ADAPTIVE FUNCTION ESTIMATION IN NONPARAMETRIC REGRESSION WITH ONE-SIDED ERRORS

BY MORITZ JIRAK[1], ALEXANDER MEISTER AND MARKUS REISS

*Humboldt Universität zu Berlin, Universität Rostock and
Humboldt Universität zu Berlin*

We consider the model of nonregular nonparametric regression where smoothness constraints are imposed on the regression function f and the regression errors are assumed to decay with some sharpness level at their endpoints. The aim of this paper is to construct an adaptive estimator for the regression function f. In contrast to the standard model where local averaging is fruitful, the nonregular conditions require a substantial different treatment based on local extreme values. We study this model under the realistic setting in which both the smoothness degree $\beta > 0$ and the sharpness degree $\mathfrak{a} \in (0, \infty)$ are unknown in advance. We construct adaptation procedures applying a nested version of Lepski's method and the negative Hill estimator which show no loss in the convergence rates with respect to the general L_q-risk and a logarithmic loss with respect to the pointwise risk. Optimality of these rates is proved for $\mathfrak{a} \in (0, \infty)$. Some numerical simulations and an application to real data are provided.

1. Introduction. In the standard model of nonparametric regression, the data

$$(1.1) \qquad Y_j = f(x_j) + \varepsilon_j, \qquad j = 1, \ldots, n$$

are observed. In this paper, in contrast to classical theory, the observation errors (ε_j) are not assumed to be centred, but to have certain support properties. This is motivated from many applications where rather the support than the mean properties of the noise are known and where the regression function f describes some frontier or boundary curve. Below we shall discuss concrete applications to sunspot data and annual sport records. Typical economical examples include auctions where the bidders' private values are inferred from observed bids (see Guerre et al. [25] or Donald and Paarsch [10]) and note the extension to bid and ask prices in financial data. Related phenomena arise in the context of inference for deterministic production frontiers, where it is assumed that f is concave (convex) or monotone.

A pioneering contribution in this area is due to Farrell [17], who introduced data envelopment analysis (DEA), based on either the conical hull or the convex hull of the data. This was further extended by Deprins et al. [9] to the free disposal Hull (FDH) estimator, whose properties have been extensively discussed in the literature; see, for instance, Banker [2], Korostelev et al. [43], Kneip et al. [37, 38], Gijbels et al. [19], Park et al. [52, 54], Jeong and Park [34] and Daouia et al. [6]. The issue of stochastic frontier estimation goes back to the works of Aigner et al. [1] and Meeusen and van den Broeck [49]; see also the more recent contributions of Kumbhakar et al. [45], Park et al. [53] and Kneip et al. [39].

In a general nonparametric setting the accuracy of the estimator heavily depends on the average number of observations in the vicinity of the support boundary. The key quantity is the sharpness $\mathfrak{a}_x > 0$ of the distribution function F_x of ε_j at $x = x_j$, which in its simplest case has polynomial tails

$$(1.2) \qquad F_x(y) = 1 - \mathfrak{c}'_x |y|^{\mathfrak{a}_x} + \mathcal{O}(|y|^{\mathfrak{a}_x + \delta}) \qquad \text{with } \mathfrak{c}'_x, \delta > 0 \text{ as } y \to 0.$$

The cases $0 < \mathfrak{a}_x < 1$, $\mathfrak{a}_x = 1$ and $\mathfrak{a}_x > 1$ are sometimes called *sharp boundary*, *fault-type boundary* and *nonsharp boundary*. From a theoretical perspective noise models with $\mathfrak{a}_x \in (0, 2)$ are nonregular (e.g., Ibragimov and Hasminskii [33]) since they exhibit nonstandard statistical theory already in the parametric case. Chernozhukov and Hong [4] discuss extensively parametric efficiency of maximum-likelihood and Bayes estimators in this context and show their relevance in economics.

From a nonparametric statistics point of view, Korostelev and Tsybakov [44] and Goldenshluger and Zeevi [23] treat a variety of boundary estimation problems. The focus is on applications in image recovery and is mathematically and practically substantially different from ours. The optimal convergence rate $n^{(-2\beta)/(\mathfrak{a}\beta+1)}$ over β-Hölder classes of regression functions f depends heavily on $\mathfrak{a}$ (not assumed to be varying in x); for $\mathfrak{a}_x \in (0, 2)$ it is faster than for local averaging estimators in standard mean regression and can even become faster than the regular squared parametric rate n^{-1}. Hall and van Keilegom [29] study a local-linear estimator in a closer related nonparametric regression model and establish minimax optimal rates in L_2-loss if the smoothness and sharpness parameters $\beta \in (0, 2]$ and $\mathfrak{a} > 0$ are known. Earlier contributions in a related setup are due to Härdle et al. [30], Hall et al. [26, 28] and Gijbels and Peng [20]. If the support of (ε_j) is not one-sided, but symmetric like $[-a, a]$ and $\beta \leq 1$, $\mathfrak{a} = 1$, Müller and Wefelmeyer [51] have shown that mid-range estimators attain also these better rates. Recently, Meister and Reiss [50] have proved strong asymptotic equivalence in Le Cam's sense between a nonregular nonparametric regression model for $\mathfrak{a} = 1$ and a continuous-time Poisson point process experiment.

All the references above consider a theoretically optimal bandwidth choice which depends on the unknown quantities $\mathfrak{a}$ and/or β. Completely data-driven adaptive procedures have been rarely considered in the literature because the

intrinsically nonlinear inference and the nonmonotonicity of the stochastic and approximation error terms block popular concepts from mean regression like cross-validation or general unbiased risk estimation; cf. the discussion in Hall and Park [27]. Recently, Chichignoud [5] was able to produce a β-adaptive minimax optimal estimator, which, however, uses a Bayesian approach hinging on the assumption that the law of the errors (ε_j) is perfectly known in advance (in fact, after log transform a uniform law is assumed). Moreover, a log factor due to adaptation is paid, which is natural only under pointwise loss. It remained open whether under a global loss function like an L_q-norm loss adaptation without paying a log factor is possible. For regular nonparametric problems Goldenshluger and Lepski [22] study adaptive methods and convergence rates with respect to general L_q-loss which is much more involved in the general case $q \geq 1$ than for $q = 2$.

It is therefore of high interest, both from a theoretical and a practical perspective, to establish a fully data-driven estimation procedure where the error distribution and the regularity of the regression function are unknown and to analyze it under local (pointwise) and global (L_q-norm) loss. In particular, neither $\mathfrak{a}$ nor β that determine the optimal convergence rate are fixed in advance. In this paper we introduce a fully data-driven $(\mathfrak{a}, \beta)$-adaptive procedure for estimating f and prove that it is minimax optimal over $\mathfrak{a}, \beta > 0$.

To ease the presentation, we restrict to equidistant design points $x_j = j/n$ on $[0, 1]$ and regression errors (ε_j) which are concentrated on the interval $(-\infty, 0]$. Given $x \in [0, 1]$ and an open neighborhood $\mathcal{N}(x) \subseteq [0, 1]$, the function $f : \mathcal{N}(x) \to \mathbb{R}$ is supposed to lie in the Hölder class $H_{\mathcal{N}(x)}(\beta, L)$ with $\beta, L > 0$. Note that $\beta = \beta_x$ and $L = L_x$ may vary in x. The $\langle\beta\rangle$-derivatives of all $f \in H_{\mathcal{N}(x)}(\beta, L)$ satisfy

$$\left| f^{(\langle\beta\rangle)}(y) - f^{(\langle\beta\rangle)}(z) \right| \leq L|y - z|^{\beta - \langle\beta\rangle}, \qquad y, z \in \mathcal{N}(x).$$

Here $\langle\beta\rangle = \max\{m \in \mathbb{N}_0 : m < \beta\}$ is the largest integer strictly smaller than β.

We consider the case where the ε_j are independent with individual distribution function F_{x_j} and tail quantile function

$$\mathcal{U}_{x_j}(y) = F_{x_j}^{\leftarrow}(1 - 1/y),$$

where $F_{x_j}^{\leftarrow}$ denotes the generalized inverse of F_{x_j}. Weakening the polynomial tail behavior in (1.2), our key structural condition is that for each $x \in [0, 1]$, there exist $\mathfrak{a}_x, \mathfrak{c}_x > 0$, $\mathfrak{b}_x \in \mathbb{R}$ and a slowly varying function $l_x(y)$, such that

$$(1.3) \qquad \mathcal{U}_x(y) = -\mathfrak{c}_x y^{-1/\mathfrak{a}_x} l_x(y),$$

where $l_x(y)$ satisfies uniformly for $x \in [0, 1]$ condition

$$(1.4) \qquad l_x(y) = \log(y)^{\mathfrak{b}_x} + \mathcal{O}\left(\log(y)^{\mathfrak{b}_x - 1}\right) \qquad \text{as } y \to \infty.$$

If (1.2) holds, then (1.3), (1.4) are valid with $\mathfrak{b}_x = 0$ (note that $\mathfrak{c}_x \neq \mathfrak{c}'_x$ in general; see Lemma 6.2 for the precise relation). The polynomial tail condition (1.2) is one

of the standard models in the literature; see de Haan and Ferreira [7], Härdle et al. [30], Hall and van Keilegom [29] or Girard et al. [21]. In this context, so called *second order conditions* are inevitable whenever one is interested in convergence rates or limit distributions involving estimates of $\mathfrak{a}_x$; see Beirlant et al. [3], de Haan and Ferreira [7] or Falk et al. [16]. Our second order condition (1.4) is rather mild when compared to examples from the literature; cf. [3, 7, 16, 21, 29, 30]. As will be explained in Section 3.2, a more general formulation seems to be impossible.

Let us point out two main conceptual results of this paper. First, we wish to extend the existing theory beyond the limitation $\beta_x \leq 2$ imposed by locally constant or linear approximations and to have a clear notion of stochastic and deterministic error for the nonlinear estimators. To this end we develop a linear program in terms of general local polynomials, based on a quasi-likelihood method, because the definition in Hall and van Keilegom [29] does not extend to polynomials of degree 2 or more in our setup. Then Theorem 3.1 below yields for the estimator a nontrivial decomposition in approximation and stochastic error. This decomposition is a key result for our analysis, and permits us to address the adaptation problem in full generality, thus abolishing the blockade mentioned in Hall and Park [27]. We can consider not only pointwise, but also the global L_q-norm as risk measure for the whole range $q \in [1, \infty)$. Technically, the optimal L_q-adaptation is much more demanding compared to the pointwise risk. It requires very tight deviation bounds since no additional $\log n$-factor widens the margin.

For adaptive bandwidth selection, we apply a nested variant of the Lepski [48] procedure with pre-estimated critical values. Careful adaptive pre-estimation is necessary since the distribution of (ε_j) is unknown and allowed to vary in x. The fact that the underlying sample (Y_j) is inhomogeneously shifted by f adds another level of complexity for the estimation of $\mathfrak{a}_x$ and $\mathfrak{b}_x$, which needs to be addressed by translation invariant estimators. The remarkable result of Theorem 3.3 is that for general L_q-loss we obtain the rate $n^{(-2\beta)/(\mathfrak{a}\beta+1)}(\log n)^{(2\mathfrak{a}_x\mathfrak{b}_x\beta_x)/(\mathfrak{a}_x\beta_x+1)}$ of convergence, the same as in the case of known (global) Hölder regularity β and known distribution of (ε_j). For pointwise loss the rate deteriorates to $(n/\log n)^{(-2\beta_x)/(\mathfrak{a}_x\beta_x+1)}(\log n)^{(2\mathfrak{a}_x\mathfrak{b}_x\beta_x)/(\mathfrak{a}_x\beta_x+1)}$; see Theorem 3.2 below.

In Section 4 it is shown that all our rates are minimax optimal for adaptive estimation. For regular mean regression these rates, inserting $\mathfrak{a}_x = 2$ and $\mathfrak{b}_x = 0$, and particularly the payment for adaptation on β_x under pointwise loss are well known. A priori it is, however, not at all obvious that in the nonregular case with Poisson limit experiments (Meister and Reiss [50]) exactly the same factor appears. Interestingly, we do not pay in the convergence rates for not knowing $\mathfrak{a}_x$, $\mathfrak{b}_x$. The lower bound in the "default-type boundary" case $\mathfrak{a}_x > 2$ with slower rates than in regular regression requires a completely new strategy of proof where not only alternatives for the regression function, but also for the error distributions are tested against each other.

In Section 5 we provide some numerical simulations in order to evaluate the finite sample performance of the estimator. Smaller values in $\mathfrak{a}_x$ indeed lead to

significantly improved estimation results. The bandwidth selection shows a quite different behavior from the regular regression case due to taking local extremes. Applications to empirical data from sunspot observations and annual best running times on 1500 m are presented. Most proofs are deferred to Section 6, and auxiliary lemmas and details regarding the sharpness estimation are given in the supplementary material [35].

2. Methodology.
Our approach is a local polynomial estimation based on local extreme value statistics. We fix some $x \in [0, 1]$ and consider the coefficients $(\hat{b}_j)_{j=0,\dots,\beta^*}$ which minimize the objective function

$$(2.1) \qquad (b_0, \dots, b_{\beta^*}) \mapsto \sum_{|x_i - x| \leq h_k} \sum_{j=0}^{\beta^*} b_j (x_i - x)^j,$$

under the constraints $Y_i \leq \sum_{j=0}^{\beta^*} b_j (x_i - x)^j$ for all i with $|x_i - x| \leq h_k$. Set $\tilde{f}_k(x) := \hat{b}_0$. As an estimator of f we define

$$(2.2) \qquad \tilde{f}_k : x \mapsto \tilde{f}_k(x), \qquad x \in [0, 1],$$

where the bandwidth $h_k > 0$ remains to be selected.

If $-\varepsilon_j$ is exponentially distributed and the regression function a polynomial of maximal degree β^* on the interval $[x - h_k, x + h_k]$, then $\tilde{f}_k(x)$ is the maximum likelihood estimator (MLE), whence the approach can be seen as a local quasi-MLE method; see also Knight [40]. The idea of local polynomial estimators in frontier estimation was already employed, for instance, in Hall et al. [28], Hall and Park [27] and Hall and van Keilegom [29]. However, in contrast to their local linear estimators (and their higher order extensions), the sum over the evaluations at x_i in the neighborhood of x is minimized instead of the area. This marks a substantial difference and is crucial for our setup. Already in the case of quadratic polynomials p, it might occur that the minimization of just the value $p(x)$ under the support constraints yields the inappropriate estimator $\tilde{f}_k(x) = -\infty$ if x is not a design point and $\tilde{f}_k(x) = Y_i$ if $x = x_i$ because a sufficiently steep parabola always fits the constraints. This problem is visualized in Figure 1, where Figure 1(a) corresponds to estimator (2.1), and Figure 1(b) to the minimization approach employed in the above references. Note that this problem may or may not occur in practice, but it poses an obstacle for the mathematical analysis. This is why we work with the base estimators defined in (2.1).

The calculation of our estimator only requires basic linear optimization, but its error analysis will be more involved. Note that the formulation as a linear program is particularly important for implementation purposes, since our adaptive procedure requires the computation of many sequential estimators as the bandwidth h_k increases.
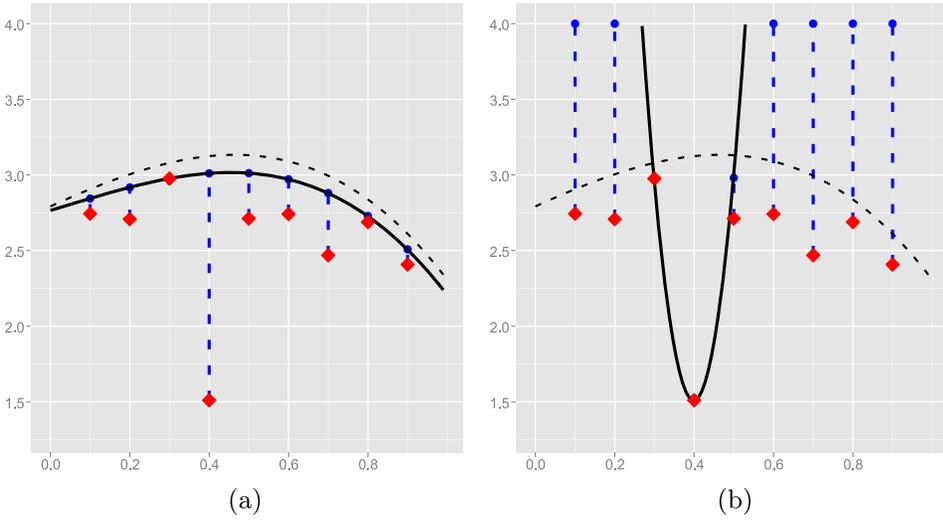
FIG. 1.   *Dashed line*: *true function*; *solid line*: *estimated function*; *red squares*: *sample points*; *estimation point*: *fourth sample point from the left*.

The adaptation problem consists of finding an (asymptotically) optimal bandwidth h_k when neither the regression function f nor the specific boundary behavior of the errors (ε_j) is known, which leads to different convergence rates. We follow the method inaugurated by Lepski [48] and consider geometrically growing bandwidths with $h_0 = n^{\mathfrak{h}_0 - 1}$, $\mathfrak{h}_0 \in (0, 1)$ and

$$h_k = h_0 \rho^k, \qquad k = 0, \ldots, K + 1$$

(2.3)

$$\text{where } \rho > 1 \text{ and } K = \lfloor \log_\rho (n^{1 - \mathfrak{h}_0}) \rfloor.$$

The purely data-driven estimator $\hat{f} := \tilde{f}_{\hat{k}}$ is defined as

$$(2.4) \qquad \hat{k} := \inf \{ k = 0, \ldots, K \, | \, \exists l \leq k : \| \tilde{f}_{k+1} - \tilde{f}_l \| > \hat{\mathfrak{z}}_l^T + \hat{\mathfrak{z}}_{k+1}^T \} \wedge K.$$

The critical values $\hat{\mathfrak{z}}_l^T$, $l = 0, \ldots, K + 1$ depend on the observations $\{Y_i\}_{1 \leq i \leq n}$, and will be specified below. The basic idea is to increase the bandwidth h_k as long as the distance (in some suitable seminorm $\| \cdot \|$) between the estimators is not significantly larger than the usual stochastic fluctuations of the estimators such that at $\hat{k}$ the bias is not yet dominating. In order to choose $\hat{\mathfrak{z}}_l^T$, the extreme-value index $\mathfrak{a}_x$, $\mathfrak{b}_x$ and the constant $\mathfrak{c}_x$ from equations (1.3) and (1.4) have to be estimated locally. For that purpose a quasi-negative-Hill method is developed in Section 3.2.

3. Asymptotic upper bounds.   In this section we will study the convergence rate of our estimator $\hat{f} = \tilde{f}_{\hat{k}}$ with $\hat{k}$ as defined in (2.4) when the sample size n tends to infinity. We will consider both the pointwise risk $\mathbb{E}_f |\hat{f}(x) - f(x)|^2$ for some fixed $x \in [0, 1]$ and the L_q-risk $\mathbb{E}_f \int_0^1 |\hat{f}(x) - f(x)|^q \, dx$ for $q \geq 1$. To deal

with the upper bounds, first some preparatory remarks and work are necessary. Throughout this section, we suppose:

ASSUMPTION 3.1. (i) $\mathfrak{c}_x, \mathfrak{b}_x, \mathfrak{a}_x \in H_{[0,1]}(\beta_0, L_0)$, where $\beta_0, L_0 > 0$ and $\inf_{x \in [0,1]} \mathfrak{a}_x, \mathfrak{c}_x > 0$,
(ii) $\max_{1 \le j \le n} \mathbb{E}[|\varepsilon_j|] < \infty$,
(iii) (ε_j) are independent, and the distribution of ε_j satisfies (1.3), (1.4).

For our theoretical treatment, an important quantity in the sequel is the approximative tail-function

$$(3.1) \qquad A_x(y) = -\mathfrak{c}_x y^{-1/\mathfrak{a}_x} \log(y)^{\mathfrak{b}_x},$$

since it asymptotically describes the quantile $\mathcal{U}_x(y)$.

3.1. *General upper bounds.* Most of our analysis relies on Theorem 3.1 and Proposition 3.1 below. These give rise to a decomposition, where the error for the implicitly defined base estimators $\tilde{f}_k(x)$ in (2.2) is split into a deterministic and a stochastic error part. Even though $\tilde{f}_k(x)$ is highly nonlinear, we obtain a relatively sharp and particularly simple upper bound.

THEOREM 3.1. *For any $x \in [0, 1]$ and $\beta_x \in (0, \beta^* + 1]$ there exist constants $c(\beta^*, L_x)$, $c(\beta^*)$ and $J(\beta^*)$, only depending on β^* and L_x, respectively, such that*

$$\big| \tilde{f}_k(x) - f(x) \big|$$
$$\le c(\beta^*, L) h_k^{\beta_x}$$
$$+ c(\beta^*) \max\{ |Z_j(h_k, x)| : j = 1, \dots, 2J(\beta^*), x + h_k \mathcal{I}_j \subseteq [0, 1] \}$$

holds true for all $f \in H_{\mathcal{N}(x)}(\beta, L)$ where

$$Z_j(h_k, x) := \max\{ \varepsilon_i : x_i \in x + h_k \mathcal{I}_j \} \quad \text{and}$$
$$\mathcal{I}_j := \big[-1 + (j-1)/J(\beta^*), -1 + j/J(\beta^*) \big].$$

REMARK 1. Interestingly, this decomposition holds true for any underlying distribution function F and dependence structure within (ε_j). Its proof is entirely based on nonprobabilistic arguments and has an interesting connection to algebra. A generalization to arbitrary dimensions or other basis functions than polynomials seems challenging.

We continue the range of the indices j of the (x_j, ε_j) from $\{1, \dots, n\}$ to $\mathbb{Z}$ while the equidistant location of the x_j and the independence of the ε_j is maintained. Then, Theorem 3.1 yields that with $c^* = c(\beta^*, L)$

$$(3.2) \quad \big| \tilde{f}_k(x) - f(x) \big| \le c^* h_k^{\beta_x} + c(\beta^*) \max\{ |Z_j(h_k, x)| : j = 1, \dots, 2J(\beta^*) \},$$

$$(3.3) \qquad \| \tilde{f}_k - f \|_q \le c^* h_k^{\beta_x} + c(\beta^*) \big\| \max\{ |Z_j(h_k, \cdot)| : j = 1, \dots, 2J(\beta^*) \} \big\|_q,$$

where $\| \cdot \|_q$ denotes the $L_q([0, 1])$-norm, $q \geq 1$. To pursue adaptivity, suppose that in terms of some seminorm $\| \cdot \|$, we can bound the error via

$$(3.4) \qquad \| \tilde{f}_k - f \| \leq R_k + B_k \qquad \forall k = 0, \ldots, K + 1, \, f \in H_{\mathcal{N}(x)}(\beta, L),$$

for some nonnegative random variables B_k, R_k, where B_k increases in k and R_k decreases in k. Neither the B_k nor the R_k depend on f, only on β_x and L_x. In the sequel B_k will be a bias upper bound while R_k is a bound on the stochastic error, which here—in contrast to usual mean regression—decays in k for each noise realisation. The following fundamental proposition addresses both, the pointwise and the L_q-risk of the adaptive estimator, since the pointwise distance of function values at some x as well as the L_q-distance of functions on $[0, 1]$ define seminorms for $q \geq 1$.

PROPOSITION 3.1. *Let $\| \cdot \|$ denote some seminorm, and let $\tilde{f}_k$, f lie in the corresponding normed space. Assume* (3.4) *and that the $\hat{\mathfrak{z}}_k^T$ decrease a.s. in k. Defining the oracle-type index*

$$(3.5) \qquad \hat{k}^* := \inf\{k = 0, \ldots, K - 1 : B_{k+1} > \hat{\mathfrak{z}}_{k+1}^T/2\} \wedge K,$$

we obtain for $q \geq 1$:

(a) $\mathbb{E}_f\big[\|\hat{f} - \tilde{f}_{\hat{k}^*}\|^q \mathbf{1}(\hat{k} > \hat{k}^*)\big]^{1/q} \leq \mathbb{E}_f\big[(\hat{\mathfrak{z}}_{\hat{k}^*}^T)^q\big],$

(b) $\mathbb{E}_f\big[\|\hat{f} - \tilde{f}_{\hat{k}^*}\|^q \mathbf{1}(\hat{k} < \hat{k}^*)\big]^{1/q}$

$$\leq 2^{(2q-1)/q} \mathbb{E}_f\big[\hat{\mathfrak{z}}_{\hat{k}^*}^q\big]^{1/q}$$

$$+ 2^{(2q-1)/q} \sum_{k=0}^{K-1} \mathbb{E}_f\big[R_k^q \mathbf{1}(\exists l \leq k : R_l > \hat{\mathfrak{z}}_l^T/2)\big]^{1/q}.$$

3.2. *Critical values and their estimation.* Our adaptive procedure and particularly the question of optimality crucially hinge on the (estimated) critical values $\hat{\mathfrak{z}}_k^T$, and thereby as a quantile for the distribution function $F_x(-y^{-1})$ as $y \to \infty$. In the literature (de Haan and Ferreira [7]), the standard, nonparametric quantile estimator is constructed via the approximation

$$(3.6) \qquad \mathcal{U}_x(ty) \approx \mathcal{U}_x(y) + a_x(t)\big(y^{-1/\mathfrak{a}_x} - 1\big)\mathfrak{a}_x \qquad \text{as } t, y \to \infty,$$

where the function $a_x(t)$ is a so-called *first-order scale function*. Unfortunately, this approach fails in our setup. The reason for this failure is the severely shifted sample (Y_j) (we do not observe ε_j) and the particular type of interpolation used in (3.6), which leads to an insufficient rate of convergence in the above approach. The bias that is induced by the shift will be present in any estimation method. This

fact makes us believe that under model (1.1), quantile estimation for general regular varying distributions is not possible. Since for any $t > 0$ we have the relation

$$F_x\big(A_x(n/t)\big)^n\big(1 + \mathcal{O}(1)\big) = F_x\big(\mathcal{U}_x(n/t)\big)^n$$

(3.7)
$$= (1 - t/n)^n$$

$$= e^{-t}\big(1 + \mathcal{O}(1)\big) \qquad \text{as } n \to \infty,$$

a viable alternative is provided by a plug-in estimator $\widehat{A}_x(y)$, based on suitable estimates $\hat{\mathfrak{a}}_x$, $\hat{\mathfrak{b}}_x$ and $\hat{\mathfrak{c}}_x$. Here, the shift may be overcome by location invariant estimators for these quantities. The fact that these parameters additionally vary in x with unknown smoothness degree adds another level of complexity and needs to be dealt with in a localized, adaptive manner. At this stage, it is worth mentioning that our adaptive procedure does not hinge on any particular type of quantile estimator. As a matter of fact, we only require the following property of an admissible quantile estimator $\widehat{A}_x(y)$.

DEFINITION 1. Given $x \in [0, 1]$, let $\mathcal{Y}_x = [\log n, n^{4/\mathfrak{a}_x}]$ and $s \in \{0, 1\}$. We call $\widehat{A}_x(y)$ admissible if for any fixed $v \in \mathbb{N}$ and constants $c_1^- < 1 < c_1^+$, which may be arbitrarily close to one, we have

$$P_f\bigg(c_1^- \le \sup_{y \in \mathcal{Y}_x}\bigg|\frac{\widehat{A}_x^{(s)}(y)}{A_x^{(s)}(y)}\bigg| \le c_1^+\bigg) = 1 - \mathcal{O}(n^{-v}),$$

uniformly over $f \in H_{\mathcal{N}(x)}(\beta, L)$, where $g^{(s)}(\cdot)$ denotes the sth derivative of a function $g(\cdot)$.

REMARK 2. Admissibility for $s = 1$ is only required in case of the L_q-norm loss.

Now we shall construct an admissible estimator under Assumption 3.1. Even though the class of potential estimators seems to be quite large under Assumption 3.1, verifying the conditions of Definition 1 leads to quite technical and tedious calculations. Moreover, the requirement of location invariance rules out many prominent estimators from the literature. Regarding the shape parameter $\mathfrak{a}_x$, this eliminates, for instance, Hill-type estimators as possible candidates; see Alves [18] and de Haan and Ferreira [7]. Possible alternatives are Pickand's estimator (cf. Pickand [55] and Drees [11]) or the probability weighted moment estimator by Hosking and Wallis [32]. These may, however, exhibit a poor performance in practice; see, for instance, de Haan and Peng [8] for a comparison. In [15], Falk proposed the negative Hill estimator, which, unlike to its positive counter part, is also location invariant; see also de Haan and Ferreira [7]. Transferring this approach to our setup, we construct estimators $\hat{\mathfrak{a}}_x$, $\hat{\mathfrak{b}}_x$ and $\hat{\mathfrak{c}}_x$ that are

location invariant, and also inherit the favorable variance property of Hill's estimator. Based on these estimates, we can use the plug-in estimator

$$(3.8) \qquad \widehat{A}_x(y) = -\hat{\mathfrak{c}}_x (\log y)^{\hat{\mathfrak{b}}_x} y^{-1/\hat{\mathfrak{a}}_x}.$$

To construct the estimators $\hat{\mathfrak{a}}_x$, $\hat{\mathfrak{b}}_x$ and $\hat{\mathfrak{c}}_x$ for fixed $x \in [0, 1]$, consider the neighborhoods $\mathcal{N}_k(x) = \{y : |x - y| \le h_k\}$ for $k = 0, \ldots, K - 1$. Introduce the sets $\mathcal{S}_k(x) = \{Y_i : i/n \in \mathcal{N}_k(x)\}$, and note that its cardinality $\bar{n}_k(x) := \#\mathcal{S}_k(x)$ satisfies $nh_k \le \bar{n}_k(x) \le 2nh_k + 1$. Let us rearrange the sample in $\mathcal{S}_k(x)$ as

$$(3.9) \qquad Y_{1,\bar{n}_k(x)}, Y_{2,\bar{n}_k(x)}, \ldots, Y_{\bar{n}_k(x),\bar{n}_k(x)},$$

where $Y_{j,\bar{n}_k(x)}$ denotes the jth largest $Y_i \in \mathcal{S}_k(x)$. For each $k = 0, \ldots, K - 1$, let $m_k = m(\bar{n}_k(x))$ such that $m_k/\bar{n}_k \to 0$, where $\bar{n}_k = \bar{n}_k(x)$ to lighten the notation. In the literature, a common parametrization of m_k is $m_k = \bar{n}_k^{\mathfrak{m}}$ for $0 < \mathfrak{m} \le 1$. Before discussing the important issue of possible choices of $\mathfrak{m}$, we formally introduce our estimation procedure. Apart from the necessary location invariance, an estimator of $A_x(y)$ should also adapt to the unknown smoothness degree of the parameters $\mathfrak{a}_x$, $\mathfrak{b}_x$ and $\mathfrak{c}_x$. A related issue is dealt with in the literature; see, for instance, Drees [13] or Grama and Spokoiny [24]. In order to achieve this adaptivity, we apply a Lepski-type procedure to select among appropriate base estimators. We first tackle the problem of estimating $\mathfrak{a}_x$. Using Falk's idea in [15], we define

$$(3.10) \qquad \frac{1}{\hat{\mathfrak{a}}_x(m_k)} = \frac{1}{m_k} \sum_{i=2}^{m_k-1} \log\left(\frac{Y_{m_k,\bar{n}_k} - Y_{1,\bar{n}_k}}{Y_{i,\bar{n}_k} - Y_{1,\bar{n}_k}}\right), \qquad k = 0, 1, \ldots, K - 1.$$

Note that this estimator is clearly location invariant. For $\rho > 1$ select the index $\hat{k}_{\mathfrak{a}}(x)$ via

$$(3.11) \qquad \begin{aligned} \hat{k}_{\mathfrak{a}}(x) &:= \inf\{k = 0, \ldots, K - 1 | \exists l \le k : \\ & \quad |\hat{\mathfrak{a}}_x^{-1}(m_{k+1}) - \hat{\mathfrak{a}}_x^{-1}(m_l)| > \rho^{-k}(\log n)^{-1}\} \wedge K. \end{aligned}$$

As a final estimator, we put

$$(3.12) \qquad \hat{\mathfrak{a}}_x^{-1} = \hat{\mathfrak{a}}_x^{-1}(m_{\hat{k}_{\mathfrak{a}}}) \qquad \text{where } \hat{k}_{\mathfrak{a}} = \hat{k}_{\mathfrak{a}}(x).$$

For the estimation of $\mathfrak{b}_x$, we proceed in a similar manner. For $k = 0, 1, \ldots, K - 1$, we put

$$(3.13) \qquad \hat{\mathfrak{b}}_x(m_k) = \frac{1}{m_k \log\log\bar{n}_k} \sum_{i=2}^{m_k-1} \log\left(\frac{Y_{i,\bar{n}_k} - Y_{1,\bar{n}_k}}{(\bar{n}_k/i)^{-1/\hat{\mathfrak{a}}_x(m_k)} - (\bar{n}_k/1)^{-1/\hat{\mathfrak{a}}_x(m_k)}}\right),$$

and select the index $\hat{k}_{\mathfrak{b}}(x)$ via

$$(3.14) \qquad \begin{aligned} \hat{k}_{\mathfrak{b}}(x) &:= \inf\{k = 0, \ldots, \hat{k}_{\mathfrak{a}}(x) | \exists l \le k : \\ & \quad |\hat{\mathfrak{b}}_x(m_{k+1}) - \hat{\mathfrak{b}}_x(m_l)| > \rho^{-k}(\log\log n)^{-1}\} \wedge K. \end{aligned}$$

As final estimator, we then put

$$(3.15) \qquad \hat{\mathfrak{b}}_x = \hat{\mathfrak{b}}_x(m_{\hat{k}_\mathfrak{b}}) \qquad \text{where } \hat{k}_\mathfrak{b} = \hat{k}_\mathfrak{b}(x).$$

Interestingly, it turns out that $\hat{\mathfrak{b}}_x = \mathfrak{b}_x + \frac{\log \mathfrak{c}_x}{\log \log n h_{\hat{k}_\mathfrak{b}}}(1 + \mathcal{O}_P(1))$. Since this implies that

$$(\log n h_k)^{\hat{\mathfrak{b}}_x} = \mathfrak{c}_x (\log n h_k)^{\mathfrak{b}_x}(1 + \mathcal{O}_P(1)) \qquad \text{for } k = 0, \dots, K - 1,$$

there is no need to specifically estimate $\mathfrak{c}_x$, it is included in the bias for free. We are thus lead to the definition of our estimator

$$(3.16) \qquad \widehat{A}_x(y) = -(\log y)^{\hat{\mathfrak{b}}_x} y^{-1/\hat{\mathfrak{a}}_x}.$$

For the consistent estimation of $\widehat{A}_x(\cdot)$, we need a relation between the initial bandwidth h_0 and the bias, induced by the parameter β_x. Note that such an assumption is inevitable, since any adaptive estimation procedure needs to start off with some initial bandwidth. Thus in the sequel, we will assume that

$$(3.17) \qquad h_0^{\beta_x} |A_x(m_0)|^{-1} = \mathcal{O}((\log n)^{-1}).$$

If $\mathfrak{h}_0, \mathfrak{m} > 0$ is such that

$$(3.18) \qquad \mathfrak{m}\mathfrak{h}_0 < (1 - \mathfrak{h}_0)\mathfrak{a}_0\beta_0,$$

for some lower bounds

$$(3.19) \qquad \mathfrak{a}_0 \leq \mathfrak{a}_x \quad \text{and} \quad \beta_0 \leq \beta_x$$

on the unknown parameters, then (3.17) is valid. In the supplementary material [35] we prove the following result under the more general Assumption 10.1, which is implied by Assumption 3.1.

PROPOSITION 3.2. *Grant Assumption* 10.1, *and suppose that* (3.17) *is valid. Then* $\widehat{A}_x(y)$ *defined in* (3.16) *is admissible.*

In practice the negative Hill estimator works well for $\mathfrak{a}_x \in (0, 3/2]$ (and small $\mathfrak{b}_x$), but has increasing (asymptotically negligible) bias for $\mathfrak{a}_x > 3/2$ and $\mathfrak{b}_x \neq 0$, which should be corrected in applications; see also Section 5, paragraph (B). Also note that our assumptions in Assumption 3.1 include cases where a CLT for an estimator $\hat{\mathfrak{a}}_x$ fails to hold, and only slower rates of convergence than $m_k^{-1/2}$ are possible. This is particularly the case if $\mathfrak{b}_x \neq 0$; we refer to de Haan and Ferreira [7] for details. In practice, the choice of the actual bandwidth m_k (and hence $\mathfrak{m}$) is of significant relevance, and much research has been devoted to this subject; see, for instance, Drees [12] and Drees et al. [14]. In [18], Alves addresses this question for a related (positive) location invariant Hill-type estimator both in theory (Theorem 2.2) and practice (concluding remarks and algorithm). Transferring the practical aspects, this amounts to the choice $m_k = 2\bar{n}_k^{\mathfrak{m}}$, $\mathfrak{m} = 2/3$ in our case. Still, any other choice also leads to the total optimal rates presented in Theorems 3.2 and 3.3, as long as $0 < \mathfrak{m} < 1$ holds.

3.3. *Pointwise adaptation.* Throughout this subsection we fix a point $x \in [0, 1]$. For the seminorm in Proposition 3.1 we take $\|f\| := |f(x)|$. According to Theorem 3.1, we set

(3.20)
$$B_k := c(\beta^*, L) h_k^\beta,$$
$$R_k := c(\beta^*) \max\{|Z_j(h_k, x)| : j = 1, \ldots, 2J(\beta^*)\},$$

in the notation of (3.4). The nonnegativity and monotonicity constraints on B_k and R_k are satisfied since h_k increases. We define the oracle and estimated critical values as

$$\mathfrak{z}_k(x) = 4c(\beta^*) \left| A_x \left(\frac{\mathfrak{a}_x n h_k}{4J(\beta^*) \log n} \right) \right|, \qquad \hat{\mathfrak{z}}_k(x) = 4c(\beta^*) \left| \widehat{A}_x \left(\frac{\hat{\mathfrak{a}}_x n h_k}{4J(\beta^*) \log n} \right) \right|,$$

for $k = 0, \ldots, K - 1$ and set $\mathfrak{z}_K(x) = \hat{\mathfrak{z}}_K(x) := 0$. To lighten the notation, we often drop the index x and write $\mathfrak{z}_k$ and $\hat{\mathfrak{z}}_k$. As outlined earlier in (3.7), this definition is motivated by the fact that $\mathcal{U}_x(y) \approx A_x(y)$ as $y \to \infty$. The critical values can thus be viewed as an appropriate estimate for certain extremal quantiles. The additional $\log n$-factor turns out to be the price to pay for adaption. We proceed by introducing the estimated truncated critical values as

(3.21)
$$\hat{\mathfrak{z}}_k^T = \min\{\hat{\mathfrak{z}}_k, 1\}.$$

The truncation of the estimator $\hat{\mathfrak{z}}_k^T$ is required to exclude a possible pathological behavior both in theory and practice. Note that this does not affect its proximity to $\mathfrak{z}_k$ if $\hat{\mathfrak{z}}_k$ is consistent, since $\mathfrak{z}_k \to 0$ uniformly in $k = 0, \ldots, K - 1$. We have the following pointwise result.

THEOREM 3.2. *Fix $x \in [0, 1]$, and suppose $\mathfrak{a}_x, \mathfrak{b}_x, \mathfrak{c}_x$ and $\beta_x \in (0, \beta^* + 1]$ are unknown with $\mathfrak{h}_0 < \beta_x \mathfrak{a}_x / (\beta_x \mathfrak{a}_x + 1)$. If Assumption 3.1 holds, then*

$$\sup_{f \in H_{\mathcal{N}(x)}(\beta, L)} \mathbb{E}_f \big[(\hat{f}(x) - f(x))^2 \big]$$
$$= \mathcal{O} \big((n / \log n)^{(-2\beta_x)/(\mathfrak{a}_x \beta_x + 1)} (\log n)^{(2\mathfrak{a}_x \mathfrak{b}_x \beta_x)/(\mathfrak{a}_x \beta_x + 1)} \big).$$

As will be demonstrated in Section 4, this result is optimal in the minimax sense.

3.4. *L_q-adaptation.* Let us consider the $L_q([0, 1])$-norm as seminorm in Proposition 3.1. Due to (3.3) we can choose

(3.22)
$$B_k := c(\beta^*, L) h_k^\beta,$$

(3.23)
$$R_k := c(\beta^*) \big\| \max\{|Z_j(h_k, \cdot)| : j = 1, \ldots, 2J(\beta^*)\} \big\|_q$$

in the notation of (3.4). We verify that the nonnegativity and monotonicity constraints on B_k and R_k are satisfied for $\rho > 1$ in (2.3) since for any $x \in [0, 1]$

each interval $x + h_k \mathcal{I}_j$, $j = 1, \ldots, 2J(\beta^*)$ is included in $x + h_{k+1}\mathcal{I}_{j'}$ for some $j' = 1, \ldots, 2J(\beta^*)$ for any k. Throughout this paragraph, we assume that the parameters $\mathfrak{a}_x, \mathfrak{b}_x, \mathfrak{c}_x$ remain constant for $x \in [0, 1]$. We denote these with $\mathfrak{a}_F, \mathfrak{b}_F, \mathfrak{c}_F$, and the corresponding $A_x(\cdot)$ with $A_F(\cdot)$.

The construction of the critical values is more intricate compared to the pointwise case, and relies on the following quantity. Introduce

$$(3.24) \quad \widehat{IU}_n(s, q) = \left(\int_{n^{-2/\hat{\mathfrak{a}}_F}}^{n^{1/2}} ((-\widehat{A}_F)^q (s/y))^{(1)} \exp(-y) \, dy \right)^{1/q}, \qquad q \geq 1,$$

and the corresponding version $IU_n(s, q)$ where we replace $\widehat{A}_F(\cdot)$ by $A_F(\cdot)$ and $\hat{\mathfrak{a}}_F$ by $\mathfrak{a}_F$ [recall that $g^{(s)}(\cdot)$ denotes the sth derivative of a function $g(\cdot)$]. For $k = 0, \ldots, K - 1$, we introduce the critical values as

$$\mathfrak{z}_k = \sqrt{5}c(\beta^*) \left| IU_n\left(\frac{nh_k}{6J(\beta^*)}, q \right) \right|, \qquad \hat{\mathfrak{z}}_k = \sqrt{5}c(\beta^*) \left| \widehat{IU}_n\left(\frac{nh_k}{6J(\beta^*)}, q \right) \right|,$$

and set $\mathfrak{z}_K = \hat{\mathfrak{z}}_K := 0$. Moreover, we define the corresponding truncated values as

$$(3.25) \qquad\qquad\qquad \hat{\mathfrak{z}}_k^T = \min\{\hat{\mathfrak{z}}_k, 1\}.$$

Unlike the pointwise case, the critical values do not correspond to an extremal quantile, but they can be considered as an estimate of $\mathbb{E}[R_k^q]^{1/q}$. This already indicates that the L_q-case is substantially different from the pointwise situation, and indeed additional, more refined arguments are necessary to prove the result given below.

THEOREM 3.3. *Suppose $\mathfrak{a}_F > 0$ and $\beta \in (0, \beta^* + 1]$ are unknown with $\beta\mathfrak{a}_F/(\beta\mathfrak{a}_F + 1) > \mathfrak{h}_0$. We select $\rho \in \mathbb{N}$ with $\rho > 1$. If $q \geq 1$, then the adaptive estimator $\hat{f}$ from Section 2 satisfies*

$$\sup_{f \in H_{[0,1]}(\beta, L)} \mathbb{E}_f\big[\|\hat{f} - f\|_q^q\big] = \mathcal{O}\big(n^{(-q\beta)/(\mathfrak{a}_F\beta + 1)}(\log n)^{(q\beta\mathfrak{a}_F\mathfrak{b}_F)/(\mathfrak{a}_F\beta + 1)}\big).$$

REMARK 3. If one allows for $\mathfrak{a}_x, \mathfrak{b}_x, \mathfrak{c}_x \in H(\beta_0, L)$ for $x \in [0, 1]$, the above result remains valid if one takes the supremum over the above bound. This result is also optimal in the minimax sense.

Theorem 3.3 shows that the estimator $\hat{f}$ is L_q-adaptive; that is, it attains the minimax rates, which are optimal in the oracle setting of known $\mathfrak{a}_F$ and β, although it does not use these constants in its construction; see Theorem 4.2 below for the lower bound.

4. Asymptotic lower bounds. We show that the logarithmic loss in the convergence rate in Theorem 3.2 is unavoidable with respect to any estimator sequence of f. First, we treat the case $\mathfrak{a}_x \in (0, 2]$ for which we derive a lower bound, even for a known error distribution. It suffices to treat the case where $\mathfrak{a} = \mathfrak{a}_x$, $\mathfrak{b} = \mathfrak{b}_x$ and $\mathfrak{c} = \mathfrak{c}_x$ remain constant for $x \in [0, 1]$. We maintain this convention throughout this section.

We assume that the ε_j have a Lebesgue density f_ε which is continuous and strictly positive on $(-\infty, 0)$, and vanishes on $[0, \infty]$. Moreover, we impose that the χ^2-distance for the parametric location problem satisfies

$$(4.1) \quad \int_{-\infty}^0 |f_\varepsilon(x + \vartheta) - f_\varepsilon(x)|^2 / f_\varepsilon(x)\, dx \le c_\varepsilon \vartheta^{\mathfrak{a}} |\log \vartheta|^{-\mathfrak{a}\mathfrak{b}} \qquad \forall \vartheta \in (0, 1),$$

for some $\mathfrak{a} \in (0, 2]$ and $\mathfrak{b} \in \mathbb{R}$. Note that $\mathfrak{a}$ and $\mathfrak{b}$ correspond to $\mathfrak{a}_x$ and $\mathfrak{b}_x$, respectively, in (1.3) and (1.4) with uniform x. As examples for such error densities with $\mathfrak{b} = 0$, we consider the reflected gamma-densities

$$f_\lambda(x) := \frac{1}{\Gamma(\lambda)} (-x)^{\lambda - 1} \exp(x) \mathbf{1}_{(-\infty, 0)}(x), \qquad x \in \mathbb{R},$$

for $\lambda \in [1, 2)$. Thus, by $|(1 + \vartheta/x)^{\lambda - 1} - 1| \le \vartheta/|x|$ for $x \le -\vartheta$ we have

$$\int_{-\infty}^0 |f_\lambda(x + \vartheta) - f_\lambda(x)|^2 / f_\lambda(x)\, dx$$

$$= \int_{-\infty}^{-\vartheta} \left| \left(1 + \frac{\vartheta}{x}\right)^{\lambda - 1} \exp(\vartheta) - 1 \right|^2 f_\lambda(x)\, dx + \int_{-\vartheta}^0 f_\lambda(x)\, dx$$

$$\le 2(\exp(\vartheta) - 1)^2 + 2\exp(2\vartheta)\vartheta^2$$

$$\quad + 2\exp(2\vartheta)\vartheta^2 \int_{-1}^{-\vartheta} x^{-2} f_\lambda(x)\, dx + \int_{-\vartheta}^0 f_\lambda(x)\, dx$$

$$\le \mathcal{O}(\vartheta^2) + \frac{2}{(2 - \lambda)\Gamma(\lambda)} \exp(2\vartheta)\vartheta^\lambda (1 - \vartheta^{2-\lambda}) + \vartheta^\lambda / \Gamma(\lambda + 1)$$

$$= \mathcal{O}(\vartheta^\lambda).$$

Therefore, the reflected gamma-density satisfies (4.1) when putting $\mathfrak{a} = \lambda$. Note that (4.1) implies (1.3), (1.4) under the Assumption 3.1(i). The following theorem together with the upper bound in Theorem 3.2 shows that pointwise adaptation causes a logarithmic loss in the convergence rates, which is known from regular regression when inserting $\mathfrak{a} = 2$.

THEOREM 4.1. *Assume condition* (4.1), *and fix some arbitrary* $x_0 \in [0, 1]$, $\beta_1 > \beta_2 > 0$ *and* $C_0, C_1 > 0$. *Let* $\{\hat{f}_n(x_0)\}_n$ *be any sequence of estimators of* $f(x_0)$ *based on the data* $Y_1, \ldots, Y_n$ *which satisfies*

$$\sup_{f \in \mathcal{H}_{[0,1]}(\beta_1, C_0)} \mathbb{E}_f |\hat{f}_n(x_0) - f(x_0)|^2 = \mathcal{O}(n^{-2\beta_2/(1+\beta_2\mathfrak{a})} n^{-\xi}),$$

for some $\xi > 0$. Then this estimator sequence suffers from the lower bound

$$\liminf_{n \to \infty} (n / \log n)^{(2\beta_2)/(1+\mathfrak{a}\beta_2)} (\log n)^{(-2\mathfrak{a}\mathfrak{b}\beta_2)/(1+\mathfrak{a}\beta_2)}$$

$$\times \sup_{f \in \mathcal{H}_{[0,1]}(\beta_2, C_1)} \mathbb{E}_f |\hat{f}_n(x_0) - f(x_0)|^2 > 0.$$

For completeness we also derive the L_q-minimax optimality of the convergence rates established by our estimator $\hat{f}$ in Theorem 3.3. This rectifies a conjecture after Theorem 3 in Hall and van Keilegom [29] for general smoothness degrees.

THEOREM 4.2.   *Assume condition* (4.1), *and let* $\{\hat{f}_n\}_n$ *be any sequence of estimators of f based on the data $Y_1, \ldots, Y_n$. Then, for any fixed $q \geq 1$, we have*

$$\liminf_{n \to \infty} n^{\beta_2/(1+\mathfrak{a}\beta_2)} (\log n)^{(-\mathfrak{a}\mathfrak{b}\beta_2)/(1+\mathfrak{a}\beta_2)} \sup_{f \in \mathcal{H}_{[0,1]}(\beta_2, C_1)} \mathbb{E}_f [\| \hat{f}_n - f \|_q] > 0.$$

Now we focus on the case $\mathfrak{a} > 2$. To simplify some of the technical arguments in the proofs, we restrict to the case $\mathfrak{b} = 0$. If $\mathfrak{a} > 2$, the convergence rates become slower than in the Gaussian case. Instead of the convenient conditions (1.3) and (1.4), we choose the slightly different Definition 2, under which the upper bound proofs obviously still hold true.

DEFINITION 2.   Let $\mathfrak{a} > 2$, $0 < \mathfrak{h}_0 \leq 1$, and denote with $\mathcal{D}_n(\mathfrak{a}, \mathfrak{h}_0)$ the set of all error distribution functions whose quantile functions $\mathcal{U}^{(n)}$ satisfy:

(i) $\displaystyle \sup_{y \in (0,\infty]} \left| \frac{\mathcal{U}^{(n)}(y)}{A(y/2)} \right| \leq 1$ where $A(y) = -y^{-1/\mathfrak{a}}$,

(ii) $\displaystyle \sup_n \sup_{y \in [\log N, N]} \left| \frac{\mathcal{U}^{(n)}(y)}{A(y)} - 1 \right| |\log y| \leq (\log n)^{-2}$, $N = n^{\mathfrak{h}_0}$.

Note that we have $\mathcal{D}_n(\mathfrak{a}, \mathfrak{h}_0) \subseteq \mathcal{D}_n(\mathfrak{a}, \mathfrak{h}_0')$ if $\mathfrak{h}_0 > \mathfrak{h}_0'$.

The above conditions particularly imply that the distribution function $F(y) = F^{(n)}(y)$ [or likewise $\mathcal{U}(y) = \mathcal{U}^{(n)}(y)$] of the errors ε_j may depend on n. While the lower bound results for $\mathfrak{a} \leq 2$ still hold true if the error distribution is known and independent of the design point, here two competing types of regression errors have to be considered in the proof. Note that the probability measure thus depends on both the regression function f and the distribution function F, which we mark as $P_{f,F}$.

THEOREM 4.3.   *Fix some arbitrary $x_0 \in [0, 1]$, $\beta_1 > \beta_2 > 0$ and $C_0, C_1 > 0$. Let $\mathfrak{a} > 2$, and suppose that $\mathfrak{h}_0 < \frac{\beta_2}{\mathfrak{a}\beta_2+1}$. Let $\{\hat{f}_n(x_0)\}_n$ be any sequence of estimators of $f(x_0)$ based on the data $Y_1, \ldots, Y_n$ which satisfies*

$$\sup_{f \in \mathcal{H}_{[0,1]}(\beta_1, C_0)} \sup_{F \in \mathcal{D}_n(\mathfrak{a}, \mathfrak{h}_0)} \mathbb{E}_{f,F} |\hat{f}_n(x_0) - f(x_0)|^2 = \mathcal{O}(n^{-2\beta_2/(1+\mathfrak{a}\beta_2)} n^{-\xi}),$$

for some $\xi > 0$. Then this estimator sequence suffers from the lower bound

$$\liminf_{n \to \infty}(n/\log n)^{(2\beta_2)/(1+\mathfrak{a}\beta_2)} \sup_{f \in \mathcal{H}_{[0,1]}(\beta_2, C_1)} \sup_{F \in \mathcal{D}_n(\mathfrak{a}, \mathfrak{h}_0)} \mathbb{E}_{f,F}|\hat{f}_n(x_0) - f(x_0)|^2 > 0.$$

The proofs of Theorems 4.1, 4.2 and 4.3 are given in the supplementary material [35]. Theorem 4.2 can be extended in a similar way to $\mathfrak{a} > 2$, a detailed proof is omitted.

5. Numerical simulations and real data application. The aim of this section is to highlight some of the theoretical findings with numerical examples. We will briefly touch on the following points:

(A) Performance of the estimator on different function types and the corresponding effect on adaptive bandwidth selection.

(B) The effect of different parameters $\mathfrak{a}_x$, $\mathfrak{b}_x$ and $\mathfrak{c}_x$.

(C) Application: wolf sunspot-number.

(D) Application: yearly best men's outdoor 1500 times.

In order to illustrate the behavior of the estimation procedure, we consider three different regression functions, displayed in black in Figures 2 and 3(a),

$$f_1(x) = -2 \cdot \mathbf{1}(x < 1/3) - 3 \cdot \mathbf{1}(1/3 \le x < 2/3) - \mathbf{1}(2/3 < x), \qquad x \in [0, 1],$$

$$f_2(x) = -2 + 2\cos(2\pi x) + 0.3\sin(19\pi x), \qquad x \in [0, 1].$$

They are similar to those discussed in Chichignoud [5]. Comments on the implementation and setup are given in the supplementary material [35], together with a
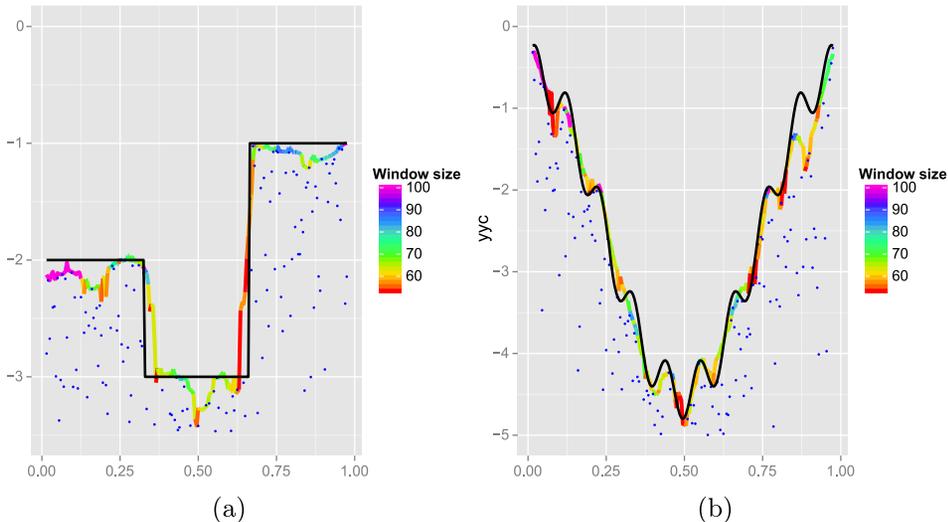

FIG. 2. (a) *Function* $f = f_1$, $\beta^* = 2$, $n = 200$, $\varepsilon_j \sim \exp(1)$; (b) *function* $f = f_2$, $\beta^* = 2$, $n = 200$, $\varepsilon_j \sim \exp(1)$.
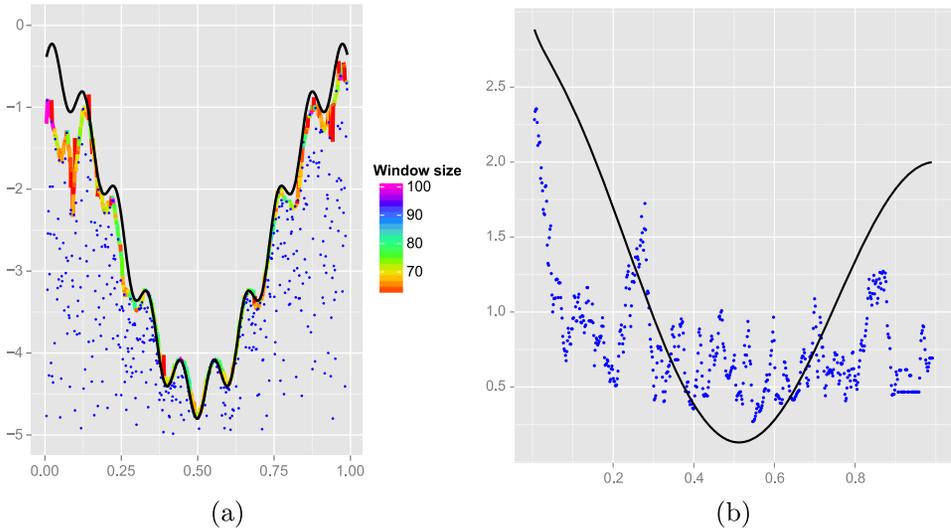
FIG. 3. (a) *Function* $f = f_2$, $\beta^* = 4$, $n = 600$, $\varepsilon_j \sim \Gamma(\mathfrak{a}_x, 1)$; (b) $\mathfrak{a}_x$ (*black line*) *and* $\hat{\mathfrak{a}}_x$ (*blue points*).

numerical comparison to oracle estimators and additional simulations. All of the results can be reproduced by R-code, available at [36].

(A) Figure 2 gives a first impression on the behavior and accuracy of our estimation procedure. In both cases, the errors ε_j follow an exponential distribution $\exp(1)$, and the sample size is $n = 200$. The window size in Figure 2 corresponds to the local sample size, chosen by the adaptive procedure. Even though n is only of moderate size, the estimation procedure achieves good results by essentially recovering the shape of the underlying regression, also in the wiggly case of function f_2. Simulations of other nonparametric (adaptive) estimators that do not take the nonregularity into account (cf. Lepski and Spokoiny [47] and the R-packages crs, gam, smooth-spline, etc.) often fail to do so (with mean correction).

The effect of the shape (type) of the function on the bandwidth selection is highlighted by a color-scheme, ranging from dark red (low) to dark violet (high). In order to understand the "coloring of the estimator," one has to recall that the estimation procedure always tries to fit a local polynomial which "stays above the observations." At first sight, this can lead to a surprisingly large bandwidth selection at particular spots. The bandwidth size is not necessarily an indicator for estimation accuracy. The reason for this effect is the maximum function: additional observations are taken into account as long as this does not substantially change the maximum, which can lead to a surprisingly large bandwidth selection.

(B) Here, the setup is different from paragraph (A). We consider a sample size of $n = 600$, and we let the parameters vary in x. The impact of $\mathfrak{b}_x$, $\mathfrak{c}_x$ (and their estimates) is rather insignificant on the total estimator. This is not unexpected

and can be explained by the very definition in (3.1). We therefore focus only on the parameter $\mathfrak{a}_x$ in this paragraph. We consider the setup where the errors follow a Gamma distribution $\varepsilon_j \sim \Gamma(\mathfrak{a}_x, 1)$ and $\mathfrak{a}_x$ varies according to the function $\mathfrak{a}_x = \sin(2\pi x + \pi/2) - \sqrt{(1 - (x - 1)^2)} + 2$. We only discuss function f_2 here; a more comprehensive comparison including an additional function f_3 is given in the supplementary material [35]. As can be clearly seen in Figure 3, there is a considerable increase in estimation accuracy as $\mathfrak{a}_x$ gets closer to zero. Generally speaking, for larger $\mathfrak{a}_x$ the bias can be pronounced, and this is indeed the case at the left top of Figure 3(a). It simply turns out that there are no observations at all near the regression function f_2, which leads to the large gap. An approximate bias correction [e.g., by $\widehat{IU}(nh_k, 1)$, Section 3.4] could be applied, but we do not pursue this any further. Figure 3 also reveals that the estimator $\hat{\mathfrak{a}}_x$ (blue) has a large variance and problems with quickly oscillating regression functions f (compare with the supplementary material [35]). On the other hand, it seems to capture the general trend of decrease and increase to some extent. We would like to point out, however, that these estimations are very sample dependent, and due to the relatively small, local sample size, the actual behavior of local samples may deviate significantly from a large sample of $\Gamma(\mathfrak{a}_x, 1)$-distributed random variables. Significant overestimation leads to critical values that are too large, which in turn results in a slight overestimation of the regression function; see the very center of Figure 3(a) (0.4 to 0.6). The opposite effect can be observed at both endpoints of Figure 3(a), where an underestimation is present, which leads to critical values that are too small. Also note that the negative Hill estimator generally tends to underestimate $\mathfrak{a}_x$ if $\mathfrak{a}_x \geq 3/2$, which is due to an (asymptotic negligible) bias; cf. de Haan and Ferreira [7]. A thorough bias correction requires a precise second order asymptotic expansion of the limit distribution of the negative Hill estimator, which is beyond the scope of this paper. Note, however, that a rudimentary bias correction is available in our implemented code. Another, more practical option would be to consider the estimation of $\mathfrak{a}_x$ itself as a regression problem with one-sided errors, treating the local estimates $\hat{\mathfrak{a}}_x$ as "sample."

A similar behavior appears when considering function f_1, but, as can be expected, the estimates $\hat{\mathfrak{a}}_x$ are more accurate.

(C) The Wolf sunspot number (often also referred to as *Zürich number*), is a measure for the number of sunspots and groups of sunspots present on the surface of the sun. Initiated by Rudolf Wolf in 1848 in Zürich, this famous time series has been studied for decades by physicists, astronomers and statisticians. The relative sunspot number R_t is computed via the formula

$$(5.1) \qquad\qquad R_t = K_t(10g_t + s_t),$$

where s_t is the number of individual spots observed at time t, g_t is the number of groups observed at time t and K_t is the *observatory factor* or *personal reduction coefficient*. The factor K_t (always positive and usually smaller than one) depends

on the individual observatories around the world and is intended to convert the data to Wolf's original scale, but also to correct for seeing conditions and other diversions. In general, we have the relationship

(5.2)   observed data = observed fraction × true value,

where we always have that the random variable observed fraction $\in (0, 1]$. Therefore, the factor K_t can be viewed as an aggregated individual estimate for the right scaling. Over the last century, many different models have been fit to the sunspot data; we refer to He [31] and Solanki et al. [56] for an overview. In particular, the study of the sunspots has attracted people long before 1848. Recorded observations are, for instance, due to Thomas Harriot, Johannes and David Fabricius (in the 17th century), Edward Maunder and many more. However, much uncertainty lies in these data, and the sunspot time series before 1850 is usually referred to as "unreliable" or "poor." It is therefore interesting to reconstruct the "true time series" or at least reduce some uncertainty. We attempt do so for the period from 1749 to 1810, based on monthly observations. Let us reconsider model (5.1). Given R_t, we may then postulate the model

(5.3) $$R_t = X_t \mathcal{S}(10g_t^\circ + s_t^\circ),$$

where g_t°, s_t° denote the corresponding true sunspot values, and $X_t \in (0, 1]$. This means we concentrate all random components in X_t, which is in spirit of model (5.2). We point out that this is only one possible way from a modeling perspective; we refer to Kneip et al. [39] or Koenker et al. [42] and the references therein for alternatives and more general models. In our setup, the parameter $\mathcal{S} > 0$ reflects the support of the "misjudgment" of the observer. For example, $\mathcal{S} \leq 1$ is equivalent with the assumption that every observer always reports less than the true value. As we see below, it incorporates the systematic bias of the observers. By using a log-transformation, we have the additive model

(5.4) $$\log R_t = \log X_t + \log(10g_t^\circ + s_t^\circ) + \log \mathcal{S},$$

which can be interpreted as a nonparametric regression problem with stochastic error $\log X_t \in (-\infty, 0]$. The goal is to estimate the function $f(t) = \log(10g_t^\circ + s_t^\circ)$, the "true" relative sunspot number. Such estimation results can serve as input to structural physical models for sunspot activity like the time series approaches mentioned above. Unfortunately, one can only estimate $f(t) + \log \mathcal{S}$, where the bias $\log \mathcal{S}$ cannot be removed without any further assumptions. This is clear from the nonidentifiability in model (5.3). Generally $\mathcal{S}$ is a systematic (intrinsic) bias, which has to be overcome using other sources of information (expert judgement). Any other statistical approach will also suffer from such a global bias.

The results of the estimated sunspot number is given in Figure 4, where we plotted an estimate corresponding to $\mathcal{S} = 1$. Given that observation techniques where much less advanced and coordinated in the 18th and 19th centuries, it is reason-
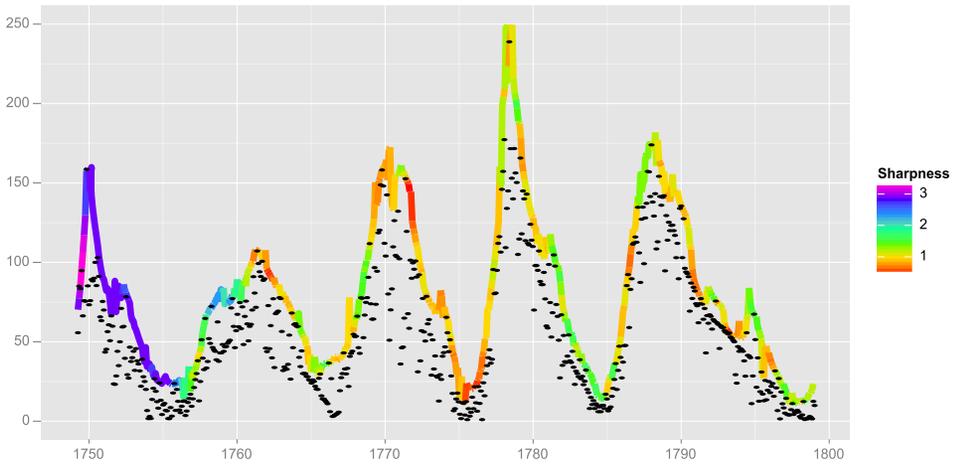
FIG. 4. *Estimated Wolf number with $\mathcal{S} = 1$, $\beta^* = 5$.*

able to assume $\mathcal{S} \leq 1$. Apart from the estimated sunspot number itself, our estimation procedure provides a map from the uncertainty level $\mathcal{S}$ to the true sunspot number $f(t)$. The sharpness $\mathfrak{a}_x$ seems to mainly vary within the interval $[0, 3.5]$. Finally, we would like to comment on the "peaks" around 1768 and 1774. These peaks are artifacts and originate from a too large initial bandwith selection at these particular points. However, for the sake of reproducibility, we have kept them and did not make any ad-hoc, data-dependent changes.

(D) As another example we discuss the yearly best men's outdoor 1500 m times starting from 1966, depicted in Figure 5 with estimated lower boundary. Following
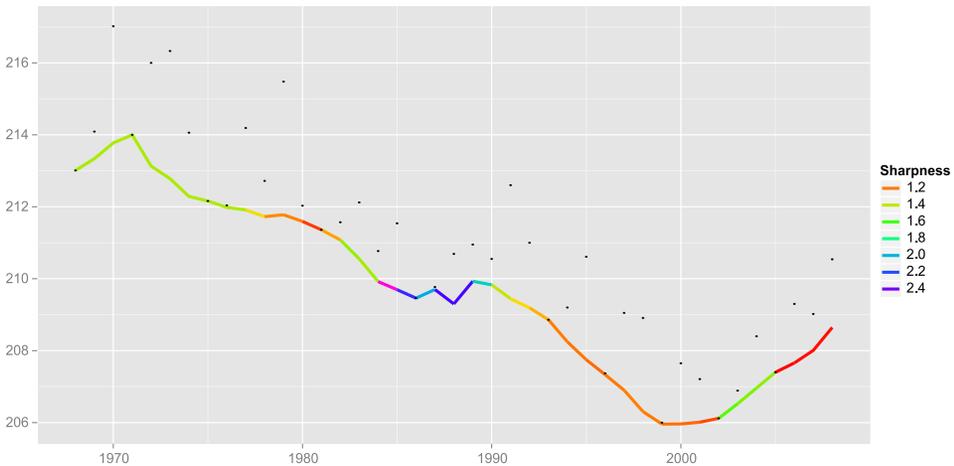

FIG. 5. *Yearly best men's outdoor 1500 m times in seconds with estimated boundary ($\beta^* = 2$).*

Knight [41], the boundary can be interpreted as the best possible time for a given year. This data set displays an interesting behavior. As can be clearly seen from Figure 5, the boundary steadily decreases from 1970 until around the year 2000, followed by a sudden and sharp increase. This event leaves room for speculation. Let us mention that until the year 2000, it had been very difficult to distinguish between the biological and synthetical EPO. The breakthrough was achieved by Lasne and Ceaurriz [46], and since then, more and more refined and efficient doping tests have been developed. It seems plausible that this change and advance in doping controls has lead to the sudden increase, but it might as well be attributed to some other reason.

6. Proof of the main results. Throughout the proofs, we make the following convention. For two sequences of positive numbers a_n and b_n, we write $a_n \gtrsim b_n$ when $a_n \geq C b_n$ for some absolute constant $C > 0$, and $a_n \lesssim b_n$ when $b_n \gtrsim a_n$. Finally, we write $a_n \sim b_n$ when both $a_n \lesssim b_n$ and $a_n \gtrsim b_n$ hold.

PROOF OF THEOREM 3.1.   Throughout the proof, we fix some arbitrary $x \in [0, 1]$ and write $\beta = \beta_x$ to lighten the notation. The data Y_i, $i = 1, \ldots, n$, can be written as

$$Y_i = \sum_{j=0}^{\beta^*} b_j (x_i - x)^j + \varepsilon_i + \Delta_i,$$

where $\Delta_i := f(x_i) - \sum_{j=0}^{\beta^*} b_j (x_i - x)^j$. Putting $\Delta := \max\{|\Delta_i| : |x_i - x| \leq h_k\}$, the coefficients b_j are chosen as the Taylor coefficients $b_j = f^{(j)}(x)/j!$ for $j \leq \langle \beta \rangle$ and $b_j = 0$ otherwise, such that by the Hölder condition on $f^{(\langle \beta \rangle)}$ in the Taylor remainder term

(6.1) $$\Delta \leq L h_k^\beta / (\langle \beta \rangle + 1)!.$$

Selecting $b_0^* := b_0 + \Delta$, $b_j^* := b_j$, $j > 0$, we realize that

$$\sum_{j=0}^{\beta^*} b_j^* (x_i - x)^j = \sum_{j=0}^{\beta^*} b_j (x_i - x)^j + \Delta \geq Y_i$$

$$\forall i = 1, \ldots, n \text{ with } |x - x_i| \leq h_k,$$

so that by the definition of the $\hat{b}_j$, $j = 0, \ldots, \beta^*$, we have

(6.2)
$$\sum_{|x_i - x| \leq h_k} \sum_{j=0}^{\beta^*} \hat{b}_j (x_i - x)^j \leq \sum_{|x_i - x| \leq h_k} \sum_{j=0}^{\beta^*} b_j^* (x_i - x)^j$$

$$= \sum_{|x_i - x| \leq h_k} \left\{ \sum_{j=0}^{\beta^*} b_j (x_i - x)^j + \Delta \right\}.$$

We define the polynomial

$$Q(y) := \sum_{j=0}^{\beta^*} (\hat{b}_j - b_j)(y - x)^j - \Delta.$$

Then inequality (6.2) implies that

(6.3)
$$\inf_{n,k,f} \int Q(y) \, d\lambda_n(y) \leq 0,$$

where λ_n denotes the uniform probability measure on the discrete set $\{x_i : |x_i - x| \leq h_k\}$ inside the interval $[x - h_k, x + h_k] \cap [0, 1]$. We introduce the sets $Q^{\pm}$ of all $y \in [x - h_k, x + h_k] \cap [0, 1]$ such that $Q(y)$ is nonnegative or negative, respectively. Our first task is to show that

(6.4)
$$\inf_{n,k,f} \lambda_n(Q^-) > 0 \quad \text{or} \quad Q = 0 \qquad \text{identically.}$$

In the latter case Theorem 3.1 is trivially true; hence we focus on the case where $Q \neq 0$. As Q^- is the complement of Q^+ with respect to $[x - h_k, x + h_k] \cap [0, 1]$, we have $\lambda_n(Q^+) \geq 1/2$ or $\lambda_n(Q^-) \geq 1/2$. Clearly, we have (6.4) in the second case, so let us study the situation where $\lambda_n(Q^+) \geq 1/2$.

As Q is a polynomial with degree $\leq \beta^*$ the set Q^+ equals the union of at most $\beta^* + 1$ disjoint sub-intervals of $[x - h_k, x + h_k] \cap [0, 1]$. The number of all design points in $[x - h_k, x + h_k]$ is denoted by m_k. Hence, there exists at least one interval $I_0^+ \subseteq Q^+$ such that $\lambda_n(I_0^+) \geq 1/(2\beta^* + 2)$. At least $\lceil m_k/(2\beta^* + 2) \rceil$ of the x_i lie in I_0^+ so that, due to the equidistant location of the design points, the length of I_0^+ is larger or equal to

$$\{\lceil m_k/(2\beta^* + 2) \rceil - 1\}/n \geq \{\lceil \lfloor nh_k \rfloor/(2\beta^* + 2) \rceil - 1\}/n \geq c_1(\beta^*) \cdot h_k,$$

for n sufficiently large and some uniform constant $c_1(\beta^*) > 0$ which does not depend on n or k, but only on β^*. The polynomial Q takes only nonnegative values on the interval I_0^+. By Lemma 6.1 below there exists some interval $I_1^+ \subseteq I_0^+$ with the length $c_2(\beta^*)h_k$ such that

$$\inf_{y \in I_1^+} |Q(y)| \geq c_3(\beta^*) \cdot \sup_{|z-x| \leq h_k} |Q(z)|,$$

where the constants $c_2(\beta^*), c_3(\beta^*) > 0$ only depend on β^*. It follows from there that

$$\int_{Q^+} Q(y) \, d\lambda_n(y) \geq \int_{I_1^+} Q(y) \, d\lambda_n(y)$$

$$\geq \lambda_n(I_1^+) \cdot \inf_{y \in I_1^+} |Q(y)|$$

$$\geq \lambda_n(I_1^+) c_3(\beta^*) \cdot \sup_{|z-x| \leq h_k} |Q(z)|.$$

On the other hand we learn from (6.3) that

$$\int_{Q^+} Q(y)\,d\lambda_n(y) \le \int_{Q^-} |Q(y)|\,d\lambda_n(y) \le \lambda_n(Q^-) \cdot \sup_{|z-x|\le h_k} |Q(z)|,$$

so that

$$\lambda_n(Q^-) \ge \lambda_n(I_1^+)c_3(\beta^*) \ge c_3(\beta^*) \cdot (c_2(\beta^*)h_k n - 1)/m_k$$

$$\ge c_3(\beta^*) \cdot (c_2(\beta^*) - n^{-\mathfrak{h}_0})/(2 + n^{-\mathfrak{h}_0}),$$

unless $Q = 0$ identically. Thus (6.4) has been shown.

Using the arguments as above, we can now find some interval $I_0^- \subseteq Q^-$ whose length is bounded from below by a constant (only depending on β^*) times h_k. By Lemma 6.1 there exists an interval $I_1^- \subseteq I_0^-$, whose length is also bounded from below by a constant (only depending on β^*) times h_k and on which $|Q|$ is bounded from below by a uniform multiple of

$$\sup_{|z-x|\le h_k} |Q(z)| \ge |Q(x)| \ge |\hat{b}_0 - b_0| - \Delta.$$

This implies that

$$(6.5) \qquad \inf_{y\in I_1^-} (-Q(y)) \ge c_4(\beta^*)(|\hat{b}_0 - b_0| - \Delta).$$

On the other hand, for all $x_i \in I_1^-$ we have

$$(6.6) \qquad Q(x_i) = \sum_{j=0}^{\beta^*} \hat{b}_j(x_i - x)^j + \Delta_i - f(x_i) - \Delta \ge Y_i - f(x_i) - 2\Delta.$$

Combining the inequalities in (6.5) and (6.6), we conclude that

$$|\tilde{f}_k(x) - f(x)| = |\hat{b}_0 - b_0| \le -c^*(\beta^*)\max\{\varepsilon_i : x_i \in I_1^-\} + c^*(\beta^*)\Delta$$

for some positive constant $c^*(\beta^*)$. Choosing $J(\beta^*)$ sufficiently large (regardless of k, n and f) there exists some $l = 1, \ldots, 2J(\beta^*)$ such that $x + h_k\mathcal{I}_l \subseteq I_1^-$, and hence

$$|\tilde{f}_k(x) - f(x)| \le c^*(\beta^*)\Delta - c^*(\beta^*) \cdot Z_j(h_k, x),$$

which completes the proof. $\quad\square$

LEMMA 6.1. *Let Q by any polynomial with the degree $\le \beta^*$ and $I \subseteq [x - h_k, x + h_k]$ be an interval with the length $\ge c_5(\beta^*)h_k$ for some constant $c_5(\beta^*) > 0$. Then there exist some finite constants $c_6(\beta^*), c_7(\beta^*) > 0$ which only depend on β^* and some interval $I^* \subseteq I$ with the length $\ge c_6(\beta^*)h_k$ such that*

$$\inf_{y\in I^*} |Q(y)| \ge c_7(\beta^*) \cdot \sup_{|z-x|\le h_k} |Q(z)|.$$

PROOF. If Q is a constant function, the assertion is satisfied by putting $c = 1$. Otherwise, by the fundamental theorem of algebra, Q can be represented by

$$Q(y) = \alpha_Q \prod_{j=1}^{\beta'} (y - y_j),$$

where $1 \le \beta' \le \beta^*$, the y_j denote the complex-valued roots of Q. By the pigeon hole principle there exists some square $I_1^+ \times [-c_5(\beta^*)h_k/(2\beta^* + 2), c_5(\beta^*)h_k/(2\beta^* + 2)]$ in the complex plane which does not contain any y_j where $I_1^+ \subseteq I$ has the length $c_5(\beta^*)h_k/(\beta^* + 1)$. Now we shrink that square by the factor $1/2$ where the center of the square does not change, leading to the square $I_2^+ \times [-c_5(\beta^*)h_k/(4\beta^* + 4), c_5(\beta^*)h_k/(4\beta^* + 4)]$. Thus, for any y in this shrinked square, the distance between y and any y_j is bounded from below by $c_5(\beta^*)h_k/(4\beta^* + 4)$ and by $|y_j - x| - h_k$. If the latter bound dominates, we have $|y_j - x| \ge \{1 + c_5(\beta^*)/(4\beta^* + 4)\} \cdot h_k$. Then the distance between any $z \in [x - h_k, x + h_k]$ and y_j has the upper bound

$$|y_j - x| + h_k \le |y_j - y| + 2h_k \le \{1 + (8\beta^* + 8)/c_5(\beta^*)\} \cdot |y_j - y|,$$

when applying the first bound. Otherwise, if the first bound dominates, we have

$$|z - y_j| \le |y_j - x| + h_k \le \{2 + c_5(\beta^*)/(4\beta^* + 4)\} \cdot h_k$$
$$\le |y - y_j| \cdot (4\beta^* + 4)\{2/c_5(\beta^*) + 1/(4\beta^* + 4)\}.$$

In both cases $|z - y_j|$ is bounded from above by a uniform constant $c_6(\beta^*)$ times $|y - y_j|$. Then we learn from the root-decomposition of the polynomial Q that

$$\inf_{y \in I^*} |Q(y)| \ge c_7(\beta^*) \cdot \sup_{|z-x| \le h_k} |Q(z)|,$$

for some deterministic constant $c_7(\beta^*) > 0$, which only depends on β^*. $\square$

PROOF OF PROPOSITION 3.1. Part (a) follows directly from the definition of Lepski's method. For part (b) we obtain from (3.4), and repeated application of the triangle and Jensen's inequality that

$$\mathbb{E}_f\big[\|\hat{f} - \tilde{f}_{\hat{k}^*}\|^q \mathbf{1}(\hat{k} < \hat{k}^*)\big]^{1/q}$$

$$\le \mathbb{E}_f\big[\|\tilde{f}_{\hat{k}} - f\|^q \mathbf{1}(\hat{k} < \hat{k}^*)\big]^{1/q} + \mathbb{E}_f\big[\|\tilde{f}_{\hat{k}^*} - f\|^q \mathbf{1}(\hat{k} < \hat{k}^*)\big]^{1/q}$$

$$\le 2^{(q-2)/q}$$

(6.7)
$$\times \big(\mathbb{E}_f\big[(R_{\hat{k}}^q + B_{\hat{k}}^q)\mathbf{1}(\hat{k} < \hat{k}^*)\big]^{1/q} + \mathbb{E}_f\big[(R_{\hat{k}^*}^q + B_{\hat{k}^*}^q)\mathbf{1}(\hat{k} < \hat{k}^*)\big]^{1/q}\big)$$

$$\le 2^{(2q-1)/q}\big(\mathbb{E}_f\big[B_{\hat{k}^*}^q\big]^{1/q} + \mathbb{E}_f\big[R_{\hat{k}}^q \mathbf{1}(\hat{k} < \hat{k}^*)\big]^{1/q}\big)$$

$$\le 2^{(2q-1)/q}\bigg(\mathbb{E}_f\big[\hat{\mathfrak{z}}_{\hat{k}^*}^q\big]^{1/q} + \sum_{k=0}^{K-1} \mathbb{E}_f\big[R_k^q \mathbf{1}(\hat{k} = k, k < \hat{k}^*)\big]^{1/q}\bigg),$$

where we also used that R_k decreases in k and B_k increases in k. Note that

$$
\begin{aligned}
\mathbf{1}(\hat{k} = k, k < \hat{k}^*) &\leq \mathbf{1}(\exists l \leq k : \|\tilde{f}_{k+1} - \tilde{f}_l\| > \hat{\mathfrak{z}}_l^T + \hat{\mathfrak{z}}_{k+1}^T) \cdot \mathbf{1}(k < \hat{k}^*) \\
&\leq \mathbf{1}(\exists l \leq k : \|\tilde{f}_{k+1} - f\| + \|f - \tilde{f}_l\| > \hat{\mathfrak{z}}_l^T + \hat{\mathfrak{z}}_{k+1}^T) \cdot \mathbf{1}(k < \hat{k}^*) \\
&\leq \mathbf{1}(\exists l \leq k : R_l + B_{k+1} > \hat{\mathfrak{z}}_l^T) \cdot \mathbf{1}(B_{k+1} \leq \hat{\mathfrak{z}}_{k+1}^T/2) \\
&\leq \mathbf{1}(\exists l \leq k : R_l > \hat{\mathfrak{z}}_l^T/2).
\end{aligned}
$$

Inserting this inequality into (6.7) completes the proof. $\quad\square$

In the sequel, the following three lemmas will be useful. The proofs are given in the supplementary material [35].

LEMMA 6.2. *If $y, t \to \infty$ and*

$$
y = \mathfrak{c}(\log t)^{\mathfrak{b}} t^{\mathfrak{a}}(1 + \mathcal{O}(1)), \qquad \mathfrak{c}, \mathfrak{a} > 0, \mathfrak{b} \in \mathbb{R},
$$

then

$$
t = (\mathfrak{c}^{-1}(\log y^{1/\mathfrak{a}})^{-\mathfrak{b}} y)^{1/\mathfrak{a}} + \mathcal{O}(1).
$$

In particular, if we have $v = \mathcal{U}_x(y)$ with $v \to 0$, then

$$
F_x(v) = 1 - \mathfrak{c}_x^{-\mathfrak{a}_x}(\log |v|^{-1/\mathfrak{a}_x})^{-\mathfrak{b}_x \mathfrak{a}_x} |v|^{\mathfrak{a}_x}(1 + \mathcal{O}(1)).
$$

LEMMA 6.3. *For $1 \leq j_0, j_1 \leq n$, let $\mathcal{J} = \{j_0, \dots, j_1\}$ such that $|j_0 - j_1|/n = \mathcal{O}(n^{-\rho_0})$ for some $0 < \rho_0 < 1$. If $u \to 0$, $u \leq -n^{-\rho_1}$ for some $\rho_1 > 0$, then*

$$
\prod_{j \in \mathcal{J}} P(\varepsilon_j \leq A_{x_{j_0}}(-u^{-1})) \leq e^{\#\mathcal{J} c_3^- u},
$$

where $c_3^- < 1$ may be chosen arbitrarily close to one.

LEMMA 6.4. *Let $(q_n)_n$ be a real-valued sequence which satisfies $q_n \in [1, \log n]$ for all integer n, and denote with $F(\cdot)$ the c.d.f. of ε. Then we have*

$$
\mathbb{E}|\max\{\varepsilon_1, \dots, \varepsilon_n\}|^{q_n} \leq (1 + \mathcal{O}(1)) \int_0^{n^{1/2}} ((-\mathcal{U})^{q_n}(n/y))^{(1)} \exp(-y) \, dy.
$$

If $\mathcal{U}(\cdot)$ is not differentiable, replace $\mathcal{U}(\cdot)$ with $c_2^+ A(\cdot)$ in the above inequality, where $c_2^+ > 1$ can be chosen arbitrarily close to one. If $q_n = q$ is finite and independent of n, we obtain that

$$
\mathbb{E}|\max\{\varepsilon_1, \dots, \varepsilon_n\}|^q = \mathcal{O}((\log n)^{q \mathfrak{b}_F} n^{-q/\mathfrak{a}_F}).
$$

For arbitrary $q_n \in [1, \log n]$ we have

$$
\mathcal{O}(n^{-c_2^+ \mathfrak{a}_F/q_n}) \leq \int_0^\infty F(-x^{1/q_n})^n \, dx \leq \mathcal{O}(n^{-c_2^- \mathfrak{a}_F/q_n}),
$$

where $0 < c_2^- < 1 < c_2^+$ can be chosen arbitrarily close to one.

PROOF OF THEOREM 3.2. In the course of the proof we will frequently apply Proposition 3.2. We may do so since condition $\mathfrak{h}_0 < \beta_x \mathfrak{a}_x/(\beta_x \mathfrak{a}_x + 1)$ implies (3.18). The general strategy is the following. By the triangle inequality and Jensen's inequality, we have

$$(6.8) \quad \mathbb{E}_f\big[(\hat{f}(x) - f(x))^2\big] \le 2\mathbb{E}_f\big[(\hat{f}(x) - \tilde{f}_{\hat{k}*})^2\big] + 2\mathbb{E}_f\big[(\tilde{f}_{\hat{k}*} - f(x))^2\big],$$

and we will treat both quantities separately. In order to deal with the first, Proposition 3.1 implies that it suffices to consider

$$\sup_{f \in H_{\mathcal{N}(x)}(\beta, L)} \mathbb{E}_f\big[(\hat{\mathfrak{z}}_{\hat{k}*}^T)^2\big]^{1/2} + \sup_{f \in H_{\mathcal{N}(x)}(\beta, L)} \sum_{k=0}^{K-1} \mathbb{E}_f\big[R_k^2 \mathbf{1}(\exists l \le k : R_l > \hat{\mathfrak{z}}_l^T/2)\big]^{1/2}$$

$$=: I + \sum_{k=0}^{K-1} II_k.$$

To treat I, we require the following simple lemma; the proof is given in the supplementary material [35].

LEMMA 6.5. *Let $q \ge 1$. Under the assumptions of Theorem* 3.2, *we have uniformly over $f \in H_{\mathcal{N}(x)}(\beta, L)$*

$$\mathbb{E}_f\big[(\hat{\mathfrak{z}}_{\hat{k}*}^T)^q\big] \le (c_1^+)^q \mathbb{E}_f\big[(\mathfrak{z}_{\hat{k}*}^T)^q\big] + \mathcal{O}(n^{-q/\mathfrak{a}_x}),$$

where $c_1^+ > 1$.

Applying the above result with $q = 2$ we obtain

$$(6.9) \qquad I^2 \le (c_1^+)^2 \mathbb{E}_f\big[(\mathfrak{z}_{\hat{k}*}^T)^2\big] + \mathcal{O}(n^{-2/\mathfrak{a}_x}),$$

and it remains to deal with $\mathbb{E}_f[(\mathfrak{z}_{\hat{k}*}^T)^2]$. We define

$$k^{\pm} := \inf\{k = 0, \ldots, K - 1 : B_{k+1} > c_2^{\pm} \mathfrak{z}_{k+1}^T/2\} \wedge K.$$

On the event $\mathcal{A}_n = \{c_2^- \mathfrak{z}_k \le \hat{\mathfrak{z}}_k \le c_2^+ \mathfrak{z}_k \text{ for all } k = 0, \ldots, K - 1\}$ we have $k^- \le \hat{k}* \le k^+$. From Proposition 3.2 we infer $P(\mathcal{A}_n^c) = \mathcal{O}(n^{-\upsilon} \log n)$ due to $K = \mathcal{O}(\log n)$. Since $\mathfrak{z}_k$ decreases monotonically in k, we deduce $\mathbb{E}[(\mathfrak{z}_{\hat{k}*}^T)^2] \le \mathfrak{z}_{k^-}^2 + \mathcal{O}(n^{-2/\mathfrak{a}_x})$. Note that the deterministic sequences $(h_k^{\pm})$ satisfy (see Lemma 6.2 for details)

$$(6.10) \qquad \begin{aligned} &\mathfrak{z}_{k^-} \sim -(h_k^{\pm})^{\beta_x} \quad \text{and} \\ &h_{k^{\pm}} \sim (n/\log n)^{-1/(\mathfrak{a}_x \beta_x + 1)} (\log n)^{(\mathfrak{a}_x \mathfrak{b}_x)/(\mathfrak{a}_x \beta_x + 1)}, \end{aligned}$$

under our assumption $\hbar_0 < \beta_0 \mathfrak{a}_0 / (\beta_0 \mathfrak{a}_0 + 1)$. We obtain

(6.11)
$$\sup_{f \in H_{\mathcal{N}(x)}(\beta, L)} \mathbb{E}_f\big[(\hat{\mathfrak{z}}_{\hat{k}^*}^T)^2\big]$$
$$= \mathcal{O}\big((n/\log n)^{(-2\beta_x)/(\mathfrak{a}_x \beta_x + 1)} (\log n)^{(2\mathfrak{a}_x \mathfrak{b}_x \beta_x)/(\mathfrak{a}_x \beta_x + 1)}\big),$$

and it remains to deal with the second part. Let $\mathcal{B}_k = \{\exists l \leq k : R_l > \hat{\mathfrak{z}}_l^T / 2\}$. Then for $\delta_0 = n^{-\hbar_0 / \mathfrak{a}_x}$, we have

$$II_k^2 = \int_0^\infty P_f\big(R_k^2 \mathbf{1}(\mathcal{B}_k) \geq x\big) \, dx \leq \int_0^{\delta_0} P_f(\mathcal{B}_k) \, dx + \int_{\delta_0}^\infty P\big(R_k^2 \geq x\big) \, dx$$

$$\leq \delta_0 \sum_{l=0}^k P_f\big(R_l > \hat{\mathfrak{z}}_l^T / 2\big) + \int_{\delta_0}^\infty P\big(R_k^2 \geq x\big) \, dx =: III_k + IV_k.$$

We first deal with III_k. Recall that for $j = 1, \ldots, 2J(\beta^*)$ we have

$$Z_j(h_k, x) = \max\{\varepsilon_i : x_i \in x + h_k \mathcal{I}_j\}$$

and

$$\mathcal{I}_j := \big[-1 + (j-1)/J(\beta^*), -1 + j/J(\beta^*)\big].$$

Put $\mathcal{J}_j(l) = \{0, 1/n, 2/n, \ldots, 1\} \cap \{x + h_l \mathcal{I}_j\}$ for $j = 1, \ldots, 2J(\beta^*)$, and note that $\#\mathcal{J}_j(l) \geq n h_l / 2J(\beta^*)$. An application of Proposition 3.2 and Theorem 3.1 then yields that

(6.12)
$$P_f\big(R_l > \hat{\mathfrak{z}}_l^T / 2\big) \leq \sum_{j=1}^{J(\beta^*)} P\big(|Z_j(h_l, x)| > c_1^- \mathfrak{z}_l\big) + \mathcal{O}\big(n^{-2/\mathfrak{a}_x}\big)$$

$$\leq \sum_{j=1}^{2J(\beta^*)} \prod_{i \in \mathcal{J}_j(l)} P\big(\varepsilon_i < -c_1^- \mathfrak{z}_l\big) + \mathcal{O}\big(n^{-2/\mathfrak{a}_x}\big),$$

where $c_1^- < 1$ may be chosen arbitrarily close to one. Arguing as in Lemma 6.3 we obtain

(6.13)
$$\prod_{i \in \mathcal{J}_j} P\big(\varepsilon_i > -c_1^- \mathfrak{z}_l\big) = \mathcal{O}\big(n^{-2c_2^- / \mathfrak{a}_x}\big),$$

where $c_2^- < 1$ may be chosen arbitrarily close to one. Hence we obtain that

(6.14)
$$III_k \leq \delta_0 \sum_{l=0}^k \sum_{j=1}^{2J(\beta^*)} \mathcal{O}\big(n^{-2c_2^- / \mathfrak{a}_x}\big) = \mathcal{O}\big(K \delta_0 n^{-2c_2^- / \mathfrak{a}_x}\big) = \mathcal{O}\big(n^{-2/\mathfrak{a}_x}\big),$$

since $K = \mathcal{O}(\log n)$. For dealing with IV_k, set $\eta_0 = \exp(n^{\hbar_0/4})$. Let $u_x = A_x^{-1}(\delta_0)$. Then Lemma 6.2 implies that $u_x < -n^{-\hbar_0/2}$ for large enough n. Then as in (6.13),

it follows from Lemma 6.3 that for sufficiently large n,

$$\int_{\delta_0}^{\eta_k} P(R_k^2 \geq x)\, dx = \int_{\delta_0}^{\eta_k} P(R_k \leq -x^{1/2})\, dx$$

$$\leq \eta_k \sum_{j=1}^{2J(\beta^*)} \prod_{i \in \mathcal{J}_j(k)} P(\varepsilon_i < A_x(-n^{-\mathfrak{h}_0/2}))$$

$$\leq \eta_k \sum_{j=1}^{2J(\beta^*)} \exp(-\#\mathcal{J}_{j(k)} n^{-\mathfrak{h}_0/2})$$

$$\leq \eta_k \sum_{j=1}^{2J(\beta^*)} \exp(-n^{\mathfrak{h}_0/2}) = \mathcal{O}(\exp(-n^{\mathfrak{h}_0/4})).$$

Let $p > 2$. Since $\mathbb{E}[|R_k|^p] = \mathcal{O}(1)$ by Lemma 6.4, it follows from the Markov inequality that

$$\int_{\eta_k}^{\infty} P(R_k^2 \geq x)\, dx \leq \int_{\eta_k}^{\infty} x^{-p/2} \mathbb{E}[|R_k|^p]\, dx = \frac{2}{p-2} \eta_k^{-p/2+1} \mathcal{O}(1).$$

Combining the above and (6.14), it follows that $IV_k = \mathcal{O}(\exp(-n^{\mathfrak{h}_0/8}))$, which in turn yields

$$(6.15) \quad II_k^2 = III_k + IV_k = \mathcal{O}(n^{-2/\mathfrak{a}_x}) + \mathcal{O}(\exp(-n^{\mathfrak{h}_0/8})) = \mathcal{O}(n^{-2/\mathfrak{a}_x}).$$

We thus conclude

$$\sum_{k=0}^{K-1} II_k = \mathcal{O}(Kn^{-1/\mathfrak{a}_x}) = \mathcal{O}(\log n\, n^{-1/\mathfrak{a}_x})$$

$$= \mathcal{O}((n\log n)^{-\beta_x/(\mathfrak{a}_x\beta_x+1)} (\log n)^{(\mathfrak{a}_x \mathfrak{b}_x)/(\mathfrak{a}_x\beta_x+1)}).$$

Piecing everything together and taking squares, we arrive at

$$(6.16) \quad \sup_{f \in H_{\mathcal{N}(x)}(\beta,L)} \mathbb{E}_f[(\hat{f}(x) - \tilde{f}_{\hat{k}*})^2]$$

$$= \mathcal{O}((n\log n)^{-\beta_x/(\mathfrak{a}_x\beta_x+1)} (\log n)^{(\mathfrak{a}_x \mathfrak{b}_x)/(\mathfrak{a}_x\beta_x+1)}).$$

To complete the proof, it remains to deal with $\mathbb{E}_f[(\tilde{f}_{\hat{k}*} - f(x))^2]$. Let $p' > 2$. Then by (3.20) and the triangle, Jensen and Hölder inequalities, we have

$$(6.17) \quad \mathbb{E}_f[(\hat{f}(x) - \tilde{f}_{\hat{k}*})^2] \leq 2\mathbb{E}_f[R_{\hat{k}*}^2 + B_{\hat{k}*}^2]$$

$$\leq 2\mathbb{E}_f[R_{\hat{k}*}^2 \mathbf{1}(\mathcal{A}_n) + B_{\hat{k}*}^2 \mathbf{1}(\mathcal{A}_n)]$$

$$(6.18) \quad + P(\mathcal{A}_n^c)^{(p'-2)/p'} (\mathbb{E}[R_{k^-}^{p'}]^{1/p'} + \mathcal{O}(L_x)).$$

Hence Proposition 3.2, Lemma 6.4 and (6.10) imply that the above is of order

$$\mathbb{E}_f\big[R_{\hat{k}*}^2 + B_{\hat{k}*}^2\big] \leq 2\mathbb{E}_f\big[R_{k-}^2\big] + B_{k+}^2 + \mathcal{O}(n^{-2/\mathfrak{a}_x})$$

(6.19)
$$= \mathcal{O}\big(n_{k-}^{-2/\mathfrak{a}_x}(\log n_{k-})^{2\mathfrak{b}_x} + \mathfrak{z}_{k+}^2 + n^{-2/\mathfrak{a}_x}\big)$$

$$= \mathcal{O}\big((n\log n)^{(-2\beta_x)/(\mathfrak{a}_x\beta_x+1)}(\log n)^{(2\mathfrak{a}_x\mathfrak{b}_x)/(\mathfrak{a}_x\beta_x+1)}\big).$$

The above bound is uniform over $H_{\mathcal{N}(x)}(\beta, L)$, and the proof is complete. □

PROOF OF THEOREM 3.3. For the proof we require the following lemma, which provides a sub-polynomial upper bound on the probability that R_k exceeds the threshold $\hat{\mathfrak{z}}_l^T/2$.

LEMMA 6.6. *Suppose* $h_k \leq \exp(-c_H \log^\gamma n)$ *for fixed constants* $\gamma \in (0, 1)$, $c_H > 0$. *Grant Assumption* 3.1, *and let* $\mathfrak{m}, \mathfrak{a}_0, \beta_0, \mathfrak{h}_0$ *satisfy* (3.18) *in view of* (3.19). *Then*

$$\sup_{k=0,\dots,K-1} P\big(R_k > \hat{\mathfrak{z}}_l^T/2\big) = \mathcal{O}\big(\exp(-c_H \log^{1+\gamma} n/2q)\big),$$

as $n \to \infty$.

Let $p > q$. Then Lemma 6.4 gives

(6.20)
$$\mathbb{E}\big[R_k^p\big]^{q/p} = \mathcal{O}\big((nh_k)^{-q/\mathfrak{a}_F}(\log nh_k)^{q\mathfrak{b}_F}\big).$$

An application of Hölder's inequality, (6.20) and Lemma 6.6 yields that

$$\sup_{f \in H_{[0,1]}(\beta,L)} \sum_{k=0}^{K-1} \mathbb{E}_f\big[R_k^q \mathbf{1}(\exists l \leq k : R_l > \hat{\mathfrak{z}}_l^T/2)\big]^{1/q}$$

$$\leq \sum_{k=0}^{K}(k+1)^{(p-q)/p} \mathbf{1}_{[0,\exp(-c_H \log^\gamma n)]}(h_k)$$

(6.21)
$$\times \mathcal{O}\big(n^{-\mathfrak{h}_0/\mathfrak{a}_F} \cdot \exp(-c_H/q(p-q)\log^{1+\gamma} n/[2qp])\big)$$

$$+ \sum_{k=0}^{K} \mathbf{1}_{(\exp(-c_H \log^\gamma n),\infty)}(h_k) \cdot (nh_k)^{-1/\mathfrak{a}_F}$$

$$= \mathcal{O}\big(n^{-c_2^-/\mathfrak{a}_F} \exp([c_H/\mathfrak{a}_F]\log^\gamma n) \cdot \log n\big).$$

Choosing $c_H, \gamma > 0$ sufficiently small, the above is of order $\mathcal{O}(n^{-c_3^-/\mathfrak{a}_F})$, where $c_3^- < 1$ can be chosen arbitrarily close to one. According to Proposition 3.1 it remains to bound the expectation $\mathbb{E}_f[(\hat{\mathfrak{z}}_{\hat{k}*}^T)^q]$ uniformly over $f \in H_{[0,1]}(\beta, L)$. Applying Lemma 6.5, we obtain that uniformly over $f \in H_{[0,1]}(\beta, L)$

$$\mathbb{E}\big[(\hat{\mathfrak{z}}_{\hat{k}*}^T)^q\big] \leq (Tc_1^+)^q \mathbb{E}_f\big[(\mathfrak{z}_{\hat{k}*})^q\big] + \mathcal{O}(n^{-q/\mathfrak{a}_F}).$$

To deal with $\mathbb{E}_f[(\mathfrak{z}_{\hat{k}*})^q]$, we introduce

$$k^{\pm} := \inf\{k = 0, \ldots, K - 1 : B_{k+1} > c_2^{\pm} \mathfrak{z}_{k+1}^T/2\} \wedge K.$$

On the event $\mathcal{A}_n = \{c_2^- \mathfrak{z}_k \leq \hat{\mathfrak{z}}_k \leq c_2^+ \mathfrak{z}_k$ for all $k = 0, \ldots, K - 1\}$ we have $k^- \leq \hat{k}^* \leq k^+$. From Proposition 3.2 we infer $P(\mathcal{A}_n^c) = \mathcal{O}(n^{-q/\mathfrak{a}_F})$. Since $\mathfrak{z}_k$ decreases monotonically in k, we find $\mathbb{E}[\mathfrak{z}_{\hat{k}*}^q] \leq \mathfrak{z}_{k^-}^q + \mathcal{O}(n^{-q/\mathfrak{a}_F})$. Note that the deterministic sequences $(h_k^{\pm})$ satisfy

(6.22)
$$h_{k^{\pm}} \sim (n)^{-1/(\mathfrak{a}_F\beta+1)}(\log n)^{(\mathfrak{a}_F\mathfrak{b}_F)/(\mathfrak{a}_F\beta+1)} \quad \text{and}$$
$$\mathfrak{z}_{k^-} \sim n^{-\beta/(\mathfrak{a}_F\beta+1)}(\log n)^{(\beta\mathfrak{a}_F\mathfrak{b}_F)/(\mathfrak{a}_F\beta+1)},$$

provided that $\mathfrak{h}_0 < \beta_0 \mathfrak{a}_0/(\beta_0 \mathfrak{a}_0 + 1)$. For computational details, refer to Lemma 6.2. Moreover, condition $\mathfrak{h}_0 < \beta_0 \mathfrak{a}_0/(\beta_0 \mathfrak{a}_0 + 1)$ also implies (3.18). We obtain

(6.23)
$$\sup_{f \in H_{[0,1]}(\beta,L)} \mathbb{E}_f[(\hat{\mathfrak{z}}_{\hat{k}*}^T)^q] = \mathcal{O}(n^{(-q\beta)/(\mathfrak{a}_F\beta+1)}(\log n)^{(q\beta\mathfrak{a}_F\mathfrak{b}_F)/(\mathfrak{a}_F\beta+1)}).$$

Combining this result with (6.21), Proposition 3.1 yields that

$$\sup_{f \in H_{[0,1]}(\beta,L)} \mathbb{E}_f[\|\hat{f} - \tilde{f}_{\hat{k}*}\|_q^q] = \mathcal{O}(n^{(-q\beta)/(\mathfrak{a}_F\beta+1)}(\log n)^{(q\beta\mathfrak{a}_F\mathfrak{b}_F)/(\mathfrak{a}_F\beta+1)}).$$

Arguing similarly as in (6.19), by (3.22) and (3.23) we deduce that

(6.24)
$$\mathbb{E}_f[\|\tilde{f}_{\hat{k}*} - f\|_q^q] \leq 2^q \mathbb{E}_f[B_{\hat{k}*}^q] + 2^q \mathbb{E}_f[R_{\hat{k}*}^q]$$
$$\leq 2^q B_{k^+}^q + \mathcal{O}(n^{-q/\mathfrak{a}_F}) + 2^q \mathbb{E}_f[R_{k^-}^q]$$
$$= \mathcal{O}(n^{(-q\beta)/(\mathfrak{a}_F\beta+1)}(\log n)^{(q\beta\mathfrak{a}_F\mathfrak{b}_F)/(\mathfrak{a}_F\beta+1)}),$$

uniformly with respect to $f \in H_{[0,1]}(\beta, L)$, by conditioning on the event $\mathcal{A}_n$ and using (6.20). The proof is complete. $\square$

The proofs of Theorems 4.1, 4.2 and 4.3 are given in the supplementary material [35].

SUPPLEMENTARY MATERIAL

Additional simulations, proof of lower bound, technical lemmas and sharpness estimation (DOI: 10.1214/14-AOS1248SUPP; .pdf). In the supplementary material we provide additional simulations and the proofs of the lower bound results as well as technical lemmas and sharpness estimation.

REFERENCES

[1] AIGNER, D., LOVELL, C. A. K. and SCHMIDT, P. (1977). Formulation and estimation of stochastic frontier production function models. *J. Econometrics* **6** 21–37. MR0448782

[2] BANKER, R. D. (1993). Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. *Manage. Sci.* **39** 1265–1273.

[3] BEIRLANT, J., GOEGEBEUR, Y., TEUGELS, J. and SEGERS, J. (2004). *Statistics of Extremes*: *Theory and Applications*. Wiley, Chichester. MR2108013

[4] CHERNOZHUKOV, V. and HONG, H. (2004). Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica* **72** 1445–1480. MR2077489

[5] CHICHIGNOUD, M. (2012). Minimax and minimax adaptive estimation in multiplicative regression: Locally Bayesian approach. *Probab. Theory Related Fields* **153** 543–586. MR2948686

[6] DAOUIA, A., FLORENS, J.-P. and SIMAR, L. (2010). Frontier estimation and extreme value theory. *Bernoulli* **16** 1039–1063. MR2759168

[7] DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory*: *An introduction*. *Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR2234156

[8] DE HAAN, L. and PENG, L. (1998). Comparison of tail index estimators. *Stat. Neerl.* **52** 60–70. MR1615558

[9] DEPRINS, D., SIMAR, L. and TULKENS, H. (1984). Measuring labor-efficiency in post offices. In *Public Goods*, *Environmental Externalities and Fiscal Competition* (P. Chander, J. Drèze, C. K. Lovell and J. Mintz, eds.) 285–309. Springer, New York.

[10] DONALD, S. G. and PAARSCH, H. J. (1993). Piecewise pseudo-maximum likelihood estimation in empirical models of auctions. *Internat. Econom. Rev.* **34** 121–148. MR1201732

[11] DREES, H. (1995). Refined Pickands estimators of the extreme value index. *Ann. Statist.* **23** 2059–2080. MR1389865

[12] DREES, H. (1998). Optimal rates of convergence for estimates of the extreme value index. *Ann. Statist.* **26** 434–448. MR1608148

[13] DREES, H. (2001). Minimax risk bounds in extreme value theory. *Ann. Statist.* **29** 266–294. MR1833966

[14] DREES, H., DE HAAN, L. and RESNICK, S. (2000). How to make a Hill plot. *Ann. Statist.* **28** 254–274. MR1762911

[15] FALK, M. (1995). Some best parameter estimates for distributions with finite endpoint. *Statistics* **27** 115–125. MR1377501

[16] FALK, M., HÜSLER, J. and REISS, R.-D. (1994). *Laws of Small Numbers*: *Extremes and Rare Events*. Birkhäuser, Basel. MR1296464

[17] FARRELL, M. J. (1957). The measurement of productive efficiency. *J. R. Stat. Soc. A*, *General* **120** 253–290.

[18] FRAGA ALVES, M. I. (2001). A location invariant Hill-type estimator. *Extremes* **4** 199–217 (2002). MR1907061

[19] GIJBELS, I., MAMMEN, E., PARK, B. U. and SIMAR, L. (1999). On estimation of monotone and concave frontier functions. *J. Amer. Statist. Assoc.* **94** 220–228. MR1689226

[20] GIJBELS, I. and PENG, L. (2000). Estimation of a support curve via order statistics. *Extremes* **3** 251–277 (2001). MR1856200

[21] GIRARD, S., GUILLOU, A. and STUPFLER, G. (2013). Frontier estimation with kernel regression on high order moments. *J. Multivariate Anal.* **116** 172–189. MR3049899

[22] GOLDENSHLUGER, A. and LEPSKI, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. MR2850214

[23] GOLDENSHLUGER, A. and ZEEVI, A. (2006). Recovering convex boundaries from blurred and noisy observations. *Ann. Statist.* **34** 1375–1394. MR2278361

[24] GRAMA, I. and SPOKOINY, V. (2008). Statistics of extremes by oracle estimation. *Ann. Statist.* **36** 1619–1648. MR2435450

[25] GUERRE, E., PERRIGNE, I. and VUONG, Q. (2000). Optimal nonparametric estimation of first-price auctions. *Econometrica* **68** 525–574. MR1769378

[26] HALL, P., NUSSBAUM, M. and STERN, S. E. (1997). On the estimation of a support curve of indeterminate sharpness. *J. Multivariate Anal.* **62** 204–232. MR1473874

[27] HALL, P. and PARK, B. U. (2004). Bandwidth choice for local polynomial estimation of smooth boundaries. *J. Multivariate Anal.* **91** 240–261. MR2087845

[28] HALL, P., PARK, B. U. and STERN, S. E. (1998). On polynomial estimators of frontiers and boundaries. *J. Multivariate Anal.* **66** 71–98. MR1648521

[29] HALL, P. and VAN KEILEGOM, I. (2009). Nonparametric "regression" when errors are positioned at end-points. *Bernoulli* **15** 614–633. MR2555192

[30] HÄRDLE, W., PARK, B. U. and TSYBAKOV, A. B. (1995). Estimation of nonsharp support boundaries. *J. Multivariate Anal.* **55** 205–218. MR1370400

[31] HE, L. (2001). Modeling and prediction of sunspot cycles. ProQuest LLC, Ann Arbor, MI. Ph.D. thesis, Massachusetts Institute of Technology. Available at http://dspace.mit.edu/handle/1721.1/8226. MR2717027

[32] HOSKING, J. R. M. and WALLIS, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* **29** 339–349. MR0906643

[33] IBRAGIMOV, I. A. and HASMINSKIĬ, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer, New York. Translated from the Russian by Samuel Kotz. MR0620321

[34] JEONG, S.-O. and PARK, B. U. (2006). Large sample approximation of the distribution for convex-hull estimators of boundaries. *Scand. J. Stat.* **33** 139–151. MR2255114

[35] JIRAK, M., MEISTER, A. and REISS, M. (2014). Supplement to "Adaptive function estimation in nonparametric regression with one-sided errors." DOI:10.1214/14-AOS1248SUPP.

[36] JIRAK, M. and MEISTER, A. Available at https://www.mathematik.hu-berlin.de/for1735/Publ/r_code_and_data.zip.

[37] KNEIP, A., PARK, B. U. and SIMAR, L. (1998). A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory* **14** 783–793. MR1666696

[38] KNEIP, A., SIMAR, L. and WILSON, P. W. (2008). Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. *Econometric Theory* **24** 1663–1697. MR2456542

[39] KNEIP, L., SIMAR, L. and VAN KEILEGOM, I. (2010). Boundary estimation in the presence of measurement error with unknown variance. Unpublished manuscript.

[40] KNIGHT, K. (2001). Limiting distributions of linear programming estimators. *Extremes* **4** 87–103 (2002). MR1893873

[41] KNIGHT, K. (2006). Asymptotic theory for *M*-estimators of boundaries. In *The Art of Semiparametrics* (S. Sperlich, W. Härdle and G. Aydinli, eds.) 1–21. Physica-Verlag, Heidelberg. MR2234873

[42] KOENKER, R., NG, P. and PORTNOY, S. (1994). Quantile smoothing splines. *Biometrika* **81** 673–680. MR1326417

[43] KOROSTELËV, A. P., SIMAR, L. and TSYBAKOV, A. B. (1995). Efficient estimation of monotone boundaries. *Ann. Statist.* **23** 476–489. MR1332577

[44] KOROSTELËV, A. P. and TSYBAKOV, A. B. (1993). *Minimax Theory of Image Reconstruction. Lecture Notes in Statistics* **82**. Springer, New York. MR1226450

[45] KUMBHAKAR, S. C., PARK, B. U., SIMAR, L. and TSIONAS, E. G. (2007). Nonparametric stochastic frontiers: A local maximum likelihood approach. *J. Econometrics* **137** 1–27. MR2347943

[46] LASNE, F. and DE CEAURRIZ, J. (2000). Recombinant erythropoietin in urine. *Nature* **405** 1084–1087.

[47] LEPSKI, O. V. and SPOKOINY, V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* **25** 2512–2546. MR1604408

[48] LEPSKIĬ, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatn. Primen.* **35** 459–470. MR1091202

[49] MEEUSEN, W. and VAN D. BROECK, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *Internat. Econom. Rev.* **18** 435–444.

[50] MEISTER, A. and REISS, M. (2013). Asymptotic equivalence for nonparametric regression with nonregular errors. *Probab. Theory Related Fields* **155** 201–229. MR3010397

[51] MÜLLER, U. U. and W., WEFELMEYER (2010). Estimation in nonparametric regression with nonregular errors. *Comm. Statist. Theory Methods* **39** 1619–1629.

[52] PARK, B. U., JEONG, S.-O. and SIMAR, L. (2010). Asymptotic distribution of conical-hull estimators of directional edges. *Ann. Statist.* **38** 1320–1340. MR2662344

[53] PARK, B. U., SICKLES, R. C. and SIMAR, L. (2007). Semiparametric efficient estimation of dynamic panel data models. *J. Econometrics* **136** 281–301. MR2328594

[54] PARK, B. U., SIMAR, L. and WEINER, CH. (2000). The FDH estimator for productivity efficiency scores. *Econometric Theory* **16** 855–877. MR1803713

[55] PICKANDS, J. III (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3** 119–131. MR0423667

[56] SOLANKI, S. K., USOSKIN, I. G., KROMER, B., SCHÜSSLER, M. and BEER, J. (2004). Unusual activity of the sun during recent decades compared to the previous 11,000 years. *Nature* **431** 1084–1087.

M. JIRAK
M. REISS
INSTITUT FÜR MATHEMATIK
HUMBOLDT-UNIVERSITÄT ZU BERLIN
UNTER DEN LINDEN 6
D-10099 BERLIN
GERMANY
E-MAIL: jirak@math.hu-berlin.de
        mreiss@mathematik.hu-berlin.de

A. MEISTER
INSTITUT FÜR MATHEMATIK
UNIVERSITÄT ROSTOCK
D-18051 ROSTOCK
GERMANY
E-MAIL: alexander.meister@uni-rostock.de