Estimation and hypotheses testing in boundary regression models

HOLGER DREES[*], NATALIE NEUMEYER[**] and LEONIE SELK[†]

Department of Mathematics, University of Hamburg, Bundesstrasse 55, 20146 Hamburg, Germany.
E-mail: [*]holger.drees@math.uni-hamburg.de; [**]natalie.neumeyer@math.uni-hamburg.de; [†]leonie.selk@
math.uni-hamburg.de

Consider a nonparametric regression model with one-sided errors and regression function in a general Hölder class. We estimate the regression function via minimization of the local integral of a polynomial approximation. We show uniform rates of convergence for the simple regression estimator as well as for a smooth version. These rates carry over to mean regression models with a symmetric and bounded error distribution. In such a setting, one obtains faster rates for irregular error distributions concentrating sufficient mass near the endpoints than for the usual regular distributions. The results are applied to prove asymptotic $\sqrt{n}$-equivalence of a residual-based (sequential) empirical distribution function to the (sequential) empirical distribution function of unobserved errors in the case of irregular error distributions. This result is remarkably different from corresponding results in mean regression with regular errors. It can readily be applied to develop goodness-of-fit tests for the error distribution. We present some examples and investigate the small sample performance in a simulation study. We further discuss asymptotically distribution-free hypotheses tests for independence of the error distribution from the points of measurement and for monotonicity of the boundary function as well.

Keywords: goodness-of-fit testing; irregular error distribution; one-sided errors; residual empirical distribution function; uniform rates of convergence

1. Introduction

We consider boundary regression models of the form

$$Y_i = g(x_i) + \varepsilon_i, \qquad i = 1, \ldots, n,$$

with negative errors ε_i whose survival function $1 - F(y)$ near the origin behaves like a multiple of $|y|^\alpha$ for some $\alpha > 0$. Such models naturally arise in image analysis, analysis of auctions and records, or in extreme value analysis with covariates. For such a boundary regression model with multivariate random covariates and twice differentiable regression function, Hall and Van Keilegom [17] establish a minimax rate for estimation of $g(x)$ (for fixed x) under quadratic loss and determine pointwise asymptotic distributions of an estimator which is defined as a solution of a linear optimization problem (cf. Remark 2.7). Relatedly, Müller and Wefelmeyer [25] consider a mean regression model with (unknown) symmetric support of the error distribution and Hölder continuous regression function. They discuss pointwise MSE rates for estimators of the regression function that are defined as the average of local maxima and local minima. Meister and Reiß [24] consider a regression model with known bounded support of the errors. They show

asymptotic equivalence in the strong LeCam sense to a continuous-time Poisson point process model when the error density has a jump at the endpoint of its support. For a regression model with error distribution that is one-sided and regularly varying at 0 with index $\alpha > 0$, Jirak et al. [19] suggest an estimator for the boundary regression function which adapts simultaneously to the unknown smoothness of the regression function and to the unknown extreme value index α. Reiß and Selk [29] construct efficient and unbiased estimators of linear functionals of the regression function in the case of exponentially distributed errors as well as in the limiting Poisson point process experiment by Meister and Reiß [24].

Closely related to regression estimation in models with one-sided errors is the estimation of a boundary function g based on a sample from (X, Y) with support $\{(x, y) \in [0, 1] \times [0, \infty] | y \leq g(x)\}$. For such models, Härdle et al. [18] and Hall et al. [16] proved minimax rates both for pointwise estimation of $g(x)$ and for the L_1-distance between g and its estimator. Moreover, they showed that an approach using local polynomial approximations of g yields this optimal rate. Explicit estimators in terms of higher order moments were proposed and analyzed by Girard and Jacob [14] and Girard et al. [13].

Our first aim is to develop uniform rates of convergence for a local polynomial regression estimator in boundary models. To our knowledge, uniform rates so far have only been shown by Daouia et al. [6] who consider spline estimation of a support frontier curve, but obtain slower rates. Our results can also be applied to mean regression models with bounded symmetric error distribution. For regression functions g in a Hölder class of order β, we obtain the rate $((\log n)/n)^{\beta/(\alpha\beta+1)}$. Thus, for tail index $\alpha \in (0, 2)$ of the error distribution, the rate is faster than the typical rate one has in mean regression models with regular errors. For pointwise and L^p-rates of convergence, it has been known in the literature that faster rates are possible for nonparametric regression estimation in models with irregular error distribution, see, for example, Gijbels and Peng [12], Hall and Van Keilegom [17], or Müller and Wefelmeyer [25].

The uniform rate of convergence for the regression estimator enables us to derive asymptotic expansions for residual-based empirical distribution functions and to prove weak convergence of the residual-based (sequential) empirical process. We state conditions under which the influence of the regression estimation is negligible such that the same results are obtained as in the case of observable errors.

Our results allow us to develop several hypotheses tests for boundary regression models. In particular, we will suggest asymptotically distribution-free tests for

- parametric classes of error distributions (goodness-of-fit),
- independence of the error distribution from the points of measurement,
- monotonicity of the boundary function.

Testing for a parametric class of distributions is a classical topic. The approach to compare the empirical distribution function with the parametrically estimated c.d.f. dates back to Darling [7] and is presented in detail in Chapter 5.5 of Shorack and Wellner [32] for i.i.d. data. Using an analogous approach for the error distribution in mean regression models (based on the residual e.d.f.) is hindered by the fact that the regression estimation strongly influences the asymptotic behavior of the empirical distribution function in these regular models (see Koul [23], for linear models and Akritas and Van Keilegom [1], for nonparametric mean regression). Therefore either resampling procedures (see Neumeyer et al. [26]) or martingale transformations (see Khmaladze

and Koul [20,21]) are employed in order to obtain asymptotically distribution-free tests. As we will show, in boundary regression models with an irregular error distribution, neither is necessary, but standard critical values from the i.i.d. theory can be used.

In production frontier models, x_i corresponds to the input and Y_i to the output of the ith considered firm. Then g is the production frontier or optimal boundary, corresponding to the most efficient firms ($g(x)$ is the maximal possible output with input x). Thus, $\varepsilon_i = Y_i - g(x_i)$ corresponds to the (in)efficiency of firm i, see, for example, Daouia et al. [6]. So the hypothesis of independence of errors from covariates means independence of the (in)efficiency from the input level. This hypothesis is certainly of interest for applications in economics. It may be violated, for example, because higher input levels may correspond to larger firms with better managers and thus more efficiency. For a detailed discussion, see Wilson [37] who considers several possible tests for assumptions of independence in a related model, but does not derive asymptotic distributions. Moreover, those independence assumptions are not only of intrinsic interest, but are also needed to prove validity of bootstrap procedures for nonparametric frontier models and are thus crucial in applications; see Simar and Wilson [33]. Tests for independence in nonparametric mean and quantile regression models that are similar to the test we will consider are suggested by Einmahl and Van Keilegom [8] and Birke et al. [4].

There is an extensive literature on regression with one-sided error distributions and similar models (in particular production frontier models) which assume monotonicity of the boundary function; see Gijbels et al. [11], the literature cited therein and the monotone nonparametric maximum likelihood estimator in Reiß and Selk [29]. In a production frontier model as above the production axiom of free disposability of input and output says that if x can produce y and $x' \geq x$, $y' \leq y$, then x' can produce y'. This means that the boundary curve g is increasing. This assumption is important for applications, but may often not be fulfilled; see, for example, Färe and Grosskopf [9]. To give a simple example, the use of more fertilizer will increase crop only until it is too much. To mention an entirely different area of application for monotonicity tests consider annual sport records. Here Y_i corresponds to the record in year i and g corresponds the best possible result for a given year, see Knight [22]. Non-monotonicity of g might be connected to changes in doping control procedures, see Jirak et al. [19] who consider yearly best men's outdoor 1500 m times. We are not aware of hypothesis tests for monotonicity or other shape constraints in the context of boundary regression, but would like to mention Gijbels' [10] review on testing for monotonicity in mean regression. Tests similar in spirit to the one we are suggesting here were considered by Birke and Neumeyer [3] and Birke et al. [4] for mean and quantile regression models, respectively.

The remainder of the article is organized as follows. In Section 2, the regression model under consideration is presented and model assumptions are formulated. The regression estimator is defined and uniform rates of convergence are given. A smooth modification of the estimator is considered and uniform rates of convergence for this estimator as well as its derivative are shown. In Section 3, residual empirical distribution functions based on both regression estimators are investigated. Conditions are stated under which the influence of regression estimation is asymptotically $\sqrt{n}$-negligible. Furthermore, an expansion of the residual empirical distribution function is shown that is valid under more general conditions. In Section 4, goodness-of-fit tests for the error distribution are discussed in general and in some detailed examples. We investigate the finite sample performance of the tests in a small simulation study. We further discuss hypotheses tests for independence of the error distribution from the design points, a test for monotonicity

of the boundary function and tests on the parameter α which governs the (ir)regularity of the error distribution as well. All proofs are given in the Appendix.

2. The regression function: Uniform rates of convergence

We consider a regression model with fixed equidistant design and one-sided errors,

$$Y_i = g\left(\frac{i}{n}\right) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{2.1}$$

under the following assumptions:

(F1) The errors $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed and supported on $(-\infty, 0]$. The error distribution function fulfills

$$F(y) = 1 - c|y|^\alpha + r(y), \qquad y < 0,$$

for some $\alpha > 0$, with $r(y) = o(|y|^\alpha)$ for $y \nearrow 0$.

(G1) The regression function g belongs to some Hölder class of order $\beta \in (0, \infty)$, i.e. g is $\lfloor \beta \rfloor$-times differentiable on $[0, 1]$ and the $\lfloor \beta \rfloor$th derivative satisfies

$$c_g := \sup_{\substack{t, x \in [0, 1] \\ t \neq x}} \frac{|g^{(\lfloor \beta \rfloor)}(t) - g^{(\lfloor \beta \rfloor)}(x)|}{|t - x|^{\beta - \lfloor \beta \rfloor}} < \infty.$$

In Figure 1, some scatter plots of data according to model (2.1) are shown for different tail indices α of the error distribution.

Remark 2.1. Strictly speaking, we consider a triangular scheme in (2.1), and the errors ε_i depend on n too, as the ith regression point i/n varies with n. For notational simplicity, we suppress the second index, because the distribution of the errors does not depend on n.
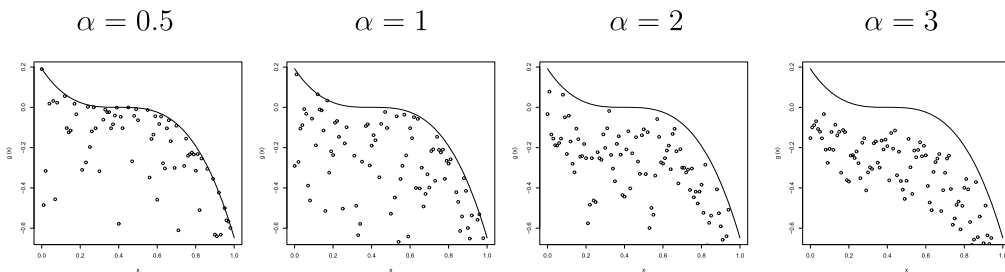

Figure 1. Scatter plots of $(\frac{i}{n}, Y_i)$, $i = 1, \ldots, n$, and the true regression function $g(x) = -3(x - 0.4)^3$. The error distribution is Weibull $F(y) = \exp(-(|y|/\theta)^\alpha)I_{(-\infty,0)}(y) + I_{[0,\infty)}(y)$ with scale $\theta = 0.3$ and shape parameter α.

We consider an estimator that locally approximates the regression function by a polynomial while lying above the data points. Polynomial estimators for boundary curves were already considered by Hall et al. [16], Hall and Park [15], Hall and Van Keilegom [17] and Jirak et al. [19]. For $x \in [0, 1]$, let

$$\hat{g}_n(x) := \hat{g}(x) := p(x),$$

where p is a polynomial of degree $\beta^* \in \mathbb{N}_0$ and minimizes the local integral

$$\int_{x-h_n}^{x+h_n} p(t)\, dt \tag{2.2}$$

under the constraints $p(\frac{j}{n}) \geq Y_j$ for all $j \in \{1, \ldots, n\}$ such that $|\frac{j}{n} - x| \leq h_n$. This estimation procedure can be realized by linear programming. For the asymptotic analysis of this estimator, we need the following assumption:

(H1) Let $(h_n)_{n \in \mathbb{N}}$ be a sequence of positive bandwidths that satisfies $\lim_{n \to \infty} h_n = 0$ and $\lim_{n \to \infty} nh_n/\log n = \infty$.

We obtain the following uniform rates of convergence.

Theorem 2.2. *In model* (2.1), *under the assumptions* (F1), (G1) *with $\beta \in (0, \beta^* + 1]$, and* (H1), *we have*

$$\sup_{x \in [h_n, 1-h_n]} |\hat{g}(x) - g(x)| = O(h_n^{\beta}) + O_P\left(\left(\frac{|\log h_n|}{nh_n}\right)^{1/\alpha}\right).$$

Note that the deterministic part $O(h_n^{\beta})$ arises from approximating the regression function by a polynomial, whereas the random part originates from the observational error. Balancing the two sources of error by setting $h_n \asymp ((\log n)/n)^{\frac{1}{\alpha\beta+1}}$ gives

$$\sup_{x \in [h_n, 1-h_n]} |\hat{g}(x) - g(x)| = O_P\left(\left(\frac{\log n}{n}\right)^{\frac{\beta}{\alpha\beta+1}}\right). \tag{2.3}$$

(Here $a_n \asymp b_n$ means that $0 < \liminf_{n \to \infty} |a_n/b_n| \leq \limsup_{n \to \infty} |a_n/b_n| < \infty$.)

This result is of particular interest in the case of irregular error distributions, i.e. $\alpha \in (0, 2)$, when the rate improves upon the typical optimal rate $O_P(((\log n)/n)^{\beta/(2\beta+1)})$ for estimating mean regression functions in models with regular errors. This improvement for small values of α is intuitively expected since then many data very close to g are observed (see Figure 1).

Remark 2.3. If the degree β^* of the polynomial in our estimator is chosen smaller than $\lceil\beta\rceil - 1$, then we obtain the rate

$$\sup_{x \in [h_n, 1-h_n]} |\hat{g}(x) - g(x)| = O(h_n^{\beta^*+1}) + O_P\left(\left(\frac{|\log h_n|}{nh_n}\right)^{1/\alpha}\right).$$

This is a direct conclusion from Theorem 2.2 and the fact that if g is Hölder of order β then it is also Hölder of order β' for all $\beta' \leq \beta$.

Remark 2.4. Jirak et al. [19] consider a similar boundary regression estimator while replacing the integral in (2.2) by its Riemann approximation $\sum_{i=1}^{n} p(\frac{i}{n})I\{|\frac{i}{n} - x| \leq h_n\}$. In particular, using the Lepskii method, they propose a fully data-driven bandwidth (merely assuming an upper bound on β, but no information about α) such that the resulting estimator converges pointwise with the optimal rate. (Proposition 3.2 in that paper is not fully correct, but the error can be fixed by using an estimator of α of the type considered in Section 4.4 of the present paper instead of their estimator (3.10) and modifying the estimator (3.13) accordingly.) We conjecture that an analogous adaptive choice of the bandwidth leads to an estimator which converges at the rate $((\log n)/n)^{\beta/(\alpha\beta+1)}$ w.r.t. the supremum norm on $[h_n, 1 - h_n]$, too. Indeed, one of the main steps in the proof of adaptivity has been established in Proposition 3.1 of Jirak et al. [19] for a general seminorm.

Remark 2.5. For local constant approximation (i.e. $\beta^* = 0$) the estimator reduces to a local maximum $\hat{g}(x) = \max\{Y_i | i = 1, \ldots, n \text{ s.t. } |\frac{i}{n} - x| \leq h_n\}$. In this case we obtain the rate of convergence as given in Theorem 2.2 uniformly over the whole unit interval.

Remark 2.6. Müller and Wefelmeyer [25] consider a mean regression model $Y_i = m(X_i) + \eta_i$, $i = 1, \ldots, n$, with symmetric error distribution supported on $[-a, a]$ (with a unknown); see the left panel of Figure 2. The error distribution function fulfills $F(a - y) \sim 1 - y^{\alpha}$ for $y \searrow 0$. The local empirical midrange of responses, that is,

$$\hat{m}(x) = \frac{1}{2}\left(\min_{\substack{i \in \{1,\ldots,n\} \\ |X_i - x| \leq h_n}} Y_i + \max_{\substack{i \in \{1,\ldots,n\} \\ |X_i - x| \leq h_n}} Y_i \right)$$

is shown to have pointwise rate of convergence $O(h_n^{\beta}) + O_P((nh_n)^{-1/\alpha})$ to $m(x)$ if m is Hölder continuous with exponent $\beta \in (0, 1]$. Theorem 2.2 enables us to extend Müller's and
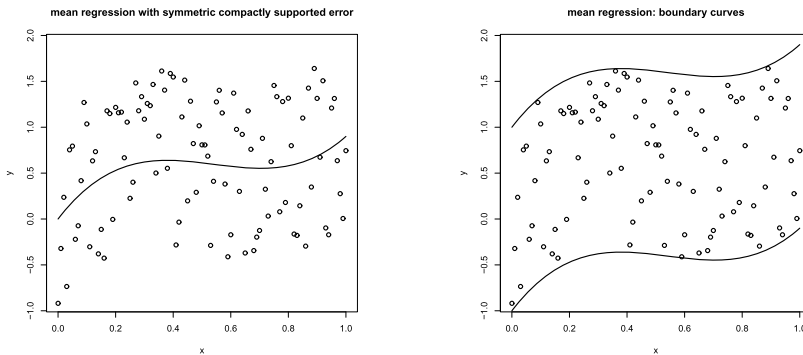

Figure 2. Example for data as in Remark 2.6.

Wefelmeyer's [25] results in two ways (in a model with fixed design $X_i = \frac{i}{n}$): we consider more general Hölder classes with general index $\beta > 0$, and we obtain uniform rates of convergence. To this end, we use the mean regression estimator $\hat{m} = (\hat{g} - \hat{\hat{g}})/2$ with $\hat{g}$ as before and $\hat{\hat{g}}$ defined analogously, but based on $(\frac{i}{n}, -Y_i)$, $i = 1, \ldots, n$; see the right panel of Figure 2. The rates obtained for $\sup_{x \in [h_n, 1-h_n]} |\hat{m}(x) - m(x)|$ are the same as in Theorem 2.2.

Remark 2.7. For $\beta \in (1, 2]$, Hall and Van Keilegom [17] consider the following local linear boundary regression estimator:

$$\check{g}(x) = \inf\left\{\alpha_0 \,\middle|\, (\alpha_0, \alpha_1) \in \mathbb{R}^2 : Y_i \leq \alpha_0 + \alpha_1\left(\frac{i}{n} - x\right) \,\forall i \in \{1, \ldots, n\} \text{ s.t. } \left|\frac{i}{n} - x\right| \leq h_n\right\}.$$

Because of $\int_{x-h_n}^{x+h_n}(\alpha_0 + \alpha_1(t - x)) \, dt = 2\alpha_0 h_n$ this estimator coincides with $\hat{g}$ for $\beta^* = 1$. However, in the case $\beta^* \geq 2$ replacing the linear function in the definition of $\check{g}$ by a polynomial of degree β^* renders the estimator useless. One obtains $\check{g}(x) = -\infty$ for $x \notin \{\frac{j}{n} | j = 1, \ldots, n\}$ while $\check{g}(\frac{j}{n}) = Y_j$, $j = 1, \ldots, n$. This was already observed by Jirak et al. [19].

Note that typically the estimator $\hat{g}$ is not continuous. One might prefer to consider a smooth estimator by convoluting $\hat{g}$ with a kernel. Such a modified estimator will also be advantageous when deriving an expansion for the residual based empirical distribution function in the next section. Therefore we define

$$\tilde{g}(x) = \int_{h_n}^{1-h_n} \hat{g}(z) \frac{1}{b_n} K\left(\frac{x - z}{b_n}\right) dz \tag{2.4}$$

and formulate some additional assumptions.

(K1) K is a continuous kernel with support $[-1, 1]$ and order $\beta^* + 1$, that is, $\int K(u) \, du = 1$, $\int u^r K(u) \, du = 0 \,\forall r = 1, \ldots, \beta^*$. Furthermore, K is differentiable with Lipschitz-continuous derivative K' on $(-1, 1)$.

(B1) The sequence $(b_n)_{n \in \mathbb{N}}$ of positive bandwidths satisfies $\lim_{n \to \infty} b_n = 0$.

(B2.δ)

$$h_n^\beta + \left(\frac{\log n}{n h_n}\right)^{1/\alpha} = o\left(b_n^{(1+2\delta)\vee(3-(\beta-1)(1/\delta-1))}\right)$$

$$= \begin{cases} o\left(b_n^{1+2\delta}\right) & \text{if } \delta \leq \dfrac{\beta - 1}{2}, \\ o\left(b_n^{3-(\beta-1)(1/\delta-1)}\right) & \text{if } \delta > \dfrac{\beta - 1}{2}. \end{cases}$$

Here we assume that the parameter δ, which quantifies the minimal required smoothness of the estimator of g', lies in $(0, 1 \wedge (\beta - 1))$. For example, if $\beta \in (1, 3)$ and the bandwidth $h_n \asymp ((\log n)/n)^{1/(\alpha\beta+1)}$ is chosen, then (B2.δ) is fulfilled with $\delta = (\beta - 1)/2$ for any b_n that satisfies $h_n = o(b_n)$.

The estimator $\tilde{g}$ is differentiable and we obtain the following uniform rates of convergence for $\tilde{g}$ and its derivative $\tilde{g}'$.

Theorem 2.8. *Assume the model assumptions* (2.1), (F1), (G1) *with* $\beta \in (1, \beta^* + 1]$, (H1), (K1), *and* (B1) *are valid, and define* $I_n = [h_n + b_n, 1 - h_n - b_n]$. *Then it holds that*

(i) $\sup_{x \in I_n} |\tilde{g}(x) - g(x)| = O(b_n^\beta) + O(h_n^\beta) + O_P((\frac{|\log h_n|}{nh_n})^{\frac{1}{\alpha}})$,

(ii) $\sup_{x \in I_n} |\tilde{g}'(x) - g'(x)| = O(b_n^{\beta-1}) + O(b_n^{-1} h_n^\beta) + O_P(b_n^{-1}(\frac{|\log h_n|}{nh_n})^{\frac{1}{\alpha}})$.

If additionally assumption (B2.δ) *holds for* $\delta \in (0, 1 \wedge (\beta - 1))$, *then*

(iii) $\sup_{x, y \in I_n, x \neq y} \frac{|\tilde{g}'(x) - g'(x) - \tilde{g}'(y) + g'(y)|}{|x-y|^\delta} = o_P(1)$.

Note that Theorem 2.8(ii) gives $\sup_{x \in I_n} |\tilde{g}'(x) - g'(x)| = o_P(1)$ if $h_n^\beta + (\log n/(nh_n))^{1/\alpha} = o(b_n)$. This holds in particular if (B2.δ) is fulfilled for some $\delta \in (0, 1 \wedge (\beta - 1))$.

3. The error distribution

3.1. Estimation

In this section, we consider estimators of the error distribution in model (2.1). For the asymptotic analysis, we need the following additional assumption.

(F2) The c.d.f. F of the errors is Hölder continuous of order $\alpha \wedge 1$.

We define residuals $\hat{\varepsilon}_i = Y_i - \hat{g}(\frac{i}{n})$, and a resulting modified sequential empirical distribution function by

$$\hat{F}_n(y, s) = \frac{1}{m_n} \sum_{i=1}^{\lfloor ns \rfloor} I\{\hat{\varepsilon}_i \leq y\} I\left\{h_n < \frac{i}{n} \leq 1 - h_n\right\},$$

where $m_n = \sharp\{i \in \{1, \dots, n\} | h_n < \frac{i}{n} \leq 1 - h_n\} = n - \lfloor nh_n \rfloor - \lceil nh_n \rceil$. We consider the sequential process, because it will be useful for testing hypotheses in Section 4. With slight abuse of notation, let $\hat{F}_n(y) = \hat{F}_n(y, 1)$ denote the corresponding estimator for $F(y)$.

We first treat a simple case where the influence of the regression estimation on the residual empirical process is negligible. To this end, let F_n denote the standard empirical distribution function of the unobservable errors $\varepsilon_1, \dots, \varepsilon_n$. Further, define $\bar{s}_n = (\lfloor n(s \wedge (1 - h_n)) \rfloor - \lfloor n(s \wedge h_n) \rfloor)/m_n$ and interpret $\bar{s}_n/\lfloor ns \rfloor$ as 0 for $s = 0$. Note that $\bar{s}_n = 1$ if $s = 1$ and $\bar{s}_n \to s$ as $n \to \infty$, for each fixed s.

Theorem 3.1. *Assume that the conditions* (F1), (G1), (H1) *and* (F2) *are fulfilled and that* $\sup_{x \in [h_n, 1-h_n]} |\hat{g}(x) - g(x)| = o_P(n^{-1/(2(\alpha \wedge 1))})$. *Then we have*

$$\sup_{y \in \mathbb{R}, s \in [0,1]} |\hat{F}_n(y, s) - \bar{s}_n F_{\lfloor ns \rfloor}(y)| = o_P(n^{-1/2}).$$

Thus the process $\{\sqrt{n}(\hat{F}_n(y, s) - \bar{s}_n F(y)) | s \in [0, 1], y \in \mathbb{R}\}$ *converges weakly to a Kiefer process* K_F, *a centered Gaussian process with covariance function* $((s_1, y_1), (s_2, y_2)) \mapsto (s_1 \wedge s_2)(F(y_1 \wedge y_2) - F(y_1)F(y_2))$.

Remark 3.2. The assumption $\sup_{x\in[h_n,1-h_n]}|\hat{g}(x)-g(x)|=o_P(n^{-1/(2(\alpha\wedge1))})$ is satisfied under the following condition on the bandwidth h_n:

(B3) The sequence of bandwidths satisfies $h_n=o(n^{-1/(2(\alpha\wedge1)\beta)})$, $n^{(\alpha\vee1)/2-1}\log n=o(h_n)$.

Condition (B3) can be fulfilled if and only if $\frac{1}{\beta}<\alpha<2-\frac{1}{\beta}$, which requires $\beta>1$. Note that the condition can be met for all $\alpha\in(0,2)$, provided the regression function g is sufficiently smooth (i.e. β is large enough). In this case, a bandwidth $h_n\asymp((\log n)/n)^{1/(\alpha\beta+1)}$ fulfills (B3). Moreover, an adaptive version of $\hat{g}$ as discussed in Remark 2.4 would satisfy $\sup_{x\in[h_n,1-h_n]}|\hat{g}(x)-g(x)|=o_P(n^{-1/(2(\alpha\wedge1))})$.

Remark 3.3. Theorem 3.1 implies that for $\alpha\in(1/\beta,2-1/\beta)$ the estimation of the regression function has no asymptotic impact on the estimation of the irregular error distribution. This is remarkably different from corresponding results on the estimation of the error distribution in mean regression models with regular error distributions. Here the empirical distribution function of residuals, say $\check{F}_n$, is not asymptotically $\sqrt{n}$-equivalent to the empirical distribution function of true errors. The process $\sqrt{n}(\check{F}_n-F)$ converges to a Gaussian process whose covariance structure depends on the error distribution in a complicated way; cf. Theorem 2 in Akritas and Van Keilegom [1]. In the simple case of a mean regression model with equidistant design and an error distribution F with bounded density f one has

$$\sqrt{n}\big(\check{F}_n(y)-F_n(y)\big)=\frac{f(y)}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i+o_P(1)$$

uniformly with respect to $y\in\mathbb{R}$ when the regression function is estimated by a local polynomial estimator, under appropriate bandwidth conditions (see Proposition 3 in Neumeyer and Van Keilegom [27]).

In order to obtain asymptotic results for estimators of the error distribution for $\alpha\geq2-\frac{1}{\beta}$, a finer analysis is needed. In what follows, we will use the smooth regression estimator $\tilde{g}$ defined in (2.4). Let $\tilde{F}_n$ denote the empirical distribution function based on residuals $\tilde{\varepsilon}_j=Y_j-\tilde{g}(\frac{j}{n})$, that is,

$$\tilde{F}_n(y)=\frac{1}{m_n}\sum_{j=1}^{n}I\{\tilde{\varepsilon}_j\leq y\}I\left\{\frac{j}{n}\in I_n\right\}, \tag{3.1}$$

where $I_n=[h_n+b_n,1-h_n-b_n]$ and $m_n=\sharp\{j\in\{1,\ldots,n\}|h_n+b_n\leq\frac{j}{n}\leq1-h_n-b_n\}=n-2\lceil n(h_n+b_n)\rceil+1$. Then the following asymptotic expansion is valid.

Theorem 3.4. *If the conditions* (F1), (F2), (G1) *with* $\beta > 1$, (H1), (K1), (B1), *and* (B2.δ) *for some* $\delta \in (0 \vee (1/\alpha - 1), 1 \wedge (\beta - 1))$ *are fulfilled, then*

$$
\tilde{F}_n(y) = \frac{1}{n} \sum_{j=1}^{n} I\{\varepsilon_j \leq y\} + \frac{1}{m_n} \sum_{j=1}^{n} \left(F\left(y + (\tilde{g} - g)\left(\frac{j}{n}\right) \right) - F(y) \right) I\left\{ \frac{j}{n} \in I_n \right\}
$$
$$
+ o_P\left(\frac{1}{\sqrt{n}} \right)
$$

(3.2)

uniformly for all $y \in \mathbb{R}$.

Remark 3.5. One can choose bandwidths h_n and b_n such that the conditions (H1), (B1) and (B2.δ) are fulfilled for some $\delta \in (0 \vee (1/\alpha - 1), 1 \wedge (\beta - 1))$ if this interval is not empty, which in turn is equivalent to $\alpha > 1/(\beta \wedge 2)$. Thus, the expansion given in Theorem 3.4 is also valid for regular error distributions.

If one assumes (B2.δ) for some $\delta \in (0, 1 \wedge (\beta - 1))$, but drops the condition $\delta > 1/\alpha - 1$ and, in addition, replaces (F2) with the assumption that F is Lipschitz continuous on $(-\infty, \kappa]$ for some $\kappa < 0$, then expansion (3.2) still holds uniformly on $(-\infty, \tilde{\kappa}]$ for all $\tilde{\kappa} < \kappa$. In particular, this holds if F has a bounded density on $(-\infty, \kappa]$.

Next, we examine under which conditions the additional term in (3.2) depending on the estimation error is asymptotically negligible. Beyond that the expansion in Theorem 3.5 might be the starting point for a more general investigation of the empirical residual process in the future. Here, we focus on those arguments y which are bounded away from 0, because in this setting weaker conditions on α and β are needed. Moreover, for the analysis of the tail behavior of the error distribution at 0, tail empirical processes are better suited and will be considered in future work.

Note that the estimator $\hat{g}$ tends to underestimate the true function because it is defined via a polynomial which is minimal under the constraint that it lies above all observations $(i/n, Y_i)$, which in turn all lie below the true boundary function. As this systematic underestimation does not vanish from (local or global) averaging, we first have to introduce a bias correction.

Let $E_{g \equiv 0}$ denote the expectation if the true regression function is identical 0. For the remaining part of this section, we assume that $\mu = E_{g \equiv 0}(\hat{g}(1/2))$ is known or that it can be estimated sufficiently accurately. For example, if the empirical process of residuals shall be used to test a simple null hypothesis $H_0 : F = F_0$, then one may calculate or simulate μ under the given null distribution. If the process is used for testing for a parametric class $H_0 : F \in \{F_\vartheta | \vartheta \in \Theta\}$ (see Section 4.1), then $\mu(\vartheta) = E_{\vartheta, g \equiv 0}(\hat{g}(1/2))$ can be estimated by $\hat{\mu} = \mu(\hat{\vartheta}_n)$, where $\hat{\vartheta}_n$ denotes a consistent estimator for the true parameter under the null. If the expectation has no explicit form, it can be approximated via Monte Carlo methods. The results of Theorem 3.7 hold true under the condition $\mu(\hat{\vartheta}_n) - \mu(\vartheta) = o_{P_\vartheta}(n^{-1/2})$ if one defines $\tilde{g}_n^*(x) := \tilde{g}(x) - \mu(\hat{\vartheta}_n)$.

We define a bias corrected version of the smoothed estimator by

$$
\tilde{g}_n^*(x) := \tilde{g}(x) - E_{g \equiv 0}\big(\hat{g}(1/2)\big),
$$

for $x \in I_n$. Note that $E_{g \equiv 0}(\hat{g}_n(x)) = E_{g \equiv 0}(\hat{g}_n(1/2))$ for any $x \in [h_n, 1 - h_n]$ because under $g \equiv 0$ we observe i.i.d. data $Y_i = \varepsilon_i$, $i = 1, \ldots, n$. The following lemma ensures that the above results for $\tilde{g}$ carry over to this variant if the following condition on the lower tail of F holds:

(F3) There exists $\tau > 0$ such that $F(-t) = o(t^{-\tau})$ as $t \to \infty$.

Lemma 3.6. *If model* (2.1) *holds with* g *identical* 0 *and the conditions* (F1), (F3), (G1), *and* (H1) *are fulfilled, then for all* $x \in [h_n, 1 - h_n]$

$$E_{g \equiv 0}\big(|\hat{g}_n(x)|\big) = O\left(\left(\frac{\log n}{n h_n}\right)^{1/\alpha}\right).$$

We need some additional conditions on the rates at which the bandwidths h_n and b_n tend to 0:

(H2) $h_n = o(n^{-1/(2\beta)} \wedge n^{-1/(\alpha\beta+1)})$, $n^{\alpha/4-1} \log n = o(h_n)$,
(B4) $b_n = o(n^{-1/(2\beta)} \wedge (h_n^{-2\beta} n^{-1}) \wedge ((\frac{nh_n}{\log n})^{2/\alpha} n^{-1}))$.

In particular, these assumptions ensure that the bias terms of order $h_n^{\beta} + b_n^{\beta}$ are of smaller order than $n^{-1/2}$ and $(nh_n)^{-1/\alpha}$ and hence asymptotically negligible, and that quadratic terms in the estimation error are uniformly negligible, that is, $\sup_{x \in I_n} |\tilde{g}_n^*(x) - g(x)|^2 = o_P(n^{-1/2})$.

Theorem 3.7. *Suppose the model assumptions* (2.1) *with* $\alpha \in (0, 2)$, $\beta > 1$, (F1), (F3), (G1), (H1), (H2), (K1), (B1), (B2.δ) *for some* $\delta > 0$, *and* (B4) *hold and* F *has a bounded density on* $(-\infty, \kappa]$ *for some* $\kappa < 0$. *Then*

$$\sup_{y \in (-\infty, \kappa]} \left| \frac{1}{m_n} \sum_{j=1}^{n} \left(F\left(y + (\tilde{g}_n^* - g)\left(\frac{j}{n}\right)\right) - F(y) \right) I\left\{ \frac{j}{n} \in I_n \right\} \right| = o_P\left(n^{-1/2}\right).$$

Remark 3.8. The conditions on h_n and b_n used in Theorem 3.7 can be fulfilled if and only if $\alpha < 2\beta - 1$. In particular, this theorem is applicable if $\beta \geq 3/2$ and the error distribution is irregular, i.e., $\alpha < 2$. A possible choice of bandwidths is

$$h_n \asymp \left(n^{-1/(2\beta)} \wedge n^{-1/(\alpha\beta+1)}\right)/\log n, \qquad b_n \asymp n^{-\lambda} \qquad \text{for some } \lambda \in \left(\frac{1}{2\beta}, \frac{\beta}{\alpha\beta+1} \wedge \frac{2\beta - 1}{2\alpha\beta}\right).$$

We obtain asymptotic equivalence of the empirical process of residuals (restricted to $(-\infty, \kappa]$) to the empirical process of the errors. To formulate the result, let $\tilde{F}_n^*$ be defined analogously to $\tilde{F}_n$, but with $\tilde{g}$ replaced by $\tilde{g}^*$.

Corollary 3.9. *Under the assumptions of Theorems* 3.4 *and* 3.7, *we have* $\sup_{y \in (-\infty, \kappa]} |\tilde{F}_n^*(y) - F_n(y)| = o_P(n^{-1/2})$. *Thus, the process* $(\sqrt{n}(\tilde{F}_n^*(y) - F(y)))_{y \in (-\infty, \kappa]}$ *converges weakly to a centered Gaussian process with covariance function* $(y_1, y_2) \mapsto F(y_1 \wedge y_2) - F(y_1)F(y_2)$, $y_1, y_2 \in (-\infty, \kappa]$.

Note that for the corollary one needs the condition $1/(\beta \wedge 2) < \alpha < (2\beta - 1) \wedge 2$.

4. Hypotheses testing

4.1. Goodness-of-fit testing

Let $\mathcal{F} = \{F_\vartheta | \vartheta \in \Theta\}$ denote a continuously parametrized class of error distributions such that for each $\vartheta \in \Theta$, $F_\vartheta(y) = 1 - c_\vartheta |y|^{\alpha_\vartheta} + r_\vartheta(y)$ with $r_\vartheta(y) = o(|y|^{\alpha_\vartheta})$ for $y \nearrow 0$. Our aim is to test the null hypothesis $H_0 : F \in \mathcal{F}$. We assume that $\alpha_\vartheta \in (1/\beta, 2 - 1/\beta)$ for all $\vartheta \in \Theta$, such that Theorem 3.1 can be applied under H_0. Let $\hat{\vartheta}_n$ denote an estimator for ϑ based on residuals $\hat{\varepsilon}_i = Y_i - \hat{g}(\frac{i}{n})$, $i = 1, \ldots, n$. The goodness-of-fit test is based on the empirical process

$$S_n(y) = \sqrt{n}(\hat{F}_n(y) - F_{\hat{\vartheta}_n}(y)), \qquad y \in \mathbb{R},$$

where, as before, $\hat{F}_n(y) = \hat{F}_n(y, 1)$. Under any fixed alternative that fulfills (F1) for some α, $\hat{g}$ still uniformly consistently estimates g, and thus $\hat{F}_n$ is a consistent estimator of the error distribution F. If $\hat{\vartheta}_n$ converges to some $\vartheta^* \in \Theta$ under the alternative, too, then a consistent hypothesis test is obtained by rejecting H_0 for large values of, for example, a Kolmogorov–Smirnov test statistic $\sup_{y \in \mathbb{R}} |S_n(y)|$. Note that under H_0 it follows from Theorem 3.1 that

$$S_n(y) = \sqrt{n}(F_n(y) - F_\vartheta(y)) - \sqrt{n}(F_{\hat{\vartheta}_n}(y) - F_\vartheta(y)) + o_P(1),$$

where ϑ denotes the true parameter. We consider two examples.

Example 4.1. Consider the mean regression model $Y_i = m(\frac{i}{n}) + \eta_i$, $i = 1, \ldots, n$, with symmetric error c.d.f. F and $\beta > 1$, and define $\hat{m}$ as in Remark 2.6 with some bandwidth h_n that fulfills (B3). We want to test the null hypothesis $H_0 : F \in \mathcal{F} = \{F_\vartheta | \vartheta \in \Theta\}$ for some $\Theta \subset (0, \infty)$, where F_ϑ denotes the distribution function of the uniform distribution on $[-\vartheta, \vartheta]$ (with $\alpha_\vartheta = 1$ for all $\vartheta > 0$). Define residuals $\hat{\eta}_i = Y_i - \hat{m}(\frac{i}{n})$, $i = 1, \ldots, n$, and let

$$\hat{\vartheta}_n = \max\left(\max_{nh_n \leq i \leq n - nh_n} \hat{\eta}_i, - \min_{nh_n \leq i \leq n - nh_n} \hat{\eta}_i \right)$$

$$= \max_{nh_n \leq i \leq n - nh_n} |\hat{\eta}_i|.$$

Then $|\hat{\vartheta}_n - \vartheta|$ is bounded by $|\max_{nh_n \leq i \leq n - nh_n} |\eta_i| - \vartheta| + \sup_{x \in [h_n, 1 - h_n]} |\hat{m}(x) - m(x)| = o_P(n^{-1/2})$. Since $F_\vartheta(y) = \frac{y + \vartheta}{2\vartheta} I_{[-\vartheta, \vartheta]}(y) + I_{(\vartheta, \infty)}(y)$, one may conclude $\sup_{y \in \mathbb{R}} |F_{\hat{\vartheta}_n}(y) - F_\vartheta(y)| = o_P(n^{-1/2})$. Thus, the process S_n converges weakly to a Brownian bridge B composed with F. The Kolmogorov–Smirnov test statistic $\sup_{y \in \mathbb{R}} |S_n(y)|$ converges in distribution to $\sup_{t \in [0,1]} |B(t)|$. Thus, although our testing problem requires the estimation of a nonparametric function and we have a composite null hypothesis, the same asymptotic distribution arises as in the Kolmogorov–Smirnov test for the simple hypothesis $H_0 : F = F_0$ based on an i.i.d. sample with distribution F.

Example 4.2. Again assume that the Hölder coefficient β is greater than 1. Consider the null hypothesis $H_0 : F \in \mathcal{F} = \{F_\vartheta | \vartheta \in (0, \infty)\}$, where $F_\vartheta(y) = e^{-(-\vartheta y)^\alpha} I_{(-\infty, 0)}(y) + I_{[0, \infty)}(y)$ de-

notes a Weibull distribution with some fixed shape parameter $\alpha \in (1/\beta, 2 - 1/\beta)$ and unknown scale parameter ϑ. Note that F_ϑ satisfies (F1) with $c = \vartheta^\alpha$.

Define the moment estimator $\hat{\vartheta}_n = (\frac{1}{m_n} \sum_{j=1}^n (-\hat{\varepsilon}_j)^\alpha I\{h_n < \frac{j}{n} \le 1 - h_n\})^{-\frac{1}{\alpha}}$ which is motivated by $E_\vartheta[(-\varepsilon_1)^\alpha] = \vartheta^{-\alpha}$. A Taylor expansion of $x \mapsto x^\alpha$ at $x = -\varepsilon_j$ yields

$$
\hat{\vartheta}_n^\alpha - \vartheta^\alpha = -(\hat{\vartheta}_n \vartheta)^\alpha \left(\frac{1}{m_n} \sum_{j=1}^n ((-\varepsilon_j)^\alpha - \vartheta^{-\alpha}) I\left\{ h_n < \frac{j}{n} \le 1 - h_n \right\} \right.
$$

$$
\left. + \frac{\alpha}{m_n} \sum_{j=1}^n (-\xi_j)^{\alpha-1} \left(\hat{g}\left(\frac{j}{n}\right) - g\left(\frac{j}{n}\right) \right) I\left\{ h_n < \frac{j}{n} \le 1 - h_n \right\} \right)
$$

$$
= -\vartheta^{2\alpha} \frac{1}{m_n} \sum_{j=1}^n ((-\varepsilon_j)^\alpha - \vartheta^{-\alpha}) I\left\{ h_n < \frac{j}{n} \le 1 - h_n \right\} + o_{P_\vartheta}(n^{-1/2})
$$

$$
= O_{P_\vartheta}(n^{-1/2})
$$

for some ξ_j between $\hat{\varepsilon}_j$ and ε_j, where in the last steps we have applied Theorem 2.2, the law of large numbers and a central limit theorem. With a Taylor expansion of $G(\theta, y) = e^{-(-y)^\alpha \theta}$ in $\theta = \hat{\vartheta}_n^\alpha$ around $\theta = \vartheta^\alpha$ one obtains

$$
F_{\hat{\vartheta}_n}(y) - F_\vartheta(y) = -(-y)^\alpha e^{-(-\vartheta y)^\alpha} (\hat{\vartheta}_n^\alpha - \vartheta^\alpha) + O_{P_\vartheta}(n^{-1})
$$

uniformly for all $y \in (-\infty, 0]$. Now analogously to the proof of Theorem 19.23 in van der Vaart [35], we can conclude weak convergence of

$$
S_n(y) = \sqrt{n}\left(F_n(y) - F(y)\right)
$$

$$
- e^{-(-\vartheta y)^\alpha} \vartheta^{2\alpha} (-y)^\alpha \frac{\sqrt{n}}{m_n} \sum_{j=1}^n \left((-\varepsilon_j)^\alpha - \frac{1}{\vartheta^\alpha} \right) I\left\{ h_n < \frac{j}{n} \le 1 - h_n \right\}
$$

$$
+ o_{P_\vartheta}(1),
$$

$y \in \mathbb{R}$, to a Gaussian process with covariance function $(y_1, y_2) \mapsto F_\vartheta(y_1 \wedge y_2) - F_\vartheta(y_1) F_\vartheta(y_2) - e^{-(-\vartheta)^\alpha (y_1^\alpha + y_2^\alpha)} (y_1 y_2)^\alpha \vartheta^{2\alpha}$, where the covariance function follows by simple calculations and the fact that $E_\vartheta[I\{\varepsilon_1 \le y\}((-\varepsilon_1)^\alpha - \vartheta^{-\alpha})] = (-y)^\alpha e^{-(-\vartheta y)^\alpha}$. For the special case of a test for exponentially distributed errors ($\alpha = 1$), the asymptotic quantiles for the Cramér–von-Mises test statistic $\int S_n(y)^2 \, dF_{\hat{\vartheta}_n}(y)$ are tabled in Stephens [34].

Simulations

To study the finite sample performance of our goodness-of-fit test, we investigate its behaviour on simulated data according to Examples 4.1 and 4.2 for samples of size 50, 100, 200 and 500. In both settings, the regression function is given by $g(x) = 0.5 \sin(2\pi x) + 4x$. We use the local linear estimator (corresponding to $\beta = 2$) with bandwidth $n^{-\frac{1}{3}}$, which corresponds to the rate
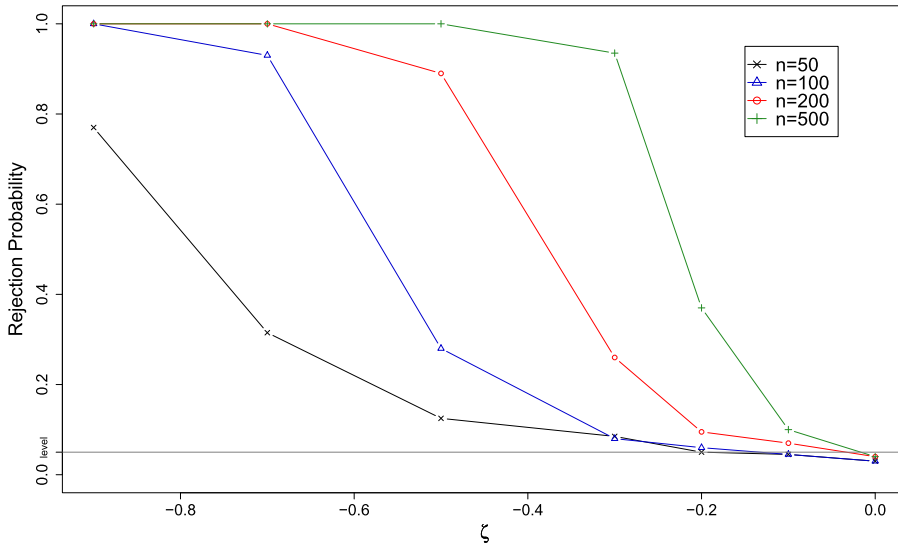
Figure 3. Monte-Carlo simulations for Example 4.1.

considered in (2.3) for $\alpha = 1$ and $\beta = 2$ (up to a log term). The hypothesis tests are based on the adjusted Cramér–von-Mises test statistic $\frac{m_n}{n} \int S_n(y)^2 \, dF_{\hat{\vartheta}_n}(y)$ and have nominal size 5%. The results reported below are based on 200 Monte Carlo simulations for each model.

In the situation of Example 4.1, the errors are drawn according to the density $f_\varepsilon(y) = 0.5(\zeta + 1)(1 - |y|)^\zeta I_{[-1,1]}(y)$ for different values of $\zeta \in (-1, 0]$. Note that the null hypothesis $H_0 : \exists \vartheta : \varepsilon_i \sim U[-\vartheta, \vartheta]$ holds if and only if $\zeta = 0$. Figure 3 shows the empirical power of the Cramér–von-Mises type test. The actual size is close to the nominal level for all sample sizes and the power function is monotone both in ζ and the sample size n. For parameter values $\zeta \in [-0.2, 0)$, one needs rather large sample sizes to detect the alternative, as the error distribution is too similar to the uniform distribution.

In the setting of Example 4.2, we simulate Weibull(ϑ, α) distributed errors for $\vartheta = 1$ and different values of $\alpha > 0$. We test the null hypothesis $H_0 : \exists \vartheta : -\varepsilon_i \sim \text{Exp}(\vartheta)$ of exponentiality, which is only fulfilled for $\alpha = 1$. In Figure 4, the empirical power function of our test is displayed for different sample sizes. Again the actual size is close to the 5% and the power increases with α departing from one as well as with increasing n.

To examine the influence of the bandwidth choice, in addition we have simulated the same models with $h_n = c \cdot n^{-\frac{1}{3}}$ for different values of c ranging from $c = 0.2$ to $c = 1.2$. The results for the test of uniformity in Example 4.1 are similar to those displayed in Figure 3 for all these bandwidths. In the situation of Example 4.2, we obtain similar power functions as reported above for c between 0.8 and 1.2, whereas for smaller bandwidths the actual size of the test exceeds its nominal value substantially.
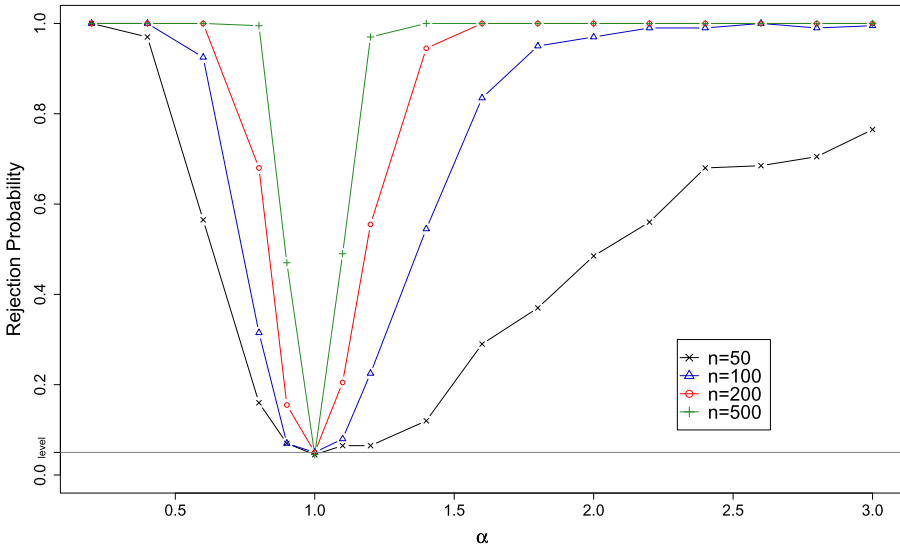
Figure 4. Monte-Carlo simulations for Example 4.2.

4.2. Test for independence

In model (2.1), we assume that the distributions of the errors ε_i ($i = 1, \ldots, n$) do not depend on the point of measurement $x_i = i/n$. We can test this assumption by comparing the sequential empirical distribution function $\hat{F}_n(y, s)$ with the estimator $\bar{s}_n \hat{F}_n(y)$, which should behave similarly if the errors are i.i.d. The following corollary to Theorem 3.1 describes the asymptotic behavior of the Kolmogorov–Smirnov type test statistic

$$T_n = \sup_{s \in [0,1], y \in \mathbb{R}} \sqrt{n} \left| \hat{F}_n(y, s) - \bar{s}_n \hat{F}_n(y) \right|$$

under the null hypothesis of i.i.d. errors.

Corollary 4.3. *Assume model* (2.1) *with* (F1), (F2), (G1) *and that* $\sup_{x \in [h_n, 1-h_n]} |\hat{g}(x) - g(x)| = o_P(n^{-1/(2(\alpha \wedge 1))})$ *holds. Then* T_n *converges in distribution to* $\sup_{s \in [0,1], z \in [0,1]} |G(s, z)|$ *where* G *is a completely tucked Brownian sheet, that is, a centered Gaussian process with covariance function* $((s_1, z_1), (s_2, z_2)) \mapsto (s_1 \wedge s_2 - s_1 s_2)(z_1 \wedge z_2 - z_1 z_2)$.

The proof is given in the Appendix. Note that under the assumptions of the corollary the limit of the test statistic T_n is distribution free. The asymptotic quantiles tabled by Picard [28] can be used to determine the critical value for a given asymptotic size of the test.

4.3. Test for monotone boundary functions

We consider model (2.1) and aim at testing the null hypothesis

$$H_0 : g \text{ is increasing,}$$

which is a common assumption in boundary models. Let $\tilde{g}$ denote the smooth local polynomial estimator for g defined in (2.4). Such an unconstrained estimator can be modified to obtain an increasing estimator $\tilde{g}_I$. To this end, for any function $h : [0, 1] \to \mathbb{R}$ define the increasing rearrangement on $[a, b] \subset [0, 1]$ as the function $\Gamma(h) : [a, b] \to \mathbb{R}$ with

$$\Gamma(h)(x) = \inf\left\{ z \in \mathbb{R} \,\Big|\, a + \int_a^b I\{h(t) \leq z\}\, dt \geq x \right\}.$$

Denote by Γ_n the operator Γ with $[a, b] = I_n$. We define the increasing rearrangement of $\tilde{g}$ as $\tilde{g}_I = \Gamma_n(\tilde{g})$, so that $\tilde{g}_I = \tilde{g}$ if $\tilde{g}$ is nondecreasing (see Anevski and Fougères [2], and Chernozhukov et al. [5]). We now consider residuals obtained from the monotone estimator: $\hat{\varepsilon}_{I,i} = Y_i - \tilde{g}_I(\frac{i}{n})$, $i = 1, \ldots, n$. Under the null hypothesis, these residuals should be approximately i.i.d., whereas under the alternative they show a varying behavior for $\frac{i}{n}$ in different subintervals of $[0, 1]$. For illustration, see Figure 5 where we have generated a data set (upper panel) with true non-monotone boundary curve g (dashed curve). The solid curve is the increasing rearrangement g_I. The lower left panel shows the errors ε_i, $i = 1, \ldots, n$, with i.i.d.-behaviour. The lower right panel shows $\varepsilon_{I,i} = Y_i - g_I(\frac{i}{n})$, $i = 1, \ldots, n$, with a clear non-i.i.d. pattern.

Similarly as in Section 4.2, we compare the sequential empirical distribution function

$$\tilde{F}_{I,n}(y, s) = \frac{1}{m_n} \sum_{j=1}^{\lfloor ns \rfloor} I\{\tilde{\varepsilon}_{I,j} \leq y\} I\left\{ \frac{j}{n} \in I_n \right\}$$

based on the increasing estimator $\tilde{g}_I$ with the product estimator $\bar{s}_n \tilde{F}_n(y)$, where again $I_n := [h_n + b_n, 1 - h_n - b_n]$ and $m_n := n - 2\lceil n(h_n + b_n) \rceil + 1$ and $\tilde{F}_n$ is defined in (3.1). Let

$$\tilde{G}_n(s, y) = \sqrt{n}\big(\tilde{F}_{I,n}(y, s) - \bar{s}_n \tilde{F}_n(y)\big), \qquad s \in [0, 1], y \in \mathbb{R}.$$

To derive its limit distribution under the null hypothesis, we need additional assumptions:

(I1) Let $\inf_{x \in [0,1]} g'(x) > 0$.

(B5) Let $b_n^\beta = o(n^{-1/(2(\alpha \wedge 1))})$ and $h_n^\beta + (\log n/(n h_n))^{1/\alpha} = o(b_n)$.

Note that under (B5), from Theorem 2.8(ii), we obtain uniform consistency of $\tilde{g}'$.

Theorem 4.4. *Assume model* (2.1) *with* (F1), (F2), (G1), (K1), (I1) *and* (B5), *and that* $\sup_{x \in [h_n, 1-h_n]} |\hat{g}(x) - g(x)| = o_P(n^{-1/(2(\alpha \wedge 1))})$ *holds. Then,*

$$\sup_{y \in \mathbb{R}, s \in [0,1]} \big| \tilde{F}_{I,n}(y, s) - \bar{s}_n F_{\lfloor ns \rfloor}(y) \big| = o_P\big(n^{-1/2}\big). \tag{4.1}$$
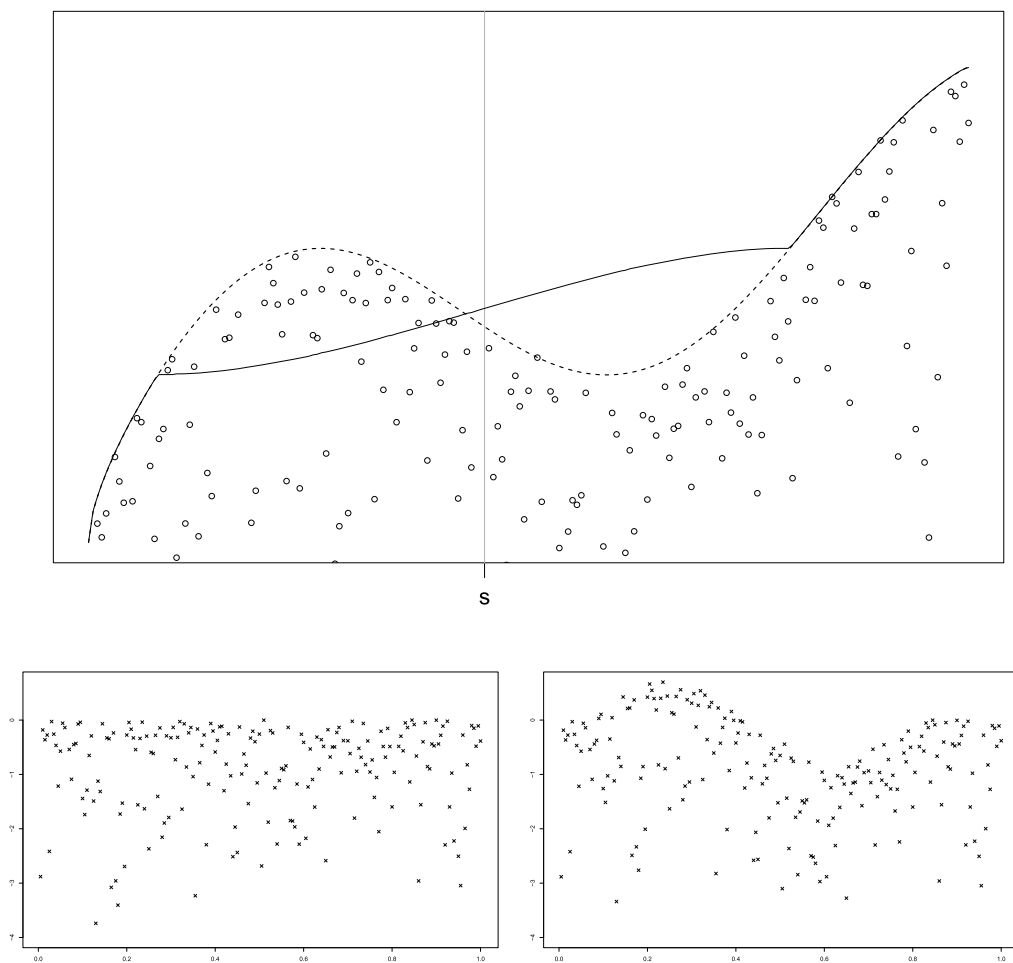
Figure 5. The upper panel shows data points and the true boundary function (dashed curve) as well as the increasing rearrangement (solid curve). The lower left panel shows the errors. The lower right panel shows residuals built with respect to the increasing rearrangement.

Thus the Kolmogorov–Smirnov test statistic $\sup_{s\in[0,1],y\in\mathbb{R}} |\tilde{G}_n(s,y)|$, converges in distribution to $\sup_{s\in[0,1],z\in[0,1]} |G(s,z)|$ where G is the completely tucked Brownian sheet (see Corollary 4.3).

Remark 4.5. A test that rejects H_0 for large values of the Kolmogorov–Smirnov test statistic $T_n = \sup_{s\in[0,1],y\in\mathbb{R}} |\tilde{G}_n(s,y)|$ is consistent. To see this, note that by Theorem 1 of Anevski and Fougères [2], $\sup_{x\in I_n} |\tilde{g}_I(x) - g_I(x)| \leq \sup_{x\in I_n} |\tilde{g}(x) - g(x)| = o_P(1)$ with g_I denoting the

increasing rearrangement of g. Thus, $n^{-1/2}T_n$ converges to

$$T = \sup_{s \in [0,1],\, y \in \mathbb{R}} \left| \int_0^s F\big(y + (g_I - g)(x)\big)\, dx - sF(y) \right|.$$

Since $T > 0$ under the alternative hypothesis $g \neq g_I$, the test statistic T_n converges to infinity.

4.4. Tests for irregularity

For some of our results (e.g., Corollary 3.9), we assume an irregular model, that is condition (F1) with $\alpha < 2$. In this subsection, we discuss tests of this assumption and related hypotheses. We focus on a rather simple test statistic which directly uses the observations Y_i locally in the neighborhood of some fixed x_0. A more refined approach based on an approximation of the tail behavior of the empirical c.d.f. of all residuals is the topic of future research.

Fix some $x_0 \in (0, 1)$ and let $\tilde{Y}_i := Y_{\lfloor nx_0 \rfloor + i}$, $1 \leq i \leq l_n \leq n - \lfloor nx_0 \rfloor$, for some intermediate sequence $(l_n)_{n \in \mathbb{N}}$, that is, $l_n \to \infty$, but $l_n/n \to 0$. Denote the corresponding order statistics by $\tilde{Y}_{1:l_n} \leq \tilde{Y}_{2:l_n} \leq \cdots \leq \tilde{Y}_{l_n:l_n}$. For another intermediate sequence $k_n = o(l_n)$ let

$$\hat{\alpha}_n := -\left[\sum_{i=2}^{k_n} d_{ni} \log \frac{\tilde{Y}_{l_n:l_n} - \tilde{Y}_{l_n-i+1:l_n}}{\tilde{Y}_{l_n:l_n} - \tilde{Y}_{l_n-k_n:l_n}} \right]^{-1},$$

with $d_{ni} = k_n^{-1} d(i/k_n)$ for some continuously differentiable function $d : [0, 1] \to \mathbb{R}$ such that

$$\int_0^1 d(t) |\log t|\, dt = -1 \tag{4.2}$$

and that there exists $\lambda > 0$ such that $d(t) = 0$ for all $t \in [0, \lambda)$.

The basic assumption (F1) on the error c.d.f. F is equivalent to

$$F^{\leftarrow}(1 - t) = -(t/c)^{1/\alpha} + \tilde{r}(t) \tag{4.3}$$

for some function $\tilde{r}(t) = o(t^{1/\alpha})$ as $t \downarrow 0$. Here $F^{\leftarrow}$ denotes the quantile function (generalized inverse) pertaining to F.

Proposition 4.6. *Suppose that the condition* (F1) *and* (G1) *hold for some* $\beta \geq 1$, *that* $\alpha < 2$,

$$k_n^{1/2 - 1/\alpha} l_n^{1+1/\alpha} = o(n) \tag{4.4}$$

and, for some $\iota > 0$,

$$k_n^{1/2 - 1/\alpha} l_n^{1/\alpha} \sup_{0 < t \leq (1+\iota)k_n/l_n} \left| \tilde{r}(t) \right| \to 0. \tag{4.5}$$

Then

$$k_n^{1/2}(\hat{\alpha}_n - \alpha) \to \alpha \int_\lambda^1 d(t) \frac{B(t)}{t}\, dt \qquad weakly \tag{4.6}$$

for a Brownian bridge B. The limit has a centered normal distribution with variance

$$\sigma_\alpha^2 = 2\alpha^2 \int_\lambda^1 \int_t^1 s^{-1} d(s) d(t) ds dt.$$

Remark 4.7. (i) In extreme value theory, it is common to assume a second order condition of the type

$$\tilde{r}(t) = O\big(t^{(1+\rho)/\alpha}\big) \qquad \text{as } t \downarrow 0$$

for some $\rho > 0$. Then condition (4.5) is fulfilled if $k_n = o(l_n^{2\rho/(2\rho+\alpha)})$. This condition can be jointly fulfilled with (4.4) if and only if $l_n^{3/2+(\alpha+1)/(2\rho)} = o(n)$. In particular, for $\alpha < 2$, this last condition is satisfied for $l_n = O(n^{2\rho/(3(\rho+1))})$.

(ii) A close inspection of the proof of Proposition 4.6 reveals that $\hat{\alpha}_n$ is a consistent estimator of α (not assumed to be less than 2), provided the basic assumptions (F1) and (G1) hold with $\beta \geq 1$ and $k_n^{-1/\alpha} l_n^{1+1/\alpha} = o(n)$.

Now tests of hypotheses on α can be easily deduced. The following exemplary result is an immediate consequence of Proposition 4.6 and Remark 4.7(ii).

Corollary 4.8. *Consider the problem of testing the null hypothesis $\alpha \in (0, \alpha_0]$ for some $\alpha_0 < 2$. If the conditions of Proposition 4.6 are fulfilled with $\alpha = \alpha_0$, then the test that rejects the null hypothesis if $\hat{\alpha}_n > \alpha_0 + k_n^{-1/2}\sigma_{\alpha_0}\Phi^{-1}(1-\tau)$ is consistent and has asymptotic size $\tau \in (0, 1)$.*

Appendix: Proofs

A.1. Auxiliary results

Proposition A.1. *Assume that model* (2.1) *holds and that the regression function g fulfills condition* (G1) *for some $\beta \in (0, \beta^* + 1]$ and some $c_g \in [0, c^*]$. Then there exist constants $L_{\beta^*, c^*}, L_{\beta^*} > 0$ and a natural number j_{β^*} (depending only on the respective subscripts) such that*

$$\big|\hat{g}(x) - g(x)\big| \leq L_{\beta^*, c^*} h_n^\beta + L_{\beta^*} \max_{1 \leq j \leq 2j_{\beta^*}} \Big(\min_{i: -1+(j-1)/j_{\beta^*} \leq (i/n-x)/h_n \leq -1+j/j_{\beta^*}} |\varepsilon_i| \Big).$$

This proposition can be verified by an obvious modification of the proof of Theorem 3.1 by Jirak et al. [19].

Lemma A.2. *Under assumptions* (F1) *and* (H1) *for any fixed set $I_1, \ldots, I_m$ of disjoint non-degenerate subintervals of $[-1, 1]$ we have*

$$\sup_{x \in [h_n, 1-h_n]} \max_{1 \leq j \leq m} \min_{\substack{i \in \{1, \ldots, n\} \\ (i/n-x)/h_n \in I_j}} |\varepsilon_i| = O_P\left(\left(\frac{|\log h_n|}{nh_n} \right)^{1/\alpha} \right).$$

Proof. Let $r_n := (|\log h_n|/(nh_n))^{1/\alpha}$. Obviously it suffices to prove that for all non-degenerate subintervals $I \subset [-1, 1]$ there exists a constant L such that

$$\lim_{n\to\infty} P\left\{\sup_{\substack{x\in[h_n,1-h_n]}} \min_{\substack{i\in\{1,\dots,n\}\\(i/n-x)/h_n\in I}} |\varepsilon_i| > Lr_n\right\} = 0.$$

Denote by $d = \sup I - \inf I > 0$ the diameter of I and let $d_n := \lceil nh_n d\rceil - 1$ and $l_n := \lfloor n/d_n\rfloor$. Then for all $y > 0$

$$P\left\{\sup_{\substack{x\in[h_n,1-h_n]}} \min_{\substack{i\in\{1,\dots,n\}\\(i/n-x)/h_n\in I}} |\varepsilon_i| > y\right\}$$

$$\leq P\left\{\max_{j\in\{1,\dots,n-d_n\}} \min_{i\in\{j,\dots,j+d_n\}} |\varepsilon_i| > y\right\} \tag{A.1}$$

$$\leq P\left\{\max_{\substack{l\in\{0,\dots,l_n\}\\l \text{ even}}} M_{n,l} > y\right\} + P\left\{\max_{\substack{l\in\{0,\dots,l_n\}\\l \text{ odd}}} M_{n,l} > y\right\}$$

with

$$M_{n,l} := \max_{j\in\{ld_n+1,\dots,(l+1)d_n\}} \min_{i\in\{j,\dots,j+d_n\}} |\varepsilon_i|.$$

Since the random variables $M_{n,l}$ for l even are i.i.d., we have

$$P\left\{\max_{\substack{l\in\{0,\dots,l_n\}\\l \text{ even}}} M_{n,l} > y\right\} = 1 - \left(1 - P\{M_{n,0} > y\}\right)^{\lfloor l_n/2\rfloor+1},$$

and an analogous equation holds for the maxima over the odd numbered block maxima $M_{n,l}$.

Let G be the c.d.f. of $|\varepsilon_i|$. If $M_{n,0}$ exceeds y, then there is a smallest index $j \in \{1, \dots, d_n\}$ for which $\min_{i\in\{j,\dots,j+d_n\}} |\varepsilon_i| > y$. Hence,

$$P\{M_{n,0} > y\} = P\left\{\min_{i\in\{1,\dots,1+d_n\}} |\varepsilon_i| > y\right\} + \sum_{j=2}^{d_n} P\left\{|\varepsilon_{j-1}| \leq y, \min_{i\in\{j,\dots,j+d_n\}} |\varepsilon_i| > y\right\}$$

$$= \left(1 - G(y)\right)^{d_n+1} + (d_n - 1)G(y)\left(1 - G(y)\right)^{d_n+1}$$

$$\leq \left(1 + d_n G(y)\right)\left(1 - G(y)\right)^{d_n}.$$

To sum up, we have shown that

$$P\left\{\sup_{\substack{x\in[h_n,1-h_n]}} \min_{\substack{i\in\{1,\dots,n\}\\(i/n-x)/h_n\in I}} |\varepsilon_i| > Lr_n\right\} \leq 2\left(1 - \left(1 - \left(1 + d_n G(Lr_n)\right)\left(1 - G(Lr_n)\right)^{d_n}\right)^{\lfloor l_n/2\rfloor+1}\right).$$

It remains to be shown that the right hand side tends to 0 for sufficiently large L which is true if and only if

$$\left(1 + d_n G(Lr_n)\right)\left(1 - G(Lr_n)\right)^{d_n} = o(1/l_n).$$

This is an immediate consequence of $1/l_n \sim dh_n$ and

$$G(Lr_n) = cL^\alpha \frac{|\log h_n|}{nh_n}(1 + o(1))$$

$$\implies \quad (1 - G(Lr_n))^{d_n} = \exp\left(-nh_n dcL^\alpha \frac{|\log h_n|}{nh_n}(1 + o(1))\right)$$

$$\implies \quad (1 + d_n G(Lr_n))(1 - G(Lr_n))^{d_n}$$
$$= O\left(|\log h_n| \exp\left(-cdL^\alpha |\log h_n|(1 + o(1))\right)\right) = o(h_n)$$

if $cdL^\alpha > 1$. □

A.2. Proof of Theorem 2.2

The assertion directly follows from Proposition A.1 and Lemma A.2.

A.3. Proof of Theorem 2.8

(i) Using Theorem 2.2, a Taylor expansion of g of order $\lfloor \beta \rfloor$ and assumption (K1), one can show by direct calculations that for some $\tau_u \in (0, 1)$

$$\sup_{x \in I_n} |\tilde{g}(x) - g(x)|$$

$$\leq \sup_{x \in I_n} \left| \int_{h_n}^{1-h_n} (\hat{g}(z) - g(z)) \frac{1}{b_n} K\left(\frac{x - z}{b_n}\right) dz \right|$$

$$\quad + \sup_{x \in I_n} \left| \int_{h_n}^{1-h_n} (g(z) - g(x)) \frac{1}{b_n} K\left(\frac{x - z}{b_n}\right) dz \right|$$

$$\leq \sup_{z \in [h_n, 1-h_n]} |\hat{g}(z) - g(z)| O(1) \tag{A.2}$$

$$\quad + \sup_{x \in I_n} \left| \int_{-1}^{1} (g(x - ub_n) - g(x)) K(u) \, du \right|$$

$$\leq O(h_n^\beta) + O_P\left(\left(\frac{|\log h_n|}{nh_n}\right)^{\frac{1}{\alpha}}\right)$$

$$\quad + b_n^{\lfloor \beta \rfloor} \sup_{x \in I_n} \left| \frac{1}{\lfloor \beta \rfloor!} \int_{-1}^{1} u^{\lfloor \beta \rfloor} (g^{(\lfloor \beta \rfloor)}(x - \tau_u ub_n) - g^{(\lfloor \beta \rfloor)}(x)) K(u) \, du \right|. \tag{A.3}$$

Now the Hölder property of g combined with (K1) yields the desired result.

(ii) Since g is bounded on $[h_n, 1 - h_n]$ and $\sup_{x \in [h_n, 1-h_n]} |\hat{g}(x) - g(x)| = o_P(1)$, $\hat{g}$ is eventually bounded on $[h_n, 1 - h_n]$ too. Note that the partial derivative of $\hat{g}(z)b_n^{-1}K((x - z)/b_n)$ with respect to x is continuous and bounded (for fixed n). Thus, we can exchange integration and differentiation and obtain

$$\sup_{x \in I_n} |\tilde{g}'(x) - g'(x)|$$

$$= \sup_{x \in I_n} \left| \int_{h_n}^{1-h_n} \hat{g}(z) \frac{1}{b_n^2} K'\left(\frac{x - z}{b_n} \right) dz - g'(x) \right|.$$

Integration by parts yields

$$\int_{h_n}^{1-h_n} g(z) \frac{1}{b_n^2} K'\left(\frac{x - z}{b_n} \right) dz$$

$$= \int_{h_n}^{1-h_n} g'(z) \frac{1}{b_n} K\left(\frac{x - z}{b_n} \right) dz$$

since $K(-1) = K(1) = 0$. Therefore,

$$\sup_{x \in I_n} |\tilde{g}'(x) - g'(x)| \leq \sup_{x \in I_n} \left| \int_{h_n}^{1-h_n} (\hat{g}(z) - g(z)) \frac{1}{b_n^2} K'\left(\frac{x - z}{b_n} \right) dz \right|$$

$$+ \sup_{x \in I_n} \left| \int_{h_n}^{1-h_n} (g'(z) - g'(x)) \frac{1}{b_n} K\left(\frac{x - z}{b_n} \right) dz \right|$$

$$\leq \sup_{z \in [h_n, 1-h_n]} |\hat{g}(z) - g(z)| O(b_n^{-1})$$

$$+ \sup_{x \in I_n} \left| \int_{-1}^{1} (g'(x - ub_n) - g'(x)) K(u) du \right|.$$

Similarly as in the proof of (i), assertion (ii) follows by Theorem 2.2, a Taylor expansion of g' of order $\lfloor \beta \rfloor - 1$ and the Assumptions (K1) and (G1).

(iii) We distinguish the cases $|x - y| > a_n$ and $|x - y| \leq a_n$ for some suitable sequence $(a_n)_{n \in \mathbb{N}}$ with $\lim_{n \to \infty} a_n = 0$ specified later. In the first case, we obtain

$$\sup_{x, y \in I_n, |x-y| > a_n} \frac{|\tilde{g}'(x) - g'(x) - \tilde{g}'(y) + g'(y)|}{|x - y|^\delta}$$

$$\leq 2 \sup_{x \in I_n} |\tilde{g}'(x) - g'(x)| a_n^{-\delta} \tag{A.4}$$

$$= \left(O(b_n^{\beta-1}) + \left(O(h_n^\beta) + O_P\left(\left(\frac{|\log h_n|}{nh_n} \right)^{\frac{1}{\alpha}} \right) \right) b_n^{-1} \right) a_n^{-\delta}.$$

Let T denote an index set. P^* and E^* denote outer probability and outer expectation. Weak convergence of a sequence $(X_n)_{n\in\mathbb{N}}$ in $\ell^\infty(T)$ to a tight limit X is equivalent to convergence of the finite dimensional distributions and the existence of a semimetric ρ on T, such that T is totally bounded and

$$\forall \epsilon, \eta > 0 \; \exists \delta > 0 : \limsup_{n\to\infty} P^*\left(\sup_{\substack{s,t\in T \\ \rho(s,t)<\delta}} \left|X_n(s) - X_n(t)\right| > \epsilon \right) < \eta.$$

The latter condition means that $(X_n)_{n\in\mathbb{N}}$ is asymptotically equicontinuous with respect to ρ. If it holds, then for sequences $\delta_n \searrow 0$ one has

$$\sup_{\substack{s,t\in T \\ \rho(s,t)<\delta_n}} \left|X_n(s) - X_n(t)\right| = o_P(1). \tag{A.6}$$

For stochastic processes $Z_{n1}, \ldots, Z_{nn}$ with index set $\mathcal{F}$ and for $\epsilon > 0$ the bracketing number $N_{[]}(\epsilon, \mathcal{F}, L_2^n)$ is defined as the minimal number N_ϵ of sets in a partition $\mathcal{F} = \bigcup_{j=1}^{N_\epsilon} \mathcal{F}_{\epsilon j}^n$ of the index set $\mathcal{F}$ such that

$$\sum_{i=1}^n E\left[\sup_{\varphi,\psi\in\mathcal{F}_{\epsilon j}^n} \left|Z_{ni}(\varphi) - Z_{ni}(\psi)\right|^2 \right] \le \epsilon^2$$

for every partitioning set $\mathcal{F}_{\epsilon j}^n$. We now state Theorem 2.11.9 of van der Vaart and Wellner [36] (p. 211) that will be applied in Section A.6.

Theorem. *For each $n \in \mathbb{N}$ let $Z_{n1}, \ldots, Z_{nn}$ be independent stochastic processes with finite second moments indexed by a totally bounded semimetric space $(\mathcal{F}, \rho)$. Suppose that*

$$\sum_{i=1}^n E^*\left[\sup_{\varphi\in\mathcal{F}} \left|Z_{ni}(\varphi)\right| I\left\{ \sup_{\varphi\in\mathcal{F}} \left|Z_{ni}(\varphi)\right| > \eta \right\} \right] \to 0 \qquad \text{for all } \eta > 0, \tag{A.7}$$

$$\sup_{\rho(\varphi,\psi)<\delta_n} \sum_{i=1}^n E\left[\left(Z_{ni}(\varphi) - Z_{ni}(\psi)\right)^2\right] \to 0 \qquad \text{for every sequence } \delta_n \searrow 0, \tag{A.8}$$

$$\int_0^{\delta_n} \sqrt{\log N_{[]}\left(\epsilon, \mathcal{F}, L_2^n\right)}\, d\epsilon \to 0 \qquad \text{for every sequence } \delta_n \searrow 0. \tag{A.9}$$

Then the sequence $\sum_{i=1}^n (Z_{ni} - E[Z_{ni}])$ converges weakly in $\ell^\infty(\mathcal{F})$ provided it converges marginally.

A.5. Proof of Theorem 3.1

By our assumptions, there exists a sequence $a_n = o(n^{-1/(2(\alpha\wedge 1))})$ such that

$$P\left(\sup_{x\in[h_n, 1-h_n]} \left|\hat{g}(x) - g(x)\right| \le a_n \right) \xrightarrow[n\to\infty]{} 1.$$

For this rate a_n we then have by Hölder continuity of F,

$$\left| F(y + a_n) - F(y) \right| = O\left(a_n^{\alpha \wedge 1}\right) = o\left(n^{-1/2}\right). \tag{A.10}$$

Let $\bar{F}_n(y, s) := \frac{1}{m_n} \sum_{j=1}^{\lfloor ns \rfloor} I\{\varepsilon_j \leq y\} I\{h_n < \frac{j}{n} \leq 1 - h_n\}$. Since

$$\hat{F}_n(y, s) = \frac{1}{m_n} \sum_{j=1}^{\lfloor ns \rfloor} I\left\{\varepsilon_j \leq y + (\hat{g} - g)\left(\frac{j}{n}\right)\right\} I\left\{h_n < \frac{j}{n} \leq 1 - h_n\right\}$$

we may conclude

$$\sqrt{n}\left(\bar{F}_n(y - a_n, s) - \bar{s}_n F_{\lfloor ns \rfloor}(y)\right) \leq \sqrt{n}\left(\hat{F}_n(y, s) - \bar{s}_n F_{\lfloor ns \rfloor}(y)\right)$$

$$\leq \sqrt{n}\left(\bar{F}_n(y + a_n, s) - \bar{s}_n F_{\lfloor ns \rfloor}(y)\right)$$

for all $y \in \mathbb{R}$ and $s \in [0, 1]$ with probability converging to 1.

We take a closer look at the bounds. The sequential empirical process

$$E_n(y, s) = n^{-1/2} \sum_{j=1}^{\lfloor ns \rfloor} \left(I\{\varepsilon_j \leq y\} - F(y)\right), \qquad y \in \mathbb{R}, s \in [0, 1], \tag{A.11}$$

converges weakly to a Kiefer process; see, for example, Theorem 2.12.1 in van der Vaart and Wellner [36]. Now, $n \sim m_n$, the asymptotic equicontinuity of the process E_n (see (A.6)), the Hölder continuity (F2) and (A.10) imply

$$\sqrt{n}\left(\bar{F}_n(y \pm a_n, s) - \bar{s}_n F_{\lfloor ns \rfloor}(y)\right)$$

$$= \frac{n}{m_n}\left(E_n\left(y \pm a_n, s \wedge (1 - h_n)\right) - E_n\left(y, s \wedge (1 - h_n)\right)\right.$$

$$\left. - E_n\left(y \pm a_n, s \wedge h_n\right) + E_n(y, s \wedge h_n)\right)$$

$$+ \sqrt{n}\bar{s}_n\left(F(y \pm a_n) - F(y)\right) + \sqrt{n}\left(\bar{F}_n(y, s) - \bar{s}_n F_{\lfloor ns \rfloor}(y)\right)$$

$$= o_P(1) + \sqrt{n}\left(\bar{F}_n(y, s) - \bar{s}_n F_{\lfloor ns \rfloor}(y)\right)$$

uniformly for all $y \in \mathbb{R}$, $s \in [0, 1]$.

It remains to be shown that

$$\sqrt{n}\left(\bar{F}_n(y, s) - \bar{s}_n F_{\lfloor ns \rfloor}(y)\right)$$

$$= \frac{\sqrt{n}}{m_n} \sum_{j=1}^{\lfloor ns \rfloor} \left(I\{\varepsilon_j \leq y\} - F(y)\right) I\left\{h_n < \frac{j}{n} \leq 1 - h_n\right\} - \frac{\sqrt{n}\bar{s}_n}{\lfloor ns \rfloor} \sum_{j=1}^{\lfloor ns \rfloor} \left(I\{\varepsilon_j \leq y\} - F(y)\right)$$

$$= \left(\frac{n}{m_n} - 1\right)\left(E_n\left(y, s \wedge (1 - h_n)\right) - E_n(y, s \wedge h_n)\right) \tag{A.12}$$

$$- \left(\frac{n\bar{s}_n}{\lfloor ns \rfloor} - 1 \right) E_n(y, s)$$

$$+ \left(E_n\big(y, s \wedge (1 - h_n)\big) - E_n(y, s \wedge h_n) - E_n(y, s) \right)$$

tends to 0 in probability uniformly for all $y \in \mathbb{R}$, $s \in [0, 1]$.

The first term vanishes asymptotically, because E_n is uniformly stochastically bounded and $n \sim m_n$.

Next note that $\bar{s}_n = 0$ for $s < h_n$, while for $s \geq h_n$

$$\frac{n\bar{s}_n}{\lfloor ns \rfloor} - 1 = \frac{\lfloor n(s \wedge (1 - h_n)) \rfloor - \lfloor nh_n \rfloor}{(1 - 2h_n + O(n^{-1}))\lfloor ns \rfloor} - 1 = \frac{O(nh_n)}{(1 - 2h_n + O(n^{-1}))\lfloor ns \rfloor}, \tag{A.13}$$

which is uniformly bounded for all $s \in [h_n, 1]$ and tends to 0 uniformly with respect to $s \in [h_n^{1/2}, 1]$. Moreover, E_n is uniformly stochastically bounded and $\sup_{0 \leq s \leq h_n^{1/2}, y \in \mathbb{R}} |E_n(y, s)| = o_P(1)$, because E_n is asymptotically equicontinuous with $E_n(y, 0) = 0$ (cf. (A.6)). Hence, the second term in (A.12) converges to 0 in probability, too. Likewise, the convergence of the last term to 0 follows from the asymptotic equicontinuity of E_n and (A.6). This concludes the proof.

A.6. Proof of Theorem 3.4 and of Remark 3.5

For any interval $I \subset \mathbb{R}$ and constants $k > 0$, $\delta \in (0, 1)$, define the following class of differentiable functions:

$$C_k^{1+\delta}(I) = \left\{ d : I \to \mathbb{R} \,\middle|\, \max\left\{ \sup_{x \in I} |d(x)|, \sup_{x \in I} |d'(x)|, \sup_{x, y \in I, x \neq y} \frac{|d'(x) - d'(y)|}{|x - y|^\delta} \right\} \leq k \right\}.$$

Then Theorem 2.8 yields $P((\tilde{g} - g) \in C_{1/2}^{1+\delta}(I_n)) \to 1$ as $n \to \infty$. Hence, there exist random functions $d_n : [0, 1] \to \mathbb{R}$ such that $d_n(x) = (\tilde{g} - g)(x)$ for all $x \in I_n$ and $P(d_n \in C_1^{1+\delta}([0, 1])) \to 1$ for $n \to \infty$. (For instance, one may extrapolate $\tilde{g} - g$ linearly on $[0, h_n]$ and on $[1 - h_n, 1]$.)

On the space $\mathcal{F} := \mathbb{R} \times C_1^{1+\delta}([0, 1])$ we define the semimetric

$$\rho\big((y, d), (y^*, d^*)\big)$$

$$= \max\left\{ \sup_{x \in [0,1]} \sup_{\gamma \in C_1^{1+\delta}([0,1])} \left| F\big(y + \gamma(x)\big) - F\big(y^* + \gamma(x)\big) \right|, \sup_{x \in [0,1]} \left| d(x) - d^*(x) \right| \right\}.$$

For $\varphi = (y, d) \in \mathcal{F}$ let

$$Z_{nj}(\varphi) := \frac{\sqrt{n}}{m_n} I\left\{ \varepsilon_j \leq y + d\left(\frac{j}{n}\right) \right\} I\left\{ \frac{j}{n} \in I_n \right\} - \frac{1}{\sqrt{n}} I\{\varepsilon_j \leq y\}$$

and

$$G_n(\varphi) := \sum_{j=1}^{n} \big(Z_{nj}(\varphi) - E[Z_{nj}(\varphi)] \big).$$

Note that

$$G_n(y, d_n) = \frac{\sqrt{n}}{m_n} \sum_{j=1}^{n} I\{\varepsilon_j \leq y - g(j/n) + \tilde{g}(j/n)\} I\{j/n \in I_n\}$$

$$- \frac{\sqrt{n}}{m_n} \sum_{j=1}^{n} F(y + (\tilde{g} - g)(j/n)) I\{j/n \in I_n\} - \frac{1}{\sqrt{n}} \sum_{j=1}^{n} I\{\varepsilon_j \leq y\} + \sqrt{n} F(y)$$

$$= \sqrt{n} \left(\tilde{F}_n(y) - \frac{1}{n} \sum_{j=1}^{n} I\{\varepsilon_j \leq y\} \right.$$

$$\left. - \frac{1}{m_n} \sum_{j=1}^{n} \left(F\left(y + (\tilde{g} - g)\left(\frac{j}{n}\right) \right) - F(y) \right) I\left\{ \frac{j}{n} \in I_n \right\} \right).$$

In the following, we will apply Theorem 2.11.9 of van der Vaart and Wellner [36] (see Section A.4) to show that the process $(G_n(\varphi))_{\varphi \in \mathcal{F}}$ converges to a (Gaussian) limiting process. In particular, G_n is asymptotically equicontinuous. Note that for

$$G_n(y, 0) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\left(\frac{n}{m_n} - 1 \right) I\{\varepsilon_j \leq y\} I\{j/n \in I_n\} - I\{\varepsilon_j \leq y\} I\{j/n \notin I_n\} \right)$$

we have

$$\mathrm{Var}(G_n(y, 0)) = \frac{1}{n} \sum_{j=1}^{n} \left(\left(\frac{n}{m_n} - 1 \right) I\{j/n \in I_n\} - I\{j/n \notin I_n\} \right)^2 F(y)(1 - F(y))$$

$$\leq \frac{1}{2} \left(\left(\frac{n}{m_n} - 1 \right)^2 + \frac{n - m_n}{n} \right) \to 0$$

for $n \to \infty$. Now note that $\sup_{y \in \mathbb{R}} \rho((y, d_n), (y, 0)) = \sup_{x \in [0,1]} |d_n(x)| = o_P(1)$. From the asymptotic equicontinuity of G_n it follows that $G_n(y, d_n) = o_P(1)$ uniformly in y (see (A.6)) and the assertion holds.

One may proceed as in the proof of Lemma 3 in Neumeyer and Van Keilegom [27] (see the online supporting information to that article) to prove that the conditions of Theorem 2.11.9 of van der Vaart and Wellner [36] are fulfilled. The proof of (A.7) and (A.8) are analogous. The only difference is that Neumeyer and Van Keilegom [27] assume a bounded error density while we use Hölder continuity of F, see assumption (F2). Next, we show that the bracketing entropy condition (A.9) is fulfilled and that $(\mathcal{F}, \rho)$ is totally bounded.

To this end, let $d_m^L \leq d_m^U$, $m = 1, \ldots, M$, be brackets for $C_1^{1+\delta}([0, 1])$ of length $\eta^{2/(\alpha \wedge 1)}$ w.r.t. the supremum norm. According to van der Vaart and Wellner [36], Theorem 2.7.1 and Corollary 2.7.2, $M = O(\exp(\kappa \eta^{-2/((1+\delta)(\alpha \wedge 1))}))$ brackets are needed. For each m define $F_m^L(y) := n^{-1} \sum_{j=1}^{n} F(y + d_m^L(j/n))$ and choose $y_{m,k}^L$, $k = 1, \ldots, K = O(\eta^{-2})$ such that $F_m^L(y_{m,k}^L) - F_m^L(y_{m,k-1}^L) < \eta^2$ for all $k \in \{1, \ldots, K + 1\}$ with $y_{m,0}^L := -\infty$ and $y_{m,K+1}^L := \infty$.

Define F_m^U and $y_{m,k}^U$ analogously, $\tilde{y}_{m,k}^L := y_{m,k}^L$ and denote by $\tilde{y}_{m,k}^U$ the smallest $y_{m,l}^U$ larger than or equal to $y_{m,k+1}^L$. Then $\mathcal{F}$ is covered by

$$\mathcal{F}_{mk} := \left\{ (y,d) \in \mathcal{F} \mid \tilde{y}_{m,k}^L \leq y \leq \tilde{y}_{m,k}^U, d_m^L \leq d \leq d_m^U \right\}, \qquad m = 1, \ldots, M, k = 1, \ldots, K.$$

Check that by condition (F2)

$$\sup_{y \in \mathbb{R}} \left| F_m^U(y) - F_m^L(y) \right| \leq \sup_{y \in \mathbb{R}} n^{-1} \sum_{j=1}^n \left| F\left(y + d_m^U(j/n) \right) - F\left(y + d_m^L(j/n) \right) \right|$$

$$\leq L_F \sup_{x \in \mathbb{R}} \left| d_m^U(x) - d_m^L(x) \right|^{\alpha \wedge 1} \leq L_F \eta^2 \tag{A.14}$$

with L_F denoting the Hölder constant of F. Thus,

$$\frac{1}{n} \sum_{j=1}^n E \left[\sup_{(y,d),(y^*,d^*) \in \mathcal{F}_{mk}} \left| I\{\varepsilon_j \leq y + d(j/n)\} - I\{\varepsilon_j \leq y^* + d^*(j/n)\} \right| \right]^2$$

$$\leq \frac{1}{n} \sum_{j=1}^n E \left[I\{\varepsilon_j \leq \tilde{y}_{m,k}^U + d_m^U(j/n)\} - I\{\varepsilon_j \leq \tilde{y}_{m,k}^L + d_m^L(j/n)\} \right]^2$$

$$\leq F_m^U(\tilde{y}_{m,k}^U) - F_m^L(\tilde{y}_{m,k}^L)$$

$$\leq \left| F_m^U(\tilde{y}_{m,k}^U) - F_m^U(\tilde{y}_{m,k+1}^L) \right| + \left| F_m^U(\tilde{y}_{m,k+1}^L) - F_m^L(\tilde{y}_{m,k+1}^L) \right| + \left| F_m^L(\tilde{y}_{m,k+1}^L) - F_m^L(\tilde{y}_{m,k}^L) \right|$$

$$\leq (2 + L_F)\eta^2,$$

where the last step follows from (A.14) and the definitions of $\tilde{y}_{m,k}^L$ and $\tilde{y}_{m,k}^U$. Hence we obtain for the squared diameter of $\mathcal{F}_{mk}$ w.r.t. L_2^n

$$\sum_{j=1}^n E \left[\sup_{(y,d),(y^*,d^*) \in \mathcal{F}_{mk}} \left| Z_{nj}(y,d) - Z_{nj}(y^*,d^*) \right| \right]^2$$

$$\leq 2 \frac{n}{m_n^2} \sum_{j=1}^n E \left[\sup_{(y,d),(y^*,d^*) \in \mathcal{F}_{mk}} \left| I\{\varepsilon_j \leq y + d(j/n)\} - I\{\varepsilon_j \leq y^* + d^*(j/n)\} \right| \right]^2 I\{j/n \in I_n\}$$

$$+ \frac{2}{n} \sum_{j=1}^n E \left[\sup_{(y,d),(y^*,d^*) \in \mathcal{F}_{mk}} \left| I\{\varepsilon_j \leq y\} - I\{\varepsilon_j \leq y^*\} \right| \right]^2$$

$$\leq 3(2 + L_F)\eta^2$$

for sufficiently large n. This shows that the bracketing number satisfies $\log N_{[]}(\eta, \mathcal{F}, L_2^n) = O(\log M + \log K) = O(\eta^{-2/((1+\delta)(\alpha \wedge 1))})$, and (A.9) follows from $\delta > 1/\alpha - 1$.

It remains to show that $(\mathcal{F}, \rho)$ is totally bounded, that is, that, for all $\eta \in (0,1)$, the space $\mathcal{F}$ can be covered by finitely many sets with ρ-diameter less than 5η. To this end, choose d_m^L

and d_m^U as above. For each $m \in \{1, \ldots, M\}$ and $j \in \{0, \ldots, J := \lceil \eta^{-1} \rceil\}$, let $s_j := j\eta^{1/(\alpha \wedge 1)} \wedge 1$ and $F_{jm}(y) := P(\varepsilon_1 \leq y + d_m^L(s_j))$, and choose an increasing sequence $y_{jm,k}$, $k = 1, \ldots, K := \lfloor \eta^{-1} \rfloor$, such that $F_{jm}(y_{jm,k}) - F_{jm}(y_{jm,k-1}) < \eta$ for all $k \in \{1, \ldots, K+1\}$ with $y_{jm,0} := -\infty$ and $y_{jm,K+1} := \infty$. Denote by $\bar{y}_l$, $1 \leq l \leq L$, all points $y_{jm,k}$, $j \in \{0, \ldots, J\}$, $m \in \{1, \ldots, M\}$, $k \in \{1, \ldots, K\}$, in increasing order. We show that all sets $\mathcal{F}_{ml} := \{(y, d) | \bar{y}_{l-1} \leq y \leq \bar{y}_l, d_m^L \leq d \leq d_m^U\}$ have ρ-diameter less than 5η. Check that, for all $1 \leq l \leq L$, one has

$$
\sup_{x \in [0,1]} \sup_{\gamma \in C_1^{1+\delta}([0,1])} \left| F\left(\bar{y}_l + \gamma(x)\right) - F\left(\bar{y}_{l-1} + \gamma(x)\right) \right|
$$

$$
\begin{aligned}
\leq \max_{1 \leq j \leq J} \sup_{s_{j-1} \leq x \leq s_j} \max_{1 \leq m \leq M} \sup_{d_m^L \leq \gamma \leq d_m^U} \Big[&\left| F\left(\bar{y}_l + \gamma(x)\right) - F\left(\bar{y}_l + \gamma(s_j)\right) \right| \\
&+ \left| F\left(\bar{y}_l + \gamma(s_j)\right) - F\left(\bar{y}_l + d_m^L(s_j)\right) \right| \\
&+ \left| F\left(\bar{y}_l + d_m^L(s_j)\right) - F\left(\bar{y}_{l-1} + d_m^L(s_j)\right) \right| \\
&+ \left| F\left(\bar{y}_{l-1} + d_m^L(s_j)\right) - F\left(\bar{y}_{l-1} + \gamma(s_j)\right) \right| + \left| F\left(\bar{y}_{l-1} + \gamma(s_j)\right) - F\left(\bar{y}_{l-1} + \gamma(x)\right) \right| \Big] \\
< \max_{1 \leq j \leq J} &\left[(s_j - s_{j-1})^{\alpha \wedge 1} + \eta^2 + \eta + \eta^2 + (s_j - s_{j-1})^{\alpha \wedge 1} \right] \\
\leq 5\eta. &
\end{aligned}
\tag{A.15}
$$

Therefore, for all $(y, d), (y^*, d^*) \in \mathcal{F}_{ml}$

$$
\rho\left((y, d), (y^*, d^*)\right)
$$

$$
\leq \max \left\{ \sup_{x \in [0,1]} \sup_{\gamma \in C_1^{1+\delta}([0,1])} \left| F\left(\bar{y}_l + \gamma(x)\right) - F\left(\bar{y}_{l-1} + \gamma(x)\right) \right|, \ \sup_{x \in [0,1]} d_m^U(x) - d_m^L(x) \right\}
$$

$$
\leq \max \left\{ 5\eta, \eta^{2/(\alpha \wedge 1)} \right\} = 5\eta,
$$

which concludes the proof of Theorem 3.4.

If we drop the assumption $\delta > 1/\alpha - 1$ but require F to be Lipschitz continuous, then we use brackets for $C_1^{1+\delta}([0, 1])$ of length η^2 (instead of $\eta^{2/(\alpha \wedge 1)}$) and replace (A.14) with

$$
\sup_{y \in \mathbb{R}} \left| F_m^U(y) - F_m^L(y) \right| \leq \sup_{y \in \mathbb{R}} n^{-1} \sum_{j=1}^{n} \left| F\left(y + d_m^U(j/n)\right) - F\left(y + d_m^L(j/n)\right) \right|
$$

$$
\leq L_F \sup_{x \in \mathbb{R}} \left| d_m^U(x) - d_m^L(x) \right| \leq L_F \eta^2
$$

with L_F denoting the Lipschitz constant of F to prove $\log N_{[]}(\eta, \mathcal{F}, L_2^n) = O(\eta^{-2/(1+\delta)})$, which again yields (A.9). Likewise, in the last part of the proof, one defines $s_j := j\eta \wedge 1$ and replaces (A.15) with $\max_{1 \leq j \leq J} [(s_j - s_{j-1}) + \eta^2 + \eta + \eta^2 + (s_j - s_{j-1})] \leq 5\eta$.

In the remaining proofs to Section 3, we use the index n for the estimators to emphasize the dependence on the sample size and to distinguish between estimators and polynomials corre-

sponding to a given sample on the one hand and corresponding objects in a limiting setting on the other hand.

A.7. Proof of Lemma 3.6

Proposition A.1 and the proof of Lemma A.2 show that there exist constants $d, \tilde{d} > 0$ depending only on β and c_g such that $E(\hat{g}_n(x)) \leq \tilde{d} E(M_{n,0})$ and $P\{M_{n,0} > t\} \leq (1 + dnh_n(1 - F(-t)))(F(-t))^{dnh_n}$ for all $t > 0$.

Let $a_n := a(\log n/(nh_n))^{1/\alpha}$ for a suitable constant $a > 0$ and fix some $t_0 > 0$ such that $(1 - F(-t))/(ct^{\alpha}) \in (1/2, 2)$ for all $t \in (0, t_0]$. Then

$$
\begin{aligned}
E(M_{n,0}) \\
= \int_0^{\infty} P\{M_{n,0} > t\} \, dt \\
\leq a_n + \int_{a_n}^{t_0} \left(1 + dnh_n(1 - F(-t))\right)(F(-t))^{dnh_n} \, dt \\
+ (1 + dnh_n) \int_{t_0}^{\infty} (F(-t))^{dnh_n} \, dt.
\end{aligned}
$$

Now, for sufficiently large n,

$$
\begin{aligned}
\int_{a_n}^{t_0} \left(1 + dnh_n(1 - F(-t))\right)(F(-t))^{dnh_n} \, dt \\
\leq \int_{a_n}^{t_0} \left(1 + 2cdnh_n t^{\alpha}\right)\left(1 - \frac{c}{2}t^{\alpha}\right)^{dnh_n} \, dt \\
\leq (1 + 2cd)nh_n \int_{a_n}^{t_0} t^{\alpha} \exp\left(-\frac{c}{2} dnh_n t^{\alpha}\right) dt \\
\leq (1 + 2cd)nh_n \frac{t_0}{\alpha} \int_{a_n^{\alpha}}^{t_0^{\alpha}} \exp\left(-\frac{c}{2} dnh_n u\right) du \\
\leq \frac{2(1 + 2cd)nh_n t_0}{\alpha cdnh_n} \exp\left(-\frac{c}{2} da^{\alpha} \log n\right) \\
= o(n^{-\xi})
\end{aligned}
$$

for all $\xi > 0$ if a is chosen sufficiently large. Hence the assertion follows from (H1) and (F3) which imply

$$
\int_{t_0}^{\infty} (F(-t))^{dnh_n} \, dt \leq nh_n (F(-t_0))^{dnh_n} + \int_{nh_n}^{\infty} t^{-d\tau nh_n} \, dt = o(n^{-\xi})
$$

for all $\xi > 0$.

A.8. Proof of Theorem 3.7

As the density f is bounded and Lipschitz continuous, one has

$$\left| F\left(y + (\tilde{g}_n^* - g)(j/n)\right) - F(y) - f(y)(\tilde{g}_n^* - g)(j/n) \right|$$

$$= \left| \int_0^{(\tilde{g}_n^* - g)(j/n)} f(y + t) - f(y)\, dt \right|$$

$$= O\left((\tilde{g}_n^* - g)^2(j/n)\right)$$

uniformly for $y \le y_0$ and $j/n \in I_n$. Hence, the remainder term can be approximated by a sum of estimation errors as follows:

$$\left| \frac{1}{m_n} \sum_{j=1}^n \left(F\left(y + (\tilde{g}_n^* - g)(j/n)\right) - F(y)\right) I\{j/n \in I_n\} \right.$$

$$\left. - \frac{f(y)}{m_n} \sum_{j=1}^n (\tilde{g}_n^* - g)(j/n) I\{j/n \in I_n\} \right|$$

$$= O\left(\frac{1}{m_n} \sum_{j=1}^n (\tilde{g}_n^* - g)^2(j/n) I\{j/n \in I_n\} \right)$$

$$= O_P\left(h_n^{2\beta} + b_n^{2\beta} + \left(\frac{\log n}{nh_n}\right)^{2/\alpha} \right) = o_P\left(n^{-1/2}\right),$$

where for the last conclusions we have used Theorem 2.2, Lemma 3.6 and the assumptions (H2) and (B4). Thus the assertion follows if we show that

$$\frac{1}{m_n} \sum_{j=1}^n (\tilde{g}_n^* - g)(j/n) I\{j/n \in I_n\} = o_P\left(n^{-1/2}\right).$$

To this end, note that $\tilde{g}_n^*(x)$ and $\tilde{g}_n^*(y)$ are independent for $|x - y| > 2(h_n + b_n)$. For simplicity, we assume that $2n(h_n + b_n) =: k_n$ is a natural number. If we split the whole sum into blocks with k_n consecutive summands, then all blocks with odd numbers are independent and all blocks with even numbers are independent. It suffices to show that

$$\frac{1}{m_n} \sum_{\ell=1}^{\lfloor n/(2k_n) \rfloor} \Delta_{n,2\ell-1} = o_P\left(n^{-1/2}\right),$$

$$\frac{1}{m_n} \sum_{\ell=1}^{\lfloor n/(2k_n) \rfloor} \Delta_{n,2\ell} = o_P\left(n^{-1/2}\right),$$

where $\Delta_{n,l} = \sum_{j=(l+1)k_n}^{(l+2)k_n-1} (\tilde{g}_n^* - g)(j/n)$, $1 \leq \ell \leq \lfloor n/k_n \rfloor$. We only consider the second sum, because the first convergence obviously follows by the same arguments.

It suffices to verify

$$E\left(\Delta_{n,2\ell}^2\right) = o(k_n), \tag{A.16}$$

$$E(\Delta_{n,2\ell}) = o\left(n^{-1/2}k_n\right) = o\left(n^{1/2}(h_n + b_n)\right) \tag{A.17}$$

uniformly for all $1 \leq \ell \leq \lfloor n/(2k_n) \rfloor$, since then

$$E\left(\sum_{\ell=1}^{\lfloor n/(2k_n) \rfloor} \Delta_{n,2\ell}\right)^2 = \sum_{\ell=1}^{\lfloor n/(2k_n) \rfloor} \text{Var}(\Delta_{n,2\ell}) + \left(\sum_{\ell=1}^{\lfloor n/(2k_n) \rfloor} E(\Delta_{n,2\ell})\right)^2 = o(n),$$

which implies the assertion.

To prove (A.16), note that according to the definition of $\tilde{g}_n^*$ and Lemma 3.6

$$\sup_{x \in I_n} \left|\tilde{g}_n^*(x) - g(x)\right|$$

$$\leq \sup_{x \in I_n} \left|\tilde{g}(x) - g(x)\right| + O\left(\left(\frac{\log n}{nh_n}\right)^{1/\alpha}\right)$$

$$\leq c_0 \sup_{x \in [h_n, 1-h_n]} \left|\hat{g}(x) - g(x)\right| + O\left(b_n^\beta\right) + O\left(\left(\frac{\log n}{nh_n}\right)^{1/\alpha}\right)$$

by (A.2) and (A.3) in the proof of Theorem 2.8(i). (Here and in the sequel, c_i denote generic constants depending only on β, c_g and the kernel K.) Further, by Proposition A.1 and (A.1) in the proof of Lemma A.2, we obtain

$$\sup_{x \in I_n} \left|\tilde{g}_n^*(x) - g(x)\right| \leq O\left(h_n^\beta\right) + O\left(b_n^\beta\right) + O\left(\left(\frac{\log n}{nh_n}\right)^{1/\alpha}\right) + c_1 \max\left(M_1^*, M_2^*\right),$$

where $M_1^* = \max_{l \in \{0,\ldots,l_n\}, l \text{ even}} M_{n,l}$ and $M_2^* = \max_{l \in \{0,\ldots,l_n\}, l \text{ odd}} M_{n,l}$.

Because $k_n(h_n^\beta + b_n^\beta + (\log n/(nh_n))^{1/\alpha}) = o(k_n^{1/2})$ by (H2) and (B4), it suffices to show that

$$E\left((M_i^*)^2\right) = \int_0^\infty P\left\{M_i^* > t^{1/2}\right\} dt = o(1/k_n). \tag{A.18}$$

Recall from the proof of Lemma A.2 that $P\{M_i^* > t\} \leq 1 - (1 - P\{M_{n,0} > t\})^{c_2(h_n+b_n)/h_n}$ with

$$P\{M_{n,0} > t\} \leq \left(1 + c_3 nh_n\left(1 - F(-t)\right)\right)\left(F(-t)\right)^{c_3 nh_n}.$$

Fix some $t_0 \in (0, (2c)^{-2/\alpha})$ such that $(1 - F(-t))/(ct^\alpha) \in (1/2, 2)$ for all $t \in (0, t_0]$. In what follows, d denotes a generic constant (depending only on β, c_g, c and K) which may vary from

456   H. Drees, N. Neumeyer and L. Selk

line to line. Applying the inequalities $\exp(-2\rho u) \leq (1-u)^\rho \leq \exp(-\rho u)$, which holds for all $\rho > 0$ and $u \in (0, 1/2)$, we obtain for $(nh_n/\log n)^{-2/\alpha} < t \leq t_0$ and sufficiently large n

$$P\{M_i^* > t^{1/2}\} \leq 1 - \big[1 - \big(1 + c_3 nh_n 2ct^{\alpha/2}\big)\big(1 - ct^{\alpha/2}/2\big)^{c_3 nh_n}\big]^{c_2(h_n+b_n)/h_n}$$

$$\leq 1 - \big[1 - 3c_3 cnh_n t^{\alpha/2} \exp\big(-c_3 cnh_n t^{\alpha/2}/2\big)\big]^{c_2(h_n+b_n)/h_n}$$

$$\leq 1 - \exp\big(-dn(h_n+b_n)t^{\alpha/2}\exp\big(-c_3 cnh_n t^{\alpha/2}/2\big)\big)$$

$$\leq dn(h_n+b_n)t^{\alpha/2}\exp\big(-c_3 cnh_n t^{\alpha/2}/2\big).$$

Therefore, for sufficiently large $a > 0$,

$$\int_0^{t_0^2} P\{M_i^* > t^{1/2}\}\, dt$$

$$\leq a\left(\frac{nh_n}{\log n}\right)^{-2/\alpha} + dt_0 n(h_n+b_n)\int_{a(nh_n/\log n)^{-2/\alpha}}^{t_0} t^{\alpha/2-1}\exp\big(-c_3 cnh_n t^{\alpha/2}/2\big)\, dt$$

$$\leq o\big(1/(n(h_n+b_n))\big) + dt_0 n(h_n+b_n)\exp\big(-c_3 ca^{\alpha/2}\log n/2\big)$$ \tag{A.19}

$$= o\big(1/(n(h_n+b_n))\big),$$

where in the last but one step we apply the conditions (B4) and (H2). Now, assertion (A.18) (and hence (A.16)) follows from

$$\int_{t_0^2}^\infty P\{M_i^* > t^{1/2}\}\, dt \leq \int_{t_0^2}^\infty 1 - \big[1 - c_3 nh_n\big(F\big(-t^{1/2}\big)\big)^{c_3 nh_n}\big]^{c_2(h_n+b_n)/h_n}\, dt$$

$$\leq \int_{t_0^2}^\infty 1 - \exp\big(-dn(h_n+b_n)\big(F\big(-t^{1/2}\big)\big)^{c_3 nh_n}\big)\, dt$$

$$\leq dn(h_n+b_n)\left(nh_n\big(F(-t_0)\big)^{c_3 nh_n} + \int_{nh_n}^\infty t^{-\tau c_3 nh_n/2}\, dt\right)$$

$$= o\big(n^{-\xi}\big)$$

for all $\xi > 0$ and sufficiently large n, where we have used (H2) and (F3).

To establish (A.17), first note that for a kernel K of order greater than or equal to $\lceil\beta\rceil$

$$E\big(\tilde{g}_n(x) - g(x)\big) = E\left(\int_{-1}^1 \left(\hat{g}_n(x+b_n u) - \sum_{j=0}^{\lceil\beta\rceil-1}\frac{g^{(j)}(x)}{j!}(b_n u)^j\right)K(u)\, du\right)$$

$$= \int_{-1}^1 E\big(\hat{g}_n(x+b_n u) - g(x+b_n u)\big)K(u)\, du + O\big(b_n^\beta\big)$$

uniformly for all $x \in [h_n + b_n, 1 - h_n - b_n]$. In view of (K1), (H2) and (B4), it thus suffices to show that

$$\left| E(\hat{g}_n(x) - g(x)) - E_{g \equiv 0}(\hat{g}_n(1/2)) \right| = \left| E(\hat{g}_n(x) - g(x)) - E_{g \equiv 0}(\hat{g}_n(x)) \right|$$
$$= o(n^{-1/2}) \tag{A.20}$$

uniformly for Lebesgue almost all $x \in [h_n, 1 - h_n]$. Note that the distribution of $\hat{g}_n(x)$ does not depend on x if g equals 0.

Recall that $\hat{g}_n(x) = \tilde{p}_n(0)$ where $\tilde{p}_n$ is a polynomial on $[-1, 1]$ of degree β^* that solves the linear optimization problem

$$\int_{-1}^{1} \tilde{p}_n(t)\, dt \to \min!$$

under the constraints

$$\tilde{p}_n\left(\frac{i/n - x}{h_n}\right) \geq Y_i, \qquad \forall i \in [n(x - h_n), n(x + h_n)].$$

Define polynomials

$$q_x(t) := \sum_{k=0}^{\lfloor \beta \rfloor} \frac{1}{k!} g^{(k)}(x)(h_n t)^k, \qquad p_n(t) := (nh_n)^{1/\alpha}(\tilde{p}_n(t) - q_x(t)), \qquad t \in [-1, 1].$$

Then $q_x((u - x)/h_n)$ is the Taylor expansion of order $\lfloor \beta \rfloor$ of $g(u)$ at x and the estimation error can be written as

$$\hat{g}_n(x) - g(x) = (nh_n)^{-1/\alpha} p_n(0). \tag{A.21}$$

Note that p_n is a polynomial of degree β^* that solves the linear optimization problem

$$\int_{-1}^{1} p_n(t)\, dt \to \min!$$

subject to

$$p_n\left(\frac{i/n - x}{h_n}\right) \geq (nh_n)^{1/\alpha} \bar{\varepsilon}_i, \qquad \forall i \in [n(x - h_n), n(x + h_n)], \tag{A.22}$$

with

$$\bar{\varepsilon}_i := \varepsilon_i + g(i/n) - q_x\left(\frac{i/n - x}{h_n}\right).$$

We now use point process techniques to analyze the asymptotic behavior of this linear program. Denote by

$$N_n := \sum_{i \in [n(x - h_n), n(x + h_n)]} \delta_{((i/n - x)/h_n, (nh_n)^{1/\alpha} \bar{\varepsilon}_i)}$$

a point process of standardized error random variables. Then the constraints (A.22) can be reformulated as $N_n(A_{p_n}) = 0$ where $A_f := \{(t, u) \in [-1, 1] \times \mathbb{R} | u > f(t)\}$ denotes the open epigraph of a function f.

Since by (H2) $|\bar\varepsilon_i - \varepsilon_i| = g(i_n) - q_x((i/n - x)/h_n) = O(h_n^\beta) = o((nh_n)^{-1/\alpha})$ uniformly for all $i \in [n(x - h_n), n(x + h_n)]$, one has

$$E\big(N_n\big([-1, 1] \times (-1, \infty)\big)\big) \sim 2nh_n P\big\{\bar\varepsilon_1 > -(nh_n)^{-1/\alpha}\big\} \to 2c.$$

Therefore, N_n converges weakly to a Poisson process N on $[-1, 1] \times \mathbb{R}$ with intensity measure $2cU_{[-1,1]} \otimes \nu_\alpha$ where ν_α has Lebesgue density $x \mapsto \alpha|x|^{\alpha-1} I(-\infty, 0)$ (see, e.g., Resnick [31], Theorem 6.3). By Skorohod's representation theorem, we may assume that the convergence holds a.s.

Next, we analyze the corresponding linear program in the limiting model to minimize $\int_{-1}^1 p(t)\,dt$ over polynomials of degree β^* subject to $N(A_p) = 0$. In what follows, we use a representation of the Poisson process as $N = \sum_{i=1}^\infty \delta_{(T_i, Z_i)}$ where T_i are independent random variables which are uniformly distributed on $[-1, 1]$.

First, we prove by contradiction that the optimal solution is almost surely unique. Suppose that there exist more than one solution. From the theory of linear programs, it is known that then there exists a solution p such that $J := \{j \in \mathbb{N} | p(T_j) = Z_j\}$ has at most β^* elements. Because p is bounded and N has a.s. finitely many points in any bounded set, $\eta := \inf\{|p(T_i) - Z_i| | i \in \mathbb{N} \setminus J\} > 0$ a.s. Since p is an optimal solution, all polynomials Δ of degree β^* such that $\Delta(T_j) = 0$, $j \in J$, and $\|\Delta\|_\infty < \eta$ must satisfy $\int_{-1}^1 \Delta(t)\,dt = 0$, because both $p + \Delta$ and $p - \Delta$ satisfy the constraints $N(A_{p\pm\Delta}) = 0$. In particular, for all polynomials q of degree $\beta^* - |J|$, $\Delta(t) = \tau \prod_{i \in J}(t - T_i)q(t)$ is of that type if $\tau > 0$ is sufficiently small. Write $\prod_{i \in J}(t - T_i)$ in the form $t^{|J|} + \sum_{l=0}^{|J|-1} a_l t^l$. Then necessarily

$$\int_{-1}^1 \prod_{i \in J}(t - T_i)t^j\,dt = \frac{2}{|J| + j + 1} I\{|J| + j \text{ even}\} + \sum_{l=0}^{|J|-1} \frac{2a_l}{l + j + 1} I\{l + j \text{ even}\} = 0,$$

for all $j \in \{0, \dots, \beta^* - |J|\}$. This implies that $(T_i)_{i \in J}$ lies on a manifold $M_{|J|,\beta^*}$ of dimension $|J| - (\beta^* - |J| + 1) = 2|J| - \beta^* - 1$ which only depends on $|J|$ and β^*. However, by Proposition A.1, $\|p\|_\infty \le K_{\beta^*} Z_{max}$ where

$$Z_{max} := \max_{1 \le i \le j_{\beta^*}} \min\big\{|Z_i| | T_i \in \big[-1 + (j-1)/j_{\beta^*}, -1 + j/j_{\beta^*}\big]\big\}.$$

The above conclusion contradicts $P\{Z_{max} > K\} \to 0$ as $K \to \infty$, since

$$P\Big\{\exists J \subset \mathbb{N} : |J| \le \beta^*, (T_j)_{j \in J} \in M_{|J|,\beta^*}, \max_{j \in J} |Z_j| \le K_{\beta^*} K\Big\} = 0$$

for all $K > 0$ (i.e., the fact that among finitely many values T_i a.s. there does not exist a subset which lies on a given manifold of lower dimension).

Therefore the solution p must be a.s. unique which in turn implies that it is a basic feasible solution, that is, $|J| \ge \beta^* + 1$. On the other hand, because the intensity measure of N is absolutely continuous, $|J| \le \beta^* + 1$ a.s. and thus $|J| = \beta^* + 1$. Because of $N_n \to N$ a.s., one has

$N_n([-1, 1] \times [-K_{\beta^*} Z_{\max}, \infty)) = N([-1, 1] \times [-K_{\beta^*} Z_{\max}, \infty)) =: M$ for sufficiently large n. Moreover, one can find a numeration of the points $(T_{n,i}, Z_{n,i})$, $1 \le i \le M$, of N_n and (T_i, Z_i), $1 \le i \le M$, of N in $[-1, 1] \times [-K_{\beta^*} Z_{\max}, \infty)$ such that $(T_{n,i}, Z_{n,i}) \to (T_i, Z_i)$.

Next we prove that the solution to the linear program to minimize $\int_{-1}^{1} p_n(t)\, dt$ subject to $N_n(A_{p_n}) = 0$ is eventually unique with $p_n \to p$ a.s. Since any optimal solution can be written as a convex combination of basic feasible solutions, w.l.o.g. we may assume that $J_n := \{1 \le i \le M \mid p_n(T_{n,i}) = Z_{n,i}\}$ has at least $\beta^* + 1$ elements. The polynomial p_n is uniquely determined by this set J_n. Suppose that along a subsequence n' the set $J_{n'}$ is constant, but not equal to J. Then p'_n converges uniformly to the polynomial $\bar{p}$ of degree β^* that is uniquely determined by the conditions $\bar{p}(T_i) = Z_i$ for all $i \in J_{n'}$. In particular, $\bar{p}$ is different from the unique optimal polynomial p for the limit Poisson process N, but it satisfies the constraints $N(A_p) = 0$. Thus, $\int_{-1}^{1} \bar{p}(t)\, dt > \int_{-1}^{1} p(t)\, dt$. On the other hand, for all $\eta > 0$ the polynomial $p + \eta$ eventually satisfies the constraints $N_n(A_{p+\eta}) = 0$ and thus $\int_{-1}^{1} p(t) + \eta\, dt \ge \int_{-1}^{1} \bar{p}_n(t)\, dt$, which leads to a contradiction.

Hence, $J_n = J$ for all sufficiently large n and the optimal solution p_n for N_n is unique and it converges uniformly to the optimal solution p for the Poisson process N. Moreover, using the relation $(p_n(T_{n,j}))_{j \in J} = (Z_{n,j})_{j \in J}$ (which is a system of linear equation in the coefficients of p_n), $p_n(0)$ can be calculated as $w_n^t (Z_{n,j})_{j \in J}$ for some vector w_n which converges to a limit vector w (corresponding to the analogous relation for p).

Exactly the same arguments apply if we replace $\bar{\varepsilon}_i$ with ε_i, which corresponds to the case that g is identical 0. Since the points $(\tilde{T}_{n,i}, \tilde{Z}_{n,i})$ of the pertaining point process equal $(T_{n,i}, Z_{n,i} - (nh_n)^{1/\alpha}(g(i/n) - q_x((i/n) - x)/h_n))$ and thus $|\tilde{Z}_{n,i} - Z_{n,i}| \le c_g (nh_n)^{1/\alpha} h_n^\beta$, the difference of the resulting values for optimal polynomial at 0 is bounded by a multiple of $(nh_n)^{1/\alpha} h_n^\beta$. In view of (A.21) and (H2), we may conclude that the difference between the estimation errors can be bounded by a multiple of $h_n^\beta = o(n^{-1/2})$, which finally yields (A.20) and thus the assertion.

A.9. Proof of Corollary 4.3

Note that $\frac{\lfloor ns \rfloor}{\sqrt{n}}(F_{\lfloor ns \rfloor}(y) - F_n(y)) = E_n(y, s) - \frac{\lfloor ns \rfloor}{n} E_n(y, 1)$ with E_n defined in (A.11). A similar reasoning as in the proof of Theorem 3.1 (see (A.13)) shows that

$$\sup_{y \in \mathbb{R}, s \in [0,1]} \left| \left(\frac{n \bar{s}_n}{\lfloor ns \rfloor} - 1 \right) \frac{\lfloor ns \rfloor}{\sqrt{n}} (F_{\lfloor ns \rfloor}(y) - F_n(y)) \right| = o_P(1).$$

Hence, by Theorem 3.1, uniformly for all $y \in \mathbb{R}$, $s \in [0, 1]$,

$$\begin{aligned}
&\sqrt{n}(\hat{F}_n(y, s) - \bar{s}_n \hat{F}_n(y)) \\
&= \sqrt{n}\big(\hat{F}_n(y, s) - \bar{s}_n F_{\lfloor ns \rfloor}(y)\big) - \bar{s}_n \sqrt{n}\big((\hat{F}_n(y) - F_n(y))\big) + \bar{s}_n \sqrt{n}\big(F_{\lfloor ns \rfloor}(y) - F_n(y)\big) \\
&= \frac{n \bar{s}_n}{\lfloor ns \rfloor} \frac{\lfloor ns \rfloor}{\sqrt{n}}\big(F_{\lfloor ns \rfloor}(y) - F_n(y)\big) + o_P(1)
\end{aligned}$$

$$= E_n(y, s) - \frac{\lfloor ns \rfloor}{n} E_n(y, 1) + o_P(1)$$

$$= E_n(y, s) - s E_n(y, 1) + o_P(1)$$

which converges weakly to $K_F(y, s) - s K_F(y, 1)$ for the Kiefer process K_F defined in Theorem 3.1. Check that this Gaussian process has the same law as $G(s, F(y))$, because they have the same covariance function. Thus, the Kolmogorov–Smirnov statistic T_n converges weakly to $\sup_{s \in [0,1], y \in \mathbb{R}} |G(s, F(y))| = \sup_{s \in [0,1], z \in [0,1]} |G(s, z)|$, where the last equality holds by the continuity of F.

A.10. Proof of Theorem 4.4

Note that under the given assumptions, the statements of Theorem 2.8(i) and (ii) are valid with rate $o_P(1)$. Let $\Omega_n := \{\inf_{x \in I_n} \tilde{g}'(x) > 0\}$. From assumption (I1) and Theorem 2.8(ii) it follows that $P(\Omega_n) \to 1$ for $n \to \infty$. But on Ω_n the estimators $\tilde{g}_I$ and $\tilde{g}$ are identical, and thus $\tilde{F}_{I, \lfloor ns \rfloor} = \tilde{F}_{\lfloor ns \rfloor}$. Now (4.1) can be concluded as in the proof of Theorem 3.1, because as in Theorem 2.8(i) we may conclude from our assumptions that $\sup_{x \in I_n} |\tilde{g}(x) - g(x)| = o_P(n^{-1/(2(\alpha \wedge 1))})$. The convergence of the Kolmogorov–Smirnov test statistic then follows exactly as in the proof of Corollary 4.3.

A.11. Proof of Proposition 4.6

Let $\tilde{\varepsilon}_i := \varepsilon_{\lfloor nx_0 \rfloor + i}$, $1 \le i \le l_n$. By assumption (G1) we have $\tilde{Y}_i = g(x_0) + \tilde{\varepsilon}_i + O(l_n/n)$. Let U_i, $1 \le i \le l_n$, be i.i.d. standard uniform random variables so that $(\tilde{\varepsilon}_i)_{1 \le i \le l_n}$ has the same distribution as $(F^{\leftarrow}(1 - U_i))_{1 \le i \le l_n}$. Let $\tilde{R}_n := \sup_{0 < t \le (1+\iota)k_n/l_n} |\tilde{r}(t)|$. Since $(l_n/k_n)U_{k_n:l_n} \to 1$ in probability and $U_{1:k_n} = O_P(k_n^{-1})$, we obtain

$$\frac{1}{\hat{\alpha}_n} =^d -\sum_{i=2}^{k_n} d_{ni} \log \frac{\tilde{\varepsilon}_{l_n:l_n} - \tilde{\varepsilon}_{l_n-i+1:l_n} + O(l_n/n)}{\tilde{\varepsilon}_{l_n:l_n} - \tilde{\varepsilon}_{l_n-k_n:l_n} + O(l_n/n)}$$

$$= -\sum_{i=2}^{k_n} d_{ni} \log \frac{U_{i:l_n}^{1/\alpha} - U_{1:l_n}^{1/\alpha} + O_P(l_n/n + \tilde{R}_n)}{U_{k_n+1:l_n}^{1/\alpha} - U_{1:l_n}^{1/\alpha} + O_P(l_n/n + \tilde{R}_n)}$$

$$= -\sum_{i=2}^{k_n} d_{ni} \log \frac{(U_{i:l_n}/U_{k_n+1:l_n})^{1/\alpha} - (U_{1:l_n}/U_{k_n+1:l_n})^{1/\alpha} + O((k_n/l_n)^{-1/\alpha}(l_n/n + \tilde{R}_n))}{1 - (U_{1:l_n}/U_{k_n+1:l_n})^{1/\alpha} + O((k_n/l_n)^{-1/\alpha}(l_n/n + \tilde{R}_n))}$$

$$=^d -\sum_{i=2}^{k_n} d_{ni} \log \frac{U_{i:k_n}^{1/\alpha} - U_{1:k_n}^{1/\alpha} + o_P(k_n^{-1/2})}{1 - U_{1:k_n}^{1/\alpha} + o_P(k_n^{-1/2})}$$

$$= -\sum_{i=2}^{k_n} d_{ni} \log \left[U_{i:k_n}^{1/\alpha} + o_P(k_n^{-1/2}) \right],$$

where in the last but one step we have used Corollary 1.6.2 of Reiss [30] and the conditions (4.4) and (4.5).

For suitable versions of U_i and B, one has

$$\sup_{1 \le i \le k_n} \left| k_n^{1/2} (U_{i:k_n} - i/k_n) - B(i/k_n) \right| \to 0 \qquad \text{a.s.}$$

(see, e.g., Shorack and Wellner [32], Theorem 3.1.1, p. 93). Therefore,

$$U_{i:k_n}^{1/\alpha} = \left(\frac{i}{k_n} \right)^{1/\alpha} \left[1 + \frac{k_n^{1/2}}{i} B(i/k_n) + o\left(k_n^{-1/2}\right) \right]^{1/\alpha}$$

$$= \left(\frac{i}{k_n} \right)^{1/\alpha} \left[1 + \frac{k_n^{1/2}}{\alpha i} B(i/k_n) + o\left(k_n^{-1/2}\right) \right]$$

uniformly for $\lambda k_n \le i \le k_n$. Thus,

$$\frac{1}{\hat{\alpha}_n} =^d \frac{1}{\alpha} \sum_{i=\lceil \lambda k_n \rceil}^{k_n} d_{ni} \left| \log \frac{i}{k_n} \right| - \sum_{i=\lceil \lambda k_n \rceil}^{k_n} d_{ni} \log \left[1 + \frac{k_n^{1/2}}{\alpha i} B(i/k_n) + o\left(k_n^{-1/2}\right) \right]$$

$$= \frac{1}{\alpha} \int_\lambda^1 d(t) |\log t| \, dt + O\left(k_n^{-1}\right) - \frac{1}{\alpha} \sum_{i=\lceil \lambda k_n \rceil}^{k_n} d_{ni} \left[\frac{k_n^{1/2}}{i} B(i/k_n) + o\left(k_n^{-1/2}\right) \right].$$

In view of (4.2), we conclude

$$k_n^{1/2} \left(\frac{1}{\hat{\alpha}_n} - \frac{1}{\alpha} \right) \to -\frac{1}{\alpha} \int_\lambda^1 d(t) t^{-1} B(t) \, dt \qquad \text{weakly,}$$

from which assertion (4.6) easily follows.

The asymptotic variance can be calculated as $\alpha^2 \int_0^1 \int_0^1 d(s) \, d(t) (st)^{-1} \text{Cov}(B(s), B(t)) \, ds \, dt$.

Acknowledgement

References

[1] Akritas, M.G. and Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scand. J. Stat.* **28** 549–567. MR1858417

[2] Anevski, D. and Fougères, A.-L. (2007). Limit properties of the monotone rearrangement for density and regression function estimation. Available at arXiv:0710.4617v1.

[3] Birke, M. and Neumeyer, N. (2013). Testing monotonicity of regression functions – an empirical process approach. *Scand. J. Stat.* **40** 438–454.

[4] Birke, M., Neumeyer, N. and Volgushev, S. (2017). The independence process in conditional quantile location-scale models and an application to testing for monotonicity. *Statist. Sinica* **27** 1815–1839.

[5] Chernozhukov, V., Fernández-Val, I. and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika* **96** 559–575. MR2538757

[6] Daouia, A., Noh, H. and Park, B.U. (2016). Data envelope fitting with constrained polynomial splines. *J. R. Stat. Soc.*, *B* **78** 3–30.

[7] Darling, D.A. (1955). The Cramér-Smirnov test in the parametric case. *Ann. Math. Stat.* **26** 1–20.

[8] Einmahl, J.H.J. and Van Keilegom, I. (2008). Specification tests in nonparametric regression. *J. Econometrics* **143** 88–102.

[9] Färe, R. and Grosskopf, S. (1983). Measuring output efficiency. *European J. Oper. Res.* **13** 173–179.

[10] Gijbels, I. (2005). Monotone regression. In *The Encyclopedia of Statistical Sciences*, 2nd ed. (N. Balakrishnan, S. Kotz, C.B. Read and B. Vadakovic, eds.). Hoboken, NJ: Wiley.

[11] Gijbels, I., Mammen, E., Park, B. and Simar, L. (1999). On estimation of monotone and concave frontier functions. *J. Amer. Statist. Assoc.* **94** 220–228.

[12] Gijbels, I. and Peng, L. (1999). Estimation of a support curve via order statistics. *Extremes* **3** 251–277. MR1856200

[13] Girard, S., Guillou, A. and Stupfler, G. (2013). Frontier estimation with kernel regression on high order moments. *J. Multivariate Anal.* **116** 172–189.

[14] Girard, S. and Jacob, P. (2008). Frontier estimation via kernel regression on high power-transformed data. *J. Multivariate Anal.* **99** 403–420.

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

[16] Hall, P., Park, B.U. and Stern, S.E. (1998). On polynomial estimators of frontiers and boundaries. *J. Multivariate Anal.* **66** 71–98.

[17] Hall, P. and Van Keilegom, I. (2009). Nonparametric "regression" when errors are positioned at endpoints. *Bernoulli* **15** 614–633. MR2555192

[18] Härdle, W., Park, B.U. and Tsybakov, A.B. (1995). Estimation of non-sharp support boundaries. *J. Multivariate Anal.* **55** 205–218.

[19] Jirak, M., Meister, A. and Reiß, M. (2014). Adaptive function estimation in nonparametric regression with one-sided errors. *Ann. Statist.* **42** 1970–2002. MR3262474

[20] Khmaladze, E.V. and Koul, H.L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.* **32** 995–1034.

[21] Khmaladze, E.V. and Koul, H.L. (2009). Goodness-of-fit problem for errors in nonparametric regression: Distribution free approach. *Ann. Statist.* **37** 3165–3185. MR2549556

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

[23] Koul, H.L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*, 2nd ed. New York: Springer.

[24] Meister, A. and Reiß, M. (2013). Asymptotic equivalence for nonparametric regression with non-regular errors. *Probab. Theory Related Fields* **155** 201–229.

[25] Müller, U.U. and Wefelmeyer, W. (2010). Estimation in nonparametric regression with non-regular errors. *Comm. Statist. Theory Methods* **39** 1619–1629.

[26] Neumeyer, N., Dette, H. and Nagel, E.-R. (2006). Bootstrap tests for the error distribution in linear and nonparametric regression models. *Aust. N. Z. J. Stat.* **48** 129–156. MR2253914

[27] Neumeyer, N. and Van Keilegom, I. (2009). Change-point tests for the error distribution in non-parametric regression. *Scand. J. Stat.* **36** 518–541. Online supporting information available at http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9469.2009.00639.x/suppinfo.

[28] Picard, D. (1985). Testing and estimating change-points in time series. *Adv. in Appl. Probab.* **17** 841–867.

[29] Reiß, M. and Selk, L. (2017). Efficient estimation of functionals in nonparametric boundary models. *Bernoulli* **23** 1022–1055.

[30] Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics*: *With Applications to Nonparametric Statistics*. New York: Springer.

[31] Resnick, S.I. (2007). *Heavy-Tail Phenomena. Springer Series in Operations Research and Financial Engineering*. New York: Springer. MR2271424

[32] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*. New York: Wiley.

[33] Simar, L. and Wilson, P.W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Manage. Sci.* **44** 49–61.

[34] Stephens, M.A. (1976). Asymptotic results for goodness-of-fit statistics with unknown parameters. *Ann. Statist.* **4** 357–369.

[35] van der Vaart, A.W. (2000). *Asymptotic Statistics*. Cambridge: Cambridge Univ. Press.

[36] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer.

[37] Wilson, P.W. (2003). Testing independence in models of productive efficiency. *J. Product. Anal.* **20** 361–390.