UNIVERSITÄT HAMBURG

A bootstrap version of the residual-based smooth empirical distribution function

Natalie Neumeyer

Preprint–No. 2006-08 November 2006

FACHBEREICH MATHEMATIK

SCHWERPUNKT MATHEMATISCHE STATISTIK UND STOCHASTISCHE PROZESSE

BUNDESSTR. 55, D – 20146 HAMBURG

A bootstrap version of the residual-based smooth empirical distribution function

Natalie Neumeyer
Universität Hamburg
Department Mathematik, Schwerpunkt Stochastik
Bundesstrasse 55, 20146 Hamburg, Germany
e-mail: neumeyer@math.uni-hamburg.de

November 3, 2006

Abstract

In this paper we consider estimating the error distribution in a nonparametric regression model by a smooth version of the empirical distribution function of residuals. We show that a classical residual bootstrap version of the resulting residual-based empirical process joins the same limiting distribution. From this result consistency of various goodness-of-fit tests in nonparametric regression models is obtained.

AMS Classification: 62G09, 62G08
Keywords and Phrases: error distribution, kernel estimator, nonparametric regression, resampling

1 Introduction

Assume we have independent observations $(X_1, Y_1), \dots, (X_n, Y_n)$ from a nonparametric regression model

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, \dots, n, \tag{1.1}$$

with unkown regression function m, and we are interested in the centered distribution of Y given the covariate, i. e. the distribution F of the unobservable errors $\varepsilon_1, \dots, \varepsilon_n$, which are assumed to be independent and identically distributed. The canonical idea seems to be to consider the empirical distribution function

$$\hat{F}_n(y) \;=\; \frac{1}{n}\sum_{i=1}^{n} I\{\hat{\varepsilon}_i \le y\}, \quad y \in I\!R, \tag{1.2}$$

based on nonparametrically estimated residuals

$$\hat{\varepsilon}_i \;=\; Y_i - \hat{m}(X_i), \quad i = 1, \ldots, n. \tag{1.3}$$

Similar empirical distribution functions and weak convergence of the corresponding empirical processes based on residuals from linear models and parametric regression models have been studied intensively by Koul (1970, 2002), Loynes (1980), Shorack (1984), Portnoy (1986), Bai (1996) and Mammen (1996), among others. In a nonparametric regression context Akritas and Van Keilegom (2001) investigate asymptotic properties of the empirical distribution function of nonparametrically estimated residuals and show weak convergence of the corresponding empirical process $\sqrt{n}(\hat{F}_n - F)$. As related literature we should mention Cheng (2002), who showed uniform consistency of a histogram density estimator and the empirical distribution function of nonparametric residuals, whereas Cheng (2004) gives similar results for a residual-based kernel density estimator. Müller, Schick and Wefelmeyer (2004a, 2004b) consider smooth linear functionals based on nonparametric regression residuals.
In this paper we follow the approach by Müller, Schick and Wefelmeyer (2004c), who consider a smooth version of the empirical distribution function of nonparametric residuals, i. e.

$$\overline{F}_n(y) \;=\; \frac{1}{n}\sum_{i=1}^{n} L\Big(\frac{y - \hat{\varepsilon}_i}{b_n}\Big), \quad y \in I\!R, \tag{1.4}$$

where b_n denotes a sequence of bandwidths and L some distribution function. Smoothing an empirical distribution function by kernel methods was already considered by Nadaraya (1964b) and Falk (1983). Fernholz (1991) and van der Vaart (1994) showed that the smoothed and not smoothed empirical processes based on a sample of independent and identically distributed random variables have the same asymptotic law, when the underlying distribution function is Lipschitz continuous; see also Winter (1973) for a Glivenko–Cantelli type result. Azzalini (1981), Reiss (1981), Hill (1985), among others pointed out that the smooth version of the empirical distribution function can have a better performance than the non-smooth version for suitable kernel and bandwidths.
Under suitable assumptions Müller et al. (2004c) prove weak convergence of the process $\sqrt{n}(\overline{F}_n - F)$ to some centered Gaussian process G. The covariance structure of the limiting

distribution depends heavily on the unknown distribution function F and its density. To apply the process, for example, to construct testing procedures such as a test for equality of error distributions in different regression models, it is therefore sensible to use resampling procedures. We will show in this paper that the (classical) residual bootstrap as was introduced by Härdle and Bowman (1988) can be applied here. In particular we show that, conditional on an initial sample, the bootstrap version of the process $\sqrt{n}(\overline{F}_n - F)$ converges weakly to G in probability. This paper extends the work by Neumeyer (2006), who proves consistency of a smooth residual bootstrap version of the (non-smooth) residual-based empirical distribution considered by Akritas and Van Keilegom (2001). We will also extend Müller et al.'s (2004c) results to a heteroscedastic setting, where then a version of the classical residual bootstrap can be applied that takes into account the heteroscedasticity. Finally, we will discuss how the results apply to construct new consistent versions of various hypotheses tests in nonparametric regression models including goodness-of-fit tests. Here we consider in detail tests for symmetry of F and tests for equality of error distributions. We also consider briefly how our new results can be applied to obtain new versions of goodness-of-fit tests for the regression function [see Van Keilegom, González–Manteiga and Sánchez Sellero (2004)] or the variance function in a heteroscedastic model [see Dette and Van Keilegom (2005)], as well as tests for equality of regression functions [see Pardo–Fernández, Van Keilegom and González–Manteiga (2004)].

The remainder of the paper is organized as follows. Some asymptotic results for the processes defined in (1.2) and (1.4) are given in section 2. In section 3 the main results concerning the residual bootstrap are given, which are extended to the heteroscedastic model in section 4. Applications to model tests are discussed in section 5 and some simulation results are presented in section 6. Finally, the proofs of the main results are given in the appendix.

2 Assumptions and first results

In the definition (1.3) of the residuals let $\hat{m}(x)$ denote the Nadaraya-Watson [Nadaraya (1964a) and Watson (1964)] kernel regression estimator for $m(x)$, that is

(2.1) $$\hat{m}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}K\Big(\frac{X_i - x}{h_n}\Big)Y_i\frac{1}{\hat{f}_X(x)}$$

with the kernel density estimator $\hat{f}_X(x)$ for the design density $f_X(x)$, i. e.

(2.2) $$\hat{f}_X(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}K\Big(\frac{X_i - x}{h_n}\Big).$$

Here, K denotes a kernel function and h_n a sequence of bandwidths. Note, that other uniformly consistent nonparametric function estimators such as local polynomial estimators [see, e. g. Fan and Gijbels (1996)] can be applied as well and very similar results can be obtained, compare Remark 2.2 in Neumeyer (2006). Here, we impose the following assumptions.

A2.1 The univariate design points $X_1, \ldots, X_n$ 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 in $(0,1)$.

A2.2 The errors $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed with distribution function F. They are centered, i. e. $E[\varepsilon_i] = 0$, with variance $\sigma^2 = \mathrm{Var}(\varepsilon_i) \in (0, \infty)$, existing fourth moment, $E[\varepsilon_i^4] < \infty$, and are independent of the design points. F is twice continuously differentiable with positive density f such that $\sup_{y \in \mathbb{R}} f(y) < \infty$ and $\sup_{y \in \mathbb{R}} |f'(y)| < \infty$.

A2.3 Let K denote a twice continuously differentiable symmetric density with compact support $[-1,1]$, say, such that $\int uK(u)\,du = 0$ and $K(-1) = K(1) = 0$.

A2.4 Let h_n denote a sequence of positive bandwidths such that for $n \to \infty$, $nh_n^4 = o(1)$, and there exists a constant $\alpha > 0$ such that $nh_n^{3+\alpha} \to \infty$.

The constant α in assumption A2.4 can be chosen arbitrarily small. It has only relevance according to techniques of proofs. Note that Akritas and Van Keilegom's (2001) bandwidth condition (A1)(i),

$$\frac{nh_n^{3+2\delta}}{\log(h_n^{-1})} \to \infty, \tag{2.3}$$

follows from our assumption A2.4 for $\delta = \alpha/4$, for instance.

We first give the weak convergence result for the process $\sqrt{n}(\hat{F}_n - F)$, which is due to Akritas and Van Keilegom, who consider a heteroscedastic model, where the regression and variance functions are defined as certain L-functionals depending on a score function J in $[0,1]$. Here and throughout the paper $I\{\cdot\}$ denotes the indicator function and $y \wedge z$ denotes the minimum of y and z.

Theorem 2.1 *Under model (1.1) with assumptions A2.1–4 the process* $\sqrt{n}(\hat{F}_n(y) - F(y))$, $y \in \mathbb{R}$, *converges weakly to a centered Gaussian process* G *with covariance*

$$\mathrm{Cov}(G(y), G(z)) \;=\; F(y \wedge z) - F(y)F(z) + f(y)f(z)\sigma^2 + f(y)U(z) + f(z)U(y),$$

where $U(y) = E[\varepsilon_1 I\{\varepsilon_1 \leq y\}]$.

For the smooth version $\overline{F}_n$ of the residual-based empirical distribution function defined in (1.4) we need the following assumptions.

A2.5 Let ℓ be a symmetric density function with compact support, say $[-1,1]$, such that $\int \ell(u)u\,du = 0$, and let $L(y) = \int_{-\infty}^{y} \ell(t)\,dt$ denote the corresponding distribution function.

A2.6 Let b_n denote a sequence of positive bandwidths such that for $n \to \infty$, $nb_n^4 = o(1)$, $nb_n \to \infty$.

We have the following weak convergence result.

Theorem 2.2 *Under model (1.1) with assumptions A2.1–6 the process $\sqrt{n}(\overline{F}_n(y) - F(y))$, $y \in \mathbb{R}$, converges weakly to a centered Gaussian process G with the same covariance as defined in Theorem 2.1.*

Theorem 2.2 was shown by Müller et al. (2004c) under slightly different assumptions. However, we give a much shorter proof here that presents ideas used later [compare also the proof of Th. 2.7 and Prop. 2.8 in the aforementioned reference].

Proof of Theorem 2.2. We have by definition

$$\overline{F}_n(y) \;=\; \int L\Big(\frac{y-z}{b_n}\Big)\,d\hat{F}_n(z) \;=\; \int \hat{F}_n(y - b_n u)\ell(u)\,du$$

with $\hat{F}_n$ from (1.2). Let $R_n = \sqrt{n}(\hat{F}_n - F)$ denote the empirical process with limit distribution as in Theorem 2.1. Then,

$$\begin{aligned}
&\sqrt{n}\sup_{y\in\mathbb{R}}\Big|\hat{F}_n(y) - F(y) - \overline{F}_n(y) + \int F(y - b_n u)\ell(u)\,du\Big| \\
&= \sup_{y\in\mathbb{R}}\Big|R_n(y) - \int R_n(y - b_n u)\ell(u)\,du\Big| \\
&= \sup_{y\in\mathbb{R}}\Big|\int_{-1}^{1}(R_n(y) - R_n(y - b_n u))\ell(u)\,du\Big| \\
&\le \sup_{\substack{y,z\in\mathbb{R}\\ |y-z|\le b_n}}\Big|R_n(y) - R_n(z)\Big| \;=\; o_p(1)
\end{aligned}$$

which follows from asymptotic equicontinuity of the process R_n [compare Theorem 2.1 and van der Vaart and Wellner (1996), p. 89, 90] and $b_n \to 0$. Further, we have by Taylor's expansion

$$
\begin{aligned}
&\sqrt{n} \sup_{y\in I\!R} \Big| F(y) - \int F(y - b_n u)\ell(u)\, du \Big| \\
&= \sqrt{n} \sup_{y\in I\!R} \Big| \int (F(y) - F(y - b_n u))\ell(u)\, du \Big| \\
&\le \sqrt{n} \Big(\sup_{y\in I\!R} \Big| b_n f(y) \int \ell(u) u\, du \Big| + b_n^2 \sup_{y\in I\!R} |f'(y)| \int |\ell(u)| u^2\, du \Big) \\
&= O(\sqrt{n} b_n^2) \;=\; o(1)
\end{aligned}
$$

by assumption A2.6. Therefore, the process $\sqrt{n}(\overline{F}_n - F)$ has the same limit distribution as R_n and the assertion follows from Theorem 2.1. □

Note that the limit distribution G from Theorem 2.2 is not known, but depends on the unknown error distribution, its density and the unknown variance. When applying the limit law for statistical inference (compare section 5) either the covariance of G has to be estimated nonparametrically or resampling procedures have to be applied. In similar problems in nonparametric regression, resampling procedures can lead to much more precise critical values for small and moderate sample sizes. We will show in the following section that a residual bootstrap can be applied here.

3 Residual bootstrap

In this section we consider a bootstrap method for imitating the distribution of statistics based on the smooth version of the empirical distribution function based on residuals. We will prove that a classical residual bootstrap is applicable. To this end denote by $\mathcal{Y}_n$ the original sample $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, and let $\tilde{\varepsilon}_1^*, \ldots, \tilde{\varepsilon}_n^*$ denote a random sample drawn with replacement from the centered residuals $\{\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n\}$ defined by

$$
\tilde{\varepsilon}_i \;=\; \hat{\varepsilon}_i - \frac{1}{n} \sum_{j=1}^{n} \hat{\varepsilon}_j, \tag{3.1}
$$

with $\hat{\varepsilon}_i$ from (1.3). We have $E[\tilde{\varepsilon}_i^* \mid \mathcal{Y}_n] = 0$ because of the centering of $\tilde{\varepsilon}_j$. Denote by $\widehat{\tilde{F}}_n$ the empirical distribution function of the centered residuals $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$. We define bootstrap observations as

$$
\tilde{Y}_i^* \;=\; \hat{m}(X_i) + \tilde{\varepsilon}_i^*, \tag{3.2}
$$

where the regression estimator $\hat{m}$ is defined in (2.1). From the bootstrap sample $(X_1, \tilde{Y}_1^*), \ldots,$ $(X_n, \tilde{Y}_n^*)$ we obtain the estimator $\tilde{m}^*$ for the regression function, analogous to (2.1), and residuals

$$\widehat{\tilde{\varepsilon}}_i^* = \tilde{Y}_i^* - \tilde{m}^*(X_i). \tag{3.3}$$

The bootstrap residual-based smooth empirical distribution function (a residual bootstrap version for $\overline{F}_n$) is defined as

$$\overline{F}_n^*(y) = \frac{1}{n}\sum_{i=1}^{n} L\Big(\frac{y - \widehat{\tilde{\varepsilon}}_i^*}{b_n}\Big). \tag{3.4}$$

We impose the following assumptions.

A3.1 Let ℓ be a symmetric probability density with compact support, $\int \ell(u)u\,du = 0$, with λ continuous bounded derivatives inside the support, which are of bounded variation. Let $L(y) = \int_{-\infty}^{y} \ell(t)\,dt$ denote the corresponding distribution function.

A3.2 Let b_n be a sequence of positive bandwidths such that $nb_n^4 \to 0$, $\log n/(nb_n^3) \to 0$,

$$\frac{\log h_n^{-1}}{nh_n b_n^{2+\tilde{\alpha}}} \to 0\,, \quad \frac{(\log h_n^{-1})^{1+1/\lambda}}{nh_n^{1+\frac{1}{\lambda}} b_n^{2+\frac{2}{\lambda}}} \to 0\,, \quad \frac{b_n^{8/(\lambda+1)} \log h_n^{-1}}{h_n} \to 0\,, \quad \frac{h_n^{2\nu+2}}{b_n^{(2\nu-1)\frac{\lambda+3}{\lambda+1}}} = O(1)$$

for $\nu = 1, 2$, where $\tilde{\alpha} = \frac{4+8\delta(\lambda+3)/3}{\lambda+1}$ with δ from (2.3), λ from assumption A3.1, and h_n fulfils bandwidth conditions A2.4.

A3.3 Let $E[|\varepsilon_1|^{\upsilon}] < \infty$ for some $\upsilon > 1 + \frac{2}{\delta}$.

Remark 3.1 Compare Remark 6.9 in Neumeyer (2006) on how condition A3.3 can be omitted by applying boundary adapted kernel estimators.

Note that the sophisticated bandwidth conditions as in assumption A3.2 are stated exactly as they appear in the rather technical proof. More primitive sufficient conditions could be imposed and we give two examples. For instance, when a very smooth kernel is applied with $\lambda = 6$, both h_n and b_n can be chosen to be of the same order, say $n^{-\beta}$. Then we only have to impose condition $0.25 < \beta < 0.28$ instead of assumptions A2.4 and A3.2 (where we assume that δ can be chosen arbitrarily small because assumption A3.3 can be omitted).

There are various possibilities to derive different bandwidth conditions for b_n and h_n under which the following Theorem 3.2 is valid. The bandwidth conditions in A3.2 originate from the choice $a_n = b_n^{(\lambda+3)/(\lambda+1)}$ for some auxiliary bandwidth in the proof of the theorem, where $a_n = o(1)$ has to fulfil

$$\frac{nh_n a_n^{2+8\delta/3}}{\log h_n^{-1}} \to \infty, \quad \frac{h_n^{2\nu+2}}{a_n^{2\nu-1}} = O(1) \quad \text{for } \nu = 1, 2. \tag{3.5}$$

Different choices of a_n in dependence of b_n will lead to different bandwidth conditions. Further we obtain much less restrictive bandwidth conditions when the density ℓ is chosen such that $\ell^{(\lambda_0)} \equiv 0$ for some λ_0 [for example a truncated polynomial]. Then, bandwidth condition $(\log h_n^{-1})^{1+1/\lambda}/(nh_n^{1+1/\lambda} b_n^{2+2/\lambda}) \to 0$ is not required. For example, when the auxiliary bandwidth a_n in the proof of Theorem 3.2 is then chosen as $a_n = b_n h_n^{\omega+1/4}$ the bandwidth conditions in A3.2 can be relaxed to

$$nb_n^4 \to 0\,, \quad \frac{\log n}{nb_n^3} \to 0\,, \quad \frac{\log h_n^{-1}}{nh_n^{\frac{3}{2}+2\omega+\delta+8\delta\omega/3} b_n^{2+8\delta/3}} \to 0\,, \quad \frac{b_n^2}{h_n^{\omega+1/4}} = O(1)$$

for some $\omega \in (0, \frac{5}{12})$.
For instance, when h_n is of order $n^{-\beta}$ and b_n is of order $n^{-\gamma}$, then (for small δ and ω) instead of assumptions A2.4 and A3.2 we only have to impose the constraints $\beta, \gamma \in (\frac{1}{4}, \frac{1}{3})$, where $3\beta + 4\gamma < 2$. ■

We now state the main result which shows asymptotic validity of the bootstrap procedure. The proof of Theorem 3.2 is given in the appendix. Applications to testing problems are discussed in section 5.

Theorem 3.2 *Assume model (1.1) is valid under assumptions A2.1–4 and A3.1–3. Then, conditional on the sample $\mathcal{Y}_n$, the process $\sqrt{n}(\overline{F}_n^*(y) - \hat{\tilde{F}}_n(y))$, $y \in I\!R$, converges weakly to a centered Gaussian process G with covariance* $\mathrm{Cov}(G(y), G(z))$ *defined in Theorem 2.2, in probability.*

Remark 3.3 The process is centered by $\hat{\tilde{F}}_n$, the empirical distribution function of the centered residuals $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$, i.e. the conditional distribution function of the bootstrap errors $\tilde{\varepsilon}_i^*$. Nevertheless, the same theorem is valid when the process is centered by the empirical distribution function of uncentered residuals $\hat{\varepsilon}_i$, i.e. $\hat{F}_n$ defined in (1.2). The process $\sqrt{n}(\overline{F}_n^*(y) - \hat{F}_n(y))$ has the same limit distribution as $\sqrt{n}(\overline{F}_n^*(y) - \hat{\tilde{F}}_n(y))$ with a reasoning based on the fact that

$$\frac{1}{n}\sum_{j=1}^{n} \hat{\varepsilon}_j \;=\; o_p(\frac{1}{\sqrt{n}})$$

[compare Müller et al. (2004b) and Kiwitt, Nagel and Neumeyer (2005)]. This rate shows that under our assumptions the difference of centered and not centered residuals is of negligible order.
With the same argumentation it can be shown that the process $\sqrt{n}(\hat{\tilde{F}}_n(y) - F(y))$ has the same limit distribution as given in Theorem 2.1. ■

Validity of bootstrap procedures can be deduced as follows. Let T_n and its bootstrap version T_n^* be continuous functionals ψ evaluated in the stochastic processes $S_n = \sqrt{n}(\overline{F}_n(y) - F(y))$ and $S_n^* = \sqrt{n}(\overline{F}_n^*(y) - \widehat{\overline{F}}_n(y))$, respectively, such that T_n converges in distribution to $T = \psi(G)$, and, conditional on $\mathcal{Y}_n$, T_n^* converges in distribution to T, in probability. Define $c_{n,\alpha}^*$ by $P(T_n^* \leq c_{n,\alpha}^* \mid \mathcal{Y}_n) = \alpha$. Then we have the following result [see Beran, LeCam and Millar (1987) or Delgado and González Manteiga (2001)].

Corollary 3.4 *Under the assumptions of Theorem 3.2 it holds that* $P(T_n \leq c_{n,\alpha}^*) = \alpha + o(1)$.

4 Extension to a heteroscedastic model

In this section we consider the heteroscedastic analogue of model (1.1), that is

$$(4.1) \qquad Y_i \;=\; m(X_i) + \sigma(X_i)\varepsilon_i, \quad i = 1, \ldots, n,$$

under the following assumptions.

A4.1 The univariate design points $X_1, \ldots, X_n$ 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 and the variance function σ^2 are twice continuously differentiable in $(0, 1)$ such that $\inf_{x\in[0,1]} \sigma^2(x) > 0$.

A4.2 The errors $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed with distribution function F. They are centered, i. e. $E[\varepsilon_i] = 0$, with unit variance $\mathrm{Var}(\varepsilon_i) = 1$, existing fourth moment, $E[\varepsilon_i^4] < \infty$, and are independent of the design points. F is twice continuously differentiable with positive density f such that $\sup_{y\in\mathbb{R}} |yf(y)| < \infty$ and $\sup_{y\in\mathbb{R}} |y^2 f'(y)| < \infty$.

The same kernel regression estimator $\hat{m}$ as in the previous section is used, compare (2.1). Further we estimate the variance function by a similar kernel approach, i. e.

$$(4.2) \qquad \hat{\sigma}^2(x) \;=\; \frac{1}{nh_n} \sum_{i=1}^{n} K\Big(\frac{X_i - x}{h_n}\Big)(Y_i - \hat{m}(x))^2 \frac{1}{\hat{f}_X(x)}$$

with kernel density estimator $\hat{f}_X$ defined in (2.2). Now the standardized residuals from the nonparametric fit are defined by

$$(4.3) \qquad \hat{\varepsilon}_i \;=\; \frac{Y_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)}, \quad i = 1, \ldots, n,$$

where $\hat{\sigma} = (\hat{\sigma}^2)^{1/2}$ is the scale function estimator. The residual-based empirical distribution function $\hat{F}_n$ is defined as in (1.2). Due to the estimation of the variance function the asymptotic distribution of $\hat{F}_n$ changes compared to Theorem 2.1. We state the weak convergence result in the next theorem.

Theorem 4.1 *Under the assumptions A2.3–4 and A4.1–2 the process* $\sqrt{n}(\hat{F}_n(y) - F(y))$, $y \in I\!R$, *converges weakly to a centered Gaussian process* G *with covariance*

$$\begin{aligned}\operatorname{Cov}(G(y), G(z)) \;=\; & F(y \wedge z) - F(y)F(z) + f(y)f(z) + f(y)U(z) + f(z)U(y) \\ & + \frac{1}{2} y f(y) V(z) + \frac{1}{2} z f(z) V(y) + \frac{1}{2} f(y) f(z) E[\varepsilon_1^3](y+z) \\ & + \frac{1}{4} y f(y) z f(z) \operatorname{Var}(\varepsilon_1^2),\end{aligned}$$

where $U(y) = E[\varepsilon_1 I\{\varepsilon_1 \leq y\}]$ *and* $V(y) = E[(\varepsilon_1^2 - 1) I\{\varepsilon_1 \leq y\}]$.

Compare Akritas and Van Keilegom (2001) and Neumeyer (2006) for the validity of Theorem 4.1. Müller et al. (2004c) only considered the homoscedastic model (1.1) but the results are easily generalized to the heteroscedastic model (4.1). In the same way as Theorem 2.2 is deduced from Theorem 2.1, Theorem 4.2 below follows from Theorem 4.1. To this end let model (4.1) be valid and let $\overline{F}_n$ be defined as in (1.4), but based on residuals $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$ defined in (4.3).

Theorem 4.2 *Under the assumptions A2.3–6 and A4.1–2 the process* $\sqrt{n}(\overline{F}_n(y) - F(y))$, $y \in I\!R$, *converges weakly to a centered Gaussian process* G *with the same covariance as defined in Theorem 4.1.*

The bootstrap in the heteroscedastic model is slightly different to take into account heteroscedasticity. The standardised residuals are defined as

$$\tilde{\varepsilon}_i \;=\; \frac{\hat{\varepsilon}_i - \frac{1}{n}\sum_{j=1}^n \hat{\varepsilon}_j}{\sqrt{\frac{1}{n}\sum_{l=1}^n (\hat{\varepsilon}_l - \frac{1}{n}\sum_{j=1}^n \hat{\varepsilon}_j)^2}},$$

and $\tilde{\varepsilon}_1^*, \ldots, \tilde{\varepsilon}_n^*$ are drawn randomly with replacement from these residuals. Bootstrap observations are then defined as

$$\tilde{Y}_i^* \;=\; \hat{m}(X_i) + \hat{\sigma}(X_i)\tilde{\varepsilon}_i^*$$

with $\hat{m}$ and $\hat{\sigma}^2$ from (2.1) and (4.2), respectively. Analogous to (2.1) and (4.2) we obtain regression and variance function estimators $\tilde{m}^*$ and $\tilde{\sigma}^{2*}$, respectively, from the new bootstrap sample $\{(X_1,\tilde{Y}_1^*),\ldots,(X_n,\tilde{Y}_n^*)\}$. Residuals estimated from the bootstrap sample are then defined as

$$\widehat{\tilde{\varepsilon}}_i^* = \frac{\tilde{Y}_i^* - \tilde{m}^*(X_i)}{\tilde{\sigma}^*(X_i)}.$$

Now let $\overline{F}_n^*$ denote the smooth empirical distribution function of bootstrap residuals $\widehat{\tilde{\varepsilon}}_1^*,\ldots,\widehat{\tilde{\varepsilon}}_n^*$ analogous to (3.4) and let $\widehat{\tilde{F}}_n$ denote the empirical distribution function of $\tilde{\varepsilon}_1,\ldots,\tilde{\varepsilon}_n$. To obtain weak convergence of the process $\sqrt{n}(\overline{F}_n^* - \widehat{\tilde{F}}_n)$, conditional on the sample $\mathcal{Y}_n$ in probability, we need an additional assumption.

<u>A4.3</u> Let $E[|\varepsilon_1|^{\max(2(\lambda-1),\lambda+1,6)}] < \infty$, where λ is defined in assumption A3.1. For the bandwidths we further assume that $b_n^{2+4/(\lambda+1)}/h_n^{1+\delta} = o(1)$.

The asymptotic result, from which asymptotic validity of the bootstrap procedure follows, is stated in the next theorem. The proof is deferred to the appendix.

Theorem 4.3 *Assume model (4.1) is valid under assumptions A2.3–4, A3.1–3 and A4.1–3. Then, conditional on the sample $\mathcal{Y}_n$, the process $\sqrt{n}(\overline{F}_n^*(y)-\widehat{\tilde{F}}_n(y))$, $y \in \mathbb{R}$, converges weakly to a centered Gaussian process G with covariance* $\mathrm{Cov}(G(y),G(z))$ *defined in Theorem 4.2, in probability.*

5 Applications to model testing

Assume we want to test a null hypothesis H_0 against an alternative H_1. Given is a test statistic T_n that is a continuous functional evaluated in a process R_n, i. e. $T_n = \psi(R_n)$, where R_n is an empirical process of residuals. Suppose that under the null hypothesis the weak convergence of the process R_n to a process R with continuous sample paths is shown. From the Continuous Mapping Theorem it follows that under H_0, T_n converges in distribution to $T = \psi(R)$. Further, T_n is chosen such that T_n converges to infinity in probability under the alternative H_1. The most important examples are Kolmogorov–Smirnov statistics $T_n = \sup_y |R_n(y)|$ or Cramér–von Mises functionals $T_n = \int R_n^2(y)\,dH(y)$. If t_n is the realisation of the test statistic T_n based on the sample $\mathcal{Y}_n$, then a level α–test is obtained by rejecting H_0 whenever $t_n > c_{1-\alpha}$, where $P_{H_0}(T_n > c_{1-\alpha}) = \alpha$. The critical value $c_{1-\alpha}$ could be estimated from the asymptotic distribution, i. e. the distribution of T. In the context considered here, however, the distribution of T depends on unknown features of the data-generating process that have to be estimated nonparametrically. Therefore, bootstrap procedures can lead to more accurate critical values. The bootstrap procedure

is based on a bootstrap sample $\mathcal{Y}_n^*$ built in a way such that it fulfils the null hypothesis. We consider the process R_n^* that is built from $\mathcal{Y}_n^*$ in the same way as R_n from $\mathcal{Y}_n$. For the bootstrap version R_n^* we show that both under the null hypothesis and under any fixed alternative it converges, conditional on the sample $\mathcal{Y}_n$, to a process R^*, in probability, whose distribution under H_0 coincides with the distribution of R. From the bootstrap process we build the test statistic $T_n^* = \psi(R_n^*)$. From the weak convergence of the bootstrap process and the Continuous Mapping Theorem we obtain the result that $P(T_n^* > c \mid \mathcal{Y}_n)$ converges to $P(T > c)$ in probability. The quantile $c_{1-\alpha}$ can now be approximated by the bootstrap quantile $c_{1-\alpha}^*$ defined by $P(T_n^* > c_{1-\alpha}^* \mid \mathcal{Y}_n) = \alpha$ and we obtain an asymptotic level α–test by rejecting the null hypothesis whenever $t_n > c_{1-\alpha}^*$. This test is also consistent because under the alternative T_n converges to infinity in probability, but T_n^* converges, conditional on the sample, in distribution to some random variable (by "consistent" in the following it is meant that a test has both asymtotic level α and power converging to one). The value $c_{1-\alpha}^*$ is usually estimated by $T^*_{n,(\lfloor B(1-\alpha)\rfloor)}$, where $T^*_{n,(b)}$ denotes the bth order statistic of B replications $T^*_{n,1}, \dots, T^*_{n,B}$ of bootstrap experiments all based on the same sample $\mathcal{Y}_n$.
In the following applications of the methods presented in sections 2–4 we concentrate primarily on the heteroscedastic regression model; homoscedastic models can be considered analogously.

5.1 Testing for a symmetric error distribution

Testing for symmetry of the conditional distribution around its mean is an important problem in nonparametric regression, see Neumeyer and Dette (2004) for motivation and references. In this section we consider a bootstrap procedure for testing symmetry of the error distribution, i. e. testing the hypotheses

$$H_0 : F(y) = 1 - F(-y) \text{ for all } y \in \mathbb{R} \quad vs \quad H_1 : F(y) \neq 1 - F(-y) \text{ for some } y \in \mathbb{R}$$

in the heteroscedastic regression model (4.1). Neumeyer and Dette (2004) investigated the estimated empirical symmetry process of residuals $\hat{\varepsilon}_i$ defined in (4.3), given by

$$\hat{S}_n(y) \;=\; \hat{F}_n(y) - (1 - \hat{F}_n(-y-)) \;=\; \frac{1}{n}\sum_{i=1}^n \Big(I\{\hat{\varepsilon}_i \le y\} - I\{-\hat{\varepsilon}_i \le y\}\Big). \tag{5.1}$$

Under assumptions A2.3–4 and A4.1–2 under the null hypothesis of symmetry they showed weak convergence of the process $\sqrt{n}\hat{S}_n(y)$, $y \in \mathbb{R}$, to a centered Gaussian process G with covariance structure

$$\operatorname{Cov}(G(y), G(z)) \;=\; 2F(-(|y| \vee |z|)) + 4f(y)f(z) + 4f(y)U(z) + 4f(z)U(y), \tag{5.2}$$

where U is defined in Theorem 2.1. They further proposed a symmetric wild bootstrap version of this process and showed consistency of corresponding Kolmogorov-Smirnov and Cramér-von Mises tests

$$\sqrt{n}\sup_{t\in I\!R}|\hat{S}_n(t)| \quad \text{and} \quad n\int \hat{S}_n^2(t)\, d\hat{H}_n(t), \tag{5.3}$$

where $\hat{H}_n$ is the empirical distribution function of $|\hat{\varepsilon}_1|,\ldots,|\hat{\varepsilon}_n|$. As an alternative Neumeyer (2006) explained the applicability of a smooth residual bootstrap version of the process S_n. Alternatively, the smooth version of the empirical symmetry process (5.1) can be used, i. e.

$$\overline{S}_n(y) \;=\; \overline{F}_n(y) - (1-\overline{F}_n(-y-)) \;=\; \frac{1}{n}\sum_{i=1}^{n}\Big[L\Big(\frac{y-\hat{\varepsilon}_i}{b_n}\Big) + L\Big(\frac{-y-\hat{\varepsilon}_i}{b_n}\Big) - 1\Big] \tag{5.4}$$

with $\overline{F}_n$ from (1.4), and under the assumptions of Theorem 2.2 the process $\sqrt{n}\overline{S}_n$ has the same limit distribution as $\sqrt{n}\hat{S}_n$ (compare the proof of Theorem 2.2). Then a classical residual bootstrap version of process (5.4) can be used, where the bootstrap errors $\tilde{\varepsilon}_1^*,\ldots,\tilde{\varepsilon}_n^*$ are obtained under the null hypothesis of symmetry, i. e. drawn randomly with replacement from $\{\tilde{\varepsilon}_1,\ldots,\tilde{\varepsilon}_n,-\tilde{\varepsilon}_1,\ldots,-\tilde{\varepsilon}_n\}$ and bootstrap observations are defined as $\tilde{Y}_i^* = \hat{m}(X_i) + \hat{\sigma}(X_i)\tilde{\varepsilon}_i^*$. Let the process $\overline{S}_n^*$ be built from the bootstrap sample in the same way as $\overline{S}_n$ is built from the original sample. With similar methods as in the proof of Theorem 3.2 we can prove the following result, from which consistency of the bootstrap test can be deduced.

Theorem 5.1 *Assume model (4.1) is valid under assumptions A2.3–4, A3.1–3 and A4.1–3. Then, conditional on the sample $\mathcal{Y}_n$, the process $\sqrt{n}\overline{S}_n^*(y)$, $y\in I\!R$, converges weakly to a centered Gaussian process G with covariance*

$$\operatorname{Cov}(G(y),G(z)) \;=\; 2F_s(-(|y|\vee|z|)) + 4f_s(y)f_s(z) + 4f_s(y)U_s(z) + 4f_s(z)U_s(y),$$

where $U_s(y)=\int_{-\infty}^{y}\varepsilon f_s(\varepsilon)\,d\varepsilon$, $f_s(y)=(f(y)+f(-y))/2$ and $F_s(y)=(F(y)+1-F(-y))/2$. Under the null hypothesis of a symmetric error distribution we have $f_s=f$ and $F_s=F$ and the covariance coincides with the covariance given in (5.2).

A detailed proof is omitted for the sake of brevity. In section 6 we compare the finite sample performances of the smooth residual bootstrap version of $\hat{S}_n$, the classical residual bootstrap version of $\overline{S}_n$ and a wild bootstrap version of $\hat{S}_n$ in a small simulation study.

5.2 Testing for equality of error distributions

In this section we consider two independent nonparametric regression models of type (4.1), they are

$$Y_{1j} \;=\; m_1(X_{1j}) + \sigma_1(X_{1j})\varepsilon_{1j}, \quad j=1,\ldots,n_1, \tag{5.5}$$

$$(5.6) \qquad Y_{2j} = m_2(X_{2j}) + \sigma_2(X_{2j})\varepsilon_{2j}, \quad j = 1, \ldots, n_2,$$

where the regression function m_i, variance function σ_i^2, design points X_{ij} with distribution f_{X_i} and errors ε_{ij} ($j = 1, \ldots, n_i$) with error distribution F_i of model $i \in \{1, 2\}$ fulfil assumptions A4.1–2 that were formulated for functions m, σ^2 and F. We consider tests for equality of the error distributions F_1 and F_2; the hypotheses we are interested in are

$$H_0 : F_1 = F_2 \quad vs \quad H_1 : F_1 \neq F_2.$$

Similar two-sample problems in parametric models were considered by Koul and Sen (1985) and Mora and Neumeyer (2005). In applications the comparison of two samples is often of interest when two standardised variables are to be compared or when it is to be tested whether two observable variables belong to the same location-scale family. Equality of the error distributions means that the standardised conditional distributions of observations Y_{ij}, given the covariates, are the same.

We consider Kolmogorov–Smirnov and Cramér–von Mises statistics based on the difference of empirical distribution functions $\hat{F}_{1n_1} - \hat{F}_{2n_2}$ of residuals in the two samples. Assume that the estimators in both individual sample fulfil assumptions as imposed for Theorem 4.1. Further assume for the sample sizes that

$$(5.7) \qquad \frac{n_i}{n_1 + n_2} \longrightarrow p_i > 0 \quad \text{for} \quad i = 1, 2.$$

Under the null hypothesis $H_0 : F_1 = F_2 = F$ weak convergence of the process $G_n(y) = (n_1 n_2/(n_1 + n_2))^{1/2}(\hat{F}_{1n_1}(y) - \hat{F}_{2n_2}(y))$, $y \in I\!R$, to a centered Gaussian process G with covariance

$$\begin{aligned}(5.8) \quad \operatorname{Cov}(G(y), G(z)) &= F(y \wedge z) - F(y)F(z) + f(y)f(z) + f(y)U(z) + f(z)U(y) \\ &\quad + \frac{1}{2} y f(y) V(z) + \frac{1}{2} z f(z) V(y) + \frac{1}{2} f(y) f(z) E[\varepsilon_1^3](y + z) \\ &\quad + \frac{1}{4} y f(y) z f(z) \operatorname{Var}(\varepsilon_1^2),\end{aligned}$$

where U and V are defined in Theorem 4.1, can be deduced easily from Akritas and Van Keilegom's (2001) results. With results by Neumeyer (2006) it can be shown that a smooth residual bootstrap version of this process leads to consistent testing procedures.

Alternatively, a test for equality of the error distributions can be based on the process $\overline{G}_n = (n_1 n_2/(n_1 + n_2))^{1/2}(\overline{F}_{1n_1} - \overline{F}_{2n_2})$ based on the smooth empirical distribution functions $\overline{F}_{in_i}$, $i = 1, 2$. Under assumptions as before the process $\overline{G}_n$ has the same limit distribution as G_n (compare the proof of Theorem 2.2).

As bootstrap version we then suggest using the classical residual bootstrap as in sections 3 and 4, where the bootstrap observations are built as

$$
\begin{aligned}
Y_{1j}^* &= \hat{m}_1(X_{1j}) + \hat{\sigma}_1(X_{1j})\varepsilon_{1j}^*, \quad j = 1, \ldots, n_1 \\
Y_{2j}^* &= \hat{m}_2(X_{2j}) + \hat{\sigma}_2(X_{2j})\varepsilon_{2j}^*, \quad j = 1, \ldots, n_2.
\end{aligned}
$$

under the null hypothesis. This is accomplished by drawing the bootstrap errors ε_{ij}^*, $j = 1, \ldots, n_i$, for the ith sample with replacement from all standardized residuals from both samples, $\{\tilde{\varepsilon}_{11}, \ldots, \tilde{\varepsilon}_{1n_1}, \tilde{\varepsilon}_{21}, \ldots, \tilde{\varepsilon}_{2n_2}\}$ $(i = 1, 2)$. Based on that bootstrap sample the process $\overline{G}_n^*$ is built and the validity of the following theorem can be shown similarly to the proof of Theorem 3.2.

Theorem 5.2 *Assume models (5.5), (5.6) under assumptions as in Theorem 4.3 for both samples and (5.7). Then the process $\overline{G}_n^*(y)$, $y \in \mathbb{R}$, conditional on $\mathcal{Y}_n = \{(X_{ij}, Y_{ij}) \mid j = 1, \ldots, n_i, i = 1, 2\}$, converges weakly in probability to a centered Gaussian process G with covariance*

$$
\begin{aligned}
\mathrm{Cov}(G(y), G(z)) = \; & F(y \wedge z) - F(y)F(z) + f(y)f(z) + f(y)U(z) + f(z)U(y) \\
& + \frac{1}{2} y f(y) V(z) + \frac{1}{2} z f(z) V(y) + \frac{1}{2} f(y) f(z) E[\varepsilon^3](y + z) \\
& + \frac{1}{4} y f(y) z f(z) \mathrm{Var}(\varepsilon^2),
\end{aligned}
$$

where $U(y) = E[\varepsilon I\{\varepsilon \leq y\}]$, $V(y) = E[(\varepsilon^2 - 1)I\{\varepsilon \leq y\}]$ and ε denotes a random variable with distribution $F = p_1 F_1 + p_2 F_2$ and density $f = p_1 f_1 + p_2 f_2$. When the null hypothesis is valid f and F coincide with the error density and distribution function in both samples and the covariance coincides with (5.8).

Both alternative bootstrap procedures for testing the equality of error distributions are compared in a simulation study in the next section.

5.3 Goodness-of-fit tests for regression and variance functions

Tests for a parametric form of the regression and variance function, respectively, based on empirical processes of estimated residuals are proposed by Van Keilegom et al. (2004) and Dette and Van Keilegom (2005). These authors show weak convergence of processes based on the difference of estimators for the error distribution on the one hand estimated nonparametrically and, on the other hand, estimated under the parametric alternative. In practise,

they propose to apply smooth residual bootstrap versions of their tests and Neumeyer (2006) shows consistency of these bootstrap procedures.
Alternatively, the empirical distribution function (1.2) in these procedures can be replaced by the smooth version (1.4). Consistency of a residual bootrap version of the obtained test can be proved analogously to the results of the present paper.
A similar argumentation applies to the test for equality of regression functions proposed by Pardo-Fernández et al. (2004).

6 Finite sample properties

For the sake of brevity we only present simulation results for the tests for symmetry as considered in section 5.1 and the test for equality of error distributions as in section 5.2.
For a test for symmetry of the error distribution we consider the following versions of test statistics (5.3):

(a) a smooth residual bootstrap version of the test statistics based on the non-smooth process $\sqrt{n}\hat{S}_n$ with smoothing parameter $a_n = n^{-1/4}$

(b) a symmetric wild bootstrap version of the test statistics based on the non-smooth process $\sqrt{n}\hat{S}_n$

(c) the classical residual bootstrap based on the smooth version of the residual-based empirical processes $\sqrt{n}\tilde{S}_n$ as in Theorem 5.1.

For (a) compare Neumeyer (2006) who showed that the performance of the smooth residual bootstrap does not depend very heavily on the choice of the smoothing parameter a_n. Compare Neumeyer and Dette (2004) for the symmetric wild bootstrap.
We estimate the regression function by the Nadaraya–Watson estimator with Gaussian kernel and simple rule-of-thumb bandwidth, where we set $h_n = (\hat{s}/n)^{0.3}$. The rate corresponds to the bandwidth conditions A2.4. Here, $\hat{s}$ denotes a simple estimator for the variance, see Rice (1984). The presented estimated quantiles are based on $B = 200$ bootstrap replications in each of 1000 simulation runs.
We consider the homoscedastic model (1.1) with uniform design and regression function $m(x) = 2x^2$. Under the null hypothesis we choose the error distributions to be

(i) standard normal

(ii) Student's t with three degrees of freedom

(iii) Student's t with five degrees of freedom

and results are displayed in Tables 1 and 2 for sample sizes $n \in \{30, 100\}$. The nominal levels $\alpha \in \{0.025, 0.05, 0.1\}$ are overestimated in some cases, especially for model (ii). We observe here that the quantile estimates obtained from the wild bootstrap have a tendency to larger values than the estimates obtained from the residual bootstraps.
Further we consider the following error distributions corresponding to fixed alternatives,

(iv) centered skew-normal with shape parameter 1 and skewness parameter 3

(v) centered skew-normal with shape parameter 1 and skewness parameter 7

(vi) double exponential with expectation zero and variance 0.5.

See Azzalini (1985) for the skew-normal distribution. The results are shown in Tables 3 and 4 for sample sizes $n \in \{30, 50, 100, 150\}$. We observe good and very similar power properties of all procedures, where the smooth residual procedure has slightly less power than the wild bootstrap procedure, but the difference is not very large. The Cramér–von Mises statistic can detect the considered alternatives better than the Kolmogorov–Smirnov statistic in most cases (an exception are small samples with $n = 30$). All results considered we see that (in this limited simulation study) the classical residual bootstrap procedure based on the smooth empirical distribution function has better power properties than the smooth residual bootstrap or wild bootstrap versions based on the classical empirical distribution function.
In the remainder of the section we present a small simulation study for the test for equality of error distributions as proposed in section 5.2. Here we concentrate on two models of the homoscedastic type (1.1) with uniform design. We consider (a) smooth residual bootstrap versions $T^*_{k,n} = \psi_k(G^*_n)$ of the Cramér–von Mises statistic ($k = 1$) and Kolmogorov–Smirnov statistic ($k = 2$) based on the non-smooth empirical processes

$$\begin{aligned} T_{1,n} &= \psi_1(G_n) = \frac{n_1 n_2}{(n_1+n_2)^2} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (\hat{F}_{1n_1}(\hat{\varepsilon}_{ij}) - \hat{F}_{2n_2}(\hat{\varepsilon}_{ij}))^2 \\ T_{2,n} &= \psi_2(G_n) = \sqrt{\frac{n_1 n_2}{N}} \sup_{y \in I\!R} |\hat{F}_{1n_1}(y) - \hat{F}_{2n_2}(y)| \end{aligned}$$

for smoothing parameter $a_n = n^{-1/4}$. Alternatively, we consider (b) classical residual bootstrap versions $\overline{T}^*_{k,n} = \psi_k(\overline{G}^*_n)$ of $\overline{T}_{k,n} = \psi_k(\overline{G}_n)$ ($k = 1, 2$) based on the smooth residual-based empirical distribution functions. The regression estimators $\hat{m}_i$ are the Nadaraya–Watson estimators with Gaussian kernels, and the bandwiths h_{in} are chosen according to the simple rule of thumb explained before within the ith sample, $i = 1, 2$. We estimated the critical values using $B = 200$ bootstrap replications in each of 1000 simulations. The regression functions are $m_1(x) = x$ in the first model and $m_2(x) = 1 - 2x$ in the second model. We consider levels $\alpha \in \{0.025, 0.05, 0.1\}$ and different combinations of sample sizes (n_1, n_2). First we observe how well the level is approximated in the following settings corresponding to the null hypothesis, where both error distributions are

(i) normal with expectation zero and variance one

(ii) Student's t with three degrees of freedom

(iii) centered skew-normal with scale parameter 1 and shape parameter 2.

We display values for the Cramér–von Mises statistic in Table 5 and for the Kolmogorov–Smirnov statistic in Table 6 obtained for sample sizes $(n_1, n_2) = (50, 50)$ and $(100, 100)$, results obtained for other considered sample sizes were similar. We observe values very slightly larger than the nominal levels for the Cramér–von Mises statistic in situation (a). The values obtained from the Kolmogorov–Smirnov statistic as well as for the Cramér–von Mises procedure in situation (b) the nominal levels are overestimated, only in case (ii) the approximations for the Cramér–von Mises statistic are accurate.
For simulating the power of the test we consider the same design distribution and regression functions as before with standard normally distributed errors in the first sample and the following error distributions in the second sample,

(iv) Student's t with four degrees of freedom

(v) centered skew-normal with scale parameter 1 and shape parameter 2

(vi) centered skew-normal with scale parameter 1 and shape parameter 7.

The results are displayed in Tables 7 and 8 for sample sizes $(n_1, n_2) = (30, 30)$, $(50, 50)$, $(75, 100)$, $(100, 100)$ and $(150, 150)$. The Kolmogorov–Smirnov statistic has less power than the Cramér–von Mises statistic in the situations (iv) [except for sample sizes $(50, 50)$] and (vi) for test procedure (a). For test (b) the Kolmogorov–Smirnov and Cramér–von Mises statistics have comparable power properties for the considered settings, where the latter procedure can detect alternative (iv) better. A comparison shows that the classical residual bootstrap version of the smooth empirical process has better power properties than the smooth residual bootstrap version of the non-smooth empirical process, but this is not a fair comparison as the former test is oversized.

A Proofs

A.1 Preliminaries

For the definitions

$$\begin{aligned} m_n(x) &= \frac{1}{h_n}\int K\Big(\frac{x-y}{h_n}\Big) m(y) f_X(y)\,dy\,\frac{1}{f_{X,n}(x)} \\ f_{X,n}(x) &= \frac{1}{h_n}\int K\Big(\frac{x-y}{h_n}\Big) f_X(y)\,dy \\ \tilde m_n(x) &= \frac{1}{h_n}\int K\Big(\frac{x-y}{h_n}\Big) m_n(y) f_X(y)\,dy\,\frac{1}{f_{X,n}(x)} \end{aligned} \tag{A.1}$$

note that m_n and $\tilde m_n$ are deterministic functions with

$$\sup_{x\in[0,1]} |m_n(x)-m(x)| = O(h_n) \quad \text{and} \quad \sup_{x\in[0,1]} |\tilde m_n(x)-m(x)| = O(h_n), \tag{A.2}$$

and

$$\begin{aligned} \sup_{x\in[0,1]} |\hat m(x)-m_n(x)| &= O_p\Big(\sqrt{\frac{\log(h_n^{-1})}{nh_n}}\Big) \\ \sup_{x\in[0,1]} |\hat m(x)-m(x)| &= o(1) \end{aligned} \tag{A.3}$$

almost surely. Further we have

$$\begin{aligned} |\hat m(X_i)-m(X_i)| &\le \sup_{x\in[0,1]} |\hat m(x)-m_n(x)| + \sup_{x\in[h_n,1-h_n]} |m_n(x)-m(x)| \\ &\quad + (I\{X_i<h_n\}+I\{X_i>1-h_n\}) \sup_{x\in[0,1]\setminus[h_n,1-h_n]} |m_n(x)-m(x)| \\ &\le t_n + C_1h_n^2 + C_2t_{ni}h_n \end{aligned} \tag{A.4}$$

for some constants C, C_1, C_2, some random variable $t_n = o(\alpha_n)$, for all sequences α_n of positive real numbers such that $nh_n\alpha_n \to \infty$ and $\log n/(nh_n\alpha_n^2) = o(1)$ [compare Neumeyer (2006), Proposition 6.2], and $t_{ni} = I\{X_i < h_n\} + I\{X_i > 1-h_n\}$ is non-negative with moments $E[t_{ni}^l] = O(h_n)$ for all $l > 0$.

Lemma A.1 *Under the assumptions of Theorem 3.2 we have*

$$\begin{aligned} \sup_{x\in[0,1]} |\tilde m^*(x)-\tilde m_n(x)| &= O_p\Big(\sqrt{\frac{\log h_n^{-1}}{nh_n}}\Big) \\ \sup_{x\in[0,1]} |\tilde m^{*\prime}(x)-\tilde m_n'(x)| &= o_p(1) \\ \sup_{x,t\in[0,1]} \frac{|\tilde m^{*\prime}(x)-\tilde m_n'(x)-\tilde m^{*\prime}(t)+\tilde m_n'(t)|}{|x-t|^{\delta/2}} &= o_p(1). \end{aligned}$$

Here δ denotes the constant from (2.3) and $\tilde m_n$ is defined in (A.1).

Proof. The proof follows along the lines of the proof of Lemma 6.5 by Neumeyer (2006) setting $a_n = 0$ and details are omitted for the sake of brevity. $\square$

A.2 Proof of Theorem 3.2

We take a detour via the smooth residual bootstrap applied for the non-smooth residual-based empirical distribution function considered by Neumeyer (2006). Note that bootstrap observations for this kind of bootstrap are defined as

$$Y_i^* \;=\; \hat{m}(X_i) + \varepsilon_i^* = \hat{m}(X_i) + \tilde{\varepsilon}_i^* + a_n Z_i, \tag{A.5}$$

where the $\tilde{\varepsilon}_i^*$ are randomly drawn with replacement from the centered residuals $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$ as defined in (3.1). The independent and identically distributed random variables Z_i with density k are independent from the sample $\mathcal{Y}_n$ and a_n denotes a sequence of positive bandwidths specified in a moment. Given the sample $\mathcal{Y}_n$, the $\varepsilon_i^* = \tilde{\varepsilon}_i^* + a_n Z_i$ are independent and identically distributed with density

$$\tilde{f}_n(y) \;=\; \frac{1}{na_n} \sum_{i=1}^{n} k\Big(\frac{y - \tilde{\varepsilon}_i}{a_n}\Big). \tag{A.6}$$

We choose k to be symmetric with compact support $[-1, 1]$ such that $\int k(z)z\,dz \;=\; 0$, $\int k(z)z^2\,dz \;<\; \infty$, and k has three derivatives inside the support, such that the first two derivatives are bounded and of bounded variation with $\int (k^{(\nu)}(z))^2\,dz < \infty$, $\nu = 1, 2$, and $k^{(3)} \equiv 0$ (for example a truncated polynomial kernel). The sequence of bandwidths is chosen as $a_n = b_n^{(\lambda+3)/(\lambda+1)}$ with b_n and λ from assumption A3.2. Then, a_n fulfils assumptions (3.5) in Remark 3.1 and also $na_n^4 \to 0$. Now we define auxiliary bootstrap residuals as

$$\breve{\varepsilon}_i^* \;=\; Y_i^* - \tilde{m}^*(X_i) \;=\; \tilde{Y}_i^* - \tilde{m}^*(X_i) + a_n Z_i \;=\; \hat{\tilde{\varepsilon}}_i^* + a_n Z_i \tag{A.7}$$

with Y_i^* from (A.5), $\tilde{Y}_i^*$ from (3.2), and $\tilde{m}^*$, $\hat{\tilde{\varepsilon}}_i^*$ from (3.3). Define the process

$$\sqrt{n}\Big(\frac{1}{n}\sum_{j=1}^{n} I\{\breve{\varepsilon}_i^* \le y\} - \tilde{F}_n(y)\Big), \tag{A.8}$$

where $\tilde{F}_n$ denotes the distribution function according to the residual density estimator $\tilde{f}_n$ defined in (A.6). The auxiliary bootstrap residuals $\breve{\varepsilon}_i^*$ from (A.7) have the advantage that they are built from bootstrap observations with errors ε_i^*, which have a smooth distribution. The only difference from (A.8) to the process considered in Theorem 3.2 by Neumeyer (2006) is the different regression estimator ($\tilde{m}^*$ based on $\{(X_i, \tilde{Y}_i^*) \mid i = 1, \ldots, n\}$ instead of $\{(X_i, Y_i^*) \mid i = 1, \ldots, n\})$. Following the lines of the proof of of the cited theorem and applying Lemma A.1, it is relatively straightforward to see that the assertion of Th. 3.2 in the aforementioned reference (which coincides with the assertion of Theorem 3.2 in the present paper) is also valid for the empirical process (A.8). It follows with very similar arguments as in the proof of Theorem 2.2 that the smooth version of this empirical process,

$$\text{(A.9)} \qquad \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} L\Big(\frac{y-\breve{\varepsilon}_i^*}{b_n}\Big) - \tilde{F}_n(y)\Big)$$

also joins the same limit distribution as given in Theorem 3.2. We will prove in the following that the process (A.9) is equivalent to the one we are interested in, i. e.

$$\text{(A.10)} \qquad \sqrt{n}(\overline{F}_n^*(y) - \hat{\tilde{F}}_n(y)) \;=\; \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} L\Big(\frac{y-\hat{\tilde{\varepsilon}}_i^*}{b_n}\Big) - \hat{\tilde{F}}_n(y)\Big)$$

in the sense of conditional weak convergence in probability. To this end, firstly note that,

$$\tilde{F}_n(y) \;=\; \int \hat{\tilde{F}}_n(y - a_n u)k(u)\,du$$

and therefore,

$$\begin{aligned}
&\sqrt{n}\sup_{y\in\mathbb{R}}\Big|\hat{\tilde{F}}_n(y) - \tilde{F}_n(y)\Big| \\
&\le \sup_{y\in\mathbb{R}}\Big|\int\Big(\sqrt{n}(\hat{\tilde{F}}_n(y-a_nu) - F(y-a_nu)) - \sqrt{n}(\hat{\tilde{F}}_n(y) - F(y))\Big)k(u)\,du\Big| \\
&\quad + \sqrt{n}\sup_{y\in\mathbb{R}}\Big|\int (F(y-a_nu) - F(y))k(u)\,du\Big| \\
&\le \sup_{\substack{y,z\in\mathbb{R}:\\ |y-z|\le a_n}} |\sqrt{n}(\hat{\tilde{F}}_n(y) - F(y)) - \sqrt{n}(\hat{\tilde{F}}_n(z) - F(z))| + \sqrt{n}\sup_{y\in\mathbb{R}}|f'(y)|\int (a_nu)^2k(u)\,du \\
&= o(1) + O(\sqrt{n}a_n^2) \;=\; o(1),
\end{aligned}$$

where we have made use of the fact that k has support $[-1,1]$, the asymptotic equicontinuity of the process $\sqrt{n}(\hat{\tilde{F}}_n - F)$ (compare Remark 3.3), the vanishing first moments of the kernel function, a Taylor expansion of F, and the bandwidth conditions. We obtain the result that the process (A.9) is equivalent to

$$\text{(A.11)} \qquad \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} L\Big(\frac{y-\breve{\varepsilon}_i^*}{b_n}\Big) - \hat{\tilde{F}}_n(y)\Big)$$

in terms of conditional weak convergence in probability.

Let $\Delta(1),\ldots,\Delta(n)$ be independent random variables with Laplace distribution on $\{1,\ldots,n\}$ corresponding to the drawing of $\tilde\varepsilon_1^*,\ldots,\tilde\varepsilon_n^*$ with replacement from $\tilde\varepsilon_1,\ldots,\tilde\varepsilon_n$, such that

$$\tilde\varepsilon_i^* \;=\; \sum_{l=1}^n I\{\Delta(i)=l\}\tilde\varepsilon_l \;=\; \sum_{l=1}^n I\{\Delta(i)=l\}\hat\varepsilon_l - \frac{1}{n}\sum_{j=1}^n \hat\varepsilon_j .$$

From (3.3) and (A.7) we have

$$\begin{aligned}
\breve\varepsilon_i^* &= \tilde\varepsilon_i^* + \hat m(X_i) - \tilde m^*(X_i) + a_n Z_i \qquad \text{(A.12)}\\
&= \eta_{ni} - \kappa_n(X_i) - K_{ni} + r_{ni} + a_n Z_i\\
\widehat{\tilde\varepsilon}_i^* &= \eta_{ni} - \kappa_n(X_i) - K_{ni} + r_{ni} \qquad \text{(A.13)}
\end{aligned}$$

where

$$\eta_{ni} = \sum_{l=1}^n I\{\Delta(i)=l\}\varepsilon_l$$

is drawn randomly with replacement from $\{\varepsilon_1,\ldots,\varepsilon_n\}$,

$$K_{ni} = \sum_{l=1}^n I\{\Delta(i)=l\}\kappa_{nl}$$

is drawn randomly with replacement from $\kappa_{nl} = m_n(X_l) - m(X_l)$, $l=1,\ldots,n$, the deterministic sequence of functions $\kappa_n = \tilde m_n - m_n$ converges uniformly to zero by (A.2), and

$$\begin{aligned}
r_{ni} \;&=\; \hat m(X_i) - m_n(X_i) - (\tilde m^*(X_i) - \tilde m_n(X_i)) - \frac{1}{n}\sum_{l=1}^n \varepsilon_l \qquad \text{(A.14)}\\
&\quad + \sum_{l=1}^n I\{\Delta(i)=l\}(m_n(X_l) - \hat m(X_l)) + \frac{1}{n}\sum_{l=1}^n (\hat m(X_l) - m(X_l))
\end{aligned}$$

is uniformly in i of order $O_p((\log h_n^{-1}/(nh_n))^{1/2})$ by Lemma A.1, (A.3) and (A.4), because $\sum_{l=1}^n I(\Delta(i)=l) = 1$ and $h_n^2 = O((\log h_n^{-1}/(nh_n))^{1/2})$.

To prove that the difference between (A.10) and (A.11) is small we apply Taylor expansions of both $L((y-\breve\varepsilon_i^*)/b_n)$ and $L((y-\widehat{\tilde\varepsilon}_i^*)/b_n)$ around $L((y-\eta_{ni}+\kappa_n(X_i)+K_{ni})/b_n)$ to obtain from (A.12), (A.13) that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \Big[L\Big(\frac{y-\breve\varepsilon_i^*}{b_n}\Big) - L\Big(\frac{y-\widehat{\tilde\varepsilon}_i^*}{b_n}\Big)\Big]\\
=& \sum_{\mu=1}^{\lambda}\frac{1}{n}\sum_{i=1}^n \ell^{(\mu-1)}\Big(\frac{y-\eta_{ni}+\kappa_n(X_i)+K_{ni}}{b_n}\Big)\frac{1}{b_n^\mu \mu!}[(-r_{ni}-a_nZ_i)^\mu - (-r_{ni})^\mu]\\
&+ \frac{1}{n}\sum_{i=1}^n \ell^{(\lambda)}(\xi_{y,n,i})\frac{1}{b_n^{\lambda+1}(\lambda+1)!}(-r_{ni}-a_nZ_i)^{\lambda+1}
\end{aligned}$$

$$
\begin{aligned}
& - \frac{1}{n}\sum_{i=1}^{n}\ell^{(\lambda)}(\xi'_{y,n,i})\frac{1}{b_n^{\lambda+1}(\lambda+1)!}(-r_{ni})^{\lambda+1} \\
\text{(A.15)} \ = \ & \sum_{\mu=1}^{\lambda}\sum_{\nu=0}^{\mu-1}\binom{\mu}{\nu}\frac{(-1)^\mu}{\mu!}\frac{1}{n}\sum_{i=1}^{n}\ell^{(\mu-1)}\Big(\frac{y-\eta_{ni}+\kappa_n(X_i)+K_{ni}}{b_n}\Big)\frac{r_{ni}^{\nu}(a_nZ_i)^{\mu-\nu}}{b_n^{\mu}} + r_n(y)
\end{aligned}
$$

where $r_n \equiv 0$ when $\ell^{(\lambda)} \equiv 0$ (compare Remark 3.1) and otherwise,

$$
\sup_{y\in I\!R}|r_n(y)| \ \le\ O(\frac{1}{b_n^{\lambda+1}})\frac{1}{n}\sum_{i=1}^{n}(|r_{ni}|^{\lambda+1} + a_n^{\lambda+1}|Z_i|^{\lambda+1})
$$

because $\ell^{(\lambda)}$ is bounded. Because the random variables Z_i are uniformly bounded (k has compact support) we obtain the rate

$$
\sup_{y\in I\!R}|r_n(y)| \ = \ O_p\Big(\frac{(\log h_n^{-1})^{(\lambda+1)/2}}{b_n^{\lambda+1}(nh_n)^{(\lambda+1)/2}}\Big) + O_p\Big(\frac{a_n^{\lambda+1}}{b_n^{\lambda+1}}\Big) \ = \ o_p(\frac{1}{\sqrt{n}})
$$

by (A.14), the bandwidth conditions A3.2, and the definition of $a_n = b_n^{\frac{\lambda+3}{\lambda+1}}$. The rest of the proof follows from Proposition A.3 below that shows that the terms (A.15) are of order $o_p(1/\sqrt{n})$. Hence, the process (A.10) has the same limit distribution as (A.9) (conditional on $\mathcal{Y}_n$ in probability). □

A.3 Auxiliary results

Proposition A.2 *Under the assumptions of Theorem 3.2, for $\mu = 1,\dots,\lambda$, it holds that*

$$
\sup_{y\in I\!R}\frac{1}{n}\sum_{i=1}^{n}\Big|\ell^{(\mu-1)}\Big(\frac{y-\eta_{ni}+\kappa_n(X_i)+K_{ni}}{b_n}\Big)\Big| = O_p(b_n).
$$

Proof. Note that κ_n is deterministic and converges to zero uniformly. Further we have by (A.2)

$$
\max_{i=1,\dots,n}|K_{ni}| \le \sup_{x\in[0,1]}|m_n(x)-m(x)| = o(1)
$$

and, hence, we assume $|\kappa_n(\cdot)| \le 1$ and $|K_{ni}| \le 1$, $i = 1,\dots,n$. We have $\max_{i=1,\dots,n}|\eta_{ni}| \le \max_{i=1,\dots,n}|\varepsilon_i| = o_p(n^{1/4})$. Let c denote a constant such that

$$
P\Big(b_n + 2 + \max_{i=1,\dots,n}|\eta_{ni}| > cn^{1/4}\Big) = o(1)
$$

and assume in the following that $b_n + 2 + \max_{i=1,\dots,n}|\eta_{ni}| \le cn^{1/4}$. Also, we assume for the proof that ℓ has support $[-1,1]$. Then we have $\ell^{(\mu-1)}((y-\eta_{ni}+\kappa_n(X_i)+K_{ni})/b_n) = 0$ ($\mu = 1,\dots,\lambda$) for all i whenever $|y| > cn^{1/4}$ and therefore it suffices to consider the supremum $\sup_{|y|\le cn^{1/4}}$ instead of $\sup_{y\in I\!R}$ in the assertion of the proposition.

First we consider the expectation conditional on the σ–field $\mathcal{E}_n$ generated by the sample $\mathcal{Y}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$. According to Proposition A.4 (ii) below we have for the supremum of the conditional expectations,

$$\begin{aligned}
&\sup_{|y| \le cn^{1/4}} \frac{1}{nb_n} \sum_{i=1}^n E\Big[\Big|\ell^{(\mu-1)}\Big(\frac{y - \eta_{ni} + \kappa_n(X_i) + K_{ni}}{b_n}\Big)\Big| \;\Big|\; \mathcal{E}_n\Big] \\
= &\sup_{|y| \le cn^{1/4}} \frac{1}{nb_n} \sum_{i=1}^n \frac{1}{n} \sum_{j=1}^n \Big|\ell^{(\mu-1)}\Big(\frac{y - \varepsilon_j + \kappa_n(X_i) + \kappa_{nj}}{b_n}\Big)\Big| \;=\; O_p(1)
\end{aligned}$$

[note that $\sup_{x \in [0,1]} |\kappa_n(x)| = o(1)$ and $\max_{j=1,\dots,n} |\kappa_{nj}| = o(1)$]. Further because of Proposition A.4 (iii) we can assume that

$$\sup_{|y| \le cn^{1/4}} \frac{1}{nb_n} \sum_{i=1}^n \frac{1}{n} \sum_{j=1}^n \Big[\ell^{(\mu-1)}\Big(\frac{y - \varepsilon_j + \kappa_n(X_i) + \kappa_{nj}}{b_n}\Big)\Big]^2 \;\le\; A \tag{A.16}$$

for some constant A. For each fixed y and $\epsilon > 0$ we apply Bernstein's inequality to obtain

$$P\Big(\Big|\frac{1}{n}\sum_{i=1}^n \xi_i\Big| > \epsilon \;\Big|\; \mathcal{E}_n\Big) \;\le\; 2\exp(-\log n\, w_n)$$

where, conditional on $\mathcal{E}_n$, the random variables

$$\xi_i = \frac{1}{b_n}\Big|\ell^{(\mu-1)}\Big(\frac{y - \eta_{ni} + \kappa_n(X_i) + K_{ni}}{b_n}\Big)\Big| - E\Big[\frac{1}{b_n}\Big|\ell^{(\mu-1)}\Big(\frac{y - \eta_{ni} + \kappa_n(X_i) + K_{ni}}{b_n}\Big)\Big| \;\Big|\; \mathcal{E}_n\Big]$$

are independent with expectation zero, $|\xi_i| \le M/b_n = M_n$ for some constant M, and

$$\sum_{i=1}^n \mathrm{Var}(\xi_i \mid \mathcal{E}_n) \le \frac{An}{b_n} = V_n$$

by (A.16). Further,

$$w_n^{-1} \;=\; 2\Big(V_n + \frac{M_n \epsilon n}{3}\Big)\frac{\log n}{(n\epsilon)^2} \;=\; \frac{\log n}{nb_n}\Big(\frac{2A}{\epsilon^2} + \frac{2M}{3\epsilon}\Big)$$

is not random, independent of y and converges to zero by assumption A3.2.
To obtain the assertion uniformly in $|y| \le cn^{1/4}$ we cut the interval $[-cn^{1/4}, cn^{1/4}]$ into $\rho = O(n^{1+1/4})$ subintervals with centres y_ϱ and length n^{-1}. Then,

$$\begin{aligned}
&\sup_{|y| \le cn^{1/4}} \Big|\frac{1}{nb_n}\sum_{i=1}^n \Big\{\Big|\ell^{(\mu-1)}\Big(\frac{y - \eta_{ni} + \kappa_n(X_i) + K_{ni}}{b_n}\Big)\Big| \\
&\qquad - E\Big[\frac{1}{b_n}\Big|\ell^{(\mu-1)}\Big(\frac{y - \eta_{ni} + \kappa_n(X_i) + K_{ni}}{b_n}\Big)\Big| \;\Big|\; \mathcal{E}_n\Big]\Big\}\Big| \\
\text{(A.17)}\qquad &\le \max_{\varrho = 1,\dots,\rho} \Big|\frac{1}{nb_n}\sum_{i=1}^n \Big\{\Big|\ell^{(\mu-1)}\Big(\frac{y_\varrho - \eta_{ni} + \kappa_n(X_i) + K_{ni}}{b_n}\Big)\Big|
\end{aligned}$$

$$- E\Big[\frac{1}{b_n}\Big|\ell^{(\mu-1)}\Big(\frac{y_\varrho - \eta_{ni} + \kappa_n(X_i) + K_{ni}}{b_n}\Big)\Big| \,\Big|\, \mathcal{E}_n\Big]\Big\}\Big|$$

$$\text{(A.18)} \qquad + \sup_{|y-z|\le n^{-1}} \Big|\frac{1}{nb_n}\sum_{i=1}^n \Big\{\Big|\ell^{(\mu-1)}\Big(\frac{y - \eta_{ni} + \kappa_n(X_i) + K_{ni}}{b_n}\Big)\Big| - \Big|\ell^{(\mu-1)}\Big(\frac{z - \eta_{ni} + \kappa_n(X_i) + K_{ni}}{b_n}\Big)\Big|\Big\}\Big|$$

$$\text{(A.19)} \qquad + \sup_{|y-z|\le n^{-1}} \Big|\frac{1}{nb_n}\sum_{i=1}^n E\Big[\frac{1}{b_n}\Big|\ell^{(\mu-1)}\Big(\frac{y - \eta_{ni} + \kappa_n(X_i) + K_{ni}}{b_n}\Big)\Big| \,\Big|\, \mathcal{E}_n\Big] - E\Big[\frac{1}{b_n}\Big|\ell^{(\mu-1)}\Big(\frac{z - \eta_{ni} + \kappa_n(X_i) + K_{ni}}{b_n}\Big)\Big| \,\Big|\, \mathcal{E}_n\Big]\Big|.$$

Now for (A.17) we have from the above considerations,

$$\begin{aligned} &P\Big(\max_{\varrho=1,\dots,\rho}\Big|\frac{1}{nb_n}\sum_{i=1}^n \Big\{\Big|\ell^{(\mu-1)}\Big(\frac{y_\varrho - \eta_{ni} + \kappa_n(X_i) + K_{ni}}{b_n}\Big)\Big| \\ &\qquad - E\Big[\frac{1}{b_n}\Big|\ell^{(\mu-1)}\Big(\frac{y_\varrho - \eta_{ni} + \kappa_n(X_i) + K_{ni}}{b_n}\Big)\Big| \,\Big|\, \mathcal{E}_n\Big]\Big\}\Big| > \epsilon\Big) \\ &\le O(n^{1+1/4}\exp(-\log n\, w_n)) \;=\; o(1) \end{aligned}$$

for all $\epsilon > 0$, and further for (A.18),

$$\begin{aligned} &\sup_{|y-z|\le n^{-1}} \Big|\frac{1}{nb_n}\sum_{i=1}^n \Big\{\Big|\ell^{(\mu-1)}\Big(\frac{y - \eta_{ni} + \kappa_n(X_i) + K_{ni}}{b_n}\Big)\Big| - \Big|\ell^{(\mu-1)}\Big(\frac{z - \eta_{ni} + \kappa_n(X_i) + K_{ni}}{b_n}\Big)\Big|\Big\}\Big| \\ &\le O(\frac{1}{b_n^2 n}) \;=\; o(1) \end{aligned}$$

by $||a|-|b|| \le |a-b|$, a Taylor expansion, the boundedness of $\ell^{(\mu)}$, and bandwidth assumption A3.2. A similar inequality holds for (A.19). $\square$

Proposition A.3 *Under the assumptions of Theorem 3.2, it holds that*

$$\sup_{y\in I\!R}\Big|\frac{1}{n}\sum_{i=1}^n \ell^{(\mu-1)}\Big(\frac{y - \eta_{ni} + \kappa_n(X_i) + K_{ni}}{b_n}\Big)\frac{r_{ni}^\nu (a_n Z_i)^{\mu-\nu}}{b_n^\mu}\Big| = o_p(\frac{1}{\sqrt{n}})$$

for $\mu = 1, \dots, \lambda$ and $\nu = 0, \dots, \mu - 1$, where $a_n = b_n^{(\lambda+3)/(\lambda+1)}$.

Proof. With the same considerations as in the proof of Proposition A.2 it suffices to consider the supremum $\sup_{|y|\le cn^{1/4}}$ instead of $\sup_{y\in I\!R}$ in the assertion of the Proposition.

(a) First, we consider the special case $\nu = 0$ and $\mu \ge 2$ even. Conditional on the σ–field $\mathcal{E}_n$ generated by $\mathcal{Y}_n$ we will apply Bernstein's inequality as before for each fixed y and then obtain the uniform result in the same way as in Proposition A.2. First we have to consider the expectation, conditional on $\mathcal{E}_n$, that is

$$E[Z_i^\mu]\frac{a_n^\mu}{b_n^\mu}\frac{1}{n}\sum_{i=1}^n \frac{1}{n}\sum_{j=1}^n \ell^{(\mu-1)}\Big(\frac{y-\varepsilon_j+\kappa_n(X_i)+\kappa_{nj}}{b_n}\Big) \;=\; O_p\Big(\frac{a_n^\mu}{b_n^{\mu-2}}\Big),$$

where the rate is valid uniformly with respect to y by Proposition A.4 (i) in the appendix. The rate of convergence we obtain here is $b_n^{2+2\mu/(\lambda+1)} = O(b_n^2) = o(1/\sqrt{n})$ by bandwidth condition A3.2. Further, let A denote a constant such that $|\ell^{(\mu-1)}|$ is bounded by A and $|Z_i| \le A$ for all i. We obtain for any fixed y,

$$P\Big(\Big|\frac{1}{n}\sum_{i=1}^n \xi_i\Big| > \epsilon \;\Big|\; \mathcal{E}_n\Big) \;\le\; \exp(-\log n\, w_n),$$

where the random variables

$$\begin{aligned}\xi_i \;=\; & \sqrt{n}a_n^\mu b_n^{-\mu}\Big\{Z_i^\mu \ell^{(\mu-1)}\Big(\frac{y-\eta_{ni}+\kappa_n(X_i)+K_{ni}}{b_n}\Big)\\ & - E\Big[Z_i^\mu \ell^{(\mu-1)}\Big(\frac{y-\eta_{ni}+\kappa_n(X_i)+K_{ni}}{b_n}\Big)\;\Big|\; \mathcal{E}_n\Big]\Big\}\end{aligned}$$

are bounded by $M_n = M\sqrt{n}a_n^\mu b_n^{-\mu}$ for some constant M and fulfil

$$\sum_{i=1}^n \mathrm{Var}(\xi_i \mid \mathcal{E}_n) \;\le\; \frac{A^{2\mu+2}n^2a_n^{2\mu}}{b_n^{2\mu}} \;=\; V_n.$$

We have

$$w_n^{-1} \;=\; 2(V_n + \frac{M_n\epsilon n}{3})\frac{\log n}{(n\epsilon)^2} \;=\; O\Big(\frac{\log n\, a_n^{2\mu}}{b_n^{2\mu}}\Big) + O\Big(\frac{\log n\, a_n^\mu}{\sqrt{n}b_n^\mu}\Big) \;=\; o(1).$$

In the usual manner we proceed to obtain the assertion uniformly in $|y| \le cn^{1/4}$ (compare the proof of Proposition A.2) and use the fact that $\ell^{(\mu-1)}$ is for all $\mu = 1,\ldots,\lambda$ continuously differentiable).

(b) Now, let $\nu = 0$ and μ odd. Here, we have $E[Z_i^\mu] = 0$ because we have chosen the kernel k to be symmetric. The proof is analogous to (a), but easier by utilising the vanishing expectation.

(c) Let $\nu \ge 1$ and $\mu \ge \nu+1$ such that $\mu-\nu$ is even. Because the random variables Z_i are bounded we simply bound the considered term by

$$C\max_{i=1,\ldots,n}|r_{ni}|^\nu \sup_{y\in\mathbb{R}}\frac{1}{n}\sum_{i=1}^n\Big|\ell^{(\mu-1)}\Big(\frac{y-\eta_{ni}+\kappa_n(X_i)+K_{ni}}{b_n}\Big)\Big|\frac{a_n^{\mu-\nu}}{b_n^\mu}$$

for some constant C and obtain from Proposition A.2 and (A.14) the rate

$$O_p\Big(\frac{a_n^{\mu-\nu}}{b_n^{\mu-1}}\Big(\frac{\log h_n^{-1}}{nh_n}\Big)^{\nu/2}\Big) \;=\; o_p(\frac{1}{\sqrt{n}}).$$

The last equality holds due to bandwidth conditions: for $\nu = 1$ we have to consider $\mu = 3, 5, \ldots$. The most restrictive constraint is for $\mu = 3$. We then have

$$O_p\Big(\frac{a_n^2}{b_n^2}\Big(\frac{\log h_n^{-1}}{nh_n}\Big)^{1/2}\Big) \;=\; O_p\Big(b_n^{4/(\lambda+1)}\Big(\frac{\log h_n^{-1}}{nh_n}\Big)^{1/2}\Big) \;=\; o(\frac{1}{\sqrt{n}})$$

by bandwidth condition A3.2. For $\nu = 1$, $\mu = 5, 7, \ldots$ and also for $\nu = 2, 3, \ldots$ the desired rate follows similarly.

(d) For $\nu \geq 1$ and $\mu \geq \nu + 1$ such that $\mu - \nu$ odd we can utilise the property of vanishing expectations $E[Z_i^{\mu-\nu}] = 0$. For some constant A we only consider the event A_n where

$$\max_{i=1,\ldots,n} |r_{ni}| \;\leq\; A\sqrt{\frac{\log h_n^{-1}}{nh_n}},$$

because $P(A_n^c) \to 0$ [compare (A.14)], and let further $|\ell^{(\mu-1)}|$ be bounded by A. We condition on the σ–field $\mathcal{E}_n$ generated by $\mathcal{Y}_n$ and $\tilde{Y}_1^*, \ldots, \tilde{Y}_n^*$, such that only $Z_1, \ldots, Z_n$ are random and $\tilde{m}^*$, r_{ni}, K_{ni} and η_{ni} are known. We proceed with Bernstein's inequality, conditioning on $\mathcal{E}_n$. For fixed y and $\epsilon > 0$, we have

$$P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}\xi_i\Big| > \epsilon \;\Big|\; \mathcal{E}_n\Big) \;\leq\; 2\exp(-\log n\, w_n)$$

where, conditional on $\mathcal{E}_n$, the random variables

$$\xi_i = \frac{\sqrt{n}a_n^{\mu-\nu}}{b_n^{\mu}}\ell^{(\mu-1)}\Big(\frac{y-\eta_{ni}+\kappa_n(X_i)+K_{ni}}{b_n}\Big)r_{ni}^{\nu}Z_i^{\mu-\nu}$$

are independent with expectation zero and $|\xi_i| \leq CA\sqrt{n}a_n^{\mu-\nu}b_n^{-\mu}(\frac{\log h_n^{-1}}{nh_n})^{\nu/2} = M_n$ for some constant C by boundedness of $\ell^{(\mu-1)}$ and Z_i and because event A_n occurs. Further,

$$\sum_{i=1}^{n}\mathrm{Var}(\xi_i \mid \mathcal{E}_n) \;\leq\; CA^{\nu+2}a_n^{2(\mu-\nu)}\Big(\frac{\log h_n^{-1}}{nh_n}\Big)^{\nu} n^2/b_n^{2\mu} \;=\; V_n,$$

and we have

$$w_n^{-1} \;=\; O\Big(\frac{a_n^{2(\mu-\nu)}(\log h_n^{-1})^{\nu}\log n}{(nh_n)^{\nu}b_n^{2\mu}}\Big) + O\Big(\frac{(\log h_n^{-1})^{\nu/2}a_n^{\mu-\nu}\log n}{(nh_n)^{\nu/2}\sqrt{n}b_n^{\mu}}\Big) \;=\; o(1).$$

The last equality can be deduced from the bandwidth conditions A3.2. $\square$

Proposition A.4 **(i)** *Under the assumptions of Theorem 3.2 we have*

$$\sup_{y\in\mathbb{R}}\Big|\frac{1}{n}\sum_{i=1}^{n}\ell^{(\mu-1)}\Big(\frac{y-\varepsilon_i}{b_n}\Big)\Big| = O(b_n^2) \text{ almost surely for } \mu=2,\ldots,\lambda,\ \mu \text{ even.}$$

(ii) *Under the assumptions of Theorem 3.2 we have*

$$\sup_{y\in\mathbb{R}}\frac{1}{n}\sum_{i=1}^{n}\Big|\ell^{(\mu-1)}\Big(\frac{y-\varepsilon_i}{b_n}\Big)\Big| = O(b_n) \text{ almost surely for } \mu=1,\ldots,\lambda.$$

(iii) *Under the assumptions of Theorem 3.2 we have*

$$\sup_{y\in\mathbb{R}}\frac{1}{n}\sum_{i=1}^{n}\Big(\ell^{(\mu-1)}\Big(\frac{y-\varepsilon_i}{b_n}\Big)\Big)^2 = O(b_n) \text{ almost surely for } \mu=1,\ldots,\lambda.$$

(iv) *Under the assumptions of Theorem 4.3 we have*

$$\sup_{y\in\mathbb{R}}\frac{1}{n}\sum_{i=1}^{n}\Big|\ell^{(\mu-1)}\Big(\frac{y-\varepsilon_i}{b_n}\Big)\Big|(1+|\varepsilon_i|)^{\nu} = O(b_n) \text{ almost surely for } \mu=2,\ldots,\lambda,\ \nu=1,\ldots,\mu-1.$$

Proof. Detailed proofs are omitted. They are done by first bounding the supremum of expectations straightforwardly, and then applying Theorem 37 by Pollard (1984, p. 34) for cases (i), (ii) and (iii), and a generalized version of this theorem [compare Neumeyer and Dette (2003, p. 906–908)] in case (iv) to the processes centered by their expectations. □

A.4 Proof of Theorem 4.3

The proof essentially follows the lines of the proof of Theorem 3.2 and we will only state the main differences. The last bandwidth condition in A4.3 assures that the auxiliary bandwidth $a_n = b_n^{(\lambda+3)/(\lambda+1)}$ fulfils the assumptions of Theorem 3.5 by Neumeyer (2006), where we choose a special kernel k with support $[-1,1]$ and vanishing third derivative. One can deduce that

$$\widehat{\tilde{\varepsilon}}_i^* = (\eta_{ni}+K_{ni})(1+s_n(X_i)) - \kappa_n(X_i) + r_{ni},$$

where $|r_{ni}| \le (1+|\eta_{ni}|)q_n$ and $q_n = O_p((\log h_n^{-1}/(nh_n))^{1/2})$. Here κ_n and s_n converge uniformly to zero and $\eta_{ni}+K_{ni}$ is randomly drawn from $\varepsilon_l + (Y_l - m_n(X_l))/\sigma_n(X_l)$, $l = 1,\ldots,n$. With the help of the auxiliary bootstrap residuals $\breve{\varepsilon}_i^* = \widehat{\tilde{\varepsilon}}_i^* + a_n Z_i$ the rest of the proof is then analogous to the proofs of Theorem 3.2 and Proposition A.3, but applying the following Proposition A.5 instead of Proposition A.2. □

Proposition A.5 *Under the assumptions of Theorem 4.3, for $\mu = 2,\ldots,\lambda$, $\nu = 1,\ldots,\mu-1$ it holds that*

(i) $$\sup_{y\in\mathbb{R}} \frac{1}{n}\sum_{i=1}^{n} \Big|\ell^{(\mu-1)}\Big(\frac{y-(\eta_{ni}+K_{ni})(1+s_n(X_i))+\kappa_n(X_i)}{b_n}\Big)\Big|(1+|\eta_{ni}|)^{\nu} = O_p(b_n)$$

(ii) $$\frac{1}{n}\sum_{i=1}^{n}(1+|\eta_{ni}|)^{2\nu} = O_p(1).$$

Proof. The random variables η_{ni} are randomly drawn with replacement from $\varepsilon_1, \ldots, \varepsilon_n$. Part (ii) therefore follows directly from the nonnegativity and because the (iterated) expectation is $E[(1+|\varepsilon_1|)^{2\nu}] < \infty$ by assumption. Part (i) can be proved analogously to Proposition A.2, but a detailed proof is omitted for the sake of brevity. □

Acknowledgements

This paper contains parts of the author's Habilitationsschrift "Bootstrap procedures for empirical processes of nonparametric residuals", Ruhr-Universität Bochum, 2006. The financial support of the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

References

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

A. Azzalini (1981). *A note on the estimation of a distribution function and quantiles by a kernel method.* Biometrika 68, 326–328.

A. Azzalini (1985). *A class of distributions which includes the normal ones.* Scand. J. Statist. 12, 171–178.

J. Bai (1996). *Testing for parameter constancy in linear regressions: An empirical distribution function approach.* Econometrica 64, 596–622.

R. J. Beran, L. le Cam and P. W. Millar (1987). *Convergence of stochastic empirical measures.* J. Multivariate Anal. 23, 159–168.

F. Cheng (2002). *Consistency of error density and distribution function estimation in nonparametric regression.* Statist. Probab. Lett. 59, 257–270.

F. Cheng (2004). *Weak and strong uniform consistency of a kernel error density estimator in nonparametric regression.* J. Statist. Plann. Inference 119, 95–107.

M. A. Delgado and W. González–Manteiga (2001). *Significance testing in nonparametric regression based on the bootstrap.* Ann. Statist. 29, 1469–1507.

H. Dette and I. Van Keilegom (2005). *A new test for the parametric form of the variance function in nonparametric regression.* technical report. (under revision for JRSS B) http://www.rub.de/mathematik3/preprint.htm

M. Falk (1983). *Relative efficiency and deficiency of kernel type estimators of smooth distribution functions.* Statistica Neerlandica 37, 73–83.

J. Fan and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications.* Chapman & Hall, London.

L. T. Fernholz (1991). *Almost sure convergence of smoothed empirical distribution functions.* Scand. J. Statist. 18, 255–262.

W. Härdle and A. W. Bowman (1988). *Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands.* J. Amer. Statist. Assoc. 83, 102–110.

P. D. Hill (1985). *Kernel estimation of a distribution function.* Comm. Statist. A—Theory Methods 14, 605–620.

S. Kiwitt, E.-R. Nagel and N. Neumeyer (2005). *Empirical likelihood estimators for the error distribution in nonparametric regression models.* preprint. http://www.rub.de/mathematik3/preprint.htm

H. L. Koul (1970). *Some convergence theorems for ranks and weighted empirical cumulatives.* Ann. Math. Statist. 41, 1768–1773.

H. L. Koul (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*, second edition. Springer, New York.

H. L. Koul and P. K. Sen (1985). *On a Kolmogorov-Smirnov type aligned test in linear regression.* Statist. Probab. Lett. 3, 111–115.

R. M. Loynes (1980). *The empirical distribution function of residuals from generalized regression.* Ann. Statist. 8, 285–298.

E. Mammen (1996). *Empirical process of residuals for high-dimensional linear models.* Ann. Statist. 24, 307–335.

J. Mora and N. Neumeyer (2005). *The two-sample problem with regression errors: an empirical process approach.* preprint. http://www.rub.de/mathematik3/preprint.htm

U. U. Müller, A. Schick and W. Wefelmeyer (2004a). *Estimating functionals of the error distribution in parametric and nonparametric regression.* J. Nonparametr. Statist. 16, 525–584.

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

U. Müller, A. Schick and W. Wefelmeyer (2004c). *Estimating the error distribution function in nonparametric regression.* technical report. http://www.mi.uni-koeln.de/∼wefelm/preprints.html

É. A. Nadaraya (1964a). *On non–parametric estimates of density functions and regression curves.* J. Probab. Appl. 10, 186–190.

É. A. Nadaraya (1964b). *Some new estimates for distribution functions.* Theory Probab. Appl. 9, 497–500.

N. Neumeyer (2006). *Smooth residual bootstrap for empirical processes of nonparametric regression residuals.* preprint. http://www.math.uni-hamburg.de/research/ims.html

N. Neumeyer and H. Dette (2003). *Nonparametric comparison of regression curves – an empirical process approach.* Ann. Statist. 31, 880–920.

N. Neumeyer and H. Dette (2004). *Testing for symmetric error distribution in nonparametric regression models.* Statistica Sinica, forthcoming. http://www.rub.de/mathematik3/preprint.htm

J. C. Pardo–Fernández, I. Van Keilegom and W. González–Manteiga (2004). *Comparison of regression curves based on the estimation of the error distribution.* Statistica Sinica, forthcoming. http://www.stat.ucl.ac.be/IAP/publication_tr.html

R.-D. Reiss (1981). *Nonparametric estimation of smooth distribution functions.* Scand. J. Statist. 8, 116–119.

J. Rice (1984). *Bandwidth choice for nonparametric regression.* Ann. Statist. 12, 1215–1230.

G. R. Shorack (1984) *Empirical and rank processes of observations and residuals.* Canad. J. Statist. 12, 319–332.

A. W. van der Vaart (1994). *Weak convergence of smoothed empirical processes.* Scand. J. Statist. 21, 501–504.

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 (2004). *Goodness-of-fit tests in parametric regression based on the estimation of the error distribution.* TEST, forthcoming. http://www.stat.ucl.ac.be/IAP/publication_tr.html

G. S. Watson (1964). *Smooth Regression Analysis.* Sankhya Ser. A 26, 359–372.

B. B. Winter (1973). *Strong uniform consistency of integrals of density estimators.* Can. J. Stat. 1, 247–253.

model	(i)			(ii)			(iii)		
α	0.025	0.05	0.1	0.025	0.05	0.1	0.025	0.05	0.1
(a) $n = 30$	0.027	0.056	0.116	0.031	0.044	0.089	0.042	0.073	0.125
$n = 100$	0.030	0.063	0.133	0.042	0.074	0.132	0.023	0.049	0.091
(b) $n = 30$	0.028	0.062	0.132	0.054	0.081	0.122	0.029	0.045	0.097
$n = 100$	0.035	0.071	0.142	0.054	0.079	0.120	0.023	0.049	0.099
(c) $n = 30$	0.029	0.060	0.116	0.046	0.071	0.127	0.041	0.069	0.123
$n = 100$	0.031	0.060	0.117	0.030	0.064	0.124	0.029	0.055	0.104

Table 1: Kolmogorov–Smirnov test for symmetry, (a) smooth residual bootstrap, (b) wild bootstrap, (c) classical residual bootstrap for the smooth empirical process: level simulations for the homoscedastic settings (i)–(iii)

model	(i)			(ii)			(iii)		
α	0.025	0.05	0.1	0.025	0.05	0.1	0.025	0.05	0.1
(a) $n = 30$	0.030	0.052	0.109	0.049	0.081	0.139	0.032	0.065	0.108
$n = 100$	0.026	0.047	0.101	0.045	0.075	0.141	0.033	0.058	0.111
(b) $n = 30$	0.032	0.070	0.125	0.070	0.098	0.154	0.050	0.080	0.130
$n = 100$	0.030	0.056	0.107	0.066	0.096	0.146	0.037	0.068	0.114
(c) $n = 30$	0.029	0.056	0.116	0.024	0.055	0.104	0.035	0.063	0.125
$n = 100$	0.030	0.058	0.117	0.018	0.036	0.102	0.017	0.039	0.083

Table 2: Cramér–von Mises test for symmetry, (a) smooth residual bootstrap, (b) wild bootstrap, (c) classical residual bootstrap for the smooth empirical process: level simulationsfor the homoscedastic settings (i)–(iii)

model	(iv)			(v)			(vi)		
α	0.025	0.05	0.1	0.025	0.05	0.1	0.025	0.05	0.1
(a) $n = 30$	0.094	0.131	0.229	0.159	0.222	0.336	0.144	0.201	0.303
$n = 50$	0.155	0.226	0.332	0.304	0.411	0.550	0.274	0.386	0.510
$n = 100$	0.316	0.436	0.582	0.609	0.717	0.839	0.536	0.657	0.790
$n = 150$	0.522	0.639	0.767	0.843	0.900	0.952	0.764	0.845	0.919
(b) $n = 30$	0.106	0.149	0.244	0.173	0.249	0.338	0.155	0.219	0.317
$n = 50$	0.169	0.242	0.343	0.331	0.422	0.566	0.291	0.308	0.499
$n = 100$	0.341	0.442	0.579	0.626	0.729	0.834	0.555	0.681	0.803
$n = 150$	0.528	0.647	0.768	0.852	0.905	0.953	0.783	0.847	0.917
(c) $n = 30$	0.094	0.165	0.264	0.194	0.292	0.425	0.177	0.262	0.388
$n = 50$	0.180	0.258	0.383	0.380	0.511	0.645	0.328	0.448	0.606
$n = 100$	0.428	0.531	0.668	0.756	0.840	0.906	0.654	0.782	0.868
$n = 150$	0.657	0.756	0.838	0.950	0.969	0.981	0.855	0.899	0.963

Table 3: Kolmogorov–Smirnov test for symmetry, (a) smooth residual bootstrap, (b) wild bootstrap, (c) classical residual bootstrap for the smooth empirical process: power simulations for the homoscedastic settings (iv)–(vi)

model	(iv)			(v)			(vi)		
α	0.025	0.05	0.1	0.025	0.05	0.1	0.025	0.05	0.1
(a) $n = 30$	0.084	0.137	0.217	0.158	0.250	0.360	0.134	0.205	0.310
$n = 50$	0.183	0.252	0.350	0.373	0.483	0.616	0.303	0.398	0.533
$n = 100$	0.407	0.513	0.632	0.755	0.844	0.904	0.672	0.766	0.848
$n = 150$	0.643	0.740	0.840	0.936	0.968	0.989	0.865	0.921	0.963
(b) $n = 30$	0.092	0.146	0.251	0.179	0.254	0.371	0.158	0.227	0.321
$n = 50$	0.196	0.262	0.365	0.398	0.502	0.627	0.333	0.417	0.538
$n = 100$	0.420	0.529	0.656	0.764	0.848	0.910	0.690	0.769	0.857
$n = 150$	0.651	0.742	0.848	0.934	0.965	0.988	0.874	0.922	0.966
$n = 30$	0.106	0.164	0.267	0.207	0.310	0.444	0.163	0.262	0.399
$n = 50$	0.173	0.269	0.403	0.421	0.551	0.690	0.316	0.438	0.634
$n = 100$	0.469	0.593	0.706	0.828	0.895	0.945	0.697	0.807	0.907
$n = 150$	0.728	0.822	0.899	0.978	0.991	0.993	0.891	0.950	0.976

Table 4: Cramér–von Mises test for symmetry, (a) smooth residual bootstrap, (b) wild bootstrap, (c) classical residual bootstrap for the smooth empirical process: power simulationsfor the homoscedastic settings (iv)–(vi)

model	(i)			(ii)			(iii)		
(n_1, n_2) | α	0.025	0.05	0.1	0.025	0.05	0.1	0.025	0.05	0.1
(a) (50,50)	0.031	0.055	0.118	0.034	0.054	0.126	0.029	0.059	0.109
(100,100)	0.029	0.056	0.106	0.038	0.061	0.122	0.033	0.053	0.106
(b) (50,50)	0.052	0.089	0.165	0.024	0.054	0.129	0.051	0.088	0.156
(100,100)	0.035	0.067	0.136	0.020	0.048	0.100	0.038	0.072	0.152

Table 5: Cramér–von Mises test for equality of error distributions: level simulations for the homoscedastic models (i)–(iii) for (a) smooth residual bootstrap for non-smooth empirical process and (b) the classical residual bootstrap version of the smooth empirical process

model	(i)			(ii)			(iii)		
(n_1, n_2) | α	0.025	0.05	0.1	0.025	0.05	0.1	0.025	0.05	0.1
(a) (50,50)	0.040	0.073	0.136	0.044	0.085	0.162	0.042	0.086	0.160
(100,100)	0.042	0.065	0.109	0.051	0.087	0.159	0.041	0.072	0.114
(b) (50,50)	0.056	0.106	0.174	0.050	0.090	0.194	0.061	0.095	0.160
(100,100)	0.036	0.062	0.129	0.044	0.083	0.154	0.048	0.080	0.164

Table 6: Kolmogorov–Smirnov test for equality of error distributions: level simulations for the homoscedastic models (i)–(iii) for (a) smooth residual bootstrap for non-smooth empirical process and (b) the classical residual bootstrap version of the smooth empirical process

model	(iv)			(v)			(vi)		
$(n_1, n_2) \mid \alpha$	0.025	0.05	0.1	0.025	0.05	0.1	0.025	0.05	0.1
(a) (30,30)	0.046	0.089	0.161	0.234	0.310	0.423	0.430	0.554	0.683
(50,50)	0.054	0.095	0.167	0.379	0.477	0.599	0.699	0.789	0.860
(75,100)	0.082	0.136	0.220	0.623	0.713	0.821	0.938	0.977	0.990
(100,100)	0.070	0.132	0.225	0.725	0.818	0.898	0.962	0.979	0.989
(150,150)	0.102	0.156	0.256	0.887	0.923	0.966	1.000	1.000	1.000
(b) (30,30)	0.102	0.158	0.268	0.372	0.496	0.620	0.630	0.734	0.848
(50,50)	0.106	0.176	0.294	0.596	0.766	0.792	0.874	0.920	0.956
(75,100)	0.144	0218	0.352	0.836	0.900	0.934	0.984	0.992	0.998
(100,100)	0.222	0.336	0.464	0.886	0.942	0.966	0.996	0.998	1.000
(150,150)	0.274	0.398	0.560	0.980	0.994	0.996	1.000	1.000	1.000

Table 7: Cramér–von Mises test for equality of error distributions: power simulations for the homoscedastic models (iv), (v) and (vi) for (a) smooth residual bootstrap for non-smooth empirical process and (b) the classical residual bootstrap version of the smooth empirical process

model	(iv)			(v)			(vi)		
(n_1, n_2) | α	0.025	0.05	0.1	0.025	0.05	0.1	0.025	0.05	0.1
(a) (30,30)	0.056	0.101	0.184	0.160	0.251	0.358	0.310	0.427	0.578
(50,50)	0.070	0.116	0.204	0.243	0.352	0.514	0.595	0.688	0.813
(75,100)	0.053	0.096	0.171	0.451	0.556	0.703	0.888	0.938	0.974
(100,100)	0.070	0.119	0.216	0.546	0.648	0.757	0.935	0.965	0.984
(150,150)	0.074	0.120	0.217	0.768	0.852	0.929	0.998	0.998	0.999
(b) (30,30)	0.096	0.164	0.266	0.372	0.466	0.606	0.642	0.748	0.828
(50,50)	0.112	0.156	0.274	0.564	0.664	0.766	0.866	0.924	0.958
(75,100)	0.100	0.142	0.262	0.802	0.858	0.916	0.986	0.994	0.998
(100,100)	0.170	0.242	0.366	0.858	0.906	0946	0.994	0.998	1.000
(150,150)	0.160	0.264	0.380	0.968	0.986	0.990	1.000	1.000	1.000

Table 8: Kolmogorov–Smirnov test for equality of error distributions: power simulations for the homoscedastic models (iv), (v) and (vi) for (a) smooth residual bootstrap for non-smooth empirical process and (b) the classical residual bootstrap version of the smooth empirical process