to 0$ a.s. as $\Delta \to 0$.*

Proof. Define a transformed time series $u_i = F(x_t)$, $i = 0, \pm 1, \pm 2, \ldots$. For any $x \in \mathbb{R}$, let $t = F(x) \in [0, 1]$, then $F_{N_k}(x) = F_{U,N_k}(t)$, in which

$$F_{U,N_k}(t) = N_k^{-1} \sum_{i=1}^{N_k} I\{u_i \leq t\}.$$

The α-mixing coefficients for $\{u_i\}_{i=-\infty}^{\infty}$ is the same as those for $\{x_i\}_{i=-\infty}^{\infty}$, which satisfy Assumption (A1) that $\alpha(n) \ll n^{-6-\epsilon}$, hence there exists $\tau \in (0, 1/2)$ such that $\alpha(n)^{1/2-\tau} \ll n^{-3}$, and thus $\sum_{n=1}^{\infty} n^2 \alpha(n)^{1/2-\tau} < \infty$. Then applying Lemma 3 with ξ_i replaced by u_i, one has $N_k^{1/2}\{F_{U,N_k}(t) - t\} \to Y(t)$.

Define $\zeta(x) = Y(F(x))$, then $N_k^{1/2}\left\{F_{N_k}(\cdot) - F(\cdot)\right\} \xrightarrow{d} \zeta(\cdot)$ as $k \to \infty$ and

$$\begin{aligned} \sup_{x,x' \in \mathbb{R}, |x-x'| \leq \Delta} \left|\zeta(x) - \zeta\left(x'\right)\right| &= \sup_{x,x' \in \mathbb{R}, |x-x'| \leq \Delta} \left|Y(F(x)) - Y\left(F\left(x'\right)\right)\right| \\ &\leq \sup_{t,t' \in [0,1], |t-t'| \leq \omega(F,\Delta)} \left|Y(t) - Y\left(t'\right)\right| \leq \omega(Y, \omega(F, \Delta)) \end{aligned}$$

The uniform continuity of $F(\cdot)$ is guaranteed by Assumption (A2), and almost sure uniform continuity of $Y(\cdot)$ by the fact that sample paths of $Y(\cdot)$ are almost surely continuous over the compact interval $[0, 1]$. These facts imply that $\omega(Y, \omega(F, \Delta)) \to 0$ *a.s.* as $\Delta \to 0$, thus ζ is continuous with probability one and $\omega(\zeta, \Delta) \leq \omega(Y, \omega(F, \Delta))$.

A.2 Proof of Theorem 1

Define a transformed time series $u_i = F(x_i)$, $i = 0, \pm 1, \pm 2, \ldots$ and for any $k = 1, 2, \ldots$, a finite population $\pi_{k,U} = \{u_1, u_2, \ldots, u_{N_k}\}$ together with a simple random sample $U_i = F(X_i)$, $1 \leq i \leq n_k$ from population $\pi_{k,U}$. For any $x \in \mathbb{R}$, let $t = F(x) \in [0, 1]$, then

$$F_{N_k}(x) = F_{U,N_k}(t),\ F_{n_k}(x) = F_{U,n_k}(t), \tag{15}$$

in which

$$F_{U,N_k}(t) = N_k^{-1} \sum_{i=1}^{N_k} I\{u_i \leq t\}, \tag{16}$$

$$F_{U,n_k}(t) = n_k^{-1} \sum_{i=1}^{n_k} I\{U_i \leq t\}. \tag{17}$$

By Assumption (A1), the time series $\{u_t, t = 0, \pm 1, \pm 2, \ldots\}$ is ergodic and has stationary distribution $\mathcal{U}(0, 1)$, hence almost surely $\lim_{k\to\infty} F_{U,N_k}(t) = t$ for $0 \leq t \leq 1$. As $\lim_{k\to\infty} \min(n_k, N_k - n_k) = \infty$ is contained in Assumption (A3), applying Theorem 14.1 of Rosén (1964), one obtains that as random elements taking values in the space $\mathcal{D}[0, 1]$ of cadlag functions:

$$\lambda_k \left\{F_{U,n_k}(t) - F_{U,N_k}(t)\right\} \xrightarrow{d} B(t)$$

almost surely. Lastly, Skorohod's Representation Theorem (Theorem 6.7, Billingsley 1999) provides versions B_k^* of Brownian bridge such that

$$\sup_{t\in[0,1]} \left|\lambda_k \left\{F_{U,n_k}(t) - F_{U,N_k}(t)\right\} - B_k^*(t)\right| \to 0,\ a.s.$$

which implies that

$$\sup_{x\in\mathbb{R}} \left|l_k \left\{F_{n_k}(x) - F_{N_k}(x)\right\} - B_k^*(F(x))\right| \to 0,\ a.s.$$

The Theorem 1 is proved.

A.3 Proof of Theorem 2

Lemma 5 *Under Assumptions (A1) to (A3), (A5), as $k \to \infty$,*

$$\sup_{w\in[-1,1],x\in\mathbb{R}} \left| \left\{F_{n_k}(x - hw) - F_{n_k}(x)\right\} - \left\{F_{N_k}(x - hw) - F_{N_k}(x)\right\} \right| = o_p\left(l_k^{-1}\right).$$

Proof For the Brownian bridges $B_k^*\{\cdot\}$ in Theorem 1,

$$\begin{aligned}
&\sup_{w\in[-1,1],x\in\mathbb{R}} \left|l_k\left\{F_{n_k}(x-hw)-F_{N_k}(x-hw)\right\}-l_k\left\{F_{n_k}(x)-F_{N_k}(x)\right\}\right| \\
&\le \sup_{x,x'\in\mathbb{R},|x-x'|\le h} \left|l_k\left\{F_{n_k}\left(x'\right)-F_{N_k}\left(x'\right)\right\}-l_k\left\{F_{n_k}(x)-F_{N_k}(x)\right\}\right| \\
&\le 2\sup_x\left|l_k\left\{F_{n_k}(x)-F_{N_k}(x)\right\}-B_k^*\{F(x)\}\right| \\
&\qquad+\sup_{x,x'\in\mathbb{R},|x-x'|\le h}\left|B_k^*\left\{F\left(x'\right)\right\}-B_k^*\{F(x)\}\right|. \qquad (18)
\end{aligned}$$

Since $F(\cdot)$ is uniformly continuous by Assumption (A2), and Assumption (A5) implies that $h\to 0$ as $k\to\infty$, so $\omega(F,h)\to 0$ as $k\to\infty$. Assumptions (A1), (A3) ensure Theorem 1, so $l_k\left\{F_{n_k}(\cdot)-F_{N_k}(\cdot)\right\}-B_k^*\{F(\cdot)\}\to 0\; a.s.$ as $k\to\infty$. Thus, the expression in (6) is bounded by

$$\begin{aligned}
&2\sup_{x\in\mathbb{R}}\left|l_k\left\{F_{n_k}(x)-F_{N_k}(x)\right\}-B_k^*\{F(x)\}\right|+\sup_{t,t'\in[0,1],|t-t'|\le\omega(F,h)}\left|B_k^*\left(t'\right)-B_k^*(t)\right| \\
&\quad= o_{a.s.}(1)+o_p(1)=o_p(1).
\end{aligned}$$

In other words,

$$\sup_{w\in[-1,1],x\in\mathbb{R}}\left|\left\{F_{n_k}(x-hw)-F_{n_k}(x)\right\}-\left\{F_{N_k}(x-hw)-F_{N_k}(x)\right\}\right|=o_p\left(l_k^{-1}\right). \qquad (19)$$

□

Lemma 6 *Under Assumptions (A1), (A2), (A5), as $k\to\infty$,*

$$\sup_{w\in[-1,1],x\in\mathbb{R}}\left|\left\{F_{N_k}(x-hw)-F_{N_k}(x)\right\}-\{F(x-hw)-F(x)\}\right|=o_p\left(N_k^{-1/2}\right).$$

Proof Next, since Lemma 4 implies that $N_k^{1/2}\left\{F_{N_k}(\cdot)-F(\cdot)\right\}\xrightarrow{d}\zeta(\cdot)$, Skorohod's Representation Theorem (Theorem 6.7, Billingsley 1999) provides versions $\zeta_k(\cdot)$ of $\zeta(\cdot)$ such that

$$\sup_{x\in\mathbb{R}}\left|N_k^{1/2}\left\{F_{N_k}(x)-F(x)\right\}-\zeta_k(x)\right|\to 0,\; a.s.$$

Consequently

$$\begin{aligned}
&\sup_{w\in[-1,1],x\in\mathbb{R}}\left|N_k^{1/2}\left\{F_{N_k}(x-hw)-F(x-hw)\right\}-N_k^{1/2}\left\{F_{N_k}(x)-F(x)\right\}\right| \\
&\le \sup_{x,x'\in\mathbb{R},|x-x'|\le h}\left|N_k^{1/2}\left\{F_{N_k}\left(x'\right)-F\left(x'\right)\right\}-N_k^{1/2}\left\{F_{N_k}(x)-F(x)\right\}\right| \\
&\le 2\sup_{x\in\mathbb{R}}\left|N_k^{1/2}\left\{F_{N_k}(x)-F(x)\right\}-\zeta_k(x)\right|+\omega(\zeta_k,h)=o_p(1).
\end{aligned}$$

Hence the following holds

$$\sup_{w\in[-1,1],x\in\mathbb{R}}\left|\left\{F_{N_k}(x-hw)-F_{N_k}(x)\right\}-\{F(x-hw)-F(x)\}\right|=o_p\left(N_k^{-1/2}\right). \tag{20}$$

Lemma 7 *Under the Assumptions (A2), (A5) and (A6), as $k\to\infty$,*

$$\sup_{x\in\mathbb{R}}\left|\int_{-1}^{1}\{F(x-hw)-F(x)\}K(w)\,dw\right|=o\left(l_k^{-1}\right).$$

Proof According to the assumptions of the cumulative distribution function, we discuss the problem in two cases.
Case 1: $\nu\geq 1$. Note that by Assumption (A6) $\int_{-1}^{1}K(w)\,w^r dw\equiv 0, r=1,\ldots,l-1$, and by Assumption (A2) $F(\cdot)\in C^{(\nu,\mu)}(\mathbb{R})$. Hence

$$\begin{aligned}
&\int_{-1}^{1}\{F(x-hw)-F(x)\}K(w)\,dw\\
&=\int_{-1}^{1}\left\{F(x-hw)-\sum_{r=0}^{\nu-1}\frac{F^{(r)}(x)}{r!}(-hw)^r\right\}K(w)\,dw\\
&=\int_{-1}^{1}\left\{\int_{x}^{x-hw}\frac{F^{(\nu)}(t)}{(\nu-1)!}(x-hw-t)^{\nu-1}\,dt\right\}K(w)\,dw\\
&=\int_{-1}^{1}\left\{\frac{F^{(\nu)}(x)}{(\nu-1)!}(-hw)^{\nu}+\int_{x}^{x-hw}\frac{F^{(\nu)}(t)-F^{(\nu)}(x)}{(\nu-1)!}(x-hw-t)^{\nu-1}\,dt\right\}K(w)\,dw\\
&=\int_{-1}^{1}\left\{\int_{x}^{x-hw}\frac{F^{(\nu)}(t)-F^{(\nu)}(x)}{(\nu-1)!}(x-hw-t)^{\nu-1}\,dt\right\}K(w)\,dw
\end{aligned}$$

Furthermore, by Assumption (A2) $F^{(\nu)}(\cdot)\in C^{(0,\mu)}(\mathbb{R})$ and

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\left|\int_{-1}^{1}\{F(x-hw)-F(x)\}K(w)\,dw\right|\\
&\quad\leq\sup_{x\in\mathbb{R}}\int_{-1}^{1}\left|\int_{x}^{x-hw}\frac{F^{(\nu)}(t)-F^{(\nu)}(x)}{(\nu-1)!}(x-hw-t)^{\nu-1}\,dt\right|K(w)\,dw\\
&\quad\leq\sup_{x\in\mathbb{R}}\int_{-1}^{1}\left|(hw)^{\nu}\sup_{x\leq t\leq x-hw}\frac{\left|F^{(\nu)}(t)-F^{(\nu)}(x)\right|}{(\nu-1)!}\right|K(w)\,dw\\
&\quad\leq\sup_{x\in\mathbb{R}}\int_{-1}^{1}\left|(hw)^{\nu}\sup_{x\leq t\leq x-hw}\frac{C\,|t-x|^{\mu}}{(\nu-1)!}\right|K(w)\,dw\\
&\quad\leq\sup_{x\in\mathbb{R}}\int_{-1}^{1}\left|(hw)^{\nu}\frac{C(hw)^{\mu}}{(\nu-1)!}\right|K(w)\,dw\\
&\quad\leq\sup_{x\in\mathbb{R}}\int_{-1}^{1}ch^{\nu+\mu}\,|w|^{\nu+\mu}K(w)\,dw=\mathcal{O}\left(h^{\nu+\mu}\right)=o\left(l_k^{-1}\right),
\end{aligned} \tag{21}$$

which follows from Assumption (A5) that $\lim_{k\to\infty} l_k h_{n_k}^{\nu+\mu} = 0$.

Case 2: $\nu = 0$. By Assumption (A2) $F(x) \in C^{(0,\mu)}(\mathbb{R})$. Hence

$$\begin{aligned} &\sup_{x\in\mathbb{R}} \left| \int_{-1}^{1} \{F(x-hw) - F(x)\} K(w)\, dw \right| \\ &\quad \le \sup_{x\in\mathbb{R}} \int_{-1}^{1} C(hw)^{\mu} K(w)\, dw \\ &\quad = \mathcal{O}\left(h^{\nu+\mu}\right) = o\left(l_k^{-1}\right), \end{aligned} \tag{22}$$

which follows from Assumption (A5) that $\lim_{k\to\infty} l_k h_{n_k}^{\nu+\mu} = 0$. □

Proof of Theorem 2 Define $G(x) = \int_{-\infty}^{x} K(u)\, du$. By the definition of $\hat{F}_k(x)$, one obtains

$$\hat{F}_k(x) = n^{-1} \sum_{i=1}^{n_k} \int_{-\infty}^{x} K_h(u - X_i)\, du = n_k^{-1} \sum_{i=1}^{n_k} G\left(\frac{x - X_i}{h}\right).$$

Therefore, by the definition of $F_{n_k}(x) = n_k^{-1} \sum_{i=1}^{n_k} I(X_i \le x)$ in (4)

$$\begin{aligned} \hat{F}_k(x) &= \int_{-\infty}^{+\infty} G\left(\frac{x-u}{h}\right) dF_{n_k}(u) = \int_{-\infty}^{+\infty} h^{-1} K\left(\frac{x-u}{h}\right) F_{n_k}(u)\, du \\ &= \int_{-1}^{1} K(w) F_{n_k}(x - hw)\, dw \end{aligned}$$

using integration by parts and a change of variable $w = (x-u)/h$. The following decomposition plays an important role:

$$\hat{F}_k(x) - F_{n_k}(x) = \int_{-1}^{1} \left\{F_{n_k}(x-hw) - F_{n_k}(x)\right\} K(w)\, dw. \tag{23}$$

Since Assumption (A4) requires that $n_k/N_k = o(1)$ and consequently $N_k^{-1/2} = o\left(l_k^{-1}\right)$, using Lemma 5 and Lemma 6 together with the triangle inequality imply that as $k \to \infty$

$$\left|\left\{F_{n_k}(x-hw) - F_{n_k}(x)\right\} - \{F(x-hw) - F(x)\}\right| = u_p\left(l_k^{-1}\right). \tag{24}$$

By Lemma 7 and applying (23), (24), (21) and (22), the following holds

$$\sup_{x\in\mathbb{R}} \left|\hat{F}_k(x) - F_{n_k}(x)\right| = \sup_{x\in\mathbb{R}} \left| \int_{-1}^{1} \left\{F_{n_k}(x-hw) - F_{n_k}(x)\right\} K(w)\, dw \right| = o_p\left(l_k^{-1}\right).$$

Applying Theorem 1, one has $l_k \left\{\hat{F}_k(x) - F_{N_k}(x)\right\} \xrightarrow{d} B\{F(x)\}$, proving (11).

Notice that under Assumption (A4), $n_k^{-1/2}/l_k^{-1} \to 1$, $N_k^{-1/2} = o\left(l_k^{-1}\right)$, $N_k^{-1/2} = o\left(n_k^{-1/2}\right)$ as $k \to \infty$, and that $N_k^{1/2}\left\{F_{N_k}(\cdot) - F(\cdot)\right\} \stackrel{d}{\to} \zeta(\cdot)$ by Lemma 4. Hence, as $k \to \infty$

$$n_k^{1/2} D\left(F_{N_k}, F\right) = n_k^{1/2} \mathcal{O}_p\left(N_k^{-1/2}\right) = \mathcal{O}_p\left(n_k^{1/2} N_k^{-1/2}\right) = o_p(1).$$

Likewise, $l_k D\left(F_{N_k}, F\right) = o_p(1)$. These, together with (11) establish (12). The proof of Theorem 2 is complete by applying Slutsky's Theorem.

References

Billingsley P (1999) Convergence of probability measures, 2nd edn. Wiley, New York

Brockwell PJ, Davis RA (1991) Time series: theory and methods, 2nd edn. Springer, New York

Cai L, Yang L (2015) A smooth simultaneous confidence band for conditional variance function. TEST 24:632–655

Cao G, Wang L, Li Y, Yang L (2016) Oracle-efficient confidence envelopes for covariance functions in dense functional data. Stat Sin 26:359–383

Cao G, Yang L, Todem D (2012) Simultaneous inference for the mean function based on dense functional data. J Nonparametr Stat 24:359–377

Cardot H, Degras D, Josserand E (2013) Confidence bands for Horvitz–Thompson estimators using sampled noisy functional data. Bernoulli 19:2067–2097

Cardot H, Josserand E (2011) Horvitz–Thompson estimators for functional data: asymptotic confidence bands and optimal allocation for stratified sampling. Biometrika 98:107–118

Cheng M, Peng L (2002) Regression modeling for nonparametric estimation of distribution and quantile functions. Stat Sin 12:1043–1060

Deo CM (1973) A note on empirical process of strong-mixing sequences. Ann Probab 5:870–875

Degras D (2011) Simultaneous confidence bands for nonparametric regression with functional data. Stat Sin 21:1735–1765

Falk M (1985) Asymptotic normality of the kernel quantile estimator. Ann Stat 13:428–433

Francq C, Zakoian J (2010) GARCH models: structure, statistical inference and financial applications. Wiley, Paris

Gu L, Wang L, Härdle W, Yang L (2014) A simultaneous confidence corridor for varying coefficient regression with sparse functional data. TEST 23:806–843

Gu L, Yang L (2015) Oracally efficient estimation for single-index link function with simultaneous band. Electron J Stat 9:1540–1561

Kong J, Gu L, Yang L (2018) Prediction interval for autoregressive time series via oracally efficient estimation of multi-step ahead innovation distribution function. J Time Ser Anal 39:690–708

Liu R, Yang L (2008) Kernel estimation of multivariate cumulative distribution function. J Nonparametr Stat 20:661–677

Ma S, Yang L, Carroll R (2012) A simultaneous confidence band for sparse longitudinal regression. Stat Sin 22:95–122

Reiss R (1981) Nonparametric estimation of smooth distribution functions. Scand J Stat 8:116–119

Rosén B (1964) Limit theorems for sampling from finite populations. Ark Mat 5:383–424

Shao Q, Yang L (2017) Oracally efficient estimation and consistent model selection for auto-regressive moving average time series with trend. J R Stat Soc Ser B Stat Methodol 79:507–524

Song Q, Liu R, Shao Q, Yang L (2014) A simultaneous confidence band for dense longitudinal regression. Commun Stat A-Theory Methods 43:5195–5210

Song Q, Yang L (2009) Spline confidence bands for variance functions. J Nonparametr Stat 5:589–609

Wang J, Cheng F, Yang L (2013) Smooth simultaneous confidence bands for cumulative distribution functions. J Nonparametr Stat 25:395–407

Wang J, Liu R, Cheng F, Yang L (2014) Oracally efficient estimation of autoregressive error distribution with simultaneous confidence band. Ann Stat 42:654–668

Wang J, Wang S, Yang L (2016) Simultaneous confidence bands for the distribution function of a finite population and its superpopulation. TEST 25:692–709
Wang J, Yang L (2009) Polynomial spline confidence bands for regression curves. Stat Sin 19:325–342
Xue L, Wang J (2010) Distribution function estimation by constrained polynomial spline regression. J Nonparametr Stat 22:443–457
Yamato H (1973) Uniform convergence of an estimator of a distribution function. Bull Math Stat 15:69–78
Zheng S, Liu R, Yang L, Härdle W (2016) Statistical inference for generalized additive models: simultaneous confidence corridors and variable selection. TEST 25:607–626
Zheng S, Yang L, Härdle W (2014) A smooth simultaneous confidence corridor for the mean of sparse functional data. J Am Stat Assoc 109:661–673
Zhang Y, Liu R, Shao Q, Yang L (2020) Two-step estimation for time varying arch models. J Time Ser Anal 41:551–570
Zhang Y, Yang L (2018) A smooth simultaneous confidence band for correlation curve. TEST 27:247–269