Confidence intervals in monotone regression

Groeneboom, Piet; Jongbloed, Geurt

DOI
10.1111/sjos.12730

Publication date
2024

Document Version
Final published version

Published in
Scandinavian Journal of Statistics

Citation (APA)
Groeneboom, P., & Jongbloed, G. (2024). Confidence intervals in monotone regression. *Scandinavian Journal of Statistics*, *51*(4), 1749-1781. https://doi.org/10.1111/sjos.12730

Received: 31 March 2023 | Revised: 6 January 2024 | Accepted: 16 May 2024

DOI: 10.1111/SJOS.12730

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Confidence intervals in monotone regression

Piet Groeneboom | **Geurt Jongbloed**

Delft Institute of Applied Mathematics, Delft, Netherlands

Correspondence
Piet Groeneboom, Delft Institute of Applied Mathematics, Mekelweg 4, 2628 CD Delft, The Netherlands
Email: p.groeneboom@tudelft.nl

Abstract

We construct bootstrap confidence intervals for a monotone regression function. It has been shown that the ordinary nonparametric bootstrap, based on the nonparametric least squares estimator (LSE) $\hat{f}_n$, is inconsistent in this situation. We show that an $n^{2/5}$-consistent bootstrap can be based on the smoothed $\hat{f}_n$, to be called the SLSE (Smoothed Least Squares Estimator). The asymptotic pointwise distribution of the SLSE is derived. The confidence intervals, based on the smoothed bootstrap, are compared to intervals based on the (not necessarily monotone) Nadaraya Watson estimator and the effect of Studentization is investigated. We also give a method for automatic bandwidth choice, correcting work in Sen and Xu (2015). Analogous methods for constructing confidence intervals in the current status model are discussed, improving on work in Groeneboom and Hendrickx (2018).

KEYWORDS

bandwidth choice, confidence intervals, Nadaraya Watson, smooth monotone regression, smoothed bootstrap

1 | INTRODUCTION

We consider the monotone regression setting where we observe independent pairs (X_i, Y_i) of random variables $(1 \le i \le n)$, where the X_i are i.i.d. with nonvanishing density g on $[0, 1]$ and

$$Y_i = f_0(X_i) + \varepsilon_i, \quad 1 \le i \le n. \tag{1}$$

The regression function $f_0 : [0,1] \mapsto \mathbb{R}$ is nondecreasing or nonincreasing and the ε_i are i.i.d. sub-Gaussian with expectation 0 and variance σ_0^2, independent of the X_i's. Our aim is to construct pointwise nonparametric confidence intervals for $f_0(t)$. In the discussion below we will only return to the nonincreasing functions in the example, given in Section 6, and for the rest stick to the nondecreasing functions. The theory for these cases is completely similar.

The monotone regression model is often natural to assume, when the expected value of the response naturally increases with the value of the explanatory variable. Think of Y as the height of a child of age X in a certain region. In other situations, monotonicity may not be that obvious and one may want to test monotonicity. Such tests for (local and global) monotonicity are discussed, for example, in Armstrong (2015), Hall and Van Keilegom (2005), and Groeneboom and Jongbloed (2012). A basic ingredient of such tests is a monotone estimator of the regression function.

The basic monotone least squares estimate (LSE) $\hat{f}_n$ of f_0 is the so-called isotonic regression of (X_i, Y_i). This estimator is defined as minimizer of

$$\sum_{i=1}^{n} (Y_i - f(X_i))^2,$$

over all nondecreasing functions f. This LSE can be computed via a straightforward method, using the so-called cumulative sum diagram (cusum diagram). From now on, we interpret $X_1, \ldots, X_n$ as *ordered* in the sense that $X_1 < X_2 \cdots < X_n$ and relabel the Y_i's accordingly (as Y_i related to the specific X_i). The cusum diagram is then the set of points

$$(0,0), \qquad \left(i, \sum_{j \le i} Y_j\right), \qquad i = 1, \ldots, n,$$

and the monotone least squares estimator $\hat{f}_n(X_i)$ is given by the left-continuous slope of the greatest convex minorant of the cusum diagram evaluated at i (see lemma 2.1 in Groeneboom & Jongbloed, 2014).

Constructing nonparametric pointwise bootstrap confidence intervals for $f_0(t)$ poses several difficulties. It has been proved by several authors that the straightforward bootstrap, using resampling with replacement from the pairs (X_i, Y_i) and computing the monotone least squares estimator $\hat{f}_n$ based on the bootstrap samples, is inconsistent (see, e.g., Abrevaya & Huang, 2005, Kosorok, 2008, Sen & Xu, 2015, and Sen et al., 2010 for results related to this phenomenon).

A problem closely related to least squares estimation in monotone regression, is maximum likelihood estimation in the current status model. In this model, the variable of interest is a survival variable X with distribution function F_0. Instead of observing the exact survival time X, a censoring variable $T \sim G$ is observed together with the indicator $\Delta = 1_{X \le T}$. Such data arise naturally in clinical trials when a patient can only be checked at one measurement due to destructive testing.

Within this context, it has been suggested in Sen and Xu (2015) to construct confidence intervals by using a smoothed bootstrap. In the monotone regression setting, their approach means that one fixes the values of the X_i's and generates bootstrap values of the Y_i by resampling with replacement from the residuals of the Y_i with respect to a smooth monotone estimate of f_0. This approach addresses the intrinsic cause of the inconsistency of the bootstrap method based on direct resampling of the pairs, namely the fact that the derivative f_0' cannot be estimated directly by differentiating the (piecewise constant) monotonic least squares estimator $\hat{f}_n$.

Bootstrap confidence intervals based on the smoothed maximum likelihood estimator (SMLE) for the current status model were also proposed in Groeneboom and Hendrickx (2017) and Groeneboom and Hendrickx (2018). One of the main issues in these papers is the treatment of the bias, which will be treated differently in the present paper. The usual methods for dealing with the bias are undersmoothing and correction by estimating the bias. These methods are rather unsatisfactory for the present model. Härdle and Marron (1991) suggest a third method, which we will systematically use in the sequel.

A lot of research has been published on the behavior of the maximum likelihood estimator (MLE) $\hat{F}_n$ of the distribution function F_0 in the current status model. The limiting distribution of $n^{1/3}(\hat{F}_n(t) - F_0(t))$ is, after scaling by the constant $\kappa = \{4F_0(t)(1 - F_0(t))f_0(t)/g(t)\}^{1/3}$ given by,

$$\mathbb{Z} = \arg\max_t \{W(t) - t^2\},$$

where W is a two-sided Brownian motion with $W(0) = 0$ (see Groeneboom & Wellner, 1992). Other estimators with similar asymptotic properties are Chernoff's estimator of the mode (Chernoff, 1964), the Grenander estimator (Grenander, 1956) of a nonincreasing density, Manski's maximum score estimator (Manski, 1975) and Rouseeuw's least median of squares estimator (Rousseeuw, 1984). A general framework for cube-root n asymptotics is given in Kim and Pollard (1990).

It is known that the nonparametric bootstrap is inconsistent for generating the limit distribution of the MLE. Abrevaya and Huang (2005) prove that

$$\begin{aligned} &n^{1/3}\{4F_0(t)(1 - F_0(t))f_0(t)/g(t)\}^{-1/3}\{\hat{f}_n^*(t) - \hat{F}_n(t)\} \\ &\quad \xrightarrow{\mathcal{D}} \quad \arg\max_t(W(t) + \hat{W}(t) - t^2) - \arg\max_t(W(t) - t^2), \end{aligned}$$

where $\hat{f}_n^*$ is the bootstrap MLE and W and $\hat{W}$ are two independent two-sided standard Brownian motions originating at zero. Similar results are discussed in Kosorok (2008) and in Sen et al. (2010) for the Grenander estimator. The maximum score estimator of Manski (1975) is another example of a cube-root n statistic with asymptotic distribution first derived in Kim and Pollard (1990), where inconsistency of the nonparametric bootstrap for this estimator is shown in Abrevaya and Huang (2005).

In Section 2, we define a Smoothed Least Squares Estimator (SLSE) for f_0 in the regression context. We show that, under some conditions, this estimator is asymptotically normally distributed with rate $n^{2/5}$ and derive its asymptotic bias and variance. Furthermore, we consider the smooth but not necessarily monotone Nadaraya Watson (NW) estimator of f_0. Based on the SLSE and the NW estimator, we propose bootstrap methods to construct confidence sets for $f_0(t)$ in Section 3.

We also prove a theorem stating that the bootstrap method based on the SLSE asymptotically works, with specific choices for the various bandwidths involved. In particular, we show that the method for computing the optimal bandwidth in Sen and Xu (2015) will only work if bandwidths of different order are chosen. Moreover, we empirically study the effect of Studentization on coverage probabilities. In Section 4 we address the problem of bandwidth selection in practice. Also here, we propose a smoothed bootstrap approach.

In Section 5 we treat analogous methods for the aforementioned current status model and give a much better method for treating the bias than in Groeneboom and Hendrickx (2018). We also compare with cube root n convergent methods here, based on the MLE itself. Using the MLE without smoothing has the (at first sight) attractive aspect that we do not have to specify a bandwidth. However, one still needs smoothness restrictions in applying the asymptotic convergence of the MLE to Chernoff's distribution or a variant of this distribution in the Bannerjee–Wellner confidence intervals (Banerjee & Wellner, 2005). Also, in the Sen-Xu confidence intervals (Sen & Xu, 2015), one needs the SMLE in the centering of the bootstrap intervals, for which one needs to specify a bandwidth. And finally, one alleviates the dependence on the bandwidth choice by applying an automatic method of bandwidth choice as given in Section 4.

Section 6 illustrates the method using a well known data set related to climate change. We note that the methods described in this paper can be applied using the open software (Groeneboom, 2021).

2 | SMOOTH (MONOTONE) ESTIMATION OF THE REGRESSION FUNCTION

As immediately follows from its construction, the LSE $\hat f_n$ is a (nonsmooth) step function. It can be used to define a smooth estimate. We define a particular SLSE, $\tilde f_{nh}$. For this, let K be a symmetric twice continuously differentiable nonnegative kernel with support $[-1,1]$ such that $\int K(u)\,du = 1$. Let $h>0$ be a bandwidth and define the scaled kernel K_h by

$$K_h(u) = \frac{1}{h}K\left(\frac{u}{h}\right), \quad u \in \mathbb{R}. \tag{2}$$

Then, for $t \in [h, 1-h]$, the SLSE is defined as a two-sided local average of the LSE,

$$\tilde f_{nh}(t) = \int K_h(t-x)\,\hat f_n(x)\,dx. \tag{3}$$

Note that on $[h,1-h]$ the SLSE inherits monotonicity from $\hat f_n(t)$. For $t \notin [h,1-h]$ we use a boundary correction to be discussed later. In this paper, we choose for K the triweight kernel

$$K(u) = \frac{35}{32}\left(1-u^2\right)^3 1_{[-1,1]}(u).$$

Estimator (3) is rather different from the NW estimator, which is defined by

$$\tilde f_{nh}^{NW}(t) = \frac{\int yK_h(t-x)\,d\mathbb{H}_n(x,y)}{\int K_h(t-x)\,d\mathbb{H}_n(x,y)} = \frac{\sum_{i=1}^n Y_i K_h(t-X_i)}{\sum_{i=1}^n K_h(t-X_i)}, \tag{4}$$

for $t \in [h,1-h]$, where $\mathbb{H}_n$ is the empirical distribution function of the pairs (X_i, Y_i), $i = 1, \ldots, n$. This estimator can also be differentiated, just as (3), but is not necessarily monotone.

The NW estimator (4) is seemingly simpler than SLSE (3), because it is expressed as ratio of sums over sample values, whereas (3) is an integral with respect to Lebesgue measure. However, using integration by parts, (3) can be rewritten as a simple sum. Indeed, for $t \in [h,1-h]$,

$$\tilde f_{nh}(t) = \hat f_n(0) + \int_{x\in(0,t+h]} IK_h(t-x)\,d\hat f_n(x) = \hat f_n(0) + \sum_{\tau_i \in (0,t+h]} IK_h(t-\tau_i)p_i, \tag{5}$$

where

$$IK_h(y) = \int_{-\infty}^{y} K_h(u)\,du = \int_{-\infty}^{y/h} K(u)\,du. \tag{6}$$

Here the τ_i are the locations of the jumps of the LSE $\hat{f}_n$ and the $p_i > 0$ the sizes of the jumps. This time the summation is over points at stochastic locations τ_i with stochastic masses p_i, characterizing the LSE.

It is well known that defining $\tilde{f}_{nh}$ as in (3) outside the interval $[h, 1-h]$, leads to serious bias. We define $\tilde{f}_{nh}$ therefore slightly differently on the intervals near zero and one, using quadratic Taylor approximations at the points h and $1-h$, respectively. More precisely, for $t \in [0, h)$ we define

$$\tilde{f}_{nh}(t) = \tilde{f}_{nh}(h) + (t-h)\tilde{f}'_{nh}(h) + \frac{1}{2}(t-h)^2\tilde{f}''_{nh_0}(h_0), \tag{7}$$

where

$$\tilde{f}''_{nh_0}(h_0) = \int K'_{h_0}(h_0 - x)\, d\hat{f}_n(x) = \sum_{\tau_i} K'_{h_0}(h_0 - \tau_i)p_i.$$

Here $h_0 \asymp n^{-1/9}$ and the τ_i are the points of jump of $\hat{f}_n$ with values p_i. Note that $\tilde{f}''_{nh_0}(h_0)$ converges in probability to to $f_0''(0+)$, as $n \to \infty$. We can replace $\tilde{f}''_{nh_0}(h_0)$ by any other consistent estimate of $f''(0+)$.

For $t \in [1-h, 1]$ we define analogously

$$\tilde{f}_{nh}(t) = \tilde{f}_{nh}(1-h) + (t-(1-h))\tilde{f}'_{nh}(1-h) + \frac{1}{2}(t-(1-h))^2\tilde{f}''_{nh_0}(1-h_0). \tag{8}$$

Note that we may lose monotonicity in the boundary intervals $[0, h]$ and $[1-h, 1]$ in this way.

For the NW estimator we use a different boundary correction, replacing the kernel K by a linear combination of the kernels K and $uK(u)$, because of its more complicated expression as a ratio. This type of boundary correction is for example described on pp. 210 and 211 of Groeneboom and Jongbloed (2014).

We have the following asymptotic pointwise result for the SLSE. Note that by going from $\hat{f}_n(t)$ to (3) we improve the rate of convergence $n^{1/3}$ to $n^{2/5}$ and lose the "non-standard asymptotics" behavior of $\hat{f}_n$ (for its cube root n convergence to Chernoff's distribution, see Brunk, 1970, Thm. 5.2, p. 190).

Theorem 1. *Let f_0 be a nondecreasing continuous function on $[0, 1]$. Let $X_1, X_2, \ldots$ be i.i.d. random variables with continuous density g, staying away from zero on $[0, 1]$, and where the derivative g' is continuous and bounded on $(0, 1)$. Furthermore, let $\varepsilon_1, \varepsilon_2 \ldots$ be i.i.d. random variables distributed according to a sub-Gaussian distribution with expectation zero and variance $0 < \sigma_0^2 < \infty$, independent of the X_i's. Then consider Y_i, defined by*

$$Y_i = f_0(X_i) + \varepsilon_i, \qquad i = 1, 2, \ldots$$

Suppose $t \in (0, 1)$ such that f_0 has a strictly positive derivative and a continuous second derivative $f_0''(t) \neq 0$ at t. Then, for the SLSE $\tilde{f}_{nh}$ defined by (3) based on the pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$, and $h \sim cn^{-1/5}$ for $c > 0$,

$$n^{2/5}\left\{\tilde{f}_{nh}(t) - f_0(t)\right\} \xrightarrow{\mathcal{D}} N(\beta, \sigma^2). \tag{9}$$

Here

$$\beta = \frac{1}{2}c^2 f_0''(t) \int u^2 K(u)\, du \quad \text{and} \quad \sigma^2 = \frac{\sigma_0^2}{cg(t)} \int K(u)^2\, du. \tag{10}$$

The asymptotically Mean Squared Error optimal constant c is given by:

$$c = \left\{ \frac{\sigma_0^2}{g(t)} \int K(u)^2\, du \Big/ \left\{ f_0''(t) \int u^2 K(u)\, du \right\}^2 \right\}^{1/5}. \tag{11}$$

Remark 1. For the first function in Example 1, discussed below, the conditions of Theorem 1 are satisfied, and the asymptotically optimal bandwidth is approximately $0.7n^{-1/5}$ (in particular not depending on t). It is seen from Theorem 1 that the smoothed LSE has the same rate of convergence and also the same asymptotic variance as the NW estimator under the usual conditions (see (23) below).

However, the asymptotic bias is different. For the NW estimator, the asymptotic bias up to order h^2 at $t \in (0, 1)$ is given by

$$h^2 \left\{ \frac{1}{2} f_0''(t) + f_0'(t) \frac{g'(t)}{g(t)} \right\} \int u^2 K(u)\, du, \tag{12}$$

while the asymptotic bias for the SMLE is just given by

$$\frac{1}{2} h^2 \frac{1}{2} f_0''(t). \tag{13}$$

In both cases it can happen that the coefficient of h^2 is zero and then on could take a bandwidth h tending to zero at a slower rate than $n^{-1/5}$ to get a better MSE (smaller variance).

In the current status model, to be discussed in Section 5, there is a similar distinction between the SMLE and the maximum smoothed likelihood estimator (MSLE) which have the same asymptotic variance, but the SMLE has asymptotic bias (13) and the MSLE asymptotic bias (12), with f_0 replaced by the distribution function F_0, see Groeneboom et al. (2010). It is clear that the asymptotic biases of the SLSE and the SMLE are simpler that the respective biases of the NW estimator and the MSLE, but it is clear that not one of the situation is necessarily to be preferred over the other.

The convergence result (9) still holds if $f_0''(t) = 0$, but then the optimality result via the constant (11) does no longer hold.

We now give a road map of the proof of Theorem 1. The proof itself is given in Appendix. Since the estimators are based on the LSE $\hat{f}_n$, the proof is totally different from the proofs for the NW estimator, which is a ratio of two sample averages. To prove the result, we introduce methods similar to those used in Groeneboom et al. (2010) for local smooth functionals in the current status model.

The first step is to write

$$\tilde{f}_{nh}(t) - f_0(t) = \int K_h(t - x)\{\hat{f}_n(x) - f_0(x)\}\, dx + \int K_h(t - x) f_0(x)\, dx - f_0(t), \tag{14}$$

and represent the first term at the right-hand side as functional

$$\int \psi_{t,h}(x)\left\{\hat f_n(x) - y\right\}\, dH_0(x,y), \text{ where } \psi_{t,h}(x) = \frac{K_h(t-x)}{g(x)},$$

and H_0 is the distribution function of the pairs (X_i, Y_i). Next, a piecewise constant version $\overline{\psi}_{t,h}$ of $\psi_{t,h}$ is constructed to be able to use the characterization of $\hat f_n$ as a least squares estimator, enabling us to write

$$\int \overline{\psi}_{t,h}(x)\left\{\hat f_n(x) - y\right\}\, dH_0(x,y) = \int \overline{\psi}_{t,h}(x)\left\{\hat f_n(x) - y\right\}\, d\left(H_0 - \mathbb{H}_n\right)(x,y).$$

Here $\mathbb{H}_n$ is the empirical distribution function of the pairs (X_i, Y_i). The latter expression turns out to behave as the empirical integral

$$\int \psi_{t,h}(x)\{f_0(x) - y\}\, d\left(H_0 - \mathbb{H}_n\right)(x,y),$$

for which we have a central limit theorem, after multiplying with $n^{2/5}$ and letting $h \sim cn^{-1/5}$.

The main effort goes into showing that the remainder terms are of lower order. For example, it needs to be shown that

$$\int \left\{\overline{\psi}_{t,h}(x) - \psi_{t,h}(x)\right\}\left\{\hat f_n(x) - f_0(x)\right\} g(x)\, dx = o_p\left(n^{-2/5}\right).$$

To this end we use the Cauchy–Schwarz inequality and the inequality

$$\left|\overline{\psi}_{t,h}(x) - \psi_{t,h}(x)\right| \le Kh^{-2}\left|\hat f_n(x) - f_0(x)\right|,$$

for a $K > 0$, which follows from a judicious choice of $\overline{\psi}_{t,h}$ (see (A2) and lemma A.4 on p. 379 of Groeneboom et al., 2010). We need the existence of f_0' here. This method was also used in Groeneboom and Jongbloed (2014), pp. 290 and 332, and Groeneboom and Hendrickx (2018), p. 154.

We also need L_2 bounds for $\hat f_n - f_0$, restricted to an interval $[a,b] \subset (0,1)$, These follow from the condition that the errors ε_i are sub-Gaussian and L_2-bounds for functions of uniformly bounded variation in chap. 9 of van de Geer (2000). Alternatively, we could use the "switch relation," as used in Groeneboom and Hendrickx (2018) and Sen and Xu (2015). The proof is completed by incorporating the behavior of the bias term in (14),

$$\int K_h(t-x) f_0(x)\, dx - f_0(t) = \frac{1}{2} h^2 f_0''(t) \int u^2 K(u)\, du + o\left(h^2\right).$$

3 | CONFIDENCE INTERVALS BASED ON THE SMOOTHED BOOTSTRAP

In this section, we will create confidence intervals for $f_0(t)$ based on the SLSE. As basis for the intervals, we choose the quantity

$$\tilde f_{nh}(t) - f_0(t), \tag{15}$$

as studied asymptotically in Theorem 1. The distribution of this quantity is approximated by the distribution of a related quantity based on observations $(X_1, Y_1^*), \dots, (X_n, Y_n^*)$ generated by an estimated model, which makes it a bootstrap approach. As an approximate (estimated) model, we choose to generate the Y_i^*-values by taking an *oversmoothed* SLSE, and adding noise to that. More precisely, we take $h_0 \asymp n^{-1/9}$ (we will come back to this choice in Section 4), compute $\tilde{f}_{nh_0}$ and also compute the residuals of the Y_i with respect to this estimate:

$$E_i = Y_i - \tilde{f}_{nh_0}(X_i), \qquad i = 1, \dots, n.$$

Next, we center the E_i by subtracting their mean:

$$\tilde{E}_i = E_i - n^{-1}\sum_{j=1}^{n} E_j, \qquad i = 1, \dots, n. \tag{16}$$

Using the $\tilde{E}_i$, we generate bootstrap samples

$$(X_i, Y_i^*) = \left(X_i, \tilde{f}_{nh_0}(X_i) + E_i^*\right), \qquad i = 1, \dots, n, \tag{17}$$

where the E_i^* are (discretely) uniformly drawn with replacement from the $\tilde{E}_i$, and consider the differences

$$\tilde{f}^*_{nh}(t) - \tilde{f}_{nh_0}(t). \tag{18}$$

Here $\tilde{f}^*_{nh}(t)$ is the estimate of f_0, based on a bootstrap sample, with bandwidth h as in (15). Note that we keep the X_i fixed in the bootstrap samples.

Example 1. Consider the setting used in Chakraborty and Ghosal (2021), who study a Bayesian approach to constructing confidence sets for f_0. Following their choice, we take $f_0(x) = x^2 + x/5$, $g(t) = 1_{[0,1]}(t)$ and independent normal errors ε_i with expectation 0 and variance 0.01. For a sample of size $n = 100$, the NW estimator and the SLSE are shown as blue solid curves in Figure 1. The confidence intervals, of which the construction will be explained below, are shown in Figure 1 at the points $t = 0.01, 0.02, \dots, 0.99$. The coverage is shown in Figure 2a, also for sample size $n = 100$. In Figure 2b we also show the results for the rather different function $f_0(x) = \exp\{4(x - 1/2)\}/\{1 + \exp(4(x - 1/2)\}$, for which the second derivative is not constant. The results for sample size $n = 500$ are given in Figure 3. For generating the confidence intervals and coverage percentages, we use the code in Groeneboom (2021).

The 95% bootstrap confidence intervals are given by

$$\left(\tilde{f}_{nh}(t) - Q^*_{0.975}, \tilde{f}_{nh}(t) - Q^*_{0.025}\right), \tag{19}$$

where $Q^*_{0.025}$ and $Q^*_{0.975}$ are the 2.5th and 97.5th percentiles of 1000 (bootstrap) samples of (18). Note that the percentiles $Q^*_{0.025}$ and $Q^*_{0.975}$ contain an estimate of the asymptotic bias

$$\frac{1}{2}h^2\tilde{f}''_{nh_0}(t) \sim \frac{1}{2}h^2 f_0''(t),$$

(see also Lemma 1 in Section 4) and that therefore the bias of $\tilde{f}_{nh}(t)$ drops out in (19). So we do not need undersmoothing or explicit estimation of the bias in our procedure. The oversmoothing

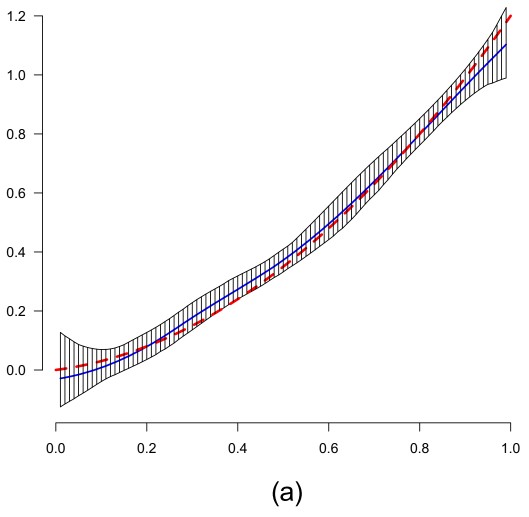

FIGURE 1 (a) The smoothed least squares estimator (SLSE) (blue, solid) and 95% confidence intervals, using the confidence intervals (19). The red dashed curve is f_0. (b) The Nadaraya Watson estimator (blue, solid) and pointwise 95% confidence intervals, for sample size $n = 100$, the dashed red curve is the function f_0. In both cases the bandwidth $h = 0.5n^{-1/5}$ and $h_0 = 0.7n^{-1/9}$.

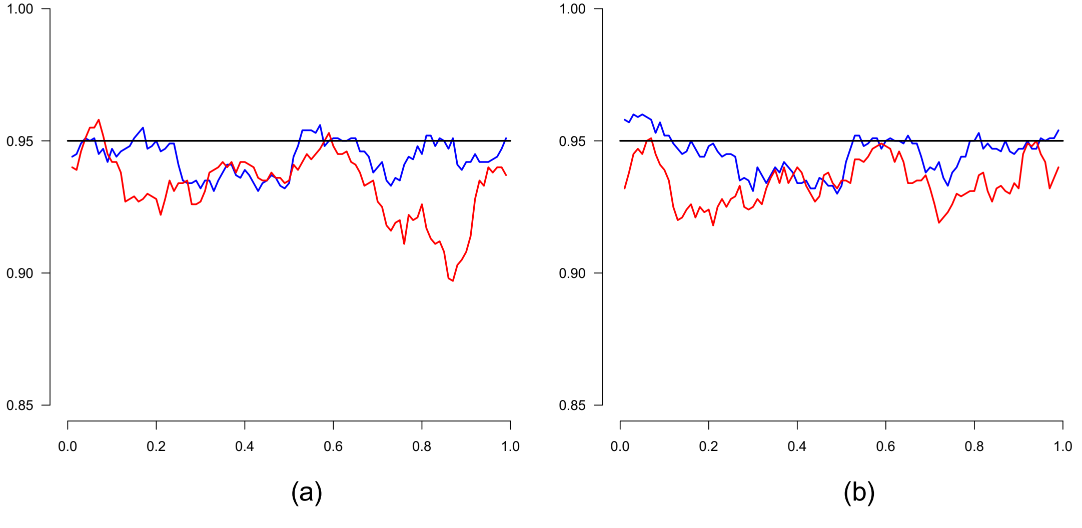

FIGURE 2 Coverage of the confidence intervals, based on the bootstrap described. (a) Fraction of the B experiments where $f_0(t)$ is in the interval (19) for the function $f_0(x) = x^2 + x/5$ and for the smoothed least squares estimator (SLSE) (blue) and the Nadaraya Watson (NW) estimator (red), and (b) fraction of the B experiments where $f_0(t)$ is in the interval (19) for the function $f_0(x) = \exp\{4(x - 1/2)\}/\{1 + \exp(4(x - 1/2)\}$ and for the SLSE (blue) and the NW estimator (red), based on $B = 1000$ samples of size $n = 100$, and $t = 0.01, \ldots, 0.99$. The chosen bandwidths h and h_0 are $h = 0.5n^{-1/5}$ and $h = 0.7n^{-1/9}$.

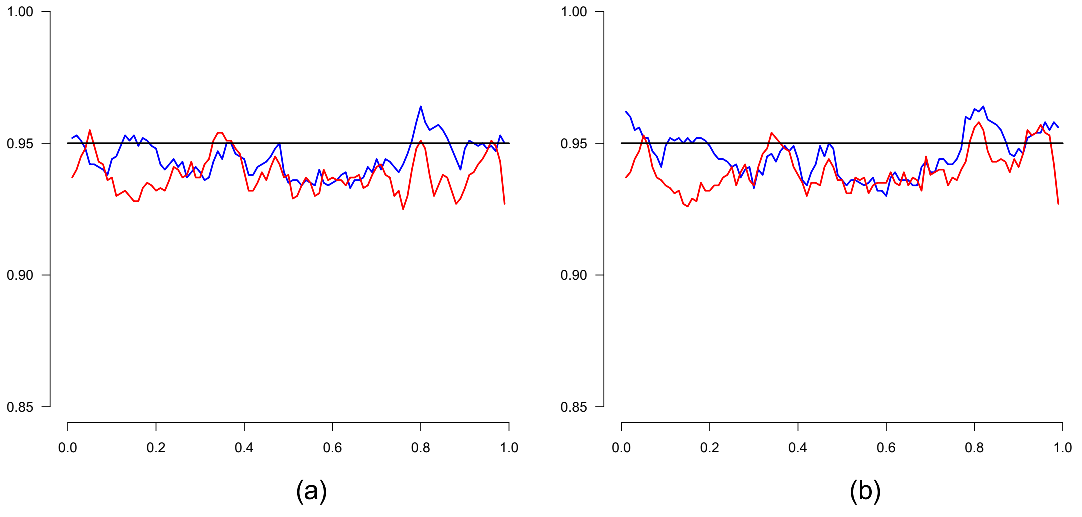

FIGURE 3 Coverage of the confidence intervals, based on the bootstrap described. (a) Fraction of the B experiments where $f_0(t)$ is in the intervals (19) for the function $f_0(x) = x^2 + x/5$ and for the smoothed least squares estimator (SLSE) (blue) and the Nadaraya Watson (NW) estimator (red), and (b) fraction of the B experiments where $f_0(t)$ is in the intervals (19) for the function $f_0(x) = \exp\{4(x-1/2)\}/\{1+\exp(4(x-1/2)\}$ and for the SLSE (blue) and the NW estimator (red), based on $B = 1000$ samples of size $n = 500$, and $t = 0.01, \ldots, 0.99$. The chosen bandwidths h and h_0 are $h = 0.5n^{-1/5}$ and $h = 0.7n^{-1/9}$.

by taking $h_0 \asymp n^{-1/9}$ (or at least a bandwidth tending to zero slower than $n^{-1/6}$) is essential here, though.

Theorem 2 shows that the bootstrap method described will asymptotically give the right coverage, in the sense that after rescaling with $n^{2/5}$ the asymptotic distribution of (18) under the estimated model (17) coincides with the asymptotic distribution of (15) under model (1).

Theorem 2. *Let the conditions of Theorem 1 be satisfied. Moreover, let $h \sim cn^{-1/5}$ and $h_0 \sim c'n^{-1/9}$, for some positive constants c and c'. Then, at $t \in (0,1)$,*

$$n^{2/5}\left\{\tilde{f}^*_{nh}(t) - \tilde{f}_{nh_0}(t)\right\} \xrightarrow{\mathcal{D}} N(\beta, \sigma^2),$$

given $(X_1, Y_1), \ldots, (X_n, Y_n)$, almost surely along sequences $(X_1, Y_1), (X_2, Y_2), \ldots$, where β and σ^2 are defined in (10).

The proof is given in Appendix. It goes through similar steps as the proof of Theorem 1 but is more complicated because we are here in the “bootstrap world” and for example have to use the Lindeberg–Feller version of the central limit theorem to deal (conditionally) with the dependence on the changing regression function $\tilde{f}_{nh_0}$ instead of f_0.

We compare the confidence intervals based on the SLSE with confidence intervals based on the NW estimator. To construct the latter, we define the (empirical) residuals by

$$E_i = Y_i - \tilde{f}^{\mathrm{NW}}_{nh_0}(X_i), \qquad i = 1, \ldots, n,$$

where $\tilde{f}^{\mathrm{NW}}_{nh_0}$ is the NW estimator with bandwidth h_0 (again of order $n^{-1/9}$), leading to the bootstrap quantity

$$\left(\tilde{f}_{nh}^{\rm NW}\right)^*(t) - \tilde{f}_{nh_0}^{\rm NW}(t). \tag{20}$$

Here $\left(\tilde{f}_{nh}^{\rm NW}\right)^*$ is the NW estimator based on (17) with $\tilde{f}_{nh_0}(X_i)$ replaced by $\tilde{f}_{nh_0}^{\rm NW}(X_i)$ and E_i^* sampled with replacement from the residuals $\tilde{E}_i^{\rm NW}$, $i = 1,\dots,n$,

$$\tilde{E}_i^{\rm NW} = E_i^{\rm NW} - \bar{E}^{\rm NW}, \qquad E_i^{\rm NW} = Y_i - \tilde{f}_{nh_0}^{\rm NW}(X_i), \qquad \bar{E}^{\rm NW} = n^{-1}\sum_{i=1}^n E_i^{\rm NW}.$$

In Hall (1992) the variance of the NW estimator, conditionally on $X_1,\dots,X_n$, in the model (1) at t is shown to be equal to

$$\tau_t^2 = \sigma_0^2\beta_t^2, \tag{21}$$

where $\sigma_0^2 = \mathrm{var}(\varepsilon_i)$ and

$$\beta_t^2 = \frac{\sum_{i=1}^n K_h(t-X_i)^2}{\left\{\sum_{i=1}^n K_h(t-X_i)\right\}^2}. \tag{22}$$

In our set-up the parameter β_t is the same in the original sample and in the bootstrap samples, so to estimate τ_t^2 we only need an estimate of σ_0^2.

Denoting the empirical distribution function of the X_i by $\mathbb{G}_n$, note that

$$\frac{1}{n}\sum_{i=1}^n K_h(t-X_i) = \int K_h(t-x)\, d\mathbb{G}_n(x) \to^P g(t),$$

whenever g is continuous at t and $h = h_n$ tends to zero such that $nh \to \infty$. Under the same conditions,

$$\frac{h}{n}\sum_{i=1}^n K_h(t-X_i)^2 = h\int K_h(t-x)^2\, d\mathbb{G}_n(x) \to_P g(t)\int K(u)^2\, du.$$

Therefore, as $n \to \infty$, the variance of $\tilde{f}_{nh}^{\rm NW}(t)$ behaves like

$$\sigma_0^2\beta_t^2 = \sigma_0^2\frac{\frac{1}{n}\sum_{i=1}^n K_h(t-X_i)^2}{n\left\{\frac{1}{n}\sum_{i=1}^n K_h(t-X_i)\right\}^2} \sim \frac{\sigma_0^2}{nhg(t)}\int K(u)^2\, du = \frac{\sigma_0^2}{cn^{4/5}g(t)}\int K(u)^2\, du, \tag{23}$$

where we use $h = cn^{-1/5}$ in the final step. In view of Theorem 1, it follows that the SLSE and the NW estimator (both rescaled and centered) have the same asymptotic variance.

A well known approach to improve the coverage of bootstrap confidence sets is Studentization. For the SLSE in the setting of this paper, this would mean that instead of using difference (15) as basis for the bootstrap, one would use a rescaled difference such that asymptotically the variance does not depend on unknown quantities anymore. In view of Theorem 1, this means

$$\left\{\tilde{f}_{nh}(t) - f_0(t)\right\}/\hat{\sigma}_{n,0}, \tag{24}$$

where the estimate of the variance σ_0^2, $\hat\sigma_{n,0}^2$, is given by

$$\hat\sigma_{n,0}^2 = n^{-1}\sum_{i=1}^{n}\tilde E_i^2. \tag{25}$$

The distribution of (24) under the true model is then approximated by the distribution of

$$\left\{\tilde f_{nh}^*(t) - \tilde f_{nh_0}(t)\right\}/\hat\sigma_{n,0}^*, \quad \text{with} \quad \left(\hat\sigma_{n,0}^*\right)^2 = n^{-1}\sum_{i=1}^{n}\left(\tilde E_i^* - \bar E^*\right)^2 \quad \text{and} \quad \bar E^* = n^{-1}\sum_{i=1}^{n}\tilde E_i^*, \tag{26}$$

where $\left(\hat\sigma_{n,0}^*\right)^2$ is the variance estimate based on a bootstrap sample, and where again $h_0 \asymp n^{-1/9}$.

A 95% confidence interval for $f_0(t)$ can then be based on the 2.5th and 97.5th percentiles $Q_{0.025}^*$ and $Q_{0.975}^*$ of 1000 bootstrap draws of (26). It is then given by

$$\left(\tilde f_{nh}(t) - Q_{0.975}^*\hat\sigma_{n,0}, \hat f_{nh}(t) - Q_{0.025}^*\hat\sigma_{n,0}\right), \tag{27}$$

where $\hat\sigma_{n,0}^2$ is defined by (25). The comparison with the bootstrap intervals based on the SLSE without Studentization is shown in Figure 4a.

For the NW estimator, an estimate of σ_0^2 is given on p. 226 of Hall (1992) (but note the typo w.r.t. the index j in Hall, 1992). We take the definition from Hall et al. (1990) and define

$$\left(\hat\sigma_n^{\mathrm{NW}}\right)^2 = (n-m)^{-1}\sum_{i=1}^{n-2}\left(\sum_{j=0}^{2} d_j Y_{i+j}\right)^2, \tag{28}$$

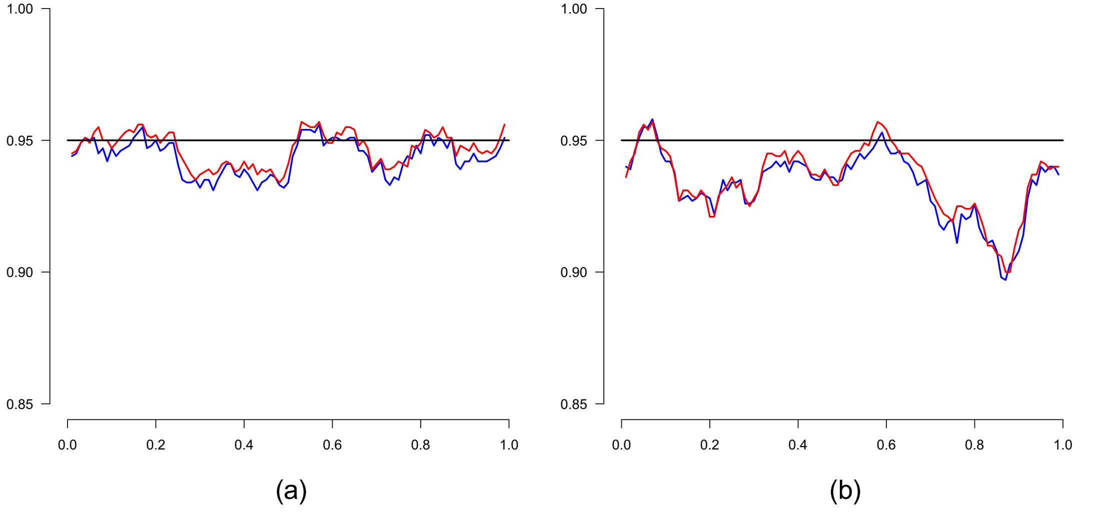

FIGURE 4 Coverage of the confidence intervals, Studentized and non-Studentized, based on the bootstrap described. Fraction of the B experiments where $f_0(t)$ is in the intervals (19) for the function $f_0(x) = x^2 + x/5$ for (a) the Studentized smoothed least squares estimator (SLSE) (red) and the non-Studentized SLSE (blue), and (b) the Studentized Nadaraya Watson (NW) estimator (red) and the non-Studentized NW estimator (blue). The figures are based on $B = 1000$ samples of size $n = 100$, and $t = 0.01, \dots, 0.99$. The chosen bandwidths h and h_0 are $h = 0.5n^{-1/5}$ and $h_0 = 0.7n^{-1/9}$.

for the variance of the ε_i, where $m = 2$ and

$$(d_0, d_1, d_2) = \left(\frac{1}{4}(\sqrt{5}+1), -\frac{1}{2}, -\frac{1}{4}(\sqrt{5}-1)\right),$$

(as recommended in Hall et al., 1990).

If we now compare the non-Studentized and Studentized confidence intervals based on the NW estimator, constructed in the same way as in the case of the SLSE, we get Figure 4b Here the non-Studentized are based on the differences (20) and the Studentized intervals on the differences

$$\left\{\left(\tilde{f}_{nh}^{\mathrm{NW}}\right)^*(t) - \tilde{f}_{nh_0}^{\mathrm{NW}}(t)\right\} / \left(\hat{\sigma}_n^{\mathrm{NW}}\right)^*, \tag{29}$$

where $\left(\hat{\sigma}_n^{\mathrm{NW}}\right)^*$ is the estimate (28) for the bootstrap samples. It is seen that in both cases there is not a great improvement.

For the NW estimator, one can also use the estimate of σ_0, based on the residuals, the type of estimate of σ_0 we used for the SLSE. In this case we get a bit more improvement for the NW estimator, see Figure 5. We still do not understand this phenomenon, based on the different ways of estimating σ_0 for the NW estimator, however.

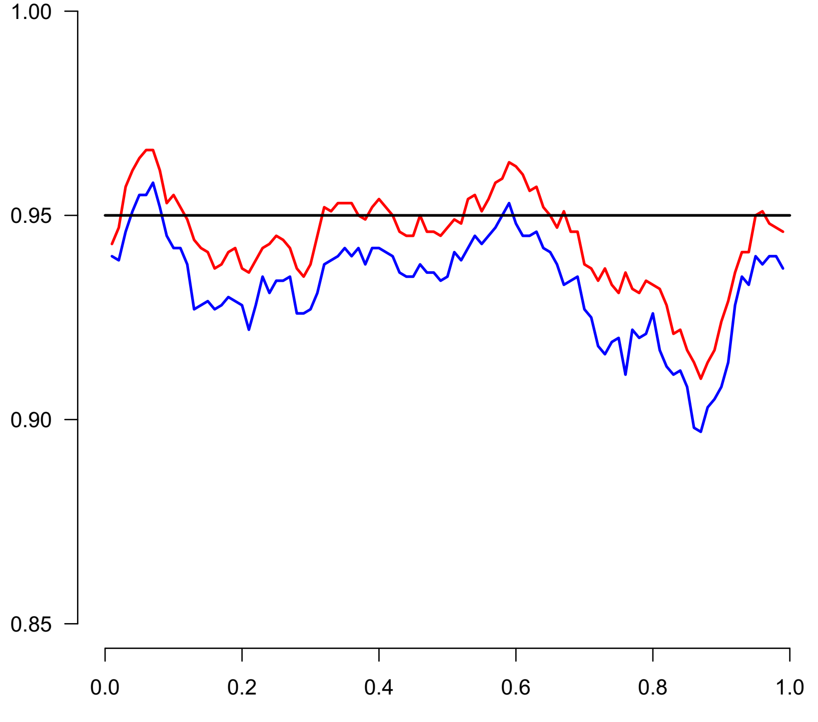

FIGURE 5 The Studentized Nadaraya Watson (NW) estimator (red), with variance estimated from the residuals, just as for the smoothed least squares estimator (SLSE), and the non-Studentized NW estimator (blue). The figures are based on $B = 1000$ samples of size $n = 100$, and $t = 0.01, \ldots, 0.99$. The chosen bandwidths h and h_0 are $h = 0.5n^{-1/5}$ and $h_0 = 0.7n^{-1/9}$.

4 | BANDWIDTH SELECTION

As seen in the previous section, for the smooth bootstrap to work, h_0 can be chosen as $c'n^{-1/9}$ and the precise value of $c' > 0$ is not crucial for this. However, the value of c in the bandwidth choice $h = cn^{-1/5}$ is important. It can be chosen such that the asymptotic MSE is minimal. However, this optimal choice depends on unknown quantities, as seen in (11). In this section, we propose a bootstrap method to find an approximately MSE optimal bandwidth for estimating $f_0(t)$ at a point $t \in (0, 1)$. The MSE to have minimized as a function of h is given by:

$$MSE_h(t) = \mathbb{E}\left\{ \left\{\tilde{f}_{nh}(t) - f_0(t)\right\}^2 \middle| X_1,\dots,X_n \right\}. \tag{30}$$

Of course, f_0 being unknown, this quantity cannot be computed as function of h. However, the analogous bootstrap quantity (again using oversmoothing, in the sense that $h_0 \asymp n^{-1/9}$) is given by,

$$\text{MSE}_h^*(t) = \mathbb{E}\left\{ \left\{\tilde{f}^*_{nh}(t) - \tilde{f}_{nh_0}(t)\right\}^2 \middle| (X_1,Y_1),\dots,(X_n,Y_n)\right\}, \tag{31}$$

where h_0 is called a “pilot” bandwidth. We shall show that (31) is asymptotically independent of the constant in the pilot bandwidth h_0 if we take $h_0 \asymp n^{-1/9}$.

We have:

$$\begin{aligned} MSE_h^*(t) = \mathbb{E}&\left\{ \left\{ \int K_h(t-x)\left\{\hat{f}^*_n(x) - \tilde{f}_{nh_0}(x)\right\}\,dx \right\}^2 \middle| (X_1,Y_1),\dots,(X_n,Y_n)\right\} \\ &+ \left\{ \int K_h(t-x)\,\tilde{f}_{nh_0}(x)\,dx - \tilde{f}_{nh_0}(t)\right\}^2. \end{aligned} \tag{32}$$

For the second term on the right we get:

$$\int K_h(t-x)\,\tilde{f}_{nh_0}(x)\,dx - \tilde{f}_{nh_0}(t) = \frac{1}{2}h^2\tilde{f}''_{nh_0}(t)\int u^2K(u)\,du + o_p\left(h^2\right),$$

so

$$\left\{ \int K_h(t-x)\,\tilde{f}_{nh_0}(x)\,dx - \tilde{f}_{nh_0}(t)\right\}^2 = \frac{1}{4}h^4\tilde{f}''_{nh_0}(t)^2\left\{\int u^2K(u)\,du\right\}^2 + o_p\left(h^4\right).$$

We have the following result.

Lemma 1. *Let the conditions of Theorem 1 be satisfied. Moreover, let $h_0 = h_{n,0} \sim c_0n^{-1/9}$, as $n \to \infty$. Then*

$$\tilde{f}''_{nh_0}(t) \xrightarrow{p} f''_0(t), \qquad n \to \infty.$$

Remark 2. Note that this convergence result does not hold if the pilot bandwidth h_0 is of order $n^{-1/5}$. For this reason the method suggested in Sen and Xu (2015), where the pilot bandwidth is chosen of order $n^{-1/5}$ will not work. Another way out is to use subsampling, as used in Groeneboom and Hendrickx (2018), but choosing the right subsample size is a rather hard problem.

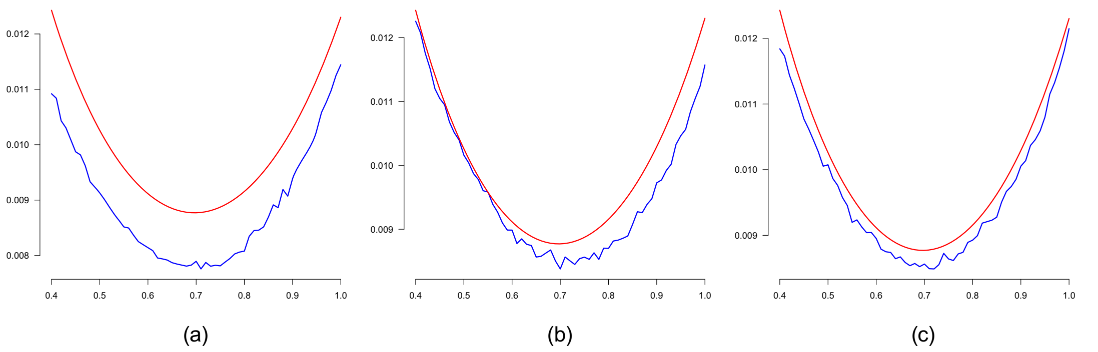

FIGURE 6 Estimated mean integrated squared error (MISE), for samples of sizes $n = 500, 1000$, and 5000 and $h = cn^{-1/5}$, $c = 0.4, 0.41, \ldots, 1$ and asymptotic MISE with $g(x) = 1_{[0,1]}(x)$ and $f_0(x) = x^2 + x/5$ on $[0, 1]$. The blue curve is based on (34), and the red curve on (35), $B = 10{,}000$, where $t_i = 0.20, 0.21, \ldots, 0.80$, and $m = 60$. (a) Estimates for $n = 500$, (b) Estimates for $n = 1000$, (c) Estimates for $n = 5000$. The asymptotically minimizing c in $h = cn^{-1/5}$ is $c_{\min} \approx 0.697601$.

We can choose to let h_0 tend to zero at any rate slower than $n^{-1/6}$ (see the proof of the Lemma, where it is shown that one needs $h_0^{-4}n^{-2/3} \to 0$ and $h_0^{-3/2}n^{-1/2} \to 0$). We choose rate $h_0 \asymp n^{-1/9}$ because this is the optimal choice for estimating the second derivative. See also Figure 10c for the disastrous effect of choosing h_0 equal to $h \asymp n^{-1/5}$, which is the bandwidth for estimating the distribution function F_0 in the original sample in the current status model in Section 5.

The proof of Lemma 1 is given in Appendix. The lemma suggests, as in Hazelton (1996), to take the pilot bandwidth $h_0 = c_0 n^{-1/9}$ for some $c_0 > 0$, taking the optimal order for a bandwidth for estimating the second derivative f_0'' in the case that the fourth derivative $f_0^{(4)}(t)$ exists and is not equal to zero. Note that for our example function $f_0(x) = x^2 + x/5$ we have $f_0^{(4)}(t) = 0$, so in this case we cannot apply the optimality criterion. The most important fact is, however, that h_0 has to tend slower to zero than $n^{-1/6}$, since otherwise the variance of $\tilde f_{nh_0}''(t)$ does not tend to zero.

For the first term on the right of (32) we get:

$$\mathbb{E}\left\{\left\{\int K_h(t-x)\left\{\hat f_n^*(x) - \tilde f_{nh_0}(x)\right\}\,dx\right\}^2 \middle| (X_1, Y_1), \ldots, (X_n, Y_n)\right\} \sim \frac{S_n}{nh} + o_p\left(\frac{1}{nh}\right),$$

where

$$S_n \xrightarrow{p} \frac{\sigma_0^2}{g(t)} \int K(u)^2\,du,$$

if $h \sim cn^{-1/5}$ (see (A14) and the argument using the Lindeberg–Feller central limit theorem part of the proof of Theorem 2).

So, asymptotically, the bandwidth h, minimizing (32) minimizes

$$\frac{\sigma_0^2}{g(t)nh}\int K(u)^2\,du + \frac{1}{4}h^4 f_0''(t)^2\left\{\int u^2 K(u)\,du\right\}^2.$$

The minimization of (30) leads asymptotically to the same minimization over h.

Instead of minimizing (31), we minimize a Monte Carlo approximation of (31):

$$B^{-1}\sum_{i=1}^{B}\left\{\tilde f_{nh}^{*,i}(t)-\tilde f_{nh_0}(t)\right\}^2, \tag{33}$$

where the $\tilde f_{nh}^{*,i}$, $i=1,\dots,B$ are the estimates in B bootstrap samples.

As we are choosing a fixed bandwidth over the range of t-values, it is then natural to aim at minimizing the Mean Integrated Squared Error (MISE) as a function of h. The asymptotically MISE optimal bandwidth can also be approximated by a smoothed bootstrap experiment, in which case one replaces (33) by

$$\widehat{\text{MISE}}_c^* = n^{4/5}B^{-1}\sum_{j=1}^{B}\sum_{i=1}^{m}\left\{\tilde f_{n,cn^{-1/5}}^{*,j}(t_i)-\tilde f_{nh_0}(t_i)\right\}^2\Delta_i, \qquad \Delta_i = t_i - t_{i-1}. \tag{34}$$

for a grid of points $0\le a=t_0<t_1<t_2<\cdots<t_m=b\le 1$. The latter global minimization produced Figure 6, where we took $a=0.21$ and $b=0.8$. We compare with the plot of the asymptotic MISE as a function of c:

$$\text{AsMISE}_c = \frac{\sigma_0^2}{c}\int K(u)^2\,du\int_{t=a}^{b}\frac{1}{g(t)}\,dt + \frac14 c^4\left\{\int u^2K(u)\,du\right\}^2\int_{t=a}^{b} f_0''(t)^2\,dt. \tag{35}$$

In this case we consider

$$\mathbb{E}\left\{\int_a^b\left\{\tilde f_{nh}^*(t)-\tilde f_{nh_0}(t)\right\}^2 dt\,\Big|\,(X_1,Y_1),\dots,(X_n,Y_n)\right\}, \tag{36}$$

where h_0 is the pilot bandwidth. We have:

$$\begin{aligned}
&\mathbb{E}\left\{\int_a^b\left\{\tilde f_{nh}^*(t)-\tilde f_{nh_0}(t)\right\}^2 dt\,\Big|\,(X_1,Y_1),\dots,(X_n,Y_n)\right\}\\
&\sim \mathbb{E}\left\{\int_{t=a}^b\left\{\int K_h(t-x)\left\{\tilde f_n^*(x)-\tilde f_{nh_0}(x)\right\}dx\right\}^2 dt\,\Big|\,(X_1,Y_1),\dots,(X_n,Y_n)\right\}\\
&+\int_{t=a}^b\left\{\int K_h(t-x)\,\tilde f_{nh_0}(x)\,dx-\tilde f_{nh_0}(t)\right\}^2 dt.
\end{aligned} \tag{37}$$

For the second term on the right we get:

$$\int_{t=a}^b\left\{\int K_h(t-x)\,\tilde f_{nh_0}(x)\,dx-\tilde f_{nh_0}(t)\right\}^2 dt \sim \frac14 h^4\int_{t=a}^b \tilde f_{nh_0}''(t)^2\,dt\left\{\int u^2K(u)\,du\right\}^2 dt + o_p\left(h^4\right).$$

Since we want $\tilde f_{nh_0}''$ to be as close as possible to f_0'', we suggest to minimize

$$\int_{t=a}^b\left\{\tilde f_{nh_0}''(t)-f_0''(t)\right\}^2 dt, \tag{38}$$

over h_0, which is a direct generalization of the locally optimal choice of h_0.

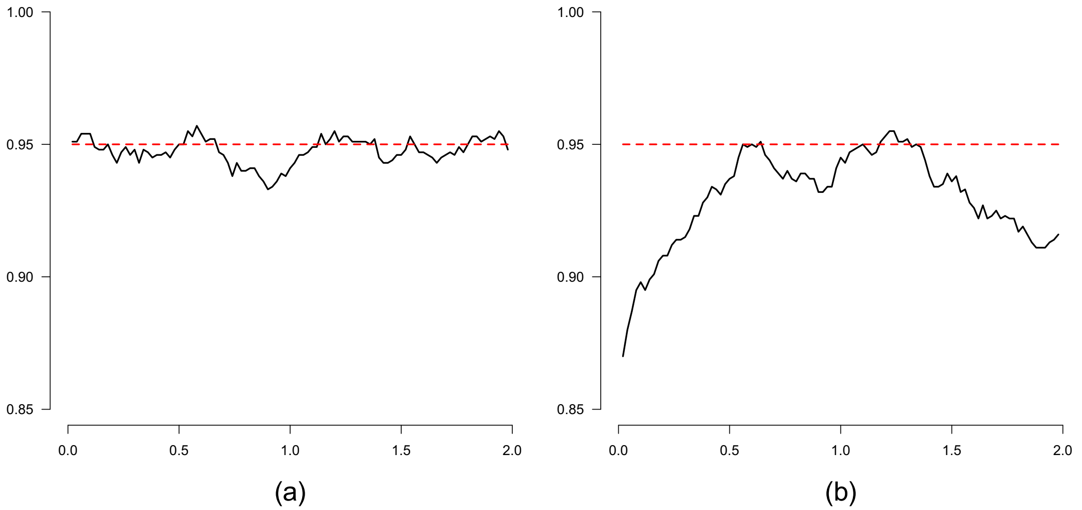

FIGURE 7 Coverage of the confidence intervals: (a) $n^{2/5}$-consistent interval, based on smoothed maximum likelihood estimator (SMLE) with Studentization, (b) $n^{2/5}$-consistent interval, based on SMLE without Studentization. Sample size is $n = 500$.

We get:

$$\tilde{f}''_{nh_0}(t) - f_0''(t) = \int K''_{h_0}(t-x)\{\hat{f}_n(x) - f_0(x)\}\,dx + \int K''_{h_0}(t-x)f_0(x)\,dx - f_0''(t),$$

and

$$\int K''_{h_0}(t-x)f_0(x)\,dx - f_0''(t) = \int K_{h_0}(t-x)f_0''(x)\,dx - f_0''(t)$$
$$\int K_{h_0}(t-x)\left\{f_0''(x) - f_0''(t)\right\}\,dx = \frac{1}{2}h_0^2 f_0^{(4)}(t)\int u^2K(u)\,du + o\left(h_0^2\right),$$

provided a bounded fourth derivative $f_0^{(4)}(t)$ exists.

This yields:

$$\int_{t=a}^{b}\{\tilde{f}''_{nh_0}(t) - f_0''(t)\}^2\,dt \sim \frac{\sigma_0^2}{nh_0^5}\int K''(u)^2\,du\int_{t=a}^{b}\frac{1}{g(t)}\,dt$$
$$+\frac{1}{4}h_0^4\left\{\int u^2K(u)\,du\right\}^2\int_{t=a}^{b}\left(f_0^{(4)}(t)\right)^2\,dt,$$

and minimizing this as a function of h_0 gives again $h_0 \asymp n^{-1/9}$ if $\int_{t=a}^{b}\left(f_0^{(4)}(t)\right)^2\,dt \neq 0$.

5 | CONFIDENCE INTERVALS FOR THE CURRENT STATUS MODEL

We now turn to confidence intervals for the current status model. In this case there is a very large choice of cube root n consistent confidence intervals. There are the Banerjee–Wellner confidence intervals, based on likelihood ratio tests (see e.g., Banerjee & Wellner, 2005), the Sen–Xu confidence intervals, based on SMLEs (see Sen & Xu, 2015), and bootstrap intervals, based on subsampling and, conceivably, the numerical bootstrap (see Hong & Li, 2020). Still other options are discussed in Banerjee and Wellner (2005).

Confidence intervals which are $n^{2/5}$-consistent are discussed in Groeneboom and Hendrickx (2017) and Groeneboom and Hendrickx (2018). For the $n^{2/5}$-consistent intervals the variance and the squared bias are of the same order $n^{-4/5}$, in contrast with the cube root n consistent intervals, where the bias vanishes in comparison with the variance. This entails that dealing with the bias is the big issue for the $n^{2/5}$-consistent intervals.

Analogously to Theorem 1 we have the following result (see thm. 4.2 in Groeneboom et al., 2010 and thm. 11.4 in Groeneboom & Jongbloed, 2014).

Theorem 3. *Let the distribution corresponding to F_0 have support $[0, M]$ and let F_0 have a density f_0 staying away from zero on $(0, M)$. Furthermore, let G have a density g with a support that contains $[0, M]$ and let g stay away from zero on $[0, M]$, with a bounded derivative g'. Finally, let t be an interior point of $[0, M]$ such that f_0 has a continuous derivative f_0' at t. Then, if $h \sim cn^{-1/5}$ and the SMLE $\hat{F}_{nh}$ be defined by*

$$\hat{F}_{nh}(t) = \int IK_h(t - x)\, d\hat{F}_n(x), \qquad IK_h(x) = IK(x/h), \tag{39}$$

where the integrated kernel IK is defined by (6) and $\hat{F}_n$ is the MLE. Then

$$n^{2/5}\left\{\tilde{F}_{nh}(t) - F_0(t)\right\} \xrightarrow{\mathcal{D}} N\left(\mu, \sigma^2\right),$$

where

$$\mu = \frac{1}{2}c^2 f_0'(t) \int u^2 K(u)\, du \quad \text{and} \quad \sigma^2 = \frac{F_0(t)\{1 - F_0(t)\}}{cg(t)} \int K(u)^2\, du.$$

Remark 3. In thm. 4.2 on p. 365 of Groeneboom et al. (2010) there is the extra condition that $f_0'(t) \neq 0$. This condition is not needed for the validity of Theorem 3, but only to ensure that a bandwidth of order $n^{-1/5}$ is the optimal choice. If $f_0'(t) = 0$, the squared bias vanishes with respect to the variance and in that situation one can choose a larger bandwidth to obtain a faster convergence than order $n^{-2/5}$. For more details, see Groeneboom et al. (2010).

Under the conditions of this theorem, we can construct confidence intervals for F_0 in the following way. From the original sample we obtain a bootstrap sample $(T_1, \Delta_1^*), \dots, (T_n, \Delta_n^*)$ by keeping the original T_i fixed and by resampling the indicators Δ_i^* from a Bernoulli distribution with probability $\tilde{F}_{nh_0}(T_i)$, where $h_0 \sim cn^{-1/9}$. A crucial difference w.r.t. the approach in Groeneboom and Hendrickx (2018) is that we base the Bernoulli probabilities on an oversmoothed estimate $\tilde{F}_{nh_0}(T_i)$ instead of an estimate using the bandwidth of order $n^{-1/5}$.

Next we compute the SMLE $\tilde{f}_{nh}^*(t)$ in the (smoothed) bootstrap samples with the usual bandwidth h of order $n^{-1/5}$ and compare this with the estimate $\tilde{F}_{nh_0}(t)$. The following bootstrap confidence interval is then defined:

$$\left[\tilde{F}_{nh}(t) - \hat{\sigma}_{nh}(t) U_{1-\alpha/2}^*(t), \tilde{F}_{nh}(t) - \hat{\sigma}_{nh}(t) U_{\alpha/2}^*(t)\right], \tag{40}$$

where $U_\alpha^*(t)$ is the αth quantile of

$$\left\{\tilde{f}^*_{nh}(t) - \tilde{F}_{nh_0}(t)\right\} / \hat{\sigma}^*_{nh}(t),$$

and

$$\sigma_{nh}(t)^2 = \tilde{F}_{nh}(t)\{1 - \tilde{F}_{nh}(t)\}, \qquad \sigma^*_{nh}(t)^2 = \tilde{f}^*_{nh}(t)\{1 - \tilde{f}^*_{nh}(t)\}.$$

The bootstrap value $\{\tilde{f}^*_{nh}(t) - \tilde{F}_{nh_0}(t)\}/\hat{\sigma}^*_{nh}(t)$ has, conditionally on $(T_1, \Delta_1), \ldots, (T_n, \Delta_n)$, the same limit behavior as $n^{2/5}\{\tilde{F}_{nh}(t) - F_0(t)\}/\hat{\sigma}_n(t)$, so the bias drops out in (40), just as in the same type of intervals for the homoscedastic regression in (18). Since we keep the observation times T_i fixed, the density g of the T_i acts in the same way on the original sample and on the bootstrap sample, but it helps to improve the intervals by a kind of Studentization by dividing by estimates of $\{F_0(t)\{1 - F_0(t)\}\}^{1/2}$. A comparison between the "non-Studentized" intervals, based on $\tilde{f}^*_{nh}(t) - \tilde{F}_{nh_0}(t)$ and the "Studentized" intervals, based on $\{\tilde{f}^*_{nh}(t) - \tilde{F}_{nh_0}(t)\}/\hat{\sigma}^*_{nh}(t)$ is shown in Figure 9.

It is clear that the intervals, exhibited in parts (a) of Figures 7 and 8, are much better than the intervals in parts (b) and (c), if the conditions of Theorem 3 are satisfied. There is also a big improvement on the intervals, proposed in Groeneboom and Hendrickx (2018). Although the latter intervals were constructed in a similar way, there were three major differences:

1. The bootstrap samples in in Groeneboom and Hendrickx (2018) were generated by the same estimate $\tilde{F}_{nh}$ of the distribution function as used in the original sample, with a bandwidth of order $n^{-1/5}$.
2. The difference $\tilde{f}^*_{nh}(t) - \tilde{F}_{nh} * \tilde{F}_{nh}(t)$ was used in Groeneboom and Hendrickx (2018) in the centering of $\tilde{f}^*_{nh}(t)$ to remove the bias, with the consequence that we had to deal with the bias in the resulting bootstrap confidence intervals around $\tilde{F}_{nh}(t)$. Getting a good estimate of the bias is extremely difficult, and it is amazing how well the oversmoothing in the resampling, using $h_0 \asymp n^{-1/9}$, and using the differences $\tilde{f}^*_{nh}(t) - \tilde{F}_{nh_0}(t)$ (where we implicitly estimate the bias in the bootstrap samples) deals with this problem.

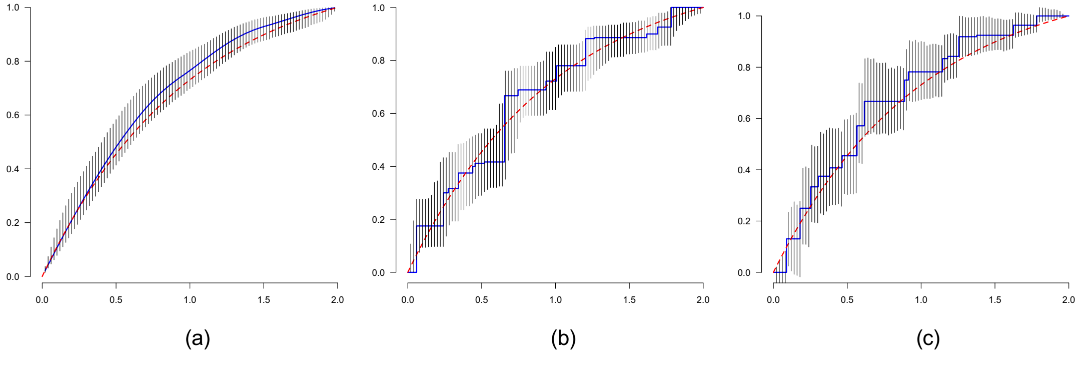

FIGURE 8 Confidence intervals, based on the bootstrap described: (a) $n^{2/5}$-consistent interval, based on smoothed maximum likelihood estimator (SMLE), (b) Banerjee–Wellner l, based on SMLE, (b) Banerjee–Wellner $n^{1/3}$-consistent intervals, based on LR test and asymptotic critical value, (c) Sen–Xu $n^{1/3}$-consistent intervals based on maximum likelihood estimator (MLE) and SMLE. Red dashed curve: real (truncated on $[0, 2]$) exponential distribution function, blue curve: SMLE for (a) and MLE for (b) and (c). Sample size is $n = 500$.

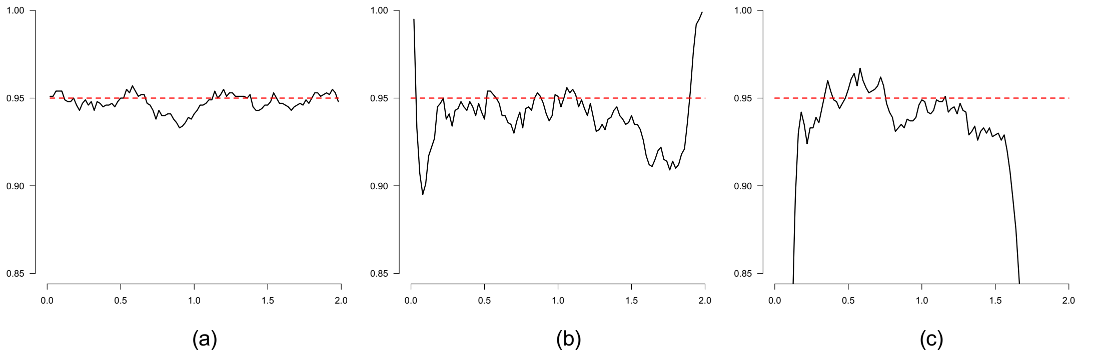

FIGURE 9 Coverage of the confidence intervals: (a) $n^{2/5}$-consistent interval, based on smoothed maximum likelihood estimator (SMLE), (b) Banerjee–Wellner intervals, based on LR test and asymptotic critical value, (c) Sen–Xu $n^{1/3}$-consistent intervals based on maximum likelihood estimator and SMLE. Sample size is $n = 500$.

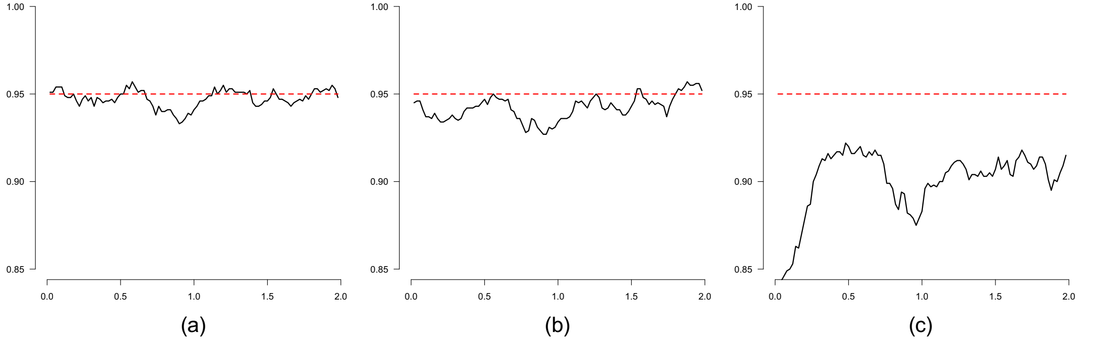

FIGURE 10 Coverage of the confidence intervals: (a) $n^{2/5}$-consistent interval, based on smoothed maximum likelihood estimator (SMLE) with Studentization and second bandwidth $h_0 = 1.9n^{-1/9}$, (b) $n^{2/5}$-consistent interval, based on SMLE with Studentization and second bandwidth $h_0 = 1.5n^{-1/9}$. (c) interval, based on SMLE with Studentization and second bandwidth $h_0 = h = 1.5n^{-1/5}$. Sample size is $n = 500$.

3. We used in the present paper a very simple type of "Studentization," where we divide in the bootstrap samples by $\{\tilde{f}^*_{nh}(t)\{1-\tilde{f}^*_{nh}(t)\}\}^{1/2}$ instead of the $S^*_{nh}(t)^{1/2}$ on p. 141 of Groeneboom and Hendrickx (2018).

We used the same boundary correction as in sec. 2.4 of Groeneboom and Hendrickx (2018), however.

For the SMLE $\tilde{F}_{nh}$ in the Sen–Xu intervals $\hat{f}^*_n(t) - \tilde{F}_{nh}(t)$ we chose $h = 2n^{-1/5}$. For the SMLE in the confidence intervals, resulting in parts (a) of Figures 7 and 8 we chose bandwidths $h = 1.5n^{-1/5}$ for the SMLE around which the confidence intervals are formed and $h_0 = 1.9n^{-1/9}$ for the bandwidth of the SMLE generating and centering the bootstrap intervals $\tilde{f}^*_{nh}(t) - \tilde{F}_{nh_0}(t)$. Note that h_0 is used in generating the bootstrap intervals and also for the centering function $\tilde{F}_{nh_0}$ in these intervals and this role is taken by $\tilde{F}_{nh}$ with $h \asymp n^{-1/5}$ in the Sen–Xu intervals.

So in both cases the generation of the bootstrap samples and the centering of the bootstrap intervals is accomplished by a function which has greater smoothness than the actual estimator itself in order to catch the bias appropriately. For the intervals in parts (a) of the figures it is absolutely necessary to use oversmoothing and a bandwidth h_0 that converges at a slower rate than $n^{-1/6}$, since otherwise the bias is not estimated correctly (in contrast with the suggestion in Sen & Xu, 2015). On the other hand the constant c_0 chosen in $h_0 = c_0 n^{-1/9}$ is not crucial for the behavior of the confidence intervals. Figure 10 shows the effect of taking $c_0 = 1.5$ instead of $c_0 = 1.9$. If $h_0 = h = 0.5n^{-1/5}$, however, the intervals are totally off. The essential feature of this procedure is that the second derivative is estimated consistently, and converges to zero at a slower rate than $n^{-1/6}$, but the constant c_0 does not show up in the asymptotic distribution of $\tilde{F}_{nh}$. All procedures (also the Banerjee–Wellner and Sen–Xu intervals) can be produced by running the R scripts, given on Groeneboom (2023).

One can also construct cube root n consistent confidence intervals different from the Banerjee–Wellner or Sen–Xu intervals, as is shown in sec. 4 of Groeneboom and Jongbloed (2023). For reasons of space we will not elaborate on that here.

6 | LAKE MENDOTA: YEARLY NUMBER OF DAYS FROZEN

As a real data application, we give confidence intervals for the Lake Mendota data, analyzed in Groeneboom and Jongbloed (2014), sec. 1.1. For 157 consecutive years, starting in 1854, the number of days that the lake was frozen was recorded. The idea is that in the presence of global warming, the number of days that the lake is frozen will show a downward trend over the years. It is the first example in Barlow et al. (1972).

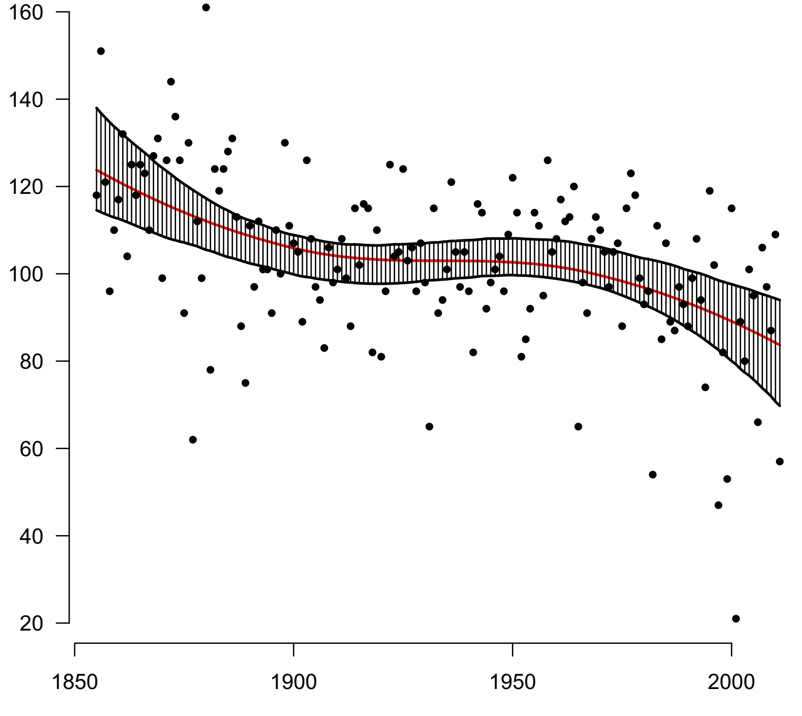

FIGURE 11 95% confidence intervals for the regression function for the Mendota data. The red curve is the SLSE, with the bandwidth chosen by the method based on the MISE-approximation (34).

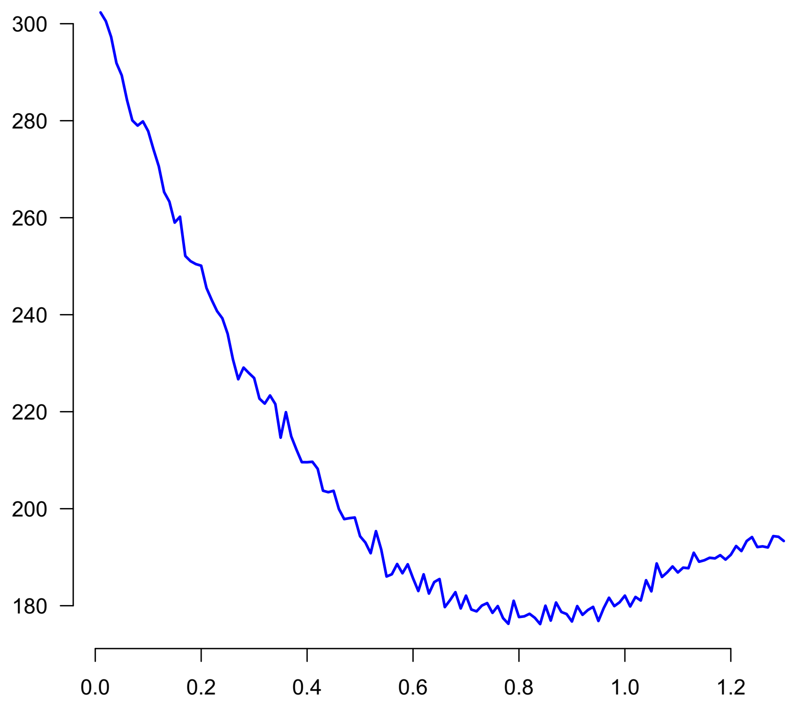

FIGURE 12 The MISE-approximation (34), as a function of c for the transformed Mendota data.

To apply the methods of the preceding sections, we first rescale the x-coordinates to $[0, 1]$ by making the transformation

$$X_i := (X_i - 1853)/158 = i/158,$$

and next by letting $Y_i := Y_{n-i+1}$. If there is a downward trend in the Y_i, there will be an upward trend in the (old) Y_{n-i+1}, where $n = 157$, and so we can apply the theory of the preceding sections to the new (X_i, Y_i).

As before (in the examples) we take the pilot bandwidth $h_0 = 0.7n^{-1/9}$, and used the bandwidth choice of Section 4, based on (34), which gave $c = 0.84$ for the optimal bandwidth $h = cn^{-1/5} \approx 0.30556$. The isotonic confidence intervals for this choice of h are shown in Figure 11. The bootstrap approximation of the MISE as a function of the constant c in the bandwidth choice $h = cn^{-1/5}$ is shown in Figure 12.

7 | CONCLUSION

We introduce the SLSE

$$\tilde{f}_{nh}(t) = \int K_h(t-x)\hat{f}_n(x)\,dx, \tag{41}$$

in the monotone regression problem, where $\hat{f}_n$ is the monotone nonparametric least squares estimator, where K is a smooth symmetric kernel with support $[-1, 1]$ and $K_h = h^{-1}K(\cdot/h)$. In contrast with $\hat{f}_n$, the smooth estimator converges at rate $n^{2/5}$, under the conditions of our Theorem 1; the monotone nonparametric regression estimator $\hat{f}_n$ only converges at rate $n^{1/3}$ in these circumstances.

We use the SLSE to construct bootstrap confidence intervals, based on sampling with replacement residuals with respect to an oversmoothed estimator of type (41). The oversmoothing has the effect that the bias is estimated correctly (which is not the case if one uses residuals w.r.t. an estimate, based on a bandwidth of order $n^{-1/5}$) and the bias drops out in the final confidence interval. The idea goes back to similar methods used for NW estimates in Härdle and Marron (1991).

We compare the monotone estimates with the NW estimates, suggesting that the monotone estimates are somewhat more stable if the underlying regression function is monotone. The method extends the construction of confidence intervals for distribution functions in interval censoring models, studied in Sen and Xu (2015) and Groeneboom and Hendrickx (2018). We think the bias problem is solved more efficiently than in Groeneboom and Hendrickx (2018), where undersmoothing or explicit estimation of the bias was suggested, which is also suggested in Hall (1992).

We describe in Section 4 a method for choosing the bandwidth automatically, correcting the method used in Sen and Xu (2015). This method is used in Section 6 to choose the bandwidth in the classical example of the Lake Mendota data, which is the first example in the book Barlow et al. (1972).

Our paper was inspired by the recent paper Chakraborty and Ghosal (2021) for Bayesian confidence intervals in this setting, which will converge at a slower rate and which is analyzed from a bootstrap perspective in Groeneboom and Jongbloed (2023). We use their example of a regression function in our examples.

All examples in our paper can be recreated using the R scripts in Groeneboom (2021).

ORCID

Piet Groeneboom https://orcid.org/0000-0001-8027-8114
Geurt Jongbloed https://orcid.org/0000-0003-4708-5868

REFERENCES

Abrevaya, J., & Huang, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica*, *73*, 1175–1204.

Armstrong, T. (2015). Adaptive testing on a regression function at a point. *Annals of Statistics*, *43*, 2086–2101. https://doi.org/10.1214/15-AOS1342

Banerjee, M., & Wellner, J. (2005). Confidence intervals for current status data. *Scandinavian Journal of Statistics*, *32*, 405–424. https://doi.org/10.1111/j.1467-9469.2005.00454.x

Barlow, R., Bartholomew, D., Bremner, J., & Brunk, H. (1972). *Statistical inference under order restrictions. The theory and application of isotonic regression Wiley Series in Probability and Mathematical Statistics*. John Wiley & Sons.

Birman, M. Š., & Solomjak, M. Z. (1967). Piecewise polynomial approximations of functions of classes W_p^α. *Mathematics of the USSR-Sbornik*, *73*(115), 331–355.

Brunk, H. D. (1970). *Estimation of isotonic regression*. In *Nonparametric techniques in statistical inference (proceedings of Symposium, Indiana University, Bloomington, Indiana, 1969* (pp. 177–197). Cambridge University Press.

Chakraborty, M., & Ghosal, S. (2021). Coverage of credible intervals in nonparametric monotone regression. *Annals of Statistics*, *49*, 1011–1028. https://doi.org/10.1214/20-aos1989

Chernoff, H. (1964). Estimation of the mode. *Annals of the Institute of Statistical Mathematics*, *16*, 31–41.

Grenander, U. (1956). On the theory of mortality measurement. *Skandinavisk Aktuarietidskrift*, *39*, 125–153.

Groeneboom, P. (2021). *Monotone regression*. https://github.com/pietg/monotone-regression

Groeneboom, P. (2023). *Confidence intervals current status examples*. https://github.com/pietg/Confidence-Intervals-current-status-examples-

Groeneboom, P., & Hendrickx, K. (2017). The nonparametric bootstrap for the current status model. *Electronic Journal of Statistics*, *11*, 3446–3484.

Groeneboom, P., & Hendrickx, K. (2018). Confidence intervals for the current status model. *Scandinavian Journal of Statistics*, *45*, 135–163. https://doi.org/10.1111/sjos.12294

Groeneboom, P., & Jongbloed, G. (2012). Isotonic L_2-projection test for local monotonicity of a hazard. *Journal of Statistical Planning and Inference*, *142*, 1644–1658. https://doi.org/10.1016/j.jspi.2012.02.004

Groeneboom, P., & Jongbloed, G. (2014). *Nonparametric estimation under shape constraints*. Cambridge University Press.

Groeneboom, P., & Jongbloed, G. (2023). *Credible intervals and bootstrap confidence intervals in monotone regression*. https://arxiv.org/abs/2307.16168

Groeneboom, P., Jongbloed, G., & Witte, B. (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *Annals of Statistics*, *38*, 352–387.

Groeneboom, P., & Wellner, J. (1992). *Information bounds and nonparametric maximum likelihood estimation*. In *DMV seminar* (Vol. *19*). Birkhäuser Verlag.

Hall, P. (1992). *The bootstrap and Edgeworth expansion Springer series in statistics*. Springer.

Hall, P., Kay, J. W., & Titterington, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, *77*, 521–528. https://doi.org/10.1093/biomet/77.3.521

Hall, P., & Van Keilegom, I. (2005). Testing for monotone increasing hazard rate. *Annals of Statistics*, *33*, 1109–1137. https://doi.org/10.1214/009053605000000039

Härdle, W., & Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Annals of Statistics*, *19*, 778–796. https://doi.org/10.1214/aos/1176348120

Hazelton, M. (1996). Bandwidth selection for local density estimators. *Scandinavian Journal of Statistics*, *23*, 221–232.

Hong, H., & Li, J. (2020). The numerical bootstrap. *Annals of Statistics*, *48*, 397–412. https://doi.org/10.1214/19-AOS1812

Kim, J., & Pollard, D. (1990). Cube root asymptotics. *Annals of Statistics*, *18*, 191–219. https://doi.org/10.1214/aos/1176347498

Kosorok, M. (2008). *Bootstrapping the Grenander estimator*. In *Beyond parametrics in interdisciplinary research: Festschrift in honor of professor Pranab K. Sen* (Vol. *1*, pp. 282–292). The Institute of Mathematical Statistics Collections.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics*, *3*, 205–228.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, *79*, 871–880.

Sen, B., Banerjee, M., & Woodroofe, M. (2010). Inconsistency of bootstrap: The Grenander estimator. *Annals of Statistics*, *38*, 1953–1977. https://doi.org/10.1214/09-AOS777

Sen, B., & Xu, G. (2015). Model based bootstrap methods for interval censored data. *Computational Statistics & Data Analysis*, *81*, 121–129. https://doi.org/10.1016/j.csda.2014.07.007

van de Geer, S. (2000). *Applications of empirical process theory Cambridge series in statistical and probabilistic mathematics* (Vol. *6*). Cambridge University Press.

How to cite this article: Groeneboom, P., & Jongbloed, G. (2024). Confidence intervals in monotone regression. *Scandinavian Journal of Statistics*, 1–33. https://doi.org/10.1111/SJOS.12730

APPENDIX

Proof of Theorem 1. Let $t \in (0,1)$, and let n be sufficiently large, so that $t \in [h, 1-h]$, where $h = h_n = cn^{-1/5}$. Then $\tilde{f}_{nh}$ is represented by the so-called local smooth functional

$$\int K_h(t-x)\,\hat{f}_n(x)\,dx.$$

We now analyze the difference of this functional with the corresponding functional of f_0,

$$\int K_h(t-x)\left\{\hat{f}_n(x) - f_0(x)\right\}\,dx.$$

Defining G as the distribution function of the X_i, we can trivially write this in the following form:

$$\int K_h(t-x)\,\frac{\hat{f}_n(x) - f_0(x)}{g(x)}\,dG(x).$$

We define

$$\psi_{t,h}(u) = \frac{K_h(t-u)}{g(u)}, \tag{A1}$$

and

$$\overline{\psi}_{t,h}(u) = \begin{cases} \psi_{t,h}(\tau_i), & \text{if } f_0(u) > \hat{f}_n(\tau_i),\ u \in [\tau_i, \tau_{i+1}), \\ \psi_{t,h}(s), & \text{if } f_0(s) = \hat{f}_n(s),\ \text{for some } s \in [\tau_i, \tau_{i+1}), \\ \psi_{t,h}(\tau_{i+1}), & \text{if } f_0(u) < \hat{f}_n(\tau_i),\ u \in [\tau_i, \tau_{i+1}), \end{cases} \tag{A2}$$

where the τ_i are successive points of jump of $\hat{f}_n$.

The characterization of the LSE $\hat{f}_n$, implies that $\hat{f}_n$ satisfies

$$\int \overline{\psi}_{t,h}(u)\{y - \hat{f}_n(x)\}\,d\mathbb{H}_n(x,y) = 0,$$

see lemma 2.1 on p. 19 of Groeneboom and Jongbloed (2014). This is a consequence of the fact that $\overline{\psi}_{t,h}$ is constant between jumps of $\hat{f}_n$.

So we get:

$$\begin{aligned}
0 &= \int \overline{\psi}_{t,h}(x)\{y - \hat{f}_n(x)\}\,d\mathbb{H}_n(x,y) \\
&= \int \psi_{t,h}(x)\{y - \hat{f}_n(x)\}\,d\mathbb{H}_n(x,y) + \int \left\{\overline{\psi}_{t,h}(x) - \psi_{t,h}(x)\right\}\{y - \hat{f}_n(x)\}\,d\mathbb{H}_n(x,y) \\
&= \int \psi_{t,h}(x)\{y - \hat{f}_n(x)\}\,d(\mathbb{H}_n - H_0)(x,y) + \int \psi_{t,h}(x)\{y - \hat{f}_n(x)\}\,dH_0(x,y) \\
&\quad + \int \left\{\overline{\psi}_{t,h}(x) - \psi_{t,h}(x)\right\}\{y - \hat{f}_n(x)\}\,d\mathbb{H}_n(x,y) \\
&= \int \psi_{t,h}(x)\{y - \hat{f}_n(x)\}\,d(\mathbb{H}_n - H_0)(x,y) + \int \psi_{t,h}(x)\{f_0(x) - \hat{f}_n(x)\}\,dG(x) \\
&\quad + \int \left\{\overline{\psi}_{t,h}(x) - \psi_{t,h}(x)\right\}\{y - \hat{f}_n(x)\}\,d\mathbb{H}_n(x,y).
\end{aligned}$$

Hence

$$\int K_h(t-x)\left\{\hat{f}_n(x)-f_0(x)\right\}dx=\int \psi_{t,h}(x)\{\hat{f}_n(x)-f_0(x)\}\,dG(x)$$
$$=\int \psi_{t,h}(x)\{y-\hat{f}_n(x)\}\,d(\mathbb{H}_n-H_0)(x,y)+\int\left\{\overline{\psi}_{t,h}(x)-\psi_{t,h}(x)\right\}\{y-\hat{f}_n(x)\}\,d\mathbb{H}_n(x,y)$$
$$=\int \psi_{t,h}(x)\{y-\hat{f}_n(x)\}\,d(\mathbb{H}_n-H_0)(x,y)+\int\left\{\overline{\psi}_{t,h}(x)-\psi_{t,h}(x)\right\}\{f_0(x)-\hat{f}_n(x)\}\,d\mathbb{G}_n(x)$$
$$+\int\left\{\overline{\psi}_{t,h}(x)-\psi_{t,h}(x)\right\}\{y-f_0(x)\}\,d\mathbb{H}_n(x,y)=:A_I+A_{II}+A_{III}.$$

We have:

$$A_I=\int \psi_{t,h}(x)\{y-f_0(x)\}\,d(\mathbb{H}_n-H_0)(x,y)+\int \psi_{t,h}(x)\{f_0(x)-\hat{f}_n(x)\}\,d(\mathbb{G}_n-G)(x).$$

The central limit theorem implies that if $h\asymp n^{-1/5}$ the first term, multiplied with $n^{2/5}$, converges in distribution to a normal distribution with expectation zero and variance

$$\sigma^2=\frac{\sigma_0^2}{cg(t)}\int K(u)^2\,du, \tag{A3}$$

where σ_0^2 is the variance of the error in the regression model.

We now consider the term

$$\int \psi_{t,h}(x)\left\{f_0(x)-\hat{f}_n(x)\right\}d(\mathbb{G}_n-G)(x).$$

Note that we may assume that $|\hat{f}_n(x)-f_0(x)|$ is bounded by a fixed constant M on an interval $[a,b]\subset(0,1)$ for all sufficiently large n (consistency of the LSE on the interior of $[0,1]$).

Let $\mathcal{F}_h$ be the collection of functions $x\mapsto h\psi_{t,h}(x)\,\{f_0(x)-f(x)\}1_{[t-h,t+h]}$ for uniformly bounded nondecreasing functions $f:[a,b]\to\mathbb{R}$. Then the bracketing entropy number for the L_2-distance $H_B(\varepsilon,\mathcal{F}_h,G)$ satisfies

$$H_B(\varepsilon,\mathcal{F}_h,G)\le\frac{c}{\varepsilon},$$

for all $\varepsilon>0$ and some $c>0$, using the results of Birman and Solomjak (1967), see (2.5) on p. 18 of van de Geer (2000). As in van de Geer (2000) we use the notation $\|\cdot\|_2$ for the L_2- norm w.r.t. the relevant probability measure, such as G or (the two-dimensional) H_0.

We can now apply for example lemma 5.13 on p. 79 of van de Geer (2000) with $\alpha=1$ and $\beta=0$, where we multiply both sides of the inequalities inside the probabilities with $n^{2/5}h^{-1}$. This implies by (5.42) of lemma 5.13 in van de Geer (2000) that there are constants c and n_0 such that for all $T\ge c$ and $n\ge n_0$, taking $\alpha=1$ and $\beta=0$,

$$\mathbb{P}\left\{n^{2/5}h^{-1}\sup_{\phi\in\mathcal{F}_{h_n},\;n^{1/3}\|\phi-\phi_0\|_2\le1}\left|\int\{\phi(x)-\phi_0(x)\}\,d(\mathbb{G}_n-G)(x)\right|\ge Tn^{2/5}h^{-1}n^{-2/3}\right\}$$
$$\le c\exp\left\{-Tn^{1/3}/c^2\right\}, \tag{A4}$$

and

$$\mathbb{P}\left\{ n^{2/5}h^{-1} \sup_{\phi\in\mathcal{F}_n,\ n^{1/3}\|\phi-\phi_0\|_2>1} \|\phi-\phi_0\|_2^{-1/2}\left|\int \{\phi(x)-\phi_0(x)\}\, d(\mathbb{G}_n - G)(x)\right| \geq Tn^{2/5}h^{-1}n^{-1/2} \right\}$$
$$\leq c\exp\left\{-T/c^2\right\}, \tag{A5}$$

where we multiply both sides of the inequalities inside the probabilities with $n^{2/5}h^{-1}$.

Taking $h \asymp n^{-1/5}$, we get from (A4),

$$\mathbb{P}\left\{ n^{3/5} \sup_{\phi\in\mathcal{F}_{h_n},\ n^{1/3}\|\phi-\phi_0\|_2\leq 1} \left|\int \{\phi(x)-\phi_0(x)\}\, d(\mathbb{G}_n - G)(x)\right| \geq Tn^{-1/15} \right\} \leq c\exp\left\{-Tn^{1/3}/c^2\right\},$$

and hence

$$n^{2/5}1_{\{n^{1/3}\|(\hat f_n - f_0)1_{[t-h,t+h]}\|_2\leq 1\}} \int \psi_{t,h}(x)\left\{f_0(x)-\hat f_n(x)\right\}\, d(\mathbb{G}_n - G)(x) = o_p(1). \tag{A6}$$

The inequality (A5) yields:

$$\mathbb{P}\left\{ n^{3/5} \sup_{\phi\in\mathcal{F}_n,\ n^{1/3}\|\phi-\phi_0\|_2>1} \frac{n^{-1/10}}{\|\phi-\phi_0\|_2^{1/2}} \left|\int \{\phi(x)-\phi_0(x)\}\, d(\mathbb{G}_n - G)(x)\right| \geq T \right\} \leq c\exp\left\{-T/c^2\right\},$$

and hence

$$\frac{n^{2/5}}{\|\psi_{t,h}\{\hat f_n - f_0\}\|_2^{1/2}} 1_{\{n^{1/3}\|(\hat f_n - f_0)1_{[t-h,t+h]}\|_2>1\}} \int \psi_{t,h}(x)\left\{f_0(x)-\hat f_n(x)\right\}\, d(\mathbb{G}_n - G)(x) = O_p(1),$$

implying

$$n^{2/5}1_{\{n^{1/3}\|(\hat f_n - f_0)1_{[t-h,t+h]}\|_2>1\}} \int \psi_{t,h}(x)\left\{f_0(x)-\hat f_n(x)\right\}\, d(\mathbb{G}_n - G)(x)$$
$$= O_p\left(\|\psi_{t,h}\{\hat f_n - f_0\}\|_2^{1/2}\right) = o_p(1). \tag{A7}$$

The last equality follows from the fact that for large n and $h \asymp n^{-1/5}$,

$$\mathbb{E}\left\|\psi_{t,h}\|(\hat f_n - f_0)\right\|_2^2 = h^{-2}\int_{t-h}^{t+h} \frac{K((t-x)/h)^2}{g(x)^2}\, \mathbb{E}\{\hat f_n(x) - f_0(x)\}^2\, dG(x = O\left(h^{-1}n^{-2/3}\right) = O\left(n^{-7/15}\right).$$

Combining (A6) and (A7) we find

$$n^{2/5}\int \psi_{t,h}(x)\left\{f_0(x)-\hat f_n(x)\right\}\, d(\mathbb{G}_n - G)(x) = o_p(1).$$

For the term A_{II} we get:

$$A_{II} = \int \left\{\overline{\psi}_{t,h}(x) - \psi_{t,h}(x)\right\}\{f_0(x) - \hat f_n(x)\}\, d(\mathbb{G}_n - G)(x)$$
$$+ \int \left\{\overline{\psi}_{t,h}(x) - \psi_{t,h}(x)\right\}\{f_0(x) - \hat f_n(x)\}\, dG(x).$$

By the same reasoning as used for A_I, the first term n the right-hand side is again of order $o_p(n^{-2/5})$. For the second term on the right-hand side we get by the Cauchy-Schwarz inequality:

$$\left|\int \left\{\overline{\psi}_{t,h}(x)-\psi_{t,h}(x)\right\}\{f_0(x)-\hat f_n(x)\}\,dG(x)\right|$$
$$\le \left\{\int \left\{\overline{\psi}_{t,h}(x)-\psi_{t,h}(x)\right\}^2 dG(x)\right\}^{1/2}\left\{\int_{t-h}^{t+h}\left\{f_0(x)-\hat f_n(x)\right\}^2 dG(x)\right\}^{1/2}.$$

We first note that, letting $[a,b]\subset(0,1)$:

$$\mathbb{E}\left\|(\overline{\psi}_{t,h}-\psi_{t,h})1_{[a,b]}\right\|_2 = O_p\left(h^{-3/2}\|(\hat f_n-f_0)1_{[a,b]}\|_2\right), \tag{A8}$$

where we use the Cauchy–Schwarz inequality.and where we use

$$\left|\overline{\psi}_{t,h}(x)-\psi_{t,h}(x)\right|\le\frac{c}{h^2}\left|\hat f_n(x)-f_0(x)\right|.$$

This relation presently belongs to the standard tools of this type of theory, and is for example proved in lemma A.4 of of Groeneboom et al. (2010) and used in Groeneboom and Jongbloed (2014), pp. 290 and 332.

So we obtain:

$$\left|\int_a^b\left\{\overline{\psi}_{t,h}(x)-\psi_{t,h}(x)\right\}\,\left\{\hat f_n(x)-f_0(x)\right\}\,dG(x)\right| = O_p\left(h^{-1}\left\|(\hat f_n-f_0)1_{[a,b]}\right\|_2^2\right).$$

Hence:

$$\left|\int_a^b\left\{\overline{\psi}_{t,h}(x)-\psi_{t,h}(x)\right\}\,\left\{\hat f_n(x)-f_0(x)\right\}\,dG(x)\right| = O_p\left(n^{1/5-2/3}\right)=O_p\left(n^{-7/15}\right)=o_p\left(n^{-2/5}\right).$$

Finally we consider A_{III}. We can write

$$A_{III}=\int\left\{\overline{\psi}_{t,h}(x)-\psi_{t,h}(x)\right\}\{y-f_0(x)\}\,d\left(\mathbb{H}_n-H_0\right)(x,y),$$

since

$$\int\left\{\overline{\psi}_{t,h}(x)-\psi_{t,h}(x)\right\}\{y-f_0(x)\}\,dH_0(x,y)=\int\left\{\overline{\psi}_{t,h}(x)-\psi_{t,h}(x)\right\}\{f_0(x)-f_0(x)\}\,dG(x)=0.$$

We now consider the class of functions

$$\mathcal{G}_h=\{\phi\,:\,\phi(x,y)=h\{\overline{\psi}_{h,f}(x)-\psi_h(x)\}(y-x),\ f\,:\,[a,b]\to\mathbb{R}\},$$

for piecewise constant right-continuous bounded monotone functions $f:[a,b]\to\mathbb{R}$, where $\overline{\psi}_{h,f}$ is chosen in the same way as $\overline{\psi}_h$ in (A2), but with $\hat f_n$ replaced by f. We now get by (5.42) of Lemma 5.13 in van de Geer (2000) that there are constants c and n_0

such that for all $T \geq c$ and $n \geq n_0$,

$$\mathbb{P}\left\{ n^{2/5}h^{-1} \sup_{\phi \in \mathcal{G}_h,\ n^{1/3}\|\phi-\phi_0\|_2 \leq 1} \left| \int \{\phi(x,y)-\phi_0(x,y)\}\ d(\mathbb{H}_n - H_0)(x,y) \right| \geq T n^{2/5}h^{-1}n^{-2/3} \right\}$$
$$\leq c \exp\left\{-Tn^{1/3}/c^2\right\}, \tag{A9}$$

and

$$\mathbb{P}\left\{ n^{2/5}h^{-1} \sup_{\phi \in \mathcal{F}_n,\ n^{1/3}\|\phi-\phi_0\|_2 > 1} \|\phi-\phi_0\|_2^{-1/2} \left| \int \{\phi(x,y)-\phi_0(x,y)\}\ d(\mathbb{H}_n - H_0)(x,y) \right| \right.$$
$$\left. \geq T n^{2/5}h^{-1}n^{-1/2} \right\}$$
$$\leq c \exp\left\{-\frac{T}{c^2}\right\}. \tag{A10}$$

Note that

$$\begin{aligned} h^2 &\int_{x\in[t-h,t+h]} \left\{\overline{\psi}_{t,h}(x)-\psi_{t,h}(x)\right\}^2 \{y-f_0(x)\}^2\ dH_0(x,y) \\ &\lesssim h^{-2} \int_{x\in[t-h,t+h]} \left\{\hat{f}_n(x)-f_0(x)\right\}^2 \{y-f_0(x)\}^2\ dH_0(x,y) \\ &= O_p\left(h^{-1}n^{-2/3}\right) = O_p\left(n^{-7/15}\right), \end{aligned} \tag{A11}$$

if $h \asymp n^{-1/5}$.

So, using the same arguments as used for A_I, and defining

$$\phi_1(x,y) = y - f_0(x)\},$$

we get:

$$n^{2/5} 1_{n^{1/3}\|h\{\overline{\psi}_{t,h}-\psi_{t,h}\}\phi_1\}\|_2 \leq 1\}} \int \{\overline{\psi}_{t,h}(x)-\psi_{t,h}(x)\ \{y-f_0(x)\}\ d(\mathbb{H}_n - H_0)(x,y) = o_p(1),$$

and

$$\begin{aligned} &n^{2/5} 1_{n^{1/3}\|h\{\overline{\psi}_{t,h}-\psi_{t,h}\}\phi_1\}\|_2 > 1\}} \int \{\overline{\psi}_{t,h}(x)-\psi_{t,h}(x)\ \{y-f_0(x)\}\ d(\mathbb{H}_n - H_0)(x,y) \\ &= O_p\left(\left\| \left(\overline{\psi}_{t,h}-\psi_{t,h}\right)\phi_1 \right\|_2^{1/2} \right) = O_p\left(h^{-1/2}n^{-7/60}\right) = o_p(1), \end{aligned}$$

using (A11) and $h \asymp n^{-1/5}$ in the last two steps. It now follows that $A_{III} = o_p(n^{-2/5})$. ■

Proof of Theorem 2. We follow the set-up of the proof of Theorem 1, take $t \in (0,1)$ and consider n sufficiently large so that $h < t < 1-h$. Then $\tilde{f}^*_{nh}$ is defined by

$$\tilde{f}^*_{nh}(t) = \int K_h(t-x)\,\hat{f}^*_n(x)\ dx,$$

where $\hat{f}^*_n$ is the LSE based on a bootstrap sample. Consider

$$\int K_h(t-x)\left\{\hat{f}^*_n(x) - \tilde{f}_{nh_0}(x)\right\}\ dx,$$

and recall that G is the distribution function of the X_i. Then:

$$\int K_h(t-x)\left\{\hat f_n^*(x)-\tilde f_{nh_0}(x)\right\}\,dx=\int K_h(t-x)\,\frac{\hat f_n^*(x)-\tilde f_{nh_0}(x)}{g(x)}\,dG(x).$$

Let $\tilde H_{n,h_0}$ be the distribution function of the pairs (X_i,Y_i^*), induced by uniform sampling with replacement from the centered residuals $\tilde E_i$. This means that the distribution function $\tilde H_{n,h_0}$ corresponds to the probability measure P_{n,h_0} of the two-dimensional random variable (X,Y^*), defined by

$$P_{n,h_0}\left\{(X,Y^*)=(X_i,\tilde f_{nh_0}(X_i)+\tilde E_i)\right\}=\frac{1}{n},\qquad i=1,\dots,n,$$

conditionally on $(X_1,Y_1),\dots,(X_n,Y_n)$.

Defining $\psi_{t,h}$ as in (A1), define

$$\overline{\psi}^*_{t,h}(x)=\begin{cases}\psi_{t,h}(\tau_i^*), & \text{if } \tilde f_{nh_0}(x)>\hat f_n^*(\tau_i),\ u\in[\tau_i,\tau_{i+1}),\\ \psi_{t,h}(s), & \text{if } \tilde f_{nh_0}(s)=\hat f_n^*(s),\ \text{for some } s\in[\tau_i,\tau_{i+1}),\\ \psi_{t,h}(\tau_{i+1}^*), & \text{if } \tilde f_{nh_0}(x)<\hat f_n^*(\tau_i),\ u\in[\tau_i,\tau_{i+1}),\end{cases}$$

where the τ_i^* are successive points of jump of $\hat f_n^*$. We have:

$$\int y\,\overline{\psi}^*_{t,h}(x)\,d\tilde H_{nh_0}(x,y)=n^{-1}\sum_{i=1}^n\left\{\tilde f_{nh_0}(X_i)+\tilde E_i\right\}\overline{\psi}^*_{t,h}(X_i)$$
$$=n^{-1}\sum_{i=1}^n\tilde f_{nh_0}(X_i)\overline{\psi}^*_{t,h}(X_i)=\int\tilde f_{nh_0}(x)\overline{\psi}^*_{t,h}(x)\,d\mathbb{G}_n(x),$$

where $\mathbb{G}_n$ is the empirical distribution function of the X_i, using the fact that the residuals $\tilde E_i$ are centered. (have mean zero). Hence

$$\int\{y-\tilde f_{nh_0}(x)\}\ \psi_{t,h}(x)\ d\tilde H_{n,h_0}(x,y)=0. \tag{A12}$$

By the characterization of the LSE $\hat f_n^*$, we also have

$$\int\{y-\hat f_n^*(x)\}\overline{\psi}^*_{t,h}(x)\,d\mathbb{H}_n^*(x,y)=0,$$

where H_n^* is the empirical distribution function of the bootstrap sample $(X_1,Y_1^*),\dots,(X_n,Y_n^*)$.

So we get:

$$0=\int\{y-\hat f_n^*(x)\}\overline{\psi}^*_{t,h}(x)\,d\mathbb{H}_n^*(x,y)=\int\{y-\hat f_n^*(x)\}\psi_{t,h}(x)\,d\mathbb{H}_n^*(x,y)$$
$$+\int\{y-\hat f_n^*(x)\}\left\{\overline{\psi}^*_{t,h}(x)-\psi_{t,h}(x)\right\}\,d\mathbb{H}_n^*(x,y)$$

$$= \int \{y - \tilde f_{nh_0}(x)\}\psi_{t,h}(x)\, d\left(\mathbb{H}_n^* - \tilde H_{n,h_0}\right)(x,y) + \int \{\tilde f_{nh_0}(x) - \hat f_n^*(x)\}\psi_{t,h}(x)\, d\mathbb{H}_n^*(x,y)$$
$$+ \int \{y - \hat f_n^*(x)\}\left\{\overline{\psi}_{t,h}^*(x) - \psi_{t,h}(x)\right\}\, d\mathbb{H}_n^*(x,y),$$

using (A12) to replace $d\mathbb{H}_n^*$ by $d(\mathbb{H}_n^* - \tilde H_{n,h_0})$ in the first integral after the last equality sign. This implies

$$\int K_h(t-x)\left\{\hat f_n^*(x) - \tilde f_{nh_0}(x)\right\}\, dx = \int \psi_{t,h}(x)\left\{\hat f_n^*(x) - \tilde f_{nh_0}(x)\right\}\, dG(x)$$
$$= \int \psi_{t,h}(x)\left\{\hat f_n^*(x) - \tilde f_{nh_0}(x)\right\}\, dG(x) + \int \{y - \hat f_n^*(x)\}\overline{\psi}_{t,h}^*(x)\, d\mathbb{H}_n^*(x,y)$$
$$= \int \{y - \tilde f_{nh_0}(x)\}\psi_{t,h}(x)\, d\left(\mathbb{H}_n^* - \tilde H_{n,h_0}\right)(x,y) + \int \{\tilde f_{nh_0}(x) - \hat f_n^*(x)\}\psi_{t,h}(x)\, d\left(\mathbb{G}_n - G\right)(x)$$
$$+ \int \{y - \hat f_n^*(x)\}\left\{\overline{\psi}_{t,h}^*(x) - \psi_{t,h}(x)\right\}\, d\mathbb{H}_n^*(x,y). \tag{A13}$$

The Lindeberg–Feller central limit theorem implies that

$$n^{2/5}\int \{y - \tilde f_{nh_0}(x)\}\psi_{t,h}(x)\, d\left(\mathbb{H}_n^* - \tilde H_{n,h_0}\right)(x,y) \xrightarrow{\mathcal{D}} N(0,\sigma^2),$$

given $(X_1,Y_1),\dots,(X_n,Y_n)$, almost surely along sequences $(X_1,Y_1),(X_2,Y_2),\dots$. Note that

$$n^{2/5}\int \{y - \tilde f_{nh_0}(x)\}\psi_{t,h}(x)\, d\left(\mathbb{H}_n^* - \tilde H_{n,h_0}\right)(x,y) = n^{-3/5}\sum_{i=1}^n \{Y_i^* - \tilde f_{nh_0}(x)\}\psi_{t,h}(X_i),$$

which has conditional expectation zero and conditional variance

$$n^{-1/5}\sum_{i=1}^n \left\{n^{-1}\sum_{j=1}^n \tilde E_j^2\right\}\psi_{t,h}(X_i)^2 \sim \frac{\sigma_0^2}{cg(t)}\int K(u)^2\, du, \tag{A14}$$

given $(X_1,Y_1),\dots,(X_n,Y_n)$, if $h \sim cn^{-1/5}$.

Note that

$$n^{-1}\sum_{j=1}^n \tilde E_j^2 \sim n^{-1}\sum_{j=1}^n \left\{Y_j - f_0(X_i)\right\}^2 + n^{-1}\sum_{j=1}^n \left\{f_0(X_i) - \tilde f_{nh_0}(X_i)\right\}^2 \to \sigma_0^2,$$

almost surely, as $n \to \infty$.

We now turn to the second term after the last equality sign in (A13). By lemma 4.1 and corol. 4.1 from Groeneboom and Jongbloed (2023), added to the monotonicity of f_n^*, we may assume that

$$\mathcal{F}_n = \left\{\{\hat f_n^* - \tilde f_{nh_0}\}1_{[a,b]}\right\},$$

is of uniformly bounded variation for an interval $[a,b] \subset (0,1)$ and hence has entropy with bracketing $H(\varepsilon, \mathcal{F}_n, \mathbb{G}_n) \le c\varepsilon^{-1}$ for the L_2-distance and some $c > 0$, conditionally on $(X_1,Y_1),\dots,(X_n,Y_n)$, along all sequences $(X_1,Y_1),(X_2,Y_2),\dots$. Moreover, the L_2-distance $\|\{\hat f_n^* - \tilde f_{nh_0}\}1_{[a,b]}\|_2$ is of order $n^{-1/3}$, just as the L_2-distance $\|\{\hat f_n - f_0\}1_{[a,b]}\|_2$. So we find, if $h \asymp n^{-1/5}$,

$$\int \{\tilde f_{nh_0}(x) - \hat f_n^*(x)\}\psi_{t,h}(x)\, d(\mathbb{G}_n - G)(x) = O_p^*\left(h^{-1}n^{-2/3}\right) = O_p^*\left(n^{-7/15}\right) = o_p^*\left(n^{-2/5}\right),$$

where we add the star $*$ to the O_p and o_p symbol to indicate the conditional meaning of these symbols for the bootstrap samples.

Finally, the third term after the last equality sign in (A13) can be written

$$\begin{aligned}
&\int \{y - \hat f_n^*(x)\}\left\{\overline{\psi}_{t,h}^*(x) - \psi_{t,h}(x)\right\}\, d\left(\mathbb{H}_n^* - \tilde H_{n,h_0}\right)(x,y)\\
&\quad + \int \left\{\tilde f_{nh_0}(x) - \hat f_n^*(x)\right\}\left\{\overline{\psi}_{t,h}^*(x) - \psi_{t,h}(x)\right\}\, d\tilde H_{n,h_0}(x,y)\\
&= \int \{y - \hat f_n^*(x)\}\left\{\overline{\psi}_{t,h}^*(x) - \psi_{t,h}(x)\right\}\, d\left(\mathbb{H}_n^* - \tilde H_{n,h_0}\right)(x,y)\\
&\quad + \int \left\{\tilde f_{nh_0}(x) - \hat f_n^*(x)\right\}\left\{\overline{\psi}_{t,h}^*(x) - \psi_{t,h}(x)\right\}\, d\mathbb{G}_n(x).
\end{aligned}$$

Using

$$\left|\overline{\psi}_{t,h}^*(x) - \psi_{t,h}(x)\right| \lesssim h^{-2}\left|\hat f_n^*(x) - \tilde f_{nh_0}(x)\right|,$$

for all x in a neighborhood of t, we find again that these terms are $o_p(n^{-2/5})$.

Moreover, again conditionally and almost surely,

$$\int K_h(t-x)\tilde f_{nh_0}(x)\, dx = \tilde f_{nh_0}(t) + \frac{1}{2}h^2 f_0''(t)\int u^2K(u)\, du + o\left(h^2\right).$$

Hence

$$\begin{aligned}
&E\left\{\tilde f_{nh}^*(t) - \tilde f_{nh_0}(t)\middle|(X_1,Y_1),\dots,(X_n,Y_n)\right\}\\
&= E\left\{\int K_h(t-x)\left\{\hat f_n^*(x) - \tilde f_{nh_0}(x)\right\}\, dx\middle|(X_1,Y_1),\dots,(X_n,Y_n)\right\}\\
&\quad + \frac{1}{2}h^2 f_0''(t)\int u^2K(u)\, du + o\left(h^2\right).
\end{aligned}$$

This gives the expression for the mean of the conditional limit distribution of $\tilde f_{nh}^*(t) - \tilde f_{nh_0}(t)$. Note that the bias drops out in the construction of the bootstrap confidence intervals. ∎

Proof of Lemma 1. We have:

$$\tilde f_{nh_{n,0}}''(t) - h_{n,0}^{-3}\int K''\left((t-x)/h_{n,0}\right) f_0(x)\, dx = h_{n,0}^{-3}\int K''\left((t-x)/h_{n,0}\right)\left\{\hat f_n(x) - f_0(x)\right\}\, dx,$$

and

$$\begin{aligned}
\int K''\left((t-x)/h_{n,0}\right)\left\{\hat f_n(x) - f_0(x)\right\}\, dx &= \int \frac{K''\left((t-x)/h_{n,0}\right)}{g(x)}\left\{\hat f_n(x) - f_0(x)\right\}\, dG(x)\\
&= \int \frac{K''\left((t-x)/h_{n,0}\right)}{g(x)}\left\{\hat f_n(x) - y\right\}\, dH_0(x,y),
\end{aligned}$$

where G is the distribution function of the X_i and H_0 the distribution function of the pairs (X_i, Y_i). We now define the function $\psi_{t,h}$ by

$$\psi_{t,h}(x) = \frac{K''((t-x)/h)}{h^3 g(x)},$$

and introduce a piecewise constant version $\overline{\psi}_{t,h}$ of $\psi_{t,h}$ as

$$\overline{\psi}_{t,h}(x) = \begin{cases} \psi_{t,h}(\tau_i), & \text{if } f_0(x) > \hat{f}_n(\tau_i),\ u \in [\tau_i, \tau_{i+1}), \\ \psi_{t,h}(s), & \text{if } f_0(s) = \hat{f}_n(s),\ \text{for some } s \in [\tau_i, \tau_{i+1}), \\ \psi_{t,h}(\tau_{i+1}), & \text{if } f_0(x) < \hat{f}_n(\tau_i),\ u \in [\tau_i, \tau_{i+1}), \end{cases}$$

where the τ_i are successive points of jump of $\hat{f}_n$. So we can write:

$$\begin{aligned} h_{n,0}^{-3} \int K''\big((t-x)/h_{n,0}\big) \left\{\hat{f}_n(x) - f_0(x)\right\} dx &= h_{n,0}^{-3} \int \frac{K''\big((t-x)/h_{n,0}\big)}{g(x)} \left\{\hat{f}_n(x) - y\right\} dH_0(x,y) \\ &= \int \psi_{t,h_{n,0}}(x) \left\{\hat{f}_n(x) - y\right\} dH_0(x,y) \\ &= \int \left\{\psi_{t,h_{n,0}}(x) - \overline{\psi}_{t,h_{n,0}}(x)\right\} \left\{\hat{f}_n(x) - y\right\} dH_0(x,y) + \int \overline{\psi}_{t,h_{n,0}}(x) \left\{\hat{f}_n(x) - y\right\} dH_0(x,y) \\ &= \int \left\{\psi_{t,h_{n,0}}(x) - \overline{\psi}_{t,h_{n,0}}(x)\right\} \left\{\hat{f}_n(x) - y\right\} dH_0(x,y) \\ &\quad + \int \overline{\psi}_{t,h_{n,0}}(x) \left\{\hat{f}_n(x) - y\right\} d(H_0 - \mathbb{H}_n)(x,y), \end{aligned}$$

using the characterization of the LSE $\hat{f}_n$ in the last step. We now get:

$$\begin{aligned} &\int \left\{\psi_{t,h_{n,0}}(x) - \overline{\psi}_{t,h_{n,0}}(x)\right\} \left\{\hat{f}_n(x) - y\right\} dH_0(x,y) \\ &\quad = \int \left\{\psi_{t,h_{n,0}}(x) - \overline{\psi}_{t,h_{n,0}}(x)\right\} \left\{\hat{f}_n(x) - f_0(x)\right\} dG(x) = O_p\left(h_{n,0}^{-4} n^{-2/3}\right) = O_p\left(n^{-2/9}\right), \end{aligned}$$

and

$$\int \overline{\psi}_{t,h_{n,0}}(x) \left\{\hat{f}_n(x) - y\right\} d(H_0 - \mathbb{H}_n)(x,y) = O_p\left(h_{n,0}^{-3/2} n^{-1/2}\right) = O_p\left(n^{-1/3}\right).$$

So the conclusion is:

$$\tilde{f}''_{nh_{n,0}}(t) - h_{n,0}^{-3} \int K''\big((t-x)/h_{n,0}\big) f_0(x)\, dx = O_p\left(n^{-2/9}\right).$$

But under the conditions on f_0 of Theorem 1 we have:

$$h_{n,0}^{-3} \int K''\big((t-x)/h_{n,0}\big) f_0(x)\, dx = f_0''(t) + o(1), \qquad n \to \infty.$$

■