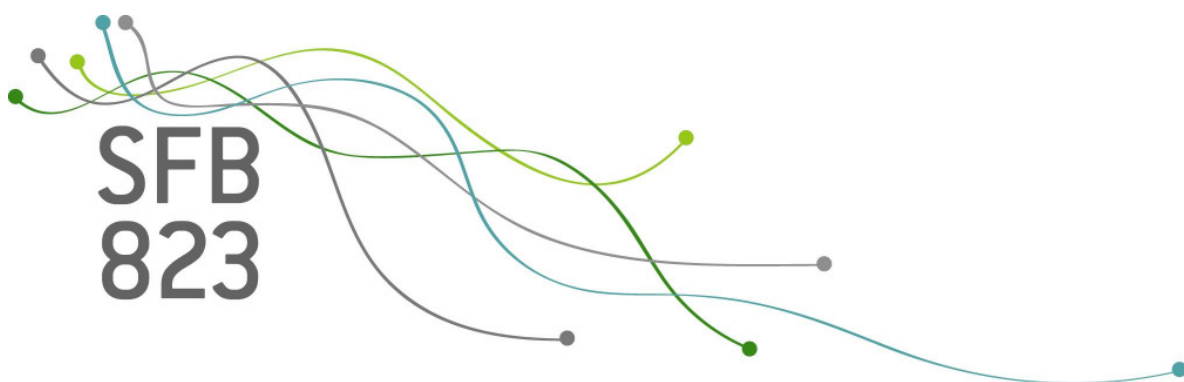# Testing monotonicity of regression functions – an empirical process approach

Melanie Birke, Natalie Neumeyer

Testing monotonicity of regression functions – an empirical process approach

Melanie Birke

Ruhr-Universität Bochum

Fakultät für Mathematik

Universitätsstraße 150

44780 Bochum, Germany

e-mail: melanie.birke@ruhr-uni-bochum.de

Natalie Neumeyer

Universität Hamburg

Department Mathematik

Bundesstraße 55

20146 Hamburg, Germany

e-mail: neumeyer@math.uni-hamburg.de

April 6, 2010

Abstract

We propose several new tests for monotonicity of regression functions based on different empirical processes of residuals. The residuals are obtained from recently developed simple kernel based estimators for increasing regression functions based on increasing rearrangements of unconstrained nonparametric estimators. The test statistics are estimated distance measures between the regression function and its increasing rearrangement. We discuss the asymptotic distributions, consistency, and small sample performances of the tests.

AMS Classification: 62G10, 62G08, 62G30
Keywords and Phrases: Kolmogorov-Smirnov test, model test, monotone rearrangements, nonparametric regression, residual processes

1   Introduction

In a nonparametric regression context with regression function m we consider the important problem of testing for monotonicity of the regression function, i. e. testing for validity of the null hypothesis

$$H_0 : \text{``m is increasing''}.$$

First literature on testing for monotonicity of regression function is given by Schlee (1982) who proposes a test which is based on estimates of the derivative of the regression function. Bowman, Jones and Gijbels (1998) used Silverman's (1981) "critical bandwidth" approach to construct a bootstrap test for monotonicity, while Gijbels, Hall, Jones and Koch (2000) considered the length or runs of consecutive negative values of observation differences. Hall and Heckman (2000) suggested to fit straight lines through subsequent groups of consecutive points and reject monotonicity for too large negative values of the slopes. Other recent work on testing monotonicity can be found in Goshal, Sen and Van der Vaart (2000), Dümbgen (2002), Durot (2003), Baraud, Huet and Laurent (2003) and Domínguez-Menchero, González-Rodríguez and López-Palomo (2005). Birke and Dette (2007) consider a test for strict monotonicity based on the L^2-distance of the distribution function of the unconstrained estimator evaluated at the unconstrained estimator to the identity (see section 3 for more comments on that test). Most tests for monotonicity suffer from the problem of underestimating the level because they are calibrated with the most difficult null model which is a constant regression function. Gijbels (2005) gives a thorough review on tests for monotonicity of regression functions and suggests as alternative "to base a test statistic on a measure of the distance between an unconstrained and a constrained estimate of the regression function". However, to the authors' best knowlegde in the paper at hand for the first time those tests for monotonicity are investigated. For similar testing problems (for instance, testing whether m belongs to a parametric class of functions) such tests based on an estimated norm difference between an estimator under H_0 and an unconstrained estimator are very popular. In our context, such tests would be based, for example, on an estimated L^2-distance between a completely nonparametric regression estimator $\hat{m}$ and an increasing estimator $\hat{m}_I$ constructed under the null hypothesis; see Härdle and Mammen (1993) for such a test in the goodness-of-fit context. Other test statistics could be constructed similar to the goodness-of-fit tests by Stute (1997) or Van Keilegom, González-Manteiga and Sánchez Sellero (2008) based on estimated empirical processes of residuals (see section 3 for exact definitions of those processes). To do so one needs an estimator for the regression function under the null hypothesis H_0 of monotonicity. Such increasing regression estimators were proposed by Mammen (1991), Hall and Huang (2001), and Dette, Neumeyer and Pilz (2006), among others. The mentioned methods have in common that they are based on a preliminary unconstrained estimator $\hat{m}$ and are (under appropriate assumptions) first order asymptotically equivalent to each other and to the unconstrained estimator $\hat{m}$. This is a nice and desirable property for estimation purposes, but it limits the application of such estimators for testing monotonicity by distance based tests as suggested before. It turns out that those typical distance based test that are so popular in testing for different model assumptions have, when testing for monotonicity, degenerate limit distributions under the

null hypothesis and, hence, are not suitable for testing.

The intuitive idea we follow in the present paper instead is as follows. We investigate the behaviour of pseudo-residuals built under the hypothesis H_0 of monotonicity, which estimate pseudo-errors that coincide with the true errors in general only under the null hypothesis. Whereas the true errors are assumed to be independent and identically distributed, the pseudo-errors behave differently. We construct several test statistics to detect these different behaviours. The test statistics are based on several empirical processes of (pseudo-)residuals. To build the pseudo-residuals, we estimate the regression function under H_0 by applying the simple kernel based increasing estimator by Dette, Neumeyer and Pilz (2006) [see also Birke and Dette (2008) for further discussion of this estimator]. Under the null hypothesis we show weak convergence of the empirical processes to Gaussian processes. The asymptotic distributions are independent of the regression function m and, hence, the tests need not be calibrated using a most difficult null model. For normal regression models we can even obtain asymptotically distribution free tests. The test statistics turn out to be estimators for certain distance measures between the true regression function m and an "increasing version" m_I of m, and hence, are consistent. Moreover, the tests can detect local alternatives of convergence rate $n^{-1/2}$. To the authors' best knowledge those are the first tests in literature for testing monotonicity of regression functions with this property. We compare the small sample behavior of the empirical process approaches to that of the L^2-test in Birke and Dette (2007) and observe, that they are less conservative and can in fact better detect local alternatives of convergence rate $n^{-1/2}$.

The paper is organized as follows. In section 2 we define the monotone regression estimator and list the model assumptions. Section 3 motivates and defines the test statistics, for which the asymptotic distributions are stated in section 4. In section 5 we explain bootstrap versions of the tests and investigate the small sample behaviour, whereas some concluding remarks are given in section 6. All proofs are given in an appendix.

2 Model and assumptions

Consider the nonparametric regression model

$$Y_i = m(X_i) + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where (X_i, Y_i), $i = 1, \ldots, n$, is a bivariate sample of i.i.d. observations. If there is evidence that the regression function m is increasing we define

$$\hat{m}_I^{-1}(t) = \frac{1}{b_n} \int_0^1 \int_{-\infty}^t k\Big(\frac{\hat{m}(v) - u}{b_n}\Big) du dv$$

as an estimate of $m^{-1}(t)$, where $\hat{m}$ denotes a local linear estimator for m with kernel K and bandwidth h_n. [By increasing throughout the paper we mean nondecreasing in distinction from strictly increasing.]

The estimator $\hat{m}_I$ is defined as the generalized inverse of $\hat{m}_I^{-1}$, that is

$$\hat{m}_I(x) = \inf\{t \in I\!\!R | \hat{m}_I^{-1}(t) \geq x\}.$$

Under H_0 and under the following assumptions $\hat{m}_I$ is asymptotically first order equivalent to the unconstrained estimator $\hat{m}$, see Dette, Neumeyer and Pilz (2006). This is a smoothed version of the monotone rearrangement (see e.g. Hardy, Littlewood and Polyá, 1952 or Lieb and Loss, 2001)

(A1) The covariates $X_1, \ldots, X_n$ are independent and identically distributed with distribution function F_X on compact support, say $[0, 1]$. F_X has a twice continuously differentiable density f_X such that $\inf_{x \in [0,1]} f_X(x) > 0$. The regression function m is twice continuously differentiable.

(A2) K and k are symmetric, twice continuously differentiable kernels of order 2 with bounded supports, say $(-1, 1)$, such that $K(-1) = K(1) = k(-1) = k(1) = 0$.

(A3) The bandwidths fulfill h_n, $b_n \to 0$, nh_n, $nb_n \to \infty$ and

$$nh_n^4 \to 0, \quad nb_n^4 \to 0, \quad \frac{b_n^2 \log(h_n^{-1})}{h_n} \to 0, \quad \frac{\log(h_n^{-1})}{nh_n^3 b_n^{4\delta}} \to 0, \quad \frac{\log(h_n^{-1})}{nh_n b_n^2} \to 0$$

for some $\delta \in (0, \frac{1}{2})$ and for $n \to \infty$.

(A4) The errors $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed, independent from the covariates, with strictly increasing distribution function F_ε and bounded density f_ε, which has one bounded continuous derivative. The errors are centered, i.e. $E[\varepsilon_i] = 0$ with variance $\sigma^2 > 0$ and existing fourth moment.

(A5) The errors have median zero, i.e. $F_\varepsilon(0) = \frac{1}{2}$.

(A6) The errors $\varepsilon_1, \ldots, \varepsilon_n$ are independent and normally distributed with expectation zero and variance $\sigma^2 > 0$, independent from the covariates.

(A7) The error density f_ε is unimodal.

Conditions (A1)–(A4) are assumed to be valid throughout the paper, whereas it is stated explicitly when (A5), (A6) or (A7) are assumed.

We restrict to the homoscedastic case with random covariates for the moment, but other cases will be discussed in Remarks 4.4 and 4.5.

3 Test statistics

In general the estimator $\hat{m}_I$ estimates the increasing rearrangement m_I of m. Only under the hypothesis H_0 of an increasing regression function we have $m = m_I$. We build (pseudo-) residuals

$$\hat{\varepsilon}_{i,I} = Y_i - \hat{m}_I(X_i),$$

which estimate pseudo-errors $\varepsilon_{i,I} = Y_i - m_I(X_i)$ that coincide with the true errors $\varepsilon_i = Y_i - m(X_i)$ $(i = 1, \ldots, n)$ in general only under H_0. Let further

$$\hat{\varepsilon}_i = Y_i - \hat{m}(X_i)$$

denote the unconstrained residuals. Under H_0 both $\hat{m}$ and $\hat{m}_I$ join the same first order asymptotic expansion. This for estimation purposes very desirable property limits the possibilities to apply the estimator $\hat{m}_I$ for hypotheses testing. Test statistics based on estimated empirical processes such as

$$(3.1) \qquad \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\varepsilon}_{i,I} I\{X_i \leq \cdot\} \ \text{ or } \ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(I\{\hat{\varepsilon}_{i,I} \leq \cdot\} - I\{\hat{\varepsilon}_i \leq \cdot\} \right)$$

[compare to Stute (1997) and Van Keilegom, González-Manteiga and Sánchez Sellero (2008)] are of convergence order $o_P(1)$ and not suitable for the testing problem considered here. For the estimated L^2-distance

$$(3.2) \qquad n h_n^{1/2} \int (\hat{m}_I - \hat{m})^2$$

[cf. Härdle and Mammen (1993)] the same problem arises. One could try to rescale the test statistics and apply second order asymptotic expansions to derive an nondegenerate limit distribution. Whereas this seems not possible for the second empirical process in (3.1) with methods of proofs typically applied for such processes, it might work for the first process in (3.1) as well as for the L^2-distance test (3.2). However, the resulting tests typically react rather sensitive to the choice of smoothing parameters. Birke and Dette (2007) follow this approach by considering a suitably scaled version of the test statistic $\int (\hat{m}_I^{-1}(\hat{m}(x)) - x)^2 \, dx$ and by applying second order asymptotic expansions.

The idea we follow in the present paper instead is the following. Whereas the true errors $\varepsilon_1, \ldots, \varepsilon_n$ are assumed to be independent and identically distributed, the pseudo-errors $\varepsilon_{1,I}, \ldots, \varepsilon_{n,I}$ behave differently. If the true function m is not monotone (e.g. like in Figure 1) and we calculate the pseudo-residuals from m_I too many of them are positive (see solid dots with black lines) on some subinterval of $[0, 1]$ and too many are negative (see open dots with grey lines) on another subinterval. Therefore, they are no longer identically distributed if H_0 is not fulfilled.
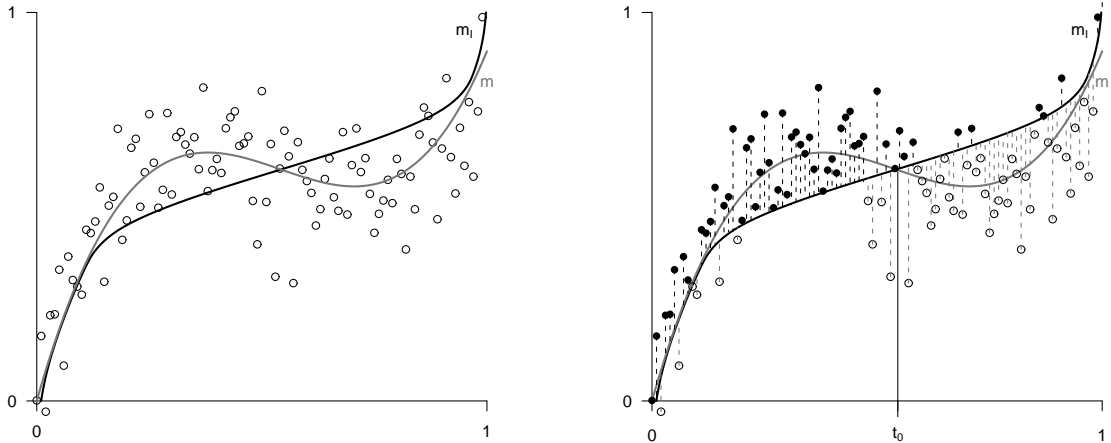
Figure 1: Left part: True function m (grey line), monotonized function m_I (black line) together with observations. Right part: Pseudo-residuals (positive ones solid with black dashed lines, negative ones open with grey dashed lines)

We construct test statistics from several estimated empirical processes to detect this different behaviour. The first process we will consider is defined as

$$S_n(t) \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(I\{\hat{\varepsilon}_{i,I} > 0\} I\{X_i \leq t\} - \frac{1}{2}\hat{F}_{X,n}(t) \right)$$

where $t \in [0,1]$ and $\hat{F}_{X,n}$ denotes the empirical distribution function of the covariates $X_1, \ldots, X_n$. For every $t \in [0,1]$ it counts how many pseudo-residuals are positive up to covariates $\leq t$. This term is then centered with respect to the estimated expectation under H_0 and scaled with $n^{-1/2}$. Under assumptions (A1)–(A5), $S_n(t)$ consistently estimates the expectation

$$\sqrt{n}\left(E[I\{\varepsilon_{i,I} > 0\} I\{X_i \leq t\}] - \frac{1}{2}F_X(t) \right)$$

(3.3)
$$= \sqrt{n}\left(E[I\{\varepsilon_i > (m_I - m)(X_i)\} I\{X_i \leq t\}] - (1 - F_\varepsilon(0))F_X(t) \right)$$

$$= \sqrt{n} \int_0^t \left(F_\varepsilon(0) - F_\varepsilon((m_I - m)(x)) \right) f_X(x)\, dx.$$

This term is zero for all $t \in [0,1]$ if and only if $m_I = m$ is valid F_X-a.s. To obtain this equivalence one especially uses that F_ε is strictly increasing, whereas equality (3.3) applies assumption (A5). As we have seen, for instance a Kolmogorov-Smirnov type statistic $s_n = \sup_{t \in [0,1]} |S_n(t)|$ estimates a distance measure between m_I and m and, to obtain a consistent testing procedure, the null hypothesis should be rejected for large values of s_n.

6

To avoid assumption (A5) one can alternatively consider the process

$$\tilde{S}_n(t) = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} I\{\hat{\varepsilon}_{i,I} > 0\}I\{X_i \le t\} - (1 - \hat{F}_{\varepsilon,n}(0))\hat{F}_{X,n}(t)\Big),$$

where $\hat{F}_{\varepsilon,n}$ denotes the empirical distribution function of $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$.

The application of tests based on S_n and $\tilde{S}_n$ may not lead to good power in cases where m and m_I are quite similar and the variance is large. Hence it seems sensible to not only take into account the sign of the estimated pseudo-errors, but also their value, i.e. to consider tests based on

$$\frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_{i,I}I\{\hat{\varepsilon}_{i,I} > 0\}I\{X_i \le t\}.$$

The estimated expectation of this term under H_0, i.e. $E[\varepsilon_i I\{\varepsilon_i > 0\}I\{X_i \le t\}]$ is known to be $\frac{\sigma}{\sqrt{2\pi}}F_X(t)$ under assumption (A6), and then can be estimated by $\frac{\hat{\sigma}}{\sqrt{2\pi}}\hat{F}_{X,n}(t)$, where $\hat{\sigma} = (n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_i^2)^{1/2}$. This leads to the process

$$V_n(t) = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_{i,I}I\{\hat{\varepsilon}_{i,I} > 0\}I\{X_i \le t\} - \frac{\hat{\sigma}}{\sqrt{2\pi}}\hat{F}_{X,n}(t)\Big).$$

To avoid assumption (A6) one can alternatively consider the process

$$\tilde{V}_n(t) = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_{i,I}I\{\hat{\varepsilon}_{i,I} > 0\}I\{X_i \le t\} - \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i I\{\hat{\varepsilon}_i > 0\}\hat{F}_{X,n}(t)\Big).$$

Lemma 3.1 *Tests that for some constant $c > 0$ reject H_0 whenever $\sup_{t\in[0,1]}|V_n(t)| > c$ or $\sup_{t\in[0,1]}|\tilde{V}_n(t)| > c$ are consistent under assumptions (A1)–(A3),(A6) and (A1)–(A4), respectively.*

Because the proof of this statement requires a longer argumentation we defer it to the appendix.

Now let $\hat{R}_{i,I}$ denote the fractional rank of $\hat{\varepsilon}_{i,I}$ with respect to $\hat{\varepsilon}_{1,I}, \ldots, \hat{\varepsilon}_{n,I}$, i.e.

$$\hat{R}_{i,I} = \frac{1}{n}\sum_{j=1}^{n} I\{\hat{\varepsilon}_{j,I} \le \hat{\varepsilon}_{i,I}\}$$

and consider the term

$$\frac{1}{n}\sum_{i=1}^{n} \hat{R}_{i,I}I\{\hat{\varepsilon}_{i,I} > 0\}I\{X_i \le t\},$$

which (under H_0) estimates the expectation

$$E\Big[\frac{1}{n}\sum_{j=1}^{n} I\{\varepsilon_j \le \varepsilon_i\}I\{\varepsilon_i > 0\}I\{X_i \le t\}\Big] = E[F_\varepsilon(\varepsilon_i)I\{\varepsilon_i > 0\}]F_X(t) + o(1)$$

$$= \int_{F_\varepsilon(0)}^{1} x\,dx\, F_X(t) + o(1) = \Big(\frac{1}{2} - \frac{(F_\varepsilon(0))^2}{2}\Big)F_X(t) + o(1) = \frac{3}{8}F_X(t) + o(1),$$

where the last equality holds under assumption (A5). This expectation can be estimated by $\frac{3}{8}\hat{F}_{X,n}(t)$ under (A5) and by $\frac{1}{2}(1-(\hat{F}_{\varepsilon,n}(0))^2)\hat{F}_{X,n}(t)$ otherwise, which leads to the empirical processes

$$R_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\hat{R}_{i,I}I\{\hat{\varepsilon}_{i,I}>0\}I\{X_i \le t\} - \frac{3}{8}\hat{F}_{X,n}(t)\right)$$

$$\tilde{R}_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\hat{R}_{i,I}I\{\hat{\varepsilon}_{i,I}>0\}I\{X_i \le t\} - \frac{1}{2}(1-(\hat{F}_{\varepsilon,n}(0))^2)\hat{F}_{X,n}(t)\right).$$

Lemma 3.2 *Tests that for some constant $c>0$ reject H_0 whenever $\sup_{t\in[0,1]}|R_n(t)|>c$ or $\sup_{t\in[0,1]}|\tilde{R}_n(t)|>c$ are consistent under assumptions (A1)–(A5),(A7) and (A1)–(A4),(A7), respectively.*

Again, the proof of this result needs a longer argumentation and is defered to the appendix.

4   Main asymptotic results

In the following theorem we state weak convergence results for the processes defined before. Note that we have to assume a strictly increasing regression function to derive the asymptotic distributions. Nevertheless, the monotone regression estimator $\hat{m}_I$ can also be applied for monotone regression functions with flat parts, see Dette and Pilz (2006).

Theorem 4.1 *Assume that m' is positive in $[0,1]$.*
(i) *Under assumptions (A1)–(A5) the process S_n converges weakly in $\ell^\infty([0,1])$ to a Gaussian process S with covariance*

$$Cov(S(s),S(t)) = F_X(s \wedge t)\left(\frac{1}{4}+\sigma^2 f_\varepsilon^2(0)+2f_\varepsilon(0)E[\varepsilon_1 I\{\varepsilon_1 \le 0\}]\right)$$
$$= F_X(s \wedge t)(\frac{1}{4}-\frac{1}{2\pi}),$$

where the last equality holds under the additional assumption (A6).
(ii) *Under assumptions (A1)–(A4) the process $\tilde{S}_n$ converges weakly in $\ell^\infty([0,1])$ to a Gaussian process $\tilde{S}$ with covariance*

$$Cov(\tilde{S}(s),\tilde{S}(t))$$
$$= (F_X(s \wedge t) - F_X(s)F_X(t))\left(F_\varepsilon(0)(1-F_\varepsilon(0))+\sigma^2 f_\varepsilon^2(0)+2f_\varepsilon(0)E[\varepsilon_1 I\{\varepsilon_1 \le 0\}]\right)$$
$$= (F_X(s \wedge t) - F_X(s)F_X(t))(\frac{1}{4}-\frac{1}{2\pi}),$$

where the last equality holds under the additional assumption (A6).
(iii) *Under assumptions (A1)–(A3) and (A6) the process V_n converges weakly in $\ell^\infty([0,1])$*

to a Gaussian process V with covariance

$$Cov(V(s), V(t)) = F_X(s \wedge t)(\frac{1}{4} - \frac{1}{2\pi})\sigma^2 - F_X(s)F_X(t)\frac{\sigma^2}{4\pi}.$$

(iv) *Under assumptions (A1)–(A4) the process $\tilde{V}_n$ converges weakly in $\ell^\infty([0,1])$ to a Gaussian process $\tilde{V}$ with covariance*

$$Cov(\tilde{V}(s), \tilde{V}(t))$$
$$= (F_X(s \wedge t) - F_X(s)F_X(t))\Big((2F_\varepsilon(0) - 1)E[\varepsilon_1^2 I\{\varepsilon_1 \le 0\}] - (E[\varepsilon_1 I\{\varepsilon_1 \le 0\}])^2 + \sigma^2(1 - F_\varepsilon(0))^2\Big)$$
$$= (F_X(s \wedge t) - F_X(s)F_X(t))(\frac{1}{4} - \frac{1}{2\pi})\sigma^2,$$

where the last equality holds under the additional assumption (A6).
(v) *Under assumptions (A1)–(A5) the process R_n converges weakly in $\ell^\infty([0,1])$ to a Gaussian process R with covariance*

$$Cov(R(s), R(t))$$
$$= (F_X(s \wedge t) - F_X(s)F_X(t))\Big(\frac{29}{192} + \sigma^2(f_\varepsilon(0)F_\varepsilon(0) - E[f_\varepsilon(\varepsilon_1)I\{\varepsilon_1 > 0\}])^2$$
$$+ 2E[F_\varepsilon(\varepsilon_1)\varepsilon_1 I\{\varepsilon_1 \le 0\}](f_\varepsilon(0)F_\varepsilon(0) - E[f_\varepsilon(\varepsilon_1)I\{\varepsilon_1 > 0\}])\Big) + \frac{1}{16}F_X(s)F_X(t).$$

(vi) *Under assumptions (A1)–(A4) the process $\tilde{R}_n$ converges weakly in $\ell^\infty([0,1])$ to a Gaussian process $\tilde{R}$ with covariance*

$$Cov(\tilde{R}(s), \tilde{R}(t))$$
$$= (F_X(s \wedge t) - F_X(s)F_X(t))\Big(E[F_\varepsilon^2(\varepsilon_1)I\{\varepsilon_1 > 0\}] - (E[F_\varepsilon(\varepsilon_1)I\{\varepsilon_1 > 0\}])^2$$
$$+ \sigma^2(f_\varepsilon(0)F_\varepsilon(0) + E[f_\varepsilon(\varepsilon_1)I\{\varepsilon_1 > 0\}])^2$$
$$- 2E[F_\varepsilon(\varepsilon_1)\varepsilon_1 I\{\varepsilon_1 \le 0\}](f_\varepsilon(0)F_\varepsilon(0) + E[f_\varepsilon(\varepsilon_1)I\{\varepsilon_1 > 0\}])\Big).$$

The proof is given in the appendix.

Remark 4.2 For a normal regression model, i.e. under assumption (A6) we can obtain asymptotically distribution free tests, because then

$$\sup_{t \in [0,1]} |S_n(t)| = \sup_{s \in (0,1)} |S_n(F_X^{-1}(s))|$$

converges in distribution to $(\frac{1}{4} - \frac{1}{2\pi}) \sup_{s \in [0,1]} |W(s)|$ for a Brownian motion W. Similarly, $\sup_{t \in [0,1]} |\tilde{S}_n(t)|$ converges in distribution to $(\frac{1}{4} - \frac{1}{2\pi}) \sup_{s \in [0,1]} |B(s)|$, where B is a Brownian bridge.

9

Remark 4.3 The proposed tests can detect local alternatives of the form

$$H_{1,n} : m(x) = m_I(x) + \frac{\Delta(x)}{\sqrt{n}},$$

where $\Delta \neq 0$ on an interval in $[0,1]$ of positive length. Consider S_n for simplicity. From (3.3) we see that the asymptotic expectation of $S_n(t)$ under $H_{1,n}$ is

$$\sqrt{n} \int_0^t \Big(F_\varepsilon(0) - F_\varepsilon((m_I - m)(x)) \Big) f_X(x)\, dx = f_\varepsilon(0) \int_0^t \Delta(x) f_X(x)\, dx + o(1).$$

With similar arguments as in the proof of Theorem 4.1 one can show that under $H_{1,n}$, S_n converges in distribution to the process $f_\varepsilon(0) \int_0^t \Delta(x) f_X(x)\, dx + S(t)$, $t \in [0,1]$. The Kolmogorov-Smirnov test statistic $\sup_{t \in [0,1]} |S_n(t)|$ constructed from Theorem 4.1 detects $H_{1,n}$ because $f_\varepsilon(0) \sup_{t \in [0,1]} | \int_0^t \Delta(x) f_X(x)\, dx| > 0$.

Remark 4.4 Assume a heteroscedastic regression model

$$Y_i = m(X_i) + \sigma(X_i)\varepsilon_i, \qquad i = 1, \ldots, n,$$

where X_i and ε_i are independent, $E[\varepsilon_i^2] = 1$, $E[\varepsilon_i^4] < \infty$, the regression function m, error distribution F_ε and covariate distribution F_X fulfill assumptions as before, whereas the variance function σ^2 is twice continuously differentiable and bounded away from zero. Then similar tests for monotonicity of the regression function m can be constructed by replacing residuals $\hat{\varepsilon}_i$ and pseudo-residuals $\hat{\varepsilon}_{i,I}$ from before by

$$\hat{\varepsilon}_i = \frac{Y_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)}, \quad \hat{\varepsilon}_{i,I} = \frac{Y_i - \hat{m}_I(X_i)}{\hat{\sigma}(X_i)},$$

where $\hat{\sigma}^2$ denotes a Nadaraya-Watson estimator for σ^2 with kernel K and bandwidth h_n based on "observations" $(Y_i - \hat{m}(X_i))^2$. With these changes the same processes as before can be considered for testing H_0. Weak convergence to Gaussian processes can be obtained with methods as in Akritas and Van Keilegom (2001), where the asymptotic covariances change in comparison to Theorem 4.1 due to the estimation of the variance function.

Remark 4.5 Assume a (homoscedastic) fixed design regression model

$$Y_i = m(x_{ni}) + \varepsilon_i, \qquad i = 1, \ldots, n,$$

with assumptions as before but with nonrandom covariates $x_{n1} \leq \ldots \leq x_{nn}$ such that there exists a distribution function F_X with support $[0,1]$ so that $F_X(x_{ni}) = \frac{i}{n}$, $i = 1, \ldots, n$, and F_X fulfills assumptions as before. Then similar tests for monotonicity of m can be derived by considering sequential empirical processes. For instance, instead of $\tilde{S}_n$ we would consider

$$\bar{S}_n(t) = \sqrt{n}\Big(\frac{1}{n} \sum_{i=1}^{\lfloor nt \rfloor} I\{\hat{\varepsilon}_{i,I} > 0\} - (1 - \hat{F}_{\varepsilon,n}(0))t \Big),$$

10

where $\lfloor nt \rfloor$ is the largest integer $\leq nt$. Weak convergence of the processes similar to the results in Theorem 4.1 can be obtained with methods as in Neumeyer and Van Keilegom (2009).

5   Bootstrap method and simulation results

Since the asymptotic distributions of the test statistics still depend on the unknown functions m and f we use the bootstrap procedures to construct tests based on the above statistics. We build bootstrap observations that fulfill the null hypothesis by defining

$$Y_i^* = \hat{m}_I(X_i) + \varepsilon_i^*, \quad i = 1, \ldots, n.$$

Here, under assumption (A6) we can generate the bootstrap errors $\varepsilon_1^*, \ldots, \varepsilon_n^*$ by the normal distribution $N(0, \hat{\sigma}^2)$, where $\hat{\sigma}^2$ is the estimated variance from residuals $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$.

Without assumption (A6) instead we apply a nonparametric smoothed residual bootstrap. To this end, we randomly draw $\tilde{\varepsilon}_i^*$ with replacement from centered residuals $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$, where $\tilde{\varepsilon}_j = \hat{\varepsilon}_j - n^{-1} \sum_{k=1}^n \hat{\varepsilon}_k$. Let further a denote a small smoothing parameter and let $Z_1, \ldots, Z_n$ be independent and standard normally distributed. Then, $\varepsilon_i^* = \tilde{\varepsilon}_i^* + aZ_i$, $i = 1, \ldots, n$, are independent, given the original sample $\mathcal{Y}_n = \{(X_i, Y_i) \mid i = 1, \ldots, n\}$ and have a distribution function $\tilde{F}_{n,\varepsilon}$ with density

$$\tilde{f}_{n,\varepsilon}(y) = \frac{1}{na} \sum_{i=1}^n \varphi\Big(\frac{\tilde{\varepsilon}_i - y}{a}\Big).$$

From the bootstrap observations calculate the constrained and unconstrained regression estimators $\hat{m}_I^*$ and $\hat{m}^*$ and build residuals $\hat{\varepsilon}_{i,I}^* = Y_i^* - \hat{m}_I^*(X_i)$ and $\hat{\varepsilon}_i^* = Y_i^* - \hat{m}(X_i)$, respectively. Let $\hat{F}_{\varepsilon,n}^*$ denote the empirical distribution function of $\hat{\varepsilon}_1^*, \ldots, \hat{\varepsilon}_n^*$ and $\hat{R}_{i,I}^*$ the fractional rank of $\hat{\varepsilon}_{i,I}^*$ with respect to $\hat{\varepsilon}_{1,I}^*, \ldots, \hat{\varepsilon}_{n,I}^*$. The bootstrap versions of the considered processes are defined as follows,

$$S_n^*(t) = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^n I\{\hat{\varepsilon}_{i,I}^* > 0\}I\{X_i \leq t\} - (1 - \tilde{F}_{\varepsilon,n}(0))\hat{F}_{X,n}(t)\Big)$$

$$\tilde{S}_n^*(t) = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^n I\{\hat{\varepsilon}_{i,I}^* > 0\}I\{X_i \leq t\} - (1 - \hat{F}_{\varepsilon,n}^*(0))\hat{F}_{X,n}(t)\Big)$$

$$V_n^*(t) = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_{i,I}^* I\{\hat{\varepsilon}_{i,I}^* > 0\}I\{X_i \leq t\} - \frac{\hat{\sigma}^*}{\sqrt{2\pi}}\hat{F}_{X,n}(t)\Big)$$

$$\tilde{V}_n^*(t) = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_{i,I}^* I\{\hat{\varepsilon}_{i,I}^* > 0\}I\{X_i \leq t\} - \frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_i^* I\{\hat{\varepsilon}_i^* > 0\}\hat{F}_{X,n}(t)\Big)$$

$$R_n^*(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(\hat{R}_{i,I}^* I\{\hat{\varepsilon}_{i,I}^* > 0\}I\{X_i \leq t\} - \frac{1}{2}(1 - (\tilde{F}_{\varepsilon,n}(0))^2)\hat{F}_{X,n}(t)\Big)$$

11

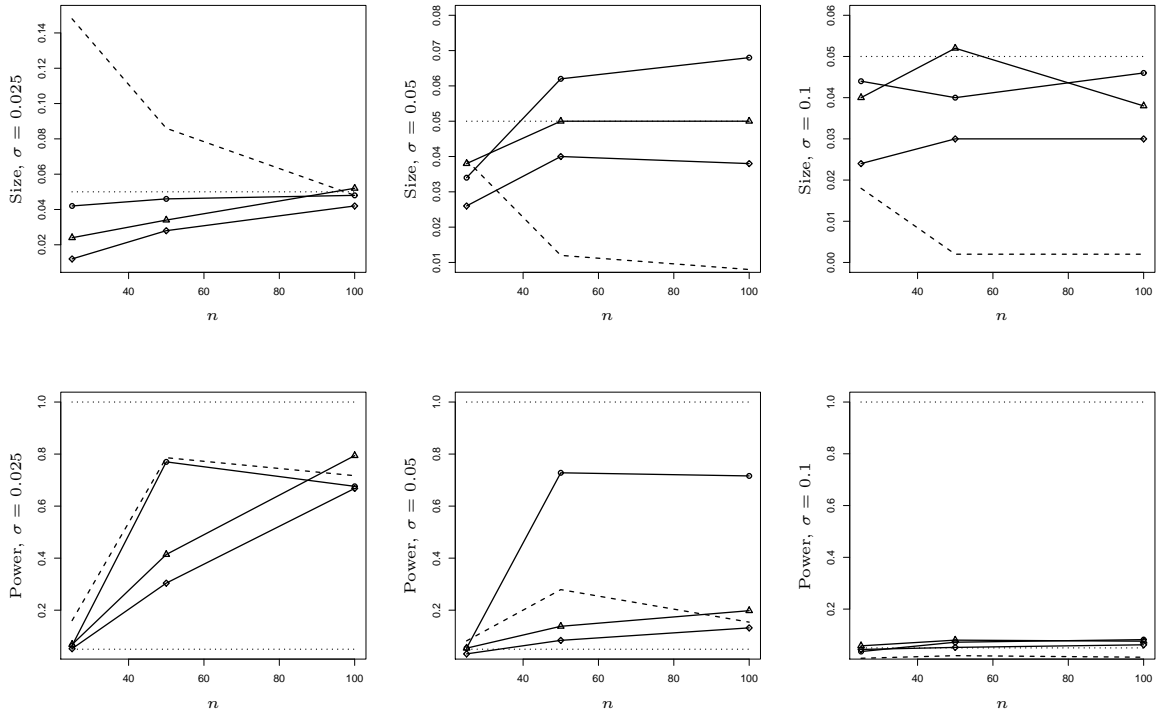Figure 2: Simulated size and power in dependence of n for the tests φ_{s_n} (diamonds), φ_{v_n} (dots) and φ_{r_n} (triangles) compared to the test φ_{L^2} (dashed line) for different standard deviations σ (left $\sigma = 0.025$, middle $\sigma = 0.05$, right $\sigma = 0.1$). First row m_1, second row $m_{2,n}$.

$$\tilde{R}_n^*(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\hat{R}_{i,I}^* I\{\hat{\varepsilon}_{i,I}^* > 0\} I\{X_i \leq t\} - \frac{1}{2}(1 - (\hat{F}_{\varepsilon,n}^*(0))^2)\hat{F}_{X,n}(t) \right),$$

where only in the case of V_n^* under assumption (A6) we use the parametric bootstrap applying the normal distribution as explained above, where then $\hat{\sigma}^*$ is the empirical standard deviation of $\hat{\varepsilon}_1^*, \ldots, \hat{\varepsilon}_n^*$. Note that the bootstrap processes are centered in a slightly different way than the original statistics with the aim to obtain processes that are asymptotically centered with respect to the conditional expectation $E[\cdot \mid \mathcal{Y}_n]$. In the appendix we sketch a proof for validity of the bootstrap procedures.

Since it turned out in a simulation study in Birke and Dette (2007), that their test and the test developed by Bowman, Jones and Gijbels (1998) behave very similar we will compare the tests described here only to the one by Birke and Dette (2007). More precisely we use the Kolmogorov-type statistics $s_n = \sup|S_n(t)|$, $v_n = \sup|V_n(t)|$, $r_n = \sup|R_n(t)|$, $\tilde{s}_n = \sup|\tilde{S}_n(t)|$, $\tilde{v}_n = \sup|\tilde{V}_n(t)|$ and $\tilde{r}_n = \sup|\tilde{R}_n(t)|$ and denote the corresponding tests by φ_{s_n}, φ_{v_n}, φ_{r_n}, $\tilde{\varphi}_{s_n}$, $\tilde{\varphi}_{v_n}$ and $\tilde{\varphi}_{r_n}$. We show the behavior of all tests under the null hypothesis
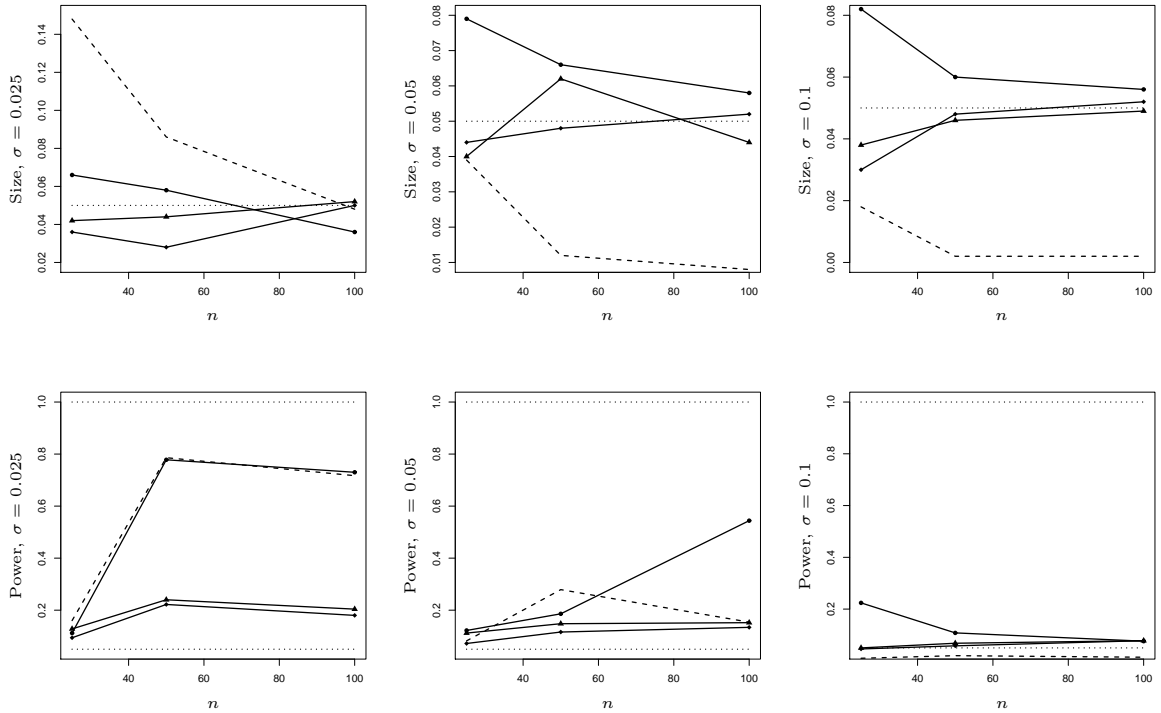
12

Figure 3: Simulated size and power in dependence of n for the tests $\tilde{\varphi}_{s_n}$ (diamonds), $\tilde{\varphi}_{v_n}$ (dots) and $\tilde{\varphi}_{r_n}$ (triangles) compared to the test φ_{L^2} (dashed line) for different standard deviations σ (left $\sigma = 0.025$, middle $\sigma = 0.05$, right $\sigma = 0.1$). First row m_1, second row $m_{2,n}$.

as well as under local alternatives and compare it to the behavior of the L^2-test φ_{L^2}. To this end we simulate from the regression model

$$Y_i = m(X_i) + \sigma \varepsilon_i$$

with different regression functions

$$
\begin{aligned}
m_1(x) &= x, \ x \in [0,1] \\
m_{2,n}(x) &= x + \frac{1}{2\sqrt{n}} \sin(10\pi x), \ x \in [0,1]
\end{aligned}
$$

and standard normal errors for the sample sizes $n = 25, 50$ and 100 and standard deviations $\sigma = 0.025, 0.05$ and 0.1. Those errors fulfill all conditions (A4)-(A7) from section 2 and should give acceptable results for all test statistics. We perform 500 simulation runs with each 200 bootstrap repetitions to estimate the size and power of the tests. Note that m_1 corresponds to the null hypothesis while $m_{2,n}$ corresponds to a local alternative as described in Remark 4.3 which converges for increasing sample size to the null hypothesis with rate
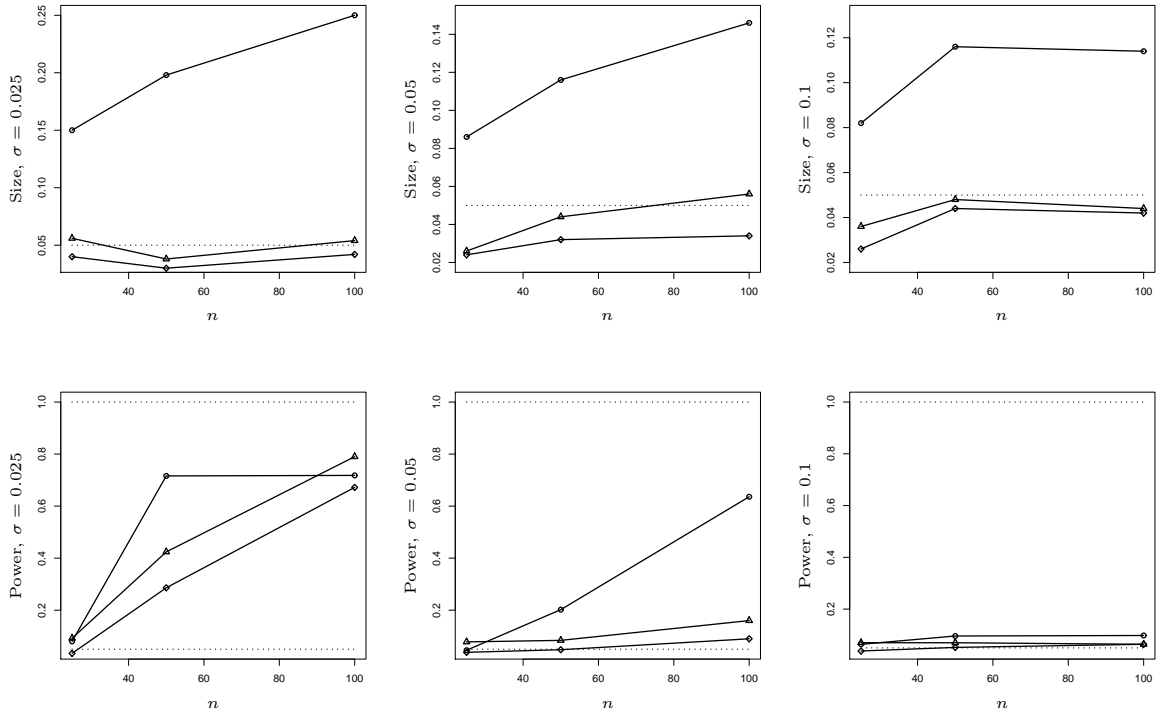
13

Figure 4: Simulated size and power in dependence of n for the tests φ_{s_n} (diamonds), φ_{v_n} (dots) and φ_{r_n} (triangles) compared to the test φ_{L^2} (dashed line) for different standard deviations σ (left $\sigma = 0.025$, middle $\sigma = 0.05$, right $\sigma = 0.1$) and t-distributed errors. First row m_1, second row $m_{2,n}$.

$1/\sqrt{n}$ and should therefore be harder to detect by an L^2-test than by the empirical process approach discussed here. The bandwith for the unconstrained estimator is chosen by cross validation while the bandwith for monotonizing is chosen as $b_n = h_n^{1.2}$. For generating the bootstrap data we use a slightly larger bandwidth $h_{n,b} = h_n^{0.5}$ to guarantee the consistency (see also Härdle, 1990 for that). For the smoothed residual bootstrap we use for the test statistics S_n, R_n, $\tilde{S}_n$, $\tilde{V}_n$ and $\tilde{R}_n$ we need an additional smoothing parameter a which we choose as $a = 0.2\hat{\sigma}n^{-0.15}$.

Figure 2 shows the simulated size (first row) for m_1 and the simulated power (second row) for $m_{2,n}$ of the tests φ_{s_n}, φ_{v_n} and φ_{r_n}. We compare this to the results for the L^2-test proposed by Birke and Dette (2007) (dashed line). The behavior of the tests heavily depends on the standard deviation σ. For all standard deviations, the tests φ_{s_n}, φ_{v_n} and φ_{r_n} are less conservative than the L^2-test. Let us now consider the behavior under the alternative. For a small standard deviation ($\sigma = 0.025$) all tests behave very similar with some advantages for the L^2-test for a sample size of $n = 25$ and advantages for both the L^2-test and the
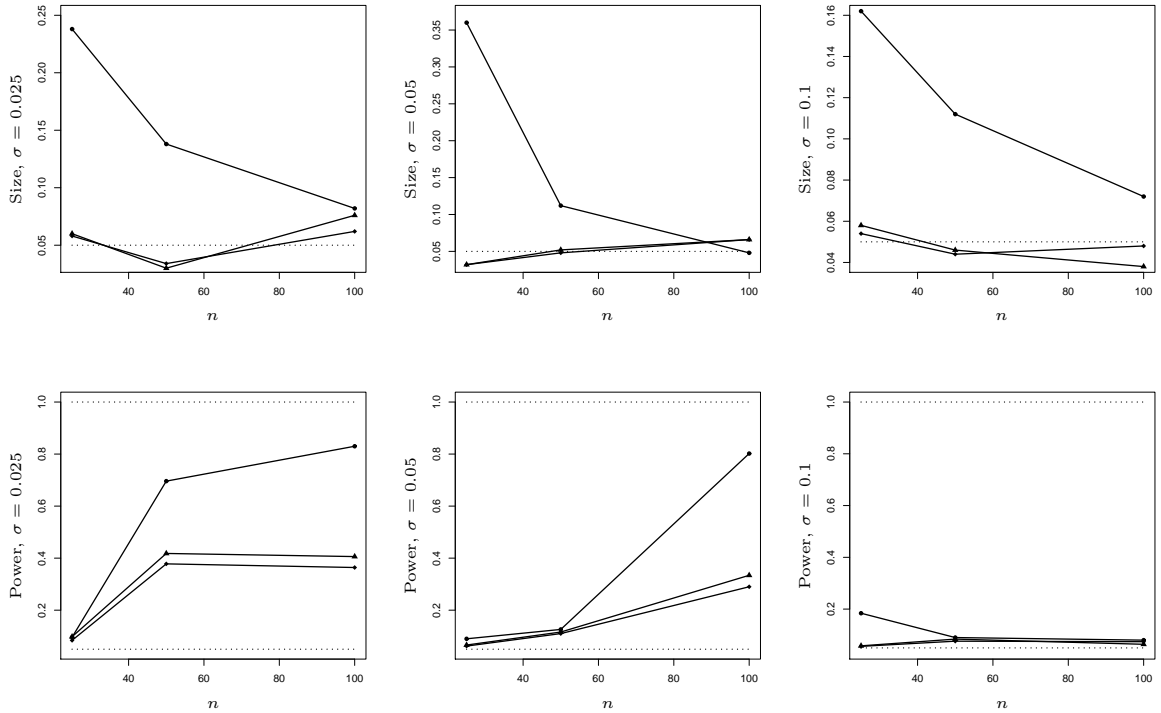
Figure 5: Simulated size and power in dependence of n for the tests $\tilde{\varphi}_{s_n}$ (diamonds), $\tilde{\varphi}_{v_n}$ (dots) and $\tilde{\varphi}_{r_n}$ (triangles) compared to the test φ_{L^2} (dashed line) for different standard deviations σ (left $\sigma = 0.025$, middle $\sigma = 0.05$, right $\sigma = 0.1$) and t-distributed errors. First row m_1, second row $m_{2,n}$.

test based on V_n for the sample size $n = 50$. But for $n = 100$ the power of all tests is comparable and satisfactorily high for the local alternative. Now, for moderate standard deviation ($\sigma = 0.05$) we see clear advantages in the power of the test φ_{v_n} while the results for the tests based on φ_{s_n} and φ_{r_n} are still lower than that of the L^2-test. For a comparable high standard deviation the tests again behave very similar with small advantages for the tests proposed here which have for $n = 100$ still a power larger than the size of $\alpha = 0.05$. This is not the case for the L^2-test. To conclude the above discussion we note that our assumption in section 3, that φ_{v_n} exhibits the best power, is confirmed for this simulation example.

We already mentioned in section 3, that φ_{s_n}, φ_{v_n} and φ_{r_n} need the restrictive assumption (A5). Furthermore an asymptotic consideration of V_n additionally needs the assumption (A6) of normal errors and we therefore used the parametric bootstrap in this case. It would now be interesting to see how the more general test statistics behave for the same simulation setting. The results are shown in Figure 3. As before we see that the tests $\tilde{\varphi}_{s_n}$, $\tilde{\varphi}_{v_n}$ and $\tilde{\varphi}_{r_n}$

15

approximate the size better than the L^2-test and are therefore less conservative. The behavoir under local alternatives is similar to that of the tests φ_{s_n}, φ_{v_n} and φ_{r_n}. Again the test $\tilde{\varphi}_{v_n}$ has the best power under the tests $\tilde{\varphi}_{s_n}$, $\tilde{\varphi}_{v_n}$ and $\tilde{\varphi}_{r_n}$. But the different standardisation without using assumption (A5) seems to result in a slightly lower power.

To show the limits concerning the different types of error distributions of the different test statistics, we simulate from the same regression models but now with the following two error distributions

(i) The errors are generated as $\varepsilon_i = \sqrt{6/8}\sigma t_i$, $i = 1, \ldots, n$ where t_i, $i = 1, \ldots, n$ are drawn independently from a t-distribution with 8 degrees of freedom.

(ii) The errors are generated as $\varepsilon_i = \sigma(e_i - 1)$, $i = 1, \ldots, n$ where e_i, $i = 1, \ldots, n$ are drawn independently from an exponential distribution with parameter 1.

(i) Note that in this case, the errors fulfill assumptions (A4), (A5) and (A7) but not (A6) and the expectation is that this choice results in a failure of the test based on V_n since we need the assumption of normality for deriving the asymptotic distribution while all other test statistics should perform right. Figure 4 shows the results for the tests φ_{s_n}, φ_{v_n} and φ_{r_n}. As expected we observe, that for m_1 and the tests based on φ_{s_n} and φ_{r_n} the size is approximated very well while for the test based on φ_{v_n}, the size is much to large and gets even larger for increasing sample size. Concerning the power of the tests there are no large differences to the case of normal errors. We constructed the further test statistic $\tilde{V}_n$ to avoid assumption (A6) and therefore the test based on $\tilde{V}_n$ (and, of course, also the tests based on $\tilde{S}_n$ and $\tilde{R}_n$) should perform better for those errors. The results are shown in Figure 5. We observe, that the test $\tilde{\varphi}_{v_n}$ still has problems to approximate the size for small sample sizes but tends to the right size for larger sample sizes and has the best power of the three different tests. The tests based on $\tilde{\varphi}_{s_n}$ and $\tilde{\varphi}_{r_n}$ perform very well with the typical effect that the power gets lower the larger the standard deviation is.

(ii) The centered exponential errors fulfill assumptions (A4) and (A7) but not (A5) and (A6). Therefore we would expect from the theoretical results that the tests φ_{s_n}, φ_{v_n} and φ_{r_n} can no longer be used while the tests based on $\tilde{\varphi}_{s_n}$, $\tilde{\varphi}_{v_n}$ and $\tilde{\varphi}_{r_n}$ still behave well. We see the results in Figure 6 where we show the estimated size for S_n, V_n and R_n in the first row which is much too large and increasing for all tests. In the second and third row we show the estimated size respectively power of the tests $\tilde{\varphi}_{s_n}$, $\tilde{\varphi}_{v_n}$ and $\tilde{\varphi}_{r_n}$. Again, the approximated size of the test $\tilde{\varphi}_{v_n}$ is too large for small sample sizes but seems to approximate it better for larger sample sizes while the other two tests approximate the size very well. All tests perform satisfactorily concerning the power. Here again $\tilde{\varphi}_{v_n}$ provides the best power.
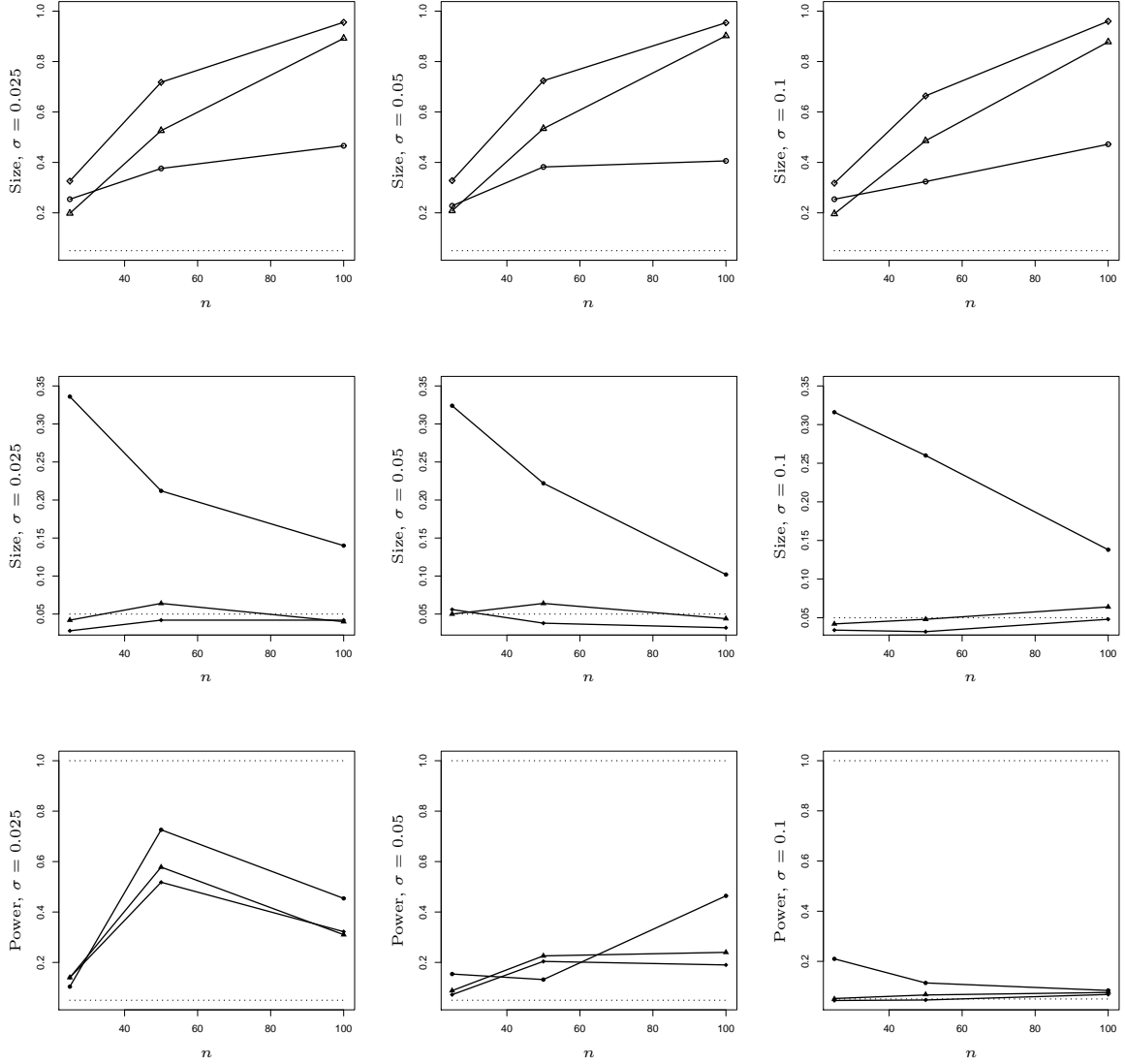
16

Figure 6: Behavior of the tests for different standard deviations σ (left $\sigma = 0.025$, middle $\sigma = 0.05$, right $\sigma = 0.1$) and exponenial errors. First row: Behavior of the tests φ_{s_n}, φ_{v_n} and φ_{r_n} for m_1; second and third row: Behavior of the tests $\tilde{\varphi}_{s_n}$, $\tilde{\varphi}_{v_n}$ and $\tilde{\varphi}_{r_n}$ for m_1 respectively $m_{2,n}$.

17

6 Conclusion

In this paper we have considered the problem of testing for monotonicity of regression functions. We have demonstrated that typical distance based tests, which are popular in goodness-of-fit testing in regression models, are not applicable here. As alternative we suggested several non-standard, but intuitive distance based tests constructed from Kolmogorov-Smirnov type statistics of empirical processes of residuals estimated both under the null hypothesis of monotonicity and under the general nonparametric model. We presented the asymptotic distributions as well as the small sample performance of bootstrap versions of the tests. We discussed differences in the behaviours of the various tests and compared with the results of the L^2-test proposed in Birke and Dette (2007). We have seen, that all empirical process approaches lead to less conservative testing procedures than does the L^2-test while having a comparable power. It turned out, that the power is even better for local alternatives of order $1/\sqrt{n}$ and relatively large standard deviations. But we also observed, that some of the tests fail for non-normal or non-symmetric error distributions.

It is a topic of current research to investigate whether similar tests can be applied to test for monotonicity of quantile regression curves.

A Proofs

A.1 Consistency proofs

Proof of Lemma 3.1. To check whether Kolmogorov-Smirnov statistics $\sup_{t \in [0,1]} |V_n(t)|$ or $\sup_{t \in [0,1]} |\tilde{V}_n(t)|$ can lead to a consistent testing procedure we consider the estimated expectation of the processes and show, that it is 0 if and only if H_0 holds. With $m_I(x) - m(x) = \delta(x)$ we have

$$
\sqrt{n}\Big(E[\varepsilon_{i,I} I\{\varepsilon_{i,I} > 0\} I\{X_i \leq t\}] - E[\varepsilon_i\{\varepsilon_i > 0\} I\{X_i \leq t\}]\Big)
$$
$$
= \sqrt{n}\Big(E[(\varepsilon_i - \delta(X_i)) I\{\varepsilon_i > \delta(X_i)\} I\{X_i \leq t\}] - E[\varepsilon_i\{\varepsilon_i > 0\} I\{X_i \leq t\}]\Big)
$$
$$
= \sqrt{n} \int_0^t \Big(\int_{\delta(x)}^\infty (y - \delta(x)) f_\varepsilon(y) dy - \int_0^\infty y f_\varepsilon(y) dy \Big) f_X(x) dx
$$

the latter expression is 0 for all $t \in [0,1]$ if and only if f_X-a.s.

$$
0 = \int_{\delta(x)}^\infty (y - \delta(x)) f_\varepsilon(y) dy - \int_0^\infty y f_\varepsilon(y) dy
$$
$$
= - \int_0^{\delta(x)} y f_\varepsilon(y) dy - \delta(x) \int_{\delta(x)}^\infty f_\varepsilon(y) dy
$$

18

$$= -\int_0^{\delta(x)} y f_\varepsilon(y) dy + \delta(x) \int_0^{\delta(x)} f_\varepsilon(y) dy - \delta(x)(1 - F_\varepsilon(0))$$

If we define

$$G(z) = \frac{1}{(1 - F_\varepsilon(0))} \int_0^z (z - y) f_\varepsilon(y) dy - z$$

the above equation is equivalent to $G(\delta(x)) = 0$ f_X-a.s. Consistency now follows if we can show that $\delta(x) = 0$ is the only solution of the above equation. To this end note, that with assumption (A4)

$$\frac{\partial}{\partial z} G(z) = \frac{1}{1 - F_\varepsilon(0)} \int_0^z f_\varepsilon(y) dy - 1 = \frac{F_\varepsilon(z) - F_\varepsilon(0)}{1 - F_\varepsilon(0)} - 1 < 0$$

and therefore G is strictly decreasing. Since $\delta(x) = 0$ is obviously a solution, this is also the only one. This means $m_I = m$ $f_X - a.s.$ which is only the case if m is increasing. Otherwise m_I and m differ on a set with positive measure. $\qquad\square$

Proof of Lemma 3.2. To prove the consistency of $\sup_{t \in [0,1]} |R_n(t)|$ or $\sup_{t \in [0,1]} |\tilde{R}_n(t)|$ we again consider the estimated expectation of both processes which should be 0 if and only if H_0 is true. Let again $\delta(x) = m_I(x) - m(x)$. Both $R_n(t)$ and $\tilde{R}_n(t)$ estimate the expectation

$$P\Big(\varepsilon_1 - \delta(X_1) \le \varepsilon_2 - \delta(X_2), \varepsilon_2 > \delta(X_2), X_2 \le t\Big) - \frac{1}{2}(1 - F_\varepsilon^2(0))$$

$$= \int_0^t \int_0^1 \Big(\int_{\mathbb{R}} F_\varepsilon(y + \delta(x) - \delta(z)) I_{\{y > \delta(z)\}} f_\varepsilon(y) dy - \int_{\mathbb{R}} F_\varepsilon(y) I_{\{y > 0\}} f_\varepsilon(y) dy \Big) f_X(x) dx f_X(z) dz$$

and this equals 0 for all $t \in [0,1]$ if and only if f_X-a.s.:

(A.1) $$\frac{1 - F_\varepsilon^2(0)}{2} = \int_0^1 \Big(\int_{\mathbb{R}} F_\varepsilon(y + \delta(x) - \delta(z)) I_{\{y > \delta(z)\}} f_\varepsilon(y) dy f_X(x) dx$$

$$= \int_0^\infty \int_0^1 F_\varepsilon(u + \delta(x)) f_X(x) dx f_\varepsilon(u + \delta(z)) du =: H(\delta(z))$$

Now assume, that $\delta(x)$ takes on different values for some different $x \in [0,1]$. Then we have positive as well as negative values for $\delta(z)$ because it is the difference between m_I and m which have to cross if and only if m is not monotone. For a positive value v of $\delta(z)$ the derivative of the function

$$H(v) = \int_0^\infty K(u) f_\varepsilon(u + v) du$$

with $K(u) = \int_0^\infty \int_0^1 F_\varepsilon(u + \delta(x)) f_X(x) dx f_\varepsilon(u + v) du$ is

$$\frac{\partial}{\partial v} H(v) = \int_0^\infty K(u) f_\varepsilon'(u + v) du < 0$$

for a unimodal density f_ε centered in 0 (assumptions (A4) and (A7)). That is H is strictly decreasing for $v \in [0, \infty)$. This is a contradiction to equation (A.1) which means that H is constant. We conclude, that for some $d^+ > 0$, $\delta(x^+) = d^+$ for all $x^+ \in I^+ = \{x \in [0, 1] \mid \delta(x) > 0\}$. Since two different negative values of δ cause two different positive value of δ, $\delta(x^-) = d^-$ for some $d^- < 0$ and for all $x^- \in I^- = \{x \in [0, 1] \mid \delta(x) < 0\}$. Because $d^+ \neq d^-$, δ would have at least one point of discontinuity which is not possible because m_I and m are both continuous functions. Therefore, $I^+ = I^- = \emptyset$, which means $\delta = 0$ $f_X - a.s.$ $\qquad \square$

A.2 Auxiliary results

Lemma A.1 *Under assumptions (A1)–(A5) under the null hypothesis of an increasing regression function m, it holds that*

$$\sup_{x \in [0,1]} |\hat{m}_I(x) - m(x)| = o_P(1), \quad \sup_{x \in [0,1]} |\hat{m}_I'(x) - m'(x)| = o_P(1)$$

$$\sup_{x,t \in [0,1]} \frac{|\hat{m}_I'(x) - m'(x) - \hat{m}_I'(t) + m'(t)|}{|x - t|^\delta} = o_P(1).$$

Proof of Lemma A.1. The first assertion directly follows from Theorem 3.3 in Birke and Dette (2008). For the second assertion we decompose

$$|\hat{m}_I'(x) - m'(x)| = \left| \frac{1}{(\hat{m}_I^{-1})'(\hat{m}_I(x))} - \frac{1}{(m^{-1})'(m(x))} \right|$$

$$\leq \left| \frac{(\hat{m}_I^{-1})'(\hat{m}_I(x)) - (m^{-1})'(\hat{m}_I(x))}{(\hat{m}_I^{-1})'(\hat{m}_I(x))(m^{-1})'(\hat{m}_I(x))} \right| + \left| \frac{1}{(m^{-1})'(\hat{m}_I(x))} - \frac{1}{(m^{-1})'(m(x))} \right|$$

and use Theorem 3.3 in Birke and Dette (2008) again and the uniform continuity of $(m^{-1})'$. It remains to establish the Lipschitz condition. To this end we distinguish two different cases where $|s - t| > b_n^2$ and $|s - t| \leq b_n^2$ for the sequence of bandwidths $h_n \to 0$ for $n \to \infty$. In the first case we derive

$$\sup_{|s-t|>b_n^2} \frac{|\hat{m}_I'(s) - m'(s) - (\hat{m}_I'(t) - m'(t))|}{|s - t|^\delta} \leq \frac{2 \sup_{s \in [0,1]} |\hat{m}_I'(s) - m'(s)|}{b_n^{2\delta}}$$

$$= O_P \left(\frac{\log h_n^{-1}}{n h_n^3 b_n^{4\delta}} \right)^{1/2} = o_P(1),$$

see Blondin (2007). In the second case we decompose

$$D_n(s,t) = |\hat{m}_I'(s) - m'(s) - (\hat{m}_I'(t) - m'(t))| \leq |\hat{m}_I'(s) - \hat{m}_I'(t)| + |m'(s) - m'(t)|$$
$$= D_n^{(1)}(s,t) + D_n^{(2)}(s,t),$$

$$D_n^{(1)}(s,t) = \frac{|(\hat{m}_I^{-1})'(\hat{m}_I(s)) - (\hat{m}_I^{-1})'(\hat{m}_I(t))|}{(\hat{m}_I^{-1})'(\hat{m}_I(s))(\hat{m}_I^{-1})'(\hat{m}_I(t))}$$

and define

$$D_n^{(1)} = \sup_{|s-t|\leq b_n^2} \frac{D_n^{(1)}(s,t)}{|s-t|^\delta} \leq C_1 \sup_{|u-v|\leq Cb_n^2} \frac{|(\hat{m}_I^{-1})'(u) - (\hat{m}_I^{-1})'(v)|}{|\hat{m}_I^{-1}(u) - \hat{m}_I^{-1}(v)|^\delta}.$$

The last inequality follows because $\hat{m}_I$ is continuously differentiable and, hence, Lipschitz continuous. Under the assumptions (A3) we obtain by using Taylor expansions

$$|\hat{m}_I^{-1}(u) - \hat{m}_I^{-1}(v)| = \frac{1}{b_n}\int_0^1 k\Big(\frac{m(x)-v}{b_n}\Big)dx|u-v|(C_2 + o_P(1))$$

$$|(\hat{m}_I^{-1})'(u) - (\hat{m}_I^{-1})'(v)| = \frac{1}{b_n}\Big|k\Big(\frac{m(1)-v}{b_n}\Big) - k\Big(\frac{m(0)-v}{b_n}\Big)\Big||u-v|(C_3 + o_P(1))$$

That means for $D_n^{(1)}$

$$D_n^{(1)} \leq C_1 \sup_{|u-v|\leq Cb_n^2} \frac{\frac{1}{b_n}\Big|k\Big(\frac{m(1)-v}{b_n}\Big) - k\Big(\frac{m(0)-v}{b_n}\Big)\Big||u-v|(C_3+o_P(1))}{\Big(\frac{1}{b_n}\int_0^1 k\Big(\frac{m(x)-v}{b_n}\Big)dx\Big)^\delta |u-v|^\delta(C_2+o_P(1))}$$

$$= O_P\Big(\sup_{|u-v|\leq Cb_n^2} \frac{|u-v|^{1-\delta}}{b_n}\Big) = O_P(b_n^{1-2\delta}) = o_P(1).$$

The regression function m is two times continuously differentiable and therefore m' is Lipschitz continuous on $[0,1]$. That means

$$\sup_{|s-t|\leq b_n^2} \frac{D_n^{(2)}(s,t)}{|s-t|^\delta} = O(b_n^{2-2\delta}) = o(1).$$

$\square$

Lemma A.2 *Under assumptions (A1)–(A5) under the null hypothesis of an increasing regression function m, it holds that*

$$\int_0^t (\hat{m}_I(x) - \hat{m}(x))f_X(x)\,dx = o_P\Big(\frac{1}{\sqrt{n}}\Big)$$

uniformly with respect to $t \in [0,1]$.

Proof of Lemma A.2. We use the representation

$$\hat{m}_I(x) - m(x) = -\frac{\hat{m}_I^{-1} - m^{-1}}{(m^{-1})'}(m(x)) + \tilde{B}_n(x)$$

(see Birke and Dette, 2008) with

$$\tilde{B}_n(x) = (\hat{m}_I^{-1}(m(x)) - m^{-1}(m(x)))\Big(\frac{1}{(m^{-1})'(m(x))} - \frac{1}{\hat{m}_I^{-1}(\eta_n(x))}\Big)$$

21

and $|\eta_n(x) - m(x)| \le |\hat{m}_I(x) - m(x)|$ for all x to rewrite

$$\int_0^t (\hat{m}_I(x) - m(x)) f_X(x) dx = -\int_{m(0)}^{m(t)} (\hat{m}_I^{-1}(u) - m^{-1}(u)) f_X(m^{-1}(u)) du + \int_0^t \tilde{B}_n(x) dx$$

$$= A_n(t) + B_n(t),$$

where

<div style="text-align:left">(A.2)</div>
$$A_n(t) = A_{n,1}(t) + A_{n,2}(t) + A_{n,3}(t)$$

and

$$A_{n,1}(t) = -\int_0^t (m_I(x) - m(x)) f_X(x) dx = 0$$

$$A_{n,2}(t) = -\frac{1}{b_n} \int_0^1 \int_{m(0)}^{m(t)} k\left(\frac{m(v) - u}{b_n}\right) f_X(m^{-1}(u)) du (\hat{m}(v) - m(v)) dv$$

$$A_{n,3}(t) = -\frac{1}{b_n} \int_0^1 \int_{m(0)}^{m(t)} k'\left(\frac{\xi(v) - u}{b_n}\right) f_X(m^{-1}(u)) du (\hat{m}(v) - m(v))^2 dv.$$

One obtains the expansion

$$A_{n,2}(t) = \left(\int_0^t (\hat{m}(v) - m(v)) f_X(v) dv - \int_0^{m^{-1}(m(0)+b_n)} (\hat{m}(v) - m(v)) f_X(v) dv \right.$$

$$\left. - \int_{m^{-1}(m(0)+b_n)}^{m^{-1}(m(t)-b_n)} (\hat{m}(v) - m(v)) f_X(v) dv \right)(1 + o_P(b_n^2))$$

$$= \left(\int_0^t (\hat{m}(v) - m(v)) f_X(v) dv - R_{n,1}(t) - R_{n,2}(t)\right)(1 + o_P(b_n^2)).$$

The two remainders $R_{n,1}(t)$ and $R_{n,2}(t)$ can be handled in the same way. We show the estimation of $R_{n,1}(t)$ here.

$$R_{n,1}(t) \le \sup |\hat{m}(v) - m(v)| \sup f_X(v) m^{-1}(m(0) + b_n) = O_P\left(\frac{b_n^2 \log h_n^{-1}}{n h_n}\right)^{1/2} = o_P\left(\frac{1}{\sqrt{n}}\right)$$

since $b_n^2 \log h_n^{-1}/h_n \to 0$. The third term $A_{n,3}(t)$ in the decomposition (A.2) can be estimated by similar means as $\Delta_n^{(2)}$ in Dette, Neumeyer and Pilz (2006) as

$$A_{n,3}(t) = O_P\left(\frac{1}{n h_n}\right) = o_P\left(\frac{1}{\sqrt{n}}\right),$$

the first one as $A_{n,1}(t) = O(b_n^2) = o(1/\sqrt{n})$ and

$$\int_0^t \tilde{B}_n(x) dx = o_P\left(\frac{1}{\sqrt{n}}\right)$$

by using similar arguments as for estimating the deterministic part and $B_{n,j}(x)$ in Birke and Dette (2007). This means

$$\int_0^t (\hat{m}_I(x) - m(x)) f_X(x) dx = \int_0^t (\hat{m}(x) - m(x)) f_X(x) dx + o_P\Big(\frac{1}{\sqrt{n}}\Big).$$

wich proves the assertion. □

Lemma A.3 *Under assumptions (A1)–(A5) under the null hypothesis of an increasing regression function m, it holds that*

$$\int_0^t (\hat{m}(x) - m(x)) f_X(x)\, dx \;=\; \frac{1}{n} \sum_{i=1}^n \varepsilon_i I\{X_i \le t\} + o_P(\frac{1}{\sqrt{n}})$$

uniformly with respect to $t \in [0,1]$.

Proof of Lemma A.3. From the proof of Proposition 2.10 by Neumeyer and Van Keilegom (2009) it follows that

$$\int_0^t (\hat{m}(x) - m(x)) f_X(x)\, dx \;=\; \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int_0^t \frac{1}{h_n} K\Big(\frac{X_i - x}{h_n}\Big)\, dx + o_P(\frac{1}{\sqrt{n}})$$

uniformly with respect to $t \in [0,1]$. Applying Theorem 2.11.23 in Van der Vaart and Wellner (1996, p. 221) (similar to, but simpler than the proof of Th. 2.7 in the aforementioned paper) one shows that the process

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i \Big(I\{X_i \le t\} - \int_0^t \frac{1}{h_n} K\Big(\frac{X_i - x}{h_n}\Big)\, dx\Big), \quad t \in [0,1],$$

converges weakly to a degenerated Gaussian process with vanishing covariances. This proves the assertion. □

A.3 Proof of main results

Proof of Theorem 4.1.
(i). The process S_n has the following simple form,

$$S_n(t) \;=\; \sqrt{n}\Big(\frac{1}{n} \sum_{i=1}^n I\{X_i \le t\} - \frac{1}{n} \sum_{i=1}^n I\{\hat{\varepsilon}_{i,I} \le 0\} I\{X_i \le t\} - \frac{1}{2}\hat{F}_{X,n}(t)\Big)$$

(A.3) $$\;=\; \sqrt{n}\Big(\frac{1}{2}\hat{F}_{X,n}(t) - \hat{F}_{X,\varepsilon_I,n}(t,0)\Big)$$

where $\hat{F}_{X,\varepsilon_I,n}$ denotes the empirical distribution function of $(X_i, \hat{\varepsilon}_{i,I})$, $i = 1, \dots, n$. Further let $F_{X,\varepsilon,n}$ denote the empirical distribution function of (X_i, ε_i), $i = 1, \dots, n$. Analogous to

23

the proof of Lemma A.2 in Neumeyer and Van Keilegom (2009) applying Lemma A.1 it holds that

$$\hat{F}_{X,\varepsilon_I,n}(t,y) = F_{X,\varepsilon,n}(t,y) + f_\varepsilon(y)\int_0^t (\hat{m}_I(x) - m(x))f_X(x)\,dx + o_P(\frac{1}{\sqrt{n}})$$

uniformly with respect to $t \in [0,1]$ and $y \in \mathbb{R}$. Applying Lemma A.2 and (A.3) it follows that

$$(A.4) \quad \hat{F}_{X,\varepsilon_I,n}(t,y) = \frac{1}{n}\sum_{i=1}^n I\{X_i \le t\}I\{\varepsilon_i \le y\} + f_\varepsilon(y)\frac{1}{n}\sum_{i=1}^n \varepsilon_i I\{X_i \le t\} + o_P(\frac{1}{\sqrt{n}})$$

and

$$S_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^n I\{X_i \le t\}\left(\frac{1}{2} - I\{\varepsilon_i \le 0\} - \varepsilon_i f_\varepsilon(0)\right) + o_P(1)$$

uniformly with respect to $t \in [0,1]$. Weak convergence to the asserted Gaussian process now follows by standard arguments. □

(ii). The proof for $\tilde{S}_n$ is very similar to (i). We have

$$\tilde{S}_n(t) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n I\{X_i \le t\} - \frac{1}{n}\sum_{i=1}^n I\{\hat{\varepsilon}_{i,I} \le 0\}I\{X_i \le t\} - (1 - \hat{F}_{\varepsilon,n}(0))\hat{F}_{X,n}(t)\right)$$

$$= \sqrt{n}\left(-\hat{F}_{X,\varepsilon_I,n}(t,0) + \hat{F}_{\varepsilon,n}(0)\hat{F}_{X,n}(t)\right)$$

Similarly to (A.4) one has

$$(A.5) \qquad \hat{F}_{\varepsilon,n}(y) = \frac{1}{n}\sum_{i=1}^n I\{\varepsilon_i \le y\} + f_\varepsilon(y)\frac{1}{n}\sum_{i=1}^n \varepsilon_i + o_P(\frac{1}{\sqrt{n}})$$

(this follows from Akritas and Van Keilegom (2001), see also Neumeyer and Van Keilegom (2009), for instance). Applying (A.4) and (A.5) we obtain the expansion

$$\tilde{S}_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\hat{F}_{X,n}(t) - I\{X_i \le t\}\right)\left(I\{\varepsilon_i \le 0\} + \varepsilon_i f_\varepsilon(0)\right) + o_P(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(F_X(t) - I\{X_i \le t\}\right)\left(I\{\varepsilon_i \le 0\} + \varepsilon_i f_\varepsilon(0)\right)$$

$$\qquad + \sqrt{n}(\hat{F}_{X,n}(t) - F_X(t))(F_\varepsilon(0) + o_P(1)) + o_P(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(I\{X_i \le t\} - F_X(t)\right)\left(F_\varepsilon(0) - I\{\varepsilon_i \le 0\} - \varepsilon_i f_\varepsilon(0)\right) + o_P(1)$$

uniformly with respect to $t \in [0,1]$. Weak convergence to a Gaussian process with the asserted covariance structure follows by standard arguments. □

(iii). From Lemma A.1 it follows that $P(m - \hat{m}_I \in \mathcal{C}) \to 1$ for $n \to \infty$, where the class $\mathcal{C} = C_1^{1+\delta}[0, 1]$ of smooth functions is defined in Van der Vaart and Wellner (1996, p. 154), and its bracketing number fulfills

$$\log N_{[]}(\epsilon, \mathcal{C}, || \cdot ||_\infty) \leq K\epsilon^{-1/(1+\delta)}$$

for some $K > 0$ and all $\epsilon > 0$, where $|| \cdot ||_\infty$ denotes the supremum norm (see Th. 2.7.1 in the same reference). Note that for the process

$$\bar{V}_n(t, h) = \frac{1}{n} \sum_{i=1}^n (\varepsilon_i + h(X_i)) I\{\varepsilon_i + h(X_i) > 0\} I\{X_i \leq t\}, \quad t \in [0, 1], h \in \mathcal{C},$$

we have

$$\bar{V}_n(t, m - \hat{m}_I) = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{i,I} I\{\hat{\varepsilon}_{i,I} > 0\} I\{X_i \leq t\}$$

(compare to the definition of V_n). Now consider the empirical process

$$\sqrt{n}(\bar{V}_n(t, h) - E[\bar{V}_n(t, h)]) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(g_{h,t}(X_i, \varepsilon_i) - E[g_{h,t}(X_i, \varepsilon_i)]\Big),$$

where the functions $g_{h,t}$ vary over the pairwise products of the function classes $\mathcal{G}$ and $\mathcal{H}$ defined as

$$\mathcal{G} = \Big\{\varepsilon \mapsto (\varepsilon + h(x)) I\{\varepsilon + h(x) > 0\} \,\Big|\, g \in \mathcal{C}\Big\}$$
$$\mathcal{H} = \Big\{x \mapsto I\{x \leq t\} \,\Big|\, t \in [0, 1]\Big\}.$$

$\mathcal{H}$ is Donsker by standard empirical process theory with bracketing numbers $N_{[]}(\epsilon, \mathcal{H}, || \cdot ||_\infty) = O(\epsilon^{-1})$. In the following we calculate bracketing numbers for $\mathcal{G}$. Let $\epsilon > 0$ and let $[h_j^L, h_j^U]$ $(j = 1, \ldots, m)$ build ϵ^2-brackets for $\mathcal{C}$, where $m = N_{[]}(\epsilon^2, \mathcal{C}, || \cdot ||_\infty)$. Then for $h_j^L \leq h \leq h_j^U$ one has

$$(\varepsilon + h_j^L(x)) I\{\varepsilon + h_j^L(x) > 0\} \leq (\varepsilon + h(x)) I\{\varepsilon + h(x) > 0\} \leq (\varepsilon + h_j^U(x)) I\{\varepsilon + h_j^U(x) > 0\}$$

and such a bracket has L^2-length

$$\Big(E\Big[\Big((\varepsilon_1 + h_j^U(X_1)) I\{\varepsilon_1 + h_j^U(X_1) > 0\} - (\varepsilon_1 + h_j^L(X_1)) I\{\varepsilon_1 + h_j^L(X_1) > 0\}\Big)^2\Big]\Big)^{1/2}$$

$$\leq \Big(2E\Big[(|\varepsilon_1| + 1)^2 (I\{\varepsilon_1 + h_j^U(X_1) > 0\} - I\{\varepsilon_1 + h_j^L(X_1) > 0\})^2 + (h_j^U(X_1) - h_j^L(X_1))^2\Big]\Big)^{1/2}$$

$$\leq \Big(2 \int \int_{-h_j^U(x)}^{-h_j^L(x)} (|y| + 1)^2 f_\varepsilon(y) \, dy \, f_X(x) \, dx + 4||h_j^U - h_j^L||_\infty\Big)^{1/2}$$

$$\leq \Big(2||h_j^U - h_j^L||_\infty E[(|\varepsilon_1| + 1)^2] + 4||h_j^U - h_j^L||_\infty\Big)^{1/2} \leq c\epsilon$$

for some constant c, where we have used that $|h_j^L| \leq 1$, $|h_j^U| \leq 1$.

The function class $\mathcal{HG}$ of pairwise products has a square-integrable envelope and the covering numbers with respect to the L^2-norm $||\cdot||_2$ fulfill

$$
\begin{aligned}
\log N_{[]}(\epsilon, \mathcal{HG}, ||\cdot||_2) &\leq \log\Big(N_{[]}(\epsilon, \mathcal{G}, ||\cdot||_2)N_{[]}(\epsilon, \mathcal{H}, ||\cdot||_2)\Big) \\
&\leq c_1 \log\Big(N_{[]}(c_2\epsilon^2, \mathcal{C}, ||\cdot||_\infty)N_{[]}(\epsilon, \mathcal{H}, ||\cdot||_2)\Big) \\
&\leq c_3 \log(\exp(\epsilon^{-2/(1+\delta)})\epsilon^{-1})
\end{aligned}
$$

for some constants c_1, c_2, c_3, for $\epsilon \leq 2(E[(|\varepsilon_1| + 1)^2])^{1/2}$, whereas for larger ϵ one bracket is sufficient, because

$$
|(\varepsilon + h(x))I\{\varepsilon + h(x) > 0\}I\{x \leq t\}| \leq |\varepsilon| + 1
$$

for all $h \in \mathcal{C}$, $t \in [0,1]$. Hence,

$$
\int_0^\infty \Big(\log N_{[]}(\epsilon, \mathcal{HG}, ||\cdot||_2)\Big)^{1/2} d\epsilon < \infty
$$

and $\mathcal{GH}$ is Donsker [see Van der Vaart and Wellner (1996, p. 129)].

This yields weak convergence of the empirical process

$$
\begin{aligned}
\check{V}_n(h,t) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big\{ &\Big((\varepsilon_i + h(X_i))I\{\varepsilon_i + h(X_i) > 0\} - \varepsilon_i I\{\varepsilon_i > 0\}\Big)I\{X_i \leq t\} \\
&- E\Big[\Big((\varepsilon_1 + h(X_1))I\{\varepsilon_1 + h(X_1) > 0\} - \varepsilon_1 I\{\varepsilon_1 > 0\}\Big)I\{X_1 \leq t\}\Big] \Big\},
\end{aligned}
$$

$t \in [0,1]$, $h \in \mathcal{C}$, to a Gaussian process. For the expectation we obtain the expansion

$$
\begin{aligned}
E\Big[\Big((\varepsilon_1 &+ h(X_1))I\{\varepsilon_1 + h(X_1) > 0\} - \varepsilon_1 I\{\varepsilon_1 > 0\}\Big)I\{X_1 \leq t\}\Big] \\
&= \int\int (y + h(x))I\{y + h(x) > 0\}I\{x \leq t\}f_X(x)f_\varepsilon(y)\,dxdy \\
&\quad - F_X(t)\int yI\{y > 0\}f_\varepsilon(y)\,dy \\
(\text{A.6}) \quad &= \int\int zI\{z > 0\}I\{x \leq t\}f_X(x)(f_\varepsilon(z - h(x)) - f_\varepsilon(z))\,dxdz
\end{aligned}
$$

which tend to zero if $\sup_{x\in[0,1]}|h(x)| \to 0$. Similarly it follows that the covariances $\text{Cov}(\check{V}_n(h,s), \check{V}_n(h,t))$ tend to zero if $\sup_{x\in[0,1]}|h(x)| \to 0$. Hence, with Van der Vaart (1998, Le. 19.24 and proof of Le. 19.26, p. 280) we obtain that $\sup_{t\in[0,1]}|\check{V}_n(m - \hat{m}_I, t)| = o_P(1)$. Inserting $h = m - \hat{m}_I$ into (A.6) for the expectation and applying Taylor's expansion we obtain

$$
\int\int zI\{z > 0\}I\{x \leq t\}f_X(x)(f_\varepsilon(z + \hat{m}_I(x) - m(x)) - f_\varepsilon(z))\,dxdz
$$

26

$$
= \int \int z I\{z > 0\} I\{x \le t\} f_X(x) f'_\varepsilon(z) (\hat{m}_I(x) - m(x)) \, dx \, dz + o_P(\frac{1}{\sqrt{n}})
$$

$$
= \int_0^\infty z f'_\varepsilon(z) \, dz \int_0^t (\hat{m}_I(x) - m(x)) f_X(x) \, dx + o_P(\frac{1}{\sqrt{n}})
$$

$$
= (F_\varepsilon(0) - 1) \frac{1}{n} \sum_{i=1}^n \varepsilon_i I\{X_i \le t\} + o_P(\frac{1}{\sqrt{n}}),
$$

where the last equality follows from integration by parts and Lemmata A.2 and A.3. Combining all results we obtain the asymptotic expansion

$$
\text{(A.7)} \qquad \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\varepsilon}_{i,I} I\{\hat{\varepsilon}_{i,I} > 0\} I\{X_i \le t\}
$$

$$
= \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i \Big(I\{\varepsilon_i > 0\} - (1 - F_\varepsilon(0)) \Big) I\{X_i \le t\} + o_P(1).
$$

For the asymptotic expansion of V_n now consider $\hat{\sigma} - \sigma = (\hat{\sigma}^2 - \sigma^2)/(\hat{\sigma} + \sigma) = (\hat{\sigma}^2 - \sigma^2)/2\sigma + o_P(1/\sqrt{n})$, from which with results by Müller, Schick and Wefelmeyer (2004) it follows that

$$
\hat{\sigma} - \sigma = \frac{1}{2\sigma n} \sum_{i=1}^n (\varepsilon_i^2 - \sigma^2) + o_P(\frac{1}{\sqrt{n}})
$$

and

$$
\frac{\hat{\sigma}}{\sqrt{2\pi}} \hat{F}_{X,n}(t) = \frac{1}{2\sqrt{2\pi}\sigma n} \sum_{i=1}^n (\varepsilon_i^2 - \sigma^2) F_X(t) + \frac{\sigma}{\sqrt{2\pi}} \frac{1}{n} \sum_{i=1}^n (I\{X_i \le t\} - F_X(t))
$$

$$
+ \frac{\sigma}{\sqrt{2\pi}} F_X(t) + o_P(\frac{1}{\sqrt{n}}).
$$

Now we have

$$
V_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(\varepsilon_i I\{\varepsilon_i > 0\} I\{X_i \le t\} - \frac{\sigma}{\sqrt{2\pi}} F_X(t) - \varepsilon_i (1 - F_\varepsilon(0)) I\{X_i \le t\}
$$

$$
- (\varepsilon_i^2 - \sigma^2) \frac{F_X(t)}{2\sqrt{2\pi}\sigma} - \frac{\sigma}{\sqrt{2\pi}} (I\{X_i \le t\} - F_X(t)) \Big) + o_P(1)
$$

$$
= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(\Big(\varepsilon_i I\{\varepsilon_i > 0\} - \frac{\sigma}{\sqrt{2\pi}} - \frac{1}{2}\varepsilon_i \Big) I\{X_i \le t\} - (\varepsilon_i^2 - \sigma^2) \frac{F_X(t)}{\sqrt{8\pi}\sigma} \Big) + o_P(1)
$$

The proof of weak convergence is omitted for the sake of brevity. The calculation of the covariances uses that under (A6),

$$
E[\varepsilon_1 I\{\varepsilon_1 > 0\}] = \frac{\sigma}{\sqrt{2\pi}}, \quad E[\varepsilon_1^2 I\{\varepsilon_1 > 0\}] = \frac{\sigma^2}{2}, \quad E[\varepsilon_1^3 I\{\varepsilon_1 > 0\}] = \frac{2\sigma^3}{\sqrt{2\pi}}.
$$

$\square$

(iv). We only give a sketch of the asymptotic expansion for the process $\tilde{V}_n$. Similarly to (A.7) one obtains

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\varepsilon}_i I\{\hat{\varepsilon}_i > 0\} \hat{F}_{X,n}(t)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \Big(I\{\varepsilon_i > 0\} - (1 - F_\varepsilon(0)) \Big) \hat{F}_{X,n}(t) + o_P(1)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \Big(I\{\varepsilon_i > 0\} - (1 - F_\varepsilon(0)) \Big) F_X(t)$$

$$- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (I\{X_i \leq t\} - F_X(t)) \Big(E[\varepsilon_1 I\{\varepsilon_1 > 0\}] + o_P(1) \Big) + o_P(1)$$

and in combination with (A.7) this yields the expansion

$$\tilde{V}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\varepsilon_i I\{\varepsilon_i > 0\} - E[\varepsilon_1 I\{\varepsilon_1 > 0\}] - \varepsilon_i (1 - F_\varepsilon(0)) \Big) \Big(I\{X_i \leq t\} - F_X(t) \Big)$$
$$+ o_P(1).$$

$\square$

(v). We first consider the process

$$\bar{R}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{R}_{i,I} I\{\hat{\varepsilon}_{i,I} > 0\} I\{X_i \leq t\},$$

where $\hat{R}_{i,I} = \hat{F}_{X,\varepsilon_I,n}(0, \hat{\varepsilon}_{i,I})$. On has the expansion $\bar{R}_n = \bar{R}_{1,n} + \bar{R}_{2,n}$, where

$$\bar{R}_{1,n}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} F_\varepsilon(\hat{\varepsilon}_{i,I}) I\{\hat{\varepsilon}_{i,I} > 0\} I\{X_i \leq t\}$$

$$\bar{R}_{2,n}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\hat{F}_{X,\varepsilon_I,n}(1, \hat{\varepsilon}_{i,I}) - F_\varepsilon(\hat{\varepsilon}_{i,I}) \Big) I\{\hat{\varepsilon}_{i,I} > 0\} I\{X_i \leq t\}.$$

Analogously to the proof of (iii) [see (A.7)] we obtain

$$\bar{R}_{1,n}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} F_\varepsilon(\varepsilon_i) I\{\varepsilon_i > 0\} I\{X_i \leq t\}$$

$$+ \sqrt{n} \int f_\varepsilon'(z) F_\varepsilon(z) I\{z > 0\} \, dz \int_0^t (\hat{m}_I(x) - m(x)) f_X(x) \, dx + o_P(1)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(F_\varepsilon(\varepsilon_i) I\{\varepsilon_i > 0\} - \varepsilon_i \Big(f_\varepsilon(0) F_\varepsilon(0) + E[f_\varepsilon(\varepsilon_1) I\{\varepsilon_1 > 0\}] \Big) \Big) I\{X_i \leq t\} + o_P(1).$$

28

Now, for the second term we have

$$\bar{R}_{2,n}(t) = \sqrt{n} \int \int \left(\hat{F}_{X,\varepsilon_I,n}(1,y) - F_\varepsilon(y) \right) I\{y > 0\} I\{x \le t\} \, dF_X(x) \, dF_\varepsilon(y)$$

$$+ \sqrt{n} \int (\hat{F}_{X,\varepsilon_I,n} - F_X \otimes F_\varepsilon)(1,y) I\{y > 0\} I\{x \le t\} \, d(\hat{F}_{X,\varepsilon_I,n} - F_X \otimes F_\varepsilon)(x,y).$$

The last integral converges to zero uniformly in probability, because the process $\sqrt{n}(\hat{F}_{X,\varepsilon_I,n} - F_X \otimes F_\varepsilon)$ converges to a Gaussian process, see the proof of (i). Inserting the expansion (A.4) we further have

$$\bar{R}_{2,n}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int \left(I\{\varepsilon_i \le y\} - F_\varepsilon(y) + f_\varepsilon(y)\varepsilon_i \right) I\{y > 0\} \, dF_\varepsilon(y) \, F_X(t) + o_P(1)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(1 - F_\varepsilon(0) + I\{\varepsilon_i > 0\}(F_\varepsilon(0) - F_\varepsilon(\varepsilon_i)) - E[F_\varepsilon(\varepsilon_1)I\{\varepsilon_1 > 0\}]$$

$$+ \varepsilon_i E[f_\varepsilon(\varepsilon_1)I\{\varepsilon_1 > 0\}] \Big) F_X(t) + o_P(1).$$

We obtain the expansion

$$R_n(t) = \bar{R}_{1,n}(t) + \bar{R}_{2,n}(t) - E[F_\varepsilon(\varepsilon_1)I\{\varepsilon_1 > 0\}] \frac{1}{\sqrt{n}} \sum_{i=1}^n I\{X_i \le t\}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(I\{X_i \le t\} - F_X(t) \Big) \Big(F_\varepsilon(\varepsilon_i)I\{\varepsilon_i > 0\} - E[F_\varepsilon(\varepsilon_1)I\{\varepsilon_1 > 0\}]$$

$$+ \varepsilon_i \Big(f_\varepsilon(0)F_\varepsilon(0) - E[f_\varepsilon(\varepsilon_1)I\{\varepsilon_1 > 0\}] \Big) \Big)$$

$$+ \Big(1 - F_\varepsilon(0) + I\{\varepsilon_i > 0\}F_\varepsilon(0) - 2E[F_\varepsilon(\varepsilon_1)I\{\varepsilon_1 > 0\}] \Big) F_X(t) + o_P(1).$$

$\square$

(vi). With (A.5) one obtains

$$\hat{F}_{\varepsilon,n}^2(0) = \frac{2F_\varepsilon(0)}{n} \sum_{i=1}^n (I\{\varepsilon_i \le 0\} + f_\varepsilon(0)\varepsilon_i) - F_\varepsilon^2(0) + o_P(\frac{1}{\sqrt{n}}).$$

Further we have

$$\frac{1}{2}(1 - \hat{F}_{\varepsilon,n}^2(0))\hat{F}_X(t) = \frac{1}{2} \Big(\hat{F}_X(t) - \hat{F}_{\varepsilon,n}^2(0)F_X(t) - F_\varepsilon^2(0)(\hat{F}_X(t) - F_X(t)) \Big) + o_P(\frac{1}{\sqrt{n}})$$

$$= \frac{1}{n} \sum_{i=1}^n \Big(\frac{1}{2}(1 - F_\varepsilon^2(0))I\{X_i \le t\} - F_X(t)F_\varepsilon(0)(I\{\varepsilon_i \le 0\} + f_\varepsilon(0)\varepsilon_i) \Big)$$

$$+ F_\varepsilon^2(0)F_X(t) + o_P(\frac{1}{\sqrt{n}}),$$

29

and hence, with notations from the proof of (v),

$$
\begin{aligned}
\tilde{R}_n(t) \; = \; & \bar{R}_{1,n}(t) + \bar{R}_{2,n}(t) - \frac{\sqrt{n}}{2}(1 - \hat{F}^2_{\varepsilon,n}(0))\hat{F}_X(t) \\
= \; & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (I\{X_i \le t\} - F_X(t))\Big(F_\varepsilon(\varepsilon_i)I\{\varepsilon_i > 0\} - E[F_\varepsilon(\varepsilon_1)I\{\varepsilon_1 > 0\}] \\
& - \varepsilon_i(f_\varepsilon(0)F_\varepsilon(0) + E[f_\varepsilon(\varepsilon_1)I\{\varepsilon_1 > 0\}])\Big) + o_P(1).
\end{aligned}
$$

$\square$

A.4 Validity of bootstrap

In the following we sketch the proof of bootstrap consistency for the process

$$
\tilde{S}^*_n(t) \; = \; \sqrt{n}\Big(-\hat{F}^*_{X,\varepsilon_I,n}(t,0) + \hat{F}^*_{\varepsilon,n}(0)\hat{F}_{X,n}(t) \Big)
$$

(the other processes are treated similarly). From Neumeyer (2009, Lemma 2) (confer this reference also for restrictions on the bandwidth a) we have

$$
\hat{F}^*_{\varepsilon,n}(y) \; = \; \frac{1}{n} \sum_{i=1}^{n} I\{\varepsilon^*_i \le y\} + \tilde{f}_{\varepsilon,n}(y)\frac{1}{n} \sum_{i=1}^{n} \varepsilon^*_i + o_P(\frac{1}{\sqrt{n}}).
$$

Combining the methods shown in the paper at hand with the methods by Neumeyer (2009) one can show similarly to (A.4) that

$$
\hat{F}^*_{X,\varepsilon_I,n}(t,y) \; = \; \frac{1}{n} \sum_{i=1}^{n} I\{X_i \le t\}I\{\varepsilon^*_i \le y\} + \tilde{f}_{\varepsilon,n}(y)\frac{1}{n} \sum_{i=1}^{n} \varepsilon^*_i I\{X_i \le t\} + o_P(\frac{1}{\sqrt{n}}).
$$

Hence we obtain the expansion $\tilde{S}^*_n(t) = \tilde{S}^{*,1}_n(t) + o_P(1)$, where

$$
\tilde{S}^{*,1}_n(t) \; = \; \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\hat{F}_{X,n}(t) - I\{X_i \le t\} \Big)\Big(I\{\varepsilon^*_i \le 0\} - \tilde{F}_{\varepsilon,n}(0) + \varepsilon^*_i \tilde{f}_{\varepsilon,n}(0) \Big)
$$

uniformly with respect to $t \in [0,1]$. Note that bootstrap results are formulated conditionally on the original sample $\mathcal{Y}_n = \{(X_i, Y_i) \mid i = 1, \dots, n\}$ and hence, that in this expansion $\hat{F}_{X,n}$ and $\tilde{F}_{\varepsilon,n}, \tilde{f}_{\varepsilon,n}$ are "not random". Further, for the conditional expectation we have $E[\tilde{S}^{*,1}_n(t) \mid \mathcal{Y}_n] = 0$. To derive weak convergence we first consider the conditional covariances

$$
\begin{aligned}
\mathrm{Cov}(\tilde{S}^{*,1}_n(t), \tilde{S}^{*,1}_n(s) \mid \mathcal{Y}_n) \; = \; & \Big(\hat{F}_{X,n}(s \wedge t) - \hat{F}_{X,n}(s)\hat{F}_{X,n}(t) \Big)\Big(\tilde{F}_{\varepsilon,n}(0)(1 - \tilde{F}_{\varepsilon,n}(0)) \\
& + \tilde{f}^2_{\varepsilon,n}(0)\int y^2 \tilde{f}_{\varepsilon,n}(y)\,dy + 2\tilde{f}_{\varepsilon,n}(0)\int yI\{y \le 0\}\tilde{f}_{\varepsilon,n}(y)\,dy \Big),
\end{aligned}
$$

which (under regularity conditions) converge almost surely to $\mathrm{Cov}(\tilde{S}(t), \tilde{S}(s))$ as defined in Theorem 4.1.

Using Cramér-Wold's device and Lindeberg's condition one can show convergence of the finite dimensional distributions analogously to the proof of Theorem 3 by Neumeyer (2009). To show tightness we follow the proof of Lemma A.3 by Stute, González Manteiga and Presedo Quindimil (1998). To this end, note that the dominating term of the process has the form

$$\tilde{S}_n^{*,1}(t) \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\hat{F}_{X,n}(t) - I\{X_i \le t\} \right) Z_{n,i}^*$$

where $E[Z_{n,i}^* \mid \mathcal{Y}_n] = 0$ and $E[(Z_{n,i}^*)^4 \mid \mathcal{Y}_n]$ is independent of i and converges for $n \to \infty$ to positive constants with probability one. For $0 \le s \le t \le r \le 1$ one can easily obtain the bound

$$E\left[\left(\tilde{S}_n^{*,1}(t) - \tilde{S}_n^{*,1}(s) \right) \left(\tilde{S}_n^{*,1}(r) - \tilde{S}_n^{*,1}(t) \right) \; \Big| \; \mathcal{Y}_n \right] \;\le\; C \left(\hat{F}_{X,n}(r) - \hat{F}_{X,n}(s) \right)^2 E[(Z_{n,i}^*)^4 \mid \mathcal{Y}_n]$$

for some constant C. Tightness follows as in the aforementioned proof from monotonicity of $\hat{F}_{X,n}$, which converges almost surely to F_X.

Altogether we obtain that the process $\tilde{S}_n^*$, conditional on $\mathcal{Y}_n$ converges weakly to the process $\tilde{S}$ as defined in Theorem 4.1, in probability. □

B   References

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

Y. Baraud, S. Huet and B. Laurent (2003). *Adaptive tests of qualitative hypotheses* ESAIM, Probab. Statist. 7, 147–159.

M. Birke and H. Dette (2007). *Testing strict monotonicity in nonparametric regression.* Math. Meth. Statist. 16, p. 110–123.

M. Birke and H. Dette (2008). *A note on estimating a smooth monotone regression by combining kernel and density estimates.* J. Nonparam. Statist. 20, 679–691.

D. Blondin (2007). *Rates of strong uniform consistency for local least squares kernel regression estimators.* Stat. Probab. Lett. 77, 1526–1534.

A. W. Bowman, M. C. Jones and I. Gijbels (1998). *Testing monotonicity of regression.* J. Computational and Graphical Statist. 7, 489–500.

H. Dette, N. Neumeyer and K. F. Pilz (2006). *A simple nonparametric estimator of a strictly monotone regression function.* Bernoulli 12 , 469–490.

H. Dette and K. F. Pilz (2006). *A comparative study of monotone nonparametric kernel estimates.* J. Stat. Comput. Simul. 76, 41–56.

J. Domínguez-Menchero, G. González-Rodríguez and M. J. López-Palomo (2005). *An L2 point of view in testing monotone regression* J. Nonparam. Statist. 17, 135–153.

L. Dümbgen (2002). *Application of local rank tests to nonparametric regression.* J. Nonparametr. Stat. 14, 511–537.

C. Durot (2003). *A Kolmogorov-type test for monotonicity of regression.* Statist. Probab. Lett. 63, 425–433.

I. Gijbels (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.

I. Gijbels, P. Hall, M. C. Jones and I. Koch (2000). *Tests for monotonicity of a regression mean with guaranteed level.* Biometrika 87, 663–673.

S. Ghosal, A. Sen, A. W. van der Vaart (2000). *Testing monotonicity of regression.* Ann. Statist. 28, 1054–1082.

P. Hall and N. E. Heckman (2000). *Testing for monotonicity of a regression mean by calibrating for linear functions.* Ann. Statist. 28, 20–39.

P. Hall and L.-S. Huang (2001). *Nonparametric kernel regression subject to monotonicity constraints.* Ann. Statist. 29, 624–647.

W. Härdle (1990). *Applied nonparametric regression.* Cambridge Univ. Press., Cambridge

W. Härdle and E. Mammen (1993). *Comparing nonparametric versus parametric regression fits.* Ann. Statist. 21, 1926–1947.

Hardy, G.H., Littlewood, J.E. and Pólya, G. (1952). *Inequalities.* 2nd ed., Cambridge University Press.

Lieb, E.H. and Loss, M. (2001). *Analysis.* 2nd ed., Graduate Studies in Mathematics. 14. Providence, RI: American Mathematical Society

E. Mammen (1991). *Estimating a smooth monotone regression function.* Ann. Statist. 19, 724–740.

U. U. Müller, A. Schick, W. Wefelmeyer (2004). *Estimating linear functionals of the error distribution in nonparametric regression.* J. Statist. Plann. Inference 119, 75–93.

N. Neumeyer (2009). *Smooth residual bootstrap for empirical processes of nonparametric regression residuals.* Scand. J. Statist. 36, 204–228.

N. Neumeyer and I. Van Keilegom (2009). *Changepoint tests for the error distribution in nonparametric regression.* Scand. J. Statist. 36, 518 - 541.

W. Schlee (1982). *Nonparametric tests of the monotony and convexity of regression.* Nonparametric statistical inference, Vol. I, II, 823–836, Colloq. Math. Soc. János Bolyai, 32, North-Holland, Amsterdam.

B. W. Silverman (1981). *Using kernel density estimates to investigate multimodality.* J. Roy. Statist. Soc. Ser. B 43, 97–99.

W. Stute (1997). *Nonparametric model checks for regression.* Ann. Statist. 25, 613–641.

W. Stute, W. González Manteiga and M. Presedo Quindimil (1998). *Bootstrap Approximation in Model Checks for Regression.* J. Amer. Statist. Assoc. 93, 141–149.

A. W. Van der Vaart (1998). *Asymptotic statistics.* Cambridge University Press, Cambridge.

A. W. Van der Vaart and J. A. Wellner (1996). *Weak convergence and empirical processes.* Springer, New York.

I. Van Keilegom, W. González–Manteiga and C. Sánchez Sellero (2008). *Goodness-of-fit tests in parametric regression based on the estimation of the error distribution.* TEST 17, 401–415.