Hypotheses tests in boundary regression models

Holger Drees, Natalie Neumeyer and Leonie Selk*
University of Hamburg, Department of Mathematics
Bundesstrasse 55, 20146 Hamburg, Germany

October 12, 2016

Abstract

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.

AMS 2010 Classification: Primary 62G08; Secondary 62G10, 62G30, 62G32

Keywords and Phrases: 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

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

with negative errors ε_i whose survival function $1-F(y)$ behaves like a multiple of $|y|^\alpha$ for some $\alpha > 0$ near the origin. 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 (2009) 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 2.6). Relatedly, Müller and Wefelmeyer (2010) 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 (2013) 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. (2014) 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 (2016+) 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 (2013).

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 $\{(x, y) \in [0, 1] \times [0, \infty] \mid y \leq g(x)\}$. For such models, Härdle et al. (1995) and Hall et al. (1998) proved minimax rates both for $g(x)$ 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 (2008) and Girard et al. (2013). Daouia et al. (2016) consider spline estimation of a support frontier curve and obtain uniform rates of convergence.

The aim of the paper is to develop tests for model assumptions in 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.

The test statistics are based on (sequential) empirical processes of residuals. To investigate these, we need uniform rates of convergence for the regression estimator, which are of interest on its own. To our knowledge, uniform rates so far have only been shown by Daouia et al. (2016) who do not obtain optimal 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 e.g. Gijbels and Peng (2000), Hall and Van Keilegom (2009), or Müller and Wefelmeyer (2010).

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 distribution function. 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. We apply the results to derive goodness-of-fit tests for parametric classes of error distributions. Asymptotic properties of residual empirical distribution functions in mean regression models were investigated by Akritas and Van Keilegom (2001), among others. As the regression estimation strongly influences the asymptotic behavior of the empirical distribution function in these regular models, asymptotic distributions of goodness-of-fit test statistics are involved, and typically bootstrap is applied to obtain critical values, see Neumeyer et al. (2006). In contrast, in the present situation with an irregular error distribution, standard critical values can be used.

In nonparametric frontier models, Wilson (2003) discusses several possible tests for assumptions of independence, for instance independence between input levels and output inefficiency. Those assumptions are needed to prove validity of bootstrap procedures and are thus crucial in applications, but they may be violated; see Simar and Wilson (1998). Wilson (2003) points out the analogy to tests for independence between errors and covariates in regression models, but no asymptotic distributions are derived. 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 (2008) and Birke et al. (2016+).

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. (1999), the literature cited therein and the monotone nonparametric maximum likelihood estimator in Reiß and Selk (2016+). Monotonicity of a production frontier function in each component is given under the strong disposability assumption, but may often not be fulfilled; see e.g. Färe and Grosskopf (1983). 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' (2005) 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 (2013) and Birke et al. (2016+) 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 based 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 as well as a test for monotonicity of the boundary function. 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,

$$Y_i \;=\; g(\tfrac{i}{n}) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

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

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

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. ■

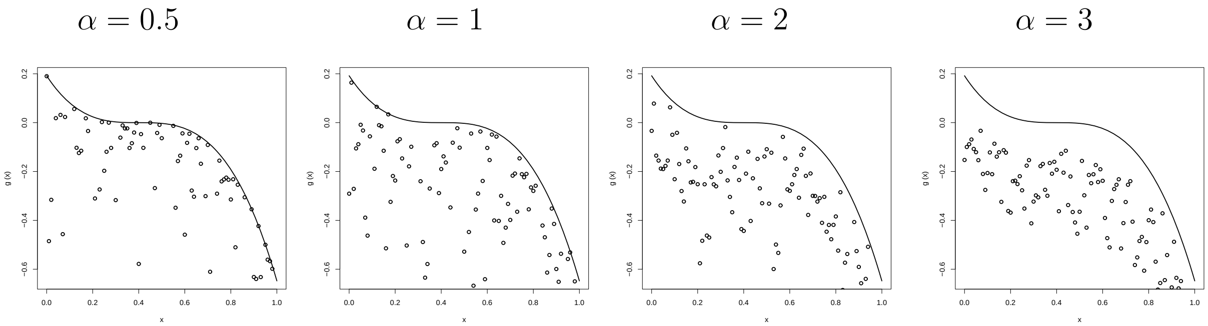

Figure 1: *Scatter plots of* $(\frac{i}{n}, Y_i)$*,* $i = 1, \ldots, n$*, and the true regression function* $g(x) = -3(x - 0.4)^3$*. The error distribution is Weibull* $F(y) = \exp(-(|y|/\theta)^\alpha) I_{(-\infty,0)}(y) + I_{[0,\infty)}(y)$ *with scale* $\theta = 0.3$ *and shape parameter* α*.*

We consider an estimator that locally approximates the regression function by a polynomial while lying above the data points. More specifically, for $x \in [0, 1]$, let

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

where p is a polynomial of order $\lceil \beta \rceil - 1$ 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 $|\frac{j}{n} - x| \leq h_n$. 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**), and (**H1**), we have*

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

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

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

(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 $O_P(((\log n)/n)^{\frac{\beta}{2\beta+1}})$ for estimating mean regression functions in models with regular errors.

Remark 2.3 Jirak et al. (2014) consider a similar boundary regression estimator while replacing the integral in (2.2) by its Riemann approximation $\sum_{i=1}^n p(\frac{i}{n})I\{|\frac{i}{n}-x| \le h_n\}$. In particular, they use the Lepski method to construct a data-driven bandwidth that satisfies $h_n \asymp ((\log n)/n)^{\frac{1}{\alpha\beta+1}}$ in probability. For this modified estimator, we obtain the same uniform rate of convergence as in Theorem 2.2 by replacing Proposition A.1 in the proof of Theorem 2.2 by Theorem 3.1 in Jirak et al. (2014). ■

Remark 2.4 For Hölder continuous regression functions with exponent $\beta \in (0,1]$ the estimator reduces to a local maximum, i. e. $\hat{g}(x) = \max\{Y_i \mid i = 1,\dots,n \text{ s. t. } |\frac{i}{n}-x| \le h_n\}$. In this case we obtain the rate of convergence as given in Theorem 2.2 uniformly over the whole unit interval. ■

Remark 2.5 Müller and Wefelmeyer (2010) consider a mean regression model $Y_i = m(X_i)+\eta_i$, $i = 1,\dots,n$, 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, i. e.

$$\hat{m}(x) = \frac{1}{2}\Big(\min_{\substack{i\in\{1,\dots,n\}\\ |X_i-x|\le h_n}} Y_i + \max_{\substack{i\in\{1,\dots,n\}\\ |X_i-x|\le h_n}} Y_i\Big)$$

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 Wefelmeyer's (2010) 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 $\hat{m} = (\hat{g} - \hat{\tilde{g}})/2$ with $\hat{g}$ as before and $\hat{\tilde{g}}$ defined analogously, but based on $(\frac{i}{n}, -Y_i)$, $i = 1,\dots,n$; 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. ■

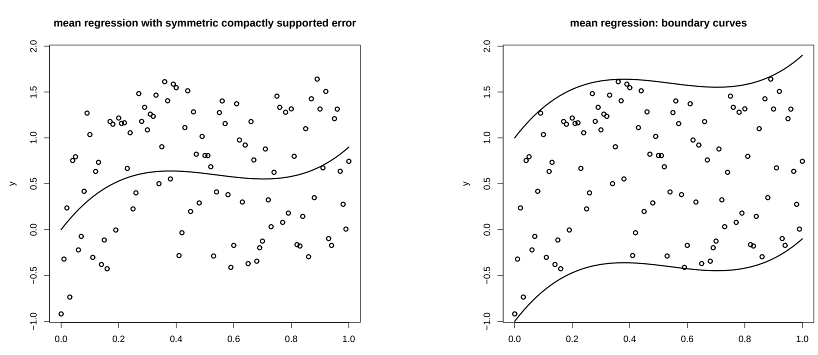

Figure 2: *Example for data as in Remark 2.5.*

Remark 2.6 For $\beta \in (1,2]$, Hall and Van Keilegom (2009) consider the following local linear boundary regression estimator:

$$\check{g}(x) = \inf\left\{\alpha_0 \,\middle|\, (\alpha_0,\alpha_1) \in \mathbb{R}^2 : Y_i \le \alpha_0 + \alpha_1\left(\tfrac{i}{n} - x\right) \forall i \in \{1,\dots,n\} \text{ s.t. } \left|\tfrac{i}{n} - x\right| \le h_n\right\}. \tag{2.4}$$

Because of $\int_{x-h_n}^{x+h_h}(\alpha_0 + \alpha_1(t-x))\,dt = 2\alpha_0 h_n$ this estimator coincides with $\hat{g}$ for $\beta \in (1,2]$. However, in the case $\beta > 2$ replacing the linear function in (2.4) by a polynomial of order $\lceil\beta\rceil - 1$ renders the estimator $\check{g}$ useless. One obtains $\check{g}(x) = -\infty$ for $x \notin \{\frac{j}{n} \mid j = 1,\dots,n\}$ while $\check{g}(\frac{j}{n}) = Y_j$, $j = 1,\dots,n$. This was already observed by Jirak et al. (2014). ■

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

$$\tilde{g}(x) = \int_{h_n}^{1-h_n} \hat{g}(z)\frac{1}{b_n}K\left(\frac{x-z}{b_n}\right)dz \tag{2.5}$$

and formulate some additional assumptions.

(K1) K is a continuous kernel with support $[-1,1]$ and order $\lfloor\beta\rfloor + 1$, i.e. $\int K(u)\,du = 1$, $\int u^r K(u)\,du = 0\ \forall r = 1,\dots,\lfloor\beta\rfloor$. 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.$\boldsymbol{\delta}$) $h_n^\beta + \left(\dfrac{\log n}{nh_n}\right)^{1/\alpha} = o\big(b_n^{(1+2\delta)\vee(3-(\beta-1)(1/\delta-1))}\big) = \begin{cases} o\big(b_n^{1+2\delta}\big) & \text{if } \delta \le \frac{\beta-1}{2} \\ o\big(b_n^{3-(\beta-1)(1/\delta-1)}\big) & \text{if } \delta > \frac{\beta-1}{2}. \end{cases}$

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 $\beta < 3$ and the optimal bandwidth $h_n \asymp ((\log n)/n)^{1/(\alpha\beta+1)}$ is chosen, then (**B2.$\boldsymbol{\delta}$**) 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 2.7 *If the model assumptions (2.1), **(F1)**, **(G1)** with $\beta > 1$, **(H1)**, **(K1)**, and **(B1)** hold, then for $I_n = [h_n + b_n, 1 - h_n - b_n]$*

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

(ii) $\sup_{x\in I_n} |\tilde{g}'(x) - g'(x)| = O(b_n^{\beta-1}) + O\left(b_n^{-1}h_n^\beta\right) + O_P\Big(b_n^{-1}\Big(\dfrac{|\log h_n|}{nh_n}\Big)^{\frac{1}{\alpha}}\Big).$
*If $h_n^\beta + (\log n/(nh_n))^{1/\alpha} = o(b_n)$, then $\sup_{x\in I_n}|\tilde{g}'(x) - g'(x)| = o_P(1)$; in particular this holds if (**B2.$\boldsymbol{\delta}$**) is fulfilled for some $\delta \in (0, 1\wedge(\beta-1))$.*

(iii) *For all* $\delta \in (0, 1 \wedge (\beta - 1))$, *under the additional assumption* (**B2.δ**),

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

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 cdf 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

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

where $m_n = \sharp\{i \in \{1, \dots, n\} \mid h_n < \frac{i}{n} \leq 1 - h_n\} = n - \lfloor nh_n \rfloor - \lceil nh_n \rceil$. 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 $\varepsilon_1, \dots, \varepsilon_n$. Furthermore, define $\bar{s}_n = \big(\lfloor n(s \wedge (1 - h_n)) \rfloor - \lfloor n(s \wedge h_n) \rfloor\big)/m_n$ and interpret $\bar{s}_n / \lfloor ns \rfloor$ as 0 for $s = 0$. Note that $\bar{s}_n = 1$ if $s = 1$ and $s_n \to s$ as $n \to \infty$, for each fixed s.

Theorem 3.1 *Assume that the conditions (**F1**), (**G1**), and (**F2**) are fulfilled with* $\beta > 1$. *Furthermore, assume* $\frac{1}{\beta} < \alpha < 2 - \frac{1}{\beta}$ *and* $h_n \asymp ((\log n)/n)^{1/(\alpha\beta + 1)}$. *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 $\{\sqrt{n}(\hat{F}_n(y, s) - \bar{s}_n F(y)) \mid s \in [0, 1], y \in \mathbb{R}\}$ *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 assertion of Theorem 3.1 holds true under the following weaker conditions on the (possibly random) bandwidth:

$$h_n = o_P\big(n^{-1/(2(\alpha \wedge 1)\beta)}\big), \quad n^{(\alpha \vee 1)/2 - 1} \log n = o_P(h_n). \tag{3.1}$$

In particular, one may use the adaptive bandwidth proposed by Jirak et al. (2014).
Condition (3.1) can be fulfilled if and only if $\frac{1}{\beta} < \alpha < 2 - \frac{1}{\beta}$, which in turn can be satisfied for all $\alpha \in (0, 2)$, provided the regression function g is sufficiently smooth. It ensures that one can choose a rate a_n of larger order than the uniform bound on the estimation error established in Theorem 2.2 such that

$$|F(y + a_n) - F(y)| = O(a_n^{\alpha \wedge 1}) = o(n^{-1/2}). \quad \blacksquare$$

Remark 3.3 Theorem 3.1 implies that for $\alpha \in (1/\beta, 2 - 1/\beta)$ the estimation of the regression function has no 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 (2001). In the simple case of a mean regression model with equidistant design and an error distribution F with bounded density f one has

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

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 (2009)). $\blacksquare$

In order to obtain asymptotic results for estimators of the error distribution for $\alpha \geq 2 - \frac{1}{\beta}$, a finer analysis is needed. In what follows, we will use the smooth regression estimator $\tilde{g}$ defined in (2.5). Let $\tilde{F}_n$ denote the empirical distribution function based on residuals $\tilde{\varepsilon}_j = Y_j - \tilde{g}(\frac{j}{n})$, i. e.

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

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

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

$$\tilde{F}_n(y) = \frac{1}{n} \sum_{j=1}^{n} I\{\varepsilon_j \leq y\} + \frac{1}{m_n} \sum_{j=1}^{n} \left(F\left(y + (\tilde{g} - g)(\tfrac{j}{n})\right) - F(y)\right) I\{\tfrac{j}{n} \in I_n\} + o_P\Big(\frac{1}{\sqrt{n}}\Big) \quad (3.2)$$

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.$\boldsymbol{\delta}$**) are fulfilled for some $\delta \in (1/\alpha - 1, 1 \wedge (\beta - 1))$ 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.$\boldsymbol{\delta}$**) 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. 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 $E_{g\equiv 0}(\hat{g}(1/2))$ 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, then one may calculate or simulate this expectation under the given null distribution. We define a bias corrected version of the smoothed estimator by

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

for $x \in I_n$. 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 $x \in [h_n, 1 - h_n]$*

$$E_{g\equiv 0}(|\hat{g}_n(x)|) = E_{g\equiv 0}(|\hat{g}_n(1/2)|) = O\Big(\Big(\frac{\log n}{nh_n}\Big)^{1/\alpha}\Big).$$

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

(**H2**) $h_n = o\big(n^{-1/(2\beta)} \wedge n^{-1/(\alpha\beta+1)}\big), \quad n^{\alpha/4-1} \log n = o(h_n)$

(**B3**) $b_n = o\Big(n^{-1/(2\beta)} \wedge \big(h_n^{-2\beta} n^{-1}\big) \wedge \Big(\Big(\frac{nh_n}{\log n}\Big)^{2/\alpha} n^{-1}\Big)\Big)$

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* ***(B3)*** *hold and* F *has a bounded density on* $(-\infty, \kappa]$ *for some* $\kappa < 0$. *Then*

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

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

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

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 $\mathcal{F} = \{F_\vartheta \mid \vartheta \in \Theta\}$ 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 $\hat{\vartheta}$ denote an estimator for ϑ based on residuals $\hat{\varepsilon}_i = Y_i - \hat{g}(\frac{i}{n})$, $i = 1, \dots, n$. The goodness-of-fit test is based on the empirical process

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

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 $\hat{\vartheta}$ 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, e. g., 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

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

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

Example 4.1 Consider the mean regression model $Y_i = m(\frac{i}{n}) + \eta_i$, $i = 1, \dots, n$, with symmetric error cdf F and $\beta > 1$, and define $\hat{m}$ with some bandwidth $h_n \asymp ((\log n)/n)^{1/(\alpha\beta+1)}$ as in Remark 2.5. We want to test the null hypothesis $H_0 : F \in \mathcal{F} = \{F_\vartheta \mid \vartheta \in \Theta\}$ 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 $\hat{\eta}_i = Y_i - \hat{m}(\frac{i}{n})$, $i = 1, \dots, n$, and let

$$\hat{\vartheta}_n = \max\Big(\max_{nh_n\le i\le n-nh_n} \hat{\eta}_i, -\min_{nh_n\le i\le n-nh_n} \hat{\eta}_i\Big) = \max_{nh_n\le i\le n-nh_n} |\hat{\eta}_i|.$$

Then $|\hat{\vartheta}_n - \vartheta|$ is bounded by $|\max_{nh_n\le i\le n-nh_n} |\eta_i| - \vartheta| + \sup_{x\in[h_n,1-h_n]} |\hat{m}(x) - m(x)| = o_P(n^{-1/2})$. 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 iid sample with distribution F. ■

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

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

$$\begin{aligned}
\hat{\vartheta}_n^\alpha - \vartheta^\alpha &= -(\hat{\vartheta}_n\vartheta)^\alpha\Big(\frac{1}{m_n}\sum_{j=1}^n((-\varepsilon_j)^\alpha - \vartheta^{-\alpha}) I\{h_n < \frac{j}{n} \le 1 - h_n\} \\
&\qquad + \frac{\alpha}{m_n}\sum_{j=1}^n (-\xi_j)^{\alpha-1}\Big(\hat{g}(\frac{j}{n}) - g(\frac{j}{n})\Big) I\{h_n < \frac{j}{n} \le 1 - h_n\}\Big) \\
&= -\vartheta^{2\alpha}\frac{1}{m_n}\sum_{j=1}^n((-\varepsilon_j)^\alpha - \vartheta^{-\alpha}) I\{h_n < \frac{j}{n} \le 1 - h_n\} + o_{P_\vartheta}(n^{-1/2})
\end{aligned}$$

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

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.

For all $z, \tilde{z} \in \mathbb{R}$ one has $|e^{-z} - e^{-\tilde{z}} - (z-\tilde{z})e^{-z}| = e^{-z}|e^{z-\tilde{z}} - 1 - (z-\tilde{z})| \leq e^{-z \wedge \tilde{z}}(z-\tilde{z})^2$. Thus

$$\left|F_{\hat{\vartheta}_n}(y) - F_\vartheta(y) - e^{-(-\vartheta y)^\alpha}\left(\hat{\vartheta}_n^\alpha - \vartheta^\alpha\right)^2\right| \leq e^{-(-(\hat{\vartheta}_n \wedge \vartheta)y)^\alpha}\left((\hat{\vartheta}_n^\alpha - \vartheta^\alpha)(-y)^\alpha\right)^2 = O_{P_\vartheta}(n^{-1})$$

uniformly for all $y \in (-\infty, 0]$. Now analogously to the proof of Theorem 19.23 in van der Vaart (2000) we can conclude weak convergence of

$$S_n(y) \; = \; \sqrt{n}(F_n(y) - F(y)) - e^{-(-\vartheta y)^\alpha}\vartheta^{2\alpha}(-y)^\alpha \frac{\sqrt{n}}{m_n}\sum_{j=1}^n((-\varepsilon_j)^\alpha - \frac{1}{\vartheta^\alpha})I\{h_n < \frac{j}{n} \leq 1 - h_n\}$$
$$+ o_{P_\vartheta}(1),$$

$y \in \mathbb{R}$, 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 $E_\vartheta[I\{\varepsilon_1 \leq y\}((-\varepsilon_1)^\alpha - \vartheta^{-\alpha})] = (-y)^\alpha e^{-(-\vartheta y)^\alpha}$.

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 (1976). ■

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 is up to a log term of optimal rate for $\alpha = 1$ and $\beta = 2$. 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 $H_0 : \exists\vartheta : -\varepsilon_i \sim Exp(\vartheta)$ 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.

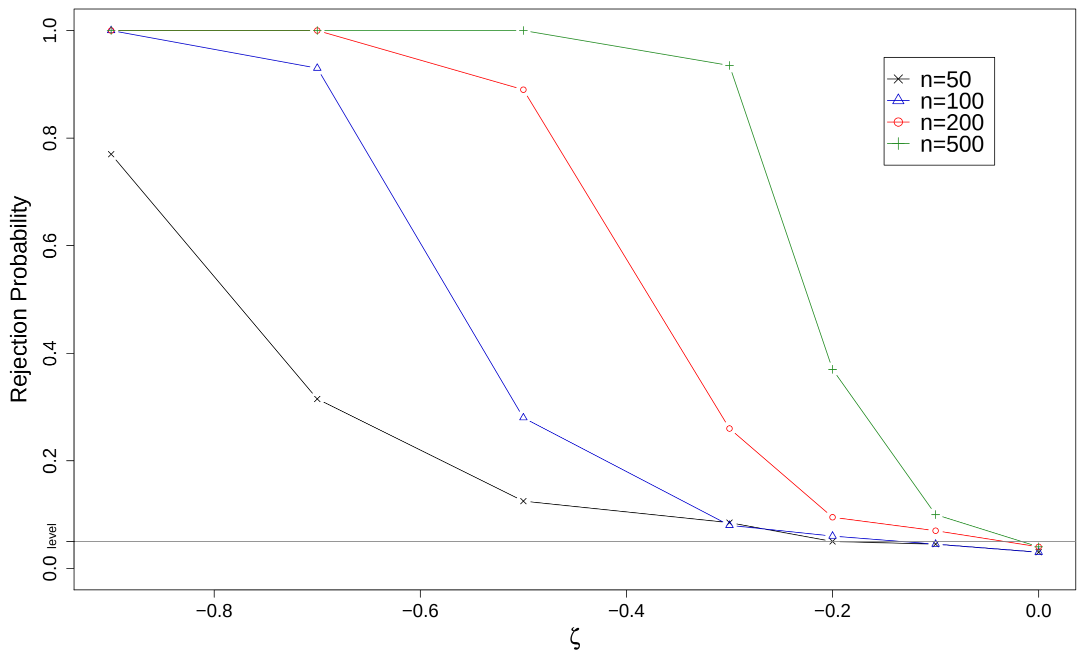

Figure 3: Monte-Carlo simulations for Example 4.1

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 $\hat{F}_n(y, s)$ for the residuals with the estimator $\bar{s}_n \hat{F}_n(y)$, which should behave similarly if the errors are iid. The following corollary to Theorem 3.1 describes the asymptotic behavior of the Kolmogorov-Smirnov type test statistic

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

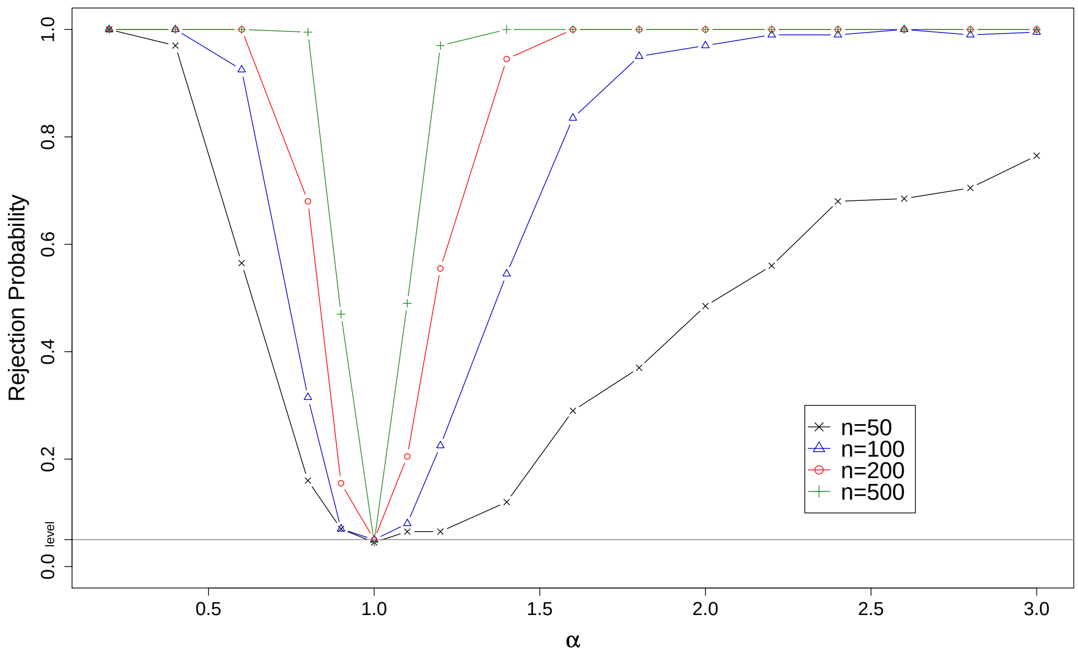

Figure 4: Monte-Carlo simulations for Example 4.2

under the null hypothesis of iid errors.

Corollary 4.3 *Assume model (2.1) with **(F1)**, **(F2)**, **(G1)**, and* $\frac{1}{\beta} < \alpha < 2 - \frac{1}{\beta}$. *Choose a bandwidth* $h_n \asymp ((\log n)/n)^{1/(\alpha\beta+1)}$.

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, i. e. 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 (1985) 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

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

which is a common assumption in boundary models. Let $\tilde{g}$ denote the smooth local polynomial estimator for g defined in (2.5). 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

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

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, 2007, and Chernozhukov et al., 2009). 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 iid, 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 iid-behaviour. The lower right panel shows $\varepsilon_{I,i} = Y_i - g_I(\frac{i}{n})$, $i = 1, \ldots, n$, with a clear non-iid pattern.

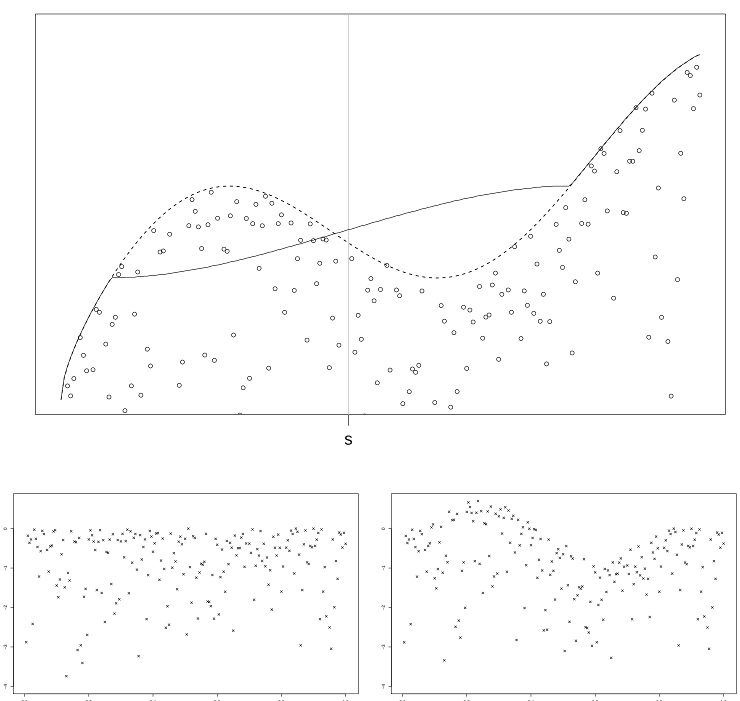

Figure 5: *The upper panel shows data points and the true boundary function (dashed curve) as well as the increasing rearrangement (solid curve). The lower left panel shows the errors. The lower right panel shows residuals built with respect to the increasing rearrangement.*

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

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

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

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

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

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

Theorem 4.4 *Assume model (2.1) with (**F1**), (**F2**), (**G1**), (**K1**), (**I1**), $\beta > 1$ and $\frac{1}{\beta} < \alpha < 2 - \frac{1}{\beta}$. If $h_n \asymp ((\log n)/n)^{1/(\alpha\beta+1)}$ and $b_n \asymp ((\log n)/n)^{1/(\alpha\beta+1)}$, then*

$$\sup_{y \in \mathbb{R}, s \in [0,1]} |\tilde{F}_{I,n}(y,s) - \bar{s}_n F_{\lfloor ns \rfloor}(y)| = o_P(n^{-1/2}). \tag{4.1}$$

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).

The conditions on the bandwidths can be substantially relaxed; cf. Remark 3.2.

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 (2007), $\sup_{x \in I_n} |\tilde{g}_I(x) - g_I(X)| \leq \sup_{x \in I_n} |\tilde{g}(x) - g(x)| = o_P(1)$ with g_I denoting the increasing rearrangement of g. Thus $n^{-1/2} T_n$ converges to

$$T = \sup_{s \in [0,1], y \in \mathbb{R}} \Big| \int_0^s F_\varepsilon(y + (g_I - g)(x))\, dx - s F_\varepsilon(y) \Big|.$$

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

A 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 $\beta \in (0, \beta^*]$ 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*

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

This proposition can be verified by an obvious modification of the proof of Theorem 3.1 by Jirak et al. (2014).

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

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

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

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

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 $x > 0$

$$\begin{aligned} P\Big\{\sup_{x\in[h_n,1-h_n]} \min_{\substack{i\in\{1,\dots,n\}\\ (i/n-x)/h_n\in I}} |\varepsilon_i| > x\Big\} &\le P\Big\{\max_{j\in\{1,\dots,n-d_n\}} \min_{i\in\{j,\dots,j+d_n\}} |\varepsilon_i| > x\Big\} \\ &\le P\Big\{\max_{\substack{l\in\{0,\dots,l_n\}\\ l \text{ even}}} M_{n,l} > x\Big\} + P\Big\{\max_{\substack{l\in\{0,\dots,l_n\}\\ l \text{ odd}}} M_{n,l} > x\Big\} \end{aligned}$$

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 iid, we have

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

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

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

$$\begin{aligned} P\{M_{n,0} > x\} &= P\Big\{\min_{i\in\{1,\dots,1+d_n\}} |\varepsilon_i| > x\Big\} + \sum_{j=2}^{d_n} P\Big\{|\varepsilon_{j-1}| \le x, \min_{i\in\{j,\dots,j+d_n\}} |\varepsilon_i| > x\Big\} \\ &= (1-G(x))^{d_n+1} + (d_n-1)G(x)(1-G(x))^{d_n+1} \\ &\le (1+d_nG(x))(1-G(x))^{d_n}. \end{aligned}$$

To sum up, we have shown that

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

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

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

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

$$\begin{aligned}
&G(Lr_n) = cL^\alpha \frac{|\log h_n|}{nh_n}(1 + o(1)) \\
&\implies (1 - G(Lr_n))^{d_n} = \exp\Big(- nh_n dcL^\alpha \frac{|\log h_n|}{nh_n}(1 + o(1))\Big) \\
&\implies (1 + d_n G(Lr_n))(1 - G(Lr_n))^{d_n} = O\Big(|\log h_n| \exp\big(- cdL^\alpha |\log h_n|(1 + o(1))\big)\Big) = o(h_n)
\end{aligned}$$

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 2.7

(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 $\tau_u \in (0, 1)$

$$\begin{aligned}
\sup_{x \in I_n} |\tilde{g}(x) - g(x)| &\leq \sup_{x \in I_n} \left| \int_{h_n}^{1-h_n} (\hat{g}(z) - g(z)) \frac{1}{b_n} K\left(\frac{x - z}{b_n}\right) dz \right| \\
&\quad + \sup_{x \in I_n} \left| \int_{h_n}^{1-h_n} (g(z) - g(x)) \frac{1}{b_n} K\left(\frac{x - z}{b_n}\right) dz \right| \\
&\leq \sup_{z \in [h_n, 1-h_n]} |\hat{g}(z) - g(z)| O(1) + \sup_{x \in I_n} \left| \int_{-1}^{1} (g(x - ub_n) - g(x)) K(u)\, du \right| \\
&\leq O(h_n^\beta) + O_P\Big(\Big(\frac{|\log h_n|}{nh_n}\Big)^{\frac{1}{\alpha}}\Big) \\
&\quad + b_n^{\lfloor \beta \rfloor} \sup_{x \in I_n} \left| \frac{1}{\lfloor \beta \rfloor !} \int_{-1}^{1} u^{\lfloor \beta \rfloor} (g^{(\lfloor \beta \rfloor)}(x - \tau_u u b_n) - g^{(\lfloor \beta \rfloor)}(x)) K(u) du \right|.
\end{aligned}$$

Now the Hölder property of g combined by (**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

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

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

$$\begin{aligned}
\sup_{x\in I_n}|\tilde g'(x)-g'(x)| \;\le\; & \sup_{x\in I_n}\left|\int_{h_n}^{1-h_n}(\hat g(z)-g(z))\frac{1}{b_n^2}K'\left(\frac{x-z}{b_n}\right)dz\right| \\
& + \sup_{x\in I_n}\left|\int_{h_n}^{1-h_n}(g'(z)-g'(x))\frac{1}{b_n}K\left(\frac{x-z}{b_n}\right)dz\right| \\
\le\; & \sup_{z\in[h_n,1-h_n]}|\hat g(z)-g(z)|O(b_n^{-1}) + \sup_{x\in I_n}\left|\int_{-1}^{1}(g'(x-ub_n)-g'(x))K(u)\,du\right|.
\end{aligned}$$

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 $|x-y|\le a_n$ 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

$$\begin{aligned}
& \sup_{x,y\in I_n,|x-y|>a_n}\frac{|\tilde g'(x)-g'(x)-\tilde g'(y)+g'(y)|}{|x-y|^\delta} \\
& \le\; 2\sup_{x\in I_n}|\tilde g'(x)-g'(x)|a_n^{-\delta} \\
& =\; \left(O(b_n^{\beta-1})+\left(O(h_n^\beta)+O_P\left(\left(\frac{|\log h_n|}{nh_n}\right)^{\frac{1}{\alpha}}\right)\right)b_n^{-1}\right)a_n^{-\delta}.
\end{aligned}\tag{A.1}$$

In the second case, we use a decomposition like in the proof of (ii):

$$\begin{aligned}
& \sup_{x,y\in I_n,0<|x-y|\le a_n}\frac{|\tilde g'(x)-g'(x)-\tilde g'(y)+g'(y)|}{|x-y|^\delta} \\
& \le \sup_{x,y\in I_n,0<|x-y|\le a_n}\frac{\left|\int_{h_n}^{1-h_n}(\hat g(z)-g(z))\frac{1}{b_n^2}\left(K'\left(\frac{x-z}{b_n}\right)-K'\left(\frac{y-z}{b_n}\right)\right)dz\right|}{|x-y|^\delta} \\
& \quad + \sup_{\substack{x,y\in I_n\\ 0<|x-y|\le a_n}}\frac{|g'(x)-g'(y)|}{|x-y|^\delta} + \sup_{\substack{x,y\in I_n\\ 0<|x-y|\le a_n}}\frac{\left|\int_{h_n}^{1-h_n}g'(z)\frac{1}{b_n}\left(K\left(\frac{x-z}{b_n}\right)-K\left(\frac{y-z}{b_n}\right)\right)dz\right|}{|x-y|^\delta}.
\end{aligned}$$

By Lipschitz continuity of K' and Theorem 2.2, the first term on the right hand side is of the order

$$\left(O(h_n^\beta)+O_P\Big(\Big(\frac{|\log h_n|}{nh_n}\Big)^{\frac{1}{\alpha}}\Big)\right)\frac{1}{b_n^3}O(a_n^{1-\delta}).\tag{A.2}$$

For $\beta\ge 2$, the second term is of the order $a_n^{1-\delta}$ as g' is Lipschitz continuous, while for $\beta\in(1,2)$ assumption (**G1**) yields the rate $a_n^{\beta-1-\delta}$. In both cases, condition ($\mathbf{B2.\boldsymbol{\delta}}$) ensures

that the second term converges to 0.
The last term on the right hand side can be rewritten as

$$\sup_{\substack{x,y\in I_n\\ 0<|x-y|\le a_n}} \frac{\left|\int_{-1}^{1}(g'(x-h_nu)-g'(y-h_nu))K(u)\,du\right|}{|x-y|^{\delta}}$$

and is thus of the same order as the second term by assumption (**G1**).

To conclude the proof, one needs to find a sequence $a_n = o(1)$ such that (A.1) and (A.2) tend to 0 in probability, i.e.

$$b_n^{\beta-1} + \frac{\vartheta_n}{b_n} = o(a_n^{\delta}) \text{ and } a_n^{1-\delta} = o\Big(\frac{b_n^3}{\vartheta_n}\Big)$$

with $\vartheta_n := h_n^{\beta} + (|\log h_n|/(nh_n))^{1/\alpha}$. Obviously, such a sequence a_n exists if and only if

$$b_n^{\beta-1} + \frac{\vartheta_n}{b_n} = o\Big(\Big(\frac{b_n^3}{\vartheta_n}\Big)^{\frac{\delta}{1-\delta}}\Big),$$

which in turn is equivalent to condition ($\mathbf{B2.}\boldsymbol{\delta}$). $\square$

A.4 Proof of Theorem 3.1

The assumptions about α ensure that $\beta/(\alpha\beta+1) > 1/(2(\alpha\wedge 1))$, and so in view of (2.3) the uniform estimation error of $\hat{g}$ is stochastically of smaller order than $n^{-1/(2(\alpha\wedge 1))}$. Hence there exists a sequence

$$a_n = o(n^{-\frac{1}{2(\alpha\wedge 1)}}) \tag{A.3}$$

such that

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

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

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

we may conclude

$$\begin{aligned}\sqrt{n}(\bar{F}_n(y-a_n,s) - \bar{s}_nF_{\lfloor ns\rfloor}(y)) \;&\le\; \sqrt{n}(\hat{F}_n(y,s) - \bar{s}_nF_{\lfloor ns\rfloor}(y))\\ &\le\; \sqrt{n}(\bar{F}_n(y+a_n,s) - \bar{s}_nF_{\lfloor ns\rfloor}(y))\end{aligned}$$

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

$$E_n(y,s) = n^{-1/2}\sum_{j=1}^{\lfloor ns\rfloor}(I\{\varepsilon_j\le y\} - F(y)), \quad y\in\mathbb{R}, s\in[0,1], \tag{A.4}$$

converges weakly to a Kiefer process; see e.g. Theorem 2.12.1 in van der Vaart and Wellner (1996). Now, $n \sim m_n$, the asymptotic equicontinuity of the process E_n, the Hölder continuity (**F2**) and (A.3) imply

$$\begin{aligned}
&\sqrt{n}\left(\bar{F}_n(y \pm a_n, s) - \bar{s}_n F_{\lfloor ns \rfloor}(y)\right) \\
&= \frac{n}{m_n}\Big(E_n(y \pm a_n, s \wedge (1-h_n)) - E_n(y, s \wedge (1-h_n)) - E_n(y \pm a_n, s \wedge h_n) + E_n(y, s \wedge h_n)\Big) \\
&\quad + \sqrt{n}\bar{s}_n(F(y \pm a_n) - F(y)) + \sqrt{n}(\bar{F}_n(y,s) - \bar{s}_n F_{\lfloor ns \rfloor}(y)) \\
&= o_P(1) + \sqrt{n}(\bar{F}_n(y,s) - \bar{s}_n F_{\lfloor ns \rfloor}(y))
\end{aligned}$$

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

It remains to be shown that

$$\begin{aligned}
&\sqrt{n}(\bar{F}_n(y,s) - \bar{s}_n F_{\lfloor ns \rfloor}(y)) \\
&= \frac{\sqrt{n}}{m_n}\sum_{j=1}^{\lfloor ns \rfloor}(I\{\varepsilon_j \le y\} - F(y)) I\{h_n < \frac{j}{n} \le 1 - h_n\} - \frac{\sqrt{n}\bar{s}_n}{\lfloor ns \rfloor}\sum_{j=1}^{\lfloor ns \rfloor}(I\{\varepsilon_j \le y\} - F(y)) \\
&= \Big(\frac{n}{m_n} - 1\Big)\big(E_n(y, s \wedge (1-h_n)) - E_n(y, s \wedge h_n)\big) \\
&\quad - \Big(\frac{n\bar{s}_n}{\lfloor ns \rfloor} - 1\Big)E_n(y,s) \\
&\quad + \big(E_n(y, s \wedge (1-h_n)) - E_n(y, s \wedge h_n) - E_n(y,s)\big) \qquad \text{(A.5)}
\end{aligned}$$

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 $s \ge h_n$

$$\frac{n\bar{s}_n}{\lfloor ns \rfloor} - 1 = \frac{\lfloor n(s \wedge (1-h_n)) \rfloor - \lfloor nh_n \rfloor}{(1 - 2h_n + O(n^{-1}))\lfloor ns \rfloor} - 1 = \frac{O(nh_n)}{(1 - 2h_n + O(n^{-1}))\lfloor ns \rfloor}, \qquad \text{(A.6)}$$

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 $\sup_{0 \le s \le h_n^{1/2}, y \in \mathbb{R}} |E_n(y,s)| = o_P(1)$, because E_n is asymptotically equicontinuous with $E_n(y,0) = 0$. Hence, the second term in (A.5) converges to 0 in probability, too. Likewise, the convergence of the last term to 0 follows from the asymptotic equicontinuity of E_n, which concludes the proof. □

A.5 Proof of Theorem 3.4 and of Remark 3.5

For any interval $I \subset \mathbb{R}$ and constant $k > 0$, define the following class of differentiable functions:

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

Then Theorem 2.7 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 $P\left(d_n\in C_1^{1+\delta}([0,1])\right)\to 1$ 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

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

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

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

and

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

Note that

$$\begin{aligned}
&G_n(y,d_n)\\
&=\frac{\sqrt{n}}{m_n}\sum_{j=1}^n I\{\varepsilon_j\le y-g(j/n)+\tilde{g}(j/n)\}I\{j/n\in I_n\}\\
&\quad-\frac{\sqrt{n}}{m_n}\sum_{j=1}^n F(y+(\tilde{g}-g)(j/n))I\{j/n\in I_n\}-\frac{1}{\sqrt{n}}\sum_{j=1}^n I\{\varepsilon_j\le y\}+\sqrt{n}F(y)\\
&=\sqrt{n}\Big(\tilde{F}_n(y)-\frac{1}{n}\sum_{j=1}^n I\{\varepsilon_j\le y\}-\frac{1}{m_n}\sum_{j=1}^n\left(F\left(y+(\tilde{g}-g)(\tfrac{j}{n})\right)-F(y)\right)I\{\tfrac{j}{n}\in I_n\}\Big).
\end{aligned}$$

We will apply Theorem 2.11.9 of van der Vaart and Wellner (1996) 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, which readily yields the assertion, because $\sup_{y\in\mathbb{R}}\rho((y,d_n),(y,0))=\sup_{x\in[0,1]}|d_n(x)|=o_P(1)$ and the variance of

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

tends to 0, implying that $G_n(y,0)=o_P(1)$ uniformly in y. Thus $G_n(y,d_n)=o_P(1)$ uniformly in y and the assertion holds.

One may proceed as in the proof of Lemma 3 in Neumeyer and Van Keilegom (2009) (see the online supporting information to that article) to prove that the conditions of Theorem 2.11.9 of van der Vaart and Wellner (1996) are fulfilled. The proof of the first two displayed formulas of this theorem are analogous. The only difference is that Neumeyer and Van

Keilegom (2009) assume a bounded error density while we use Hölder continuity of F, see assumption (**F2**). Next we show that the bracketing entropy condition (i.e., the last displayed condition in Theorem 2.11.9 of van der Vaart and Wellner, 1996) 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 (1996), Theorem 2.7.1 and Corollary 2.7.2, $M = O\big(\exp(\kappa\eta^{-2/((1+\delta)(\alpha\wedge 1))})\big)$ 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

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

Check that by condition (**F2**)

$$\begin{aligned}
\sup_{y\in\mathbb{R}} |F_m^U(y) - F_m^L(y)| &\leq \sup_{y\in\mathbb{R}} n^{-1}\sum_{j=1}^n |F(y + d_m^U(j/n)) - F(y + d_m^L(j/n))| \\
&\leq L_F \sup_{x\in\mathbb{R}} |d_m^U(x) - d_m^L(x)|^{\alpha\wedge 1} \leq L_F\eta^2
\end{aligned} \tag{A.7}$$

with L_F denoting the Hölder constant of F. Thus

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^n E\Big[\sup_{(y,d),(y^*,d^*)\in\mathcal{F}_{mk}} \big|I\{\varepsilon_j \leq y + d(j/n)\} - I\{\varepsilon_j \leq y^* + d^*(j/n)\}\big|\Big]^2 \\
&\leq \frac{1}{n}\sum_{j=1}^n E\big[I\{\varepsilon_j \leq \tilde{y}_{m,k}^U + d_m^U(j/n)\} - I\{\varepsilon_j \leq \tilde{y}_{m,k}^L + d_m^L(j/n)\}\big]^2 \\
&\leq F_m^U(\tilde{y}_{m,k}^U) - F_m^L(\tilde{y}_{m,k}^L) \\
&\leq |F_m^U(\tilde{y}_{m,k}^U) - F_m^U(\tilde{y}_{m,k+1}^L)| + |F_m^U(\tilde{y}_{m,k+1}^L) - F_m^L(\tilde{y}_{m,k+1}^L)| + |F_m^L(\tilde{y}_{m,k+1}^L) - F_m^L(\tilde{y}_{m,k}^L)| \\
&\leq (2 + L_F)\eta^2
\end{aligned}$$

where the last step follows from (A.7) 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

$$\begin{aligned}
&\sum_{j=1}^n E\Big[\sup_{(y,d),(y^*,d^*)\in\mathcal{F}_{mk}} |Z_{nj}(y,d) - Z_{nj}(y^*,d^*)|\Big]^2 \\
&\leq 2\frac{n}{m_n^2}\sum_{j=1}^n E\Big[\sup_{(y,d),(y^*,d^*)\in\mathcal{F}_{mk}} \big|I\{\varepsilon_j \leq y + d(j/n)\} - I\{\varepsilon_j \leq y^* + d^*(j/n)\}\big|\Big]^2 I\{j/n \in I_n\} \\
&\quad + \frac{2}{n}\sum_{j=1}^n E\Big[\sup_{(y,d),(y^*,d^*)\in\mathcal{F}_{mk}} \big|I\{\varepsilon_j \leq y\} - I\{\varepsilon_j \leq y^*\}\big|\Big]^2
\end{aligned}$$

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

for sufficiently large n. This shows that the bracketing number satisfies $\log N_{[\,]}(\eta, \mathcal{F}, L_2^n) = O(\log M + \log K) = O\big(\eta^{-2/((1+\delta)(\alpha\wedge 1))}\big)$, and the last displayed condition of Theorem 2.11.9 of van der Vaart and Wellner (1996) follows from $\delta > 1/\alpha - 1$.

It remains to show that $(\mathcal{F}, \rho)$ is totally bounded, i.e. 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 $\mathcal{F}_{ml} := \{(y,d) \mid \bar{y}_{l-1} \leq y \leq \bar{y}_l, d_m^L \leq d \leq d_m^U\}$ have ρ-diameter less than 5η. Check that, for all $1 \leq l \leq L$, one has

$$\begin{aligned}
&\sup_{x\in[0,1]} \sup_{\gamma \in C_1^{1+\delta}([0,1])} |F(\bar{y}_l + \gamma(x)) - F(\bar{y}_{l-1} + \gamma(x))| \\
&\leq \max_{1\leq j\leq J} \sup_{s_{j-1}\leq x \leq s_j} \max_{1\leq m\leq M} \sup_{d_m^L \leq \gamma \leq d_m^U} \Big[|F(\bar{y}_l + \gamma(x)) - F(\bar{y}_l + \gamma(s_j))| \\
&\qquad + |F(\bar{y}_l + \gamma(s_j)) - F(\bar{y}_l + d_m^L(s_j))| + |F(\bar{y}_l + d_m^L(s_j)) - F(\bar{y}_{l-1} + d_m^L(s_j))| \\
&\qquad + |F(\bar{y}_{l-1} + d_m^L(s_j)) - F(\bar{y}_{l-1} + \gamma(s_j))| + |F(\bar{y}_{l-1} + \gamma(s_j)) - F(\bar{y}_{l-1} + \gamma(x))| \Big] \\
&< \max_{1\leq j\leq J} \Big[(s_j - s_{j-1})^{\alpha\wedge 1} + \eta^2 + \eta + \eta^2 + (s_j - s_{j-1})^{\alpha\wedge 1} \Big] \\
&\leq 5\eta.
\end{aligned}$$

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

$$\begin{aligned}
&\rho((y,d),(y^*,d^*)) \\
&\leq \max\Big\{ \sup_{x\in[0,1]} \sup_{\gamma\in C_1^{1+\delta}([0,1])} |F(\bar{y}_l + \gamma(x)) - F(\bar{y}_{l-1} + \gamma(x))|, \sup_{x\in[0,1]} d_m^U(x) - d_m^L(x) \Big\} \\
&\leq \max\{5\eta, \eta^{2/(\alpha\wedge 1)}\} = 5\eta,
\end{aligned}$$

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.7) with

$$\begin{aligned}
\sup_{y\in\mathbb{R}} |F_m^U(y) - F_m^L(y)| &\leq \sup_{y\in\mathbb{R}} n^{-1} \sum_{j=1}^n |F(y + d_m^U(j/n)) - F(y + d_m^L(j/n))| \\
&\leq L_F \sup_{x\in\mathbb{R}} |d_m^U(x) - d_m^L(x)| \leq L_F \eta^2
\end{aligned}$$

with L_F denoting the Lipschitz constant of F to prove $\log N_{[\,]}(\eta, \mathcal{F}, L_2^n) = O\big(\eta^{-2/(1+\delta)}\big)$, which again yields the third condition of Theorem 2.11.9 of van der Vaart and Wellner

(1996). Likewise, in the last part of the proof, one defines $s_j := j\eta \wedge 1$ and replaces (A.8) with $\max_{1\le j\le J}\big[(s_j - s_{j-1}) + \eta^2 + \eta + \eta^2 + (s_j - s_{j-1})\big] \le 5\eta$. □

In the remaining proofs to Section 3, we use the index n for the estimators to emphasis the dependence on the sample size and to distinguish between estimators and polynomials corresponding to a given sample on the one hand and corresponding objects in a limiting setting on the other hand.

A.6 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 $E(\hat{g}_n(x)) \le \tilde{d}E(M_{n,0})$ and $P\{M_{n,0} > t\} \le \big(1 + dnh_n(1 - F(-t))\big)(F(-t))^{dnh_n}$ 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

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

Now, for sufficiently large n,

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

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

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

for all $\xi > 0$. □

A.7 Proof of Theorem 3.7

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

$$\begin{aligned}\big|F\big(y+(\tilde g_n^*-g)(j/n)\big)-F(y)-f(y)(\tilde g_n^*-g)(j/n)\big| &= \Big|\int_0^{(\tilde g_n^*-g)(j/n)} f(y+t)-f(y)\,dt\Big| \\ &= O\big((\tilde g_n^*-g)^2(j/n)\big)\end{aligned}$$

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:

$$\begin{aligned}&\Big|\frac{1}{m_n}\sum_{j=1}^n\big(F\big(y+(\tilde g_n^*-g)(j/n)\big)-F(y)\big)I\{j/n\in I_n\}-\frac{f(y)}{m_n}\sum_{j=1}^n(\tilde g_n^*-g)(j/n)I\{j/n\in I_n\}\Big| \\ &= O\Big(\frac{1}{m_n}\sum_{j=1}^n(\tilde g_n^*-g)^2(j/n)I\{j/n\in I_n\}\Big)=O_P\Big(h_n^{2\beta}+b_n^{2\beta}+\Big(\frac{\log n}{nh_n}\Big)^{2/\alpha}\Big)=o_P(n^{-1/2})\end{aligned}$$

where for the last conclusions we have used Theorem 2.2, Lemma 3.6 and the assumptions (**H2**) and (**B3**). Thus the assertion follows if we show that

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

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

$$\begin{aligned}\frac{1}{m_n}\sum_{\ell=1}^{\lfloor n/(2k_n)\rfloor}\Delta_{n,2\ell-1} &= o_P(n^{-1/2})\\ \frac{1}{m_n}\sum_{\ell=1}^{\lfloor n/(2k_n)\rfloor}\Delta_{n,2\ell} &= o_P(n^{-1/2})\end{aligned}$$

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

It suffices to verify

$$E\big(\Delta_{n,2\ell}^2\big) = o(k_n) \tag{A.8}$$

$$E\big(\Delta_{n,2\ell}\big) = o\big(n^{-1/2}k_n\big)=o\big(n^{1/2}(h_n+b_n)\big) \tag{A.9}$$

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

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

which implies the assertion.

To prove (A.8), note that according to Lemma 3.6, Proposition A.1 and the proofs of Lemma A.2 and of Theorem 2.7(i), there exist constants $c_1, c_2, c_3 > 0$ (depending only on β, c_g and the kernel K) such that

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

where M_1^*, M_2^* are independent random variables such that $P\{M_i^* > t\} \le 1-(1-P\{M_{n,0} > t\})^{c_2(h_n+b_n)/h_n}$ with

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

Because $k_n\big(h_n^\beta + b_n^\beta + (\log n/(nh_n))^{1/\alpha}\big) = o(k_n^{1/2})$ by (**H2**) and (**B3**), it suffices to show that

$$E\big((M_i^*)^2\big) = \int_0^\infty P\{M_i^* > t^{1/2}\}\, dt = o(1/k_n). \tag{A.10}$$

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

$$\begin{aligned}
P\{M_i^* > t^{1/2}\} \;&\le\; 1 - \big[1-(1+c_3 nh_n 2ct^{\alpha/2})(1-ct^{\alpha/2}/2)^{c_3 nh_n}\big]^{c_2(h_n+b_n)/h_n} \\
&\le\; 1 - \big[1 - 3c_3 cnh_n t^{\alpha/2} \exp\big(-c_3 cnh_n t^{\alpha/2}/2\big)\big] \\
&\le\; 1 - \exp\Big(-dn(h_n+b_n)t^{\alpha/2} \exp\big(-c_3 cnh_n t^{\alpha/2}/2\big)\Big) \\
&\le\; dn(h_n+b_n)t^{\alpha/2} \exp\big(-c_3 cnh_n t^{\alpha/2}/2\big).
\end{aligned}$$

Therefore, for sufficiently large $a > 0$,

$$\begin{aligned}
&\int_0^{t_0^2} P\{M_i^* > t^{1/2}\}\, dt \\
&\le\; a\Big(\frac{nh_n}{\log n}\Big)^{-2/\alpha} + dt_0 n(h_n+b_n) \int_{a(nh_n/\log n)^{-2/\alpha}}^{t_0} t^{\alpha/2-1} \exp\big(-c_3 cnh_n t^{\alpha/2}/2\big)\, dt \\
&\le\; o(1/(n(h_n+b_n))) + dt_0 n(h_n+b_n) \exp\big(-c_3 ca^{\alpha/2} \log n/2\big) \\
&=\; o(1/(n(h_n+b_n)))
\end{aligned} \tag{A.11}$$

where in the last but one step we apply the conditions (**B3**) and (**H2**). Now, assertion (A.10) (and hence (A.8)) follows from

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

$$
\begin{aligned}
&\le \int_{t_0^2}^\infty 1-\exp\Big(-dn(h_n+b_n)(F(-t^{1/2}))^{c_3 nh_n}\Big)\,dt\\
&\le dn(h_n+b_n)\Big(nh_n(F(-t_0))^{c_3nh_n}+\int_{nh_n}^\infty t^{-\tau c_3 nh_n/2}\,dt\Big)\\
&= o(n^{-\xi})
\end{aligned}
$$

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

To establish (A.9), first note that for a kernel K of order $d+1$ with $d:=\lfloor\beta\rfloor$

$$
\begin{aligned}
E(\tilde g_n(x)-g(x)) &= E\Big(\int_{-1}^1\Big(\hat g_n(x+b_nu)-\sum_{j=0}^d\frac{g^{(j)}(x)}{j!}(b_nu)^j\Big)K(u)\,du\\
&= \int_{-1}^1 E(\hat g_n(x+b_nu)-g(x+b_nu))K(u)\,du+O(b_n^\beta)
\end{aligned}
$$

uniformly for all $x\in[h_n+b_n,1-h_n-b_n]$. In view of (**K1**), (**H2**) and (**B3**), it thus suffices to show that

$$
\big|E(\hat g_n(x)-g(x))-E_{g\equiv 0}(\hat g_n(1/2))\big|=\big|E(\hat g_n(x)-g(x))-E_{g\equiv0}(\hat g_n(x))\big|=o(n^{-1/2}) \quad (A.12)
$$

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 d that solves the linear optimization problem

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

under the constraints

$$
\tilde p_n\Big(\frac{i/n-x}{h_n}\Big)\ge Y_i,\quad \forall\, i\in[n(x-h_n),n(x+h_n)].
$$

Define polynomials

$$
q_x(t):=\sum_{k=0}^d\frac{1}{k!}g^{(k)}(x)(h_nt)^k,\quad p_n(t):=(nh_n)^{1/\alpha}(\tilde p_n(t)-q_x(t)),\quad t\in[-1,1].
$$

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

$$
\hat g_n(x)-g(x)=(nh_n)^{-1/\alpha}p_n(0). \quad (A.13)
$$

Note that p_n is a polynomial of degree d that solves the linear optimization problem

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

subject to

$$p_n\Big(\frac{i/n-x}{h_n}\Big) \geq (nh_n)^{1/\alpha}\bar{\varepsilon}_i, \quad \forall\, i \in [n(x-h_n), n(x+h_n)], \tag{A.14}$$

with

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

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.14) can be reformulated as $N_n(A_{p_n}) = 0$ where $A_f := \{(t,u) \in [-1,1]\times\mathbb{R} \mid u > f(t)\}$ 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

$$E\big(N_n([-1,1]\times(-1,\infty))\big) \sim 2nh_n P\big\{\bar{\varepsilon}_1 > -(nh_n)^{-1/\alpha}\big\} \to 2c.$$

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 (2007), 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 d 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 $J := \{j \in \mathbb{N} \mid p(T_j) = Z_j\}$ has at most d elements. Because p is bounded and N has a.s. finitely many points in any bounded set, $\eta := \inf\{|p(T_i) - Z_i| \mid i \in \mathbb{N}\setminus J\} > 0$ a.s. Since p is an optimal solution, all polynomials Δ of degree d 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 $d-|J|$, $\Delta(t) = \tau\prod_{i\in J}(t-T_i)q(t)$ 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 $j \in \{0, \dots d - |J|\}$. This implies that $(T_i)_{i\in J}$ lies on a manifold $M_{|J|,d}$ of dimension $|J| - (d - |J| + 1) = 2|J| - d - 1$ which only depends on $|J|$ and d. However, by Proposition A.1, $\|p\|_\infty \le K_d Z_{\max}$ where

$$Z_{\max} := \max_{1\le i\le j_d} \min\{|Z_i| \mid T_i \in [-1 + (j-1)/j_d, -1 + j/j_d]\}.$$

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

$$P\bigl\{\exists J \subset \mathbb{N} : |J| \le d, (T_j)_{j\in J} \in M_{|J|,d}, \max_{j\in J} |Z_j| \le K_d K\bigr\} = 0$$

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, i.e., $|J| \ge d+1$. On the other hand, because the intensity measure of N is absolutely continuous, $|J| \le d+1$ a.s. and thus $|J| = d+1$. Because of $N_n \to N$ a.s., one has $N_n([-1,1] \times [-K_d Z_{\max}, \infty)) = N([-1,1] \times [-K_d Z_{\max}, \infty)) =: M$ 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 $[-1,1] \times [-K_d Z_{\max}, \infty)$ 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 $d+1$ 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 d 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, 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 $\bigl(T_{n,i}, Z_{n,i} - (nh_n)^{1/\alpha}(g(i/n) - q_x((i/n) - x)/h_n)\bigr)$ 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.13) 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.12) and thus the assertion. □

A.8 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.4). A similar reasoning as in the proof of Theorem 3.1 (see (A.6)) shows that

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

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

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

which converges weakly to $K_F(y,s)-sK_F(y,1)$ 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.9 Proof of Theorem 4.4

Note that under the given assumptions, the statements of Theorem 2.7 (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 2.7 (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 Theorem 2.7 (i) yields $\sup_{x\in I_n}|\tilde{g}(x)-g(x)| = o_P(n^{-1/(2(\alpha\wedge 1))})$. The convergence of the Kolmogorov-Smirnov test statistic then follows exactly as in the proof of Corollary 4.3. □

Acknowledgement

Financial support by the DFG (Research Unit FOR 1735 *Structural Inference in Statistics: Adaptation and Effciency*) is gratefully acknowledged.

References

Akritas, M. and Van Keilegom, I. (2001). Nonparametric estimation of the residual distribution. *Scand. J. Statist.* **28**, 549–567.

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

Birke, M. and Neumeyer, N. (2013). Testing Monotonicity of Regression Functions - An Empirical Process Approach. *Scand. J. Statist.* **40**, 438–454.

Birke, M., Neumeyer, N. and Volgushev, S. (2016+). The independence process in conditional quantile location-scale models and an application to testing for monotonicity. Statistica Sinica, to appear.

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

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

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

Färe, R. and Grosskopf, S. (1983). Measuring output efficiency. *European Journal of Operational Research* **13**, 173–179.

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

Gijbels, I., Mammen, E., Park, B. and Simar, L. (2000). On estimation of monotone and concave frontier functions. *J. Amer. Statist. Assoc.* **94**, 220–228.

Gijbels, I. and Peng, L. (2000). Estimation of a support curve via order statistics. *Extremes* **3**, 251–277.

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

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

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

Hall, P. and Van Keilegom, I. (2009). Nonparametric "regression" when errors are positioned at end-points. *Bernoulli* **15**, 614–633.

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

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

Meister, A. and Reiß, M. (2013). Asymptotic equivalence for nonparametric regression with non-regular errors. *Probab. Th. Rel. Fields* **155**, 201–229.

Müller, U.U. and Wefelmeyer W. (2010). Estimation in Nonparametric Regression with Non-Regular Errors. *Comm. Statist. Theory Methods* **39**, 1619–1629.

Neumeyer, N. and Van Keilegom, I. (2009). Change-Point Tests for the Error Distribution in Nonparametric Regression. *Scand. J. Statist.* **36**, 518–541.
online supporting information available at
http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9469.2009.00639.x/suppinfo

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

Reiß, M. and Selk, L. (2016+). Efficient nonparametric functional estimation for one-sided regression. Bernoulli, to appear.

Resnick, S.I. (2007). *Heavy-Tail Phenomena.* Springer.

Simar, L. and Wilson, P.W. (1998). Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Management Science* **44**, 49–61.

Stephens, M.A. (1976). Asymptotic Results for Goodness-of-Fit Statistics with Unknown Parameters. *Ann. Statist.* **4**, 357–369.

van der Vaart, A.W. (2000). *Asymptotic Statistics.* Cambridge University Press.

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

Wilson, P.W. (2003). Testing independence in models of productive efficiency. *Journal of Productivity Analysis* **20**, 361–390.