Computational Statistics and Data Analysis 155 (2021) 107106

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Smooth simultaneous confidence band for the error distribution function in nonparametric regression

Lijie Gu [a], Suojin Wang [b,*], Lijian Yang [c]

[a] Soochow College & School of Mathematical Sciences, Soochow University, Suzhou 215006, China
[b] Department of Statistics, Texas A&M University, College Station, TX 77843, USA
[c] Center for Statistical Science & Department of Industrial Engineering, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 17 December 2019
Received in revised form 14 September 2020
Accepted 22 September 2020
Available online 9 October 2020

Keywords:
Confidence band
Error distribution
Kernel
Local linear
Residuals

ABSTRACT

A smooth simultaneous confidence band (SCB) is constructed for the distribution of unobserved errors in a nonparametric regression model based on a plug-in kernel distribution estimator. The normalized estimation error process is shown to converge to a Gaussian process. Simulation experiments indicate that the proposed SCB not only strikes an intelligent balance between coverage probability and precision, but also achieves surprisingly as much as double efficiency of the classical infeasible SCB. Furthermore, extensive empirical studies are carried out to compare the proposed method with the smooth residual bootstrap method in order to demonstrate the usefulness of each of these methods. As an illustration, the proposed SCB is applied to the Old Faithful geyser data for testing the error distribution.

1. Introduction

Consider the following nonparametric regression model with random design:

$$Y_i = m(X_i) + Z_i, \tag{1}$$

in which (X_i, Y_i), $i = 1, \ldots, n$, are independent identically distributed (i.i.d.) copies of (X, Y), and Z_i, $i = 1, \ldots, n$, are i.i.d. random errors independent of X, satisfying $\mathrm{E}(Z_i) = 0$, $\mathrm{E}\left(Z_i^2\right) = \sigma^2$, with an unknown probability density function (pdf) $f(z)$ and cumulative distribution function (cdf) $F(z) = \int_{-\infty}^{z} f(u)\,du$.

Due to its flexibility, nonparametric regression model (1) has been well explored. This is especially the case regarding the estimation of the regression function using kernel or spline methods. The estimation of the distribution function of the unobserved errors has also attracted attention. Akritas and Van Keilegom (2001) proposed an estimator of the error distribution based on non-parametric regression residuals with weak convergence, leading to prediction intervals and goodness-of-fit tests. Cheng (2002) obtained uniform consistency with the convergence rate of the nonparametric residual empirical distribution function. Müller et al. (2004) showed that a kernel smoothed error distribution function based on residuals is asymptotically equivalent to the true error distribution plus an appropriate correction term, whereas analogous results for the residual empirical distribution function in a semiparametric model were also obtained by Müller et al. (2007). For more applications and tests about the error distribution in a semi- or nonparametric regression model, see Cheng (2005), Dette et al. (2007), and Müller et al. (2012).

* Corresponding author.
E-mail address: sjwang@stat.tamu.edu (S. Wang).

https://doi.org/10.1016/j.csda.2020.107106
0167-9473/

Our goal is to construct a smooth simultaneous confidence band (SCB) for the error distribution $F(z)$ by deriving the asymptotic distribution of the maximal deviation. As well known, an SCB is a powerful tool for making statistical inferences on the global shape of an unknown curve with a quantifiable error probability. Investigators can easily use SCBs to test whether a function has a specific feature. For instance, in Section 5, the proposed SCB is used to test whether the regression error follows a pre-specified distribution for the Old Faithful geyser data with corresponding p-values. Härdle (1989), Xia (1998), and Wang and Yang (2009) investigated the SCB for the unknown nonparametric regression function $m(\cdot)$ using kernel and spline methods. For recent theoretical developments and applications of SCBs in various contexts, see, for instance, Krivobokova et al. (2010), Degras (2011), Cao et al. (2012), Zhu et al. (2012), Wang et al. (2013), Gu et al. (2014), Wang et al. (2014), Zheng et al. (2014), Gu and Yang (2015), Wang et al. (2016), Zheng et al. (2016), Zhang and Yang (2018), Cai et al. (2019) and Gu et al. (2019).

In model (1), if the regression function $m(\cdot)$ were known, one could easily compute the well-known empirical distribution function (EDF) of the errors,

$$F_n(z) = n^{-1}\sum_{j=1}^{n} I\left(Z_j \le z\right), \quad z \in \mathbb{R}, \tag{2}$$

and the accompanying infeasible nonsmooth SCB for $F(z)$ could be constructed as

$$\left[F_n(z) \pm d_{1-\alpha}/\sqrt{n}\right] \cap [0,1], \quad z \in \mathbb{R}, \tag{3}$$

where $d_{1-\alpha} = D^{-1}(1-\alpha), \alpha \in (0,1)$ is the $100(1-\alpha)$th percentile of the Kolmogorov distribution function $D(t) \equiv 1-2\sum_{j=1}^{\infty}(-1)^{j-1}\exp\left(-2j^2t^2\right), t > 0$. Clearly, $F_n(z)$ is a discontinuous step function regardless of $F(z)$ being continuous or discrete. To remedy this deficiency of $F_n(z)$, Xue and Wang (2010) proposed a smooth monotone polynomial spline estimator for the cdf. Yamato (1973) proposed a kernel distribution estimator (KDE) $\tilde{F}(z)$ as follows:

$$\tilde{F}(z) = n^{-1}\sum_{j=1}^{n}\int_{-\infty}^{z} K_h\left(u - Z_j\right)du, \quad z \in \mathbb{R},$$

in which $K(\cdot)$ is a kernel function and $K_h(\cdot) = h^{-1}K(\cdot/h)$ with bandwidth h. Then, according to Wang et al. (2013), the infeasible smooth SCB for $F(z)$ could be constructed as

$$\left[\tilde{F}(z) \pm d_{1-\alpha}/\sqrt{n}\right] \cap [0,1], \quad z \in \mathbb{R}. \tag{4}$$

The SCBs in (3) and (4) are termed "infeasible", as one observes only $\{(X_i, Y_i)\}_{i=1}^n$, not $\{Z_i\}_{i=1}^n$, so that $F_n(z)$, $\tilde{F}(z)$ are both infeasible since the regression function $m(x)$ is unknown. A standard approach is to use a two-step procedure to construct an estimator of the error distribution $F(z)$ given below.

Step 1: Estimate the regression function $m(x)$. We construct the local linear estimator $\hat{m}(x)$ of $m(x)$ by solving the least squares problem

$$\begin{aligned}\left(\hat{m}(x), \hat{m}'(x)\right) &= \left(\hat{\beta}_0, \hat{\beta}_1\right) \\ &= \arg\min_{\beta_0,\beta_1}\sum_{i=1}^{n}\{Y_i - \beta_0 - \beta_1(X_i - x)\}^2 W_\lambda(X_i - x),\end{aligned} \tag{5}$$

where $W_\lambda(\cdot) = \lambda^{-1}W(\cdot/\lambda)$ is a rescaled kernel function with bandwidth λ.

Step 2: Replace the unobservable errors $\left\{Z_j\right\}_{j=1}^n$ by the regression residuals

$$\hat{Z}_j = Y_j - \hat{m}\left(X_j\right), \quad j = 1, \ldots, n, \tag{6}$$

to estimate $F(z)$ by a plug-in KDE $\hat{F}(z)$ or EDF estimator $\hat{F}_n(z)$ as follows:

$$\hat{F}(z) = n^{-1}\sum_{j=1}^{n}\int_{-\infty}^{z} K_h\left(u - \hat{Z}_j\right)du, \quad z \in \mathbb{R}, \tag{7}$$

and

$$\hat{F}_n(z) = n^{-1}\sum_{j=1}^{n} I\left(\hat{Z}_j \le z\right), \quad z \in \mathbb{R}. \tag{8}$$

In order to obtain the asymptotic distribution of the global estimation error to be used for constructing SCBs, we refer to the weak convergence of the residual-based empirical process. As a leading method in the existing literature for constructing SCBs for the error distribution in nonparametric regression, Neumeyer (2009) proved that the empirical process $\sqrt{n}\{\hat{F}_n(z) - F(z)\}$ converges weakly to a centered Gaussian process. Meanwhile, to the best of our knowledge, there do not exist any results on the weak convergence of the residual KDE.

We adopt the main result of Müller et al. (2004) to obtain asymptotically equivalent i.i.d. representation for the process $\sqrt{n}\{\hat{F}(z)-F(z)\}$, from which a comprehensible proof can be derived to show that $\sqrt{n}\{\hat{F}(z)-F(z)\}$ converges in distribution to a Gaussian process; see Lemmas 1 and 2 in Section 2.

For the implementation of the SCBs, Neumeyer (2009) pointed out that it is difficult to use the limit distribution because the asymptotic covariance has to be estimated by plugging in some nonparametric estimators, which would cause the rate of convergence to be rather slow. Hence, she recommended to use a smooth residual bootstrap approach as a useful alternative. Moreover, it can be seen that the plug-in covariance function might not always be positive definite. In order to avoid these difficulties and to substantially save the computing time, we recommend a more straightforward procedure to estimate the covariance. We further show that the covariance estimator converges uniformly to the Gaussian covariance; see Theorem 2. Extensive simulation studies were carried out to compare the proposed method with the smooth residual bootstrap method, as well as to compare the performance of the smooth and nonsmooth SCBs. The numerical results indicate that our method generally has advantages over the bootstrap in medium to large sample sizes. The proposed smooth SCB not only strikes an intelligent balance between coverage probability and precision, but also achieves surprisingly almost as much as double efficiency of the infeasible SCBs defined in (3) and (4).

The rest of the paper is organized as follows. In Section 2, we state the main theoretical results which establish the smooth SCB for the error distribution $F(z)$ based on KDE $\hat{F}(z)$. Section 3 describes the actual steps to implement the SCB. Section 4 contains the simulation results and a real data analysis is illustrated in Section 5. Section 6 provides some concluding remarks. Some technical proofs are given in the Appendix.

2. Smooth SCB based on KDE $\hat{F}(z)$

In this section, we focus on discussing the SCB for $F(z)$ based on KDE $\hat{F}(z)$. For non-negative integer p and $\mu \in (0, 1]$, denote the space of functions whose pth order derivatives are Hölder continuous of order μ by

$$C^{(p,\mu)}(\mathbb{R}) = \left\{\varphi : \mathbb{R} \to \mathbb{R} \left| \|\varphi\|_{p,\mu} = \sup_{t_1 \neq t_2, t_1, t_2 \in \mathbb{R}} \frac{\left|\varphi^{(p)}(t_1) - \varphi^{(p)}(t_2)\right|}{|t_1 - t_2|^{\mu}} < +\infty \right.\right\}. \tag{9}$$

Throughout this paper, $a_n \sim b_n$ means $\lim_{n\to\infty} a_n/b_n = c$, where c is some nonzero constant.

The following technical assumptions are needed for our main results.

- (A1) *The error distribution function* $F(z) \in C^{(1,1)}(\mathbb{R})$ *with density* $f(z)$ *satisfying* $0 < f(z) \le C_f$, $\forall z \in \mathbb{R}$, *for some constant* $C_f > 0$. *The moment* $\mathrm{E}\,|Z|^{\rho} < \infty$ *for some* $\rho > 4$.
- (A2) *The density function* $g(\cdot)$ *of* X *is continuous on* (a, b) *with positive minimum.*
- (A3) *The regression function* $m(\cdot)$ *is twice continuously differentiable on* (a, b).
- (A4) *The kernel function* $W(\cdot)$ *for local linear regression estimation is a three times continuously differentiable symmetric density supported on* $[-1, 1]$, *and the corresponding bandwidth* $\lambda = \lambda_n$ *satisfies* $\lambda \sim n^{-1/4}$.
- (A5) *The kernel function* $K(\cdot)$ *for KDE* $\hat{F}(z)$ *is a symmetric pdf supported on* $[-1, 1]$, *and is twice continuously differentiable. The corresponding bandwidth* $h = h_n$ *satisfies* $h \sim n^{-1/4}/\log n$.
- (A6) *The kernel function* $L(\cdot)$ *for the error density* $f(z)$ *estimation is a symmetric pdf supported on* $[-1, 1]$ *with a bounded second order derivative* L''. *There exists* $\gamma \in (0, 3/16)$ *such that the bandwidth* $H = H_n \sim n^{-\gamma}$.

Under Assumptions (A1)–(A5), Müller et al. (2004) showed that

$$\sup_{z\in\mathbb{R}} \left|\hat{F}(z) - F_n(z) - f(z)\, n^{-1} \sum_{j=1}^{n} Z_j\right| = o_p\left(n^{-1/2}\right), \tag{10}$$

which implies the following Lemma 1.

Lemma 1. *Under Assumptions (A1)–(A5), as* $n \to \infty$,

$$\sup_{z\in\mathbb{R}} \sqrt{n}\left|\hat{F}(z) - F(z)\right| = \sup_{z\in\mathbb{R}} \sqrt{n}\left|F_n(z) + f(z)\, n^{-1} \sum_{j=1}^{n} Z_j - F(z)\right| + o_p(1).$$

Lemma 2. *Given a sequence of i.i.d. random variables* $Z_1, Z_2, \ldots$, *satisfying* $\mathrm{E}(Z_i) = 0$, $\mathrm{E}\left(Z_i^2\right) = \sigma^2$, *with continuous cdf* $F(z)$ *and pdf* $f(z)$, *for* $F_n(z)$ *defined in (2), the process* $\sqrt{n}\{F_n(z) + f(z)n^{-1}\sum_{j=1}^{n} Z_j - F(z)\}$ *converges weakly to a Gaussian process* $G(z)$ *with mean zero and covariance function*

$$\begin{aligned}\Sigma\left(z, z'\right) = {} & F\left(z \wedge z'\right) - F(z)\,F\left(z'\right) + f(z)\,\mathrm{E}\left\{Z_1 I\left(Z_1 \le z'\right)\right\} \\ & + f\left(z'\right)\mathrm{E}\left\{Z_1 I\left(Z_1 \le z\right)\right\} + f(z)\,f\left(z'\right)\sigma^2, \quad z, z' \in \mathbb{R}.\end{aligned} \tag{11}$$

Remark 1. Eq. (10) ensures that the process $\sqrt{n}\{\hat{F}(z) - F(z)\}$ is asymptotically equivalent to the process $\sqrt{n}\{F_n(z) + f(z)\, n^{-1}\sum_{j=1}^{n} Z_j - F(z)\}$. It also converges weakly to the Gaussian process $G(z)$ defined in Lemma 2. This result is similar to Theorem 1 in Neumeyer (2009) for the process $\sqrt{n}\{\hat{F}_n(z) - F(z)\}$, while the proof given in Appendix A.2 is more elementary and more accessible.

The following Theorem follows from Lemmas 1 and 2, Slutsky's Theorem, and the Continuous Mapping Theorem (Theorem 12 in Pollard (1984), p. 70).

Theorem 1. *Under Assumptions (A1)–(A5), as $n \to \infty$, one has*

$$\mathbb{P}\left\{\sup_{z\in\mathbb{R}} \sqrt{n}\left|\hat{F}(z) - F(z)\right| \le t\right\} \to D^*(t) = \mathbb{P}\left\{\sup_{z\in\mathbb{R}} |G(z)| \le t\right\}, t \ge 0,$$

where $D^*(t)$ represents the extreme value distribution of the Gaussian process $G(z)$ defined in Lemma 2.

Corollary 1. *Under Assumptions (A1)–(A5), for $\alpha \in (0, 1)$, an asymptotic $100(1-\alpha)\%$ SCB based on KDE $\hat{F}(z)$ for the error distribution $F(z)$ is*

$$\left[\hat{F}(z) \pm d^*_{1-\alpha}/\sqrt{n}\right] \cap [0, 1], \quad z \in \mathbb{R}, \tag{12}$$

in which $d^*_{1-\alpha} = (D^*)^{-1}(1-\alpha)$ is the $100(1-\alpha)$th percentile of $D^*(\cdot)$ with $(D^*)^{-1}(\cdot)$ being the inverse function of $D^*(\cdot)$.

Since there are some unknown functions $F(z)$, $f(z)$, $\mathrm{E}\left\{Z_1 I\left(Z_1 \le z'\right)\right\}$ in the covariance function $\Sigma\left(z, z'\right)$ of the Gaussian process $G(z)$, the critical value $d^*_{1-\alpha}$ in the SCB in (12) is infeasible and needs to be estimated. It is hard to estimate $\Sigma\left(z, z'\right)$ by plugging in some nonparametric estimators, because the convergence rate would be slow and the positive definiteness of covariance matrix could not be guaranteed. Neumeyer (2009) recommended to use a resampling procedure. Here we propose a simple but effective procedure to estimate the covariance function.

Note that

$$\begin{aligned}
&\sqrt{n}\left\{F_n(z) + f(z)\, n^{-1}\sum_{j=1}^{n} Z_j - F(z)\right\} \\
&= n^{-1/2}\sum_{j=1}^{n}\left\{I\left(Z_j \le z\right) + f(z) Z_j - F(z)\right\} \equiv n^{-1/2}\sum_{j=1}^{n}\eta_{nj}(z),
\end{aligned}$$

which is written as the sum of i.i.d. random variables. It is easy to see that $\mathrm{E}\eta_{n1}(z) = 0$ and

$$\begin{aligned}
\Sigma\left(z, z'\right) &= \mathrm{cov}\left(\eta_{n1}(z), \eta_{n1}\left(z'\right)\right) \\
&= \mathrm{E}\left[\{I(Z_1 \le z) + f(z) Z_1 - F(z)\}\left\{I\left(Z_1 \le z'\right) + f\left(z'\right) Z_1 - F\left(z'\right)\right\}\right] \\
&= F\left(z \wedge z'\right) - F(z) F\left(z'\right) + f(z)\,\mathrm{E}\left\{Z_1 I\left(Z_1 \le z'\right)\right\} + f\left(z'\right)\mathrm{E}\{Z_1 I(Z_1 \le z)\} + f(z) f\left(z'\right)\sigma^2.
\end{aligned}$$

In order to estimate $\Sigma\left(z, z'\right)$, one can first define a plug-in kernel density estimator of $f(z)$ based on residuals $\left\{\hat{Z}_j\right\}_{j=1}^{n}$ in (6) by

$$\hat{f}(z) = n^{-1}\sum_{j=1}^{n} L_H\left(\hat{Z}_j - z\right), \tag{13}$$

where $L_H(\cdot) = H^{-1}L(\cdot/H)$ is a rescaled kernel function with bandwidth H. Denote

$$\hat{\eta}_{nj}(z) = I\left(\hat{Z}_j \le z\right) + \hat{f}(z)\hat{Z}_j - \hat{F}_n(z), j = 1, \ldots, n, \tag{14}$$

in which $\hat{F}_n(z)$ is given in (8). Then we define the covariance estimator

$$\begin{aligned}
\hat{\Sigma}_n\left(z, z'\right) &= n^{-1}\sum_{j=1}^{n}\hat{\eta}_{nj}(z)\,\hat{\eta}_{nj}\left(z'\right) \\
&= \hat{F}_n\left(z \wedge z'\right) - \hat{F}_n(z)\hat{F}_n\left(z'\right) + \hat{f}(z)\, n^{-1}\sum_{j=1}^{n}\hat{Z}_j I\left(\hat{Z}_j \le z'\right) \\
&\quad + \hat{f}\left(z'\right) n^{-1}\sum_{j=1}^{n}\hat{Z}_j I\left(\hat{Z}_j \le z\right) + \hat{f}(z)\hat{f}\left(z'\right) n^{-1}\sum_{j=1}^{n}\hat{Z}_j^2 \\
&\quad - \hat{F}_n(z)\hat{f}\left(z'\right) n^{-1}\sum_{j=1}^{n}\hat{Z}_j - \hat{F}_n\left(z'\right)\hat{f}(z)\, n^{-1}\sum_{j=1}^{n}\hat{Z}_j.
\end{aligned} \tag{15}$$

The following Theorem 2 states that the covariance estimator $\hat{\Sigma}\left(z, z'\right)$ is uniformly convergent to the Gaussian covariance function $\Sigma\left(z, z'\right)$.

Theorem 2. *Under Assumptions (A1)–(A6), for* $\Sigma\left(z, z'\right)$ *and* $\hat{\Sigma}_n\left(z, z'\right)$ *given in (11) and (15), as* $n \to \infty$, *one has*

$$\sup_{z,z' \in \mathbb{R}} \left| \hat{\Sigma}_n\left(z, z'\right) - \Sigma\left(z, z'\right) \right| = o_p(1).$$

The proof of Theorem 2 is given in Appendix A.3. It makes use of the asymptotic properties of EDF $\hat{F}_n(z)$ shown in Müller et al. (2007) and kernel density estimator $\hat{f}(z)$. Furthermore, denote by $\hat{d}^*_{1-\alpha}$ the $100(1-\alpha)$th percentile of the extreme value distribution of Gaussian process $\hat{G}(z)$ with mean zero and covariance function $\hat{\Sigma}_n\left(z, z'\right)$. For a Gaussian process, Theorem 2 and the Continuous Mapping Theorem imply that the $d^*_{1-\alpha}$ in (12) can be replaced by $\hat{d}^*_{1-\alpha}$ with a negligible error. Hence, one obtains the following main result based on KDE $\hat{F}(z)$:

Theorem 3. *Under Assumptions (A1)–(A6), as* $n \to \infty$, *a smooth feasible asymptotic* $100(1-\alpha)\%$ *SCB based on KDE* $\hat{F}(z)$ *for the error distribution* $F(z)$ *is given by*

$$\left[\hat{F}(z) \pm \hat{d}^*_{1-\alpha}/\sqrt{n}\right] \cap [0, 1], \quad z \in \mathbb{R}. \tag{16}$$

Remark 2. Neumeyer (2009) showed that the process $\sqrt{n}\left\{\hat{F}_n(z) - F(z)\right\}$ also converges weakly to the Gaussian process $G(z)$ defined in Lemma 2 , one can also use a procedure similar to that stated in this section to estimate the Gaussian covariance $\Sigma\left(z, z'\right)$ by $\hat{\Sigma}_n\left(z, z'\right)$ and get the estimated quantile $\hat{d}^*_{1-\alpha}$ of maximum values for the Gaussian process with covariance $\hat{\Sigma}_n\left(z, z'\right)$. Hence, a nonsmooth feasible asymptotic $100(1-\alpha)\%$ SCB based on EDF $\hat{F}_n(z)$ for the error distribution $F(z)$ can be constructed as

$$\left[\hat{F}_n(z) \pm \hat{d}^*_{1-\alpha}/\sqrt{n}\right] \cap [0, 1], \quad z \in \mathbb{R}. \tag{17}$$

3. Implementation

In this section, we outline the procedures to construct the smooth SCB given in (16), to obtain the nonsmooth SCB in (17), and to implement the smooth residual bootstrap method. Firstly, the kernel functions and the corresponding bandwidths are chosen as follows:

(i) For the smooth KDE $\hat{F}(z)$ defined in (7), the triweight kernel $W(u) = 35(1-u^2)^3/32$, $|u| \le 1$, is used for the regression estimator with a data-driven bandwidth $\lambda = 2\hat{\sigma}_X n^{-1/4}$, which satisfies Assumption (A4), with $\hat{\sigma}_X =$ the sample standard deviation of X. Meanwhile, according to Assumption (A5), the KDE $\hat{F}$ is also computed with the triweight kernel $K(u)$ and bandwidth $h = \text{IQR}n^{-1/4}/\log n$, where IQR denotes the sample interquartile range of the estimated residuals $\{\hat{Z}_j\}_{j=1}^n$.

(ii) For the nonsmooth EDF $\hat{F}_n(z)$ defined in (8), we still use the triweight kernel $W(u)$ to obtain the regression estimator $\hat{m}(x)$ with a data-driven bandwidth $\lambda = 2\hat{\sigma}_X(n\log n)^{-1/4}$.

(iii) For the density estimator $\hat{f}(z)$ defined in (13), the biweight kernel $L(u) = 15(1-u^2)^2/16$, $|u| \le 1$ with bandwidth $H = \text{IQR}n^{-1/6}$ is used to satisfy Assumption (A6).

Secondly, obtaining the critical values $\hat{d}^*_{1-\alpha}$ in (16) is the key step of the procedure to construct the SCBs. After estimating the regression function $m(x)$ by the local linear kernel method, the estimated regression residuals $\{\hat{Z}_j\}_{j=1}^n$ are used to compute the EDF $\hat{F}_n(z)$, KDE $\hat{F}(z)$, and $\hat{f}(z)$ at k equal distance grid points $\min_{j=1}^n\{\hat{Z}_j\} = z_1 < z_2 < \cdots < z_k = \max_{j=1}^n\{\hat{Z}_j\}$. According to the definition of $\hat{\eta}_{nj}(z), j = 1, \ldots, n$ given in (14), it is easy to construct a sequence of random vectors $\left\{\hat{\eta}_{nj}(z_1), \ldots, \hat{\eta}_{nj}(z_k)\right\}, j = 1, \ldots, n$, to compute the sample covariance matrix $\{\hat{\Sigma}_n(z_l, z_{l'})\}_{l,l'=1}^k$, and then to generate k-dimensional normal random variables with mean zero and covariance matrix $\{\hat{\Sigma}_n(z_l, z_{l'})\}_{l,l'=1}^k$ repeatedly for Λ times, where Λ is a preset large integer. Lastly, for each replication one takes the maximal absolute value of the k normal random variables, sorts these Λ maximum values according to ascending order, and obtains the critical value $\hat{d}^*_{1-\alpha}$ by the $\Lambda(1-\alpha)$th value as the empirical quantile of these sorted Λ maximum values for any $\alpha \in (0, 1)$. The default values are set to be $k = 401$, $\Lambda = 2000$.

For comparisons, we also adopt the smooth residual bootstrap procedure recommended by Neumeyer (2009) to obtain the critical values of SCBs. Let $\left\{\left(\hat{Z}^*_{1,b}, \ldots, \hat{Z}^*_{n,b}\right), b = 1, \ldots, B\right\}$ be B bootstrap samples of the residuals from the bootstrap observations. Define

$$\hat{d}^B_{1-\alpha} \equiv (1-\alpha)\text{ th quantile of } \left\{\sup_{z\in\mathbb{R}} \sqrt{n}\left|\hat{F}^*_{n,b}(z) - \tilde{F}^B_n(z)\right|, b = 1, \ldots, B\right\},$$

Table 1
Coverage frequencies, width, and width ratio of SCBs for error distribution $N(0, 1/2)$ based on the EDF $\hat{F}_n$ (left) and KDE $\hat{F}$ (right) with infeasible error EDF F_n (middle) over 2000 replications.

n	Method	SCB	$1-\alpha=0.90$	$1-\alpha=0.95$	$1-\alpha=0.99$
125	Proposed	Coverage	0.867 (0.915) 0.912	0.935 (0.960) 0.951	0.985 (0.992) 0.989
		Width	0.130 (0.180) 0.131	0.140 (0.199) 0.142	0.161 (0.233) 0.163
		Width ratio	0.718 (1.000) 0.728	0.706 (1.000) 0.716	0.691 (1.000) 0.700
	Bootstrap	Coverage	0.918 (0.915) 0.941	0.959 (0.960) 0.975	0.994 (0.992) 0.994
		Width	0.135 (0.180) 0.138	0.146 (0.199) 0.148	0.167 (0.233) 0.170
		Width ratio	0.750 (1.000) 0.767	0.735 (1.000) 0.746	0.717 (1.000) 0.730
250	Proposed	Coverage	0.879 (0.919) 0.895	0.934 (0.958) 0.947	0.991 (0.992) 0.988
		Width	0.092 (0.128) 0.093	0.099 (0.141) 0.100	0.114 (0.166) 0.115
		Width ratio	0.717 (1.000) 0.725	0.703 (1.000) 0.711	0.686 (1.000) 0.694
	Bootstrap	Coverage	0.914 (0.919) 0.930	0.965 (0.958) 0.970	0.994 (0.992) 0.995
		Width	0.095 (0.128) 0.096	0.102 (0.141) 0.103	0.117 (0.166) 0.119
		Width ratio	0.737 (1.000) 0.745	0.723 (1.000) 0.730	0.705 (1.000) 0.717
500	Proposed	Coverage	0.887 (0.903) 0.910	0.941 (0.953) 0.953	0.987 (0.989) 0.987
		Width	0.065 (0.091) 0.065	0.070 (0.101) 0.071	0.081 (0.119) 0.081
		Width ratio	0.714 (1.000) 0.720	0.700 (1.000) 0.705	0.682 (1.0000) 0.687
	Bootstrap	Coverage	0.917 (0.903) 0.923	0.954 (0.953) 0.960	0.990 (0.989) 0.992
		Width	0.066 (0.091) 0.067	0.072 (0.101) 0.072	0.082 (0.119) 0.083
		Width ratio	0.728 (1.000) 0.734	0.713 (1.000) 0.713	0.689 (1.000) 0.697
1000	Proposed	Coverage	0.892 (0.912) 0.908	0.945 (0.961) 0.953	0.989 (0.994) 0.991
		Width	0.046 (0.064) 0.046	0.050 (0.071) 0.050	0.057 (0.084) 0.057
		Width ratio	0.711 (1.000) 0.715	0.696 (1.000) 0.700	0.678 (1.000) 0.681
	Bootstrap	Coverage	0.908 (0.912) 0.910	0.953 (0.961) 0.954	0.994 (0.994) 0.992
		Width	0.046 (0.064) 0.047	0.050 (0.071) 0.051	0.058 (0.084) 0.058
		Width ratio	0.722 (1.000) 0.725	0.706 (1.000) 0.710	0.687 (1.000) 0.691

where $\hat{F}^*_{n,b}(z) = n^{-1}\sum_{j=1}^n I(\hat{Z}^*_{j,b} \le z)$, and $\tilde{F}^B_n(z) = \int_{-\infty}^z (na_n)^{-1}\sum_{j=1}^n \mathcal{K}\{(u-\tilde{Z}_j)/a_n\}du$ with centered residuals $\tilde{Z}_j = \hat{Z}_j - n^{-1}\sum_{i=1}^n \hat{Z}_i$. We present the results for the choice of $\mathcal{K}(\cdot)$ to be the Gaussian kernel with bandwidth $a_n = n^{-1/4}$, and $\hat{d}^B_{1-\alpha}$ to be realized as the empirical quantile of $B = 2000$ maximal absolute values on $k = 401$ equal distance grid points described above. Then the smooth and nonsmooth SCBs by the smooth residual bootstrap method can be constructed by replacing $\hat{d}^*_{1-\alpha}$ in (16) and (17) with $\hat{d}^B_{1-\alpha}$. The corresponding numerical results are listed in the "right" and "left" versions of the bootstrap in Tables 1–4.

One can use an SCB for testing hypotheses in the same way as a confidence interval is used. Under the null hypothesis $H_0 : F(z) = F_0(z)$, where $F_0(z)$ is a pre-specified distribution function satisfying Assumption (A1), the $100(1-\alpha)\%$ SCBs given in (16) and (17) completely cover the distribution curve $F_0(z)$ with probability approaching $1-\alpha$. Thus one rejects the null hypothesis with significance level α, if the curve $F_0(z)$ is not contained by the SCBs for just one $z \in \mathbb{R}$. It should be cautioned that the same logic does not apply to the testing of composite null hypothesis $H_0 : F(z) = F_{\theta_0}(z)$ for some unknown $\theta_0 \in \Theta$, where $\{F_\theta, \theta \in \Theta\}$ is a family of distributions parameterized by θ. This appears to be a major limitation of the SCB approach to hypothesis testing.

Asymptotic p-values can be numerically approximated for the null hypothesis $H_0 : F(z) = F_0(z)$ based on the proposed smooth SCB based on KDE $\hat{F}$:

$$p = \max\left\{\alpha : \hat{F}(z_l) - \frac{\hat{d}^*_{1-\alpha}}{\sqrt{n}} \le F_0(z_l) \le \hat{F}(z_l) + \frac{\hat{d}^*_{1-\alpha}}{\sqrt{n}}, l = 1, \ldots, k\right\}$$
$$= 1 - \min\left\{1-\alpha : \max_{l=1}^{k}\sqrt{n}\left|\hat{F}(z_l) - F_0(z_l)\right| \le \hat{d}^*_{1-\alpha}\right\}, \tag{18}$$

where $z_l, l = 1, \ldots, k$, are equally spaced grid points and the critical values $\hat{d}^*_{1-\alpha}$ are described above. In other words, $1-p$ is the index of the minimum critical value $\hat{d}^*_{1-p}$ for which the k inequalities in (18) are simultaneously satisfied. Likewise, approximate asymptotic p-values for the nonsmooth SCB can be obtained by replacing $\hat{F}$ with $\hat{F}_n$ in (18). The corresponding approximate p-values for the bootstrap method can be obtained by $\hat{d}^B_{1-\alpha}$ instead of $\hat{d}^*_{1-\alpha}$ in (18).

4. Simulation studies

In the following, we present some results of our extensive simulation studies to illustrate the finite-sample performance of the proposed SCB in comparison to other methods, including the smooth residual bootstrap method. The data were generated from the model: $Y_i = \sin\{2\pi(X_i - 0.5)\} + Z_i, i = 1, \ldots, n$, where $X_i \sim U(0, 1)$, $Z_i \sim F(z)$ and four candidate

Table 2
Coverage frequencies, width, and width ratio of SCBs for error distribution t_8 based on the EDF $\hat{F}_n$ (left) and KDE $\hat{F}$ (right) with infeasible error EDF F_n (middle) over 2000 replications.

n	Method	SCB	$1-\alpha=0.90$	$1-\alpha=0.95$	$1-\alpha$=0.99
125	Proposed	Coverage	0.868 (0.917) 0.925	0.929 (0.962) 0.962	0.986 (0.993) 0.995
		Width	0.127 (0.174) 0.128	0.137 (0.191) 0.138	0.158 (0.224) 0.159
		Width ratio	0.731 (1.000) 0.735	0.719 (1.000) 0.723	0.705 (1.000) 0.710
	Bootstrap	Coverage	0.918 (0.917) 0.955	0.965 (0.962) 0.981	0.996 (0.993) 0.997
		Width	0.133 (0.174) 0.134	0.144 (0.191) 0.145	0.165 (0.224) 0.166
		Width ratio	0.764 (1.000) 0.770	0.754 (1.000) 0.759	0.738 (1.000) 0.742
250	Proposed	Coverage	0.878 (0.920) 0.908	0.933 (0.964) 0.955	0.987 (0.995) 0.988
		Width	0.089 (0.122) 0.090	0.096 (0.134) 0.097	0.111 (0.158) 0.111
		Width ratio	0.732 (1.000) 0.736	0.719 (1.000) 0.723	0.704 (1.000) 0.707
	Bootstrap	Coverage	0.908 (0.920) 0.934	0.957 (0.964) 0.966	0.991 (0.995) 0.997
		Width	0.092 (0.122) 0.093	0.100 (0.134) 0.101	0.115 (0.158) 0.116
		Width ratio	0.754 (1.000) 0.762	0.746 (1.000) 0.754	0.728 (1.000) 0.734
500	Proposed	Coverage	0.889 (0.924) 0.905	0.938 (0.967) 0.950	0.984 (0.991) 0.989
		Width	0.062 (0.085) 0.062	0.067 (0.094) 0.068	0.078 (0.111) 0.078
		Width ratio	0.732 (1.000) 0.734	0.719 (1.000) 0.721	0.703 (1.000) 0.704
	Bootstrap	Coverage	0.911 (0.924) 0.928	0.952 (0.967) 0.974	0.992 (0.991) 0.993
		Width	0.064 (0.085) 0.064	0.069 (0.094) 0.070	0.080 (0.111) 0.080
		Width ratio	0.752 (1.000) 0.754	0.734 (1.000) 0.745	0.723 (1.000) 0.725
1000	Proposed	Coverage	0.885 (0.907) 0.910	0.945 (0.955) 0.953	0.987 (0.990) 0.992
		Width	0.044 (0.060) 0.044	0.047 (0.066) 0.047	0.054 (0.078) 0.055
		Width ratio	0.729 (1.000) 0.730	0.716 (1.000) 0.717	0.699 (1.000) 0.700
	Bootstrap	Coverage	0.904 (0.907) 0.921	0.956 (0.955) 0.961	0.993 (0.990) 0.994
		Width	0.044 (0.060) 0.045	0.048 (0.066) 0.048	0.056 (0.078) 0.056
		Width ratio	0.733 (1.000) 0.747	0.733 (1.000) 0.734	0.717 (1.000) 0.718

error distributions $F(z)$ are normal distribution $N(0, 1/2)$, the t-distribution with 8 degrees of freedom, t_8, standardized Beta distribution $B(2, 5)$, and standardized χ^2-distribution with 10 degrees of freedom, $\chi^2(10)$. This regression function is similar to the simulation setting in Wang and Yang (2009). The number of subjects n was taken to be 125, 250, 500, 1000.

Similar to a confidence interval, an 'ideal' SCB should reach two general principles: one is to be accurate, i.e., the probability of the unknown function being contained in the SCB should be close to the prescribed nominal level $1-\alpha$. The other is to be informative, i.e., the SCB being sufficiently narrow and therefore useful in locating the unknown function. According to the above two general principles, we compared the performance of the SCBs of the proposed method and the bootstrap method constructed by using KDE $\hat{F}$ and EDF $\hat{F}_n$ at the confidence levels $1-\alpha = 0.90, 0.95, 0.99$, respectively. The infeasible SCB in (3) is used as a benchmark for comparison for both methods since F_n would be a natural estimator if it were available.

Tables 1–4 report the coverage frequencies that the true cdf curve is covered by the SCBs at the 401 equally spaced grid points from $\min_{j=1}^{n}\{\hat{Z}_j\}$ to $\max_{j=1}^{n}\{\hat{Z}_j\}$ over 2000 replications. The averages over 2000 replications of the average width of each SCB and the ratio of each average width to the average width of the infeasible SCB (called width ratio) are also listed.

For the proposed method, firstly, one observes that the coverage frequencies of the smooth SCB based on KDE $\hat{F}$ are closer to the nominal levels than those of the nonsmooth SCB based on the EDF $\hat{F}_n$, although there is essentially no difference of the average widths between the smooth and nonsmooth SCBs. This result can be understood intuitively since using the smooth version of the SCB makes more sense for covering the continuous distribution function while the bounds of the nonsmooth version are a pair of step functions and due to their discreteness they would tend to have less probability to fully cover the continuous distribution function. This phenomenon has been observed in other related studies; see, for example, Wang et al. (2016). Meanwhile, infeasible nonsmooth SCB using F_n has the coverage frequencies generally higher than $1-\alpha$ and is much wider than the smooth and nonsmooth SCBs, which confirms the well known conservative nature of the Kolmogorov statistic based on the standard EDF.

Secondly, the ratios of widths give us more intuitive comparisons. The widths of nonsmooth and smooth SCBs based on $\hat{F}_n$ and $\hat{F}$ using regression residuals are about 70% of that of the infeasible SCB using the true error EDF F_n. Note that surprisingly the efficiency of constructing SCBs using the estimated residuals can be increased as much as $\left(1/0.7^2-1\right) \times 100\% \approx 104\%$ compared to the infeasible SCB using the true errors. It is equivalent to say that the SCBs based on $\hat{F}_n$ and $\hat{F}$ constructed by using only half of a sample are as efficient as the SCB based on F_n using the whole sample. This is seen in the results in Tables 1–4 for both moderate and large sample sizes. Another way to illustrate this phenomenon is that the mean sample variance of the Gaussian process $G(z)$ defined in Lemma 2 is about 0.049 while that of the classic (infeasible) empirical process is about 0.099 in the scenario of $N(0, 1/2)$ with $n = 250$. There

Table 3
Coverage frequencies, width, and width ratio of SCBs for error distribution standardized *Beta* (2, 5) based on the EDF $\hat{F}_n$ (left) and KDE $\hat{F}$ (right) with infeasible error EDF F_n (middle) over 2000 replications.

n	Method	SCB	$1-\alpha=0.90$	$1-\alpha=0.95$	$1-\alpha=0.99$
125	Proposed	Coverage	0.842 (0.903) 0.915	0.913 (0.951) 0.954	0.976 (0.992) 0.992
		Width	0.135 (0.187) 0.136	0.146 (0.206) 0.147	0.168 (0.242) 0.170
		Width ratio	0.721 (1.000) 0.728	0.709 (1.000) 0.716	0.695 (1.000) 0.702
	Bootstrap	Coverage	0.904 (0.903) 0.952	0.954 (0.951) 0.978	0.992 (0.992) 0.996
		Width	0.143 (0.187) 0.144	0.155 (0.206) 0.156	0.178 (0.242) 0.180
		Width ratio	0.764 (1.000) 0.772	0.752 (1.000) 0.759	0.736 (1.000) 0.743
250	Proposed	Coverage	0.846 (0.916) 0.900	0.917 (0.962) 0.943	0.979 (0.990) 0.989
		Width	0.097 (0.134) 0.097	0.105 (0.148) 0.106	0.121 (0.174) 0.122
		Width ratio	0.723 (1.000) 0.729	0.719 (1.000) 0.716	0.694 (1.000) 0.700
	Bootstrap	Coverage	0.907 (0.916) 0.944	0.958 (0.962) 0.978	0.993 (0.990) 0.996
		Width	0.101 (0.134) 0.102	0.109 (0.148) 0.110	0.127 (0.174) 0.128
		Width ratio	0.754 (1.000) 0.761	0.740 (1.000) 0.746	0.727 (1.000) 0.733
500	Proposed	Coverage	0.869 (0.907) 0.907	0.929 (0.958) 0.946	0.984 (0.991) 0.990
		Width	0.069 (0.095) 0.069	0.075 (0.105) 0.075	0.086 (0.125) 0.087
		Width ratio	0.723 (1.000) 0.728	0.709 (1.000) 0.714	0.693 (1.000) 0.697
	Bootstrap	Coverage	0.910 (0.907) 0.944	0.958 (0.958) 0.976	0.991 (0.991) 0.995
		Width	0.071 (0.095) 0.072	0.078 (0.105) 0.078	0.090 (0.125) 0.090
		Width ratio	0.749 (1.000) 0.754	0.740 (1.000) 0.744	0.720 (1.000) 0.723
1000	Proposed	Coverage	0.864 (0.906) 0.900	0.930 (0.953) 0.951	0.980 (0.992) 0.986
		Width	0.049 (0.068) 0.049	0.053 (0.075) 0.054	0.062 (0.089) 0.062
		Width ratio	0.723 (1.000) 0.726	0.708 (1.000) 0.712	0.691 (1.000) 0.694
	Bootstrap	Coverage	0.898 (0.906) 0.925	0.950 (0.953) 0.960	0.990 (0.992) 0.996
		Width	0.051 (0.068) 0.051	0.055 (0.075) 0.055	0.064 (0.083) 0.064
		Width ratio	0.744 (1.000) 0.748	0.730 (1.000) 0.734	0.715 (1.000) 0.718

Table 4
Coverage frequencies, width, and width ratio of SCBs for error distribution standardized $\chi^2(10)$ based on the EDF $\hat{F}_n$ (left) and KDE $\hat{F}$ (right) with infeasible error EDF F_n (middle) over 2000 replications.

n	Method	SCB	$1-\alpha=0.90$	$1-\alpha=0.95$	$1-\alpha=0.99$
125	Proposed	Coverage	0.847 (0.928) 0.907	0.912 (0.967) 0.949	0.975 (0.992) 0.989
		Width	0.133 (0.179) 0.134	0.144 (0.197) 0.145	0.166 (0.232) 0.167
		Width ratio	0.739 (1.000) 0.746	0.728 (1.000) 0.735	0.717 (1.000) 0.723
	Bootstrap	Coverage	0.893 (0.928) 0.937	0.937 (0.967) 0.966	0.986 (0.992) 0.993
		Width	0.140 (0.179) 0.141	0.152 (0.197) 0.153	0.175 (0.232) 0.177
		Width ratio	0.782 (1.000) 0.789	0.770 (1.000) 0.777	0.756 (1.000) 0.762
250	Proposed	Coverage	0.869 (0.925) 0.907	0.925 (0.963) 0.949	0.981 (0.993) 0.986
		Width	0.094 (0.127) 0.095	0.102 (0.140) 0.103	0.118 (0.165) 0.119
		Width ratio	0.742 (1.000) 0.747	0.730 (1.000) 0.735	0.717 (1.000) 0.722
	Bootstrap	Coverage	0.882 (0.925) 0.931	0.942 (0.963) 0.967	0.987 (0.993) 0.992
		Width	0.099 (0.127) 0.099	0.107 (0.140) 0.108	0.124 (0.165) 0.125
		Width ratio	0.775 (1.000) 0.781	0.764 (1.000) 0.769	0.751 (1.000) 0.756
500	Proposed	Coverage	0.858 (0.916) 0.899	0.930 (0.956) 0.945	0.979 (0.993) 0.988
		Width	0.067 (0.090) 0.067	0.073 (0.099) 0.073	0.084 (0.117) 0.084
		Width ratio	0.744 (1.000) 0.747	0.731 (1.000) 0.735	0.718 (1.000) 0.721
	Bootstrap	Coverage	0.899 (0.916) 0.922	0.947 (0.956) 0.958	0.989 (0.993) 0.992
		Width	0.069 (0.090) 0.070	0.075 (0.099) 0.076	0.087 (0.117) 0.088
		Width ratio	0.771 (1.000) 0.775	0.759 (1.000) 0.763	0.747 (1.000) 0.749
1000	Proposed	Coverage	0.867 (0.926) 0.894	0.929 (0.965) 0.942	0.985 (0.995) 0.989
		Width	0.047 (0.064) 0.047	0.051 (0.070) 0.052	0.060 (0.083) 0.060
		Width ratio	0.743 (1.000) 0.746	0.731 (1.000) 0.733	0.717 (1.000) 0.720
	Bootstrap	Coverage	0.903 (0.926) 0.921	0.948 (0.965) 0.959	0.988 (0.995) 0.991
		Width	0.049 (0.064) 0.049	0.053 (0.070) 0.053	0.062 (0.083) 0.062
		Width ratio	0.766 (1.000) 0.769	0.755 (1.000) 0.757	0.743 (1.000) 0.746

are at least two reasons why using the estimated residuals to replace the true errors can achieve more efficiency: (1) the infeasible SCB constructed by using the true errors is well known to be conservative (see Wang et al. (2013)) and (2) more importantly, estimating unknown nuisance parameters or functions can lead to some internal negative correlation in the

Table 5
Comparisons of the computing time (minutes) of SCBs' coverage frequencies given the confidence levels $1-\alpha = 0.90, 0.95, 0.99$ based on our proposed method and the smooth bootstrap method for four candidate error distributions F over 2000 replications, respectively.

	Method	$n = 125$	$n = 250$	$n = 500$	$n = 1000$
$N(0, 1/2)$	Proposed	15.2	15.9	17.0	19.5
	Bootstrap	58.2	73.9	78.6	90.0
t_8	Proposed	16.2	17.0	17.0	19.6
	Bootstrap	60.9	86.3	98.8	110.7
Standardized $Beta(2, 5)$	Proposed	15.3	15.8	17.3	19.7
	Bootstrap	59.0	70.0	79.3	89.9
Standardized $\chi^2(10)$	Proposed	15.1	15.8	17.1	20.0
	Bootstrap	58.2	69.2	79.1	89.7

estimating equation so that the overall variance of the supernorm of the estimation difference is drastically reduced. This phenomenon has also been observed in other contexts; see, for example, Pierce (1982) and Wang et al. (1997).

Lastly, Tables 1–4 show that the bootstrap smooth (nonsmooth) SCBs have overall somewhat higher coverage frequencies and wider average widths than the proposed smooth (nonsmooth) SCBs. The coverage frequencies of the proposed smooth SCBs and the bootstrap nonsmooth SCBs are both close to the nominal level, while the width ratios of the bootstrap nonsmooth SCBs are almost always greater than that of the proposed smooth SCBs, which suggests that the bootstrap nonsmooth SCBs are somewhat less efficient than the proposed smooth SCBs.

As seen in Table 5 the computing time for the bootstrap is much longer than that for the proposed method, as expected. Therefore, it appears that when the sample size is small, the bootstrap method might be preferred since a small sample size generally does not justify our proposed asymptotic method. However, when the sample size is moderate to large, the proposed method has clear advantages over the bootstrap. One thing in common is that both methods are much more efficient than using the infeasible SCBs with true errors.

To visualize the performance of the distribution function estimates and the corresponding SCBs, Fig. 1 depicts curves of the true error distributions F (four candidate distributions), EDF $\hat{F}_n$, KDE $\hat{F}$ and error EDF F_n together with the corresponding asymptotic 95% SCBs at $n = 250$, respectively, for one simulated data set. One can see the three curves of $\hat{F}_n$, $\hat{F}$, and F_n are all close to F, while the infeasible SCB is wider than the nonsmooth and smooth SCBs. Other settings yielded similar results.

Following a reviewer's comment, we have also considered the t_6 distribution which still satisfies condition (A1) but is heavier tailed than t_8. The simulation results are quite similar to those for t_8 in Table 2. The details are omitted here.

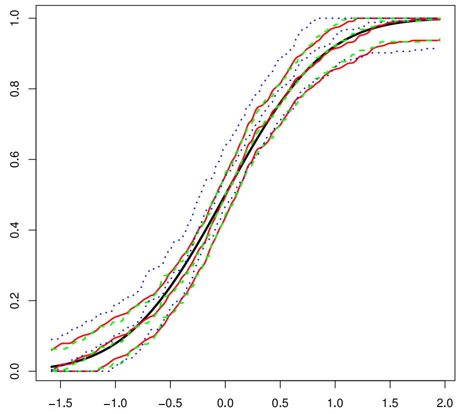
In conclusion, when the sample size is small the bootstrap method might be preferred, but when the sample size is moderate to large we recommend the smooth SCB based on KDE $\hat{F}$, which strikes an intelligent balance between coverage probability and precision the best, not to mention its clear computational advantage. The nonsmooth SCB based on the EDF $\hat{F}_n$ can be chosen for validation purposes.

5. Real data analysis

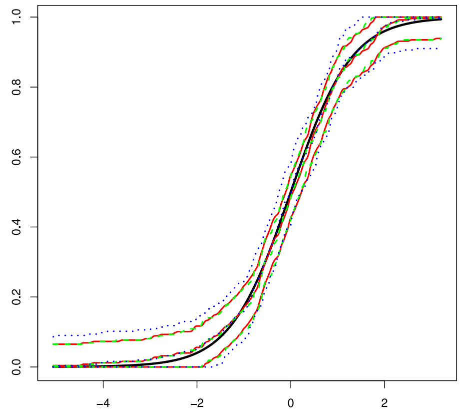
As an illustration, we applied our method to the Old Faithful geyser data with 272 observations on eruption time (explanatory variable X) and waiting time to next eruption (response variable Y) in minutes. This data set can be obtained in R by the command "data(faithful)". Via an SCB for the conditional variance function (Cai and Yang, 2015) argued that for the geyser data one cannot reject the null hypothesis of homoscedasticity. In addition, in order for identifiability to hold, the error Z in model (1) is set to have $\mathrm{E}(Z) = 0$. Therefore, the basic assumptions of zero mean and homoscedasticity are plausible for this data set.

In this illustration, we first follow the procedures described in Section 3 to estimate the regression curve and construct the SCBs for the error distribution as in our simulation studies. Then we use the constructed SCBs to conduct a simple hypothesis testing of whether the random error Z follows a specific normal distribution $N\left(0, \sigma_0^2\right)$, where σ_0^2 is given as one of reasonable candidate constants by the investigators. Note that in this setting generally rejection of the null hypothesis does not mean rejection of normality. It just means the rejection of the particular distribution stated in the simple null hypothesis. As an example, σ_0^2 is taken to be the value of the sample variance $\hat{\sigma}_Z^2 = 31.1$ of the estimated residuals $\hat{Z}$. From Fig. 2(a) with the proposed nonsmooth and smooth SCBs, obviously the null hypothesis $H_0 : Z \sim N\left(0, \sigma_0^2\right)$, where $\sigma_0^2 = 31.1$, cannot be rejected since the 95% SCBs completely cover this specific normal distribution curve.

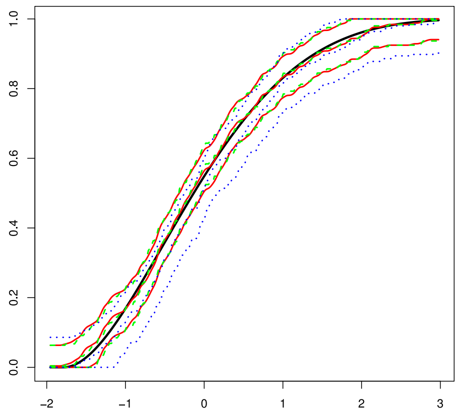
Moreover, according to (18) in Section 3, the asymptotic p-value for the proposed smooth SCB based on KDE $\hat{F}$ is calculated to be 0.782 under the null hypothesis $H_0 : Z \sim N\left(0, \sigma_0^2\right)$, where $\sigma_0^2 = 31.1$. The asymptotic p-value 0.551 for the proposed nonsmooth SCBs is obtained by replacing $\hat{F}$ with $\hat{F}_n$ in (18). Fig. 2(b) shows the narrowest SCB with the confidence level $100(1-p)\,\%$ which contains the normal distribution function $N(0, 31.1)$. That means for the geyser data one cannot reject $H_0 : Z \sim N\left(0, \sigma_0^2\right)$, where $\sigma_0^2 = 31.1$, with p-values 0.782 and 0.551 using the proposed smooth and



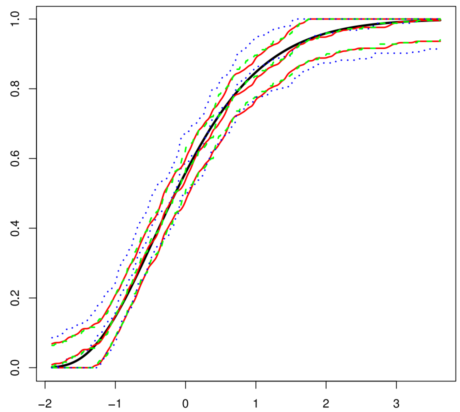
(a) $N(0,1/2)$



(b) t_8

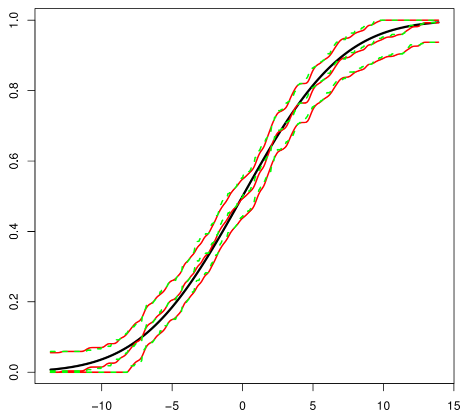


(c) standardized $Beta\,(2,5)$

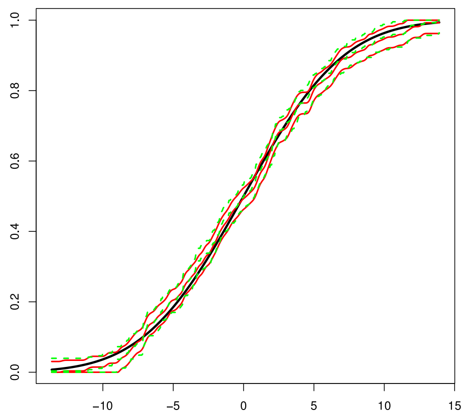


(d) standardized $\chi^2\,(10)$

Fig. 1. Plots of the true error cdf F (middle thick line), KDE $\hat{F}$ and smooth SCB (solid), EDF $\hat{F}_n$ and nonsmooth SCB (dashed), error EDF F_n and infeasible SCB (dotted) for four candidate error distributions F at $n = 250$, $1-\alpha = 0.95$, respectively, for one simulated data set.



(a)



(b)

Fig. 2. Plots of KDE $\hat{F}$ (solid), EDF $\hat{F}_n$ (dashed) for the Old Faithful geyser data with (a) $1-\alpha = 95\%$ smooth SCB (solid) and nonsmooth SCB (dashed); (b) $1-p = 21.8\%$ smooth SCB (solid) and 44.9% nonsmooth SCB (dashed), and the normal distribution (middle thick line) to be tested.

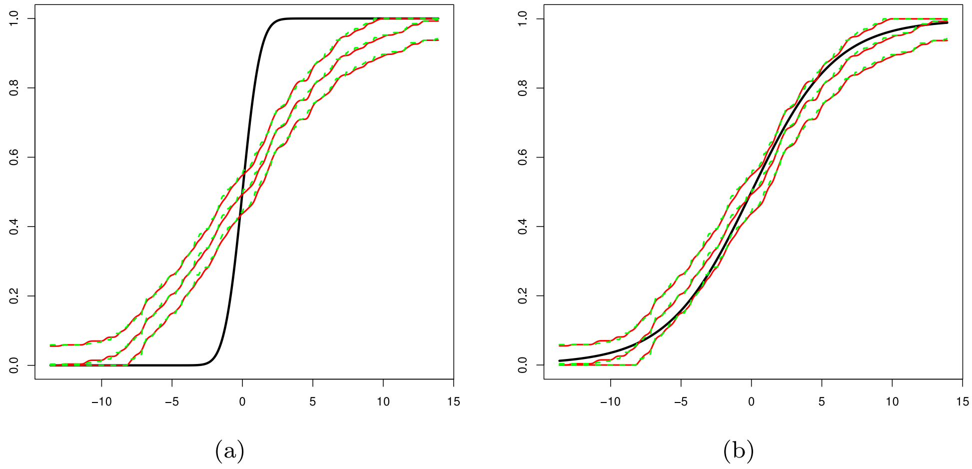


Fig. 3. Plots of $1-\alpha = 95\%$ smooth SCB and KDE $\hat{F}$ (solid), $1-\alpha = 95\%$ nonsmooth SCB and EDF $\hat{F}_n$ (dashed) for the Old Faithful geyser data with (a) the standard normal distribution and (b) the rescaled t_6 distribution (middle thick line) to be tested.

nonsmooth SCBs, respectively. Similarly, the corresponding p-values for the bootstrap method are 0.898 and 0.704 for the smooth and nonsmooth SCBs using the bootstrap quantiles $\hat{d}^{B}_{1-\alpha}$ defined in Section 3 instead of $\hat{d}^{*}_{1-\alpha}$ in (18).

In addition, two other null hypotheses are proposed: (i) $H_{01} : Z \sim N(0,1)$; (ii) $H_{02} : C_0 Z \sim t_6$, where the constant $C_0 = \left(2\hat{\sigma}_Z^2/3\right)^{-1/2} = 0.220$. Fig. 3(a) obviously shows that the standard normal distribution curve is not contained by the 95% smooth and nonsmooth SCBs so that H_{01} is rejected with a predetermined significance level 0.05. Meanwhile, Fig. 3(b) shows that the 95% SCBs does not completely cover the rescaled t_6 distribution curve either. Actually, the asymptotic p-values are 0.040 and 0.008 for the proposed smooth and nonsmooth SCBs, respectively, which imply the rejection of H_{02}. The overall conclusion is that it is plausible that the random error follows the normal distribution with mean zero and constant variance $\sigma_0^2 = 31.1$.

6. Concluding remarks

In this paper, we focused on constructing the asymptotic smooth SCB based on KDE $\hat{F}$ using regression residuals for the error distribution $F(z)$. We examined the proposed smooth SCB in comparison with the nonsmooth SCB based on the EDF $\hat{F}_n$ and the infeasible SCB. The smooth SCB is shown to perform the best in the sense that it achieves accurate coverage probabilities with high efficiency, especially compared with the infeasible SCB. Furthermore, we compared the proposed smooth SCB with the smooth residual bootstrap method. We generally recommend to use the bootstrap method when the sample size is small and to use the proposed SCB otherwise.

While our proposed SCB for the error distribution has clear advantages and shows surprisingly useful results especially in relatively large sample sizes, there are possible improvements to make. First, it might be better to consider the standardized versions of processes $\{\hat{F}(z) - F(z)\}$ and $\{\hat{F}_n(z) - F(z)\}$, which would make the SCBs more adaptive and more precise in terms of the overall confidence band area. Second, by relaxing some of the conditions imposed in this paper, basic ideas discussed in this paper might be extended from homoscedasticity to heteroscedasticity, and might be further extended to other models. These are some of interesting problems for future research.

Acknowledgments

We would like to thank the Co-Editor, an Associate Editor, and two referees for their constructive comments that substantially improved an earlier version of this paper. This research was supported in part by the National Natural Science Foundation of China Award NSFC 11701403, 11771240, the 2017 Jiangsu Overseas Visiting Scholar Program for University Prominent Young & Middle-aged Teachers and Presidents, Jiangsu Province Key-Discipline Program (Statistics), China GD10700118, and the Simons Foundation, United States Mathematics and Physical Sciences Program Award #499650.

Appendix

Throughout this Appendix, denote $a_n = o(b_n)$ as $\lim_{n\to\infty} a_n/b_n = 0$, and $a_n = O(b_n)$ as $\limsup_{n\to\infty} a_n/b_n = c$, where c is a constant. We use O_p (or o_p) to denote sequences of random variables of order O (or o) in probability. We also define $\|\varphi\|_\infty = \sup_{x\in\mathcal{D}} |\varphi(x)|$ as the supremum norm of a function $\varphi(x)$ on domain $\mathcal{D}$.

A.1. Preliminaries

Similar to Theorem 1 and Lemma 1 in Xia (1998), the following lemma gives the uniform convergence rate for the local linear estimator and its bias.

Lemma A.1. *Under Assumptions (A1)–(A4), for any compact subinterval* $[a_0, b_0]$ *in* (a, b),

$$\left\|\hat{m} - m\right\|_\infty = \sup_{x \in [a_0, b_0]} \left|\hat{m}(x) - m(x)\right| = O_p\left(\sqrt{\frac{\log n}{n\lambda}}\right) = o_p(1). \tag{A.1}$$

Lemma A.2. *Under Assumptions (A1)–(A4) and (A6),*

$$\left\|\hat{f} - f\right\|_\infty = \sup_{z \in \mathbb{R}} \left|\hat{f}(z) - f(z)\right| = o_p(1).$$

Proof. Define $\tilde{f}(z) = n^{-1} \sum_{j=1}^n L_H(Z_j - z)$ as the infeasible error density estimator. Since $\sqrt{n}\left\|F_n - F\right\|_\infty$ converges to the Kolmogorov distribution, one concludes that

$$\begin{aligned}
\left|\tilde{f}(z) - \mathrm{E}\tilde{f}(z)\right| &= \left|\int L_H(u - z)\, d\{F_n(u) - F(u)\}\right| \\
&= \left|H^{-2} \int \{F_n(u) - F(u)\} L'\left(\frac{u - z}{H}\right) du\right| \\
&\le H^{-2} \left\|F_n - F\right\|_\infty \times 2H \left\|L'\right\|_\infty = O_p\left(n^{-1/2} H^{-1}\right) = o_p(1),
\end{aligned}$$

and together with $F(z) \in C^{(1,1)}(\mathbb{R})$ under Assumption (A1) and (9), one has

$$\begin{aligned}
\left|\mathrm{E}\tilde{f}(z) - f(z)\right| &= \left|\int L_H(u - z)\{f(u) - f(z)\}\, du\right| \\
&\le \left\|F\right\|_{1,1} H \left|\int L_H(z - u)\, du\right| = \left\|F\right\|_{1,1} H = o(1).
\end{aligned}$$

Hence,

$$\sup_{z \in \mathbb{R}} \left|\tilde{f}(z) - f(z)\right| \le \sup_{z \in \mathbb{R}} \left|\tilde{f}(z) - \mathrm{E}\tilde{f}(z)\right| + \sup_{z \in \mathbb{R}} \left|\mathrm{E}\tilde{f}(z) - f(z)\right| = o_p(1).$$

On the other hand,

$$\begin{aligned}
\left|\hat{f}(z) - \tilde{f}(z)\right| &= \left|\frac{1}{nH} \sum_{j=1}^n \left\{L\left(\frac{\hat{Z}_j - z}{H}\right) - L\left(\frac{Z_j - z}{H}\right)\right\}\right| \\
&\le \frac{1}{nH} \sum_{j=1}^n \left|L'\left(\frac{Z_j - z}{H}\right)\left(\frac{\hat{Z}_j - Z_j}{H}\right) + \int_{(Z_j - z)/H}^{(\hat{Z}_j - z)/H} L''(u)\left(\frac{\hat{Z}_j - z}{H} - u\right) du\right| \\
&\le \frac{1}{nH} \sum_{j=1}^n \left\{\left\|L'\right\|_\infty H^{-1}\left|Z_j - \hat{Z}_j\right| + \frac{1}{2}\left\|L''\right\|_\infty H^{-2}\left|Z_j - \hat{Z}_j\right|^2\right\}.
\end{aligned}$$

Note that $Z_j - \hat{Z}_j = \hat{m}(X_j) - m(X_j)$, $j = 1, \ldots, n$. Then applying Lemma A.1, one obtains

$$\begin{aligned}
\sup_{z \in \mathbb{R}} \left|\hat{f}(z) - \tilde{f}(z)\right| &\le \left\|L'\right\|_\infty H^{-2} \left\|\hat{m} - m\right\|_\infty + \frac{1}{2}\left\|L''\right\|_\infty H^{-3} \left\|\hat{m} - m\right\|_\infty^2 \\
&= O_p\left(H^{-2}\sqrt{\frac{\log n}{n\lambda}} + H^{-3}\frac{\log n}{n\lambda}\right) = o_p(1).
\end{aligned}$$

Consequently,

$$\sup_{z \in \mathbb{R}} \left|\hat{f}(z) - f(z)\right| \le \sup_{z \in \mathbb{R}} \left|\hat{f}(z) - \tilde{f}(z)\right| + \sup_{z \in \mathbb{R}} \left|\tilde{f}(z) - f(z)\right| = o_p(1),$$

completing the proof of Lemma A.2. □

In addition, Lemma A.2 and the upper boundedness of $f(z)$ in Assumption (A1) entail that in probability

$$\sup_{z \in \mathbb{R}} \left|\hat{f}(z)\right| \le \sup_{z \in \mathbb{R}} \left|\hat{f}(z) - f(z)\right| + \sup_{z \in \mathbb{R}} \left|f(z)\right| \le o_p(1) + C_f = O_p(1). \tag{A.2}$$

Lemma A.3. *Given a sequence of i.i.d. random variables* $Z_1, Z_2, \ldots$*, satisfying* $\mathrm{E}(Z_1) = 0$, $\mathrm{E}\left(Z_1^2\right) = \sigma^2$*, with continuous cdf* $F(z)$ *and pdf* $f(z)$*,*

$$\sup_{z\in\mathbb{R}}\left|n^{-1}\sum_{j=1}^{n} Z_j I\left(Z_j \le z\right) - \mathrm{E}\left\{Z_1 I\left(Z_1 \le z\right)\right\}\right| = o_{a.s.}(1).$$

Proof. Define a class of functions $\mathcal{G} = \{g_z, z \in \mathbb{R} \,|\, g_z(u) = uI(u \le z), u \in \mathbb{R}\}$, and for any measurable function ϕ, denote $\mathbb{P}_n\phi = n^{-1}\sum_{j=1}^{n}\phi\left(Z_j\right)$, $\mathbb{P}\phi = \mathrm{E}\{\phi(Z_1)\}$ as in Pollard (1984) (p. 6). Let $g_0(u) \equiv 0$, $g_i(u) = g_{Z_{(i)}}(u) = uI\left(u \le Z_{(i)}\right)$, $1 \le i \le n$, where $Z_{(1)}, Z_{(2)}, \ldots, Z_{(n)}$ are the order statistics of random variables $Z_1, Z_2, \ldots, Z_n$. Define $G(u) = |u|$. Then $|g_z| \le G$ and $\mathbb{P}G = \mathrm{E}|Z_1| < \infty$. For $\forall\varepsilon > 0$,

$$\sup_{z\in\mathbb{R}}\min_{0\le i\le n} n^{-1}\sum_{j=1}^{n}\left|g_z\left(Z_j\right) - g_i\left(Z_j\right)\right| = 0 < \varepsilon.$$

Hence, the covering number $\mathcal{N}_1(\varepsilon, \mathbb{P}_n, \mathcal{G}) \le n+1$, and therefore $\log\mathcal{N}_1(\varepsilon, \mathbb{P}_n, \mathcal{G}) = o_p(n)$. Thus, according to Theorem 24 in Pollard (1984) (p. 25), $\sup_{g_z\in\mathcal{G}}|\mathbb{P}_n g_z - \mathbb{P}g_z| \to 0$ almost surely. The proof of Lemma A.3 is completed. □

A.2. Proof of Lemma 2

Define $U_j = F\left(Z_j\right), j = 1, \ldots, n$, which are i.i.d. random variables with the standard uniform distribution on [0, 1], and $u = F(z), z \in \mathbb{R}$. Then

$$\sqrt{n}\left\{F_n(z) + f(z)\,n^{-1}\sum_{j=1}^{n} Z_j - F(z)\right\} = n^{-1/2}\sum_{j=1}^{n}\left\{I\left(U_j \le u\right) + f\left(F^{-1}(u)\right)F^{-1}\left(U_j\right) - u\right\} \equiv \xi_n(u).$$

Step 1: We verify the first condition of Theorem 3 in Pollard (1984) (p. 92) for the stochastic process $\xi_n(u), u \in [0, 1]$. Since $U_j, j = 1, \ldots, n$, are i.i.d. random variables, one can easily apply the classical multivariate Central Limit Theorem to show that, for any fixed integer $k > 0$ and any $0 \le u_0 < u_1 < \cdots < u_k \le 1$, $(\xi_n(u_0), \xi_n(u_1), \ldots, \xi_n(u_k))$ is asymptotically normal with mean zero $\mathrm{E}\xi_n(u) = 0$ and covariance matrix $\mathbf{\Gamma} = (\sigma(u_l, u_{l'}))_{0\le l,l'\le k}$, where

$$\begin{aligned}\sigma(u_l, u_{l'}) &= \mathrm{E}\xi_n(u_l)\,\xi_n(u_{l'})\\ &= u_l \wedge u_{l'} - u_l u_{l'} + f\left(F^{-1}(u_l)\right)\mathrm{E}\left\{F^{-1}(U_1)\,I(U_1 \le u_{l'})\right\}\\ &\quad + f\left(F^{-1}(u_{l'})\right)\mathrm{E}\left\{F^{-1}(U_1)\,I(U_1 \le u_l)\right\} + f\left(F^{-1}(u_l)\right)f\left(F^{-1}(u_{l'})\right)\sigma^2.\end{aligned}$$

Let $u_l = F(z_l), u_{l'} = F(z_{l'})$. Then $\sigma(u_l, u_{l'}) = \Sigma(z_l, z_{l'})$.

Step 2: We verify the second condition of Theorem 3 in Pollard (1984). Denote $\xi_n^*(u) = \sqrt{n}\left\{n^{-1}\sum_{j=1}^{n} I\left(U_j \le u\right) - u\right\}$. Then

$$\xi_n(u) - \xi_n(u_l) = \xi_n^*(u) - \xi_n^*(u_l) + \left\{f\left(F^{-1}(u)\right) - f\left(F^{-1}(u_l)\right)\right\}n^{-1/2}\sum_{j=1}^{n} F^{-1}\left(U_j\right).$$

Note that $\xi_n^*(u)$ is the standard empirical process. According to Theorem 9 in Pollard (1984) (p. 96), $\xi_n^*(u)$ satisfies the second condition of Theorem 3 in Pollard (1984), i.e., for each $\varepsilon > 0$, $\delta > 0$, one can find a grid $0 = u_0^* < u_1^* < \cdots < u_{N_1}^* = 1$ such that

$$\limsup_{n\to\infty}\mathbb{P}\left\{\max_{0\le l\le N_1-1}\sup_{u_l^*\le u<u_{l+1}^*}\left|\xi_n^*(u) - \xi_n^*\left(u_l^*\right)\right| > \frac{\delta}{2}\right\} < \frac{\varepsilon}{2}.$$

Let $S_{1N_1} = \left\{u_l^* : l = 0, \ldots, N_1\right\}$ be the set of the N_1 grid points. For the same ε and δ, we take N_2 as the integer part of $\left[2\delta^{-1}\sigma\left\|\left(f\circ F^{-1}\right)'\right\|_\infty\sqrt{2/\varepsilon}\right] + 1$ to obtain the equal distance grid points $0 = u_0' < u_1' < \cdots < u_{N_2}' = 1$. Let $S_{2N_2} = \left\{u_l' : l = 0, \ldots, N_2\right\}$ be the set of the N_2 grid points. Using the differentiability and continuity of f and F, along with the Chebyshev Inequality, one has

$$\mathbb{P}\left\{\max_{0\le l\le N_2-1}\sup_{u_l'\le u<u_{l+1}'}\left|\left\{f\left(F^{-1}(u)\right) - f\left(F^{-1}\left(u_l'\right)\right)\right\}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}F^{-1}(U_j)\right| > \frac{\delta}{2}\right\}$$

$$\le \mathbb{P}\left\{\max_{0\le l\le N_2-1}\sup_{u_l'\le u<u_{l+1}'}\left\|\left(f\circ F^{-1}\right)'\right\|_\infty\left|u - u_l'\right|\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}F^{-1}\left(U_j\right)\right| > \frac{\delta}{2}\right\}$$

$$\leq \mathbb{P}\left\{\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}F^{-1}\left(U_j\right)\right| > \frac{\delta N_2}{2\left\|\left(f\circ F^{-1}\right)'\right\|_\infty}\right\}$$

$$\leq \left(\frac{2\sigma\left\|\left(f\circ F^{-1}\right)'\right\|_\infty}{\delta N_2}\right)^2 < \frac{\varepsilon}{2}.$$

Hence, letting $S_{3N_3} = S_{1N_1} \cup S_{2N_2} = \{u_l : l = 0, \ldots, N_3\}$ be the set of $N_3 = N_1 + N_2$ grid points, we obtain

$$\limsup_{n\to\infty} \mathbb{P}\left\{\max_{0\leq l\leq N_3-1}\sup_{u_l\leq u<u_{l+1}}\left|\xi_n(u)-\xi_n(u_l)\right| > \delta\right\}$$

$$\leq \limsup_{n\to\infty}\left[\mathbb{P}\left\{\max_{0\leq l\leq N_1-1}\sup_{u_l^*\leq u<u_{l+1}^*}\left|\xi_n^*(u)-\xi_n^*\left(u_l^*\right)\right| > \frac{\delta}{2}\right\} + \right.$$

$$\left.\mathbb{P}\left\{\max_{0\leq l\leq N_2-1}\sup_{u_l'\leq u<u_{l+1}'}\left|\left\{f\left(F^{-1}(u)\right)-f\left(F^{-1}\left(u_l'\right)\right)\right\}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}F^{-1}(U_j)\right| > \frac{\delta}{2}\right\}\right]$$

$$\leq \frac{\varepsilon}{2}+\frac{\varepsilon}{2} = \varepsilon.$$

Altogether, by Theorem 3 in Pollard (1984) the process $\xi_n(u), u \in [0,1]$, converges in distribution to a Gaussian process. Lemma 2 has been shown. □

A.3. *Proof of Theorem 2*

According to the definitions of $\Sigma\left(z, z'\right)$ in (11) and $\hat{\Sigma}_n\left(z, z'\right)$ in (15), $\hat{\Sigma}_n\left(z, z'\right) - \Sigma\left(z, z'\right)$ can be decomposed into seven parts as follows:

$$\begin{aligned}
&\hat{\Sigma}_n\left(z,z'\right)-\Sigma\left(z,z'\right)\\
&=\left\{\hat{F}_n\left(z\wedge z'\right)-F\left(z\wedge z'\right)\right\}-\left\{\hat{F}_n(z)\hat{F}_n\left(z'\right)-F(z)F\left(z'\right)\right\}\\
&\quad+\left[\hat{f}\left(z'\right)n^{-1}\sum_{j=1}^{n}\hat{Z}_jI\left(\hat{Z}_j\leq z\right)-f\left(z'\right)\mathrm{E}\left\{Z_1I\left(Z_1\leq z\right)\right\}\right]\\
&\quad+\left[\hat{f}(z)n^{-1}\sum_{j=1}^{n}\hat{Z}_jI\left(\hat{Z}_j\leq z'\right)-f(z)\mathrm{E}\left\{Z_1I\left(Z_1\leq z'\right)\right\}\right]\\
&\quad+\left\{\hat{f}(z)\hat{f}\left(z'\right)n^{-1}\sum_{j=1}^{n}\hat{Z}_j^2-f(z)f\left(z'\right)\sigma^2\right\}\\
&\quad-\left\{\hat{F}_n(z)\hat{f}\left(z'\right)n^{-1}\sum_{j=1}^{n}\hat{Z}_j\right\}-\left\{\hat{F}_n\left(z'\right)\hat{f}(z)n^{-1}\sum_{j=1}^{n}\hat{Z}_j\right\}\\
&\equiv I_1-I_2+I_3+I_4+I_5-I_6-I_7.
\end{aligned}$$

By Theorem 1.1 in Müller et al. (2007), one can easily obtain that $\sup_{z\in\mathbb{R}}\left|\hat{F}_n(z)-F(z)\right| = o_p(1)$. Then

$$\sup_{z,z'\in\mathbb{R}}\left|I_1\right| = \sup_{z,z'\in\mathbb{R}}\left|\hat{F}_n\left(z\wedge z'\right)-F\left(z\wedge z'\right)\right| = o_p(1),$$

and

$$\begin{aligned}
\sup_{z,z'\in\mathbb{R}}\left|I_2\right| &\leq \sup_{z,z'\in\mathbb{R}}\left|\hat{F}_n(z)\right|\left|\hat{F}_n\left(z'\right)-F\left(z'\right)\right| + \sup_{z,z'\in\mathbb{R}}\left|\hat{F}_n(z)-F(z)\right|\left|F\left(z'\right)\right|\\
&\leq \sup_{z'\in\mathbb{R}}\left|\hat{F}_n\left(z'\right)-F\left(z'\right)\right| + \sup_{z\in\mathbb{R}}\left|\hat{F}_n(z)-F(z)\right| = o_p(1).
\end{aligned}$$

In order to prove $\sup_{z,z'\in\mathbb{R}}|I_3| = o_p(1)$, we also need to decompose I_3 into several parts as follows:

$$I_3 = \hat{f}(z')\left\{n^{-1}\sum_{j=1}^{n}\hat{Z}_j I\left(\hat{Z}_j \le z\right) - n^{-1}\sum_{j=1}^{n} Z_j I\left(Z_j \le z\right)\right\} + \hat{f}(z')\left\{n^{-1}\sum_{j=1}^{n} Z_j I\left(Z_j \le z\right) - \mathrm{E}\{Z_1 I(Z_1 \le z)\}\right\}$$
$$+ \left\{\hat{f}(z') - f(z')\right\}\mathrm{E}\{Z_1 I(Z_1 \le z)\} \equiv I_{3,1} + I_{3,2} + I_{3,3}. \tag{A.3}$$

Hence, by (A.2) and Lemmas A.2 and A.3, one has

$$sup_{z,z'\in\mathbb{R}}\left|I_{3,2}\right| = \sup_{z'\in\mathbb{R}}\left|\hat{f}(z')\right|\sup_{z\in\mathbb{R}}\left|\frac{1}{n}\sum_{j=1}^{n} Z_j I\left(Z_j \le z\right) - \mathrm{E}\{Z_1 I(Z_1 \le z)\}\right| = o_p(1), \tag{A.4}$$

and

$$\sup_{z,z'\in\mathbb{R}}\left|I_{3,3}\right| = \sup_{z,z'\in\mathbb{R}}\left|\hat{f}(z') - f(z')\right|\left|\mathrm{E}\{Z_1 I(Z_1 \le z)\}\right| = o_p(1). \tag{A.5}$$

In the following, we focus on $\sup_{z,z'\in\mathbb{R}}\left|I_{3,1}\right|$. Firstly, by the Cauchy–Schwarz Inequality, one has

$$n^{-1}\sum_{j=1}^{n} Z_j\left\{I\left(Z_j \le z\right) - I\left(\hat{Z}_j \le z\right)\right\}$$
$$\le \sqrt{n^{-1}\sum_{j=1}^{n} Z_j^2 \cdot n^{-1}\sum_{j=1}^{n}\left\{I\left(Z_j \le z\right) - I\left(\hat{Z}_j \le z\right)\right\}^2}$$
$$= \sqrt{n^{-1}\sum_{j=1}^{n} Z_j^2}\cdot\sqrt{\left\{F_n(z) + \hat{F}_n(z) - 2n^{-1}\sum_{j=1}^{n} I\left(Z_j \le z, \hat{Z}_j \le z\right)\right\}}.$$

Note that $Z_j - \hat{Z}_j = \hat{m}(X_j) - m(X_j),\ j = 1, \ldots, n$. (A.1) in Lemma A.1 also shows that $\sup_{1\le j\le n}\left|\hat{Z}_j - Z_j\right| = o_p(1)$, i.e., for $\forall\varepsilon > 0$, $\mathbb{P}\left\{\sup_{1\le j\le n}\left|\hat{Z}_j - Z_j\right| > \varepsilon\right\} \to 0$, as $n \to \infty$. On the other hand,

$$n^{-1}\sum_{j=1}^{n} I\left(Z_j \le z, \hat{Z}_j \le z\right)$$
$$= n^{-1}\sum_{j=1}^{n} I\left(Z_j \le z\right) - n^{-1}\sum_{j=1}^{n} I\left(Z_j \le z, \hat{Z}_j > z\right)$$
$$= F_n(z) - \frac{1}{n}\sum_{j=1}^{n} I\left(Z_j \le z - \varepsilon, \hat{Z}_j > z\right) - \frac{1}{n}\sum_{j=1}^{n} I\left(z - \varepsilon < Z_j \le z, \hat{Z}_j > z\right).$$

For $\forall\varepsilon_0 > 0$,

$$\mathbb{P}\left\{\sup_{z\in\mathbb{R}}\left|n^{-1}\sum_{j=1}^{n} I\left(Z_j \le z - \varepsilon, \hat{Z}_j > z\right)\right| > \varepsilon_0\right\}$$
$$\le \mathbb{P}\left\{\left|n^{-1}\sum_{j=1}^{n} I\left(\left|Z_j - \hat{Z}_j\right| > \varepsilon\right)\right| > \varepsilon_0\right\}$$
$$\le \mathbb{P}\left\{\sup_{1\le j\le n}\left|\hat{Z}_j - Z_j\right| > \varepsilon\right\} \to 0, \text{ as } n \to \infty,$$

which implies that $\sup_{z\in\mathbb{R}}\left|n^{-1}\sum_{j=1}^{n} I\left(Z_j \le z - \varepsilon, \hat{Z}_j > z\right)\right| = o_p(1)$. Meanwhile,

$$n^{-1}\sum_{j=1}^{n} I\left(z - \varepsilon < Z_j \le z, \hat{Z}_j > z\right) \le n^{-1}\sum_{j=1}^{n} I\left(z - \varepsilon < Z_j \le z\right) = F_n(z) - F_n(z - \varepsilon).$$

Since $\sup_{z\in\mathbb{R}}|F_n(z)-F(z)|=o_p(1)$, $\sup_{z\in\mathbb{R}}|F_n(z)-F_n(z-\varepsilon)|\to_p F(z)-F(z-\varepsilon)$. Furthermore, the arbitrariness of small positive number ε and the continuity of F imply that $\sup_{z\in\mathbb{R}}|F_n(z)-F_n(z-\varepsilon)|\to_p 0$. Thus,

$$\sup_{z\in\mathbb{R}}\left|n^{-1}\sum_{j=1}^{n}I\left(Z_j\le z,\hat{Z}_j\le z\right)-F(z)\right|\to_p 0.$$

In addition, by the Strong Law of Large Numbers, $n^{-1}\sum_{j=1}^{n}Z_j^2\to_p\sigma^2$. Hence,

$$\sup_{z\in\mathbb{R}}\left|n^{-1}\sum_{j=1}^{n}Z_j\left\{I\left(Z_j\le z\right)-I\left(\hat{Z}_j\le z\right)\right\}\right|=o_p(1).$$

Moreover,

$$\begin{aligned}\sup_{z,z'\in\mathbb{R}}\left|I_{3,1}\right| &= \sup_{z'\in\mathbb{R}}\left|\hat{f}\left(z'\right)\right|\sup_{z\in\mathbb{R}}\left|n^{-1}\sum_{j=1}^{n}\hat{Z}_jI\left(\hat{Z}_j\le z\right)-n^{-1}\sum_{j=1}^{n}Z_jI\left(Z_j\le z\right)\right| \\ &\le O_p(1)\left\{\sup_{z\in\mathbb{R}}\left|n^{-1}\sum_{j=1}^{n}\left(\hat{Z}_j-Z_j\right)I\left(\hat{Z}_j\le z\right)\right|+\sup_{z\in\mathbb{R}}\left|n^{-1}\sum_{j=1}^{n}Z_j\left\{I\left(Z_j\le z\right)-I\left(\hat{Z}_j\le z\right)\right\}\right|\right\} \\ &\le O_p(1)\left\{\sup_{1\le j\le n}\left|\hat{Z}_j-Z_j\right|+o_p(1)\right\}=o_p(1). \end{aligned}\tag{A.6}$$

Consequently, $\sup_{z,z'\in\mathbb{R}}|I_3|=o_p(1)$ follows from (A.3), (A.4), (A.5), and (A.6). The result $\sup_{z,z'\in\mathbb{R}}|I_4|=o_p(1)$ can be shown in the exactly same fashion.

Furthermore, using the uniform boundedness of $f(z)$, $\hat{f}(z)$, and $\hat{F}_n(z)$, Lemmas A.1 and A.2, and the Strong Law of Large Numbers, one has

$$\begin{aligned}\sup_{z,z'\in\mathbb{R}}|I_5| \le& \sup_{z,z'\in\mathbb{R}}\left|\hat{f}(z)\hat{f}\left(z'\right)\right|\left|n^{-1}\sum_{j=1}^{n}\hat{Z}_j^2-n^{-1}\sum_{j=1}^{n}Z_j^2\right|+\sup_{z,z'\in\mathbb{R}}\left|\hat{f}(z)\hat{f}\left(z'\right)-f(z)f\left(z'\right)\right|\left|n^{-1}\sum_{j=1}^{n}Z_j^2\right| \\ &+\sup_{z,z'\in\mathbb{R}}\left|f(z)f\left(z'\right)\right|\left|n^{-1}\sum_{j=1}^{n}Z_j^2-\sigma^2\right|=o_p(1),\end{aligned}$$

and

$$\sup_{z,z'\in\mathbb{R}}|I_6|\le\sup_{z,z'\in\mathbb{R}}\left|\hat{F}_n(z)\hat{f}\left(z'\right)\right|\left|n^{-1}\sum_{j=1}^{n}\hat{Z}_j-n^{-1}\sum_{j=1}^{n}Z_j\right|+\sup_{z,z'\in\mathbb{R}}\left|\hat{F}_n(z)\hat{f}\left(z'\right)\right|\left|n^{-1}\sum_{j=1}^{n}Z_j-0\right|=o_p(1).$$

Likewise, $\sup_{z,z'\in\mathbb{R}}|I_7|=o_p(1)$.

Finally, Theorem 2 follows since each part in its partition is $o_p(1)$. □

References

Akritas, M., Van Keilegom, I., 2001. Non-parametric estimation of the residual distribution. Scand. J. Stat. 28, 549–567.
Cai, L., Liu, R., Wang, S., Yang, L., 2019. Simultaneous confidence bands for mean and variance functions based on deterministic design. Statist. Sinica 29, 505–525.
Cai, L., Yang, L., 2015. A smooth simultaneous confidence band for conditional variance function. TEST 24, 632–655.
Cao, G., Yang, L., Todem, D., 2012. Simultaneous inference for the mean function based on dense functional data. J. Nonparametr. Stat. 24, 359–377.
Cheng, F., 2002. Consistency of error density and distribution function estimators in nonparametric regression. Statist. Probab. Lett. 59, 257–270.
Cheng, F., 2005. Asymptotic distributions of error density and distribution function estimators in nonparametric regression. J. Statist. Plann. Inference 128, 327–349.
Degras, D., 2011. Simultaneous confidence bands for nonparametric regression with functional data. Statist. Sinica 21, 1735–1765.
Dette, H., Neumeyer, N., Van Keilegom, I., 2007. A new test for the parametric form of the variance function in non-parametric regression. J. R. Stat. Soc. Ser. B 69, 903–917.
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., Wang, S., Yang, L., 2019. Simultaneous confidence bands for the distribution function of a finite population in stratified sampling. Ann. Inst. Statist. Math. 71, 983–1005.
Gu, L., Yang, L., 2015. Oracally efficient estimation for single-index link function with simultaneous confidence band. Electron. J. Stat. 9, 1540–1561.
Härdle, W., 1989. Asymptotic maximal deviation of M-smoothers. J. Multivariate Anal. 29, 163–179.
Krivobokova, T., Kneib, T., Claeskens, G., 2010. Simultaneous confidence bands for penalized spline estimators. J. Amer. Statist. Assoc. 105, 852–863.
Müller, U.U., Schick, A., Wefelmeyer, W., 2004. Estimating the error distribution function in nonparametric regression. Statist. Decisions 1, 1–4.
Müller, U.U., Schick, A., Wefelmeyer, W., 2007. Estimating the error distribution function in semiparametric regression. Statist. Decisions 25, 1–18.

Müller, U.U., Schick, A., Wefelmeyer, W., 2012. Estimating the error distribution function in semiparametric additive regression models. J. Statist. Plann. Inference 142, 552–566.
Neumeyer, N., 2009. Smooth residual bootstrap for empirical processes of non-parametric regression residuals. Scand. J. Stat. 36, 204–228.
Pierce, D.A., 1982. The asymptotic effect of substituting estimators for parameters in certain types of statistics. Ann. Statist. 36, 475–478.
Pollard, D., 1984. Convergence of Stochastic Processes. Springer-Verlag, New York.
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. Statist. 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, C.Y., Wang, S., Zhao, L., Ou, S., 1997. Weighted semiparametric estimation in regression analysis with missing covariate data. J. Amer. Statist. Assoc. 92, 512–525.
Wang, J., Yang, L., 2009. Polynomial spline confidence bands for regression curves. Statist. Sinica 10, 325–342.
Xia, Y., 1998. Bias-corrected confidence bands in nonparametric regression. J. R. Statist. Soc. Ser. B 80, 797–811.
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.
Zhang, Y., Yang, L., 2018. A smooth simultaneous confidence band for correlation curve. TEST 27, 247–269.
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. Amer. Statist. Assoc. 109, 661–673.
Zhu, H., Li, R., Kong, L., 2012. Multivariate varying coefficient model for functional responses. Ann. Statist. 40, 2634–2666.