CONFIDENCE INTERVALS IN MONOTONE REGRESSION

BY PIET GROENEBOOM AND GEURT JONGBLOED

Delft University of Technology

Delft Institute of Applied Mathematics, Mekelweg 4, 2628 CD Delft, The Netherlands.
P.Groeneboom@tudelft.nl; G.Jongbloed@tudelft.nl

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, however, that a 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 [15]. The procedure is illustrated using a well known dataset related to climate change.

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 non-vanishing density g on $[0, 1]$ and

$$Y_i = f_0(X_i) + \varepsilon_i, \ \ 1 \le i \le n. \tag{1.1}$$

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 5, and for the rest stick to the nondecreasing functions. The theory for these cases is completely similar.

The basic monotone least squared 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

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

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 $X_1 < X_2 \ldots < X_n$ 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 [8]).

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 the bootstrap samples, is inconsistent (see, e.g., [13], [15] and [14] for results related to this phenomenon). It has been suggested in [15] in the context of interval censored data to generate intervals by using a smoothed bootstrap. In the monotone regression setting, this 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 related current status model were proposed in [5] and [6]. 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. [11] suggests a third method, which we will systematically use in the sequel.

In section 2, we define a Smoothed Least Squares Estimator (SLSE) for f_0. We show that 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 [15] 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. Section 5 illustrates the method using a well know dataset related to climate change. We note that the methods described in this paper can be applied using the open software [4].

2. Smooth (monotone) estimation of the regression function. As immediately follows from its construction, the LSE $\hat{f}_n$ is a (non-smooth) step function. It can be used to define a smooth estimate. We define a particular *Smoothed* LSE (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

$$(2.1) \qquad\qquad K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right), \ \ u \in \mathbb{R}.$$

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

$$(2.2) \qquad\qquad \tilde{f}_{nh}(t) = \int K_h(t-x)\, \hat{f}_n(x)\, dx.$$

Note that on $[h, 1-h]$ the SLSE inherits the 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

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

Estimator (2.2) is rather different from the Nadaraya-Watson (NW) estimator, which is defined by

$$(2.3) \qquad \tilde{f}_{nh}^{NW}(t) = \frac{\int y K_h(t-x) \, d\mathbb{H}_n(x,y)}{\int K_h(t-x) \, d\mathbb{H}_n(x,y)} = \frac{\sum_{i=1}^n Y_i K_h(t-X_i)}{\sum_{i=1}^n K_h(t-X_i)},$$

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 (2.2), but is not necessarily monotone.

The NW estimator (2.3) is seemingly simpler than SLSE (2.2), because it is expressed as sum over sample values, whereas (2.2) is an integral with respect to Lebesgue measure. However, using integration by parts, (2.2) can also be rewritten as a simple sum. Indeed, for $t \in [h, 1-h]$,

$$(2.4) \quad \tilde{f}_{nh}(t) = \hat{f}_n(0) + \int_{x \in (0, t+h]} I\!K_h(t-x) \, d\hat{f}_n(x) = \hat{f}_n(0) + \sum_{\tau_i \in (0, t+h]} I\!K_h(t-\tau_i) p_i,$$

where

$$(2.5) \qquad I\!K_h(y) = \int_{-\infty}^{y} K_h(u) \, du = \int_{-\infty}^{y/h} K(u) \, du.$$

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 (2.2) outside the interval $[h, 1-h]$, leads to serious bias. We define $\tilde{f}_{nh}$ threfore 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

$$(2.6) \qquad \widetilde{f}_{nh}(t) = \widetilde{f}_{nh}(h) + (t-h)\widetilde{f}'_{nh}(h) + \tfrac{1}{2}(t-h)^2 \widetilde{f}''_{n,h_0}(h_0),$$

where

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

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

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

$$(2.7) \qquad \widetilde{f}_{nh}(t) = \widetilde{f}_{nh}(h) + (t-(1-h))\widetilde{f}'_{nh}(1-h) + \tfrac{1}{2}(t-(1-h))^2 \widetilde{f}''_{n,h_0}(1-h_0).$$

Note that we may lose the monotonicity in the boundary intervals $[0, h]$ and $[1-h, 1]$ in this way, but that the probability of this phenomenon will tend to zero with increasing sample size.

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 p. 210 and 211 of [8].

We have the following asymptotic pointwise result for the SLSE. Note that by going from $\hat{f}_n(t)$ to (2.2) 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 [2], Theorem 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*

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

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 (2.2) based on the pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$, and $h \sim cn^{-1/5}$ for $c > 0$,

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

Here

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

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

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

REMARK 1.   For 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 (3.8) below).

We now give a road map of the proof of Theorem 1. The proof itself is given in Section 7. Since the estimators are based on the non-linear LSE $\hat{f}_n$, the proof is totally different from the proofs for the NW estimator. We have to introduce methods similar to those used in [7] for local smooth functionals in the current status model.

The first step is to write

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

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 $\bar{\psi}_{t,h}$ 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

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

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

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

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

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

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

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

for a $K > 0$, which follows from a judicious choice of $\bar{\psi}_{t,h}$. 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 Chapter 9 of [16]. Alternatively, we could use the "switch relation", as used in [6] and [15]. The proof is completed by incorporating the behavior of the bias term in (2.9),

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

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

$$(3.1) \qquad \tilde{f}_{nh}(t) - f_0(t)$$

as studied asymptotically in Theorem 1. The distribution of this quantity is approximated by the distribution of a related quantity based on observatons $(X_1, Y_1^*), \ldots, (X_n, Y_n^*)$ 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:

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

Next, we center the E_i by substracting their mean:

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

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

$$(3.2) \qquad (X_i, Y_i^*) = \left(X_i, \tilde{f}_{nh_0}(X_i) + E_i^* \right), \qquad i = 1, \ldots, n,$$

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

$$(3.3) \qquad \tilde{f}_{nh}^*(t) - \tilde{f}_{nh_0}(t).$$

Here $\tilde{f}_{nh}^*(t)$ is the estimate of f_0, based on a bootstrap sample, with bandwidth h as in (3.1). Note that we keep the X_i fixed in the bootstrap samples.

EXAMPLE 1. Consider the setting used in [3], who study a Bayesian approach to constructing confidence sets for f_0. 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. 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 $t = 0.01, 0.02, \ldots, 0.99$. The coverage is shown in Figure 2a. 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 [4].
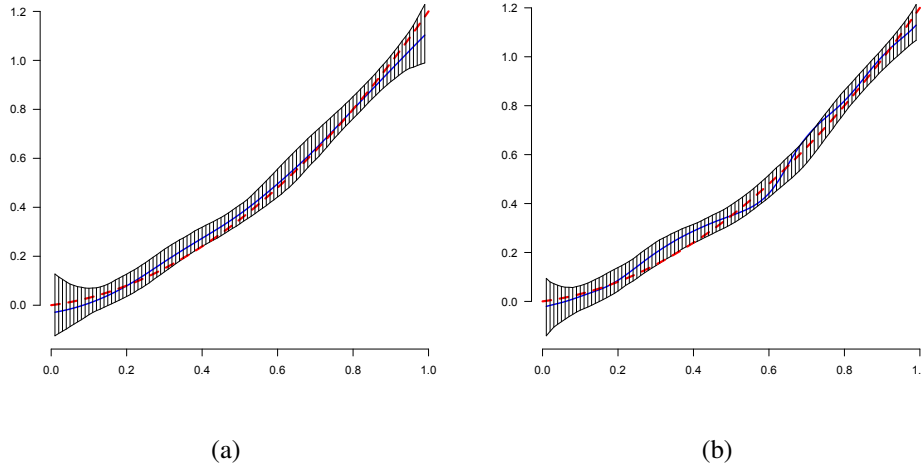

(a)   (b)

Fig 1: (a) The SLSE (blue, solid) and 95% confidence intervals for the same sample, using the confidence intervals (3.4). The red dashed curve is f_0. (b) The NW 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 $h_0 = 0.7n^{-1/9}$.

The 95% bootstrap confidence intervals are given by

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

where $Q^*_{0.025}$ and $Q^*_{0.975}$ are the 2.5th and 97.5th percentiles of 1000 (bootstrap) samples of (3.3). Note that the percentiles $Q^*_{0.025}$ and $Q^*_{0.975}$ contain an estimate of the asymptotic bias

$$\tfrac{1}{2}h^2 \tilde{f}''_{nh_0}(t) \sim \tfrac{1}{2}h^2 f''_0(t),$$

(see also Lemma 1 in Section 4) and that therefore the bias of $\tilde{f}_{nh}(t)$ drops out in (3.4). So we do not need undersmoothing or explicit estimation of the bias in our procedure. The oversmoothing by taking $h_0 \asymp n^{-1/9}$ (or at least a bandwidth tending to zero slower than $n^{-1/5}$) is essential here, though.

Theorem 2 shows that asymptotically, the bootstrap method described will asymptotically give the right coverage, in the sense that after rescaling with $n^{2/5}$ the asymptotic distribution of (3.3) under the estimated model (3.2) coincides with the asymptotic distribution of (3.1) under model (1.1).

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 $t \in (0, 1)$,*

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

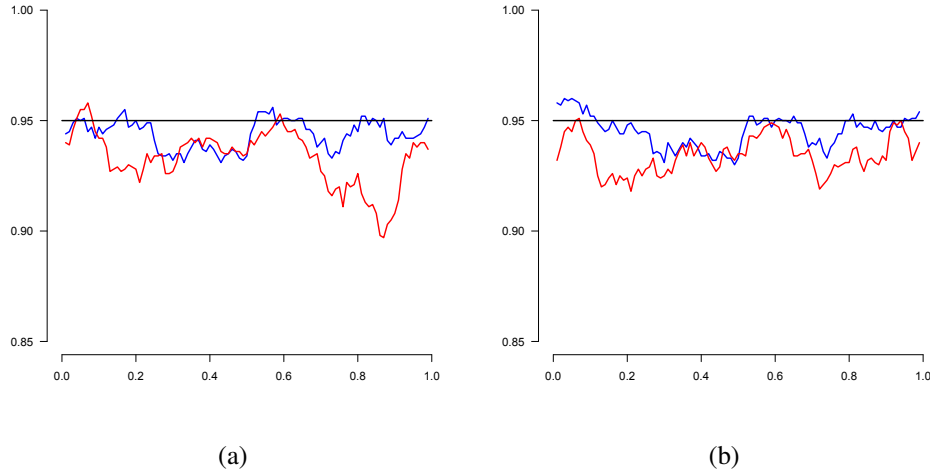(a)   (b)

Fig 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 (3.4) for the function $f_0(x) = x^2 + x/5$ and for the SLSE (blue) and the NW estimator (red), and (b) fraction of the B experiments where $f_0(t)$ is in the interval (3.4) 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 $h = 0.7n^{-1/9}$.

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 (2.8).

The proof is given in Section 7. 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

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

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

$$(3.5) \qquad \left(\tilde{f}_{nh}^{NW}\right)^*(t) - \tilde{f}_{nh_0}^{NW}(t).$$

Here $\left(\tilde{f}_{nh}^{NW}\right)^*$ is the NW estimator based on (3.2) with $\tilde{f}_{nh_0}(X_i)$ replaced by $\tilde{f}_{nh_0}^{NW}(X_i)$ and E_i^* sampled with replacement from the residuals $\tilde{E}_i^{NW}$, $i = 1, \ldots, n$,

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

In [9] the variance of the NW estimator, conditionally on $X_1, \ldots, X_n$, in the model (1.1) at t is shown to be equal to

$$(3.6) \qquad \tau_t^2 = \sigma_0^2 \beta_t^2,$$
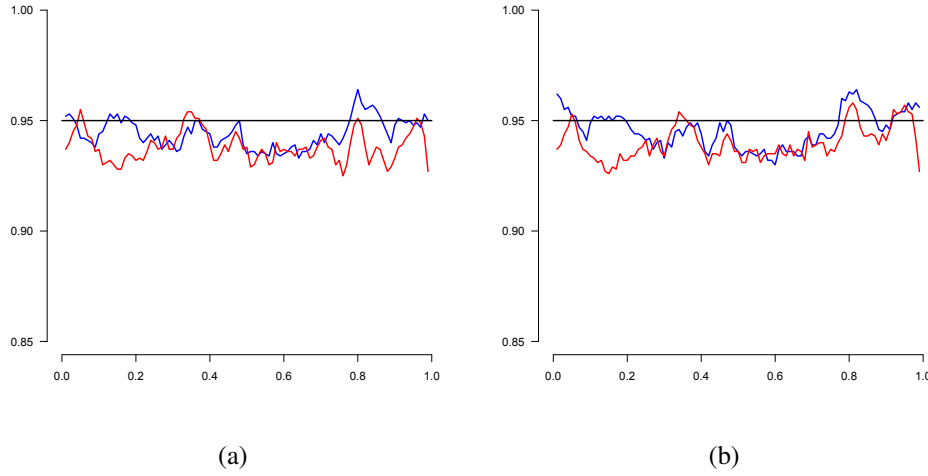
(a)   (b)

Fig 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 (3.4) for the function $f_0(x) = x^2 + x/5$ and for the SLSE (blue) and the NW estimator (red), and (b) fraction of the B experiments where $f_0(t)$ is in the intervals (3.4) 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 $h = 0.7n^{-1/9}$.

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

$$(3.7) \qquad \beta_t^2 = \frac{\sum_{i=1}^n K_h (t - X_i)^2}{\{\sum_{i=1}^n K_h(t - X_i)\}^2}.$$

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 $\tilde{f}_{nh}^{NW}(t)$ behaves like

(3.8)

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

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 (3.1) 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

$$\left\{ \tilde{f}_{nh}(t) - f_0(t) \right\} / \hat{\sigma}_{n,0} \tag{3.9}$$

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

$$\hat{\sigma}_{n,0}^2 = n^{-1} \sum_{i=1}^{n} \tilde{E}_i^2. \tag{3.10}$$

The distribution of (3.9) under the true model is then approximated by the distribution of

$$\left\{ \tilde{f}_{nh}^*(t) - \tilde{f}_{nh_0}(t) \right\} / \hat{\sigma}_{n,0}^* \tag{3.11}$$

where

$$\left(\hat{\sigma}_{n,0}^* \right)^2 = n^{-1} \sum_{i=1}^{n} \left(\tilde{E}_i^* - \bar{E}^* \right)^2, \qquad \bar{E}^* = n^{-1} \sum_{i=1}^{n} \tilde{E}_i^*,$$

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 (3.11). It is then given by

$$\left(\tilde{f}_{nh}(t) - Q_{0.975}^* \hat{\sigma}_{n,0}, \tilde{f}_{nh}(t) - Q_{0.025}^* \hat{\sigma}_{n,0} \right), \tag{3.12}$$

where $\hat{\sigma}_{n,0}^2$ is defined by (3.10). 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 [9] (but note the typo w.r.t. the index j in [9]). We take the definition from [10] and define

$$\left(\hat{\sigma}_n^{NW} \right)^2 = (n - m)^{-1} \sum_{i=1}^{n-2} \left(\sum_{j=0}^{2} d_j Y_{i+j} \right)^2, \tag{3.13}$$

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

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

(as recommended in [10]).

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 (3.5) and the Studentized intervals on the differences

$$\left\{ \left(\tilde{f}_{nh}^{NW} \right)^*(t) - \tilde{f}_{nh_0}^{NW}(t) \right\} / \left(\hat{\sigma}_n^{NW} \right)^* \tag{3.14}$$

where $\left(\hat{\sigma}_n^{NW} \right)^*$ is the estimate (3.13) for the bootstrap samples. It is seen that in both cases there is not a great improvement.

For the Nadaraya-Watson 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 Nadaraya-Watson estimator, see Figure 5. We still do not understand this phenomenon, based on the different ways of estimating σ_0 for the Nadaraya-Watson estimator, however.
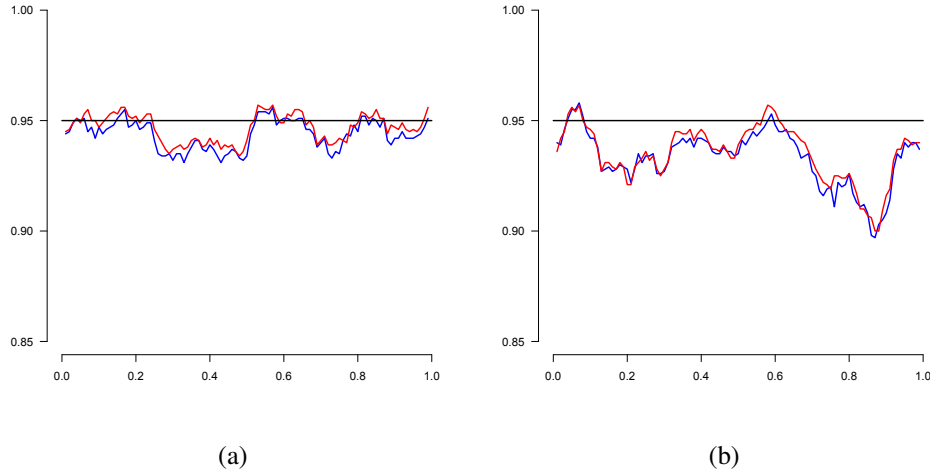
Fig 4: Coverage of the confidence intervals, Studentized and non-Studentized, based on the bootstrap described. Fraction of the B experiments where $f_0(t)$ is in the intervals (3.4) for the function $f_0(x) = x^2 + x/5$ for (a) the Studentized SLSE (red) and the non-Studentized SLSE (blue), and (b) the Studentized NW estimator (red) and the non-Studentized NW estimator (blue). The figures are based on $B = 1000$ samples of size $n = 100$, and $t = 0.01, \ldots, 0.99$. The chosen bandwidths h and h_0 are $h = 0.5n^{-1/5}$ and $h_0 = 0.7n^{-1/9}$.
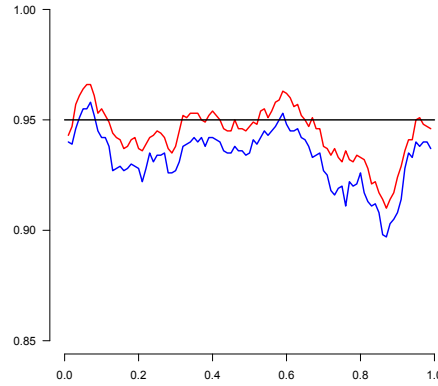

Fig 5: The Studentized NW estimator (red), with variance estimated from the residuals, just as for the SLSE, and the non-Studentized NW estimator (blue). The figures are based on $B = 1000$ samples of size $n = 100$, and $t = 0.01, \ldots, 0.99$. The chosen bandwidths h and h_0 are $h = 0.5n^{-1/5}$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

4. Bandwidth selection. 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 we want to minimize as a function of h is given by:

$$(4.1) \qquad MSE_h(t) = \mathbb{E}\left\{ \left\{ \tilde{f}_{nh}(t) - f_0(t) \right\}^2 \,\Big|\, X_1, \ldots, X_n \right\}.$$

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,

$$(4.2) \qquad MSE_h^*(t) = \mathbb{E}\left\{\left\{\tilde{f}_{nh}^*(t) - \tilde{f}_{nh_0}(t)\right\}^2 \,\middle|\, (X_1, Y_1), \ldots, (X_n, Y_n)\right\},$$

where h_0 is called a "pilot" bandwidth. We shall show that (4.2) is asymptotically independent of the constant in the pilot bandwidth h_0 if we take $h_0 \asymp n^{-1/9}$.

We have:

$$MSE_h^*(t) = \mathbb{E}\left\{\left\{\int K_h(t-x)\left\{\hat{f}_n^*(x) - \tilde{f}_{nh_0}(x)\right\}dx\right\}^2 \,\middle|\, (X_1, Y_1), \ldots, (X_n, Y_n)\right\}$$

$$(4.3) \qquad + \mathbb{E}\left\{\left\{\int K_h(t-x)\,\tilde{f}_{nh_0}(x)\,dx - f_0(t)\right\}^2 \,\middle|\, X_1, \ldots, X_n\right\}$$

For the second term on the right we get:

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

so

$$\mathbb{E}\left\{\left\{\int K_h(t-x)\,\tilde{f}_{nh_0}(x)\,dx - f_0(t)\right\}^2 \,\middle|\, X_1, \ldots, X_n\right\}$$

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

We have the following result.

LEMMA 1. *Let the conditions of Theorem 1 be satisfied. Moreover, let $h_0 = h_{n,0} \sim c_0 n^{-1/9}$, 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 [15], where the pilot bandwidth is chosen of order $n^{-1/5}$ will not work. Another way out is to use subsampling, as used in [6], but choosing the right subsample size is a rather hard problem.

The proof of Lemma 1 is given in Section 7. The lemma suggests, as in [12], 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 4th 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 $n^{-1/5}$, since otherwise the variance of $\tilde{f}_{n,h_0}''(t)$ does not tend to zero.

For the first term on the right of (4.3) 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 (7.8) and the argument using the Lindeberg-Feller central limit theorem part of the proof of Theorem 2).

So, asymptotically, the bandwidth h, minimizing (4.3) minimizes

$$\frac{\sigma_0^2}{g(t)nh} \int K(u)^2 \, du + \tfrac{1}{4}h^4 \tilde{f}_0''(t)^2 \left\{ \int u^2 K(u) \, du \right\}^2.$$

The minimization of (4.1) leads asymptotically to the same minimization over h.

Instead of minimizing (4.2) we minimize a Monte Carlo approximation of (4.2):

$$(4.4) \qquad B^{-1} \sum_{i=1}^{B} \left\{ \tilde{f}_{nh}^{*,i}(t) - \tilde{f}_{nh_0}(t) \right\}^2,$$

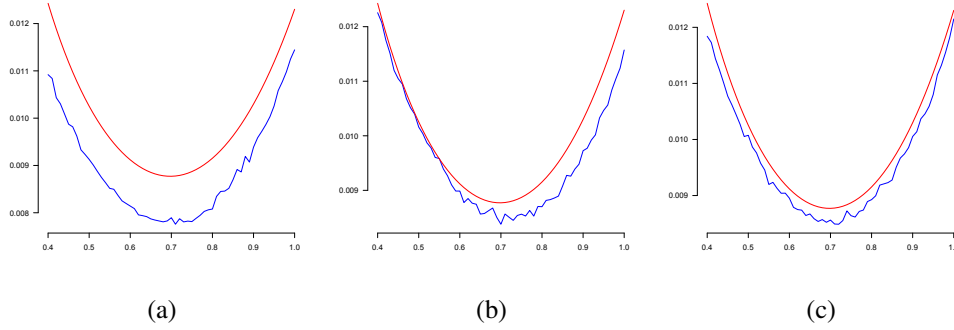where the $\tilde{f}_{nh}^{*,i}$, $i = 1, \ldots, B$ are the estimates in B bootstrap samples.

(a)   (b)   (c)

Fig 6: Estimated MISE, for $n = 500, 1000$ and 5000 and $h = cn^{-1/5}$, $c = 0.4, 0.41, \ldots, 1$ and asymptotic MISE. The blue curve is based on (4.5), and the red curve on (4.6), $B = 10,000$, 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 $c_{min} \approx 0.697601$.

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 (4.4) by

$$(4.5) \quad \widehat{MISE}_c^* = n^{4/5} B^{-1} \sum_{j=1}^{B} \sum_{i=1}^{m} \left\{ \tilde{f}_{n,cn^{-1/5}}^{*,j}(t_i) - \tilde{f}_{nh_0}(t_i) \right\}^2 \Delta_i, \qquad \Delta_i = t_i - t_{i-1}.$$

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:

$$(4.6) \quad AsMISE_c = \frac{\sigma_0^2}{c} \int K(u)^2 \, du \int_{t=a}^{b} \frac{1}{g(t)} \, dt + \tfrac{1}{4}c^4 \left\{ \int u^2 K(u) \, du \right\}^2 \int_{t=a}^{b} f_0''(t)^2 \, dt.$$

In this case we consider

$$(4.7) \qquad \mathbb{E}\left\{\int_a^b \left\{\tilde{f}_{nh}^*(t) - \tilde{f}_{nh_0}(t)\right\}^2 dt \;\middle|\; (X_1, Y_1), \ldots, (X_n, Y_n)\right\},$$

where h_0 is the pilot bandwidth. We have:

$$\mathbb{E}\left\{\int_a^b \left\{\tilde{f}_{nh}^*(t) - \tilde{f}_{nh_0}(t)\right\}^2 dt \;\middle|\; (X_1, Y_1), \ldots, (X_n, Y_n)\right\}$$

$$\sim \mathbb{E}\left\{\int_{t=a}^b \left\{\int K_h(t-x)\left\{\hat{f}_n^*(x) - \tilde{f}_{nh_0}(x)\right\} dx\right\}^2 dt \;\middle|\; (X_1, Y_1), \ldots, (X_n, Y_n)\right\}$$

$$(4.8) \qquad + \int_{t=a}^b \left\{\int K_h(t-x)\,\tilde{f}_{nh_0}(x)\, dx - \tilde{f}_{nh_0}(t)\right\}^2 dt.$$

For the second term on the right we get:

$$\int_{t=a}^b \left\{\int K_h(t-x)\,\tilde{f}_{nh_0}(x)\, dx - \tilde{f}_{nh_0}(t)\right\}^2 dt$$

$$\sim \tfrac{1}{4} h^4 \int_{t=a}^b \tilde{f}_{nh_0}''(t)^2\, dt \left\{\int u^2 K(u)\, du\right\}^2 dt + o_p\left(h^4\right)$$

Since we want $\tilde{f}_{n,h_0}''$ to be as close as possible to f_0'', we suggest to minimize

$$(4.9) \qquad \int_{t=a}^b \left\{\tilde{f}_{nh_0}''(t) - f_0''(t)\right\}^2 dt$$

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

We get:

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

and

$$\int K_{h_0}''(t-x) f_0(x)\, dx - f_0''(t)$$

$$= \int K_{h_0}(t-x) f_0''(x)\, dx - f_0''(t) = \int K_{h_0}(t-x)\left\{f_0''(x) - f_0''(t)\right\} dx$$

$$= \tfrac{1}{2} h_0^2 f_0^{(4)}(t) \int u^2 K(u)\, du + o\left(h_0^2\right),$$

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

This yields:

$$\int_{t=a}^b \{\tilde{f}_{n,h_0}''(t) - f_0''(t)\}^2 dt$$

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

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. Lake Mendota: yearly number of days frozen. As a real data application, we give confidence intervals for the Lake Mendota data, analyzed in [8], Section 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 [1].

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 (4.5), 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 7. 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 8.
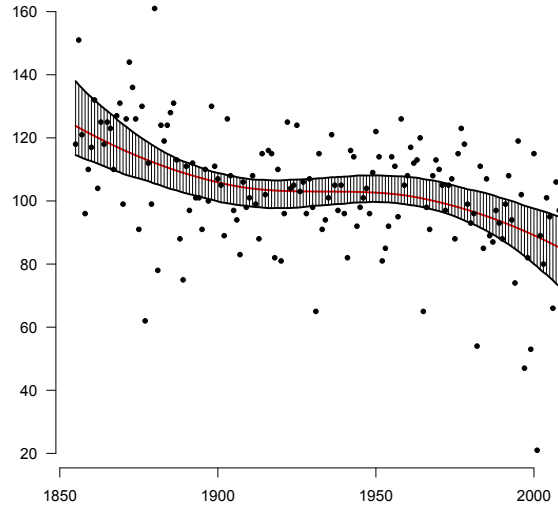

Fig 7: 95% confidence intervals for the regression function for the Mendota data. The red curve is the SLSE, with the bandwidth chosen by the method based on the MISE-approximation (4.5).

6. Conclusion. We introduce the Smoothed Least Squares Estimator (SLSE)

$$(6.1) \qquad \tilde{f}_{nh}(t) = \int K_h(t - x)\hat{f}_n(x)\, dx$$

in the monotone regression problem, where $\hat{f}_n$ is the monotone nonparametric regression estimator, where K is a smooth symmetric kernel with support $[-1, 1]$ and $K_h = h^{-1}K(\cdot/h)$.
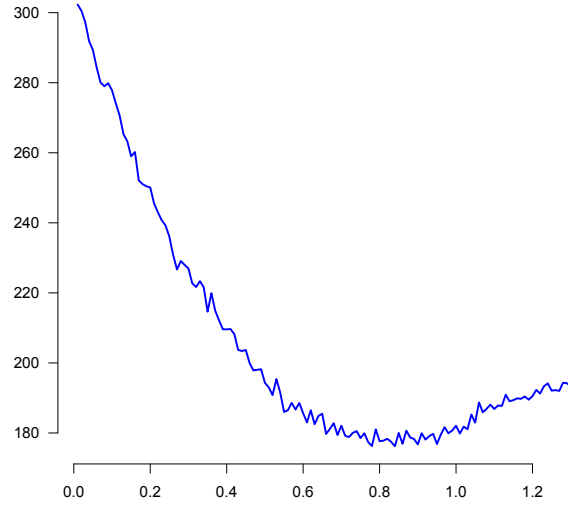
Fig 8: The MISE-approximation (4.5), as a function of c for the transformed Mendota data.

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 (6.1). 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 [11].

We compare the isotonic 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 [15] and [6]. We think the bias problem is solved more efficiently than in [6], where undersmoothing or explicit estimation of the bias was suggested, which is also suggested in [9].

We describe in Section 4 a method for choosing the bandwidth automatically, correcting the method used in [15]. This method is used in Section 5 to choose the bandwidth in the classical example of the Lake Mendota data, which was the first example in the book [1].

Our paper was inspired by the recent paper [3] for Bayesian confidence intervals in this setting, which will converge at a slower rate and which we will analyze from a bootstrap perspective in a separate paper. We used their example of a regression function in our examples.

All examples in our paper can be recreated using the R scripts in [4].

7. 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

(7.1)
$$\psi_{t,h}(u) = \frac{K_h(t-u)}{g(u)},$$

and

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

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

We now get:

$$\int K_h(t-x) \frac{\hat{f}_n(x) - f_0(x)}{g(x)} dG(x) = \int \psi_{t,h}(x) \left\{ \hat{f}_n(x) - f_0(x) \right\} dG(x)$$

(7.2)
$$= \int \bar{\psi}_{t,h}(x) \left\{ \hat{f}_n(x) - f_0(x) \right\} dG(x) + \int \left\{ \psi_{t,h}(x) - \bar{\psi}_{t,h}(x) \right\} \left\{ \hat{f}_n(x) - f_0(x) \right\} dG(x)$$

Let H_0 be the distribution function of the pairs (X_i, Y_i), with corresponding empirical distribution function $\mathbb{H}_n$. Then we get for the first term on the last line of in (7.2)

$$\int \bar{\psi}_{t,h}(x) \left\{ \hat{f}_n(x) - f_0(x) \right\} dG(x)$$

$$= \int \bar{\psi}_{t,h}(x) \left\{ \hat{f}_n(x) - y \right\} dH_0(x,y)$$

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

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

$$\qquad + \int \left\{ \bar{\psi}_{t,h}(x) - \psi_{t,h}(x) \right\} \left\{ y - \hat{f}_n(x) \right\} d(\mathbb{H}_n - H_0)(x,y)$$

where the second equality follows from the characterization of the LSE $\hat{f}_n$.

The first term on the last line can be rewritten:

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

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

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

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

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

where $\mathbb{G}_n$ is the empirical distribution of $X_1, \ldots, X_n$.

We now get, under the conditions of the theorem, that the first term on the last line, multiplied by $n^{2/5}$, converges in distribution to a normal distribution with expectation zero and variance

$$(7.3) \qquad \sigma^2 = \frac{\sigma_0^2}{cg(t)} \int K(u)^2 \, du,$$

if $h \sim cn^{-1/5}$, as $n \to \infty$, and, by the usual entropy bounds for bounded monotone functions, the last term is $o_p(n^{-1/5})$. Note that, because of the support of $\psi_{t,h}$, we can restrict ourselves to functions defined on a closed interval $[a, b] \subset (0, 1)$, whee $\hat{f}_n$ is bounded for sufficiently large n, because of its consistency on closed subintervals of $(0, 1)$.

For the last term of (7.2) we follow an argument close to the line of reasoning on p. 333 of [8]. First of all we get, using the Cauchy-Schwarz inequality:

$$(7.4) \qquad \left\| (\hat{f}_n - f_0) 1_{[t-h,t+h]} \right\|_{2,G} = O_p \left(h^{1/2} \| (\hat{f}_n - f_0) \|_2 1_{[a,b]} \right),$$

for a closed subinterval $[a, b] \subset (0, 1)$.

Using this notation, we get, and letting $[a, b] \subset (0, 1)$:

$$(7.5) \qquad \left\| (\bar{\psi}_{t,h} - \psi_{t,h}) 1_{[a,b]} \right\|_{2,G} = O_p \left(h^{-3/2} \| (\hat{f}_n - f_0) 1_{[a,b]} \|_2 \right).$$

Combining (7.4) and (7.5) and using the Cauchy-Schwarz inequality, we obtain:

$$\left| \int_a^b \left\{ \bar{\psi}_{t,h}(x) - \psi_{t,h}(x) \right\} \left\{ \hat{f}_n(x) - f_0(x) \right\} dG(x) \right|$$

$$= O_p \left(h^{-1} \left\| (\hat{f}_n - f_0) 1_{[a,b]} \right\|_2^2 \right).$$

By results in Chapter 9 of [16], $\| (\hat{f}_n - f_0) 1_{[a,b]} \|_2 = O_p(n^{-1/3})$ and if $h \asymp n^{-1/5}$, we obtain from this:

$$\left| \int_a^b \left\{ \bar{\psi}_{t,h}(x) - \psi_{t,h}(x) \right\} \left\{ \hat{f}_n(x) - f_0(x) \right\} dG(x) \right|$$

$$= O_p \left(n^{1/5 - 2/3} \right) = O_p \left(n^{-7/15} \right) = o_p \left(n^{-2/5} \right).$$

So we get as conclusion:

$$n^{2/5} \int K_h(t - x) \left\{ \hat{f}_n(x) - f_0(x) \right\} dx \xrightarrow{d} N(0, \sigma^2),$$

where $N(0, \sigma^2)$ is a normal distribution with expectation zero and variance (7.3), if $h \sim cn^{-1/5}$, as $n \to \infty$.

Since, under the conditions of the theorem, we have

$$n^{2/5} \left\{ \int K_h(t - x) \, f_0(x) \, dx - f_0(t) \right\} = \tfrac{1}{2} c^2 f_0''(t) \int u^2 K(u) \, du + o(1),$$

if $h \sim cn^{-1/5}$, as $n \to \infty$, the result follows. □

PROOF OF THEOREM 2. We follow the set-up of the proof of Theorem 1 and assume $t \in [h, 1-h]$. This time $\tilde{f}_{nh}^*$ is represented by the local smooth functional

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

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 $\tilde{E}_i$. Defining $\psi_{t,h}$ as in (7.1), define

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

where the τ_i^* are successive points of jump of $\hat{f}_n^*$. We have:

$$\int y\,\bar{\psi}_{t,h}^*(x)\,d\tilde{H}_{nh_0}(x,y) = \int \tilde{f}_{nh_0}(x)\bar{\psi}_{t,h}^*(x)\,d\tilde{H}_{n,h_0}(x,y) = \int \tilde{f}_{nh_0}(x)\bar{\psi}_{t,h}^*(x)\,d\mathbb{G}_n(x),$$

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

(7.6) $$\int \{y - \tilde{f}_{nh_0}(x)\}\,\psi_{t,h}(x)\,d\tilde{H}_{n,h_0}(x,y) = 0.$$

By the characterization of the LSE $\hat{f}_n^*$, we also have

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

So we get:

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

$$= \int \{y - \hat{f}_n^*(x)\}\psi_{t,h}(x)\,d\mathbb{H}_n^*(x,y) + \int \{y - \hat{f}_n^*(x)\}\{\bar{\psi}_{t,h}^*(x) - \psi_{t,h}(x)\}\,d\mathbb{H}_n^*(x,y)$$

$$= \int \{y - \tilde{f}_{nh_0}(x)\}\psi_{t,h}(x)\,d\left(\mathbb{H}_n^* - \tilde{H}_{n,h_0}\right)(x,y)$$

$$\quad + \int \{\tilde{f}_{nh_0}(x) - \hat{f}_n^*(x)\}\psi_{t,h}(x)\,d\mathbb{H}_n^*(x,y)$$

$$\quad + \int \{y - \hat{f}_n^*(x)\}\{\bar{\psi}_{t,h}^*(x) - \psi_{t,h}(x)\}\,d\mathbb{H}_n^*(x,y),$$

using (7.6) 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

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

$$= \int \psi_{t,h}(x)\{\hat{f}_n^*(x) - \tilde{f}_{nh_0}(x)\}\,dG(x)$$

$$= \int \psi_{t,h}(x)\{\hat{f}_n^*(x) - \tilde{f}_{nh_0}(x)\}\,dG(x) + \int \{y - \hat{f}_n^*(x)\}\bar{\psi}_{t,h}^*(x)\,d\mathbb{H}_n^*(x,y)$$

$$= \int \{y - \tilde{f}_{nh_0}(x)\}\psi_{t,h}(x)\,d\big(\mathbb{H}_n^* - \tilde{H}_{n,h_0}\big)(x,y)$$

$$+ \int \{\tilde{f}_{nh_0}(x) - \hat{f}_n^*(x)\}\psi_{t,h}(x)\,d\big(\mathbb{G}_n - G\big)(x)$$

(7.7) $$+ \int \{y - \hat{f}_n^*(x)\}\{\bar{\psi}_{t,h}^*(x) - \psi_{t,h}(x)\}\,d\mathbb{H}_n^*(x,y).$$

The Lindeberg-Feller central limit theorem implies that

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

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

$$n^{2/5}\int \{y - \tilde{f}_{nh_0}(x)\}\psi_{t,h}(x)\,d\big(\mathbb{H}_n^* - \tilde{H}_{n,h_0}\big)(x,y)$$

$$= n^{-3/5}\sum_{i=1}^{n}\{Y_i^* - \tilde{f}_{nh_0}(x)\}\psi_{t,h}(X_i),$$

which has conditional expectation zero and conditional variance

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

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

Note that

$$n^{-1}\sum_{j=1}^{n}\tilde{E}_j^2 \sim n^{-1}\sum_{j=1}^{n}\{Y_j - f_0(X_j)\}^2 + n^{-1}\sum_{j=1}^{n}\left\{f_0(X_i) - \tilde{f}_{nh_0}(X_i)\right\}^2 \longrightarrow \sigma_0^2,$$

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

We now turn to the second term after the last equality sign in (7.7). By the consistency of $\tilde{f}_{nh_0}$ and the ensuing conditional consistency of $\hat{f}_n^*$ on bounded intervals $[a,b] \subset (0,1)$, added to the monotonicity of these functions on such interval, 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 such an interval $[a,b]$ 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 $(X_1, Y_1), \ldots, (X_n, Y_n)$, along all sequences $(X_1, Y_1), (X_2, Y_2), \ldots$. 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 (7.7) can be written

$$\int \{y - \hat{f}_n^*(x)\}\{\bar{\psi}_{t,h}^*(x) - \psi_{t,h}(x)\}\, d\left(\mathbb{H}_n^* - \tilde{H}_{n,h_0}\right)(x,y)$$

$$+ \int \{\tilde{f}_{nh_0}(x) - \hat{f}_n^*(x)\}\{\bar{\psi}_{t,h}^*(x) - \psi_{t,h}(x)\}\, d\tilde{H}_{n,h_0}(x,y)$$

$$= \int \{y - \hat{f}_n^*(x)\}\{\bar{\psi}_{t,h}^*(x) - \psi_{t,h}(x)\}\, d\left(\mathbb{H}_n^* - \tilde{H}_{n,h_0}\right)(x,y)$$

$$+ \int \{\tilde{f}_{nh_0}(x) - \hat{f}_n^*(x)\}\{\bar{\psi}_{t,h}^*(x) - \psi_{t,h}(x)\}\, d\mathbb{G}_n(x).$$

Using

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

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,

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

Hence

$$E\left\{\tilde{f}_{nh}^*(t) - \tilde{f}_{nh_0}(t) \,\big|\, (X_1, Y_1), \ldots, (X_n, Y_n)\right\}$$

$$= E\left\{\int K_h(t - x)\{\hat{f}_n^*(x) - \tilde{f}_{nh_0}(x)\}\, dx \,\big|\, (X_1, Y_1), \ldots, (X_n, Y_n)\right\}$$

$$+ \tfrac{1}{2}h^2 f_0''(t)\int u^2 K(u)\, du + o\left(h^2\right).$$

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. $\square$

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

$$\int K'' \left((t-x)/h_{n,0} \right) \left\{ \hat{f}_n(x) - f_0(x) \right\} dx$$

$$= \int \frac{K'' \left((t-x)/h_{n,0} \right)}{g(x)} \left\{ \hat{f}_n(x) - f_0(x) \right\} dG(x)$$

$$= \int \frac{K'' \left((t-x)/h_{n,0} \right)}{g(x)} \left\{ \hat{f}_n(x) - y \right\} dH_0(x,y),$$

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

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

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

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

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

$$h_{n,0}^{-3} \int K'' \left((t-x)/h_{n,0} \right) \left\{ \hat{f}_n(x) - f_0(x) \right\} dx$$

$$= h_{n,0}^{-3} \int \frac{K'' \left((t-x)/h_{n,0} \right)}{g(x)} \left\{ \hat{f}_n(x) - y \right\} dH_0(x,y))$$

$$= \int \psi_{t,h_{n,0}}(x) \left\{ \hat{f}_n(x) - y \right\} dH_0(x,y)$$

$$= \int \left\{ \psi_{t,h_{n,0}}(x) - \bar{\psi}_{t,h_{n,0}}(x) \right\} \left\{ \hat{f}_n(x) - y \right\} dH_0(x,y)$$

$$+ \int \bar{\psi}_{t,h_{n,0}}(x) \left\{ \hat{f}_n(x) - y \right\} dH_0(x,y)$$

$$= \int \left\{ \psi_{t,h_{n,0}}(x) - \bar{\psi}_{t,h_{n,0}}(x) \right\} \left\{ \hat{f}_n(x) - y \right\} dH_0(x,y)$$

$$+ \int \bar{\psi}_{t,h_{n,0}}(x) \left\{ \hat{f}_n(x) - y \right\} d \left(H_0 - \mathbb{H}_n \right)(x,y),$$

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

$$\int \left\{ \psi_{t,h_{n,0}}(x) - \bar{\psi}_{t,h_{n,0}}(x) \right\} \left\{ \hat{f}_n(x) - y \right\} dH_0(x,y)$$

$$\int \left\{ \psi_{t,h_{n,0}}(x) - \bar{\psi}_{t,h_{n,0}}(x) \right\} \left\{ \hat{f}_n(x) - f_0(x) \right\} dG(x)$$

$$= O_p \left(h_{n,0}^{-4} n^{-2/3} \right) = O_p \left(n^{-2/9} \right),$$

and

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

So the conclusion is:

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

But under the conditions of Theorem 1 we have:

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

$\square$

REFERENCES

[1] BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical inference under order restrictions. The theory and application of isotonic regression.* John Wiley & Sons, London-New York-Sydney Wiley Series in Probability and Mathematical Statistics. MR0326887 (48 ##5229)

[2] BRUNK, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969)* 177–197. Cambridge Univ. Press, London. MR0277070

[3] CHAKRABORTY, M. and GHOSAL, S. (2021). Coverage of credible intervals in nonparametric monotone regression. *Ann. Statist.* **49** 1011–1028. MR4255117

[4] GROENEBOOM, P. (2021). Monotone Regression. https://github.com/pietg/monotone-regression.

[5] GROENEBOOM, P. and HENDRICKX, K. (2017). The nonparametric bootstrap for the current status model. *Electron. J. Stat.* **11** 3446-3484.

[6] GROENEBOOM, P. and HENDRICKX, K. (2018). Confidence Intervals for the Current Status Model. *Scand. J. Stat.* **45** 135-163. doi: 10.1111/sjos.12294.

[7] GROENEBOOM, P., JONGBLOED, G. and WITTE, B. I. (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *Ann. Statist.* **38** 352–387.

[8] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation under Shape Constraints.* Cambridge Univ. Press, Cambridge.

[9] HALL, P. (1992). *The bootstrap and Edgeworth expansion. Springer Series in Statistics.* Springer.

[10] HALL, P., KAY, J. W. and TITTERINGTON, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77** 521–528. MR1087842

[11] HÄRDLE, W. and MARRON, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19** 778–796. MR1105844

[12] HAZELTON, M. (1996). Bandwidth selection for local density estimators. *Scand. J. Statist.* **23** 221–232. MR1394655

[13] KOSOROK, M. R. (2008). Bootstrapping the Grenander estimator. In *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen. Inst. Math. Stat. Collect.* **1** 282–292. Inst. Math. Statist., Beachwood, OH.

[14] SEN, B., BANERJEE, M. and WOODROOFE, M. B. (2010). Inconsistency of bootstrap: the Grenander estimator. *Ann. Statist.* **38** 1953–1977. MR2676880 (2011f:62046)

[15] SEN, B. and XU, G. (2015). Model based bootstrap methods for interval censored data. *Comput. Statist. Data Anal.* **81** 121–129. MR3257405

[16] VAN DE GEER, S. A. (2000). *Applications of empirical process theory. Cambridge Series in Statistical and Probabilistic Mathematics* **6**. Cambridge University Press, Cambridge. MR1739079 (2001h:62002)