Electronic Journal of Statistics
Vol. 13 (2019) 1523–1579
ISSN: 1935-7524
https://doi.org/10.1214/19-EJS1555

Adaptive confidence sets for kink estimation

Viktor Bengs* and **Hajo Holzmann*,†**

Department of Mathematics and Computer Science.
Philipps-Universität Marburg.
Hans-Meerwein-Straße, 35032 Marburg, Germany
e-mail: bengs@mathematik.uni-marburg.de; holzmann@mathematik.uni-marburg.de

Abstract: We consider estimation of the location and the height of the jump in the γ-th derivative - a kink of order γ - of a regression curve, which is assumed to be Hölder smooth of order $s \geq \gamma + 1$ away from the kink. Optimal convergence rates as well as the joint asymptotic normal distribution of estimators based on the zero-crossing-time technique are established. Further, we construct joint as well as marginal asymptotic confidence sets for these parameters which are honest and adaptive with respect to the smoothness parameter s over subsets of the Hölder classes. The finite-sample performance is investigated in a simulation study, and a real data illustration is given to a series of annual global surface temperatures.

Keywords and phrases: Adaptive estimation, change-point estimation, limit theorem, nonparametric statistics, Lepski's method, Z-estimation.

Received October 2018.

1. Introduction

Suppose that we observe data (x_i, Y_i), $i = 1, \ldots, n$, from the regression model

$$Y_i = f(x_i) + \varepsilon_i, \quad i \in \{1, \ldots, n\}, \tag{1.1}$$

where for a given $\gamma \in \mathbb{N}$ and some unknown $\theta_f \in (0, 1)$, the regression function f is assumed to be γ times continuously differentiable on $[0, 1] \setminus \{\theta_f\}$, and the one-sided limits of $f^{(\gamma)}$ at θ_f exist and their difference $[f^{(\gamma)}]$ is non-zero. Such a jump discontinuity of the γ^{th} derivative is called a kink of order γ, and θ_f is the location and $[f^{(\gamma)}]$ is the size of the kink. Away from the kink, f is assumed to be Hölder smooth of some order $s \geq \gamma + 1$. Our purpose is to optimally estimate the parameters θ_f and $[f^{(\gamma)}]$, and to construct joint and marginal confidence sets which are honest and adaptive with respect to the smoothness parameter s over suitable subsets of the Hölder classes. In model (1.1), we shall assume that the $x_{i,n} = x_i$ are fixed, equidistant design points.

Change points and other irregularities such as kinks are important features of signals, and are of interest in various areas such as economics, medicine or the physical sciences. For example, in regression discontinuity or regression kink

*Supported by DFG Grant Ho 3260/5-1.
†Corresponding author.

designs (Card et al., 2015), the aim is to infer on the change of the level or the slope of an outcome variable from a policy change in level or slope of an assignment variable. For further reading on change point and edge detection and estimation we refer to the monographs of Carlstein et al. (1994), Korostelev and Tsybakov (1993) and Qiu (2005).

Korostelev (1988) obtained the optimal convergence rate for a change point in the white noise model over a nonparametric class of functions which are Lipschitz continuous away from the change point. In indirect estimation problems of change points, including deconvolution and kink estimation, Goldenshluger et al. (2006), Goldenshluger et al. (2008a) and Goldenshluger et al. (2008b) comprehensively studied optimal convergence rates over Sobolev-type classes in the white noise model, while Neumann (1997) considered a density deconvolution framework. Goldenshluger et al. (2006) construct their estimator of the change point in a deconvolution setting based on the zero-time-crossing technique. Cheng and Raimondo (2008) transfer this estimator to first order ($\gamma = 1$) kink estimation on compact intervals, and focus on the construction of appropriate kernel functions. Wishart (2009), Wishart and Kulik (2010) and Wishart (2011) studied convergence rates of zero-time-crossing estimators under long-range dependent errors together with corresponding lower bounds.

The asymptotic distribution of change point estimates and the construction of confidence intervals are discussed in Müller (1992); Loader (1996) among others for change points, and in Müller (1992); Eubank and Speckman (1994); Mallik et al. (2013) for kink estimation, in the latter paper even with dependent errors. In recent years, for nonparametric estimation problems the concepts of honest and adaptive confidence sets have been developed and intensively studied, see e.g. Li (1989); Low (1997); Cai and Low (2004); Giné and Nickl (2010). Confidence sets are called honest if they keep the level asymptotically uniformly over the function class under consideration, while they are called adaptive if the width is of the order of the minimax rate of estimation, up to logarithmic terms. For kink estimation, however, honest and adaptive confidence sets have apparently not yet been studied.

Our contributions in the present paper are as follows. We use the zero-crossing-time technique from Goldenshluger et al. (2006) and Cheng and Raimondo (2008) to construct estimates of the location θ_f as well as the size $[f^{(\gamma)}]$ of the kink in model (1.1). We derive optimal convergence rates over Hölder smoothness classes instead of the Sobolev-type smoothness classes in Goldenshluger et al. (2006). The proof techniques for the upper bounds differ somewhat from those in Goldenshluger et al. (2006), since we make more explicit use of the zero-crossing-time property of the estimate of the location. The lower bounds require the construction of new hypothesis functions which belong to the Hölder smoothness class. Further, we show joint asymptotic normality of the estimates, uniformly over the function classes, which allows for the construction of honest confidence sets for a given smoothness parameter s. In contrast to change point estimation, for kink estimation no additional bias arises in case of a discrete design. Finally, following Giné and Nickl (2010), based on Lepski's method we construct joint and marginal confidence sets which have an adaptive length in s

over appropriate subsets of the Hölder smoothness classes for a bounded range of the smoothness parameter s.

The paper is structured as follows. In Section 2 we introduce the estimators and present the optimal convergence rates. Section 3 contains the asymptotic normality of the estimates, and the construction of honest and adaptive confidence sets. In Section 4, we present results of a simulation study, and also give a real-data illustration to a series of annual global surface temperatures. Section 5 concludes, while outlines of the proofs are collected in Section 6. Section A of the appendix contains technical details of the proofs. In Section B we describe the construction of appropriate kernel functions and present further simulation results, in particular a comparison of the confidence intervals for the location of the kink with those in Mallik et al. (2013). Finally, in Section C and D, for the sake of completeness we recall a maximal deviation inequality for sub-Gaussian processes from Viens and Vizcarra (2007) and also collect results on weak convergence uniformly over families of probability measures.

2. Optimal kink estimation over Hölder classes

2.1. The function class

Let us introduce the function class over which we shall study the kink estimation problem in model (1.1). It is an adaptation of the function classes in definitions 1 and 2 in Goldenshluger et al. (2006), where Sobolev-type smoothness of the smooth-extension $g_{f,\gamma}$ is replaced by more conventional Hölder-smoothness.

Let C^k, $k \in \mathbb{N}_0$, denote the set of all real-valued, k-times continuously differentiable functions on $[0,1]$, and write $f^{(k)}(z) = \partial^k f(z)/\partial^k z$ to denote the k-th derivative of f. Given $s > 0$ we let $\lfloor s \rfloor = \max\{k \in \mathbb{N}_0 : \ k < s\}$, and we define the Hölder class on an interval $[a,b]$ with smoothness parameter $s > 0$ and Hölder-constant $L > 0$ by

$$\mathcal{H}^s([a,b],L) = \big\{ g \in C^{(\lfloor s \rfloor)} \,\big|\, |g^{(\lfloor s \rfloor)}(x) - g^{(\lfloor s \rfloor)}(y)| \le L\,|x-y|^{s-\lfloor s \rfloor},\ x,y \in [a,b] \big\}.$$

Given $\theta \in (0,1)$ and a continuous function $g : [0,1] \setminus \{\theta\} \to \mathbb{R}$ defined on $[0,1]$ except at θ, we denote the one-sided limits of g at θ by $g(\theta+) = \lim_{x \downarrow \theta} g(x)$, $g(\theta-) = \lim_{x \uparrow \theta} g(x)$ if these limits exist, in which case we let $[g](\theta) = g(\theta+) - g(\theta-)$ denote the jump-height at θ. We write $C^k(\{\theta\}^c)$ for the k-times continuously differentiable functions on $[0,1] \setminus \{\theta\}$, and for $L > 0$ let

$$\begin{aligned} \mathrm{Lip}(\{\theta\}^c, L) = \big\{ g \in C(\{\theta\}^c) \mid |g(x) - g(y)| \le L\,|x-y|,\ x,y \in [0,1] \setminus \{\theta\}, \\ x < y,\ \theta \notin (x,y) \big\}. \end{aligned}$$

Definition 2.1 (Regression function)**.** *Let $\gamma \in \mathbb{N}$, $s \in \mathbb{R}$ with $s \ge \gamma + 1$, let $a, L > 0$ and let $\Theta \subset (0,1)$ be a compact interval. Define the class of functions $f \in \mathcal{F}_s = \mathcal{F}_s(\gamma, a, \Theta, L)$ by assuming that $f \in C^{\gamma-1}[0,1]$, and that there is a unique $\theta_f \in \Theta$, called the location of the kink, such that* *(i)* $f^{(\gamma-1)} \in C^1(\{\theta_f\}^c)$, *and the jump height* $[f^{(\gamma)}] := [f^{(\gamma)}](\theta_f)$ *of* $f^{(\gamma)}$ *at* θ_f, *also called size of the kink, satisfies* $|[f^{(\gamma)}]| \geq a$,

(iia) *in case* $s = \gamma + 1$, *we have that* $f^{(\gamma)} \in \mathrm{Lip}(\{\theta_f\}^c, L)$,

(iib) *in case* $s > \gamma + 1$ *we actually assume that* $f^{(\gamma-1)} \in C^2(\{\theta_f\}^c)$ *with* $[f^{(\gamma+1)}](\theta_f) = 0$, *that the jump height of* $f^{(\gamma+1)}$ *is zero at* θ_f, *and that for the function*

$$g_{f,\gamma}(x) = \begin{cases} f^{(\gamma+1)}(x), & x \neq \theta_f, \\ f^{(\gamma+1)}(\theta_f+), & x = \theta_f, \end{cases}$$

we have that $g_{f,\gamma} \in \mathcal{H}^{s-(\gamma+1)}([0,1], L)$.

Remark 1. For $s > \gamma + 1$ our results actually can be shown over a slightly more general class, in which (iia) is assumed to hold, but the higher-order smoothness in (iib) is only assumed locally around the kink location θ_f.

2.2. The estimator

In this section estimators for the location and the size of the kink are introduced. Recall the motivation from Goldenshluger et al. (2006) that if $f^{(\gamma)}$ has a jump in θ_f, a smoothed version of $f^{(\gamma)}$ will have a large slope near θ_f, so that its first and second derivatives have a local maximum respectively a zero near θ_f. Following Goldenshluger et al. (2006) and Cheng and Raimondo (2008), for an appropriate kernel $K : \mathbb{R} \to \mathbb{R}$, specified in Assumption 2 below, and a bandwidth parameter $h > 0$ we introduce the probe functional

$$\psi_{h,f}(t) = h^{-(\gamma+1)} \int_0^1 f(x)\, K^{(\gamma+2)}\big(h^{-1}(x-t)\big)\, \mathrm{d}x, \tag{2.1}$$

which we estimate using a Priestley-Chao-type estimator for the fixed, equidistant design $x_{i,n}$,

$$\hat{\psi}_{h,n}(t) = n^{-1} h^{-(\gamma+1)} \sum_{i=1}^n Y_i\, K^{(\gamma+2)}(h^{-1}(x_i - t)). \tag{2.2}$$

Assumption 1 (Errors). The $\varepsilon_i = \varepsilon_{i,n}$ are centered, independent and identically distributed random variables with standard deviation $\sigma > 0$, and for any $u > 0$, $P(|\varepsilon_1| > u) \leq 2\exp(-2\,u^2/\sigma_g^2)$ for some $\sigma_g \geq \sigma$. ♢

Assumption 2 (Kernel). For parameters $\gamma, l \in \mathbb{N}$, suppose that the kernel $K : \mathbb{R} \to \mathbb{R}$ has support $\mathrm{supp}(K) = [-1,1]$, is $(\gamma+5)$-times differentiable inside its support and satisfies the following properties:

(i) $K^{(j)}(-1) = K^{(j)}(1) = 0, \quad j = 1, \ldots, \gamma+3$,

(ii) $K^{(1)}$ is an odd function, in particular $K^{(1)}(0) = 0$,

(iii) if $l \geq \gamma + 2$ then $\int_{-1}^1 x^m K^{(1)}(x)\mathrm{d}x = 0$ for $m = 0, \ldots, l - \gamma - 1$,

(iv) there are $0 < q_* < q_l < 1$ such that $K^{(1)}(x) > 0$ for $x \in [-q_l, 0)$ and $K^{(1)}$ has a unique global maximum at $-q_*$,
(v) for some $x^* \in (0,1)$ and $c_2 > 0$ we have that $|K^{(1)}(x)| \geq c_2|x|$, $x \in [-x^*, x^*]$. ♢

Remark (*Discussion of Assumption 2*). Assumption 2 is similar but more restrictive than assumption $C_{1,s}$ in Cheng and Raimondo (2008) or assumption 2 in Goldenshluger et al. (2006). In particular, Assumption 2, (v) determines the separation rate, Lemma A.1. In Section B.1 we provide an explicit construction of kernels satisfying Assumption 2 for $\gamma = 1$ and given l, and indicate how to extend it to the case $\gamma \geq 2$. ⋄

For the estimation of the location of the kink we proceed in two stages. First, an interval is constructed which contains the kink with high probability. Second, the kink is estimated by a zero of the empirical probe functional inside this interval.

In the following we shall always impose Assumption 1 and assume that the regression function f in model (1.1) satisfies $f \in \mathcal{F}_s$ as specified in Definition 2.1, and that the fixed kernel K satisfies Assumption 2 with parameters γ, $l = \lfloor s \rfloor$, and q_*, q_l and x^* in (iv) and (v).

Given $h > 0$ we then let

$$t^* = t^*(h; f) = \theta_f + hq_*, \qquad t_* = t_*(h; f) = \theta_f - hq_*. \tag{2.3}$$

Lemma 2.2. *There is an $h_0 > 0$ with $\Theta \subset [h_0, 1 - h_0]$ such that*

$$\text{for } h_0 \geq h > 0 \quad \text{there is a } \tilde{\theta} = \tilde{\theta}_{h,f} \in [t_*(h; f), t^*(h; f)] \quad \text{such that} \quad \psi_{h,f}(\tilde{\theta}) = 0. \tag{2.4}$$

Here h_0 can be chosen uniformly over $f \in \mathcal{F}_s$ and depending only on the kernel K as well as on the Lipschitz constant L and the set Θ of $\mathcal{F}_s$. In particular, we have that $|\tilde{\theta}_{h,f} - \theta_f| = O(h)$ for $h \in (0, h_0)$, uniformly over $\mathcal{F}_s$.

The proof is given in Section 6.1. Lemma 2.2 motivates the two stages of the estimation procedure: first estimate the parameters t_* and t^*, second estimate the kink-location θ_f as a zero of the empiral probe functional in the resulting interval. Thus, let

$$\begin{aligned} \hat{t}_* &= \hat{t}_*(h; n) = \min\{\arg\min_t \hat{\psi}_{h,n}(t), \arg\max_t \hat{\psi}_{h,n}(t)\}, \\ \hat{t}^* &= \hat{t}^*(h; n) = \max\{\arg\min_t \hat{\psi}_{h,n}(t), \arg\max_t \hat{\psi}_{h,n}(t)\}, \end{aligned} \tag{2.5}$$

and define the estimator for the kink-location by

$$\hat{\theta}_{h,n} \in \begin{cases} \{t \in [\hat{t}_*, \hat{t}^*] \mid \hat{\psi}_{h,n}(t) = 0\}, & \text{if the set is not empty,} \\ \{\frac{\hat{t}_* + \hat{t}^*}{2}\}, & \text{otherwise.} \end{cases} \tag{2.6}$$

In the proofs we will show that $\{t \in [\hat{t}_*, \hat{t}^*] \mid \hat{\psi}_{h,n}(t) = 0\} \neq \emptyset$ holds with high probability uniformly over $\mathcal{F}_s$, and consequently only this part of (2.6) is asymptotically relevant.

For the kink-size $[f^{(\gamma)}]$, the expansion

$$[f^{(\gamma)}] = (-1)^{\gamma+2}\, h\, \frac{\psi^{(1)}_{h,f}(\theta_f)}{K^{(2)}(0)} + O(h^{s-\gamma}),$$

see (6.5), suggests the following estimate

$$\widehat{[f^{(\gamma)}]}_{h,n} := h\, \frac{\hat{\psi}^{(1)}_{h,n}(\hat{\theta}_{h,n})}{(-1)^{\gamma+2} K^{(2)}(0)} \tag{2.7}$$

with deterministic version

$$\widetilde{[f^{(\gamma)}]}_{h} := h\, \frac{\psi^{(1)}_{h,f}(\tilde{\theta}_{h,f})}{(-1)^{\gamma+2} K^{(2)}(0)}, \tag{2.8}$$

where $\tilde{\theta}_{h,f}$ is defined in (2.4).

2.3. Optimal rates of convergence

To express the uniformity in the parameter f, we introduce the following notation. If $\eta_{h,f}$ is real-valued depending on parameters $f \in \mathcal{F}$ and $h \in (0, h_0)$ for some $h_0 > 0$, and if $\beta \in \mathbb{R}$, we write $\eta_{h,f} = O_{\mathcal{F}}(h^\beta)$ if $\sup_{f\in\mathcal{F}} \big|\eta_{h,f}\big| \le C\, h^\beta$ for some constant $C > 0$. Similarly, we write $\eta_{h,f} = o_{\mathcal{F}}(h^\beta)$ if $h^{-\beta} \sup_{f\in\mathcal{F}} \big|\eta_{h,f}\big| \to 0$, $h \to 0$.

Let $\hat{\eta}_{h,n}$ be random and depending on some $h \in (0, h_0)$ and the data in model (1.1) for the sample size n. We denote P_f to stress the dependence of the distribution on the parameter f. Then we write $\hat{\eta}_{h,n} = O_{P,\mathcal{F}}(n^\beta\, h^\alpha)$, $\alpha, \beta \in \mathbb{R}$, for uniform boundedness in probability, that is

$$\lim_{C\to\infty} \limsup_{n\to\infty} \sup_{f\in\mathcal{F}} \sup_{h\in(0,h_0)} P_f\big(\big|\hat{\eta}_{h,n}\big| \ge C\, n^\beta\, h^\alpha\big) = 0.$$

Similarly, write $\hat{\eta}_{h,n} = o_{P,\mathcal{F}}(n^\beta\, h^\alpha)$ if

$$\forall\, \epsilon > 0: \quad \limsup_{n\to\infty} \sup_{f\in\mathcal{F}} \sup_{h\in(0,h_0)} P_f\big(\big|\hat{\eta}_{h,n}\big| \ge \epsilon\, n^\beta\, h^\alpha\big) = 0.$$

In the following theorem we obtain convergence rates for the estimates of the location and the size of the kink. For two sequences (a_n) and (b_n) we write $a_n \cong b_n$ if $C_1 \le |a_n/b_n| \le C_2$ for $n \ge n_0$ and some constants $0 < C_1 < C_2$.

Theorem 2.3. *Consider model (1.1) and suppose that Assumption 1 as well as Assumption 2 with $l = \lfloor s \rfloor$ hold true. Then there exist finite constants $h_0, C > 0$ depending only on the kernel K, on σ and σ_g in Assumption 1 as well as on L and Θ of $\mathcal{F}_s = \mathcal{F}_s(\gamma, a, \Theta, L)$ such that if $h \in (0, h_0)$ and n are such that $nh^{2\gamma+1} \ge C \log(1/h)$, then*

$$\hat{\theta}_{h,n} - \theta_f = O_{P,\mathcal{F}_s}(h^{s-\gamma+1}) + O_{P,\mathcal{F}_s}((nh^{2\gamma-1})^{-1/2}) \tag{2.9}$$

and

$$\widehat{[f^{(\gamma)}]}_{h,n} - [f^{(\gamma)}] = O_{P,\mathcal{F}_s}((nh^{2\gamma+1})^{-1/2}) + O_{P,\mathcal{F}_s}(h^{s-\gamma}), \tag{2.10}$$

where the constants in the O-terms depend only on K, σ_g and L, and can be chosen uniformly over a bounded range of values of s.

Moreover, choosing h of order $n^{-1/(2s+1)}$ we obtain the convergence rates

$$\hat{\theta}_{h,n} - \theta_f = O_{P,\mathcal{F}_s}\big(n^{-(s-\gamma+1)/(2s+1)}\big), \quad \widehat{[f^{(\gamma)}]}_{h,n} - [f^{(\gamma)}] = O_{P,\mathcal{F}_s}\big(n^{-(s-\gamma)/(2s+1)}\big).$$

The proof is provided in Section 6.2. The next theorem shows that these rates are indeed optimal.

Theorem 2.4. *Let $\gamma \in \mathbb{N}$, $s \in \mathbb{R}$ with $s \geq \gamma+1$, let $a, L > 0$ and let $\Theta \subset (0,1)$ be a compact set. Then, in model (1.1) with $\varepsilon_i \sim N(0, \sigma^2)$, setting $w(x) = x/(1+x)$ it holds that*

$$\liminf_n \inf_{\hat{\theta}} \sup_{f \in \mathcal{F}_s} \mathbb{E}_f \left[w\big(n^{(s-\gamma+1)/(2s+1)} |\hat{\theta} - \theta_f|\big)\right] > 0, \tag{2.11}$$

$$\liminf_n \inf_{\hat{\theta}} \sup_{f \in \mathcal{F}_s} \mathbb{E}_f \left[w\big(n^{(s-\gamma)/(2s+1)} |\hat{\theta} - [f^{(\gamma)}]|\big)\right] > 0, \tag{2.12}$$

where w metrizes convergence in probability and the infimum is taken over all possible estimators $\hat{\theta}$ respectively.

The proof is provided in Section 6.3.

Remark (Convergence rates)*.* The convergence rates for the location of the kink in Theorem 2.3 correspond to those in Theorem 1 in Goldenshluger et al. (2006) for their function class in Definition 2, which instead of Hölder-smoothness requires that

$$\int_{\mathbb{R}} \Big| \int_{\mathbb{R}} g_{f,\gamma}(x) \exp(2\pi\omega x)\, \mathrm{d}x \Big| \, |\omega|^{m-1}\, \mathrm{d}\omega \leq L$$

for some $m > 1$. Indeed, their $m-1$ (smoothness parameter) and β (degree of ill-posedness) correspond to $s-(\gamma+1)$ resp. γ in our setting, so that $m = s-\gamma$ and $\beta = \gamma$ transforms their rate $n^{-(m+1)/(2m+2\beta+1)}$ into $n^{-(s-\gamma+1)/(2s+1)}$. Goldenshluger et al. (2008a) provide minimax rates for more conventional Sobolev-type classes and attain a rate which corresponds to $n^{-(s-\gamma+1/2)/2s}$ in our setting. As $\gamma \geq 1/2$ this rate is inferior compared to ours, which is in line with the difference between the optimal rates of pointwise estimation for Sobolev- and Hölder smoothness classes. Moreover, similar considerations for our kink-size estimate and the jump amplitude estimator in Goldenshluger et al. (2008a) lead to related observations concerning the optimal rates. Also note that the rate of convergence for the size of the kink corresponds to the minimax rate when estimating the γ-th derivative at a given point.

3. Asymptotic confidence sets

3.1. Asymptotic normality

The next theorem establishes joint asymptotic normality of the estimates of the location and the size of the kink, (2.6) and (2.7), around their deterministic counterparts (2.4) and (2.8), respectively.

Theorem 3.1. *In model (1.1) under the Assumptions 1 and 2 with $l = \lfloor s \rfloor$, if h and n are such that*

$$nh^{2\gamma+1}\log(1/h)^{-1} \to \infty, \quad \text{and} \quad nh^{4s-2\gamma+1} \to 0, \tag{3.1}$$

then for $\mathbf{x} \in \mathbb{R}^2$ we have that

$$\sup_{f\in\mathcal{F}_s}\left| P_f\left[\begin{pmatrix} \tilde{w}_n^{loc}(h)^{-1}\,(\hat{\theta}_{h,n} - \tilde{\theta}_{h,f}) \\ \tilde{w}_n^{size}(h)^{-1}\left(\widehat{[f^{(\gamma)}]}_{h,n} - \widetilde{[f^{(\gamma)}]}_h\right) \end{pmatrix} \le \mathbf{x}\right] - \Phi_2(\mathbf{x})\right| = o(1), \tag{3.2}$$

where Φ_2 denotes the bivariate standard normal distribution function and the asymptotic standard deviations for the estimates of location and size of the kink are given by

$$\tilde{w}_n^{loc}(h) = \frac{\sigma||\, K^{(\gamma+2)}\,||_2}{\sqrt{nh^{2\gamma-1}}\,[f^{(\gamma)}]\;K^{(2)}(0)}, \quad \textit{resp.} \quad \tilde{w}_n^{size}(h) = \frac{\sigma||\, K^{(\gamma+3)}\,||_2}{\sqrt{nh^{2\gamma+1}}\,K^{(2)}(0)}. \tag{3.3}$$

Here, $\|\cdot\|_2$ denotes the L_2-norm of a function on the interval $[0,1]$. Section 6.4 is devoted to the proof of the theorem.

Remark (Undersmoothing). Using similar arguments as in Theorem 2.3 one obtains the bounds

$$\tilde{\theta}_{h,f} - \theta_f = O_{\mathcal{F}_s}(h^{s-\gamma+1}), \qquad \widetilde{[f^{(\gamma)}]}_h - [f^{(\gamma)}] = O_{\mathcal{F}_s}(h^{s-\gamma}). \tag{3.4}$$

Thus, using undersmoothing, that is choosing $h \cong n^{-1/(2s+1)}\log(n)^{\zeta}$ for $\zeta < 0$, the asymptotic normality in (3.2) even holds uniformly over $\mathcal{F}_s$ for $\tilde{\theta}_{h,f}$ replaced by θ_f and $\widetilde{[f^{(\gamma)}]}_h$ by $[f^{(\gamma)}]$. Thus, Theorem 3.1 can be directly used to construct honest confidence sets for the parameters $(\theta_f, [f^{(\gamma)}])$ over $\mathcal{F}_s$.

3.2. Adaptive confidence sets

We briefly recall the definitions of *honest* and *adaptive* confidence intervals tailored to our framework by following Li (1989) resp. Cai and Low (2004).

Let $\mathcal{F} = \bigcup_{s\in\mathcal{S}}\mathcal{F}_s$, with $\mathcal{S} \subset \mathbb{R}$. A family of random intervals $\big([\,\underline{C}_n(\alpha), \overline{C}_n(\alpha)\,]\big)_{\alpha\in[0,1]}$ is called

- *honest confidence interval for θ_f* if for any $\alpha \in (0,1)$ it holds that

$$\liminf_n \inf_{f\in\mathcal{F}} P_f\big(\theta_f \in [\,\underline{C}_n(\alpha), \overline{C}_n(\alpha)\,]\big) \ge 1 - \alpha. \tag{3.5}$$

- *adaptive confidence interval over the parameter space* $\mathcal{F}$ if for every $s \in \mathcal{S}$ and $\varepsilon > 0$ there exists some constant $C = C(\alpha) > 0$ depending only on $\alpha \in (0,1)$ such that

$$\sup_{f \in \mathcal{F}_s} P_f\big(\overline{C}_n(\alpha) - \underline{C}_n(\alpha) \geq C\,\tilde{r}_n(s)\big) \leq \varepsilon, \tag{3.6}$$

  where $\tilde{r}_n(s)$ equals (up to a logarithmic term) the minimax rate of estimation θ_f over $\mathcal{F}_s$.

It is straightforward to extend the definitions to a bivariate confidence set for the parameter pair $(\theta_f, [f^{(\gamma)}])$.

Low (1997) and Cai and Low (2004) showed that honest and adaptive pointwise confidence intervals over Hölder classes do not exist. In the context of confidence bands, Giné and Nickl (2010) showed that the construction of honest and adaptive confidence bands in density estimation becomes possible by slightly reducing the function classes.

We shall follow their lead and construct honest and adaptive confidence sets for the bivariate parameter consisting of location and size of the kink. To this end, let $\gamma \in \mathbb{N}$ and $\underline{s}, \overline{s} \in \mathbb{R}_+$ be such that $\gamma + 1 \leq \underline{s} < \overline{s}$. Choose integers $k_{\min,n}$ and $k_{\max,n}$ such that

$$2^{-k_{\min,n}} \cong \Big(\frac{\log(n)}{n}\Big)^{\frac{1}{2\overline{s}+1}}, \quad 2^{-k_{\max,n}} \cong \Big(\frac{\log(n)^2}{n}\Big)^{\frac{1}{2\gamma+1}}, \tag{3.7}$$

and set $\mathcal{K}_n = [k_{\min,n}, k_{\max,n}] \cap \mathbb{N}$ as well as

$$h_k = 2^{-k}, \quad k \in \mathcal{K}_n. \tag{3.8}$$

Definition 3.2. *Let* $k_0 \in \mathbb{N}$, $0 < b_1 < b_2$, $a, L > 0$ *and let* $\Theta \subset (0,1)$ *be a compact set. Then define*

$$\widetilde{\mathcal{F}} = \widetilde{\mathcal{F}}(\underline{s}, \overline{s}, b_1, b_2, k_0, \gamma, a, \Theta, L) = \bigcup_{s \in [\underline{s}, \overline{s}]} \widetilde{\mathcal{F}_s}(b_1, b_2, k_0, \gamma, a, \Theta, L), \tag{3.9}$$

where

$$\begin{aligned} \widetilde{\mathcal{F}_s} &= \widetilde{\mathcal{F}_s}(b_1, b_2, k_0, \gamma, a, \Theta, L) \\ &= \{f \in \mathcal{F}_s(\gamma, a, \Theta, L) \,|\, b_1 h_k^{s-\gamma} \leq |\psi_{h_k,f}(\theta_f)| \leq b_2 h_k^{s-\gamma} \quad \forall k \geq k_0\} \end{aligned} \tag{3.10}$$

and where the kernel of the probe functional in (3.10) satisfies Assumption 2 with parameters γ *and* $l = \lfloor \overline{s} \rfloor + 1$. *Given* $f \in \widetilde{\mathcal{F}}$ *let* s_f *be the unique value of* s *for which* f *fulfills the bias condition in (3.10).*

For wavelet density estimation over Hölder classes, Giné and Nickl (2010) show that the minimax rates of estimation remain the same when reducing the function class in a similar fashion as in (3.9), see also the discussion in Bull (2012). In our present context, while a general result eludes us, in Section A.5 we show that the minimax rates over $\widetilde{\mathcal{F}_s}$ correspond to those over $\mathcal{F}_s$ at least for some values of the smoothness parameter $s \in [\underline{s}, \overline{s}]$.

To construct confidence sets, we divide the sample $Y_1, \ldots, Y_n$ into the two parts $S_1 = \{Y_1, Y_3 \ldots, Y_{n-1}\}$ and $S_2 = \{Y_2, Y_4, \ldots, Y_n\}$ if n is even, and similarly if n is odd. In particular, the sizes $n_j = |S_j|$ satisfy $n_j \cong n$ for $j = 1, 2$. Note that over the subsamples, (1.1) still holds with an equidistant design.

We shall use S_2 for selection of the bandwidth parameter h based on Lepski's method. For a sufficiently large constant $C_{\mathrm{Lep}} > 0$ (specified in the proof of Lemma 6.8) we let

$$\hat{k}_n = \min\{k \in \mathcal{K}_n \mid |\hat{\theta}_{h_k, n_2} - \hat{\theta}_{h_l, n_2}| \le C_{\mathrm{Lep}} \sqrt{\log(n_2)/n_2 h_l^{2\gamma-1}} \,\forall\, l > k, \quad l \in \mathcal{K}_n\}. \tag{3.11}$$

A central technical result, Lemma 6.8, states that for a function $f \in \widetilde{\mathcal{F}}$, $h_{\hat{k}_n}$ is of order $\left(\log(n_2)/n_2\right)^{1/(2s_f+1)}$ with high probability uniformly over $f \in \widetilde{\mathcal{F}}$.

We then employ undersmoothing, that is we choose the bandwidth for the estimation of θ_f as $h_{\hat{k}_n+u_n}$, and for the estimation of $[f^{(\gamma)}]$ as $h_{\hat{k}_n+v_n}$, where we assume that

$$u_n, v_n \in \mathbb{N}, \qquad h_{u_n} \cong \log(n)^{-1/(2\gamma-1)} \quad \text{and} \quad h_{v_n} \cong \log(n)^{-1/(2\gamma+1)}. \tag{3.12}$$

Furthermore, let $\hat{\sigma}_{n_1}$ be an estimate of σ based on the sample S_1 which satisfies

$$\hat{\sigma}_{n_1} - \sigma = o_{P,\widetilde{\mathcal{F}}}(1). \tag{3.13}$$

The estimates in Hall et al. (1990) or Dette et al. (1998) fulfill this assumption. Consider the estimates

$$\hat{w}^{loc} = \frac{\hat{\sigma}_{n_1} \,||\, K^{(\gamma+2)} \,||_2}{\sqrt{n_1 h_{\hat{k}_n+u_n}^{2\gamma-1}}\; \widehat{[f^{(\gamma)}]}_{h_{\hat{k}_n}, n_1} K^{(2)}(0)} \tag{3.14}$$

of the asymptotic standard deviation of the kink in (3.3), and

$$\hat{w}^{size} = \frac{\hat{\sigma}_{n_1} \,||\, K^{(\gamma+3)} \,||_2}{\sqrt{n_1 h_{\hat{k}_n+v_n}^{2\gamma+1}}\; K^{(2)}(0)} \tag{3.15}$$

of the asymptotic standard deviation of the size of the kink.

Given $\alpha \in (0, 1)$ let $q_\alpha(W)$ denote the α-quantile of $W = \max\{|X_1|, |X_2|\}$ for two independent standard normal random variables X_1 and X_2. Consider the rectangular confidence region

$$C_n^{loc}(\alpha) \times C_n^{size}(\alpha)$$

for the parameter $(\theta_f, [f^{(\gamma)}])$, where

$$C_n^{loc}(\alpha) = \big[\hat{\theta}_{h_{\hat{k}_n+u_n}, n_1} - \hat{w}^{loc}\, q_{1-\alpha}(W), \hat{\theta}_{h_{\hat{k}_n+u_n}, n_1} + \hat{w}^{loc}\, q_{1-\alpha}(W)\big], \tag{3.16}$$

$$C_n^{size}(\alpha) = \Big[\widehat{[f^{(\gamma)}]}_{h_{\hat{k}_n+v_n}, n_1} - \hat{w}^{size}\, q_{1-\alpha}(W), \widehat{[f^{(\gamma)}]}_{h_{\hat{k}_n+v_n}, n_1} + \hat{w}^{size}\, q_{1-\alpha}(W)\Big].$$

Theorem 3.3. *Consider model (1.1) under Assumption 1 and the function class $\widetilde{\mathcal{F}}$ in (3.9), and let K be a kernel satisfying Assumption 2 with γ and $l = \lfloor \overline{s} \rfloor + 1$. Then for any $\alpha \in (0,1)$,*

$$\lim_{n\to\infty} \sup_{f\in\widetilde{\mathcal{F}}} \Big| P_f\big((\theta_f, [f^{(\gamma)}])^T \in C_n^{loc}(\alpha) \times C_n^{size}(\alpha)\big) - \big(1-\alpha\big)\Big| = 0. \tag{3.17}$$

Furthermore, there exists a finite constant $C > 0$ such that

$$\lim_{n\to\infty} \sup_{f\in\widetilde{\mathcal{F}}} \Big[P_f\Big(\hat{w}^{loc} \geq C\Big(\frac{\log(n)}{n}\Big)^{\frac{s_f-\gamma+1}{2s_f+1}}\Big) + P_f\Big(\hat{w}^{size} \geq C\Big(\frac{\log(n)}{n}\Big)^{\frac{s_f-\gamma}{2s_f+1}}\Big)\Big] = 0. \tag{3.18}$$

The proof is provided in Section 6.5.

Remark. 1.Equation (3.17) shows asymptotic honesty of the confidence sets as defined in (3.5), while the adaptivity of the confidence sets as defined in (3.6) is covered by (3.18).

2. The choice of the bandwidth (3.11) is only based on the estimate of the location of the kink, but is then also used for constructing the confidence set of the size. This is possible since the optimal bandwidth resolution is the same for both estimates, see Theorem 2.3.

3. Marginal adaptive confidence intervals for either θ_f or $[f^{(\gamma)}]$ can be constructed in an analogous way (see 4.1).

4. Simulations and real data illustration

In this section we investigate the finite sample properties of the confidence sets in (3.16) as well as of the following marginal confidence intervals for θ_f,

$$\tilde{C}_n^{loc}(\alpha) = \Big[\hat{\theta}_{h_{\hat{k}_n+u_n},n_1} - \hat{w}^{loc}\, q_{1-\alpha/2}(N(0,1)), \hat{\theta}_{h_{\hat{k}_n+u_n},n_1} + \hat{w}^{loc}\, q_{1-\alpha/2}(N(0,1))\Big]. \tag{4.1}$$

The kernel K is chosen as in Section B.1 with $\gamma = 1$ and $l = 2$ so that

$$K^{(1)}(x) = \frac{8316}{832}\Big(-\frac{1}{12}x + \frac{1}{3}x^3 - \frac{1}{2}x^5 + \frac{1}{3}x^7 - \frac{1}{12}x^9\Big) 1_{[-1,1]}(x). \tag{4.2}$$

Figure 1 illustrates the first three derivatives of K. Subsection 4.1 gives detailed numerical illustrations of our methods, while Subsection 4.2 contains a real data illustration to a series of global surface temperatures. A comparison of marginal confidence intervals for the kink location with the method of Mallik et al. (2013) is presented in the appendix, Section B.2.

4.1. Numerical experiments

We consider the following two regression functions

$$f_1(x) = -2(x-\theta_n)1_{[0,\theta_n]}(x), \qquad \theta_n = 1/2 + {}^1\!/_{3n},$$

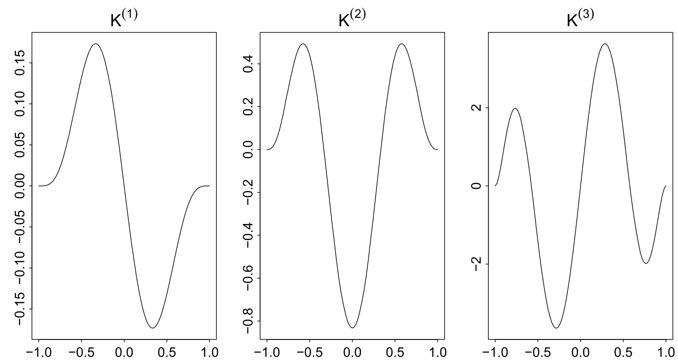

FIGURE 1. *From left to right:* $K^{(1)}$, $K^{(2)}$ *and* $K^{(3)}$ *given by (4.2).*

$$f_2(x) = |\tilde{f}^{(-2)}_{9/10,7}(x - 1/3n)|, \quad \tilde{f}_{c_1,c_2}(x) = \sum_{k=0}^{\infty} c_1^k \cos(c_2^k \pi x) 1_{[0,1]}(x). \tag{4.3}$$

Here, f_1 has a kink at θ_n of size $[f_1^{(1)}] = -2$ with infinite smoothness s outside the kink. The offset $1/3n$ is chosen so that the kink is not located on a design point $x_k = k/n$.

The regression function f_2 in (4.3) is defined as the absolute value of the second anti-derivative of the Weierstraß-function $\tilde{f}_{c_1,c_2}$ with vanishing affine linear part. By Hölder continuity of the Weierstraß-function with exponent $0 < -\log(c_1)/\log(c_2) < 1$ we have that $s = 2 - \log(c_1)/\log(c_2)$ as well as $[f_2^{(1)}] \approx 9/4$.

We use errors $\varepsilon \sim SN(\zeta, \omega, \gamma)$, that is a skew normal distribution with shape parameter $\alpha = -3$, and choose the location parameter $\zeta \in \mathbb{R}$ and the scale-parameter $\omega > 0$ so that $\mathbb{E}[\varepsilon] = 0$ and $\mathbb{E}[\varepsilon^2] = \sigma^2 = 0.2^2$. Figure 2 displays

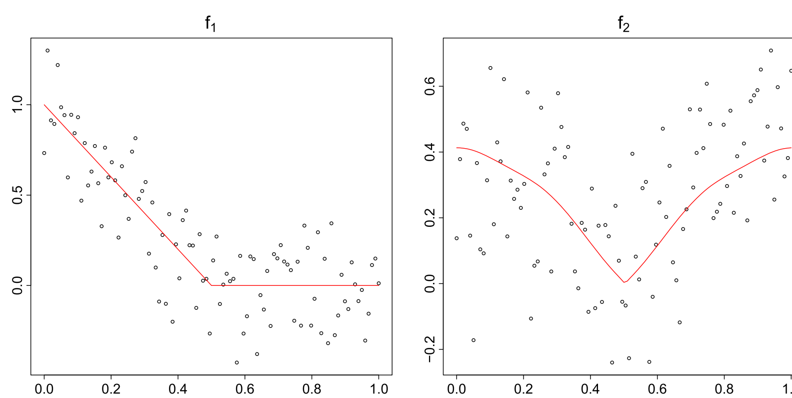

FIGURE 2. *Left:* f_1 *with noisy observations. Right:* f_2 *with noisy observations. The noise level is* $\sigma = 0.2$ *and the grid size* $n = 100$ *respectively.*

the regression functions together with samples of size $n = 100$. For the Lepski-scheme we use a grid inside the intervals $[h_{min,n}, h_{max,n}]$ as specified in Table 1, and choose the Lepski-constants in (3.11) as $C_{\text{Lep}} = 0.08$ for f_1 and as $C_{\text{Lep}} = 0.001$ for f_2. Simulations are based on a $m = 10000$ repetitions. The

Table 1
Choice of $[h_{min,n}, h_{max,n}]$ for the first scenario.

n	500	1000	2000	4000	8000
f_1	[0.49,0.55]	[0.42,0.51]	[0.39,0.45]	[0.34,0.41]	[0.29,0.39]
f_2	[0.32,0.35]	[0.31,0.34]	[0.24,0.28]	[0.22,0.26]	[0.21,0.25]

noise-level σ^2 is estimated by the simple Neumann estimator (Von Neumann, 1941)

$$\hat{\sigma}_n^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (Y_{i+1} - Y_i)^2.$$

First we investigate the accuracy of our estimates in terms of the square roots of the Mean Squared Error (RMSE) for the sample sizes $n \in \{500, 1000, 2000, 4000, 8000\}$. The results are summarized in Table 2. As can be expected from the rates in Theorem 2.3, estimates of the location of the kink are more precise than for its size. Further, in particular for f_2 the estimate of the size converges slowly.

Table 2
RMSE of the kink-location (2.6) resp. kink-size estimate (2.7) based on the Lepski choice for $\sigma = 0.2$.

n	500	1000	2000	4000	8000	true value
θ_{f_1}	0.0065	0.0003	0.0000	0.0000	0.0000	1/2 + 1/3n
$[f_1^{(1)}]$	0.6821	0.1780	0.1035	0.0727	0.0529	-2
θ_{f_2}	0.0004	0.0003	0.0007	0.0002	0.0001	1/2 - 1/3n
$[f_2^{(1)}]$	2.0778	1.3389	0.8222	0.4695	0.3001	2.25

Next we investigate the confidence sets for the location of the kink, (4.1), as well as for the joint confidence sets for location and size in (3.16) in terms of coverage and average length. The results are displayed in Tables 3 and 4. For the location of the kink in Table 3, coverage is already satisfactory for both regression functions for a sample of size $n = 500$. In contrast, for the joint confidence sets, the coverage is quite below the nominal level for sample sizes $n = 500$ and $n = 1000$, in particular for f_2. Moreover, the interval for the size of the kink is rather wide for these values of the sample size, even including zero. For larger sample sizes, the performance improves notably.

Finally, we investigate the ratio of the empirical bias and the empirical standard deviation. In contrast to change point estimation with fixed design, for kink estimation the discretization bias is asymptotically negligible, which can also be seen numerically in Table 5. Note that these findings also indicate the undersmoothing effect of the sequences u_n and v_n.

TABLE 3
Average coverage and length of the confidence intervals for the kink-location $\tilde{C}_n^{loc}$ in (4.1) for f_1 and f_2.

	90% nominal coverage		95% nominal coverage		99% nominal coverage	
	coverage	length	coverage	length	coverage	length
$n = 500$						
f_1	0.86	0.052	0.93	0.062	0.99	0.082
f_2	0.86	0.056	0.91	0.067	0.97	0.088
$n = 2000$						
f_1	0.89	0.040	0.94	0.048	0.99	0.063
f_2	0.89	0.037	0.94	0.044	0.98	0.058
$n = 8000$						
f_1	0.90	0.022	0.95	0.027	0.99	0.035
f_2	0.90	0.025	0.94	0.029	0.99	0.039

TABLE 4
Average coverage and marginal lengths of the joint confidence sets for the kink-location and -size in (3.16) for f_1 and f_2.

	90% nominal coverage			95% nominal coverage			99% nominal coverage		
	coverage	length size	length kink	coverage	length size	length kink	coverage	length size	length kink
$n = 500$									
f_1	0.69	1.914	0.062	0.91	2.197	0.071	0.99	2.737	0.089
f_2	0.70	3.991	0.067	0.79	4.581	0.076	0.92	5.707	0.095
$n = 1000$									
f_1	0.88	1.538	0.063	0.93	1.766	0.072	0.98	2.120	0.090
f_2	0.67	2.907	0.050	0.77	3.336	0.058	0.90	4.156	0.072
$n = 2000$									
f_1	0.89	1.262	0.048	0.94	1.448	0.055	0.99	1.804	0.069
f_2	0.79	2.817	0.044	0.87	3.233	0.051	0.95	4.028	0.063
$n = 4000$									
f_1	0.89	1.040	0.036	0.95	1.193	0.041	0.99	1.487	0.051
f_2	0.85	2.215	0.034	0.91	2.542	0.039	0.98	3.167	0.049
$n = 8000$									
f_1	0.89	0.859	0.027	0.94	0.986	0.030	0.99	1.229	0.038
f_2	0.88	2.069	0.029	0.94	2.375	0.034	0.98	2.958	0.042

TABLE 5
Ratios between empirical bias and standard deviation of the estimates $\hat{\theta}_{h_{\hat{k}_n+u_n}}$ and $\widehat{[f^{(\gamma)}]}_{h_{\hat{k}_n+v_n}}$.

n	500	1000	2000	4000	8000
θ_{f_1}	0.357	0.019	0.003	0.001	0.001
$[f_1^{(1)}]$	1.241	0.241	0.099	0.070	0.055
θ_{f_2}	0.021	0.021	0.061	0.022	0.031
$[f_2^{(1)}]$	1.060	1.149	0.728	0.571	0.312

4.2. Illustration to series of global surface temperature

We illustrate our method in an application to a series of changes in annual global surface temperature in degree Celsius from 1880 to 2017 relative to the average temperature for 1951 − 1980, see Figure 3. The series is available at `https://data.giss.nasa.gov/gistemp`, where further details on the data are provided. We used a grid of bandwidths inside the range $[0.05, 0.3]$.

The result are somewhat sensitive to the choice of the Lepski constant C_{Lep} in (3.11). For a value between 0.2 and 0.6, the kink location is around 1984 with a confidence interval of 11 years length at nominal level 90%, while for larger

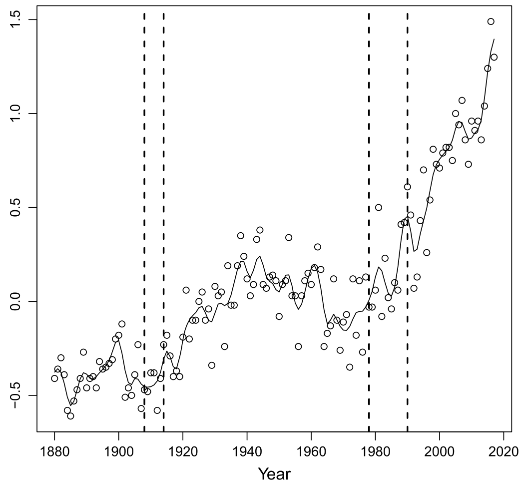

FIGURE 3. *Smooth line: Smoothed curve of the data points. Dotted lines: Adaptive 95 % confidence intervals for the global surface temperature data for different choices of the Lepski constant in (3.11).*

than 1.1, the kink location is estimated at 1912, with a confidence interval of 6 years length. Results are summarized in Table 6.

TABLE 6
Confidence intervals $\tilde{C}_n^{loc}$ in (4.1) for different significance levels and choices of the Lepski constant for the global surface temperature dataset.

C_{Lep}	90%	95%	99%	$\widehat{[f^{(\gamma)}]}_{h_{\hat{k}_n+v_n}}$	$h_{\hat{k}_n}$
0.2 – 0.6	[1979,1989]	[1978,1990]	[1977,1991]	6.467	0.100
≥ 1.1	[1909,1914]	[1908,1914]	[1908,1915]	9.314	0.195

5. Discussion

In this paper we suggest a method to construct confidence intervals for the kink location and kink size over Hölder classes without requiring a shape restriction on the regression function as in Mallik et al. (2013). The rate of convergence derived in this paper is achieved over a more conventional Hölder class which allows for local higher-order smoothness, in contrast to the nonstandard smoothness class in Goldenshluger et al. (2006) based on integrability of the Fourier transform.

Recently, there has been quite some work on change point detection and segmentation methods (Frick et al., 2014; Haynes et al., 2017). Analogous results for kink detection and corresponding segmentation algorithms would be of quite some applied interest for example in environmental or pharmacological studies, in which the function of interest can reveal changes of trend through kinks rather than jumps.

6. Proofs

We consider model (1.1) and impose the Assumptions 1 and 2 throughout this section. We shall say that a sequence of events A_n holds with high probability uniformly over $\mathcal{F}_s$ if

$$\inf_{f\in\mathcal{F}_s} P_f(A_n) \to 1, \quad n\to\infty.$$

6.1. Properties of the probe functional and first stage estimates

The proofs of the lemmas in this section are provided in Section A.1 in the appendix. Recall the definition (2.1) of the probe functional $\psi_{h,f}(t)$.

Lemma 6.1. *If $h_0 > 0$ is such that $\Theta \subset [h_0, 1-h_0]$, then for $j = 0,1,2$ and for $h \in (0,h_0)$ it holds that*

$$\begin{aligned} \psi_{h,f}^{(j)}(t) &= L_{h,j}(t) + O_{f\in\mathcal{F}_s, t\in[h,1-h]}(h^{s-\gamma-j}),\\ L_{h,j}(t) &= (-1)^{\gamma+1+j}\, h^{-j}\,[f^{(\gamma)}]\,K^{(1+j)}\big({(\theta_f-t)}/{h}\big). \end{aligned} \tag{6.1}$$

Moreover, the constant in the O-term depends only on the kernel K, parameter L and s from $\mathcal{F}_s = \mathcal{F}_s(\gamma, a, \Theta, L)$, and this constant can be chosen uniformly over a bounded range of values of s.

Proof of Lemma 2.2. We have for $L_{h,0}(t)$ in (6.1) that due to Assumption 2, (iv),

$$|L_{h,0}(t^*)| = |L_{h,0}(t_*)| = \big|[f^{(\gamma)}]\big|\big|K^{(1)}(-q_*)\big| > 0,$$

where $[f^{(\gamma)}] > 0$ is as in Assumption 2.1, and that $L_{h,0}(t^*)$ and $L_{h,0}(t_*)$ are of opposite signs since $K^{(1)}$ is odd by Assumption 2, (ii). Therefore, from (6.1) we have for sufficiently small $h_0 > 0$, depending only on K, L and Θ, that $\min\{|\psi_{h,f}(t^*)|, |\psi_{h,f}(t_*)|\} > 0$ and $\psi_{h,f}(t^*)$ and $\psi_{h,f}(t_*)$ are of opposite signs as well. The assertion follows from the continuity of the probe functional $\psi_{h,f}(t)$. □

Next, we bound the deviation of the empirical probe functional from its population counterpart as well as for their derivatives.

Lemma 6.2. *For the probe functional (2.1) and its empirical version (2.2), there exists $h_0 > 0$ such that for any $h \in (0,h_0)$ and $n\in\mathbb{N}$ we have for $j = 0,1,2$ that*

(i) $\mathbb{E}_f[\hat\psi_{h,n}^{(j)}(t)] = \psi_{h,f}^{(j)}(t) + O_{f\in\mathcal{F}_s, t\in[0,1]}\big((nh^{\gamma+1+j})^{-1}\big), \qquad t\in[0,1],$

(ii) $\sup_{t\in[0,1]}\big|\hat\psi_{h,n}^{(j)}(t) - \mathbb{E}_f[\hat\psi_{h,n}^{(j)}(t)]\big| = O_{P,\mathcal{F}_s}\Big(\Big(\frac{\log(1/h)}{nh^{2(\gamma+j)+1}}\Big)^{1/2}\Big).$

Consequently, for $j = 0,1,2$ we have for any $h\in(0,h_0)$ and $n\in\mathbb{N}$ that

$$\sup_{t\in[0,1]} |\hat\psi_{h,n}^{(j)}(t) - \psi_{h,f}^{(j)}(t)| = O_{P,\mathcal{F}_s}\Big(\Big(\frac{\log(1/h)}{nh^{2(\gamma+j)+1}}\Big)^{1/2}\Big). \tag{6.2}$$

The constants in the O-terms and the constant h_0 depend only on the kernel K, the parameter σ_g in Definition 1 as well as on the Lipschitz constant L of $\mathcal{F}_s$

as in Definition 2.1.

The discretization error contained in the remainder term in the lemma thus has the rate $(nh^{\gamma+1+j})^{-1}$ uniformly in $f \in \mathcal{F}_s$ and $t \in [0,1]$. Finally, we bound the variance of the empirical probe functional and its derivatives.

Lemma 6.3. *For $j = 0, 1, 2$ we have that for $h \in (0, h_0)$ and $t \in [h, 1-h]$ that*

$$\mathsf{Var}_f(\hat{\psi}_{h,n}^{(j)}(t)) = n^{-1} h^{-2(\gamma+j)-1} \sigma^2 || \, K^{(\gamma+2+j)} \, ||_2^2 + O_{f \in \mathcal{F}_s, t \in [0,1]}\big((nh^{\gamma+j+1})^{-2}\big).$$

The constant in the O-term depends only on the kernel K and on the Lipschitz constant L.

Next we further investigate the first stage of the zero-crossing-time-technique.

Lemma 6.4. *There exist finite constants $h_0, C > 0$ depending on K, σ, σ_g as well as on L, Θ and s of $\mathcal{F}_s$, such that if $h \in (0, h_0)$ and n are such that $nh^{2\gamma+1} \geq C \log(1/h)$, then*

1. *with high probability, uniformly in $f \in \mathcal{F}_s$ there exists a $\xi \in [\hat{t}_*, \hat{t}^*]$ such that $\hat{\psi}_{h,n}(\xi) = 0$,*
2. *we have that $|\hat{t}^* - \hat{t}_*| = O_{P,\mathcal{F}_s}(h)$,*
3. *with high probability, uniformly in $f \in \mathcal{F}_s$, we have that $\theta_f \in [\hat{t}_*, \hat{t}^*]$.*

Moreover, C can be chosen uniformly over a bounded range of values of s, while h_0 is independent of s.

6.2. Rates of convergence: Proof of Theorem 2.3

In the following we shall restrict to the event that $\hat{\psi}_{h,n}(\hat{\theta}_{h,n}) = 0$, which by Lemma 6.4 is fulfilled with high probability uniformly over $\mathcal{F}_s$. By Taylor expansion of $\hat{\psi}_{h,n}$ at θ_f we have that

$$0 = \hat{\psi}_{h,n}(\hat{\theta}_{h,n}) = \hat{\psi}_{h,n}(\theta_f) + (\hat{\theta}_{h,n} - \theta_f)\hat{\psi}_{h,n}^{(1)}(\ddot{\theta}),$$

where $\ddot{\theta} = \rho\theta_f + (1-\rho)\hat{\theta}_{h,n}$ for $\rho \in [0,1]$ is some (random) value between θ_f and $\hat{\theta}_{h,n}$, so that

$$\hat{\theta}_{h,n} - \theta_f = -\frac{h\,\hat{\psi}_{h,n}(\theta_f)}{h\,\hat{\psi}_{h,n}^{(1)}(\ddot{\theta})}. \tag{6.3}$$

Asymptotics of the scale terms

The following lemma establishes the asymptotic behavior of the denominator in (6.3).

Lemma 6.5. *Under the assumptions of Theorem 2.3, one has*

$$|h\,\hat{\psi}_{h,n}^{(1)}(\ddot{\theta}) - (-1)^{\gamma+2}[f^{(\gamma)}]\,K^{(2)}(0)| = o_{P,\mathcal{F}_s}(1).$$

The proof is given in Section A.2.

Proof of Theorem 2.3. Convergence rate for the location of the kink

To prove the statement for $\hat{\theta}_{h,n}$, consider the right side equation in (6.3). Since $K^{(2)} \neq 0$, Lemma 6.5 implies that the denominator equals a constant unequal zero and a term of order $o_{P,\mathcal{F}_s}(1)$, uniformly over $\mathcal{F}_s$. For the numerator, write $\hat{\psi}_{h,n}(\theta_f) = \hat{\psi}_{h,n}(\theta_f) - \mathbb{E}_f[\hat{\psi}_{h,n}(\theta_f)] + \mathbb{E}_f[\hat{\psi}_{h,n}(\theta_f)]$. Then, Lemma 6.2, (i) in combination with representation (6.1) for $j = 0$ in Lemma 6.1 yield for a suitable choice of h_0 (depending only on K, σ_g and L) that

$$\mathbb{E}_f[\hat{\psi}_{h,n}(\theta_f)] = O_{\mathcal{F}_s}(h^{s-\gamma}) + O_{\mathcal{F}_s}\big((nh^{\gamma+1})^{-1}\big),$$

since $K^{(1)}(0) = 0$ due to Assumption 2, (ii). Further, Lemma 6.3 and Chebychev's inequality imply $\hat{\psi}_{h,n}(\theta_f) - \mathbb{E}_f[\hat{\psi}_{h,n}(\theta_f)] = O_{P,\mathcal{F}_s}\big((nh^{2\gamma+1})^{-1/2}\big)$. Hence,

$$\frac{h\,\hat{\psi}_{h,n}(\theta_f)}{h\,\hat{\psi}^{(1)}_{h,n}(\ddot{\theta})} = O_{P,\mathcal{F}_s}(h^{s-\gamma+1}) + O_{P,\mathcal{F}_s}((nh^{2\gamma-1})^{-1/2}),$$

which yields the assertion for $\hat{\theta}_{h,n}$. Note that the constants in the O-terms depend only on K, σ_g, L as well as s and these constants can be chosen continuously in s, see Lemmas 6.1, 6.2 and 6.3.

Convergence rate of the size of the kink

By Taylor expansion of $\hat{\psi}^{(1)}_{h,n}$ in (2.7) around θ_f,

$$\widehat{[f^{(\gamma)}]}_{h,n} = h\,\frac{\big(\hat{\psi}^{(1)}_{h,n}(\theta_f) + \hat{\psi}^{(2)}_{h,n}(\grave{\theta})(\hat{\theta}_{h,n} - \theta_f)\big)}{(-1)^{\gamma+2}K^{(2)}(0)}, \tag{6.4}$$

where $\grave{\theta}$ is some value between θ_f and $\hat{\theta}_{h,n}$. It follows from (6.1) for $j = 1$ that

$$[f^{(\gamma)}] = (-1)^{\gamma+2}\,h\,\frac{\psi^{(1)}_{h,f}(\theta_f)}{K^{(2)}(0)} + O_{\mathcal{F}_s}(h^{s-\gamma}). \tag{6.5}$$

Subtracting this from (6.4) leads to

$$\big(\widehat{[f^{(\gamma)}]}_{h,n} - [f^{(\gamma)}]\big) = \frac{h\,\big(\hat{\psi}^{(1)}_{h,n}(\theta_f) - \psi^{(1)}_{h,f}(\theta_f)\big)}{(-1)^{\gamma+2}K^{(2)}(0)} + \frac{h\,\hat{\psi}^{(2)}_{h,n}(\grave{\theta})\,(\hat{\theta}_{h,n} - \theta_f)}{(-1)^{\gamma+2}K^{(2)}(0)} + O_{\mathcal{F}_s}(h^{s-\gamma}). \tag{6.6}$$

Now, Chebychev's inequality and Lemma 6.3 for $j = 1$ yield

$$h\,\big(\hat{\psi}^{(1)}_{h,n}(\theta_f) - \mathbb{E}_f[\hat{\psi}^{(1)}_{h,n}(\theta_f)]\big) = O_{P,\mathcal{F}_s}((nh^{2\gamma+1})^{-1/2}).$$

Further, by Lemma 6.2, (i), for $j = 1$ (for a suitable choice of h_0 depending if necessary on K and σ_g) derive

$$h\,\big(\mathbb{E}_f[\hat{\psi}^{(1)}_{h,n}(\theta_f)] - \psi^{(1)}_{h,f}(\theta_f)\big) = O_{\mathcal{F}_s}((nh^{\gamma+1})^{-1}),$$

such that the first term on the right-hand side in (6.6) is of order $O_{\mathcal{F}_s}((nh^{2\gamma+1})^{-1/2})$.

Concerning the second term, the convergence rate (2.9) of the location of the kink together with $h\,\hat{\psi}^{(2)}_{h,n}(\grave{\theta}) = o_{P,\mathcal{F}_s}(1)$ (from Lemma A.6 as $K^{(3)}(0) = 0$ by Assumption 2, (ii), and $|\grave{\theta} - \theta_f| = o_{P,\mathcal{F}_s}(h)$ by Lemma A.4) yield the rate $o_{P,\mathcal{F}_s}(h^{s-\gamma+1}) + o_{P,\mathcal{F}_s}((nh^{2\gamma-1})^{-1/2})$. In summary we obtain that

$$\big|\widehat{[f^{(\gamma)}]}_{h,n} - [f^{(\gamma)}]\big| = O_{P,\mathcal{F}_s}((nh^{2\gamma+1})^{-1/2}) + O_{P,\mathcal{F}_s}(h^{s-\gamma}).$$

Note that the constants in the O-terms depend only on K, σ_g, L and on s and these constants can be chosen uniformly over a bounded range of values of s, see Lemmas 6.1, 6.2 and 6.3. □

6.3. Lower bounds: Proof of Theorem 2.4

Proof of Theorem 2.4. We shall use the method of two hypothesis, see Theorem 2.2 in Tsybakov (2009). Fix some $\theta_0 \in \operatorname{int}(\Theta)$, and introduce the function

$$R_\gamma(x;a) = \frac{a}{\gamma!}(x-\theta_0)^\gamma 1_{[\theta_0,1]}(x). \tag{6.7}$$

Let us start with (2.11). We set $f_0(x) = R_\gamma(x;a)$ for $x \in [0,1]$. As $f_0^{(\gamma-1)}(x) = a(x-\theta_0)1_{[\theta_0,1]}(x)$ we have that $f_0 \in \mathcal{F}_s(\gamma, a, \Theta, L)$ for any values of $s \geq \gamma+1$ and $L \geq 0$.

As for the sequence of alternative hypotheses, let $\theta_1 = \theta_0 + r_n \in \Theta$, where $r_n \downarrow 0$ will be chosen below, and consider

$$f_1 = f_0 - (v_0 - v_n), \tag{6.8}$$

where

$$v_0(x) = \frac{a}{\gamma!}\,(x-\theta_0)^\gamma\, 1_{[\theta_0,\theta_1]}(x) + a\Big(\frac{(\theta_1-\theta_0)^\gamma}{\gamma!} + \frac{(\theta_1-\theta_0)(x-\theta_1)^{\gamma-1}}{(\gamma-1)!}\Big)\, 1_{(\theta_1,1]}(x), \tag{6.9}$$

and

$$v_n(x) = \frac{1}{b_n}\int v_0(y)\Phi\big(\frac{x-y}{b_n}\big)\,\mathrm{d}y = \frac{1}{b_n}\int v_0(x-y)\Phi\big(y/b_n\big)\,\mathrm{d}y. \tag{6.10}$$

Here, Φ is a smooth kernel of order γ with support $[-1,1]$ (Tsybakov, 2009, p. 5), the sequence $b_n \downarrow 0$ remains to be selected, and in (6.10) we extend the definition of v_0 from $(\theta_1, 1]$ to $(\theta_1, 1+b_n]$. We check the conditions *(i)* and *(iia)* or *(iib)* in Definition 2.1 for f_1.

(i). We have that

$$(f_0 - v_0)^{(\gamma-1)}(x) = a\,(x-\theta_1)1_{(\theta_1,1]}(x).$$

Hence, $(f_0-v_0)^{(\gamma-1)} \in C^1(\{\theta_1\}^c)$ and as $v_n \in C^\infty$ it follows that $f_1 \in C^1(\{\theta_1\}^c)$. Apparently, $(f_0 - v_0)^{(\gamma)}(x) = a\,1_{(\theta_1,1]}(x)$ and therefore $[f_1^{(\gamma)}](\theta_1) = a$.

(iia) or *(iib)*. We have $(f_0 - v_0)^{(\gamma)} \in \mathrm{Lip}(\{\theta_1\}^c, 0) \subset \mathrm{Lip}(\{\theta_1\}^c, L/2)$ if $s = \gamma+1$ as well as $[(f_0-v_0)^{(\gamma+1)}](\theta_1) = 0$ and $g_{f_0-v_0,\gamma} \in \mathcal{H}^{s-(\gamma+1)}([0,1], L/2)$, provided $s > \gamma+1$. Turning to ν_n, by differentiating $\lfloor s\rfloor$-times under the integral in (6.10) we obtain

$$\begin{aligned}&\big|v_n^{(\lfloor s\rfloor)}(x) - v_n^{(\lfloor s\rfloor)}(z)\big|\\ &\quad= \frac{1}{b_n^{\lfloor s\rfloor-(\gamma-1)+1}}\,\Big|\int v_0^{(\gamma-1)}(y)\Big[\Phi^{(\lfloor s\rfloor-(\gamma-1))}\big(\frac{x-y}{b_n}\big) - \Phi^{(\lfloor s\rfloor-(\gamma-1))}\big(\frac{z-y}{b_n}\big)\Big]\,\mathrm{d}y\,\Big|\\ &\quad= \frac{1}{b_n^{\lfloor s\rfloor-\gamma+1}}\,\Big|\int v_0^{(\gamma-1)}(b_n\,u)\Big[\Phi^{(\lfloor s\rfloor-(\gamma-1))}\big(\frac{x}{b_n}-u\big) - \Phi^{(\lfloor s\rfloor-(\gamma-1))}\big(\frac{z}{b_n}-u\big)\Big]\,\mathrm{d}u\,\Big|.\end{aligned} \tag{6.11}$$

Now, we have that

$$|v_0^{(\gamma-1)}(x)| = \big|a\,(x-\theta_0)\,1_{[\theta_0,\theta_1]}(x) + a\,(\theta_1-\theta_0)\,1_{(\theta_1,1]}(x)\big| \le r_n, \tag{6.12}$$

as well as

$$\Big|\Phi^{(\lfloor s\rfloor-(\gamma-1))}\big(\frac{x}{b_n}-u\big) - \Phi^{(\lfloor s\rfloor-(\gamma-1))}\big(\frac{z}{b_n}-u\big)\Big| \le \text{const.}\cdot\frac{|x-z|^{s-\lfloor s\rfloor}}{b_n^{s-\lfloor s\rfloor}} \tag{6.13}$$

for any u. Since the integral in (6.11) ranges at most over two intervals of length at most 2 by the support of Φ, it follows that

$$\big|v_n^{(\lfloor s\rfloor)}(x) - v_n^{(\lfloor s\rfloor)}(z)\big| \le \text{const.}\cdot|x-z|^{s-\lfloor s\rfloor}\,r_n\,b_n^{-(s-\gamma+1)}.$$

Hence by choosing

$$r_n \cong b_n^{s-\gamma+1} \tag{6.14}$$

we obtain that $v_n^{(\gamma)} \in \mathrm{Lip}(\{\emptyset\}^c, L/2)$ if $s = \gamma+1$ or $v_n \in \mathcal{H}^s([0,1], L/2)$ if $s > \gamma+1$, so that $f_1 \in \mathcal{F}_s(\gamma, a, \Theta, L)$.

Concerning the Kullback-Leibler divergence, note that the distribution P_j of $Y_1, \dots, Y_n$ with respect to f_j has the density

$$p_j(y_1,\dots,y_n) = \prod_{i=1}^n \varphi_\sigma(y_i - f_j(x_i)), \quad j = 0,1,$$

with respect to the Lebesgue measure on $\mathbb{R}^n$. Here φ_σ denotes the normal density with standard deviation $\sigma > 0$. Thus, the Kullback-Leibler divergence is given by

$$\begin{aligned}K(P_0,P_1) &= \sum_{i=1}^n \int \log\left(\frac{\varphi_\sigma(y-f_0(x_i))}{\varphi_\sigma(y-f_1(x_i))}\right)\varphi_\sigma(y-f_0(x_i))\,dy\\ &= \frac{1}{2\sigma^2}\sum_{i=1}^n (f_0(x_i)-f_1(x_i))^2 = \frac{1}{2\sigma^2}\sum_{i=1}^n(\nu_0(x_i)-\nu_n(x_i))^2.\end{aligned} \tag{6.15}$$

Since v_0 is a polynomial of degree $\leq \gamma$ away from θ_0 and θ_1, and since the kernel Φ of order γ reproduces polynomials of degree $\leq \gamma$, we have $v_n(x) = v_0(x)$ outside of b_n-neighborhoods of θ_0 and θ_1. Inside these neighborhoods, which contain of the order nb_n points, using Taylor-expansion up to order $\gamma - 1$ and (6.12) yields (Tsybakov, 2009, p. 6)

$$||\, v_n - v_0\,||_\infty \leq \text{const.} \cdot r_n\, b_n^{\gamma-1},$$

so that

$$\sum_{i=1}^{n} (\nu_0(x_i) - \nu_n(x_i))^2 \leq \text{const.} \cdot n\, b_n\, r_n^2\, b_n^{2\gamma-2} \cong n\, b_n^{2s+1}$$

under the choice (6.14). Choosing $b_n \cong n^{-1/(2s+1)}$, Theorem 2.2 in Tsybakov (2009) implies that the minimax lower bound over the functional class $\mathcal{F}_s$ is of order $r_n \cong b_n^{s-\gamma+1} \cong n^{-(s-\gamma+1)/(2s+1)}$.

Next, we verify (2.12). Let $a_1 > a$ be such that $a_1 - a = \tilde{r}_n \downarrow 0$, which remains to be specified. As hypotheses functions we set

$$f_0(x) = R_\gamma(x; a), \quad f_1(x) = R_\gamma(x; a_1) - \tilde{v}_n,$$

where

$$\tilde{v}_n(x) = \frac{1}{b_n} \int \Big[R_\gamma(y; a_1 - a)\Big]\, \Phi\big(\frac{x-y}{b_n}\big) \mathrm{d}y = \frac{1}{b_n} \int \Big[R_\gamma(x-y; a_1 - a)\Big]\, \Phi\big(\frac{y}{b_n}\big) \mathrm{d}y,$$

and Φ is, as above, a smooth kernel of order γ with support $[-1, 1]$, $b_n \downarrow 0$ and we extend the definition of $R_\gamma(y; a_1 - a)$ to $[\theta_0, 1 + b_n]$. Then

$$\begin{aligned}
&|\tilde{v}_n^{(\lfloor s \rfloor)}(x) - \tilde{v}_n^{(\lfloor s \rfloor)}(z)| \\
&\qquad = \frac{1}{b_n^{\lfloor s \rfloor - \gamma + 1}} \Big| \int R_\gamma^{(\gamma)}(y; a_1 - a) \Big[\Phi^{(\lfloor s \rfloor - \gamma)}\big(\frac{x-y}{b_n}\big) - \Phi^{(\lfloor s \rfloor - \gamma)}\big(\frac{z-y}{b_n}\big)\Big] \mathrm{d}y \Big| \\
&\qquad \leq \frac{a_1 - a}{b_n^{\lfloor s \rfloor - \gamma + 1}} \int \Big|\Phi^{(\lfloor s \rfloor - \gamma)}\big(\frac{x-y}{b_n}\big) - \Phi^{(\lfloor s \rfloor - \gamma)}\big(\frac{z-y}{b_n}\big)\Big| \mathrm{d}y \\
&\qquad \leq \frac{\tilde{r}_n}{b_n^{\lfloor s \rfloor - \gamma}} \int \Big|\Phi^{(\lfloor s \rfloor - \gamma)}\big(\frac{x}{b_n} - u\big) - \Phi^{(\lfloor s \rfloor - \gamma)}\big(\frac{z}{b_n} - u\big)\Big| \mathrm{d}u \\
&\qquad \leq \text{const.} \cdot |x - z|^{s - \lfloor s \rfloor}\, \tilde{r}_n\, b_n^{-(s-\gamma)}
\end{aligned}$$

by using (6.13) in the last step, which implies that $f_1 \in \mathcal{F}_s(\gamma, a, \Theta, L)$ under the choice

$$\tilde{r}_n \cong b_n^{s-\gamma}. \tag{6.16}$$

Outside of a b_n neighborhood of θ_0 we have $R_\gamma(x; a_1 - a) = \tilde{v}_n(x)$. By Taylor-expansion up to order $\gamma - 1$, using the Lipschitz continuity of $R_\gamma(x; a_1 - a)^{(\gamma-1)}$ with Lipschitz constant $a_1 - a = \tilde{r}_n$ we obtain that

$$\big|R_\gamma(x; a_1 - a) - \tilde{v}_n(x)\big| \leq \text{const.} \cdot \tilde{r}_n\, b_n^\gamma.$$

Hence, the Kullback-Leibler divergence between P_1 and P_0 is of order

$$K(P_0, P_1) \cong n\, b_n\, \tilde{r}_n^2\, b_n^{2\,\gamma} \cong n b_n^{2s+1}$$

under the choice (6.16). Inserting $b_n \cong n^{-1/(2s+1)}$ in (6.16) gives the result. $\square$

6.4. Asymptotic normality: Proof of Theorem 3.1

By Taylor expansion of $\psi_{h,f}$ at θ_f, $0 = \psi_{h,f}(\tilde{\theta}_{h,f}) = \psi_{h,f}(\theta_f) + (\tilde{\theta}_{h,f} - \theta_f)\psi^{(1)}_{h,f}(\check{\theta})$, for some $\check{\theta}$ on the line between θ_f and $\tilde{\theta}_{h,f}$, so that

$$\tilde{\theta}_{h,f} - \theta_f = -\frac{h\,\psi_{h,f}(\theta_f)}{h\,\psi^{(1)}_{h,f}(\check{\theta})}. \tag{6.17}$$

Subtracting this from (6.3) and dividing by $\tilde{w}^{loc}_n(h)$ as defined in (3.3), we get that

$$\frac{\hat{\theta}_{h,n} - \tilde{\theta}_{h,f}}{\tilde{w}^{loc}_n(h)} = \frac{[f^{(\gamma)}]\,K^{(2)}(0)}{h\,\hat{\psi}^{(1)}_{h,n}(\ddot{\theta})}\hat{S}_1(h,n) + R_1(h,n), \tag{6.18}$$

where

$$\hat{S}_1(h,n) := \frac{\sqrt{nh^{2\gamma+1}}\,(\psi_{h,f}(\theta_f) - \hat{\psi}_{h,n}(\theta_f))}{\sigma\,||\,K^{(\gamma+2)}\,||_2} \tag{6.19}$$

and

$$R_1(h,n) := \frac{h\,[\hat{\psi}^{(1)}_{h,n}(\ddot{\theta}) - \psi^{(1)}_{h,f}(\check{\theta})]\,\sqrt{nh^{2\gamma+1}}\,\psi_{h,f}(\theta_f)}{h^2\,\hat{\psi}^{(1)}_{h,n}(\ddot{\theta})\psi^{(1)}_{h,f}(\check{\theta})} \cdot \frac{[f^{(\gamma)}]\,K^{(2)}(0)}{\sigma\,||\,K^{(\gamma+2)}\,||_2}. \tag{6.20}$$

By Taylor expansion of $\psi^{(1)}_{h,f}$ in (2.8) at θ_f,

$$\widetilde{[f^{(\gamma)}]}_h = h\,\frac{\big(\psi^{(1)}_{h,f}(\theta_f) + \psi^{(2)}_{h,f}(\bar{\theta})(\tilde{\theta}_{h,f} - \theta_f)\big)}{(-1)^{\gamma+2}K^{(2)}(0)}, \tag{6.21}$$

where $\bar{\theta}$ is between $\tilde{\theta}_{h,f}$ and θ_f. Subtracting (6.21) from (6.4) and dividing by $\tilde{w}^{size}_n(h)$ leads to

$$\frac{\widehat{[f^{(\gamma)}]}_{h,n} - \widetilde{[f^{(\gamma)}]}_h}{\tilde{w}^{size}_n(h)} = \hat{S}_2(h,n) + R_2(h,n), \tag{6.22}$$

where due to definition of $\tilde{w}^{size}_n$ in (3.3) we have

$$\hat{S}_2(h,n) := \frac{\sqrt{nh^{2\gamma+3}}\,\big(\hat{\psi}^{(1)}_{h,n}(\theta_f) - \psi^{(1)}_{h,f}(\theta_f)\big)}{(-1)^{\gamma+2}\,\sigma\,||\,K^{\gamma+3}\,||_2} \tag{6.23}$$

and

$$R_2(h,n) := \frac{h^2\,\hat{\psi}^{(2)}_{h,n}(\dot{\theta})\,\sqrt{nh^{2\gamma-1}}\,(\hat{\theta}_{h,n} - \theta_f)}{(-1)^{\gamma+2}\,\sigma\,||\,K^{\gamma+3}\,||_2} - \frac{h^2\,\psi^{(2)}_{h,f}(\bar{\theta})\,\sqrt{nh^{2\gamma-1}}\,(\tilde{\theta}_{h,f} - \theta_f)}{(-1)^{\gamma+2}\,\sigma\,||\,K^{\gamma+3}\,||_2}. \tag{6.24}$$

The following lemma shows the negligibility of the remainder terms in (6.20) resp. (6.24).

Lemma 6.6. *Under the assumptions of Theorem 3.1, it holds*

$$\max\{|R_1(h,n)|, |R_2(h,n)|\} = o_{P,\mathcal{F}_s}(1).$$

The next lemma shows the joint asymptotic normality of the scores in (6.19) and (6.23).

Lemma 6.7. *Suppose the assumptions of Theorem 3.1 are fulfilled. Then, for any* $\mathbf{x} \in \mathbb{R}^2$

$$\sup_{f\in\mathcal{F}_s} \Big| P_f\Big(\big(\hat{S}_1(h,n)\,,\ \hat{S}_2(h,n)\big)^T \leq \mathbf{x}\Big) - \Phi(\mathbf{x}) \Big| = o(1).$$

The proofs of the lemmas are provided in Section A.2.

Proof of Theorem 3.1. Deduce from Lemma 6.5 that

$$\Big| \frac{[f^{(\gamma)}]^2\, K^{(2)}(0)^2}{h^2\, \hat{\psi}^{(1)}_{h,n}(\ddot{\theta})^2} - 1 \Big| = o_{P,\mathcal{F}_s}(1).$$

By combining Lemma 6.7 with the uniform Slutzky Theorem D.3 in Section D we obtain that for $\mathbf{x} \in \mathbb{R}^2$,

$$\sup_{f\in\mathcal{F}_s} |P_f\Big(\big({}^{[f^{(\gamma)}]\, K^{(2)}(0)}/_{h\, \hat{\psi}^{(1)}_{h,n}(\ddot{\theta})}\, \hat{S}_1(h,n)\,,\ \hat{S}_2(h,n)\big)^T \leq \mathbf{x}\Big) - \Phi(\mathbf{x})| = o(1).$$

With this, we conclude the proof in view of (6.18) and (6.22), Lemma 6.6 and the uniform Slutzky theorem D.3. □

6.5. Adaptive confidence sets: Proof of Theorem 3.3

From the first term in the expansion (6.3) and in Lemma 6.5, the condition for $f \in \widetilde{\mathcal{F}}_s$ in (3.10) can be written as a bias condition

$$\widetilde{b_1}\, h_k^{s-(\gamma-1)} \leq |\tilde{\theta}_{h_k,f} - \theta_f| \leq \widetilde{b_2}\, h_k^{s-(\gamma-1)} \qquad \forall\ k \geq k_0, \tag{6.25}$$

where for appropriate constants $C_{b_1} > 0$ and $0 < C_{b_2} < a\, K^{(2)}(0)$ we set

$$\widetilde{b_1} = \frac{b_1}{\big|[f^{(\gamma)}]\big|\, K^{(2)}(0) + C_{b_1}} \quad \text{and} \quad \widetilde{b_2} = \frac{b_2}{\big|[f^{(\gamma)}]\big|\, K^{(2)}(0) - C_{b_2}}. \tag{6.26}$$

Recall the definitions of $\mathcal{K}_n$ resp. h_k in (3.7) resp. (3.8) and introduce for $s \in [\underline{s}, \overline{s}]$

$$B(k,s) = \widetilde{b_2} h_k^{s-(\gamma-1)} = \widetilde{b_2} 2^{-k(s-\gamma+1)}, \qquad \sigma(n,k) = \sqrt{{}^{k}/_{(n h_k^{2\gamma-1})}} = \sqrt{{}^{2^{k(2\gamma-1)}k}/_{n}},$$

as well as

$$k_n^*(s) = \min\{k \in \mathcal{K}_n \,|\, B(k,s) \leq {}^{C_{\mathrm{Lep}}\sigma(n_2,k)}/_{8}\}.$$

The next key lemma shows that $h_{k_n^*(s)}$ is of optimal order and that under $f \in \widetilde{\mathcal{F}}$, $k_n^*(s_f)$ is essentially selected by the Lepski choice $\hat{k}_n$.

Lemma 6.8. *We have that*

$$h_{k_n^*(s)} \cong \left({}^{\log(n_2)}/{}_{n_2}\right)^{1/(2s+1)}. \tag{6.27}$$

If $C_{\mathrm{Lep}} > 0$ is chosen large enough depending only on K, σ, σ_g as well as on L and Θ of $\widetilde{\mathcal{F}}$ and if n is sufficiently large such that $k_n^*(\underline{s}) \geq 2$ then there exists a $\rho \in \mathbb{N}$ depending only on $b_1, b_2, \underline{s}$ and on C_{Lep} such that

$$\sup_{f \in \widetilde{\mathcal{F}}} P_f(\hat{k}_n \notin [k_n^*(s_f) - \rho, k_n^*(s_f)]) = o(1),$$

where s_f is given in Definition 3.2.

The proof of the lemma is deferred to the appendix, Section A.4.

Adaptive coverage: Proof of (3.17)

Lemma 6.9. *Let ρ be as in Lemma 6.8. Then we have that*

$$\lim_{n\to\infty} \sup_{f \in \widetilde{\mathcal{F}}} \max_{j \in \{0,\dots,\rho\}} \Big| P_f\Big(\max\Big[\Big|\frac{\hat{\theta}_{h_{k^*-j+u_n}, n_1} - \tilde{\theta}_{h_{k^*-j+u_n}, n_1}}{\tilde{w}_{n_1}^{loc}(h_{k^*-j+u_n})}\Big|, \\ \Big|\frac{\widehat{[f^{(\gamma)}]}_{h_{k^*-j+v_n}} - \widetilde{[f^{(\gamma)}]}_{h_{k^*-j+v_n}}}{\tilde{w}_{n_1}^{size}(h_{k^*-j+v_n})}\Big|\Big] \leq q_{1-\alpha}(W)\Big) - (1-\alpha)\Big| = 0,$$

where we abbreviate $k^* = k_n^*(s_f)$, and $\tilde{w}_{n_1}^{loc}(h)$ and $\tilde{w}_{n_1}^{size}(h)$ are defined in (3.3).

Proof of Lemma 6.9. In the proof we write n for n_1 as only the subsample S_1 is involved. Given $s \in [\underline{s}, \overline{s}]$ consider $f \in \widetilde{\mathcal{F}}_s$, so that $k^* = k_n^*(s)$. To show that the sequences $h_{k^*-j+u_n}$ and $h_{k^*-j+v_n}$, $j \in \{0, \dots, \rho\}$, satisfy the bandwidth conditions of Theorem 3.1 in (3.1) it suffices to consider $j = 0$. Then

$$h_{k^*+u_n} = h_{k^*}\, h_{u_n} \cong h_{k^*}\, \log(n)^{-1/2\gamma-1} \cong {}^{\log(n)^{\zeta}}/{}_{n^{1/(2s+1)}}, \\ \zeta = {}^{-2(s+1-\gamma)}/{}_{(2s+1)(2\gamma-1)} < 0, \tag{6.28}$$

due to choice of u_n in (3.12) and the expansion (6.27) of h_{k^*}, both of which hold true in terms of n (that is n_1). Therefore

$$n h_{k^*+u_n}^{4s-2\gamma+1} \cong \log(n)^{(4s-2\gamma+1)\,\zeta}\; n^{-2(s-\gamma)/(2s+1)} \to 0,$$

since $s \geq \underline{s} \geq \gamma + 1$, which is the first part of (3.1), and similarly,

$$n h_{k^*+u_n}^{2\gamma+1} \cong \log(n)^{(2\gamma+1)\,\zeta}\; n^{2(s-\gamma)/(2s+1)} \to \infty,$$

the second part of (3.1). Passing from u_n to v_n only changes the value of ζ, which is not relevant in the above analysis.

The lemma then follows from Theorem 3.1 together with the uniform continuous mapping theorem, Theorem D.1 in Section D. ☐

We shall denote

$$\hat{w}_k^{loc} = \frac{\hat{\sigma}_{n_1} \,||\, K^{(\gamma+2)} \,||_2}{\sqrt{n_1 h_{k+u_n}^{2\gamma-1}}\ \widehat{[f^{(\gamma)}]}_{h_k, n_1}\, K^{(2)}(0)}, \qquad \hat{w}_k^{size} = \frac{\hat{\sigma}_{n_1} \,||\, K^{(\gamma+3)} \,||_2}{\sqrt{n_1 h_{k+v_n}^{2\gamma+1}}\ K^{(2)}(0)}.$$

compare to $\hat{w}^{loc}$ in (3.14) and $\hat{w}^{size}$ in (3.15), and similarly

$$\tilde{w}_k^{loc} = \tilde{w}_{n_1}^{loc}(h_{k+u_n}), \qquad \tilde{w}_k^{size} = \tilde{w}_{n_1}^{size}(h_{k+v_n}). \tag{6.29}$$

Lemma 6.10. *Let ρ be as in Lemma 6.8. Then we have for each $j \in \{0, \ldots, \rho\}$ that*

$$\frac{\hat{\theta}_{h_{k^*-j+u_n}, n_1} - \theta_f}{\hat{w}_{k^*-j}^{loc}} - \frac{\hat{\theta}_{h_{k^*-j+u_n}, n_1} - \tilde{\theta}_{h_{k^*-j+u_n}, n_1}}{\tilde{w}_{k^*-j}^{loc}} = o_{P, \widetilde{\mathcal{F}}}(1),$$

and

$$\frac{\widehat{[f^{(\gamma)}]}_{h_{k^*-j+v_n}, n_1} - [f^{(\gamma)}]}{\hat{w}_{k^*-j}^{size}} - \frac{\widehat{[f^{(\gamma)}]}_{h_{k^*-j+v_n}} - \widetilde{[f^{(\gamma)}]}_{h_{k^*-j+v_n}}}{\tilde{w}_{k^*-j}^{size}} = o_{P, \widetilde{\mathcal{F}}}(1),$$

where again we abbreviate $k^* = k_n^*(s_f)$.

Proof of Lemma 6.10. Given $s \in [\underline{s}, \overline{s}]$ consider $f \in \widetilde{\mathcal{F}}_s$, so that $k^* = k_n^*(s)$. It will suffice to consider $j = 0$. Then

$$\begin{aligned} &\frac{\hat{\theta}_{h_{k^*+u_n}, n_1} - \theta_f}{\hat{w}_{k^*}^{loc}} - \frac{\hat{\theta}_{h_{k^*+u_n}, n_1} - \tilde{\theta}_{h_{k^*+u_n}, n_1}}{\tilde{w}_{k^*}^{loc}} \\ &\qquad = \big(\hat{\theta}_{h_{k^*+u_n}, n_1} - \theta_f\big)\Big(\frac{1}{\hat{w}_{k^*}^{loc}} - \frac{1}{\tilde{w}_{k^*}^{loc}}\Big) + \frac{\tilde{\theta}_{h_{k^*+u_n}, n_1} - \theta_f}{\tilde{w}_{k^*}^{loc}}. \end{aligned}$$

For the second term, from the definition of $\tilde{w}_{k^*}^{loc}$ in (6.29) and (3.3) and from (3.4),

$$\begin{aligned} \big(\tilde{w}_{k^*}^{loc}\big)^{-1}\big(\tilde{\theta}_{h_{k^*+u_n}, n_1} - \theta_f\big) &= O_{\widetilde{\mathcal{F}}_s}\big(\sqrt{n}\, h_{k^*+u_n}^{\gamma-1/2}\big)\, O_{\widetilde{\mathcal{F}}_s}\big(h_{k^*+u_n}^{s-\gamma+1}\big) \\ &= O_{\widetilde{\mathcal{F}}_s}\big((\log n)^{-(s+1-\gamma)/(2\gamma-1)}\big), \end{aligned}$$

where we inserted (6.28) in the last step. For the first term, since (6.28) implies undersmoothing, from (2.9) in Theorem 2.3

$$\hat{\theta}_{h_{k^*+u_n}, n_1} - \theta_f = O_{P, \widetilde{\mathcal{F}}_s}\big((\sqrt{n}\, h_{k^*+u_n}^{\gamma-1/2})^{-1}\big),$$

and therefore

$$\big(\hat{\theta}_{h_{k^*+u_n}, n_1} - \theta_f\big)\Big(\frac{1}{\hat{w}_{k^*}^{loc}} - \frac{1}{\tilde{w}_{k^*}^{loc}}\Big) = O_{P, \widetilde{\mathcal{F}}_s}(1)\Big(\frac{\widehat{[f^{(\gamma)}]}_{h_{k^*}, n_1}}{\hat{\sigma}_{n_1}} - \frac{[f^{(\gamma)}]}{\sigma}\Big) = o_{P, \widetilde{\mathcal{F}}_s}(1)$$

by using Theorem 2.3, (2.10), the Assumption (3.13) on $\hat{\sigma}_{n_1}$ together with the triangle inequality. This proves the first part of the lemma. For the second, we note that

$$h_{k^*+v_n} \cong {}^{\log(n)^\eta}/{}_{n^{1/(2s+1)}}, \qquad \eta = {}^{-2(s-\gamma)}/{}_{(2s+1)(2\gamma+1)} < 0 \tag{6.30}$$

is still an undersmoothing bandwidth. The argument then proceeds analogously. □

Proof of (3.17). Let ρ be as in Lemma 6.8, and given $s \in [\underline{s}, \overline{s}]$ consider $f \in \widetilde{\mathcal{F}}_s$, and let $k^* = k_n^*(s)$. Then from the definition of the confidence sets in (3.16) and Lemma 6.8,

$$\begin{aligned}
&P_f\Big((\theta_f, [f^{(\gamma)}])^T \in C_n^{loc}(\alpha) \times C_n^{size}(\alpha)\Big) \\
&= \sum_{k^*-\rho \le k \le k^*} P_f\Big(\max\Big[\Big|\frac{\hat{\theta}_{h_{k+u_n},n_1} - \theta_f}{\hat{w}_k^{loc}}\Big|, \Big|\frac{\widehat{[f^{(\gamma)}]}_{h_{k+v_n},n_1} - [f^{(\gamma)}]}{\hat{w}_k^{size}}\Big|\Big] \\
&\qquad\qquad \le q_{1-\alpha}(W)\,,\, \{\hat{k}_n = k\}\Big) + o_{P,\widetilde{\mathcal{F}}_s}(1) \\
&= \sum_{k^*-\rho \le k \le k^*} P_f\Big(\max\Big[\Big|\frac{\hat{\theta}_{h_{k+u_n},n_1} - \tilde{\theta}_{h_{k+u_n},n_1}}{\tilde{w}_k^{loc}}\Big|, \Big|\frac{\widehat{[f^{(\gamma)}]}_{h_{k+v_n},n_1} - \widetilde{[f^{(\gamma)}]}_{h_{k+v_n}}}{\tilde{w}_k^{size}}\Big|\Big] \\
&\qquad\qquad \le q_{1-\alpha}(W)\,,\, \{\hat{k}_n = k\}\Big) + o_{P,\widetilde{\mathcal{F}}_s}(1) \\
&= \sum_{k^*-\rho \le k \le k^*} P_f\Big(\max\Big[\Big|\frac{\hat{\theta}_{h_{k+u_n},n_1} - \tilde{\theta}_{h_{k+u_n},n_1}}{\tilde{w}_k^{loc}}\Big|, \Big|\frac{\widehat{[f^{(\gamma)}]}_{h_{k+v_n},n_1} - \widetilde{[f^{(\gamma)}]}_{h_{k+v_n}}}{\tilde{w}_k^{size}}\Big|\Big] \\
&\qquad\qquad \le q_{1-\alpha}(W)\Big)\, P_f(\hat{k}_n = k) + o_{P,\widetilde{\mathcal{F}}_s}(1),
\end{aligned}$$

where the second step follows from Lemma 6.10 and the final step from sample splitting and the independence of the subsamples. The claim then follows from Lemma 6.9. This concludes the proof of (3.17). □

Adaptive length: Proof of (3.18).

From the definition of $\hat{w}^{loc}$ in (3.14) and of $\hat{w}^{size}$ in (3.15) and the consistency of $\widehat{[f^{(\gamma)}]}_{h_{\hat{k}_n},n_1}$, it suffices to show that there exists a constant $\tilde{C} > 0$ such that

$$\begin{aligned}
&\lim_{n\to\infty} \sup_{f \in \widetilde{\mathcal{F}}} P_f\Big({}^1\big/\sqrt{n\, h_{\hat{k}_n+u_n}^{2\gamma-1}} \ge \tilde{C}\, \big({}^{\log(n)}/{}_n\big)^{(s_f-\gamma+1)/(2s_f+1)}\Big) = 0, \\
&\lim_{n\to\infty} \sup_{f \in \widetilde{\mathcal{F}}} P_f\Big({}^1\big/\sqrt{n\, h_{\hat{k}_n+v_n}^{2\gamma+1}} \ge \tilde{C}\, \big({}^{\log(n)}/{}_n\big)^{(s_f-\gamma)/(2s_f+1)}\Big) = 0.
\end{aligned} \tag{6.31}$$

As for the first display, consider $s \in [\underline{s}, \overline{s}]$ and $f \in \widetilde{\mathcal{F}}_s$ and set $k^* = k^*(s)$. From (6.28),

$$ {}^1\big/\sqrt{n\, h_{k^*+u_n}^{2\gamma-1}} \le \tilde{C}_1\, \big({}^{\log(n)}/{}_n\big)^{(s-\gamma+1)/(2s+1)}$$

for some constant $\tilde{C}_1$.

Moreover, from Lemma 6.8 for some $\tilde{C}_2$,

$$\lim_{n\to\infty} \sup_{f\in\widetilde{\mathcal{F}}_s} P_f\Big(1/\sqrt{n\, h^{2\gamma-1}_{\hat{k}_n+u_n}} \geq \tilde{C}_2\, 1/\sqrt{n\, h^{2\gamma-1}_{k^*+u_n}} \Big) = 0.$$

This implies the first display in (6.31) with $\tilde{C} = \tilde{C}_2\, \tilde{C}_1$. As for the second, from (6.30) there is a $\tilde{C}_3$ such that

$$1/\sqrt{n\, h^{2\gamma+1}_{k^*+v_n}} \leq \tilde{C}_3 \left(\log(n)/n\right)^{(s-\gamma)/(2s+1)}$$

and a $\tilde{C}_4$ such that

$$\lim_{n\to\infty} \sup_{f\in\widetilde{\mathcal{F}}_s} P_f\Big(1/\sqrt{n\, h^{2\gamma+1}_{\hat{k}_n+v_n}} \geq \tilde{C}_4\, 1/\sqrt{n\, h^{2\gamma+1}_{k^*+v_n}} \Big) = 0.$$

The second display is then clear for $\tilde{C} = \tilde{C}_3\, \tilde{C}_4$.

Appendix A: Additional technical results

A.1. Proofs for Section 6.1 and additional auxiliary results

Proof of Lemma 6.1. Let h_0 be as in the assumption and $h \in (0, h_0)$. Given $t \in [h, 1-h]$ let $\tau = (\theta_f - t)/h$. By differentiation under the integral, substitution, γ times integration by parts (note that $f^{(\gamma)}$ is absolutely continuous) and Assumption 2, (i), for $m = 2+j, \dots, \gamma+1+j$ we obtain that

$$\begin{aligned}
\psi^{(j)}_{h,f}(t) &= (-1)^j h^{-(\gamma+1+j)} \int_{[0,1]} f(x)\, K^{(\gamma+2+j)}(h^{-1}(x-t))\, \mathrm{d}x \\
&= (-1)^{\gamma+j}\, h^{-j} \int_{[-1,1]} f^{(\gamma)}(t+xh)\, K^{(2+j)}(x)\, \mathrm{d}x \\
&= (-1)^{\gamma+j}\, h^{-j} \int_{-1}^{\tau} f^{(\gamma)}(t+xh)\, K^{(2+j)}(x)\, \mathrm{d}x \\
&\quad + (-1)^{\gamma+j} \int_{\tau}^{1} f^{(\gamma)}(t+xh)\, K^{(2+j)}(x)\, \mathrm{d}x.
\end{aligned}$$

Further, since $K^{(1+j)}(-1) = K^{(1+j)}(1) = 0$ by Assumption 2, (i), we have that

$$\begin{aligned}
(-1)^{\gamma+j} f^{(\gamma)}(\theta_f-) \int_{-1}^{\tau} K^{(2+j)}(x)\mathrm{d}x + (-1)^{\gamma+j} f^{(\gamma)}(\theta_f+) \int_{\tau}^{1} K^{(2+j)}(x)\mathrm{d}x \\
= (-1)^{\gamma+j} K^{(1+j)}(\tau) \big[f^{(\gamma)}(\theta_f-) - f^{(\gamma)}(\theta_f+) \big] \\
= (-1)^{\gamma+1+j} \big[f^{(\gamma)} \big] K^{(1+j)}(\tau).
\end{aligned}$$

Thus,

$$\begin{aligned}\psi_{h,f}^{(j)}(t) &= (-1)^{\gamma+1+j}\, h^{-j}[f^{(\gamma)}]\, K^{(1+j)}(\tau)\\ &\quad + (-1)^{\gamma+j}\, h^{-j}\int_{-1}^{\tau} (f^{(\gamma)}(t+xh) - f^{(\gamma)}(\theta_f -))\, K^{(2+j)}(x)\,\mathrm{d}x\\ &\quad + (-1)^{\gamma+j}\, h^{-j}\int_{\tau}^{1} (f^{(\gamma)}(t+xh) - f^{(\gamma)}(\theta_f +))\, K^{(2+j)}(x)\,\mathrm{d}x\\ &=: (-1)^{\gamma+1+j}\, h^{-j}[f^{(\gamma)}]\, K^{(1+j)}(\tau) + J_{h,j}(t).\end{aligned}$$

Note that if $x \in (-1,\tau)$ then $t + xh < \theta_f$ and if $x \in (\tau, 1)$ then $t + xh > \theta_f$. Hence if $s - (\gamma + 1) = 0$, from the Lipschitz continuity of $f^{(\gamma)}$ outside θ_f we directly obtain that $|J_{h,j}(t)| = O_{f\in\mathcal{F}_s, t\in[h,1-h]}(h^{s-\gamma-j})$. If $s - (\gamma + 1) > 0$, then by integration by parts, $K^{(1+j)}(-1) = K^{(1+j)}(1) = 0$ and the definition of $g_{f,\gamma}$ we obtain

$$\begin{aligned}J_{h,j}(t) &= (-1)^{\gamma+1+j}\, h^{1-j}\int_{-1}^{1} g_{f,\gamma}(t+xh)\, K^{(1+j)}(x)\,\mathrm{d}x\\ &= (-1)^{\gamma+1+j}\, h^{1-j}\int_{-1}^{1} \big(g_{f,\gamma}(t+xh) - g_{f,\gamma}(t)\big)\, K^{(1+j)}(x)\,\mathrm{d}x,\end{aligned}$$

where in the second step we used that $\int_{-1}^{1} K^{(1+j)} = 0$. For $j = 0$ and $j = 2$ this follows since $K^{(1)}$ and $K^{(3)}$ are odd functions, for $K^{(2)}$ since $\int_{-1}^{1} K^{(2)} = 2\,K^{(1)}(1) = 0$. If $0 < s - (\gamma + 1) \le 1$ we can directly use the uniform Lipschitz continuity of $g_{f,\gamma}$ to get $|J_{h,j}(t)| = O_{f\in\mathcal{F}_s, t\in[h,1-h]}(h^{s-\gamma-j})$. If $s - (\gamma + 1) > 1$, first note that by integration by parts and Assumption 2, (i) and (iii) we also have

$$\int_{-1}^{1} x^k K^{(1+j)}(x)\mathrm{d}x = 0, \quad \text{for} \quad k = 1,\dots, l-\gamma-1+j. \tag{A.1}$$

Thus, by using Taylor expansion of $g_{f,\gamma}$ around t and (A.1), we also obtain that by Hölder-smoothness of $g_{f,\gamma}$,

$$|J_{h,j}(t)| = O_{f\in\mathcal{F}_s, t\in[h,1-h]}(h^{s-\gamma-j}).$$

Note that all the constants in the O-terms depend only on K, L as well as on s, where the constants are continuous in s, due to the remaining term in the Taylor expansion. □

Proof of Lemma 6.2. (i). It holds for $t \in [0, 1]$ that

$$\begin{aligned}\mathbb{E}_f[\hat\psi_{h,n}^{(j)}(t)] &= n^{-1}h^{-(\gamma+1+j)}\sum_{i=1}^{n} f(x_i)K^{(\gamma+2+j)}(h^{-1}(x_i - t))\\ &= h^{-(\gamma+1+j)}\int f(x)K^{(\gamma+2+j)}(h^{-1}(x-t))\,\mathrm{d}x + R_n(t,h),\end{aligned}$$

where $R_n(t,h)$ is an error term of order $O_{f\in\mathcal{F}_s,t\in[0,1]}\big((nh^{\gamma+1+j})^{-1}\big)$, due to Riemann-sum approximation. Indeed, let $B(n,h)$ denote the index set for which the sum in the latter display is not zero. Due to the equidistant design in model (1.1) and the support of K in Assumption 2 it holds that $|B(n,h)| \leq 2nh$. Let $A_i = [x_{i-1}, x_i)$ for $i = \{1,\dots,n-1\}$, where $x_0 = 0$ and $A_n = [x_{n-1}, x_n]$, then

$$\begin{aligned} &R_n(t,h) \\ &\leq n^{-1}h^{-(\gamma+1+j)} \sum_{i\in B(n,h)} \Big|\sup_{z_1\in A_i} f(z_1)K^{(\gamma+2+j)}(h^{-1}(z_1-t)) \\ &\qquad\qquad - \inf_{z_2\in A_i} f(z_2)K^{(\gamma+2+j)}(h^{-1}(z_2-t))\Big| \\ &\leq 4n^{-1}h^{-(\gamma+1+j)}C_{L,K}, \end{aligned}$$

where $C_{L,K} > 0$ is the Lipschitz constant of the product of f and $K^{(\gamma+2+j)}$ which, by definition of $\mathcal{F}_s$, can be chosen uniformly in $f \in \mathcal{F}_s$ and depending only on K and L, which concludes *(i)*.

(ii). Consider

$$\sqrt{nh^{2(\gamma+i)+1}}\big(\hat{\psi}^{(j)}_{h,n}(t) - \mathbb{E}_f\big[\hat{\psi}^{(j)}_{h,n}(t)\big]\big) = (nh)^{-1/2}\sum_{i=1}^{n}\varepsilon_i K^{(\gamma+2+j)}\big(h^{-1}(x_i-t)\big).$$

Then Lemma C.3 in Section C implies that there is a constant $C > 0$ depending only on K and σ_g such that

$$\mathbb{E}_f\Big[\sup_{t\in[0,1]}\big|\hat{\psi}^{(j)}_{h,n}(t) - \mathbb{E}_f[\hat{\psi}^{(j)}_{h,n}(t)]\big|\Big] \leq C\,\sqrt{\log(1/h)}/\sqrt{nh^{2(\gamma+i)+1}},$$

provided $h_0 > 0$ is chosen appropriately (depending only on K and σ_g). The claim follows by Markov's inequality. Note that all the constants in the O-terms depend if necessary only on K, L and σ_g. □

Proof of Lemma 6.3. We compute

$$\begin{aligned} \mathsf{Var}_f(\hat{\psi}^{(j)}_{h,n}(t)) &= n^{-2}h^{-2(\gamma+j+1)}\sigma^2\sum_{i=1}^{n}\big[K^{(\gamma+2+j)}(h^{-1}(x_i-t))\big]^2 \\ &= n^{-1}h^{-2(\gamma+j)-1}\sigma^2\int\big[K^{(\gamma+2+j)}(x)\big]^2\,\mathrm{d}x \\ &\quad + O_{f\in\mathcal{F}_s,t\in[0,1]}\big((nh^{\gamma+j+1})^{-2}\big), \end{aligned}$$

where the order of the discretization error is derived as in the proof of Lemma 6.2, (i). □

Before turning to Lemma 6.4, we require additional technical results. The next lemma is an adaptation of Lemma 2 in Goldenshluger et al. (2006), compare also to Lemma 1 in Cheng and Raimondo (2008).

Lemma A.1 (Separation lemma)**.** *Let $h_0 > 0$ be so small that $\Theta \subset [h_0, 1-h_0]$ and $h \in (0, h_0$.) Further, let $q \in (0, x^*)$, where x^* is as in Assumption 2, (v). Given $\delta \in (0, qh)$, let $A_{\delta,h,f} = \{t \mid \delta < |t-\theta_f| < qh\}$. Then there are constants $C_i > 0$, $i = 1,2,3$, which can be chosen uniformly over $f \in \mathcal{F}_s$, and $t \in [h, 1-h]$ such that*

(i) $|\psi_{h,f}(\theta_f)| \le C_1 h^{s-\gamma}$,
(ii) if $\delta \ge C_2 h^{s-\gamma+1}$ *then*

$$\inf_{t \in A_{\delta,h,f}} (|\psi_{h,f}(t)| - |\psi_{h,f}(\theta_f)|) \ge C_3 \delta h^{-1}.$$

Moreover, the constants C_i depend only on the kernel K as well as on the Lipschitz constant L and the smoothness parameter s of $\mathcal{F}_s$ as in Definition 2.1, and where these constants can be chosen uniformly over a bounded range of values of s.

Proof of Lemma A.1. (i). Since $K^{(1)}(0) = 0$ we have $L_{h,0}(\theta_f) = 0$ in (6.1) and it follows that $|\psi_{h,f}(\theta_f)| \le C_1 h^{s-\gamma}$, where we choose C_1 as the constant for the $O_{f\in\mathcal{F}_s, t\in[h,1-h]}$-term in (6.1).

(ii). Given $t \in A_{\delta,h,f}$, $\tau = (\theta_f - t)/h$ satisfies $q > |\tau| \ge \delta/h$. From Assumption 2, (v), it follows that $|K^{(1)}(\tau)|[f^{(\gamma)}] \ge c_2\,|\tau|\,[f^{(\gamma)}] \ge c_2\, h^{-1}\delta[f^{(\gamma)}]$, where c_2 is the kernel constant in Assumption 2, (v). From the assumption $\delta \ge C_3\, h^{s-\gamma+1}$, (6.1) and the choice of C_1 it follows that

$$|\psi_{h,f}(t)| \ge c_2\, h^{-1}\delta[f^{(\gamma)}] - C_1 h^{s-\gamma} \ge C_2\, \delta h^{-1},$$

and in fact

$$\inf_{t \in A_{\delta,h,f}} (|\psi_{h,f}(t)| - |\psi_{h,f}(\theta_f)|) \ge C_2 \delta h^{-1}$$

for $C_2 := c_2\, C_3 - 2\, C_1 > 0$ for sufficiently large C_3. Note that all the constants depend only on K, L as well as s and are continuous in s, due to Lemma 6.1. ☐

Lemma A.2. *There are finite constants $C, C_1, C_2, h_0 > 0$ which only depend on K, σ, σ_g and on the Lipschitz constant L of $\mathcal{F}_s$ such that if $h \in (0, h_0)$, $n \in \mathbb{N}$ and $\zeta_n > 0$ are such that*

$$\zeta_n \sqrt{nh^{2\gamma+1}}/\sigma - C/\sqrt{nh} \ge \zeta_n \sqrt{nh^{2\gamma+1}}/2\,\sigma > C_1 \sqrt{\log(1/h)} > 0,$$

then

$$P_f\big(\sup_{t\in[0,1]} |\hat{\psi}_{h,n}(t) - \psi_{h,f}(t)| \ge \zeta_n\big) \le 2 \exp\left(-C_2 \zeta_n^2 n h^{2\gamma+1}\right), \quad f \in \mathcal{F}_s.$$

Proof of Lemma A.2. We let

$$Z_n(t;h) = \frac{1}{\sqrt{nh}\sigma} \sum_{i=1}^n \varepsilon_i K^{(\gamma+2)}(h^{-1}(x_i - t)),$$

$$R_n(t;h) = \frac{1}{\sqrt{nh}\sigma}\big(\mathbb{E}_f[\hat{\psi}_{h,n}(t)] - \psi_{h,f}(t)\big).$$

Then

$$\frac{\sqrt{nh^{2\gamma+1}}}{\sigma}\big(\hat{\psi}_{h,n}(t) - \psi_{h,f}(t)\big) = Z_n(t;h) + R_n(t;h),$$

and choosing h_0 small enough (depending only on K and L) obtain by Lemma 6.2, (i), for any $t \in [0,1]$ that $|R_n(t;h)| \leq C/\sqrt{nh}$, where $C > 0$ depends only on K and L. Thus, for an appropriate choice of the constants in the requirements of Lemma C.2 we have that

$$\begin{aligned} P_f\big(\sup_{t\in[0,1]} |\hat{\psi}_{h,n}(t) - \psi_{h,f}(t)| \geq \zeta_n\big) & \\ &\leq P_f\Big(\sup_{t\in[0,1]} |Z_{n,1}(t)| \geq \zeta_n\sqrt{nh^{2\gamma+1}}/\sigma - C/\sqrt{nh}\Big) \\ &\leq 2\exp\left(-C_1\zeta_n^2 nh^{2\gamma+1}\right). \end{aligned}$$

□

Lemma A.3. *Let $q \in (0, x^*)$, then there are finite constants $h_0, C_1, C_2 > 0$ depending only on K, σ, σ_g as well as on the Lipschitz constant L, the set Θ and the smoothness parameter s of $\mathcal{F}_s$, such that if $h \in (0, h_0)$ and $n \in \mathbb{N}$ are such that $nh^{2\gamma+1} \geq C_1 \log(1/h)$, then it holds that*

$$\begin{aligned} \max\big\{P_f(|\hat{t}_*(h;n) - t_*(h;f)| > hq/2), P_f(|\hat{t}^*(h;n) - t^*(h;f)| > hq/2)\big\} & \\ \leq 2\exp\left(-C_2 nh^{2\gamma+1}\right). \end{aligned}$$

Moreover, C_1, C_2 can be chosen uniformly over a bounded range of values of s, while h_0 can be chosen independently of s. In particular,

$$|\hat{t}_*(h;n) - t_*(h;f)| = o_{P,\mathcal{F}_s}(h), \qquad |\hat{t}^*(h;n) - t^*(h;f)| = o_{P,\mathcal{F}_s}(h).$$

Proof of Lemma A.3. We only show $P_f(|\hat{t}_* - t_*| > hq/2)$, the other inequality can be derived analogously.

Case (i): Suppose that $(-1)^{\gamma+1}[f^{(\gamma)}] > 0$.

Then by (6.1) for $j = 0$ it holds for sufficiently small h_0 (depending on Θ) that $\psi_{h,f}(t_*) < 0$ and in this case $\hat{t}_* = \arg\min_t \hat{\psi}_{h,n}(t)$. Hence, setting $B = \{t \in [0,1] \mid |t_* - t| > hq/2\}$, we have that

$$\begin{aligned} P_f(|\hat{t}_* - t_*| > hq/2) &\leq P_f\big(\exists t \in B:\ \hat{\psi}_{h,n}(t_*) \geq \hat{\psi}_{h,n}(t)\big) \\ &= P_f\big(\exists t \in B:\ \hat{\psi}_{h,n}(t_*) - \psi_{h,f}(t_*) + \psi_{h,f}(t) - \hat{\psi}_{h,n}(t) \\ &\qquad \geq \psi_{h,f}(t) - \psi_{h,f}(t_*)\big) \\ &\leq P_f\big(2\sup_{t\in[0,1]} |\psi_{h,f}(t) - \hat{\psi}_{h,n}(t)| \geq \inf_{t\in B}\left(\psi_{h,f}(t) - \psi_{h,f}(t_*)\right)\big). \end{aligned}$$

From Lemma 6.1, obtain for h_0 small enough (depending on Θ) that

$$\psi_{h,f}(t) - \psi_{h,f}(t_*) = (-1)^{\gamma+1}[f^{(\gamma)}]\,\Big(K^{(1)}(h^{-1}(\theta_f - t)) - K^{(1)}(h^{-1}(\theta_f - t_*))\Big)$$

$$+ O_{f\in\mathcal{F}_s, t\in[h,1-h]}(h^{s-\gamma}),$$

so that for constants $\tilde{C}_i > 0$ depending only on K, L and s we have that

$$\begin{aligned}
&\inf_{t\in B} \big(\psi_{h,f}(t) - \psi_{h,f}(t_*)\big) \\
&\geq \tilde{C}_1 \inf_{t\in B} (-1)^{\gamma+1}[f^{(\gamma)}] \left(K^{(1)}(h^{-1}(\theta_f - t)) - K^{(1)}(h^{-1}(\theta_f - t_*))\right) - \tilde{C}_2\, h^{s-\gamma} \\
&\geq \tilde{C}_1 \inf_{x:|x-a_*|>q/2} \left(K^{(1)}(x) - K^{(1)}(a_*)\right) - \tilde{C}_2\, h^{s-\gamma} \geq \tilde{C}_3,
\end{aligned}$$

where the second inequality follows by substitution and properties of $K^{(1)}$ and the last inequality is due to Lemma A.1 by choosing h_0 appropriately (depending on Θ). Using Lemma A.2, for sufficiently small h_0 (depending on K, σ, σ_g) and appropriate choice of C_1 in the assumption, there exists a constant $C_2 > 0$ such that

$$\begin{aligned}
P_f(|\hat{t}_* - t_*| > {}^{hq}/_{2}) &\leq P_f\big(\sup_{t\in[0,1]} |\hat{\psi}_{h,n}(t) - \psi_{h,f}(t)| \geq \tilde{C}_3/2\big) \\
&\leq 2\exp\left(-C_2 n h^{2\gamma+1}\right).
\end{aligned}$$

Note that C_1 and C_2 can be chosen depending only on K, σ, σ_g, L as well as s and also continuous in s, due to Lemma 6.1 and A.1, while the choice of h_0 is independent of s.

Case (ii): $(-1)^{\gamma+1}[f^{(\gamma)}] < 0$, then by (6.1) for $j = 0$ it holds for sufficiently small h_0 (depending on Θ) that $\psi_{h,f}(t_*) > 0$ and in this case $\hat{t}_* = \arg\max_t \hat{\psi}_{h,n}(t)$. Thus,

$$\begin{aligned}
P_f(|\hat{t}_* - t_*| > {}^{hq}/_{2}) &\leq P_f\big(\exists t \in B:\ \hat{\psi}_{h,n}(t) \geq \hat{\psi}_{h,n}(t_*)\big) \\
&= P_f\big(\exists t \in B:\ \hat{\psi}_{h,n}(t) - \psi_{h,f}(t) + \psi_{h,f}(t_*) - \hat{\psi}_{h,n}(t_*) \\
&\qquad \geq \psi_{h,f}(t_*) - \psi_{h,f}(t)\big) \\
&\leq P_f\big(2 \sup_{t\in[0,1]} |\psi_{h,f}(t) - \hat{\psi}_{h,n}(t)| \geq \inf_{t\in B} \big(\psi_{h,f}(t_*) - \psi_{h,f}(t)\big)\big).
\end{aligned}$$

We conclude with similar arguments as in case (i). □

Proof of Lemma 6.4. Let $h_0 > 0$ be so small that Lemmas 2.2, 6.1, A.2 and A.3 apply. Assume that $(-1)^{\gamma+1}[f^{(\gamma)}] > 0$ in which case $\psi_{h,f}(t_*) < 0$ and $\psi_{h,f}(t^*) > 0$. Let $\delta > 0$, then

$$\begin{aligned}
P_f\big(\{\hat{\psi}_{h,n}(\hat{t}_*) \geq 0\}\big) &\leq P_f\big(\{|\hat{\psi}_{h,n}(\hat{t}_*) - \psi_{h,f}(\hat{t}_*)| \geq \delta/2\}\big) \\
&\quad + P_f\big(\{|\psi_{h,f}(\hat{t}_*) - \psi_{h,f}(t_*)| \geq \delta/2\}\big) \\
&\quad + 1_{\{\psi_{h,f}(t_*)\geq -\delta\}}.
\end{aligned}$$

By choosing $\zeta_n = 1/\log(1/h)$ in Lemma A.2, the first term tends to zero. By (6.1) it follows that $\psi_{h,f}$ is Lipschitz-continuous with constant of order h^{-1}. Hence the second term tends to zero by Lemma A.3. Since we consider the case $\psi_{h,f}(t_*) < 0$, (compare to (6.1) for $j = 0$) the last term tends to zero for $\delta \to 0$.

Hence $\hat{\psi}_{h,n}(\hat{t}_*) < 0$, and similarly $\hat{\psi}_{h,n}(\hat{t}^*) > 0$ with high probability, and the continuity of $\hat{\psi}_{h,n}$ implies statement 1., since all estimates hold uniformly over $f \in \mathcal{F}_s$.

Statement 2. follows since the distance between t_* and t^* is exactly of order h by definition (see (2.3)), and the distance $\hat{t}_* - t_*$ as well as $\hat{t}^* - t^*$ is of order $o_{P,\mathcal{F}_s}(h)$ by Lemma A.3. Finally, since θ_f is at distance of order h both from t_* and t^*, statement 3. also follows from Lemma A.3. □

A.2. Proofs of auxiliary results in Section 6.2 and 6.4

Notation

We extend our notation: For $\mu \in \mathbb{R}^d$ and a positive semi-definite matrix Σ let $N_d(\mu, \Sigma)$ be the d-dimensional normal distribution with expectation μ and covariance matrix Σ. In this section we denote by $|| \cdot ||_2$ the Euclidean norm on $\mathbb{R}^d$ as well as the L_2-norm on the square integrable functions.

Consistency of the kink-location estimate

By construction of t_* and t^* it holds that $|\theta_f - \tilde{\theta}_{h,f}| = O_{\mathcal{F}_s}(h)$, see Lemma 2.2. In addition, Lemma 6.4 implies $|\theta_f - \hat{\theta}_{h,n}| = O_{P,\mathcal{F}_s}(h)$. However, we need the following lemma to ensure a faster rate of convergence to analyze the term (6.3) for the proof of Theorem 2.3.

Lemma A.4. *It holds that*

$$|\tilde{\theta}_{h,f} - \theta_f| = o_{\mathcal{F}_s}(h), \quad \textit{and} \quad |\hat{\theta}_{h,n} - \theta_f| = o_{P,\mathcal{F}_s}(h).$$

For the proof of Lemma A.4, we need the following consistency result, which is an extension of Theorem 5.9 in Van der Vaart (2000).

Proposition A.5. *For $f \in \mathcal{F}_s$ define the random set $\hat{\Theta}_h = \{w \in \mathbb{R} \,|\, \theta_f + hw \in [\hat{t}_*, \hat{t}^*]\}$ and let $\hat{g}_n : [0,1] \to \mathbb{R}$ be random functions and $g_f : [0,1] \to \mathbb{R}$ be a deterministic function. Suppose that*

$$\sup_{w \in \hat{\Theta}_h} |\hat{g}_n(w) - g_f(w)| = o_{P,\mathcal{F}_s}(1), \tag{A.2}$$

and that for sufficiently small $\epsilon > 0$,

$$\inf_{f \in \mathcal{F}_s} \inf_{|w| > \epsilon} |g_f(w)| > 0. \tag{A.3}$$

Then, for any sequence of estimators $(\hat{w}_{h,n})_h$ with $\hat{w}_{h,n} \in \hat{\Theta}_h$ and $\sup_{f \in \mathcal{F}_s} \hat{g}_n(\hat{w}_{h,n}) = o_P(1)$ it holds that

$$|\hat{w}_{h,n}| = o_{P,\mathcal{F}_s}(1).$$

Proof of Proposition A.5. With (A.2) and the definition of $\hat{w}_{h,n}$ it follows that for any $\delta > 0$

$$P_f(|g_f(\hat{w}_{h,n})| > \delta) \leq P_f(|\hat{g}_n(\hat{w}_{h,n})| > \delta/2) + P_f(|\hat{g}_n(\hat{w}_{h,n}) - g_f(\hat{w}_{h,n})| > \delta/2)$$
$$\leq o_{\mathcal{F}_s}(1) + P_f(\sup_{w \in \hat{\Theta}_h} |\hat{g}_n(w) - g_f(w)| > \delta/2) = o_{\mathcal{F}_s}(1).$$

Given $\epsilon > 0$ choose $\eta > 0$ as the left side of (A.3). Then,

$$P_f(|\hat{w}_{h,n}| > \epsilon) \leq P_f(|g_f(\hat{w}_{h,n})| \geq \eta) = o_{\mathcal{F}_s}(1). \qquad \square$$

Proof of Lemma A.4. Define the dilated criterion function $\bar{\psi}_{h,n}(w) := \hat{\psi}_{h,n}(\theta_f + hw)$ and for $f \in \mathcal{F}_s$ define $\hat{\Theta}_h = \{w \in \mathbb{R} \,|\, \theta_f + hw \in [\hat{t}_*, \hat{t}^*]\}$. Note that for the zeros $\hat{w}_{h,n}$ of $\bar{\psi}_{h,n}$ over $\hat{\Theta}_h$ and the zeros $\hat{\theta}_{h,n}$ of $\hat{\psi}_{h,n}$ it holds that $\hat{w}_{h,n} = (\hat{\theta}_{h,n} - \theta_f)/h$. For any $w \in [0,1]$ define

$$\psi_f(w) := (-1)^{\gamma+1}[f^{(\gamma)}]\, K^{(1)}(w).$$

Then, by (6.1) for $j = 0$ in Lemma 6.1

$$\sup_{w \in \hat{\Theta}_h} |\psi_{h,f}(\theta_f + hw) - \psi_f(w)| = o_{P,\mathcal{F}_s}(1),$$

since $\hat{\Theta}_h \subset \{w \in \mathbb{R} \,|\, \theta_f + hw \in [h, 1-h]\}$. Thus, Lemma 6.2 implies

$$\sup_{w \in \hat{\Theta}_h} |\bar{\psi}_{h,n}(w) - \psi_f(w)| \leq \sup_{w \in \hat{\Theta}_h} |\bar{\psi}_{h,n}(w) - \psi_{h,f}(\theta_f + hw)|$$
$$+ \sup_{w \in \hat{\Theta}_h} |\psi_{h,f}(\theta_f + hw) - \psi_f(w)| = o_{P,\mathcal{F}_s}(1).$$

Further, for any $\epsilon \in (0, x^*)$ Assumption 2, (v) yields

$$\inf_{f \in \mathcal{F}_s} \inf_{|w| > \epsilon} |\psi_f(w)| \geq \inf_{|w| > \epsilon} a\, c_2 |w| \geq a\, c_2 \epsilon > 0.$$

Apply Proposition A.5, of which we have derived the assumptions in the latter two display by setting $\hat{g}_n = \bar{\psi}_{h,n}$ and $g_f = \psi_f$, to obtain

$$\frac{|\hat{\theta}_{h,n} - \theta_f|}{h} = |\hat{w}_{h,n}| = o_{P,\mathcal{F}_s}(1).$$

The assertion for $\tilde{\theta}_{h,f}$ follows analogously by noting that Proposition A.5 is also true for non-random functions $\hat{g}_n$ and deterministic $\hat{w}_{h,n}$. $\square$

Proof of Lemma 6.5

The following lemma immediately implies Lemma 6.5, due to Lemma A.4.

Lemma A.6. *Let $\check\theta, \ddot\theta \in \Theta$, where $\ddot\theta$ is random and $\check\theta$ is non-random. Then there exists an $h_0 > 0$ depending only on K, σ, σ_g, L and Θ such that if $h \in (0, h_0)$ and $n \in \mathbb{N}$, it holds for $j = 0, 1, 2$, that*

$$\begin{aligned}
|h^j\, \psi^{(j)}_{h,f}(\check\theta) - (-1)^{\gamma+1+j}[f^{(\gamma)}]\, K^{(1+j)}(0)| &= O_{\mathcal{F}_s}\big(h^{-1}|\check\theta - \theta_f|\big) + O_{\mathcal{F}_s}(h^{s-\gamma}),\\
|h^j\, \hat\psi^{(j)}_{h,n}(\ddot\theta) - (-1)^{\gamma+1+j}[f^{(\gamma)}]\, K^{(1+j)}(0)| &= O_{P,\mathcal{F}_s}\big(h^{-1}|\ddot\theta - \theta_f|\big) + O_{P,\mathcal{F}_s}(h^{s-\gamma})\\
&\quad + O_{P,\mathcal{F}_s}\Big(\sqrt{\frac{\log(n)}{nh^{2\gamma+1}}}\Big).
\end{aligned}$$

Moreover, the constants in the O-terms depend only on the kernel K as well as on the Lipschitz constant L and the smoothness parameter s of $\mathcal{F}_s$ as in Definition 2.1, where the constants are continuous in s.

Proof of Lemma A.6. Choosing h_0 appropriately, Lemma 6.1 implies

$$h^j\, \psi^{(j)}_{h,f}(\check\theta) = (-1)^{\gamma+1+j}[f^{(\gamma)}]\, K^{(1+j)}(h^{-1}(\theta_f - \check\theta)) + O_{\mathcal{F}_s}(h^{s-\gamma}).$$

Now, by the mean value theorem

$$\begin{aligned}
|K^{(1+j)}(h^{-1}(\theta_f - \check\theta)) - K^{(1+j)}(0)| &= |K^{(1+j)}(h^{-1}(\theta_f - \check\theta)) - K^{(1+j)}(h^{-1}(\theta_f - \theta_f))|\\
&\le ||\, K^{(2+j)}\, ||_\infty\, O_{P,\mathcal{F}_s}\big(h^{-1}|\check\theta - \theta_f|\big),
\end{aligned}$$

which yields the first assertion. Similarly, by Lemma 6.1 and equation (6.2) for a suitable choice of h_0, obtain

$$\begin{aligned}
h^j\, \hat\psi^{(j)}_{h,n}(\ddot\theta) &= h^j\, \psi^{(j)}_{h,f}(\ddot\theta)) + O_{P,\mathcal{F}_s}\Big(\sqrt{\frac{\log(n)}{nh^{2\gamma+1}}}\Big)\\
&= (-1)^{\gamma+1+j}[f^{(\gamma)}]\, K^{(1+j)}(h^{-1}(\theta_f - \ddot\theta)) + O_{\mathcal{F}_s}(h^{s-\gamma}) + O_{P,\mathcal{F}_s}\Big(\sqrt{\frac{\log(n)}{nh^{2\gamma+1}}}\Big).
\end{aligned}$$

Now, using a similar argumentation as before with the mean value theorem it follows that

$$|K^{(1+j)}(h^{-1}(\theta_f - \ddot\theta)) - K^{(1+j)}(0)| = O_{\mathcal{F}_s}\big(h^{-1}|\ddot\theta - \theta_f|\big),$$

which concludes the proof. Note that the constants in the O-terms depend only on K, σ_g, L as well as s and these constants can be chosen continuously in s, see Lemma 6.1. □

Neglibility of the remainder terms: proof of Lemma 6.6

Proof of Lemma 6.6. Let us start with $R_1(n,h)$ as defined in (6.20). Note that the second factor in (6.20) is a constant. Thus, we only need to investigate the first factor. By (6.2) for $j = 1$

$$h\, |\hat\psi^{(1)}_{h,n}(\ddot\theta) - \psi^{(1)}_{h,f}(\ddot\theta)| = O_{P,\mathcal{F}_s}\Big(\sqrt{\frac{\log(n)}{nh^{2\gamma+1}}}\Big).$$

Next, recall Theorem 2.3 as well as (3.4) which imply

$$|\check\theta - \ddot\theta| = O_{P,\mathcal{F}_s}(h^{s-(\gamma-1)}) + O_{P,\mathcal{F}_s}((nh^{2\gamma-1})^{-1/2}),$$

where $\breve{\theta}$ and $\ddot{\theta}$ as in (6.3) resp. (6.17). Therefore, by (6.1) for $j = 1$ we deduce that

$$\begin{aligned} h\,|\psi^{(1)}_{h,f}(\ddot{\theta}) - \psi^{(1)}_{h,f}(\breve{\theta})| &\leq [f^{(\gamma)}]|K^{(2)}(h^{-1}(\theta_f - \ddot{\theta})) - K^{(2)}(h^{-1}(\theta_f - \breve{\theta}))| + O_{\mathcal{F}_s}(h^{s-\gamma}) \\ &\leq [f^{(\gamma)}]\|\,K^{(3)}\,\|_\infty \frac{|\breve{\theta} - \ddot{\theta}|}{h} + O_{\mathcal{F}_s}(h^{s-\gamma}) \\ &= O_{P,\mathcal{F}_s}(h^{s-\gamma}) + O_{P,\mathcal{F}_s}((nh^{2\gamma+1})^{-1/2}). \end{aligned}$$

In summary, by the triangle inequality

$$h\,|\hat{\psi}^{(1)}_{h,n}(\ddot{\theta}) - \psi^{(1)}_{h,f}(\breve{\theta})| = O_{P,\mathcal{F}_s}(h^{s-\gamma}) + O_{P,\mathcal{F}_s}\Big(\sqrt{\frac{\log(n)}{nh^{2\gamma+1}}}\Big).$$

Hence, the enumerator in $R_1(h,n)$ is of order

$$O_{P,\mathcal{F}_s}(\sqrt{n}h^{2s-\gamma+1/2}) + O_{P,\mathcal{F}_s}\Big(\sqrt{\log(n)h^{2(s-\gamma)}}\Big) = o_{P,\mathcal{F}_s}(1), \tag{A.4}$$

since $\sqrt{nh^{2\gamma+1}}\,\psi_{h,f}(\theta_f) = O_{\mathcal{F}_s}(\sqrt{n}h^{s+1/2})$ by Lemma 6.1 and the assumption on the asymptotics of h and n in Theorem 3.1. Finally, $R_1(h,n)$ is $o_{P,\mathcal{F}_s}(1)$ as the denominator is asymptotically a constant unequal to zero by Lemma 6.5 resp. Lemma A.6 and Assumption 2, (ii).

Similarly, the terms in $R_2(h,n)$ are $o_{\mathcal{F}_s}(1)$ resp. $o_{P,\mathcal{F}_s}(1)$. To see this, we only analyze the enumerators in $R_2(h,n)$ as the denominators are both constant. Theorem 2.3 implies

$$\sqrt{nh^{2\gamma-1}}(\hat{\theta}_{h,n} - \theta_f) = O_{P,\mathcal{F}_s}(\sqrt{n}h^{s+1/2}) + O_{P,\mathcal{F}_s}(1).$$

Due to Lemma A.6 and Theorem 2.3

$$h^2\,\hat{\psi}^{(2)}_{h,n}(\dot{\theta}) = O_{P,\mathcal{F}_s}(h^{s-\gamma}) + O_{P,\mathcal{F}_s}((nh^{2\gamma+1})^{-1/2})),$$

such that the first term in $R_2(h,n)$ is of the same order as in (A.4). A similar argumentation shows that the second term in $R_2(h,n)$ is $o_{\mathcal{F}_s}(1)$. □

Asymptotic normality of the score vector: proof of Lemma 6.7

Proof of Lemma 6.7. By Lemma 6.2, (i), for $j = 0,1$, respectively, obtain

$$\begin{aligned} \sqrt{nh^{2(\gamma+j)+1}}\,(\psi^{(j)}_{h,f}(\theta_f) - \hat{\psi}^{(j)}_{h,n}(\theta_f)) &= (nh)^{-1/2}\sum_{i=1}^{n}\varepsilon_i K^{(\gamma+2+j)}\left(h^{-1}(x_i - \theta_f)\right) \\ &\quad + O_{\mathcal{F}_s}((nh)^{-1/2}) \\ &=: E^{(j)}_n(f) + o_{\mathcal{F}_s}(1), \end{aligned}$$

due to the assumed asymptotics of h and n. By Theorem D.2 the terms

$$\sqrt{nh^{2(\gamma+j)+1}}\,(\psi^{(j)}_{h,f}(\theta_f) - \hat{\psi}^{(j)}_{h,n}(\theta_f)) \qquad \text{and} \qquad E^{(j)}_n(f)$$

have the same asymptotic limit distribution (provided it exists and satisfies the assumption of Theorem D.2) for $j = 0,1$ respectively. For convenience set

$$\mathbf{E}_n(f) := \tilde{\Sigma}^{-1/2}\,(E^{(0)}_n(f), E^{(1)}_n(f))^T,$$

where

$$\tilde{\Sigma} = \sigma^2 \begin{pmatrix} \|\, K^{\gamma+2} \,\|_2^2 & 0 \\ 0 & \|\, K^{\gamma+3} \,\|_2^2 \end{pmatrix}.$$

Hence, we show for any $\mathbf{x} \in \mathbb{R}^2$,

$$\sup_{f \in \mathcal{F}_s} |P_f\big(\mathbf{E}_n(f) \leq \mathbf{x}\big) - \Phi_2(\mathbf{x})| = o(1), \tag{A.5}$$

which would conclude the proof. Note that $\mathbf{E}_n(f)$ depends on f only through θ_f, which is by definition of $\mathcal{F}_s$ element of Θ, a parameter of $\mathcal{F}_s$.

In order to prove (A.5), we intend to make use of the uniform version of the Lindeberg-Feller Theorem D.4, which can be applied since $\Phi_2(\cdot)$ does not depend on $\mathcal{F}_s$ and therefore $(\Phi_2(\cdot))_{f \in \mathcal{F}_s}$ fulfills the assumptions of the latter theorem. Thus, we compute the asymptotic covariance matrix of $\big(E_n^{(0)}(f), E_n^{(1)}(f)\big)^T$. By means of Lemma 6.3 for $j = 0, 1$, respectively, deduce

$$\begin{aligned} \mathsf{Var}_f(E_n^{(0)}(f)) &= \sigma^2 \|\, K^{(\gamma+2)} \,\|_2^2 + o_{\mathcal{F}_s}(1), \\ \mathsf{Var}_f(E_n^{(1)}(f)) &= \sigma^2 \|\, K^{(\gamma+3)} \,\|_2^2 + o_{\mathcal{F}_s}(1). \end{aligned}$$

Now, both $E_n^{(0)}(f)$ and $E_n^{(1)}(f)$ are centered such that their covariance is computed to be

$$\mathbb{E}_f\big[E_n^{(0)}(f)\, E_n^{(1)}(f)\big] = \frac{\sigma^2}{nh} \sum_{i=1}^n K^{(\gamma+2)}(h^{-1}(x_i - \theta_f))\, K^{(\gamma+3)}(h^{-1}(x_i - \theta_f)).$$

With a Riemann-sum approximation in a similar fashion as in the proof of Lemma 6.2 the latter term is

$$\begin{aligned} &\frac{\sigma^2}{h} \int K^{(\gamma+2)}(h^{-1}(x - \theta_f))\, K^{(\gamma+3)}(h^{-1}(x - \theta_f))\, \mathrm{d}x + O_{\mathcal{F}_s}((nh)^{-1}) \\ &\qquad = \sigma^2 \int K^{(\gamma+2)}(x)\, K^{(\gamma+3)}(x)\, \mathrm{d}x + o_{\mathcal{F}_s}(1) = o_{\mathcal{F}_s}(1), \end{aligned}$$

where the last equation holds since the function $x \mapsto K^{(\gamma+2)}(x)\, K^{(\gamma+3)}(x)$ is odd by Assumption 2, (ii). Thus,

$$\mathsf{Var}_f\Big(\big(E_n^{(0)}(f), E_n^{(1)}(f)\big)^T\Big) \to \tilde{\Sigma}$$

and the convergence holds uniformly over $\mathcal{F}_s$. Next, for any $\delta > 0$ we show that

$$\sup_{f \in \mathcal{F}_s} \sum_{i=1}^n \|\, \mathbf{a_i}(\theta_f) \,\|_2^2\, \mathbb{E}_f\Big[\varepsilon_i^2 1_{\{\|\, \mathbf{a_i}(\theta_f) \,\|_2\, |\varepsilon_i| > \delta\}}\Big] = o(1),$$

where $\mathbf{a_i}(\theta_f) := (nh)^{-1/2}\big(K^{(\gamma+2)}(h^{-1}(x_i - \theta_f)),\, K^{(\gamma+3)}(h^{-1}(x_i - \theta_f))\big)^T$ and here $\|\cdot\|_2$ denotes the euclidean distance. Note that

$$\sup_{f \in \mathcal{F}_s} \max_{1 \leq i \leq n} \|\, \mathbf{a_i}(\theta_f) \,\|_2^2 \leq (nh)^{-1} \max\{\|\, K^{(\gamma+2)} \,\|_\infty^2, \|\, K^{(\gamma+3)} \,\|_\infty^2\} = o(1).$$

Further, the computation of the covariance matrix has shown that

$$\sum_{i=1}^{n} ||\, \mathbf{a_i}(\theta_f)\, ||_2^2 \;\to ||\, K^{(\gamma+2)}\, ||_2^2 + ||\, K^{(\gamma+3)}\, ||_2^2 < \infty$$

and the convergence holds uniformly in $\mathcal{F}_s$. Hence, the former two displays lead us to

$$\sup_{f\in\mathcal{F}_s} \sum_{i=1}^{n} ||\, \mathbf{a_i}(\theta_f)\, ||_2^2\, \mathbb{E}_f\Big[\varepsilon_i^2 1_{\{||\, \mathbf{a_i}(\theta_f)\, ||_2\, |\varepsilon_i|>\delta\}}\Big]$$
$$\leq \sup_{f\in\mathcal{F}_s} \max_{1\leq i\leq n} \mathbb{E}_f\Big[\varepsilon_i^2 1_{\{||\, \mathbf{a_i}(\theta_f)\, ||_2\, |\varepsilon_i|>\delta\}}\Big] \sum_{i=1}^{n} ||\, \mathbf{a_i}(\theta_f)\, ||_2^2 = o(1). \qquad \square$$

A.3. An exponential inequality

The following exponential concentration inequality for the estimator of the location of the kink will be important for the construction of adaptive confidence sets.

Lemma A.7. *Let $\bar{C} > 0$ be some finite constant and $q \in (0, x^*)$, where x^* is as in Assumption 2, (v). There exist finite constants $C, C_1, C_2, h_0 > 0$ which only depend on K, σ, σ_g as well as on L, Θ and s of $\mathcal{F}_s$, such that if $h \in (0, h_0)$, $n \in \mathbb{N}$ and $\lambda_n > 0$ are such that*

$$\lambda_n \sqrt{nh^{2\gamma+1}}/\sigma - C/\sqrt{nh} \geq \lambda_n \sqrt{nh^{2\gamma+1}}/2\,\sigma > C_1 \sqrt{\log(1/h)} > 0,$$

and $\bar{C}\lambda_n/2 < qh$, then

$$P_f(|\tilde{\theta}_{h,f} - \hat{\theta}_{h,n}| > \bar{C}\lambda_n)$$
$$\leq 1_{\{|\tilde{\theta}_{h,f}-\theta_f|>\tau\, \bar{C}\lambda_n/2\}} + 2\exp(-C_2 \bar{C}^2 n h^{2\gamma-1} \lambda_n^2),$$

where $\tau \in (0,1)$. Moreover, C, C_1, C_2 can be chosen uniformly over a bounded range of values of s, while h_0 is independent of s.

Proof of Lemma A.7. Define the event

$$\Omega = \{|\hat{t}_*(h;n) - t_*| < {}^{hq}/_2\} \cap \{|\hat{t}^*(h;n) - t^*| < {}^{hq}/_2\}.$$

Then,

$$P_f(|\tilde{\theta}_{h,f} - \hat{\theta}_{h,n}| > \bar{C}\lambda_n)$$
$$\leq 1_{\{|\tilde{\theta}_{h,f}-\theta_f|>\tau\, \bar{C}\lambda_n/2\}} + P_f(\{|\theta_f - \hat{\theta}_{h,n}| > {}^{(1-\tau)\,\bar{C}\lambda_n}/_2\} \cap \Omega) + P_f(\Omega^c). \tag{A.6}$$

Lemma A.3 implies for sufficiently small h_0 (depending only on K, σ, σ_g, L and on Θ)

$$P_f(\Omega^c) \leq 2\exp(-\tilde{C}_1 n h^{2\gamma+1}), \tag{A.7}$$

where $\tilde{C}_1 > 0$ is some finite constant uniform for $\mathcal{F}_s$ (depending only on K, σ, σ_g, L and on Θ). Let $\delta_n = {}^{(1-\tau)\bar{C}\lambda_n}/2$, then on the event Ω it holds that $|\theta_f - \hat{\theta}_{h,n}| < hq$, such that on Ω

$$\{|\theta_f - \hat{\theta}_{h,n}| > \delta_n\} \subset \{\exists t \in A_{\delta_n,h,f} : |\hat{\psi}_{n,h}(\theta_f)| \geq |\hat{\psi}_{n,h}(t)|\},$$

where $A_{\delta_n,h,f}$ is defined as in Lemma A.1. The event $\{|\hat{\psi}_{n,h}(\theta_f)| \geq |\hat{\psi}_{n,h}(t)|\}$ can be rewritten as

$$\{|\hat{\psi}_{n,h}(\theta_f)| - |\psi_{h,f}(\theta_f)| + |\psi_{h,f}(t)| - |\hat{\psi}_{n,h}(t)| \geq |\psi_{h,f}(t)| - |\psi_{h,f}(\theta_f)|\}.$$

Hence the latter event is contained in

$$\{2 \sup_{t\in[0,1]} |\hat{\psi}_{n,h}(t) - \psi_{h,f}(t)| \geq \inf_{t\in A_{\delta_n,h,f}} (|\psi_{h,f}(t)| - |\psi_{h,f}(\theta_f)|)\}.$$

With Lemma A.1, derive for appropriate choice of h_0 (depending only on Θ) that

$$\inf_{t\in A_{\delta_n,h,f}} (|\psi_{h,f}(t)| - |\psi_{h,f}(\theta_f)|\} \geq \tilde{C}_2\lambda_n h^{-1}$$

for some constant $\tilde{C}_2 > 0$ which depends only on K, L as well as s and is continuous in s. Thus, by means of Lemma A.2, for appropriate choice of the constants in the claim,

$$\begin{aligned} P_f(\{|\theta_f - \hat{\theta}_{h,n}| > \delta_n\} \cap \Omega) &\leq P_f(2 \sup_{t\in[0,1]} |\hat{\psi}_{n,h}(t) - \psi_{h,f}(t)| \geq \tilde{C}_2\bar{C}\lambda_n h^{-1}) \\ &\leq 2\exp(-\tilde{C}_3\bar{C}^2 nh^{2\gamma-1}\lambda_n^2) \end{aligned} \tag{A.8}$$

for some finite constant $\tilde{C}_3 > 0$ depending only on K, σ, σ_g, L as well as s and continuous in s. By assumption $\bar{C}\lambda_n/2 < qh$ so that one can find a suitable constant $C_2 > 0$ (depending only on K, σ, σ_g, L and on Θ) such that with (A.7) and (A.8)

$$\begin{aligned} &P_f(\Omega^c) + P_f(\{|\theta_f - \hat{\theta}_{h,n}| > \delta_n\} \cap \Omega) \\ &\qquad \leq \exp(-C_1 nh^{2\gamma+1}) + \exp(-C_3\bar{C}^2 nh^{2\gamma-1}\lambda_n^2) \leq \exp(-C_2\bar{C}^2 nh^{2\gamma-1}\lambda_n^2), \end{aligned}$$

which shows the first claim in view of (A.6). Finally, note that the choice of h_0 did not dependent on s and furthermore C, C_1, C_2 were chosen depending only on K, σ, σ_g, L as well as s and also continuous in s, due to Lemma 6.1 and A.1. □

A.4. Proof of Lemma 6.8

Before turning to Lemma 6.8, we list some simple properties of $B(k, s)$ and $\sigma(n, k)$.

Lemma A.8. *(i)* $h_{k_{min,n}} > \ldots > h_{k_{max,n}}$,

(ii) $B(\cdot, s)$ *is decreasing, while* $\sigma(n_2, \cdot)$ *is increasing,*

(iii) $k_{min,n} \cong \log(n)$, $k_{max,n} \cong \log(n)$ *and* $k_{max,n} - k_{min,n} \cong \log(n)$,

(iv) *(6.27) holds true, hence and* $B(k_n^*(s), s) \cong \sigma\big(n_2, k_n^*(s)\big) \cong \big(\log(n_2)/n_2\big)^{\frac{s-\gamma+1}{2s+1}}$,

(v) $B(m, s) \leq C_{\text{Lep}}\sigma(n_2,k)/8$ *for* $k_n^*(s) \leq k \leq m \leq k_{max,n}$,

(vi) $h_k \geq h_{k_{max,n}} > \sigma(n_2, k)$ *for any* $k \in \mathcal{K}_n$,

(vii) $nh_k^{2\gamma+1}\log(1/h_k)^{-1} \to \infty$ *for any* $k \in \mathcal{K}_n$,

(viii) *for any* $m \leq k \in \mathcal{K}_n$ *one has* $B(k, s) = 2^{(s-\gamma+1)(m-k)}B(m, s)$,

(ix) $\sigma(n_2, m) = \sigma(n_2, m+1)\big(2^{-\frac{2\gamma-1}{2}}\sqrt{m/(m+1)}\big)$.

Proof of Lemma A.8. (i), (ii), (viii) and (ix) are clear. From (3.7) obtain

$$k_{\min,n} \cong \frac{\log(n/\log(n))}{\log(2)(2\overline{s}+1)}, \quad \text{and} \quad k_{\max,n} \cong \frac{\log(n/\log(n)^2)}{\log(2)(2\gamma+1)},$$

which immediately implies (iii). With this it is straightforward to obtain (6.27) by balancing the terms in the definition of $k_n^*(s)$.
(v) follows by (iv) and (ii), as k and m are assumed to be greater or equal to $k_n^*(s)$. From (3.7) conclude that $\sigma(n_2, k_{\max,n}) \cong \sqrt{\log(n)}/n^{1/(2\gamma+1)}$, which is of a smaller order than $h_{k_{\max,n}}$. This shows (vi) due to (i) and (ii). Next, $nh_{k_{\max,n}}^{2\gamma+1}/\log(h_{k_{\max,n}}^{-1}) \cong \log(n)$, so that by (i) we conclude (vii). ☐

The following lemma is the essential step for the the proof of Lemma 6.8.

Lemma A.9. *(i) If* C_{Lep} *is chosen large enough and depending only on* K, σ, σ_g *as well as on* L *and* Θ *of* $\widetilde{\mathcal{F}}$ *and if* n *is sufficiently large, then there exists a finite constant* $c_1 > 0$ *which is uniform in* $\widetilde{\mathcal{F}}$, *such that*

$$P_f(\hat{k}_n = k) \leq c_1 2^{-k/c_1}, \quad \forall k > k_n^*(s_f).$$

(ii) If C_{Lep} *is chosen large enough and depending only on* K, σ, σ_g *as well as on* L *and* Θ *of* $\widetilde{\mathcal{F}}$, *and also if* n *is large enough such that* $k_n^*(s_f) \geq 2$, *then there exist* $\rho \in \mathbb{N}$ *and* $c_2 > 0$, *which are both uniform in* $\widetilde{\mathcal{F}}$, *such that*

$$P_f(\hat{k}_n = k) \leq c_2 2^{-k/c_2}, \quad \forall k < k_n^*(s_f) - \rho.$$

Moreover, ρ *depends only on* $b_1, b_2, \underline{s}$ *and on* C_{Lep}.

Proof of Lemma A.9. For convenience we write n for n_2 as this lemma depends only on the subsample S_2. Using Lemma A.8, (iii), we may assume that n is so large that $k \geq k_0$ for all $k \in \mathcal{K}_n$, where k_0 is the parameter in $\widetilde{\mathcal{F}}$ in (3.9).

From the definition of $B(k, s)$, (6.25) can be written as

$$\tilde{b_1}/\tilde{b_2}\ B(k, s_f) \leq |\tilde{\theta}_{h_k,f} - \theta_f| \leq B(k, s_f), \qquad f \in \widetilde{\mathcal{F}},\ k \geq k_0. \tag{A.9}$$

(i). Fix some $k \in \mathcal{K}_n$ with $k > k_n^*(s_f)$. From the definition of $\hat{k}_n$ in (3.11),

$$P_f(\hat{k}_n = k) \leq \sum_{j \in \mathcal{K}_n :\, j \geq k} P_f\big(|\hat{\theta}_{h_{k-1},n} - \hat{\theta}_{h_j,n}| > C_{\mathrm{Lep}}\sigma(n,j)\big). \tag{A.10}$$

From now on let $j \in \mathcal{K}_n$ such that $j \geq k$. Using (A.9) we estimate

$$|\hat{\theta}_{h_{k-1},n} - \hat{\theta}_{h_j,n}| \leq |\hat{\theta}_{h_{k-1},n} - \tilde{\theta}_{h_{k-1},f}| + |\hat{\theta}_{h_j,n} - \tilde{\theta}_{h_j,f}| + B(k-1,s_f) + B(j,s_f).$$

By Lemma A.8, (ii), and the definition of $k_n^*(s_f)$ we have that

$$B(k-1,s_f)+B(j,s_f) \leq 2B(k_n^*(s_f),s_f) \leq C_{\mathrm{Lep}}\sigma\big(n,k_n^*(s_f)\big)/4 \leq C_{\mathrm{Lep}}\sigma(n,j)/4.$$

Combining the latter two displays and Lemma A.7 with $\tau = 1/3$ and $\bar{C} = C_{\mathrm{Lep}}$ leads us for sufficiently large n to

$$\begin{aligned} &P_f\big(|\hat{\theta}_{h_{k-1},n} - \hat{\theta}_{h_j,n}| > C_{\mathrm{Lep}}\sigma(n,j)\big) \\ &\quad\leq P_f\Big(|\hat{\theta}_{h_{k-1},n} - \tilde{\theta}_{h_{k-1},f}| > \frac{3C_{\mathrm{Lep}}}{8}\,\sigma(n,j)\Big) \\ &\qquad + P_f\Big(|\tilde{\theta}_{h_j,f} - \hat{\theta}_{h_j,n}| > \frac{3C_{\mathrm{Lep}}}{8}\,\sigma(n,j)\Big) \\ &\quad\leq 1_{\{|\theta_f - \tilde{\theta}_{h_{k-1},f}| > C_{\mathrm{Lep}}\sigma(n,j)/8\}} + 2\exp\big(-C_1 C_{\mathrm{Lep}}^2 n h_{k-1}^{2\gamma-1}\,\sigma^2(n,j)\big) \\ &\qquad + 1_{\{|\tilde{\theta}_{h_j,f} - \theta_f| > C_{\mathrm{Lep}}\sigma(n,j)/8\}} + 2\exp\big(-C_2 C_{\mathrm{Lep}}^2 n h_j^{2\gamma-1}\,\sigma^2(n,j)\big) \end{aligned} \tag{A.11}$$

for some absolute constants $C_i > 0$, $i = 1,2$ depending only on K, σ, σ_g as well as on L, Θ and s_f of $\widetilde{\mathcal{F}_s}$, as in (3.10). Since the constants C_1 and C_2 can be chosen continuously in s_f by Lemma A.7 and $s_f \in [\underline{s}, \overline{s}]$, we can choose these constants uniformly in $\tilde{\tilde{\mathcal{F}}}$. Using Lemma A.8, (v), the deterministic terms in (A.11) vanish for n large enough. For $j \geq k$ we have $h_{k-1} > h_j$ (Lemma A.8, (i)). Hence if $C_{\mathrm{Lep}} > 0$ is chosen large enough (depending only on K, σ, σ_g, L and Θ), such that by Lemma A.8, (vi) in the second and (vii) in the third step we estimate

$$\begin{aligned} \exp\big(-C_1 C_{\mathrm{Lep}}^2 n h_{k-1}^{2\gamma-1}\,\sigma^2(n,j)\big) &\leq \exp\Big(-C_1 C_{\mathrm{Lep}}^2 n h_j^{2\gamma-1}\,\sigma^2(n,j)\Big) \\ &\leq \exp\Big(-C_1 C_{\mathrm{Lep}}^2 n h_j^{2\gamma+1}\Big) \leq C_3\, 2^{-j/C_3}, \end{aligned}$$

for some finite constant $C_3 > 0$ depending only on K, σ, σ_g as well as on L and Θ of $\tilde{\tilde{\mathcal{F}}}$. Similarly we estimate the last term in (A.11).

(ii). Fix some $k < k_n^*(s_f) - \rho$, where $\rho \in \mathbb{N}$ will be chosen below. By definition of $\hat{k}_n$ in (3.11) and since $k < k_n^*(s_f)$,

$$P_f(\hat{k}_n = k) \leq P_f\Big(|\hat{\theta}_{h_k,n} - \hat{\theta}_{h_{k_n^*(s_f)},n}| \leq C_{\mathrm{Lep}}\sigma\big(n,k_n^*(s_f)\big)\Big). \tag{A.12}$$

Further, by means of (A.9) and the reverse triangle inequality

$$|\hat{\theta}_{h_k,n} - \hat{\theta}_{h_{k_n^*(s_f)},n}| \geq \Big(\frac{\widetilde{b_1}}{\widetilde{b_2}}\Big) B(k,s_f) - B(k_n^*(s_f),s_f) - |\hat{\theta}_{h_k,n} - \tilde{\theta}_{h_k,f} - \hat{\theta}_{h_{k_n^*(s_f)},n} + \tilde{\theta}_{h_{k_n^*(s_f)},f}|. \quad \text{(A.13)}$$

By using Lemma A.8, (viii), twice as well as the fact that $k_n^*(s_f) - k > \rho$ and $s \geq \underline{s}$ yields

$$\begin{aligned}\Big(\frac{\widetilde{b_1}}{\widetilde{b_2}}\Big) B(k,s_f) - B(k_n^*(s_f),s_f) &= \Big(\frac{\widetilde{b_1}}{\widetilde{b_2}} 2^{(s-\gamma+1)(k_n^*(s_f)-k)} - 1\Big) B(k_n^*(s_f),s_f) \\ &= \Big(\frac{\widetilde{b_1}}{\widetilde{b_2}} 2^{(s-\gamma+1)(k_n^*(s_f)-k)} - 1\Big) 2^{-(s-\gamma+1)}\, B(k_n^*(s_f)-1,s_f) \\ &> \Big(\frac{\widetilde{b_1}}{\widetilde{b_2}} 2^{(s-\gamma+1)(\rho-1)} - 2^{-(s-\gamma+1)}\Big)\, B(k_n^*(s_f)-1,s_f) \\ &\geq \Big(\frac{\widetilde{b_1}}{\widetilde{b_2}} 2^{(\underline{s}-\gamma+1)(\rho-1)} - 2^{-(\underline{s}-\gamma+1)}\Big)\, B(k_n^*(s_f)-1,s_f).\end{aligned}$$

Next, by Lemma A.8, (ii), (iv) and (ix),

$$\begin{aligned}B(k_n^*(s_f)-1,s_f) &\geq C_{\mathrm{Lep}}\sigma(n,k_n^*(s_f)-1)/8 = C_{\mathrm{Lep}}\,\sigma\big(n,k_n^*(s_f)\big) 2^{\frac{-(2\gamma+5)}{2}} \Big(\sqrt{1+k_n^*(s_f)^{-1}}\Big)^{-1} \\ &\geq 2^{-(\gamma+3)}\, C_{\mathrm{Lep}}\sigma\big(n,k_n^*(s_f)\big),\end{aligned}$$

where we used for the last inequality that due to $k_n^*(s_f) \geq 2$ we have that

$$\sqrt{(k_n^*(s_f)-1)/k_n^*(s_f)} \geq 2^{-1/2}.$$

Let

$$\tilde{C} := 2^{-(\gamma+3)}\; C_{\mathrm{Lep}} \Big(\frac{\widetilde{b_1}}{\widetilde{b_2}} 2^{(\underline{s}-\gamma+1)(\rho-1)} - 2^{-(\underline{s}-\gamma+1)}\Big),$$

which can be made arbitrarily large by choosing ρ appropriately and depending only on $b_1, b_2, \underline{s}$ and on C_{Lep}. In view of (A.13) we have just shown that

$$|\hat{\theta}_{h_k,n} - \hat{\theta}_{h_{k_n^*(s_f)},n}| \geq \tilde{C}\sigma(n,k_n^*(s_f)) - |\hat{\theta}_{h_k,n} - \tilde{\theta}_{h_k,f} - \hat{\theta}_{h_{k_n^*(s_f)},n} + \tilde{\theta}_{h_{k_n^*(s_f)},f}|. \quad \text{(A.14)}$$

Thus, using (A.14) to bound (A.12) yields

$$\begin{aligned}P_f(\hat{k}_n = k) \leq\ & P_f\Big(|\hat{\theta}_{h_k,n} - \tilde{\theta}_{h_k,f}| \geq (\tilde{C} - C_{\mathrm{Lep}})\sigma\big(n,k_n^*(s_f)\big)/2\Big) \\ & + P_f\Big(\hat{\theta}_{h_{k_n^*(s_f)},n} + \tilde{\theta}_{h_{k_n^*(s_f)},f}| \geq (\tilde{C} - C_{\mathrm{Lep}})\sigma\big(n,k_n^*(s_f)\big)/2\Big)\end{aligned}$$

and one can proceed similarly as in the first part for the term (A.11) by choosing $\tilde{C}$ suitable by choice of ρ. □

Proof of Lemma 6.8. Assume that C_{Lep} and n are sufficiently large, such that the statements of Lemma A.9 hold. Let s_f be as in Definition 3.2. On the one hand, with Lemma A.9, (i) it holds that

$$\sup_{f\in\widetilde{\mathcal{F}}} P_f(\hat{k}_n > k_n^*(s_f)) \le \sup_{f\in\widetilde{\mathcal{F}}} \sum_{k_n^*(s_f)<k\le k_{\max,n}} P_f(\hat{k}_n = k) \le c_1 \sum_{k_n^*(s_f)<k\le k_{\max,n}} 2^{-k/c_1} = o(1),$$

as c_1 is uniform in $\widetilde{\mathcal{F}}$ and due to the asymptotic behavior of the indices in $\mathcal{K}_n$, see Lemma A.8, (iii). On the other hand, $P_f(\hat{k}_n < k_n^*(s_f) - \rho) = o_{\widetilde{\mathcal{F}}}(1)$ can be shown similarly using Lemma A.9, (ii). □

A.5. Lower bounds over $\widetilde{\mathcal{F}}_s$

Theorem A.10. *Given $\gamma, k_0 \in \mathbb{N}$ and $b_1, b_2, \overline{s} \in \mathbb{R}_+$ with $b_1 < b_2$ consider $s \ge \gamma + 1$ with $\lfloor s\rfloor - \gamma + 1 \in 2\mathbb{N}_0$. Further, assume that the kernel K in the definition (3.10) satisfies $\int_0^1 x^{\lfloor s\rfloor-\gamma-1}K^{(1)}(x)\,\mathrm{d}x \neq 0$. Then, in model (1.1) with $\varepsilon_i \sim N(0,\sigma^2)$ it holds for any loss function w that*

(i) $\liminf_n \inf_{\hat\theta} \sup_{f\in\widetilde{\mathcal{F}}_s} \mathbb{E}_f\left[w\left(n^{(s-\gamma+1)/(2s+1)}\,|\hat\theta - \theta_f|\right)\right] > 0$;
(ii) $\liminf_n \inf_{\hat\theta} \sup_{f\in\widetilde{\mathcal{F}}_s} \mathbb{E}_f\left[w\left(n^{(s-\gamma)/(2s+1)}\,|\hat\theta - [f^{(\gamma)}]|\right)\right] > 0$.

Proof. We address only (i), as (ii) can be shown with a similar approach. Fix some $\theta_0 \in \operatorname{int}(\Theta)$, and consider the function

$$T_s(x) = c_1\,(x-\theta_0)^s\,1_{[\theta_0,1]}(x),$$

where $c_1 \neq 0$ will be specified below. Set

$$f_0(x) := T_s(x) + R_\gamma(x;a), \quad x\in[0,1], \tag{A.15}$$

where R_γ as in (6.7). Here, f_0 has a kink of appropriate order in θ_0 because of R_γ (see the proof of Theorem 2.4, (i)), while T_s will take care of the condition (3.10). Choosing $c_1 \le L/(2\,s!)$ and following the lines of proof of Theorem 2.4, (i), it is straightforward to show with the triangle inequality that

$$f_0 \in \widetilde{\mathcal{F}_s}(b_1, b_2, k_0, \gamma, a, \Theta, L).$$

To verify (3.10) we show that for some suitable $h_0 > 0$,

$$b_1 h^{s-\gamma} \le |\psi_{h,f_0}(\theta_0)|, \quad \forall h\in(0,h_0), \tag{A.16}$$

since the upper bound follows from Lemma 6.1. To this end, if $(1-\theta_0)/h_0 \ge 1$ we compute

$$\frac{1}{h^{\gamma+1}}\int R_\gamma(x;a)\,K^{(\gamma+2)}\big(\frac{x-\theta_0}{h}\big)\,\mathrm{d}x = \int_0^1 u^\gamma\,K^{(\gamma+2)}(u)\,\mathrm{d}u$$
$$= (-1)^\gamma\,\gamma!\int_0^1 K^{(2)}(u)\,\mathrm{d}u = 0 \quad \text{(A.17)}$$

since $K^{(1)}(1) = K^{(1)}(0) = 0$, and

$$\frac{1}{h^{\gamma+1}}\int T_s(x)\,K^{(\gamma+2)}\big(\frac{x-\theta_0}{h}\big)\,\mathrm{d}x$$
$$= c_1\,h^{s-\gamma}\int_0^1 u^s\,K^{(\gamma+2)}(u)\,\mathrm{d}u$$
$$= c_1\,h^{s-\gamma}\,(-1)^{\gamma+1}\,\frac{s!}{(s-\gamma-1)!}\int_0^1 u^{s-\gamma-1}\,K^{(1)}(u)\,\mathrm{d}u,$$

which by assumption on $K^{(1)}$ is of order $h^{s-\gamma}$, so that (A.16) is satisfied for b_1 small enough.

For the sequence of alternative hypotheses, let $\theta_1 = \theta_0 + r_n \in \Theta$, and $r_n = o(1)$ is of the same order as r_n in the proof of Theorem 2.4, (i), and consider

$$f_1 = f_0 - (\nu_0 - \nu_n) + \tilde{T}_{s,n}, \quad \text{(A.18)}$$

where ν_0 resp. ν_n as in (6.9) resp. (6.10) and

$$\tilde{T}_{s,n}(x) := c_2\,\big((x-\theta_0)(\theta_1-x)\big)^{s/2}\,1_{[\theta_0,\theta_1]}(x)$$
$$+ c_3\,\big((x-\theta_1)(2\theta_1-\theta_0-x)\big)^{s/2}\,1_{[\theta_1,2\theta_1-\theta_0]}(x),$$

where $c_2 \neq c_3 > 0$ are suitable constants such that the derivatives of f_1 have the appropriate Lipschitz- resp. Hölder-constant L. In the spirit of (A.16) we check for some suitable $h_0 > 0$

$$b_1 h^{s-\gamma} \le |\psi_{h,f_1}(\theta_1)|, \quad \forall h \le h_0. \quad \text{(A.19)}$$

Decompose the probe-functional into five parts

$$\psi_{h,f_1}(\theta_1) = \frac{1}{h^{\gamma+1}}\int f_1(x)\,K^{(\gamma+2)}(h^{-1}(x-\theta_1))\,\mathrm{d}x$$
$$= \frac{1}{h^{\gamma+1}}\int T_s(x)\,K^{(\gamma+2)}(h^{-1}(x-\theta_1))\,\mathrm{d}x$$
$$+ \frac{1}{h^{\gamma+1}}\int R_\gamma(x)\,K^{(\gamma+2)}(h^{-1}(x-\theta_1))\,\mathrm{d}x$$
$$- \frac{1}{h^{\gamma+1}}\int \nu_0(x)\,K^{(\gamma+2)}(h^{-1}(x-\theta_1))\,\mathrm{d}x$$
$$+ \frac{1}{h^{\gamma+1}}\int \nu_n(x)\,K^{(\gamma+2)}(h^{-1}(x-\theta_1))\,\mathrm{d}x$$
$$+ \frac{1}{h^{\gamma+1}}\int \tilde{T}_{s,n}(x)\,K^{(\gamma+2)}(h^{-1}(x-\theta_1))\,\mathrm{d}x$$

$$=: (\tilde{A}) + (\tilde{B}) - (\tilde{C}) + (\tilde{D}) + (\tilde{E}).$$

Without loss of generality assume that h_0 is so small that ${}^{1-\theta_0}/_{h_0} \geq 1$ and ${}^{-\theta_0}/_{h_0} \leq -1$. Now, $(\tilde{B}) = 0$ as well as $(\tilde{C}) = 0$ for sufficiently small h_0 can be shown similarly as in (A.17). Since $\theta_1 \neq \theta_0$ independently of h we have as well that $(\tilde{A}) = 0$ for sufficiently small h_0. In addition, since $\tilde{T}_{s,n}$ is a piecewise polynomial with a discontinuity in the s-th derivative at θ_1, one can show $(\tilde{E}) \cong h^{s-\gamma}$ similarly as ${}^1/_{h^{\gamma+1}} \int T_s(x)\, K^{(\gamma+2)}\big(\frac{x-\theta_0}{h}\big)\, \mathrm{d}x \cong h^{s-\gamma}$ above. Finally, since $\nu_n \in C^\infty$ it follows by integration by parts, Taylor expansion around θ_1 and since $K^{(1)}$ is of order $\lfloor \overline{s} \rfloor - \gamma$ (Assumption 2, (iii)) that

$$\begin{aligned}
(\tilde{D}) &= h \int \nu_n^{(\gamma+1)}(\theta_1 + xh)\, K^{(1)}(x)\, \mathrm{d}x \\
&= \frac{h^{\lfloor \overline{s} \rfloor - \gamma + 1}\, \nu_n^{(\lfloor \overline{s} \rfloor + 1)}(\theta_1)}{(\lfloor \overline{s} \rfloor - \gamma)!} \int x^{\lfloor \overline{s} \rfloor - \gamma}\, K^{(1)}(x)\, \mathrm{d}x + o(h^{\lfloor \overline{s} \rfloor - \gamma + 1}) \\
&= O(h^{\lfloor \overline{s} \rfloor - \gamma + 1}),
\end{aligned}$$

which is compared to $O(h^{s-\gamma})$ of a negligible order for any $s \in [\gamma + 1, \overline{s}]$. All things considered we have verified (A.19) for b_1 small enough.

Concerning the Kullback-Leibler distance between f_1 and f_0, derive similar to (6.15) that

$$\begin{aligned}
\frac{1}{2\sigma^2} \sum_{i=1}^n (\nu_0(x_i) - \nu_n(x_i) - \tilde{T}_{s,n}(x_i))^2 \leq \frac{1}{\sigma^2} \sum_{i=1}^n (\nu_0(x_i) - \nu_n(x_i))^2 \\
+ \frac{1}{\sigma^2} \sum_{i=1}^n \tilde{T}_{s,n}(x_i)^2.
\end{aligned}$$

The first term on the right-hand-side of the latter display can be dealt with as in the proof of Theorem 2.4, (i), while the second term is asymptotically negligible. Indeed, firstly obtain that

$$\begin{aligned}
||\, \tilde{T}_{s,n}\, ||_\infty = |\tilde{T}_{s,n}\big({}^{\theta_0+\theta_1}/_2\big)| &\leq \max\{c_2, c_3\} \big({}^{\theta_1-\theta_0}/_2\big)^s = \max\{c_2, c_3\} \big({}^{r_n}/_2\big)^s \\
&= \max\{c_2, c_3\} \big({}^{\tilde{C}\, b_n^{s-\gamma+1}}/_2\big)^s
\end{aligned}$$

and secondly $\tilde{T}_{s,n}$ is non-zero only inside the interval $[\theta_0, 2\theta_1 - \theta_0]$, which has Lebesgue measure $2r_n$ and consequently only up to $2nr_n$ summands in $\sum_{i=1}^n \tilde{T}_{s,n}(x_i)^2$ are not zero due to equidistant design. Therefore,

$$\frac{1}{\sigma^2} \sum_{i=1}^n \tilde{T}_{s,n}(x_i)^2 \leq \text{const.} \cdot\, n\, b_n^{2s^2 - 2s\gamma + 3s - \gamma + 1},$$

where the last term is $O\big(nb_n^{2(2s+1)}\big)$, since $s \geq \gamma + 1$, and consequently negligible for the order of the Kullback-Leibler distance between f_1 and f_0. □

Appendix B: Construction of kernels and additional simulation results

B.1. Construction of kernels satisfying Assumption 2

In this section we adapt and simplify the construction in Cheng and Raimondo (2008) to kernels which shall satisfy Assumption 2 for $\gamma = 1$ and any given l. Compared to their $C_{\alpha,s}$, for the analysis in Lemma 6.1 in case of $\gamma = 1$ we require the first four derivatives of K to vanish at the boundary points instead of merely the first and second derivatives. We also indicate how to extend the construction to the case $\gamma \geq 2$.

The construction proceeds via the second derivative of K which corresponds to the function $\tilde{L}$ in the following lemma.

Lemma B.1. *For given $k \in \mathbb{N}$ let $\tilde{l} = 2k - 1$. Then*

$$\begin{aligned}\tilde{L}(x) =& \frac{(2k+1)(2k+2)(4k+3)(4k+5)(4k+7)}{2^{2k+4}(4k+9)} \\ &\times \sum_{i=0}^{k+3} (-1)^{k+4-i} \frac{(2(k+i))!}{(k+3-i)!(k+i)!(2i)!} x^{2i} 1_{[-1,1]}(x)\end{aligned} \tag{B.1}$$

fulfills

- *(i)* $\tilde{L}$ *is even,*
- *(ii)* $\tilde{L}$ *is infinitely often differentiable inside its support* $[-1, 1]$*,*
- *(iii)* $\tilde{L}^{(j)}(\pm 1) = 0$, $j = 0, 1, 2$,
- *(iv)* $\int x^m \tilde{L}(x)\,\mathrm{d}x = 0$ *for* $m = 0, \ldots, \tilde{l}$ *and* $\int x^{\tilde{l}+1} \tilde{L}(x)\,\mathrm{d}x = (-1)^{k+1}$,
- *(v)* $\tilde{L}(0) < 0$.

By taking an antiderivative of $\tilde{L}$ with value 0 at 0, we obtain the function $\bar{L}$ which corresponds to $K^{(1)}$.

Lemma B.2. *For given $k \in \mathbb{N}$ let $\tilde{l} = 2k - 1$. Then*

$$\begin{aligned}\bar{L}(x) =& \frac{(2k+1)(2k+2)(4k+3)(4k+5)(4k+7)}{2^{2k+4}(4k+9)} \\ &\times \sum_{i=0}^{k+3} (-1)^{k+4-i} \frac{(2(k+i))!}{(k+3-i)!(k+i)!(2i+1)!} x^{2i+1} 1_{[-1,1]}(x)\end{aligned} \tag{B.2}$$

fulfills

- *(a)* $\bar{L}$ *is odd,*
- *(b)* $\bar{L}$ *is infinitely often differentiable inside its support* $[-1, 1]$*,*
- *(c)* $\bar{L}^{(j)}(\pm 1) = 0$, $j = 0, 1, 2, 3$,
- *(d)* $\int x^m \bar{L}(x)\,\mathrm{d}x = 0$ *for* $m = 0, \ldots, \tilde{l} - 1$,
- *(e) there exists* $x^* \in (0, 1)$ *and* $c_2 > 0$ *such that* $|\bar{L}(x)| \geq c_2 |x|$ *for any* $x \in (-x^*, x^*)$.

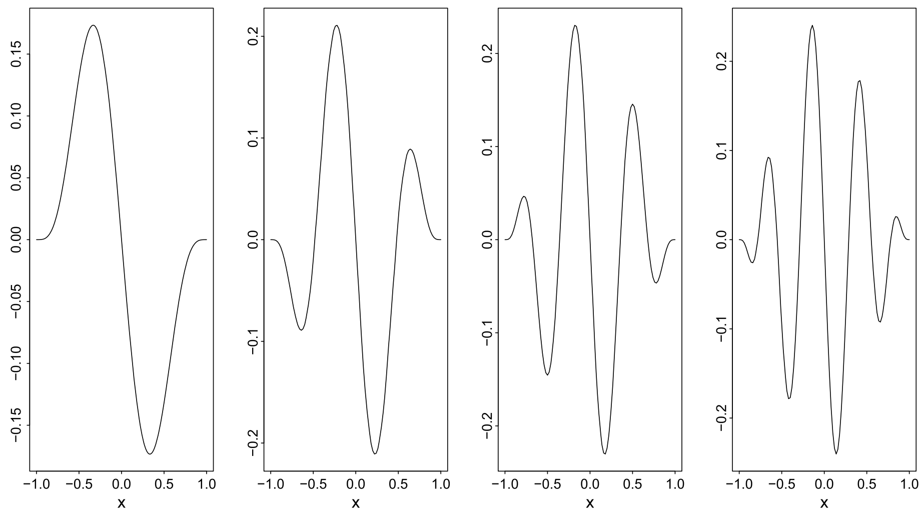

FIGURE 4. *$\bar{L}$ in (B.2) for $k = 1, 2, 3, 4$, respectively.*

Figure 4 illustrates the function $\bar{L}$ in (B.2) for different values of k. Some of the properties in the lemma are clearly visible in these plots.

Remark. The function $K^{(1)} = \bar{L}$ hence satisfies Assumption 2, (i)–(iii) and (v) for $\gamma = 1$ and $l = \tilde{l}$ as well as for $l = \tilde{l} + 1$. The condition (iv) of Assumption 2 is at least numerically true, as also the plots in Figure 4 suggest, though we did not provide a rigorous theoretical argument for this condition. ⋄

Proof of Lemma B.2. $\bar{L}$ is the anti-derivative of $\tilde{L}$ in Lemma B.1 with $\bar{L}(0) = 0$. Therefore (i) and (ii) of Lemma B.1 imply (a) and (b).

(c). Lemma B.1 (iii) means $\bar{L}^{(j)}(\pm 1) = 0$ for $j = 1, 2, 3$. Further, Lemma B.1 (iv) for $m = 0$ implies $0 = \int_{-1}^{1} \tilde{L}(x)\,\mathrm{d}x = 2\bar{L}(1)$ since $\bar{L}$ is odd, hence $\bar{L}(1) = \bar{L}(-1) = 0$.

(d). Since $\bar{L}$ is odd, the even moments vanish, thus, it suffices to consider $m \in \{0, \ldots, \tilde{l}\}$ odd. Then from Lemma B.1 (iv) we obtain

$$
\begin{aligned}
0 &= \int_{-1}^{1} x^{m+1} \tilde{L}(x)\,\mathrm{d}x = x^{m+1}\bar{L}(x)\Big|_{x=-1}^{x=1} - (m+1)\int_{-1}^{1} x^m \bar{L}(x)\,\mathrm{d}x \\
&= -(m+1)\int_{-1}^{1} x^m \bar{L}(x)\,\mathrm{d}x,
\end{aligned}
$$

since $x^{m+1}\bar{L}(x)$ is odd.

(e). As $\tilde{L}(0) \neq 0$ we can find $x^* \in (0, 1)$ and $c_2 > 0$ such that $|\bar{L}(x)| \geq c_2|x|$ for any $x \in [-x^*, x^*]$. □

Proof of Lemma B.1. We shall show that considering the expansion

$$\tilde{L}(x) = \sum_{i=0}^{k+3} p_{2i}\,\mathsf{P}_{2i}(x), \tag{B.3}$$

on $[-1, 1]$, where P_i denotes the i-th Legendre polynomial, that is

$$\mathsf{P}_i(x) = 2^{-i} \sum_{j=0}^{\lfloor i/2 \rfloor} (-1)^j \frac{(2i-2j)!}{j!(i-j)!(i-2j)!} x^{i-2j} 1_{[-1,1]}(x), \tag{B.4}$$

and requiring the conditions (i)-(v) leads to the expression in (B.1), which then actually does satisfy (i)-(v). Since the Legendre polynomials are orthogonal in $L_2([-1, 1])$ and the first m Legendre polynomials span the space of polynomials of degree $\leq m-1$, the moment condition (iv) will be satisfied if $p_{2i} = 0$, $i = 0, \ldots, k-1$, so that

$$\tilde{L}(x) = \sum_{i=k}^{k+3} p_{2i}\,\mathsf{P}_{2i}(x) = c\left(\mathsf{P}_{2k}(x) + c_1\,\mathsf{P}_{2k+2}(x) + c_2\,\mathsf{P}_{2k+4}(x) + c_3\,\mathsf{P}_{2k+6}(x)\right) \tag{B.5}$$

under the assumption that $p_{2k} = c \neq 0$. We shall arrange the coefficients c_i, $i = 1, 2, 3$ so that the boundary condition (iii) is satisfied. By (odd or even) symmetry, it suffices to satisfy the condition at $x = 1$. We have that

$$\mathsf{P}_{2i}(1) = 1, \qquad \mathsf{P}_{2i}^{(1)}(1) = {}^{2i(2i+1)}/_{2}, \quad \mathsf{P}_{2i}^{(2)}(1) = {}^{(2i-1)2i(2i+1)(2i+2)}/_{2^3} \tag{B.6}$$

Indeed, (B.6) is easily verified using Rodriguéz formula

$$\mathsf{P}_{2i} = \frac{1}{2^{2i}(2i)!} \frac{\partial^{2i}}{\partial x^{2i}} \left((x+1)^{2i}(x-1)^{2i}\right).$$

For example, for the second derivative compute

$$\begin{aligned}\frac{\partial^2 \mathsf{P}_{2i}(x)}{\partial^2 x}\Big|_{x=1} &= \frac{1}{2^{2i}(2i)!} \sum_{j=2}^{2i+2} \binom{2i+2}{j} \frac{(2i)!}{(2i-j)!}(x+1)^{2i-j} \frac{(2i)!}{(j-2)!}(x-1)^{j-2}\Big|_{x=1} \\ &= \frac{(2i-1)2i(2i+1)(2i+2)}{2^3},\end{aligned}$$

since only the term of $j = 2$ in the sum is non-zero. Now, to obtain $\tilde{L}(1) = \tilde{L}^{(1)}(1) = \tilde{L}^{(2)}(1) = 0$, insert (B.6) into (B.5) and solve the resulting linear system to obtain

$$c_1 = -\frac{3(4k+5)}{4k+9}, \quad c_2 = \frac{3(4k+3)}{4k+11}, \quad c_2 = -\frac{(4k+3)(4k+5)}{(4k+9)(4k+11)}. \tag{B.7}$$

Finally, the factor c in (B.5) is determined by the normalization in (iv),

$$(-1)^{k+1} = \int_{-1}^{1} x^{\tilde{l}+1} \tilde{L}(x)\,\mathrm{d}x = c \int_{-1}^{1} x^{2k} \mathsf{P}_{2k}(x)\,\mathrm{d}x$$

which using (B.4) and $|| \mathsf{P}_{2i} ||_2^2 = 2/(4i+1)$ yields

$$c = (-1)^{k+1} \frac{(4k+1)!}{2^{2k+1}((2k)!)^2},$$

so that explicitly

$$\tilde{L}(x) = \frac{(4k+1)!}{2^{2k+1}((2k)!)^2} \Big(\mathsf{P}_{2k}(x) - \frac{3(4k+5)}{4k+9} \mathsf{P}_{2k+2}(x) + \frac{3(4k+3)}{4k+11} \mathsf{P}_{2k+4}(x) - \frac{(4k+3)(4k+5)}{(4k+9)(4k+11)} \mathsf{P}_{2k+6}(x) \Big). \tag{B.8}$$

Inserting (B.4) into (B.8) gives the formula (B.1). Finally, using the recursion formula

$$\mathsf{P}_{2i+2}(0) = -\frac{2i+1}{2i+2} \mathsf{P}_{2i}(0), \quad i \in \mathbb{N},$$

implies that

$$\tilde{L}(0) = \frac{(4k+1)!}{2^{2k+1}((2k)!)^2} (-1)^{k+1} \mathsf{P}_{2k}(0) \Big(1 + \frac{3(2k+1)(4k+5)}{(2k+2)(4k+9)} + \frac{3(2k+1)(2k+3)(4k+3)}{(2k+2)(2k+4)(4k+11)} + \frac{(2k+1)(2k+3)(2k+5)(4k+3)(4k+5)}{(2k+2)(2k+4)(2k+6)(4k+9)(4k+11)} \Big).$$

Now, $\operatorname{sign} \mathsf{P}_{2k}(0) = (-1)^k$, so that $\tilde{L}(0) < 0$ as required. □

Remark. By using a sum from k to $k+\gamma+2$ in (B.5) the method can be extended, and (iii) can be satisfied for $\gamma \geq 2$.

B.2. Comparison with Mallik et al. (2013)

We compared our proposed confidence intervals for the kink-location with those of Mallik et al. (2013) by simulating observations within the same setting as in Section 5 of their paper. In particular, we considered the regression function $f(x) = (2(x-0.5))1_{(0.5,1]}(x)$ and normally distributed noise variables with zero mean and standard deviation $\sigma = 0.1$. The function has a kink of first order in $\theta = 0.5$. As in Mallik et al. (2013) we applied our method for over 5000 replications, where we used a grid $\mathcal{K}_n$ for every scenario such that the bandwidth values are inside an interval $[h_{min,n}, h_{max,n}]$, where the values of $h_{min,n}$ resp. $h_{max,n}$ are given in Table 7. For the Lepski-constant we used $C_{\text{Lep}} = 0.03$. The results can be found in Table 8, where we also display the results of Mallik et al. (2013) for comparison. As expected, our method (denoted by OCI) yields confidence intervals which are narrower than those of Mallik et al. (2013) (denoted by MCI) since they have milder assumptions on the smoothness of the regression function for their method. Nevertheless, f fulfills the assumptions of

TABLE 7
Choice of $[h_{min,n}, h_{max,n}]$ for the second scenario.

$n = 100$	$n = 500$	$n = 1000$	$n = 2000$
[0.52,0.57]	[0.44,0.48]	[0.36,0.41]	[0.29,0.33]

TABLE 8
Coverage probability and width (in parentheses) of the kink-location CI based on (3.16) (denoted by OCI) and the method of Mallik et al. (2013) (denoted by MCI) for the regression functions f for different grid sizes n.

n	90 % CI		95 % CI		99 % CI
	MCI	OCI	MCI	OCI	OCI
100	0.939 (0.448)	0.811 (0.060)	0.972 (0.559)	0.883 (0.071)	0.962 (0.093)
500	0.922 (0.258)	0.888 (0.039)	0.965 (0.346)	0.943 (0.047)	0.987 (0.061)
1000	0.911 (0.197)	0.897 (0.030)	0.959 (0.265)	0.951 (0.036)	0.989 (0.047)
2000	0.903 (0.153)	0.896 (0.024)	0.954 (0.205)	0.946 (0.028)	0.985 (0.037)

Mallik et al. (2013) as well as the assumptions for our setting and therefore, it seems reasonable to use our method in such cases.

Appendix C: Sub-Gaussian processes

Following Viens and Vizcarra (2007) we call a centered random variable ξ *sub-Gaussian relative to the scale* M, if for all $u > 0$

$$P(|\xi| > u) \leq 2 \exp\Big(- \frac{2\,u^2}{M^2} \Big). \tag{C.1}$$

Let $(\xi_i)_{i=1,\dots,n}$ be an i.i.d. sequence of random variables such that $\mathbb{E}[\xi_1] = 0$, $\mathbb{E}[\xi_1^2] = \sigma^2$ and ξ_1 is sub-Gaussian relative to the scale σ_g with $\sigma_g \geq \sigma > 0$. Define the process

$$Z_n(t;h) := (nh)^{-1/2} \sum_{i=1}^{n} \xi_i\, K(h^{-1}(x_i - t)), \quad t \in [0,1],$$

where $h < 1$ and $K : \mathbb{R} \to \mathbb{R}$ is a function with the following properties:

1. K is bounded and Lipschitz continuous with Lipschitz constant $C_K > 0$.
2. $\operatorname{supp}(K) = [-1, 1]$.

Define the semi-metric $\rho_{n,h} : [0,1]^2 \to \mathbb{R}_+$ by $\rho_{n,h}^2(s,t) = \mathbb{E}[(Z_n(s;h) - Z_n(t;h))^2]$. In the following we suppress the dependency of $\rho_{n,h}$ on n in the notation and just write ρ_h. This is due to our upper bound in (C.3) below. We write $N(\rho_h, T, \varepsilon)$ to denote the smallest number of ρ_h-balls of radius ε needed to cover $T \subset (0,1)$.

Lemma C.1. *The following statements are valid.*

(i) There exists a constant $c_g > 0$ depending only on K and σ_g such that for any $s, t \in [0,1]$ the random variable $c_g\big(Z_n(t;h) - Z_n(s;h)\big)$ is sub-Gaussian relative to the scale $\rho_h(s,t)$.

(ii) $diam_{\rho_h}[0,1] \leq 2\sigma_g^2 ||K||_\infty^2$.

(iii) *For any $T \subset [0,1]$, there exist some finite constants $C, h_0 > 0$ depending only on K and σ_g such that if $h \in (0, h_0)$*

$$\int_0^\infty \sqrt{\log(N(\rho_h, T, x))}\, \mathrm{d}x \le C\sqrt{\log(1/h)}.$$

Proof of Lemma C.1. Ad(i).

It holds that

$$\begin{aligned}\rho_h(s,t)^2 &= \mathbb{E}[|Z_n(s;h) - Z_n(t;h)|^2] \\ &= \sigma^2(nh)^{-1}\sum_{i=1}^n |K(h^{-1}(x_i - t)) - K(h^{-1}(x_i - s))|^2.\end{aligned} \tag{C.2}$$

Let $s, t \in [0,1]$ and $u > 0$. By means of the general Hoeffding inequality for sums of independent sub-Gaussian random variables (see for instance Theorem 2.6.3 in Vershynin (2018)) obtain

$$P(|Z_n(s;h) - Z_n(t;h)| > u) \le 2\exp\Big(-C\frac{u^2}{\rho_h(s,t)^2}\Big),$$

where $C > 0$ is some finite constant depending only on K and σ_g. Therefore, choosing $c_g > 0$ appropriately and depending only on K and σ_g we can observe from the latter display that

$$P(c_g\,|Z_n(s;h) - Z_n(t;h)| > u) \le 2\exp\Big(-2\frac{u^2}{\rho_h(s,t)^2}\Big).$$

This shows *(i)* in view of (C.1).

Ad(ii).

Now, the right-hand side of (C.2) can be bounded in two ways. On the one hand, let $A(n,s,t)$ denote the set of indices for which the latter sum is not zero. Due to the compact support of K and the design assumption we have that $|A(n,s,t)| \le C\,nh$, for some finite constant $C > 0$. Thus, with the Lipschitz continuity of K

$$\rho_h(s,t)^2 \le h^{-2}\,C\,C_K^2|t-s|^2\sigma_g^2,$$

since $\sigma_g \ge \sigma$. On the other hand, since K is bounded and due to the cardinality of $A(n,s,t)$ one has that $\rho_h(s,t)^2 \le 2\sigma_g^2||K||_\infty^2$. Hence,

$$\rho_h(s,t)^2 \le 2\sigma_g^2||K||_\infty^2 \wedge h^{-2}\,C^2\,C_K^2|t-s|^2\sigma_g^2, \quad \forall s,t \in [0,1]. \tag{C.3}$$

This yields $diam_{\rho_h}[0,1] \le 2\sigma_g^2||K||_\infty^2$.

Ad (iii).

Since the latter display relates the ρ-distance of s and t to their absolute distance it follows that for any $\varepsilon \in (0, diam_{\rho_h}[0,1])$ and any $T \subset [0,1]$ one has

$N(\rho_h, T, \varepsilon) \le C_1\, \lambda_1(T)\,(h\varepsilon)^{-1}$ for some appropriate constant $C_1 > 0$ depending only on K and σ_g. With this and if $h_0 > 0$ is chosen appropriately small depending on $diam_{\rho_h}[0,1]$ deduce for any $h \in (0, h_0)$

$$\begin{aligned}\int_0^\infty \sqrt{\log(N(\rho_h, T, x))}\,\mathrm{d}x &= \int_0^{diam_{\rho_h}[0,1]} \sqrt{\log(N(\rho_h, T, x))}\,\mathrm{d}x \\ &\le \sqrt{\log(1/h)}\, diam_{\rho_h}[0,1] \\ &\quad + \int_0^{diam_{\rho_h}[0,1]} \sqrt{\log(C_1\lambda_1(T)x^{-1})}\,\mathrm{d}x \\ &\le C_2\sqrt{\log(1/h)},\end{aligned}$$

for some finite constant $C_2 > 0$ depending only on K and σ_g. In view of (ii), the choice of h_0 depends only on K and σ_g as well which concludes the proof. □

Lemma C.2. *There exist constants $C_1, C_2, h_0 > 0$ depending only on K and σ_g, such that for any $\lambda > 0$ and $h \in (0, h_0)$ such that $\lambda > C_1\sqrt{-\log(h)}$, it holds that*

$$P\left(\sup_{t\in[0,1]} |Z_n(t;h)| \ge \lambda\right) \le 2\exp\big(-C_2\,\lambda^2\big).$$

Lemma C.3. *There exist constants $C, h_0 > 0$ depending only on K and σ_g, such that if $h \in (0, h_0)$ then*

$$\mathbb{E}\Big[\sup_{t\in[0,1]} |Z_n(t;h)|\Big] \le C\sqrt{\log(1/h)}.$$

To prove Lemma C.2 resp. Lemma C.3 we make use of Theorem 3.1 resp. Corollary 3.3 in Viens and Vizcarra (2007), of which we derived the requirements in Lemma C.1.

Proof of Lemma C.2. First, note by the triangle inequality that for any $t_0 \in [0,1]$ holds

$$\begin{aligned}P\Big(\sup_{t\in[0,1]} |Z_n(t;h)| \ge \lambda\Big) &\le P\Big(\sup_{t\in[0,1]} |Z_n(t;h) - Z_n(t_0;h)| \ge \lambda/2\Big) \\ &\quad + P\Big(|Z_n(t_0;h)| \ge \lambda/2\Big).\end{aligned} \tag{C.4}$$

Since $Z_n(t_0;h)$ is the sum of independent sub-Gaussian random variables, we can apply the general Hoeffding inequality (see for instance Theorem 2.6.3 in Vershynin (2018)) to obtain

$$P\left(|Z_n(t_0;h)| \ge \lambda/2\right) \le 2\exp(-\tilde{C}\lambda^2), \tag{C.5}$$

where $\tilde{C} > 0$ is a finite constant depending only on K and σ_g. Next, the process $Z = (c_g\, Z_n(t;h))_{t\in[0,1]}$ is separable and by Lemma C.1, (i), a sub-1th-Gaussian chaos field (see Definition 2.3 in Viens and Vizcarra (2007)) with respect to ρ.

Without loss of generality let us assume that the constant c_g in Lemma C.1, (i), is one. Otherwise, we consider the random process $\tilde{Z}_n(t,h) = c_g^{-1} Z_n(t,h)$ and incorporate the constant c_g within the constants C_1 and C_2. Thus, by Theorem 3.1 in Viens and Vizcarra (2007) for any $\lambda > \tilde{C}_1 M$, where $\tilde{C}_1 > 0$ is a finite constant depending only on K and σ_g and $M = \int_0^\infty \sqrt{\log(N(\rho_h, T, x))}\,\mathrm{d}x$ it holds for a suitable finite constant $\tilde{C}_2 > 0$ depending only on K and σ_g that

$$P\left(\sup_{t\in[0,1]} |Z_n(t;h) - Z_n(t_0;h)| \geq \lambda/2\right) \leq 2\exp\Big(-\tilde{C}_2\lambda^2\Big), \quad \forall t_0 \in [0,1]. \tag{C.6}$$

In view, of Lemma C.1, (iii), $M \leq C\sqrt{\log(1/h)}$ if $h < h_0$, where $h_0, C > 0$ are finite constants depending only on K and σ_g. Hence, choose $C_1 := C\,\tilde{C}_1$ and $C_2 := \tilde{C} + \tilde{C}_2$ to conclude the proof, due to (C.4), (C.5) and (C.6). □

Proof of Lemma C.3. As in the proof of Lemma C.2 we can assume without loss of generality that the constant c_g in Lemma C.1, (i), is one. Using Corollary 3.4. in Viens and Vizcarra (2007) yields the assertion by using the bound on the covering entropy in 3. of Lemma C.1. □

Appendix D: Uniform weak convergence theory

We use the following notation in this section adapted from Bengs and Holzmann (2019). By F_X we denote the cumulative distribution function of a random vector X and by P_X its law. Let ϕ_X be the characteristic function of a random vector X.

Moreover, we assume for this section that Θ is some arbitrary set and for any $\vartheta \in \Theta$, $(X_n^\vartheta)_{n\in\mathbb{N}}$ is a sequence of real-valued random vectors in $\mathbb{R}^d$. Likewise, for any $\vartheta \in \Theta$, let X^ϑ be random vectors in $\mathbb{R}^d$ with continuous distribution. We introduce some definitions for the remainder of this section.

Uniform convergence in distribution

We write $X_n^\vartheta \overset{D,\Theta}{\Longrightarrow} X^\vartheta$ if

$$\sup_{\vartheta\in\Theta} |F_{X_n^\vartheta}(\mathbf{x}) - F_{X^\vartheta}(\mathbf{x})| = o(1), \quad \forall \mathbf{x} \in \mathbb{R}^d.$$

In this case we say that X_n^ϑ converges uniformly over Θ in distribution to X^ϑ. We say that $(P_{X^\vartheta})_{\vartheta\in\Theta}$ is *uniformly absolutely continuous over* Θ *with respect to some continuous probability measure* Q, if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for any measurable $A \subset \mathbb{R}^d$ with $Q(A) < \delta$ one has that $\sup_{\vartheta\in\Theta} P_{X^\vartheta}(A) < \varepsilon$. Note that by continuous probability measure we mean that the measure of singletons is zero, i.e. $Q(\{x\}) = 0$ for any $x \in \mathbb{R}^d$.

Uniform weak convergence

Likewise, we define uniform weak convergence for probability measures. Let $(\mathcal{X}, d)$ be some metric space and let $\mathcal{A}$ be its Borel σ-algebra. Let for any $\vartheta \in \Theta$, $(\mu_n^\vartheta)_{n\in\mathbb{N}}$ be a sequence of probability measures on $(\mathcal{X}, \mathcal{A})$. Similarly, for any $\vartheta \in \Theta$, let μ^ϑ be a probability measure on $(\mathcal{X}, \mathcal{A})$. In the same manner as for the law of random vectors we define uniform absolute continuity over Θ for $(\mu^\vartheta)_{\vartheta\in\Theta}$, that is $(\mu^\vartheta)_{\vartheta\in\Theta}$ is *uniformly absolutely continuous over* Θ *with respect to some continuous probability measure* μ if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for any $A \in \mathcal{A}$ with $\mu(A) < \delta$ it follows that $\sup_{\vartheta\in\Theta} \mu^\vartheta(A) < \varepsilon$. Eventually, we say that μ_n^ϑ converges uniformly weakly over Θ to μ^ϑ and write $\mu_n^\vartheta \overset{w,\Theta}{\Longrightarrow} \mu^\vartheta$ if and only if

$$\sup_{\vartheta\in\Theta} \Big| \int g \,\mathrm{d}\mu_n^\vartheta - \int g \,\mathrm{d}\mu^\vartheta \Big| = o(1)$$

for any real-valued, bounded and continuous function $g : \mathcal{X} \to \mathbb{R}$. Proofs for the following results are provided in Bengs and Holzmann (2019).

Theorem D.1 (Uniform continuous mapping theorem)**.** *Let* $H : \mathbb{R}^d \to \mathbb{R}^s$ *be continuous. If* $(P_{X^\vartheta})_{\vartheta\in\Theta}$ *is uniformly absolutely continuous over* Θ *with respect to some continuous probability measure* Q *and* $X_n^\vartheta \overset{D,\Theta}{\Longrightarrow} X^\vartheta$, *then*

$$H(X_n^\vartheta) \overset{D,\Theta}{\Longrightarrow} H(X^\vartheta).$$

Theorem D.2. *Let for any* $\vartheta \in \Theta$, $(Y_n^\vartheta)_{n\in\mathbb{N}}$ *be a sequence of real-valued random vectors in* $\mathbb{R}^d$. *Suppose that* $X_n^\vartheta \overset{D,\Theta}{\Longrightarrow} X^\vartheta$ *and* $||\, Y_n^\vartheta - X_n^\vartheta \,||_2 = o_{P,\Theta}(1)$ *and in addition* $(P_{X^\vartheta})_{\vartheta\in\Theta}$ *is uniformly absolutely continuous over* Θ *with respect to some continuous probability measure* Q. *Then,*

$$Y_n^\vartheta \overset{D,\Theta}{\Longrightarrow} X^\vartheta.$$

Theorem D.3 (Uniform Slutzky's Theorem)**.** *Let for any* $\vartheta \in \Theta$, $(Y_n^\vartheta)_{n\in\mathbb{N}}$ *be a sequence of real-valued random vectors in* $\mathbb{R}^d$ *and* $(\mathbf{c}^\vartheta)_{\vartheta\in\Theta}$ *be deterministic real vectors in* $\mathbb{R}^d$ *with* $Y_n^\vartheta = \mathbf{c}^\vartheta + o_{P,\Theta}(1)$. *Furthermore, suppose* $(P_{X^\vartheta})_{\vartheta\in\Theta}$ *is uniformly absolutely continuous over* Θ *with respect to some continuous probability measure* Q, *and* $X_n^\vartheta \overset{D,\Theta}{\Longrightarrow} X^\vartheta$, *then*

$$X_n^\vartheta + Y_n^\vartheta \overset{D,\Theta}{\Longrightarrow} X^\vartheta + \mathbf{c}^\vartheta \quad \textit{and} \quad X_n^\vartheta \cdot Y_n^\vartheta \overset{D,\Theta}{\Longrightarrow} X^\vartheta \cdot \mathbf{c}^\vartheta,$$

where the multiplication is to be understood componentwise.

Theorem D.4 (Uniform Lindeberg-Feller Theorem)**.** *For each* $n \in \mathbb{N}$ *and* $\vartheta \in \Theta$ *let* $X_{n,i}^\vartheta$, $1 \le i \le n$ *be centered and independent random vectors in* $\mathbb{R}^d$. *Assume that* $(P_{X_{n,i}^\vartheta})_{\vartheta\in\Theta}$ *is uniformly absolutely continuous over* Θ *with respect to some continuous probability measure* Q. *Moreover, suppose that*

1. $\sup_{\vartheta \in \Theta} || \sum_{i=0}^{n} \mathbb{E}\big[X_{n,i}^{\vartheta}\big(X_{n,i}^{\vartheta}\big)^T\big] - \Sigma ||_2 = o(1)$, *for some semi-positive-definite matrix* $\Sigma \in \mathbb{R}^{d \times d}$;
2. *For any* $\varepsilon > 0$ *it holds that* $\limsup_n \sup_{\vartheta \in \Theta} \sum_{i=1}^{n} \mathbb{E}\big[|| X_{n,i}^{\vartheta} ||_2^2 \, 1_{|| X_{n,i}^{\vartheta} ||_2 > \varepsilon}\big] = 0$;

Then,

$$\sup_{\vartheta \in \Theta} \big| P\big(X_{n,1}^{\vartheta} + \ldots + X_{n,n}^{\vartheta} \leq \mathbf{x}\big) - \Phi_{\Sigma}(\mathbf{x}) \big| = o(1),$$

where Φ_{Σ} *is the cumulative distribution function of* $N(0, \Sigma)$.

Acknowledgments

The authors gratefully acknowledge financial support of the DFG, grant Ho 3260/5-1.

References

Bengs, V. and H. Holzmann (2019). Uniform approximation in classical weak convergence theory. https://arxiv.org/abs/1903.09864. [Online; accessed 26-March-2019].

Bull, A. D. (2012). Honest adaptive confidence bands and self-similar functions. *Electronic Journal of Statistics 6*, 1490–1516. MR2988456

Cai, T. and M. Low (2004). An adaptation theory for nonparametric confidence intervals. *The Annals of Statistics 32*(5), 1805–1840. MR2102494

Card, D., D. S. Lee, Z. Pei, and A. Weber (2015). Inference on causal effects in a generalized regression kink design. *Econometrica 83*(6), 2453–2483. MR3437431

Carlstein, E. G., H.-G. Müller, and D. Siegmund (1994). Change-point problems. IMS. MR1477909

Cheng, M. and M. Raimondo (2008). Kernel methods for optimal change-points estimation in derivatives. *Journal of Computational and Graphical Statistics 17*(1), 56–75. MR2424795

Dette, H., A. Munk, and T. Wagner (1998). Estimating the variance in nonparametric regression – what is a reasonable choice? *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60*(4), 751–764. MR1649480

Eubank, R. and P. Speckman (1994). Nonparametric estimation of functions with jump discontinuities. *Lecture Notes-Monograph Series*, 130–144. MR1477919

Frick, K., A. Munk, and H. Sieling (2014). Multiscale change point inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(3), 495–580. MR3210728

Giné, E. and R. Nickl (2010). Confidence bands in density estimation. *The Annals of Statistics 38*(2), 1122–1170. MR2604707

Goldenshluger, A., A. Juditsky, A. Tsybakov, and A. Zeevi (2008a). Change-point estimation from indirect observations. 1. minimax complexity. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 44*(5), 787–818. MR2453845

Goldenshluger, A., A. Juditsky, A. Tsybakov, and A. Zeevi (2008b). Change-point estimation from indirect observations. 2. adaptation. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 44*(5), 819–836. MR2453846

Goldenshluger, A., A. Tsybakov, and A. Zeevi (2006). Optimal change-point estimation from indirect observations. *The Annals of Statistics 34*(1), 350–372. MR2275245

Hall, P., J. Kay, and D. Titterinton (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika 77*(3), 521–528. MR1087842

Haynes, K., I. A. Eckley, and P. Fearnhead (2017). Computationally efficient changepoint detection for a range of penalties. *Journal of Computational and Graphical Statistics 26*(1), 134–143. MR3610414

Korostelev, A. (1988). On minimax estimation of a discontinuous signal. *Theory of Probability & Its Applications 32*(4), 727–730. MR0927265

Korostelev, A. and A. Tsybakov (1993). *Minimax Methods for Image Reconstruction*. Springer-Verlag New York, Inc. MR1224456

Li, K.-C. (1989). Honest confidence regions for nonparametric regression. *The Annals of Statistics 17*(3), 1001–1008. MR1015135

Loader, C. (1996). Change point estimation using nonparametric regression. *The Annals of Statistics 24*(4), 1667–1678. MR1416655

Low, M. (1997). On nonparametric confidence intervals. *The Annals of Statistics 25*(6), 2547–2554. MR1604412

Mallik, A., M. Banerjee, and B. Sen (2013). Asymptotics for p-value based threshold estimation in regression settings. *Electronic Journal of Statistics 7*(1), 2477–2515. MR3117104

Müller, H.-G. (1992). Change-points in nonparametric regression analysis. *The Annals of Statistics 20*(2), 737–761. MR1165590

Neumann, M. H. (1997). Optimal change-point estimation in inverse problems. *Scandinavian Journal of Statistics 24*(4), 503–521. MR1615339

Qiu, P. (2005). *Image Processing and Jump Regression Analysis*, Volume 599. John Wiley & Sons. MR2111430

Tsybakov, A. B. (2009). Introduction to nonparametric estimation. *Springer Series in Statistics*. MR2724359

Van der Vaart, A. (2000). *Asymptotic statistics*, Volume 3. Cambridge university press. MR1652247

Vershynin, R. (2018). *High-dimensional probability*. Cambridge university press. MR3837109

Viens, F. and A. Vizcarra (2007). Supremum concentration inequality and modulus of continuity for sub-nth chaos processes. *Journal of Functional Analysis 248*(1), 1–26. MR2329681

Von Neumann, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *The Annals of Mathematical Statistics 12*(4), 367–395. MR0006656

Wishart, J. (2009). Kink estimation with correlated noise. *Journal of the Korean Statistical Society 38*(2), 131–143. MR2649688

Wishart, J. (2011). Minimax lower bound for kink location estimators in a nonparametric regression model with long-range dependence. *Statistics & Probability Letters 81*(12), 1871–1875. MR2845903

Wishart, J. and R. Kulik (2010). Kink estimation in stochastic regression with dependent errors and predictors. *Electronic Journal of Statistics 4*, 875–913. MR2721037