Submitted to the Annals of Statistics
arXiv: math.PR/0000000

SUPPLEMENTARY MATERIAL??

BY MORITZ JIRAK[†,*], ALEXANDER MEISTER[‡] AND MARKUS REISS[†]

Humboldt Universität zu Berlin[†] *and Universität Rostock*[‡]

Additional proofs and technical results are presented.

In Section 7 of this supplement implementation details and further simulation results are discussed. The proofs of the lower bounds are given in Section 8. Technical auxiliary results are provided in Section 9, and the estimation of the critical values is analysed in Section 10.

7. Additional simulations. The aim of this section is to discuss implementation details an explore some additional features that were briefly mentioned in Section 5. The corresponding R code is available at [4]. We will illustrate the following points:

(A) Comments on implementation details.
(B) Effect of high polynomial degree on the estimators.
(C) The effect of correct estimation of $\mathfrak{a}_x$, $\mathfrak{b}_x$ and $\mathfrak{c}_x$, and the induced bias by the regression function.
(D) Comparison with oracle estimators.

We consider three different regression functions, displayed in black in Figures 1, 2 and 4b:

$$f_1(x) = -2 \cdot \mathbf{1}(x < 1/3) - 3 \cdot \mathbf{1}(1/3 \leq x < 2/3) - \mathbf{1}(2/3 < x),\ x \in [0,1],$$
$$f_2(x) = -2 + 2\cos(2\pi x) + 0.3\sin(19\pi x),\ x \in [0,1],$$
$$f_3(x) = -2 + 2\cos(2\pi x),\ x \in [0,1].$$

A. Let us comment on some implementation details. The value β^* of the approximating polynomial was set to two in most of the simulations. In case of function f_1 (Figure 1), higher order polynomial approximations may lead to a better resolution around the jump points, but also a more pronounced overshooting (Gibb's phenomena). On the other hand, imposing a linear structure

*supported by Deutsche Forschungsgemeinschaft via FOR1735 *Structural Inference in Statistics: Adaptation and Efficiency* is gratefully acknowledged.

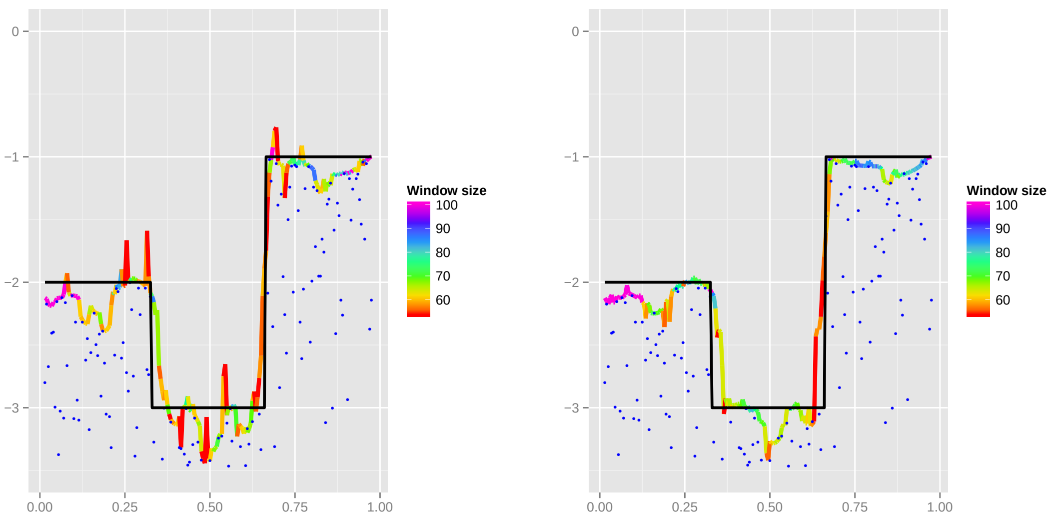
AMS 2000 subject classifications: 62G08, 62G32

Keywords and phrases: adaptive convergence rates, bandwidth selection, frontier estimation, Lepski's method, minimax optimality, negative Hill estimator, non-regular regression

($\beta^* = 1$) also slightly improves upon the results. In case of function f_2 (Figure 2), higher polynomial approximations may lead to better results. Even in case of $\beta^* = 8$, the resulting estimates are still remarkably accurate, despite the comparatively small sample size of $n = 200$. A possible way to obtain smoother results is to introduce weights in the spirit of the general local likelihood approach. For the estimation of $\mathfrak{a}_x$, we use a simple 'robustification', which in essence consists in takeing the median of the neighbouring estimates according to the number of the initial bandwidth (11 in most cases). The admissibility of such an estimator in the sense of Definition 1 can be readily established, exploiting the Hölder continuity of $\mathfrak{a}_x$.

Concerning the estimation procedure itself, we have chosen the pointwise quantiles in all simulations. The quantiles for L_q-adaptivity give, in general, similar results. The initial minimum sample size was set to 15 ($2h_0 = 15/200$), and ρ was chosen as $\rho = 20/19$. In general, ρ should be chosen close to one, otherwise the growth rate of the bandwidth might be too large. There is no danger of choosing ρ too small as long as $\rho > 1$, this might only lead to a slight increase on computation time in practice. A delicate issue for estimation procedures in practice is the actual choice of otherwise "asymptotic" constants. As is usually the case in practice, such asymptotic constants do not always give the best performance in finite samples. In our case, this concerns the constants $c(\beta^*)$ and $J(\beta^*)$ in Theorem 3.1. Further simulations show that the estimation procedure is not that sensitive to different values $4c(\beta^*)$ and $4J(\beta^*)$. In particular, simply setting $8c(\beta^*) = 4J(\beta^*) = 1$ yields a good performance of both the pointwise and L_q-adaptive estimator. One also needs to be careful about estimates of $\widehat{\mathfrak{b}}_x$. If they are too large, they may become the dominating factor in finite samples, and even lead to non-monotone quantile estimates (critical values), which is prohibited. We therefore use the truncation $|\widehat{\mathfrak{b}}_x| \leq 1/2$, which seems to produce good results.

B. As was already mentioned above in Paragraph A, the regression estimate $\widehat{f}$ can be rather stable under differing degrees of the underlying approximating polynomial, provided that the regression function f itself permits such a local approximation. In Figures 1 and 2, we compare a second order approximation with an approximation of order eight. Figure 1a reveals the well-known Gibbs-phenomena of an overshoot at jump points and neighbouring points, the remaining parts appear to similar as in Figure 1b. Note however that due to the significantly higher polynomial degree, the resulting estimate $\widehat{f}$ has more 'peaks' and 'steep valleys', since the higher polynomial order allows for much more local adaption to the data. If the order is too large - which seems to be the case in Figure 1a - the result is an over-parameterised estimate. In contrast, the estimate in Figure 2a does not yet show any significant signs of

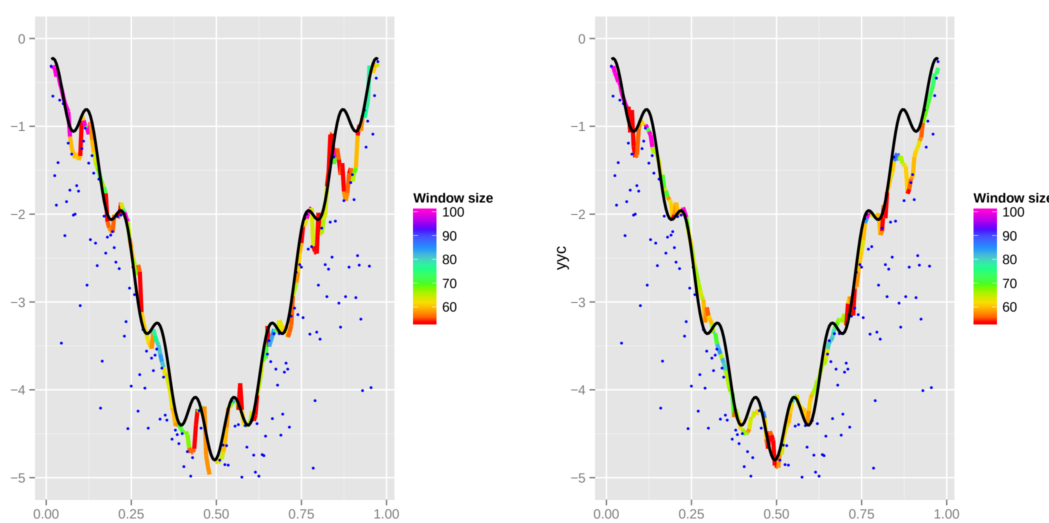


(a) (b)

Fig 1: (a) Function $f = f_1$, $\beta^* = 8$, $n = 200$, $\varepsilon_j \sim Exp(1)$; (b) Function $f = f_1$, $\beta^* = 2$, $n = 200$, $\varepsilon_j \sim Exp(1)$;

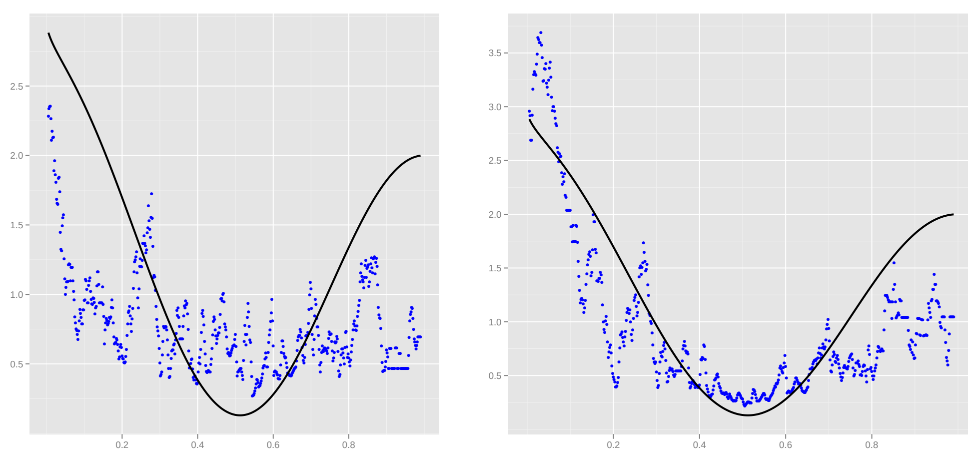
over-parametrisation. It is likely that this can be attributed to the much more complicated underlying regression function $f = f_2$, compared to the simpler case $f = f_1$ in Figure 1.

C. In this paragraph, we briefly discuss the interplay between the estimation of the critical values and the total estimate of the regression function f with respect to estimation accuracy. Figure 3 first reveals that the regression function f itself can impose a considerable bias on $\widehat{\mathfrak{a}}_x$ (and thus the critical values) if it fluctuates or changes directions often. The original random sample in both simulations is identical, the sole difference is the regression function f. As one consequence, we have a slight overestimation in the center of Figure 4a, and a rather precise estimate in Figure 4b. The (surprisingly) small deviation in the center of Figure 4a is due to the fact that a lot of observations are near the regression function, keeping the overall error small. The converse effect is visible at the top left in Figures 4a and 4b, where the underestimation of the critical values has a more visible impact on the estimation accuracy. We also note that the static initial bandwidth selection appears to be too low for large $\widehat{\mathfrak{a}}_x$ in Figure 4. This suggests that an adaptive (with respect to $\widehat{\mathfrak{a}}_x$) initial bandwidth selection might be an option. On the other hand, this can



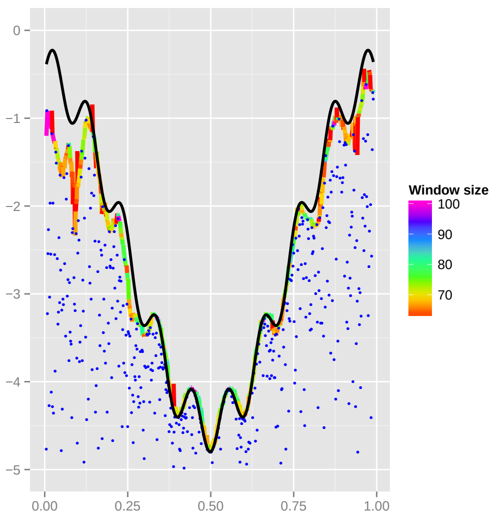
(a) (b)

Fig 2: (a) Function $f = f_2$, $\beta^* = 8$, $n = 200$, $\varepsilon_j \sim Exp(1)$; (b) Function $f = f_2$, $\beta^* = 2$, $n = 200$, $\varepsilon_j \sim Exp(1)$;

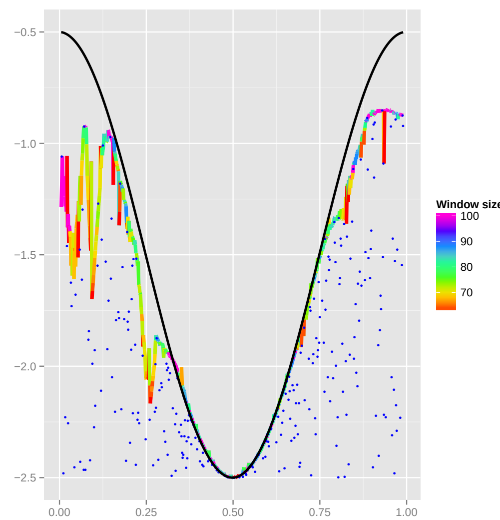


(a) (b)

Fig 3: (a) $\mathfrak{a}_x$ (black line) and $\widehat{\mathfrak{a}}_x$ (blue points), $f = f_2$; (b) $\mathfrak{a}_x$ (black line) and $\widehat{\mathfrak{a}}_x$ (blue points), $f = f_3$;



(a)



(b)

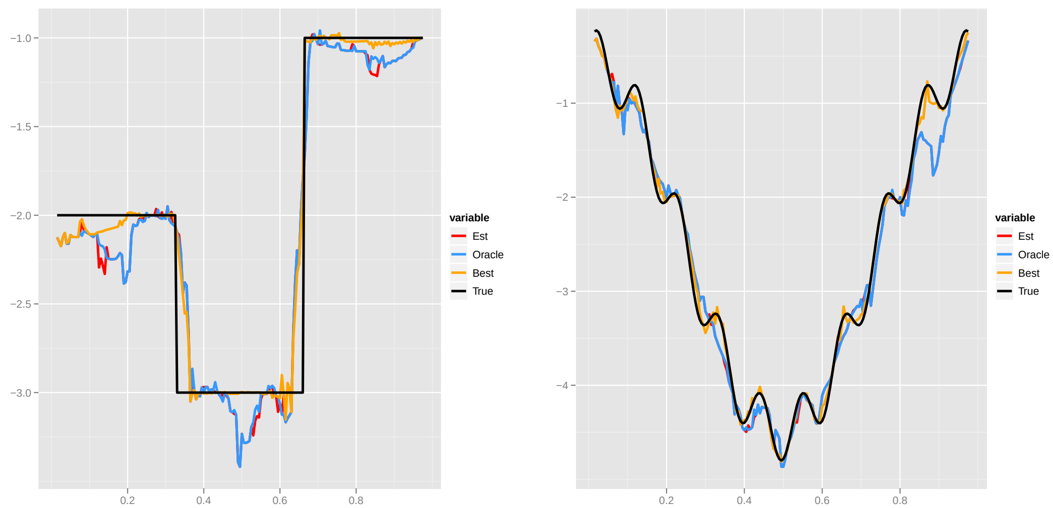
Fig 4: (a) Function $f = f_2$, $\beta^* = 4$, $n = 600$, $\varepsilon_j \sim \Gamma(\mathfrak{a}_x, 1)$; (b) Function $f = f_3$, $\beta^* = 2$, $n = 600$, $\varepsilon_j \sim \Gamma(\mathfrak{a}_x, 1)$;

also backfire if the estimator $\widehat{\mathfrak{a}}_x$ is severely off target. As a final comment: if one wonders about the two different estimation gaps in Figures 4a and 4b at the top left corners, note the different scaling.

D. This paragraph is devoted to the comparison with oracle estimators. As comparison, we consider the situation where $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}$ and hence the exact critical values are known. The corresponding oracle estimator is then constructed in exactly the same way, based on the true critical values. In addition, we consider the best possible estimator that is constructed by means of the base estimators (see Section 2). The setup we consider is that of Paragraph A, where we use $\beta^* = 2$ for all estimators. The measure of deviance is the pointwise mean

$$\Delta\big(\widehat{f}, f\big) = \frac{1}{n}\sum_{i=1}^{n}\big|\widehat{f}(i/n) - f(i/n)\big|.$$

The results are given in Table 1. Interestingly, the adaptive estimator and the oracle estimator where $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}$ are known have almost the identical risk. Figure 5 also indicates that they behave almost identically. Both have about three times the risk of the best possible base estimator, rendering a good performance.



(a) (b)

Fig 5: (a) Function $f = f_1$, $\beta^* = 2$, $n = 200$, $\varepsilon_j \sim Exp(1)$; (b) Function $f = f_2$, $\beta^* = 2$, $n = 200$, $\varepsilon_j \sim Exp(1)$;

$n = 200$	Adaptive Estimator	Oracle estimator	Best estimator
f_1	0.120	0.120	0.055
f_2	0.148	0.147	0.053
f_3	0.094	0.093	0.022
$n = 1000$			
f_1	0.058	0.057	0.039
f_2	0.057	0.056	0.022
f_3	0.036	0.035	0.011

TABLE 1

Comparison to oracle estimators with $\beta^* = 2$;

8. Proof of the lower bound. For the sake of reference, we restate the the results concerning the lower bound.

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

$$\int_{-\infty}^{0} |f_\varepsilon(x+\vartheta) - f_\varepsilon(x)|^2 / f_\varepsilon(x) dx \,\leq\, c_\varepsilon \vartheta^{\mathfrak{a}} |\log \vartheta|^{-\mathfrak{a}\mathfrak{b}}\,, \quad \forall \vartheta \in (0,1)\,, \tag{8.1}$$

for some $\mathfrak{a} \in (0, 2]$ and $\mathfrak{b} \in \mathbb{R}$.

THEOREM 8.1. *Assume condition (8.1) and fix some arbitrary $x_0 \in [0,1]$, $\beta_1 > \beta_2 > 0$ and $C_0, C_1 > 0$. Let $\{\widehat{f}_n(x_0)\}_n$ be any sequence of estimators of $f(x_0)$ based on the data $Y_1, \ldots, Y_n$ which satisfies*

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

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

$$\liminf_{n\to\infty} (n/\log n)^{\frac{2\beta_2}{1+\mathfrak{a}\beta_2}} (\log n)^{\frac{-2\mathfrak{a}\mathfrak{b}\beta_2}{1+\mathfrak{a}\beta_2}} \sup_{f \in \mathcal{H}_{[0,1]}(\beta_2, C_1)} \mathbb{E}_f \big| \widehat{f}_n(x_0) - f(x_0) \big|^2 \,>\, 0\,.$$

For completeness we also derive the L_q-minimax optimality of the convergence rates established by our estimator $\widehat{f}$ in Theorem 3.3.

THEOREM 8.2. *Assume condition (8.1) and let $\{\widehat{f}_n\}_n$ 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*

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

Now we focus on the case $\mathfrak{a} > 2$. To simplify some of the technical arguments in the proofs, we restrict to the case $\mathfrak{b} = 0$.

DEFINITION 1. *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*

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

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

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.

THEOREM 8.3. *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 $\{\widehat{f}_n(x_0)\}_n$ be any sequence of estimators of $f(x_0)$ based on the data $Y_1, \ldots, Y_n$ which satisfies*

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

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

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

PROOF OF THEOREM 8.1:. Without loss of generality assume $x_0 = 0$. We introduce the regression curves $f_0(x) = 0$ and $f_1(x) = -dh_n^{\beta_2}K(x/h_n)$ for some $h_n \downarrow 0$ to be specified, the kernel function $K(x) := \exp\big(1/(x^2-1)\big)1_{(-1,1)}(x)$ and some constant $d > 0$ sufficiently small such that $f_1 \in \mathcal{H}_{[0,1]}(\beta_2, C_1)$ is satisfied. We easily realise that $f_0 \in \mathcal{H}_{[0,1]}(\beta_1, C_0)$ so that

$$\mathbb{E}_{f_0}\big|\widehat{f}_n(0) - f_0(0)\big|^2 \;\le\; qn^{-2\beta_2/(1+\beta_2\mathfrak{a})}n^{-\xi}\,,$$

for some constant q by assumption. In the sequel we write P_f for the joint probability measure of the data $Y_1, \ldots, Y_n$ where f indicates the true regression function; and $\mathbb{E}_f$ denotes the corresponding expectation. Hence, P_f has the n-dimensional Lebesgue density

$$p_f(y) \;:=\; \prod_{j=1}^n f_\varepsilon\big(y_j - f(x_j)\big)\,, \quad y \in \mathbb{R}^n\,.$$

As the function f_1 is non-positive the probability measure P_{f_0} dominates P_{f_1} so that the Radon-Nikodym derivative dP_{f_1}/dP_{f_0} exists and

$$\begin{aligned}
&\sup_{f\in\mathcal{H}_{[0,1]}(\beta_2,C_1)} \big\{\mathbb{E}_f\big|\widehat{f}_n(0) - f(0)\big|^2\big\}^{1/2} \ge \mathbb{E}_{f_1}\big|\widehat{f}_n(0) - f_1(0)\big| \\
&\ge \big|f_0(0) - f_1(0)\big| \,-\, \mathbb{E}_{f_1}\big|\widehat{f}_n(0) - f_0(0)\big| \\
&= \big|f_0(0) - f_1(0)\big| \,-\, \mathbb{E}_{f_0}\big|\widehat{f}_n(0) - f_0(0)\big|\Big(\frac{dP_{f_1}}{dP_{f_0}} - 1\Big) \,-\, \mathbb{E}_{f_0}\big|\widehat{f}_n(0) - f_0(0)\big| \\
&\ge \big|f_0(0) - f_1(0)\big| \,-\, \big\{\mathbb{E}_{f_0}\big|\widehat{f}_n(0) - f_0(0)\big|^2\big\}^{1/2}\Big(\Big\{\mathbb{E}_{f_0}\Big(\frac{dP_{f_1}}{dP_{f_0}} - 1\Big)^2\Big\}^{1/2} + 1\Big) \\
&\ge deh_n^{\beta_2} \,-\, qn^{-\beta_2/(1+\mathfrak{a}\beta_2)}n^{-\xi/2}\big\{1 + \chi^2(P_{f_0}, P_{f_1})\big\}\,,
\end{aligned} \tag{8.2}$$

by the Cauchy-Schwarz inequality. Note that $\chi^2(P_{f_0}, P_{f_1}) := \mathbb{E}_{f_0}\Big(\frac{dP_{f_1}}{dP_{f_0}} - 1\Big)^2$ denotes the χ^2-distance between the measures P_{f_0} and P_{f_1}. We deduce that

$$\chi^2(P_{f_0}, P_{f_1}) \,=\, \mathbb{E}_{f_0}\Big(\frac{dP_{f_1}}{dP_{f_0}}\Big)^2 - 1 \,=\, \prod_{j=1}^n \Big(\int f_\varepsilon^2(y - f_1(x_j))/f_\varepsilon(y - f_0(x_j))dy\Big) - 1$$

$$= \prod_{j=1}^n \big\{1 + \chi^2\big(f_\varepsilon(\cdot - f_0(x_j)), f_\varepsilon(\cdot - f_1(x_j))\big)\big\} - 1$$

$$\leq \exp\Big\{ \sum_{j=1}^n \chi^2\big(f_\varepsilon(\cdot - f_0(x_j)), f_\varepsilon(\cdot - f_1(x_j))\big)\Big\} - 1$$

$$\leq \exp\Big\{c_\varepsilon \sum_{j=1}^n |f_1(x_j)|^{\mathfrak{a}}\Big\} - 1 \,\leq\, \exp\Big\{\text{const.}\cdot n h_n^{1+\beta_2\mathfrak{a}} |\log h_n|^{-\mathfrak{a}\mathfrak{b}}\Big\} - 1\,, \tag{8.3}$$

due to (8.1) whenever $h_n^{-1} = o(n)$ for n sufficiently large. We select

$$h_n = \{c_H n/\log n\}^{-1/(1+\mathfrak{a}\beta_2)} \cdot (\log n)^{\alpha\mathfrak{b}/(1+\alpha\beta_2)}$$

with some sufficiently small constant $c_H > 0$ so that

$$\chi^2(P_{f_0}, P_{f_1}) \,\leq\, \text{const.}\cdot n^{\xi/2}\,.$$

Inserting that into (8.2) provides the desired lower bound. □

Proof of Theorem 8.2:. With K as in the proof of Theorem 8.1, we consider the functions

$$f_\theta(x) \,:=\, -d \sum_{j=1}^{m_n-1} \theta_j m_n^{-\beta_2} K\big(2m_n(x - x_{j,n})\big)\,,$$

where $\theta = (\theta_1, \ldots, \theta_{m_n-1}) \in \{0,1\}^{(m_n-1)}$, $x_{j,n} = j/m_n$ and $(m_n)_n \uparrow \infty$ denotes some integer-valued sequence which remains to be selected. Again we may choose $d > 0$ sufficiently small such that all functions f_θ lie in $\mathcal{H}_{[0,1]}(\beta_2, L)$. Thus the supremum of the L_2-risk of the estimator $\widehat{f}_n$ taken over $f \in \mathcal{H}_{[0,1]}(\beta_2, L)$ is bounded from below by its Bayesian risk when the prior distribution of the vector θ is the uniform distribution on $\{0,1\}^{(m_n-1)}$.

According to that standard strategy we have

$$\sup_{f\in\mathcal{H}_{[0,1]}(\beta_2,L)} \mathbb{E}_f\|\hat f_n - f\|_q^q \;\geq\; \frac{1}{2}\sum_{j=1}^{m_n-1}\mathbb{E}_{\theta,-j}\sum_{b=0}^{1}\mathbb{E}_{f_{\theta,j,b}}\int_{x_{j,n}-1/(2m_n)}^{x_{j,n}+1/(2m_n)}\big|\hat f_n(x)-f_{\theta,j,b}(x)\big|^q dx$$
$$\geq\; \frac{1}{4}\sum_{j=1}^{m_n-1}\mathbb{E}_{\theta,-j}\int_{x_{j,n}-1/(2m_n)}^{x_{j,n}+1/(2m_n)}\big|f_{\theta,j,0}(x)-f_{\theta,j,1}(x)\big|^q dx$$
$$\cdot\int \min\big\{dP_{f_{\theta,j,1}}/dP_{f_{\theta,j,0}},1\big\}dP_{f,\theta,0}\,,$$

with P_f as in the proof of Theorem 8.1. Moreover, $\mathbb{E}_{\theta,-j}$ denotes the expectation with respect to the random vector θ when its jth component is removed; and $f_{\theta,j,b}$, $b=0,1$, stands for the regression function f_θ when the jth component of the index vector θ is replaced by b. We have

$$\int_{x_{j,n}-1/(2m_n)}^{x_{j,n}+1/(2m_n)}\big|f_{\theta,j,0}(x)-f_{\theta,j,1}(x)\big|^q dx \;\geq\; \frac{1}{2}d^q m_n^{-q\beta_2-1}\|K\|_q^q\,,$$

almost surely. Hence, we establish the lower bound $m_n^{-q\beta_2}$ on the convergence rate if we show that

$$\inf_{n\in\mathbb{N}}\inf_{j=1,\ldots,m_n-1}\mathbb{E}_{\theta,-j}\int\min\big\{dP_{f_{\theta,j,1}}/dP_{f_{\theta,j,0}},1\big\}dP_{f,\theta,0}\;>\;0\,. \tag{8.4}$$

For that purpose we consider that

$$\int\min\big\{dP_{f_{\theta,j,1}}/dP_{f_{\theta,j,0}},1\big\}dP_{f,\theta,0}\;\geq\;1-\sqrt{1-\Big(1-\frac{1}{2}\chi^2\big(P_{f_{\theta,j,0}},P_{f_{\theta,j,1}}\big)\Big)^2}$$
$$\geq\;1-\sqrt{1-\Big(1-\frac{1}{2}\Big[\exp\Big\{\text{const.}\cdot n m_n^{-1-\beta_2\mathfrak{a}}(\log m_n)^{-\mathfrak{a}\mathfrak{b}}\Big\}-1\Big]\Big)^2}\,,$$

analogously to (8.3). With $m_n\sim n^{1/(1+\beta_2\mathfrak{a})}(\log n)^{-\mathfrak{a}\mathfrak{b}/(1+\beta_2\mathfrak{a})}$ we can verify (8.4). □

Proof of Theorem 8.3. We follow the same strategy of proof as for Theorem 8.1. In that notation we assume that the observations at x_j of the regression curves $f_0=0$ and $f_1(x)=-dh_n^{\beta_2}K(x/h_n)$ are corrupted by noise with distribution functions $F_\varepsilon^{(0)}$ and $F_\varepsilon^{(1,f_1(x_j))}$, respectively, where for $y\leq 0$, $\delta\in(0,1)$

$$F_\varepsilon^{(0)}(y)=1-|y|^{\mathfrak{a}}1_{[-1,0]}(y)\text{ and }F_\varepsilon^{(1,\delta)}(y)=(1-|y-\delta|^{\mathfrak{a}})1_{[-1+\delta,-\delta]}(y)+(1-|2y|^{\mathfrak{a}})1_{(-\delta,0]}(y).$$

Then $F_\varepsilon^{(0)}$ is clearly in $\mathcal{D}_n(\mathfrak{a},\mathfrak{h}_0)$, but also $F_\varepsilon^{(1,f_1(x_j))} \in \mathcal{D}_n(\mathfrak{a},\mathfrak{h}_0)$ holds because the quantile function of $F_\varepsilon^{(1,\delta)}$ satisfies $\mathcal{U}^{(1,\delta)}(y) = -y^{-1/\mathfrak{a}} + \delta$ for $y \leq (2\delta)^{-\mathfrak{a}}$ and $\mathcal{U}^{(1,\delta)}(y) = -y^{-1/\mathfrak{a}}/2$ for $y > (2\delta)^{-\mathfrak{a}}$, but here

$$(2\delta)^{-\mathfrak{a}} \gtrsim \|f_1\|_\infty^{-\mathfrak{a}} \sim h_n^{-\mathfrak{a}\beta_2}$$

holds, which is of larger polynomial order in n than $n^{\mathfrak{h}_0}$ under the condition $\mathfrak{h}_0 < \mathfrak{a}\beta_2/(1+\mathfrak{a}\beta_2)$. Hence, the tail does not appear in condition (ii) for $\mathcal{D}_n(\mathfrak{a},\mathfrak{h}_0)$ and we have for $y \leq n^{\mathfrak{h}_0}$

$$\Big|\frac{\mathcal{U}^{(1,f_1(x_j))}(y)}{-y^{-1/\mathfrak{a}}} - 1\Big| = |f_1(x_j)||y|^{1/\mathfrak{a}} \lesssim h_n^{\beta_2} n^{\mathfrak{h}_0/\mathfrak{a}},$$

which tends to zero with polynomial speed in n.

In the proof of Theorem 8.1 the Lebesgue densities p_{f_0} and p_{f_1} are replaced by

$$\begin{aligned} p_{f_0}(y) &= \prod_{j=1}^n \mathfrak{a}|y_j|^{\mathfrak{a}-1}\mathbb{1}_{[-1,0]}(y_j) \\ p_{f_1}(y) &= \prod_{j=1}^n \Big(\mathfrak{a}|y_j|^{\mathfrak{a}-1}\mathbb{1}_{[-1,2f_1(x_j)]}(y_j) + \mathfrak{a}2^{\mathfrak{a}}|y_j|^{\mathfrak{a}-1}\mathbb{1}_{(2f_1(x_j),f_1(x_j)]}(y_j)\Big). \end{aligned}$$

The χ^2-distance therefore satisfies

$$\begin{aligned} \chi^2(p_{f_0},p_{f_1}) &= \exp\Big(\sum_{j=1}^n \int_{2f_1(x_j)}^{f_1(x_j)} \big(\mathfrak{a}(2^{\mathfrak{a}}-1)|y|^{\mathfrak{a}-1}\big)^2\big(\mathfrak{a}|y|^{\mathfrak{a}-1}\big)^{-1}\Big) - 1 \\ &= \exp\Big((2^{\mathfrak{a}}-1)^3\sum_{j=1}^n |f_1(x_j)|^{\mathfrak{a}}\Big) - 1 \\ &\lesssim n\int_0^1 |f_1(x)|^{\mathfrak{a}}dx \sim nh_n^{\mathfrak{a}\beta_2+1}. \end{aligned}$$

With the choice $h_n = (c_H n/\log n)^{-1/(\mathfrak{a}\beta_2+1)}$ the proof then continues exactly as for Theorem 8.1. □

9. Auxiliary lemmas. For the sake of reference, we will restate some results and notation in the following. Recall that our main assumption is

ASSUMPTION 9.1.

(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) $\max_{1\le j\le n}\mathbb{E}\big[|\varepsilon_j|\big]<\infty$,
(iii) (ε_j) *are independent and the distribution of* ε_j *satisfies* (9.1), (9.2).

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

$$\mathcal{U}_x(y)=-\mathfrak{c}_x y^{-1/\mathfrak{a}_x}l_x(y), \tag{9.1}$$

where $l_x(y)$ satisfies the condition

$$l_x(y)=\log(y)^{\mathfrak{b}_x}+\mathcal{O}\big(\log(y)^{\mathfrak{b}_x-1}\big)\quad \text{as } y\to\infty, \text{ uniformly for } x\in[0,1]. \tag{9.2}$$

Recall that

$$\|\tilde f_k-f\|\le R_k+B_k\,,\quad \forall k=0,\dots,K+1,\ f\in H_{\mathcal{N}(x)}(\beta,L)\,, \tag{9.3}$$

for some nonnegative random variables B_k,R_k, where B_k increases in k and R_k decreases in k.

PROPOSITION 9.1. *Let $\|\cdot\|$ denote some seminorm and let $\tilde f_k$, f lie in the corresponding normed space. Assume (9.3) and that the $\widehat{\mathfrak{z}}_k^T$ decrease a.s. in k. Defining the oracle-type index*

$$\widehat{k}^*:=\inf\big\{k=0,\dots,K-1:B_{k+1}>\widehat{\mathfrak{z}}_{k+1}^T/2\big\}\wedge K, \tag{9.4}$$

we obtain for $q\ge1$

$$\text{(a)}\quad \mathbb{E}_f\Big[\big\|\widehat f-\tilde f_{\widehat k^*}\big\|^q\mathbf{1}\big(\widehat k>\widehat k^*\big)\Big]^{1/q}\le \mathbb{E}_f\big[(\widehat{\mathfrak{z}}_{\widehat k^*}^T)^q\big]\,,$$

$$\text{(b)}\quad \mathbb{E}_f\Big[\big\|\widehat f-\tilde f_{\widehat k^*}\big\|^q\mathbf{1}\big(\widehat k<\widehat k^*\big)\Big]^{1/q}\le 2^{\frac{2q-1}{q}}\mathbb{E}_f[\widehat{\mathfrak{z}}_{\widehat k^*}^q]^{1/q}$$
$$+2^{\frac{2q-1}{q}}\sum_{k=0}^{K-1}\mathbb{E}_f\big[R_k^q\mathbf{1}\big(\exists l\le k\,:\,R_l>\widehat{\mathfrak{z}}_l^T/2\big)\big]^{1/q}.$$

THEOREM 9.1. *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 9.1 holds, then*

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

The concentration properties of the estimator will be derived from the following lemmas.

LEMMA 9.1. *If $y, t \to \infty$ and*

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

then

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

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

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

PROOF OF LEMMA 9.1:. Applying the logarithm and rearranging terms, it follows that

$$\log t = 1/\mathfrak{a}\big(\log y - \log \mathfrak{c} - \mathfrak{b} \log\log t\big) + \mathcal{o}(1), \tag{9.5}$$

hence we conclude $1/\mathfrak{a} \log y = \log t + \mathcal{o}(1)$. Plugging this into (9.5), we deduce that

$$\log t = 1/\mathfrak{a}\big(\log y - \log \mathfrak{c} - \mathfrak{b} \log(1/\mathfrak{a} \log y)\big) + \mathcal{o}(1),$$

hence the claim follows by applying the exponential function. The second follows by applying the first result to $-t^{-1}$. □

LEMMA 9.2. *For $1 \le j_0, j_1 \le n$, let $\mathcal{J} = \big\{j_0, ..., j_1\big\}$ such that $|j_0 - j_1|/n = \mathcal{O}\big(n^{-\rho_0}\big)$ for some $0 < \rho_0 < 1$. If $u \to 0$, $u \le -n^{-\rho_1}$ for some $\rho_1 > 0$, then*

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

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

PROOF OF LEMMA 9.2:. Since $\mathfrak{c}_x, \mathfrak{b}_x, \mathfrak{a}_x \in H_{[0,1]}(\beta_0, L)$ and $\inf_{x\in[0,1]} \mathfrak{c}_x > 0$, we conclude that

$$\max_{j\in\mathcal{J}}\left|\frac{\mathcal{U}_{x_j}(u)}{A_{x_{j_0}}(u)} - 1\right| = \mathcal{o}\big(1\big) \quad \text{as } u \to 0,\ u \le -n^{-\rho_1}, \tag{9.6}$$

see the proof of Lemma 10.4 for a related, more detailed argument. Let $c_3^+ > 1$ be arbitrarily close to one. Then for sufficiently small $u < 0$, we have from (9.6) that

$$\prod_{j\in\mathcal{J}} P\big(\varepsilon_j \le A_{x_{j_0}}(-u^{-1})\big) \le \prod_{j\in\mathcal{J}} P\big(\varepsilon_j \le \mathcal{U}_{x_j}(c_3^+ - u^{-1})\big) \le e^{\sum_{j\in\mathcal{J}}(c_3^+)^{-1}u}. \tag{9.7}$$

Hence the claim follows. □

We proceed by giving the proof of Lemma 9.3.

LEMMA 9.3. *Let $q \geq 1$. Under the Assumptions of Theorem 9.1, we have uniformly over $f \in H_{\mathcal{N}(x)}(\beta, L)$*

$$\mathbb{E}_f\big[(\widehat{\mathfrak{z}}^T_{\widehat{k}^*})^q\big] \;\leq\; (c_1^+)^q\,\mathbb{E}_f\big[(\mathfrak{z}^T_{\widehat{k}^*})^q\big] \;+\; \mathcal{O}\big(n^{-q/\mathfrak{a}_x}\big)\,,$$

where $c_1^+ > 1$.

PROOF OF LEMMA 9.3:. Proposition 10.1 and Lemma 10.1 (to deal with $\widehat{\mathfrak{a}}_x$) yield that for any $c_1^+ > 1$, we have uniformly over $f \in H_{\mathcal{N}(x)}(\beta, L)$

$$\begin{aligned}\mathbb{E}_f[(\widehat{\mathfrak{z}}^T_{\widehat{k}^*})^2] \;&\leq\; (c_1^+)^2\mathbb{E}_f[(\mathfrak{z}^T_{\widehat{k}^*})^q] + \sum_{k=0}^{K-1} P_f\big(\widehat{k}^* = k,\, \widehat{\mathfrak{z}}^T_k > c_1^+\mathfrak{z}^T_k\big)\\ &\leq\; (c_1^+)^2\,\mathbb{E}_f[(\mathfrak{z}^T_{\widehat{k}^*})^q] \;+\; \mathcal{O}\big(n^{-v}\log n\big)\,,\end{aligned}$$

since $K = \mathcal{O}\big(\log n\big)$. Choosing v sufficiently large, the claim follows. □

LEMMA 9.4. *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 cdf of ε. Then we have*

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

If $\mathcal{U}(.)$ is not differentiable, replace $\mathcal{U}(.)$ with $c_2^+A(.)$ 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

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

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

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

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

PROOF OF LEMMA 9.4:. For $0 < \delta_1 < \delta_2 < \infty$ we have

$$\begin{aligned}\mathbb{E}\big|\max\{\varepsilon_1,\dots,\varepsilon_n\}\big|^{q_n} \;&=\; \int_0^\infty P\big[\max\{\varepsilon_1,\dots,\varepsilon_n\} < -x^{1/q_n}\big]dx\\ &=\; \int_0^{\delta_1} F(-x^{1/q_n})^n dx + \int_{\delta_1}^{\delta_2} F(-x^{1/q_n})^n dx + \int_{\delta_2}^{\infty} F(-x^{1/q_n})^n dx. \qquad (9.8)\end{aligned}$$

Set $u_x = \mathcal{U}^{-1}\big(-x^{1/q_n}\big)$. Making the substitution $u_x^{-1} n = y$, the first integral in (9.8) is bounded from above as follows:

$$\begin{aligned}\int_0^{\delta_1} F(-x^{1/q_n})^n dx \;&=\; \int_0^{\delta_1} F\big(\mathcal{U}(u_x)\big)^n dx \;\le\; \int_0^{\delta_1} \exp\big(-n u_x^{-1}\big) dx \\ &\le \int_0^{n^{1/2}} n\, q_n y^{-2} \mathcal{U}^{(1)}\big(n/y\big)\big[-\mathcal{U}\big(n/y\big)\big]^{q_n - 1} \exp(-y) dy \\ &= \int_0^{n^{1/2}} \big((-\mathcal{U})^{q_n}(n/y)\big)^{(1)} \exp(-y) dy,\end{aligned}$$

for sufficiently small δ_1. If $\delta_1 \to 0$, we may replace $\mathcal{U}(.)$ with $A(.)$ in the above inequality. The above is then equal to

$$\big(1 + \mathcal{O}(1)\big) \int_0^{n^{1/2}} q_n \mathfrak{c}_F^{q_n} \big(\log(n/y)\big)^{\mathfrak{b}_F q_n - 1} \big(\mathfrak{a}_F^{-1} \log(n/y) - \mathfrak{b}_F\big) y^{-1 + q_n/\mathfrak{a}_F} n^{-q_n/\mathfrak{a}_F} \exp(-y) dy. \tag{9.9}$$

If $q_n = q$, routine computations and (9.9) lead to

$$\int_0^{\delta_1} F(-x^{1/q})^n dx = \mathcal{O}\big((\log n)^{q\mathfrak{b}_F} n^{-q/\mathfrak{a}_F}\big). \tag{9.10}$$

For arbitrary $1 \le q_n \le \log n$, we obtain, using the Gamma function and Stirling's formula, the (crude) upper and lower bounds

$$\mathcal{O}\big(n^{-c_2^+ \mathfrak{a}_F/q_n}\big) \le \int_0^{\delta_1} F(-x^{1/q_n})^n dx \le \mathcal{O}\big(n^{-c_2^- \mathfrak{a}_F/q_n}\big), \tag{9.11}$$

where $0 < c_2^- < 1 < c_2^+$ can be chosen arbitrarily close to one. For the second integral in (9.8) we obtain, proceeding as before, that

$$\int_{\delta_1}^{\delta_2} F(-x^{1/q_n})^n dx \;\le\; \delta_2 \cdot \exp\big(-n\delta_1^{c_2^- \mathfrak{a}_F/q_n}\big)\,,$$

where c_2^- is as above in (9.11). The third integral in (9.8) has the upper bound

$$\begin{aligned}\int_{\delta_2}^{\infty} F^n(-x^{1/q_n}) dx \;&=\; \int_{\delta_2}^{\infty} P(\varepsilon_1^p > x^{p/q_n})^n dx \;\le\; \int_{\delta_2}^{\infty} x^{-np/q_n} dx \cdot \{\mathbb{E}|\varepsilon_1|^p\}^n \\ &= \frac{q_n}{np - q_n} \delta_2^{1 - np/q_n} \cdot \{\mathbb{E}|\varepsilon_1|^p\}^n\,,\end{aligned}$$

by Markov's inequality for n sufficiently large. We put $\delta_2 = \exp\big(n \min\{1, \delta_1^{c_2^- \mathfrak{a}_F}\}/2\big)$. Then we may choose $\delta_1 \to 0$ such that the second and the third integral in (9.8) decay subgeometrically and are thus asymptotically negligible. □

LEMMA 9.5. *Suppose $h_k \leq \exp(-c_H \log^\gamma n)$ for fixed constants $\gamma \in (0,1)$, $c_H > 0$. Grant Assumption 9.1 and let $\mathfrak{m}, \mathfrak{a}_0, \beta_0, \mathfrak{h}_0$ satisfy (10.2) in view of (10.3). Then*

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

as $n \to \infty$.

PROOF OF LEMMA 9.5:. Proposition 10.1 and Lemma 10.1 (to deal with $\widehat{\mathfrak{a}}_F$) yield

$$\sup_{k=0,\ldots,K-1} P\big(R_k > \widehat{\mathfrak{z}}_k^T/2\big) \;\leq\; \sup_{k=0,\ldots,K-1} P\big(R_k^q > c_3^- \mathfrak{z}_k^q/2^q\big) \;+\; \mathcal{O}\Big(n^{-v}\log n\Big)\,,$$

where c_3^- is arbitrarily close to one. Choosing v sufficiently large, the term $P\big(R_k^q > c_3^- \mathfrak{z}_k^q/2^q\big)$ remains to be considered. For $I := \lceil 3J(\beta^*)/(2h_k)\rceil$ note

$$\|Z_j(h_k,\cdot)\|_q^q \;=\; \sum_{i=0}^{I-1}\int_{i/I}^{(i+1)/I} \big|\max\{\varepsilon_l : x_l \in x + h_k\mathcal{I}_j\}\big|^q dx\,.$$

We deduce that

$$K_{i,j}(h_k) := h_k\xi_j + \big[(i+1)/I, i/I + h_k/J(\beta^*)\big) \;\subseteq \bigcap_{x\in[i/I,(i+1)/I]} (x + h_k\mathcal{I}_j)\,,$$

where ξ_j denotes the left end point of the interval $\mathcal{I}_j$ so that

$$0 \;\geq\; \max\{\varepsilon_l : x_l \in x + h_k\mathcal{I}_j\} \;\geq\; \max\{\varepsilon_l : x_l \in K_{i,j}(h_k)\}$$

holds true for all $x \in [i/I,(i+1)/I]$. We conclude that

$$\|Z_j(h_k,\cdot)\|_q^q \;\leq\; \sum_{i=0}^{I-1}\int_{i/I}^{(i+1)/I} W_{i,j}(h_k)dx \;=\; \frac{1}{I}\sum_{i=0}^{I-1} W_{i,j}(h_k)\,, \tag{9.12}$$

where we introduce the random variables $W_{i,j}(h_k) := \big|\max\{\varepsilon_l : x_l \in K_{i,j}(h_k)\}\big|^q$, for all $i = 0,\ldots,I-1$. Let $n_k = nh_k$. From Lemma 9.4 we learn that

(9.13)

$$\mathbb{E}[W_{l,j}(h_k)] \leq \big(1 + o(1)\big) = c_2^+\big(1 + \mathcal{O}(1)\big)\int_{n^{-q/\mathfrak{a}_F}}^{n^{1/2}} \big|A_F^q\big(n_k/y\big)\big|^{(1)} \exp(-y)dy\,,$$

since $\int_0^{n^{-q/\mathfrak{a}_F}} F(-x^{1/q_n})dx = \mathcal{O}\big(n^{-q/\mathfrak{a}_F}\big)$. The length of each $K_{l,j}(h_k)$ has the lower bound $h_k/[3J(\beta^*)]$ and thus contains at least $\lfloor nh_k/[3J(\beta^*)]\rfloor$ of the x_l. We observe that for sufficiently large n

$$
\begin{aligned}
P\big(R_k^q > c_3^- \mathfrak{z}_k^q/2^q\big) \;&\le\; \sum_{j=1}^{2J(\beta*)} P\big(c(\beta^*)^q \|Z_j(h_k,\cdot)\|_q^q > c_3^- \mathfrak{z}_k^q/2^q\big) \\
&\le\; \sum_{j=1}^{2J(\beta*)} P\Big(\frac{1}{I}\sum_{i=0}^{I-1} W_{i,j}(h_k) > c(\beta^*)^{-q} c_3^- \mathfrak{z}_k^q/2^q\Big) \\
&\le\; \sum_{j=1}^{2J(\beta*)} P\Big(\frac{1}{I}\sum_{i=0}^{I-1} \tilde{W}_{i,j}(h_k) \ge c'\mathfrak{z}_k^q\Big)\,,
\end{aligned}
\tag{9.14}
$$

with centred random variables

$$
\tilde{W}_{l,j}(h_k) := W_{l,j}(h_k) - \mathbb{E}W_{l,j}(h_k)\,.
$$

and $c' > 0$. Note that by Lemma 9.4 we have $\mathfrak{z}_k^q \ge c'' n_k^{-qc_2^+/\mathfrak{a}_F}$ for some $c'' > 0$ and $c_2^+ > 1$ arbitrarily close to one. Markov's inequality with the even integer power $q_n \in [1, \log n]$ (to be specified) yields

$$
\begin{aligned}
&P\Big(\frac{1}{I}\sum_{i=0}^{I-1} \tilde{W}_{i,j}(h_k) \ge c''\cdot(nh_k)^{-qc_2^+/\mathfrak{a}_F}\Big) \;\le\; (c'')^{-q_n}(nh_k)^{qc_2^+ q_n/\mathfrak{a}_F}\cdot \mathbb{E}\Big|\frac{1}{I}\sum_{i=0}^{I-1}\tilde{W}_{i,j}(h_k)\Big|^{q_n} \\
&\;=\; (c'')^{-q_n}(nh_k)^{qc_2^+ q_n/\mathfrak{a}_F} I^{-q_n} \sum_{r\in\{0,\ldots,I-1\}^{(q_n)}} \mathbb{E}\prod_{i=1}^{q_n}\tilde{W}_{r_i,j}(h_k)
\end{aligned}
\tag{9.15}
$$

where r_i denotes the ith component of r. For all $r \in \{0,\ldots,I-1\}^{(q_n)}$ with at least one component r_i such that $|r_{i'} - r_i| \ge 2$ for all $i' \ne i$ the corresponding addend vanishes in (9.15). Therein note that $W_{i,j}(h_k)$ is independent of the σ-field generated by the $W_{i',j}(h_k)$, $|i'-i| \ge 2$, since all intervals $K_{i,j}(h_k)$ and $K_{i',j}(h_k)$ with $|i'-i| \ge 2$ are disjoint. Therefore the sum in (9.15) can equivalently be taken over all $r \in \{0,\ldots,I-1\}^{(q_n)}$ such that, for all i, there exists at least one $i' \ne i$ with $|r_i - r_{i'}| \le 1$. In order to provide an upper bound on the cardinality of that subset of $\{0,\ldots,I-1\}^{(q_n)}$ we consider the components of r with increasing order leading to $r_{(1)} \le \cdots \le r_{(q_n)}$. For at least $q_n/2$ of the $i = 2,\ldots,q_n$ we have $r_{(i)} \le r_{(i-1)} + 1$. Thus the summation is restricted to the set

$$
\mathcal{R}_n \;:=\; \bigcup_{K\subseteq\{2,\ldots,q_n\},\#K\ge q_n/2} \big\{r \in \{0,\ldots,I-1\}^{(q_n)} : r_{(i)} \le r_{(i-1)} + 1, \forall i \in K\big\}\,,
$$

whose cardinality is bounded from above by $q_n!\,2^{3q_n/2}\,I^{q_n/2}$. Those combinational arguments, Hölder's inequality and the assumption that $\mathfrak{h}_0 > 0$ yield that for c_2^-, c_2^+ sufficiently close to one

$$P\Big(\frac{1}{I}\sum_{i=0}^{I-1}\tilde{W}_{i,j}(h_k) \geq c'' \cdot (nh_k)^{-qc_2^+/\mathfrak{a}_F}\Big)$$

$$\leq\; q_n! \cdot (c'')^{-q_n}(nh_k)^{qc_2^+ q_n/\mathfrak{a}_F} I^{-q_n/2} 2^{3q_n/2} \sup_{r\in\mathcal{R}_n}\prod_{i=1}^{q_n}\big\{\mathbb{E}|\tilde{W}_{r_i,j}(h_k)|^{q_n}\big\}^{1/q_n}$$

$$\leq\; (D/c'')^{q_n} \cdot q_n^{(1+qc_2^+/\mathfrak{a}_F)q_n} \cdot h_k^{q_n/2^q}\,, \tag{9.16}$$

for some finite constant $D > 1$ when using Lemma 9.4 and again Stirling's formula in order to bound $q_n!$. We use $q_n = \lfloor \log n \rfloor$ and obtain

$$P\Big(\frac{1}{I}\sum_{i=0}^{I-1}\tilde{W}_{i,j}(h_k) \geq c'' \cdot (nh_k)^{-qc_2^+/\mathfrak{a}_F}\Big)$$

$$\leq\; \exp\big\{\lfloor \log n\rfloor \cdot \big(\log(D/c'') + (1+q/\mathfrak{a}_F)\log(\lfloor\log n\rfloor) + \log(h_k)/q\big)\big\}$$

$$\leq \exp\big\{-(c_H/2q)\cdot(\log n)^{1+\gamma}\big\}\,,$$

for n sufficiently large, using the upper bound imposed on h_k. □

10. Proofs for the Hill-type estimator.

$$h_0^{\beta_x}\big|A_x\big(m_0\big)\big|^{-1} = \mathcal{O}\big((\log n)^{-1}\big). \tag{10.1}$$

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

$$\mathfrak{m}\mathfrak{h}_0 < (1-\mathfrak{h}_0)\mathfrak{a}_0\beta_0, \tag{10.2}$$

for some lower bounds

$$\mathfrak{a}_0 \leq \mathfrak{a}_x \text{ and } \beta_0 \leq \beta_x \tag{10.3}$$

on the unknown parameters, then (10.1) is valid. In Section 10 below we prove the following result, under the general Assumption 10.1, which is implied by Assumption 9.1.

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

We will show the validity of Proposition 10.1, working in a general framework. For each $N \in \mathbb{N}$ and $0 < \mathfrak{h}_0 < 1$, let $x_j = j/N^{1/\mathfrak{h}_0}$ and assume that (ε_j) has distribution function $F_{x_j}^{(N)}(.)$, with quantile function $\mathcal{U}_{x_j}^{(N)}(.)$ such that $F_{x_j}^{(N)}\big(\mathcal{U}_{x_j}^{(N)}(y)\big) = 1 - 1/y$. One may think of $N^{1/\mathfrak{h}_0} = n$.

ASSUMPTION 10.1. *There exist* $\mathfrak{a}_x, \mathfrak{c}_x > 0$, $\mathfrak{b}_x \in \mathbb{R}$, *such that,*

$$(i) \quad \limsup_{y\to\infty} \sup_{x\in[0,1]} \left|\frac{\mathcal{U}_x^{(N)}(y)}{A_x(y/2)}\right| \leq 1, \quad \text{where } A_x(y) = -\mathfrak{c}_x \log(y)^{\mathfrak{b}_x} y^{-1/\mathfrak{a}_x},$$

$$(ii) \quad \limsup_{N\to\infty} \sup_{y\in[\log N, N]} \sup_{x\in[0,1]} \left|\frac{\mathcal{U}_x^{(N)}(y)}{A_x(y)} - 1\right| \log y = 0,$$

$$(iii) \quad \mathfrak{c}_x, \mathfrak{b}_x, \mathfrak{a}_x \in H_{[0,1]}(\beta_0, L), \text{ where } \beta_0 > 0 \text{ and } \inf_{x\in[0,1]} \mathfrak{a}_x, \mathfrak{c}_x > 0.$$

To lighten the notation, we use the abbreviation $\mathcal{U}_x(y) = \mathcal{U}_x^{(N)}(y)$ in the sequel. Note that our Assumptions in 10.1 include cases where a CLT for an estimator $\widehat{\mathfrak{a}}_x$ fails to hold, and only slower rates of convergence than $N^{-1/2}$ are available. This is particularly the case if $\mathfrak{b}_x \neq 0$, we refer to de Haan and Ferreira [2] for details.

We assume that we observe $(Y_j)_{1\leq j\leq N}$, specified generally by

$$Y_j = \theta_j + \varepsilon_j, \quad \theta_j \in \mathbb{R},\, j = 1, ..., N.$$

We introduce the maximal signal variation

$$\mathbf{\Theta}_N = \max_{1\leq i,j\leq N} \big|\theta_i - \theta_j\big|.$$

Recall that $Y_{1,N}, ..., Y_{N,N}$ denotes the decreasing order statistics of $(Y_j)_{1\leq j\leq N}$. Let $E_1, ..., E_N$ be a sequence of i.i.d. unit exponential random variables. Then $E_j \stackrel{d}{=} -\log\big(1 - F_{x_j}(\varepsilon_j)\big)$ holds and we may generate ε_j as

$$\varepsilon_j = F_{x_j}^{-1}\big(1 - \exp(-E_j)\big). \tag{10.4}$$

Given $x \in [0,1]$, we consider the base estimators

$$\frac{1}{\widehat{\mathfrak{a}}_x^*(m)} = \frac{1}{m}\sum_{i=2}^{m-1} \log\left(\frac{Y_{m,N} - Y_{1,N}}{Y_{i,N} - Y_{1,N}}\right), \tag{10.5}$$

and

$$\widehat{\mathfrak{b}}_x^*(m) = \frac{1}{m\log\log N}\sum_{i=2}^{m-1} \log\left(\frac{Y_{i,N} - Y_{1,N}}{(N/i)^{-1/\widehat{\mathfrak{a}}_x^*} - (N/1)^{-1/\widehat{\mathfrak{a}}_x^*}}\right). \tag{10.6}$$

Let $\Delta_{x_i,x_j} = L|x_i - x_j|^{\beta_0}$, where we suppress the indices x_i, x_j if the correspondence is clear. For notational reasons, we write $x = x_{k^*} = k^*/N$ for some $k^* \in [0, N]$ in the sequel. We then require the following conditions.

ASSUMPTION 10.2. *Let $m/N \to 0$, and put $\mathcal{I} = \big\{l, ..., N\big\} \cup \{1, k^*\}$. We require that*

(i) $\sup_{i,j\in\mathcal{I}} \Delta_{x_i,x_j} = \mathcal{O}\big((\log N)^{-2}\big)$,
(ii) $y_i = N/i\big(1 + \mathcal{O}\big(\log(N/i)^{-1}\big)\big)$ *for* $i \in \mathcal{I}$,
(iii) $B_N := \sup_{i,j\in\mathcal{I}}\big|\Theta_N A_{x_i}^{-1}(y_j)\big| = \mathcal{O}\big((\log N)^{-1}\big)$.

We then have the following two Lemmas.

LEMMA 10.1. *Suppose that $m = N^{\mathfrak{m}}$ with $0 < \mathfrak{m} < 1$. If Assumptions 10.1 and 10.2 hold, then*

$$P\left(\max_{i\in\mathcal{I}}\left|\frac{1}{\widehat{\mathfrak{a}}_x^*} - \frac{1}{\mathfrak{a}_{x_i}} - \frac{\mathfrak{b}_{x_i}}{\log N - \log m}\right| \geq \mathcal{O}\big((\log N)^{-1}\big)\right) = N\mathcal{O}\big(\exp(-(\log N)^4)\big).$$

LEMMA 10.2. *Suppose that $m = N^{\mathfrak{m}}$ with $0 < \mathfrak{m} < 1$. If Assumptions 10.1 and 10.2 hold, then*

$$P\left(\max_{i\in\mathcal{I}}\left|\widehat{\mathfrak{b}}_x^* - \mathfrak{b}_{x_i} - \frac{\log \mathfrak{c}_{x_i}}{\log\log N} - \frac{\mathfrak{b}_{x_i}}{\log\log N}\frac{\log m + \log N}{\log N - \log m}\right| \geq \mathcal{O}\big((\log\log N)^{-1}\big)\right)$$
$$= N\mathcal{O}\big(\exp(-(\log N)^4)\big).$$

Based on the above results, we introduce the main estimators, constructed via the slight bias correction

$$\widehat{\mathfrak{b}}_x = \widehat{\mathfrak{b}}_x^* - \frac{\widehat{\mathfrak{b}}_x^*}{\log\log N}\frac{\log m + \log N}{\log N - \log m}, \tag{10.7}$$

$$\frac{1}{\widehat{\mathfrak{a}}_x} = \frac{1}{\widehat{\mathfrak{a}}_x^*} - \frac{\widehat{\mathfrak{b}}_x}{\log N - \log m}. \tag{10.8}$$

We are now ready to derive Proposition 10.1.

PROOF OF PROPOSITION 10.1. Set $N = \bar{n}_0(x)$ and $m = m_0 = \bar{n}_0(x)^{\mathfrak{m}}$. Note that since $\mathfrak{h}_0, \mathfrak{m} \in (0,1)$, it follows from (10.1) that Assumption 10.2 is valid. We thus conclude from Lemma 10.1 and (10.7) that there exists a set $\mathcal{A}$, such that $P\big(\mathcal{A}^c\big) = N\mathcal{O}\big(\exp(-(\log N)^4)\big)$, and

$$\big|\widehat{\mathfrak{a}}_x\big(m_0\big) - \mathfrak{a}_x\big|\mathbf{1}(\mathcal{A}) = \mathcal{O}\big((\log N)^{-1}\big).$$

Since $q > 1$, it follows by construction that

$$\big|\widehat{\mathfrak{a}}_x - \mathfrak{a}_x\big|\mathbf{1}(\mathcal{A}) = \scriptstyle{\mathcal{O}}\displaystyle\big((\log N)^{-1}\big). \tag{10.9}$$

Similarly, it follows from Lemma 10.2 and (10.7) that there exists a set $\mathcal{B}$, such that $P\big(\mathcal{B}^c\big) = N\mathcal{O}\big(\exp(-(\log N)^4)\big)$, and

$$\big|\widehat{\mathfrak{b}}_x - \mathfrak{b}_x - \log \mathfrak{c}_x/\log\log(nh_{\widehat{k}})\big|\mathbf{1}(\mathcal{B}) = \scriptstyle{\mathcal{O}}\displaystyle\big((\log\log n)^{-1}\big),$$

where $\widehat{k}$ is a random variable, taking values $\widehat{k} = 0, ..., K-1$. Note that $n^{(\log n)^{-1}} \to e$ and $\frac{\log\log n}{\log\log nh_k} \to 1$ as $n \to \infty$, uniformly in $k = 0, ..., K-1$. Hence we conclude that on the set $\mathcal{A} \cap \mathcal{B}$, we have that

$$(nh_k)^{\widehat{\mathfrak{a}}_x} = (nh_k)^{\mathfrak{a}_x}\big(1 + \scriptstyle{\mathcal{O}}\displaystyle(1)\big) \quad \text{and} \quad (\log nh_k)^{\widehat{\mathfrak{b}}_x} = \mathfrak{c}_x(\log nh_k)^{\mathfrak{b}_x}\big(1 + \scriptstyle{\mathcal{O}}\displaystyle(1)\big). \tag{10.10}$$

Moreover, the functions $x^x, (\log x)^x$ are all differentiable for $x \neq 0$ with locally bounded derivatives, hence they are locally Lipschitz-continuous by a Taylor expansion. Together with (10.10), this yields the claim for $T = 1$, since $\big\{\widehat{\mathfrak{z}}_k \geq 1\big\} \subseteq \big\{\mathcal{A}^c \cup \mathcal{B}^c\big\}$. In the same manner, one shows that the claim is valid if T is Lipschitz-continuous. ☐

For the proofs of Lemmas 10.1 and 10.2, we require some auxiliary results. We will frequently use the following simple bound. Let $a, b, \Delta_1, \Delta_2 \in \mathbb{R}$. Then, if the expressions are well defined, we have

$$\left|\log\left(\frac{a + \Delta_1}{b + \Delta_2}\right) - \log\left(\frac{a}{b}\right)\right| = \mathcal{O}\big(|\Delta_1/a| + \big|\Delta_2/b\big|\big), \tag{10.11}$$

provided that $|\Delta_1/a| + \big|\Delta_2/b\big| = \scriptstyle{\mathcal{O}}\displaystyle(1)$.

Lemma 10.3. *Suppose that $m = N^{\mathfrak{m}}$, with $0 < \mathfrak{m} < 1$, and $\mathfrak{a}, \mathfrak{c} > 0$, $\mathfrak{b} \in \mathbb{R}$. Then*

$$\frac{1}{m}\sum_{i=2}^{m-1} \log\left(\frac{\mathfrak{c}(N/m)^{-1/\mathfrak{a}}(\log N/m)^{\mathfrak{b}} - \mathfrak{c}(N)^{-1/\mathfrak{a}}(\log N)^{\mathfrak{b}}}{\mathfrak{c}(N/i)^{-1/\mathfrak{a}}(\log N/i)^{\mathfrak{b}} - \mathfrak{c}(N)^{-1/\mathfrak{a}}(\log N)^{\mathfrak{b}}}\right) = \frac{1}{\mathfrak{a}} + \frac{\mathfrak{b}}{\log N - \log m} + \scriptstyle{\mathcal{O}}\displaystyle\big(\log(m)^{-1}\big).$$

Proof of Lemma 10.3. We have that

$$\begin{aligned}
&\frac{1}{m}\sum_{i=2}^{m-1} \log\left(\frac{\mathfrak{c}(N/m)^{-1/\mathfrak{a}}(\log N/m)^{\mathfrak{b}} - \mathfrak{c}(N)^{-1/\mathfrak{a}}(\log N)^{\mathfrak{b}}}{\mathfrak{c}(N/i)^{-1/\mathfrak{a}}(\log N/i)^{\mathfrak{b}} - \mathfrak{c}(N)^{-1/\mathfrak{a}}(\log N)^{\mathfrak{b}}}\right) \\
&= \frac{1}{m}\sum_{i=2}^{m-1} \log\left(\frac{m^{1/\mathfrak{a}}(\log N/m)^{\mathfrak{b}} - (\log N)^{\mathfrak{b}}}{i^{1/\mathfrak{a}}(\log N/i)^{\mathfrak{b}} - (\log N)^{\mathfrak{b}}}\right) =: \frac{1}{m}\sum_{i=2}^{m-1} I_i.
\end{aligned}$$

Let $m^* = m/(\log m)^3$. Then

$$\frac{1}{m}\sum_{i=2}^{m^*-1} I_i = \scriptstyle\mathcal{O}\displaystyle\big((\log m)^{-1}\big).$$

On the other hand, using (10.11) it follows that

$$\frac{1}{m}\sum_{i=m^*}^{m-1} I_i = \frac{1}{m}\sum_{i=m^*}^{m-1}\log\bigg(\frac{m^{1/\mathfrak{a}}(\log N/m)^{\mathfrak{b}}}{i^{1/\mathfrak{a}}(\log N/i)^{\mathfrak{b}}}\bigg) + \scriptstyle\mathcal{O}\displaystyle\big((\log m)^{-1}\big) \tag{10.12}$$

$$= \frac{1}{\mathfrak{a}}\frac{1}{m}\sum_{i=m^*}^{m-1}\log\big(m/i\big) + \frac{\mathfrak{b}}{m}\sum_{i=m^*}^{m-1}\log\bigg(\frac{(\log N/m)}{(\log N/i)}\bigg) + \scriptstyle\mathcal{O}\displaystyle\big((\log m)^{-1}\big). \tag{10.13}$$

By Stirling's formula, it follows that

$$\frac{1}{m}\sum_{i=1}^{m-1}\log\big(m/i\big) = \frac{1}{m}\log\bigg(\frac{(m)^{m-1}}{(m-1)!}\bigg) = \frac{1}{m}\Big(m\log m - m\log m + m + \mathcal{O}(\log m)\Big)$$

$$= 1 + \mathcal{O}\big(\log(m)/m\big). \tag{10.14}$$

Since $m^* = m/(\log m)^3$, we deduce that

$$\frac{1}{\mathfrak{a}}\frac{1}{m}\sum_{i=m^*}^{m-1}\log\big(m/i\big) = \frac{1}{\mathfrak{a}} + \scriptstyle\mathcal{O}\displaystyle\big(\log(m)^{-1}\big). \tag{10.15}$$

For the second quantity, note that

$$\frac{\mathfrak{b}}{m}\sum_{i=m^*}^{m-1}\log\bigg(\frac{(\log N/m)}{(\log N/i)}\bigg) = \frac{\mathfrak{b}}{m}\sum_{i=m^*}^{m-1}\log\bigg(\frac{\log N - \log m}{\log N - \log m - \log i/m}\bigg) =: II.$$

Let $B_N = \log N - \log m$, and note that for $m^* \leq i \leq m$ we have $(\log i/m)/B_N = \scriptstyle\mathcal{O}\displaystyle(1)$. Using the series expansion of $\log(1+x)$, we thus obtain

$$\begin{aligned}\log\bigg(\frac{B_N}{B_N - \log i/m}\bigg) &= \log\bigg(\frac{(\log i/m)/B_N}{1-(\log i/m)/B_N} + 1\bigg)\\ &= \frac{(\log i/m)/B_N}{1-(\log i/m)/B_N} + \scriptstyle\mathcal{O}\displaystyle\big((\log i/m)/B_N\big)\\ &= (\log i/m)/B_N + \scriptstyle\mathcal{O}\displaystyle\big((\log i/m)/B_N\big).\end{aligned}$$

Then using the above and (10.14), it follows that

$$II = \frac{\mathfrak{b}}{m}\sum_{i=m^*}^{m-1}\Big((\log i/m)/B_N + \scriptstyle\mathcal{O}\displaystyle\big((\log i/m)/B_N\big)\Big) = \frac{\mathfrak{b}}{B_N} + \scriptstyle\mathcal{O}\displaystyle\big(\log(m)^{-1}\big). \tag{10.16}$$

Piecing everything together, the claim follows. □

For the next result, let $\mathcal{P}(\mathcal{I})$ be the set of all permutations of $\mathcal{I}$.

LEMMA 10.4. *Suppose that $m = N^{\mathfrak{m}}$, with $0 < \mathfrak{m} < 1$. If Assumption 10.2 holds, then*

$$\sup_{\tau\in\mathcal{P}(\mathcal{I})}\left|\frac{1}{m}\sum_{i=2}^{m-1}\log\left(\frac{\mathcal{U}_{x_{\tau(m)}}(y_m)-\mathcal{U}_{x_{\tau(1)}}(y_1)+\mathcal{O}(\Theta_N)}{\mathcal{U}_{x_{\tau(i)}}(y_i)-\mathcal{U}_{x_{\tau(1)}}(y_1)+\mathcal{O}(\Theta_N)}\right)-\frac{1}{\mathfrak{a}_x}\right| = \scriptstyle{\mathcal{O}}\big((\log N)^{-1}\big).$$

PROOF OF LEMMA 10.4. For $y \geq 1$, we have

$$\max\Big\{\big|(\log y)^{\Delta}-1\big|, \big|y^{\Delta}-1\big|\Big\} = \mathcal{O}\big(\Delta\log y\big) \quad \text{if } \Delta\log y = \scriptstyle{\mathcal{O}}(1). \tag{10.17}$$

Since $\mathfrak{c}_x, \mathfrak{b}_x, \mathfrak{a}_x \in H_{[0,1]}(\beta_0, L)$ and $\inf_{x\in[0,1]}\mathfrak{c}_x > 0$, we conclude that

$$\left|\frac{A_{x_{\tau(i)}}(y)}{A_{x_{\tau(j)}}(y)}-1\right| = \mathcal{O}\big(\Delta\log y\big) \quad \text{if } \Delta\log y = \scriptstyle{\mathcal{O}}(1). \tag{10.18}$$

Consequently, we obtain that

$$\mathcal{U}_{x_{\tau(i)}}(y) = \frac{\mathcal{U}_{x_{\tau(i)}}(y)}{A_{x_{\tau(i)}}(y)}\frac{A_{x_{\tau(i)}}(y)}{A_{x_{\tau(j)}}(y)}\frac{A_{x_{\tau(j)}}(y)}{\mathcal{U}_{x_{\tau(j)}}(y)}\mathcal{U}_{x_{\tau(j)}}(y) = \mathcal{U}_{x_{\tau(j)}}(y)\Big(1+\scriptstyle{\mathcal{O}}\big((\log N)^{-1}\big)\Big). \tag{10.19}$$

Moreover, we also have from the assumptions and (10.18) that

$$\frac{\mathcal{U}_{x_{\tau(i)}}(y)}{A_{x_{\tau(j)}}(y)} = \frac{\mathcal{U}_{x_{\tau(i)}}(y)}{A_{x_{\tau(i)}}(y)}\frac{A_{x_{\tau(i)}}(y)}{A_{x_{\tau(j)}}(y)} = 1+\scriptstyle{\mathcal{O}}\big((\log N)^{-1}\big). \tag{10.20}$$

Let $\Delta(y_i) = \scriptstyle{\mathcal{O}}\big(\log(N)^{-1}\big)A_{x_m}(y_i)$. Then due to (10.19) and (10.20), we obtain that

$$\begin{aligned} I_m &:= \frac{1}{m}\sum_{i=2}^{m-1}\log\left(\frac{\mathcal{U}_{x_{\tau(m)}}(y_m)-\mathcal{U}_{x_{\tau(1)}}(y_1)+\mathcal{O}(\Theta_N)}{\mathcal{U}_{x_{\tau(i)}}(y_i)-\mathcal{U}_{x_{\tau(1)}}(y_1)+\mathcal{O}(\Theta_N)}\right) \\ &= \frac{1}{m}\sum_{i=2}^{m-1}\log\left(\frac{\mathcal{U}_{x_m}(y_m)-\mathcal{U}_{x_m}(y_1)+\mathcal{O}\big(\Delta(y_m)+\Theta_N\big)}{\mathcal{U}_{x_m}(y_i)-\mathcal{U}_{x_m}(y_1)+\mathcal{O}\big(\Delta(y_i)+\Theta_N\big)}\right). \end{aligned} \tag{10.21}$$

Let $a = \mathcal{U}_{x_m}(y_m)-\mathcal{U}_{x_m}(y_1)$ and $b = \mathcal{U}_{x_m}(y_i)-\mathcal{U}_{x_m}(y_1)$. Then an application of (10.11) in connection with (10.19) and (10.20) yields that

$$\left|\log\left(\frac{a+\mathcal{O}\big(\Delta(y_{2m})+\Theta_N\big)}{b+\mathcal{O}\big(\Delta(y_{2m})+\Theta_N\big)}\right)-\log\left(\frac{a}{b}\right)\right| = \scriptstyle{\mathcal{O}}\Big((\log N)^{-1}+(\log y_{2m})^{-1}\Big). \tag{10.22}$$

A further application of (10.11) together with Assumption 10.1 gives

(10.23)
$$\log\left(\frac{a}{b}\right) = \log\left(\frac{\mathfrak{c}_x(N/m)^{-1/\mathfrak{a}_x}(\log N/m)^{\mathfrak{b}_x} - \mathfrak{c}_x(N)^{-1/\mathfrak{a}_x}(\log N)^{\mathfrak{b}_x}}{\mathfrak{c}_x(N/i)^{-1/\mathfrak{a}_x}(\log N/i)^{\mathfrak{b}_x} - \mathfrak{c}_x(N)^{-1/\mathfrak{a}_x}(\log N)^{\mathfrak{b}_x}}\right) + \mathcal{O}\big((\log N)^{-1}\big).$$

The claim now follows from Lemma 10.3. □

Lemma 10.5. *Let $m = N^{\mathfrak{m}}$, $0 < \mathfrak{m} < 1$, and put $\mathcal{M} = \{l, ..., m\}$, $1 \leq l \leq m$. Then*

$$P\big(\max_{i\in\mathcal{M}}\big|E_{i,N} - \log(N/i)\big| \geq z/\sqrt{i}\big) = m\mathcal{O}\big(\exp(-z^2/2)\big)$$

for $0 \leq z \leq l$. Moreover, for $0 < \delta < 1$, it holds that

$$P\big(E_{1,N} \geq \delta \log N\big) = 1 - \mathcal{O}\Big(\exp(-N^{1-\delta})\Big), \quad P\big(E_{1,N} \leq N^{\delta}\big) = 1 - \mathcal{O}\Big(N\exp(-N^{\delta})\Big).$$

Proof of Lemma 10.5. For $M \leq N-1$, Renyi's representation (cf. Feller [3], p. 19) gives the relation

$$E_{M+1,N} \stackrel{d}{=} \sum_{j=1}^{N-M} \frac{1}{N-j+1}E_j.$$

For $|\tau| \leq M$, one thus readily obtains that

$$\begin{aligned}\mathbb{E}\big[\exp(\tau E_{M+1,N})\big] &= \prod_{j=1}^{N-M}\big(1-\tau/(N+1-j)\big)^{-1} = \prod_{j=1}^{N-M}\exp\big(-\log(1-\tau/(N+1-j))\big)\\ &= \exp\Big(-\sum_{j=1}^{N-M}\log\big(1-\tau/(N+1-j)\big)\Big) \leq c_1\exp\Big(\sum_{j=1}^{N-M}\tau/(N+1-j)\Big)\\ &\leq c_2\exp\big(\tau\log(N/M)\big) = c_2(N/M)^{\tau}.\end{aligned}$$

Put $\mathcal{L}_N = \sqrt{M}\big(-E_{M,N} + \log(N/M)\big)$ and consider $|\tau| \leq \sqrt{M}$. Arguing similarly as above, we derive

$$\mathbb{E}\big[\exp(\tau\mathcal{L}_N)\big] = \exp\Big(-\sum_{j=1}^{N-M}\log(1-\tau\sqrt{M}/(N+1-j)) - \tau\sqrt{M}\log(N/M)\Big)$$

$$\leq c_3\exp\Big(1/2\sum_{j=1}^{N-M}\tau^2 M/(N+1-j)^2\Big) \leq c_4\exp(\tau^2/2). \tag{10.24}$$

Hence, an application of Markov's inequality provides

$$P\big(z\mathcal{L}_N \geq z^2\big) \leq \exp(-z^2)\mathbb{E}\Big[\exp(z\mathcal{L}_N)\Big] = \mathcal{O}\big(\exp(-z^2/2)\big), \quad |z| \leq \sqrt{M}.$$

Similarly, one obtains $P\big(-\mathcal{L}_N \geq z\big) = \mathcal{O}\big(\exp(-z^2/2)\big)$, hence

$$P\big(|\mathcal{L}_N| \geq z\big) = \mathcal{O}\big(\exp(-z^2/2)\big). \tag{10.25}$$

We now introduce $\mathcal{L}_N^{(i)} = \sqrt{i}\big(-E_{i,N} + \log(N/i)\big)$, for $i \in \mathcal{M}$. Define the event $\mathcal{A} := \bigcap_{i\in\mathcal{M}}\big\{|\mathcal{L}_N^{(i)}| \leq z\big\}$. By (10.25), we have that the complement is bounded by

$$P\big(\mathcal{A}^c\big) \leq \sum_{i\in\mathcal{M}} P\big(|\mathcal{L}_N^{(i)}| > z\big) = m\mathcal{O}\big(\exp(-z^2/2)\big),$$

which proofs the first part. The second claim follows via straightforward computations. □

Proof of Lemma 10.1. We have the decomposition

$$\sum_{i=2}^{m-1} \log\left(\frac{Y_{m,N} - Y_{1,N}}{Y_{i,N} - Y_{1,N}}\right) = \sum_{i=l}^{m-1} \log\left(\frac{Y_{m,N} - Y_{1,N}}{Y_{i,N} - Y_{1,N}}\right) + \sum_{i=2}^{l-1} \log\left(\frac{Y_{m,N} - Y_{1,N}}{Y_{i,N} - Y_{1,N}}\right)$$
$$=: I_m + II_m.$$

We first deal with the second part II_m. First note the trivial fact that when computing $Y_{i,N} - Y_{j,N}$ for $i,j = 1,...,N$, $\Theta_{i,j} = \theta_{i^*} - \theta_{j^*}$ may take $N(N+1)/2$ different values, where i^*, j^* may be different from i,j. In addition, note that

$$\big|II_m\big| \leq \sum_{i=2}^{l-1}\left|\log\left(\frac{Y_{m,N} - Y_{1,N}}{Y_{2,N} - Y_{1,N}}\right)\right|.$$

Let $0 < \delta$, $l = 2(\log N)^4$ and $\mathcal{M} = \big\{l, .., m\big\}$. Put

$$\mathcal{A} = \big\{E_{1,N} \leq N^\delta\big\} \cap \big\{\max_{i\in\mathcal{M}}\big|E_{i,N} - \log(N/i)\big| \leq z/\sqrt{i}\big\}. \tag{10.26}$$

By Lemma 10.5 we have $P\big(\mathcal{A}^c\big) = m\mathcal{O}\big(e^{-l^2/2}\big)$. Lemma 10.5 and the representation given in (10.4) yield that

$$P\Big(\big\{|Y_{2,N} - Y_{1,N}| \leq z_N\big\} \cap \mathcal{A}\Big)$$
$$\leq \sum_{i,j=1}^{m} P\Big(\big\{\big|\mathcal{U}_{x_1}\big(e^{E_{i,N}}\big) - \mathcal{U}_{x_2}\big(e^{E_{j,N}}\big) + \Theta_{i,j}\big| \leq z_N\big\} \cap \mathcal{A}\Big) =: III_m,$$

where $z_N = e^{-N^\delta}$, $x_1, x_2 \in [0,1]$ are random variables. Assume without loss of generality that $i \leq j$, and let E_i^* be independent copies of E_i. Then

$$\mathcal{U}_{x_1}\big(e^{E_{j,N}}\big) \stackrel{d}{=} \mathcal{U}_{x_1}\big(e^{E_{1,j-i}^* + E_{i,N}}\big). \tag{10.27}$$

We have that

$$\begin{aligned}
&\Big\{\big|\mathcal{U}_{x_1}\big(e^{E_{i,N}}\big) - \mathcal{U}_{x_2}\big(e^{E_{j,N}}\big) + \Theta_{i,j}\big| \leq z_N\Big\} \cap \mathcal{A} \\
&= \Big\{E_{1,j-i}^* \in -E_{i,N} + \log\Big(\mathcal{U}_{x_1}^{-1}\big(\Theta_{i,j} - \mathcal{U}_{x_2}\big(e^{E_{j,N}}\big) + \big[-z_N, z_N\big]\big)\Big)\Big\} \cap \mathcal{A} \\
&\subseteq \Big\{E_{1,j-i}^* \in -E_{i,N} + RI_N\Big\},
\end{aligned} \tag{10.28}$$

where RI_N is a random interval, independent of $E_{1,j-i}^*$. Due to the event $\mathcal{A}$, Assumption 10.1 and Assumption 10.2 (i), a Taylor expansion of $\log(1-x)$, $|x| < 1$ yields that RI_N can be chosen such that the Lebesgue measure $\lambda\big(RI_N\big) = \mathcal{O}\big(z_N^\chi\big)$ for some $\chi > 0$. Denote with $d_X(x)$ the density function of a random variable X. Since $E_{1,j-i}^*$ and $E_{i,N}$ are independent, we have by the translation invariance of the Lebesgue measure that

$$\begin{aligned}
P\big(\big\{E_{1,j-i}^* \in -E_{i,N} + RI_N\big\}\big) &\leq \int_{\mathbb{R}} d_{E_{i,N}}(y) \int_{RI_N} \sup_{x\in\mathbb{R}} d_{E_{1,j-i}^*}(x)\,dx\,dy \\
&\leq \int_{\mathbb{R}} d_{E_{i,N}}(y) \int_{RI_N} \mathcal{O}\big(N\big)\,dx\,dy = \mathcal{O}\big(N z_N^\chi\big),
\end{aligned} \tag{10.29}$$

where we used that $\sup_{x\in\mathbb{R}} d_{E_{1,j-i}^*}(x) = \mathcal{O}(N)$. We thus obtain that

$$III_{m^*} \leq \sum_{i,j=1}^{m^*} = \mathcal{O}\big(N^2 z_N^\chi\big).$$

Choosing δ such that $0 < \delta\chi < \mathfrak{m}$, we conclude that on the set $\mathcal{A}$ we have

$$II_m \leq \mathcal{O}(1) \sum_{i=2}^{l-1} \log\Big(e^{N^{\delta\chi}}\Big) = \mathcal{O}\big(l\, N^{\delta\chi}\big) = o\big(m(\log N)^{-1}\big), \tag{10.30}$$

and it remains to deal with I_m. Using the representation given in (10.4), we have the relation

$$I_m \stackrel{d}{=} \frac{1}{m} \sum_{i=l}^{m-1} \log\left(\frac{\mathcal{U}_{x_m}\big(e^{E_{m,N}}\big) - \mathcal{U}_{x_1}\big(e^{E_{1,N}}\big) + \mathcal{O}\big(\Theta_N\big)}{\mathcal{U}_{x_i}\big(e^{E_{i,N}}\big) - \mathcal{U}_{x_1}\big(e^{E_{1,N}}\big) + \mathcal{O}\big(\Theta_N\big)}\right),$$

where it is important to note that x_i, $i \in \mathcal{I}$ are random variables. Note that Assumption 10.2 is valid on the set $\mathcal{A}$ uniformly, with $y_i = E_{i,N}$ for $i \in \mathcal{I}$. The claim now follows by applying Lemma 10.4, and noting that $\max_{i,j\in\mathcal{I}}\big|\mathfrak{a}_{x_i} - \mathfrak{a}_{x_j}\big| = \mathcal{O}\big((\log N)^{-1}\big)$. □

PROOF OF LEMMA 10.2. It suffices to consider a fixed x, since all the bounds in the sequel are uniform for x_i, $i \in \mathcal{I}$. The proof goes along the same lines as the one of Lemma 10.1. Recall the definition of the set $\mathcal{A}$ in the proofs of Lemma 10.1 and Lemma 10.5. In addition, let

$$\mathcal{B} = \left\{\max_{i\in\mathcal{I}}\left|\frac{1}{\widehat{\mathfrak{a}}_x^*} - \frac{1}{\mathfrak{a}_{x_i}} - \frac{\mathfrak{b}_{x_i}}{\log N - \log m}\right| \leq \mathcal{O}\big((\log N)^{-1}\big)\right\}. \tag{10.31}$$

By Lemma 10.1 and Lemma 10.5, we have $P\big(\mathcal{B}^c \cup \mathcal{A}^c\big) = N\mathcal{O}\big(e^{-(\log N)^{1+\mathfrak{l}}}\big)$. Proceeding as in the proof of Lemma 10.4, it now follows that on the set $\mathcal{A}\cap\mathcal{B}$

$$\begin{aligned}\widehat{\mathfrak{b}}_x^* &= \frac{1}{m\log\log N}\sum_{i=2}^{m-1}\log\left(\mathfrak{c}_x\frac{i^{1/\mathfrak{a}_x}\log(N/i)^{\mathfrak{b}_x} - \log(N)^{\mathfrak{b}_x}}{i^{1/\widehat{\mathfrak{a}}_x^*} - 1}\right) + \frac{\mathfrak{b}_x}{\log\log N}\frac{\log N}{\log N - \log m}\\ &\quad + \mathcal{O}\big((\log\log N)^{-1}\big) =: I_m + II_m + \mathcal{O}\big((\log\log N)^{-1}\big).\end{aligned}$$

Let $m^* = m/(\log m)^2$. Then using (10.11), it follows that

$$I_m + II_m = \frac{1}{m\log\log N}\sum_{i=m^*}^{m-1}\Big(\log\big(i^{1/\mathfrak{a}_x - 1/\widehat{\mathfrak{a}}_x^*}\big) + \mathfrak{b}_x\log\big(\log(N/i)\big)\Big) + \mathcal{O}\big((\log\log N)^{-1}\big).$$

On the set $\mathcal{A}\cap\mathcal{B}$, we have, using Stirling's formula, that

$$\frac{1}{m\log\log N}\sum_{i=m^*}^{m-1}\log\big(i^{1/\mathfrak{a}_x - 1/\widehat{\mathfrak{a}}_x^*}\big) = \frac{\mathfrak{b}_x}{\log\log N}\frac{\log m}{\log N - \log m} + \mathcal{O}\big((\log\log N)^{-1}\big). \tag{10.32}$$

On the other hand, using the $\log(1-x)$ power series, we derive that

$$\mathfrak{b}_x\sum_{i=m^*}^{m-1}\log\big(\log(N/i)\big) = \mathfrak{b}_x\sum_{i=m^*}^{m-1}\log\log N\big(1 + \mathcal{O}(1)\big). \tag{10.33}$$

Hence piecing everything together, we deduce that

$$\widehat{\mathfrak{b}}_x^* = \mathfrak{b}_x + \frac{\log\mathfrak{c}_x}{\log\log N} + \frac{\mathfrak{b}_x}{\log\log N}\frac{\log m + \log N}{\log N - \log m} + \mathcal{O}\big((\log\log N)^{-1}\big).$$

The claim now follows from property $\mathfrak{b}_x \in H_{[0,1]}(\beta_0, L)$. □

REFERENCES

[1] M. Chichignoud. Minimax and minimax adaptive estimation in multiplicative regression: locally Bayesian approach. *Probab. Theory Related Fields*, 153(3-4):543–586, 2012.

[2] L. de Haan and A. Ferreira. *Extreme value theory.* Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006. An introduction.

[3] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[4] M. Jirak, A. Meister and M. Reiß
`https://www.mathematik.hu-berlin.de/for1735/Publ/r_code_and_data.zip`

Institut für Mathematik,
Humboldt-Universität zu Berlin,
Unter den Linden 6
D-10099 Berlin, Germany
E-mail: jirak@math.hu-berlin.de
E-mail: mreiss@mathematik.hu-berlin.de

Institut für Mathematik,
Universität Rostock,
D-18051 Rostock, Germany
E-mail: alexander.meister@uni-rostock.de