UNIVERSITÄT HAMBURG

Smooth residual bootstrap for empirical processes of nonparametric regression residuals

Natalie Neumeyer

Preprint–No. 2006-07 September 2006

FACHBEREICH MATHEMATIK

SCHWERPUNKT MATHEMATISCHE STATISTIK UND STOCHASTISCHE PROZESSE

BUNDESSTR. 55, D – 20146 HAMBURG

Smooth residual bootstrap for empirical processes of nonparametric regression residuals

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

September 6, 2006

Abstract

The aim of this paper is to prove the validity of smooth residual bootstrap versions of procedures that are based on the empirical process of residuals estimated from a nonparametric regression model. From this result consistency of various model tests in nonparametric regression is deduced, such as goodness-of-fit tests for the regression and variance function, tests for equality of regression functions, and tests concerning the error distribution.

AMS Classification: 62G09, 62G08, 62G10
Keywords and Phrases: empirical distribution function, goodness-of-fit tests, kernel estimators, resampling
Running title: Smooth residual bootstrap

1 Introduction

For a few decades in statistical research nonparametric regression models have been investigated intensively. Research has focused mainly on the nonparametric estimation of the regression function and variance function and corresponding hypotheses tests. For only a few years there have existed results on the estimation of the distribution of the unobserved errors. Weak convergence of the empirical process based on nonparametrically estimated residuals was shown by Akritas and Van Keilegom (2001). The empirical distribution function of estimated errors recently turned out to be valuable for goodness-of-fit tests concerning the regression or variance function, see for instance Pardo Fernández, Van Keilegom and González Manteiga (2004), Van Keilegom, González Manteiga and Sánchez Sellero (2004), and Dette and Van Keilegom (2005). In all the problems mentioned the asymptotic distribution of estimators and test statistics depends heavily on unknown features such as the error density. In situations like these to circumvent problems with the accuracy of the critical values, often resampling procedures such as bootstrap are applied. To the present author's knowledge, in nonparametric regression bootstrap methods were first investigated by Härdle and Bowman (1988), who introduced the (classical) residual bootstrap in the context of constructing local adaptive bandwidths and confidence intervals for the regression function in a homoscedastic regression model. Härdle and Mammen (1993) introduced the so-called wild bootstrap [due to Wu (1986)] in the context of nonparametric regression. They considered the problem of testing goodness-of-fit to decide whether the regression function belongs to a certain parametric model.

In the context of residual-based empirical processes the wild bootstrap method has the disadvantage that it changes the error distribution, even asymptotically. This is discussed in detail in Neumeyer (2006). In contrast the classical residual bootstrap has the disadvantage that, conditional on the original sample, the new bootstrap observations have a discrete error distribution. The methods applied when dealing with residual-based empirical processes heavily rely on a smooth error distribution, and it has been observed that classical bootstrap methods do not work well in this context. In the context of empirical processes of residuals from a linear model this was observed by Koul and Lahiri (1994), Koul (2002, chapter 6) and De Angelis, Hall and Young (1993). In the nonparametric regression context the smooth residual bootstrap, which will be explained throughout this work, was used in simulations and data analysis by Pardo-Fernandez et al. (2004), Van Keilegom et al. (2004), and Dette and Van Keilegom (2005), but never has a proof for the consistency in such a context been given. In the different context of a linear model with fixed design the validity of smooth residual bootstrap procedures for residual-based empirical processes was shown by Koul and Lahiri (1994) and Koul (2002, chapter 6), whereas in heteroscedastic parametric regression models with random design this was shown by Mora and Neumeyer (2005) in the context of testing for equality of error distributions in parametric regression models. But methods of proof in these two parametric cases are naturally very different from the nonparametric setting.

The aim of this work is to show in a nonparametric regression context that the residual empirical process built from a bootstrap sample attained from the smooth residual bootstrap, has, conditional on the initial sample, the same asymptotic distribution as the residual empirical process of the original sample. The applicability of our new results to tests concerning the error distribution (such as tests for equality of error distributions in different models) as well as to the tests suggested by Pardo-Fernandez et al. (2004), Van Keilegom et al. (2004), and Dette and Van Keilegom (2005) (goodness-of-fit tests for the regression and variance functions that are based on empirical processes of residuals) is also discussed.

The present work is organized as follows. Section 2 summarizes some asymptotic results for the empirical distribution function based on nonparametric residuals. A smooth residual bootstrap version of the residual-based empirical distribution function is considered in section 3. Section 4 deals with applications of the results to hypotheses testing, and finite sample properties are investigated in section 5. Some of the technical details and auxiliary results are presented in section 6. Here in particular results about uniform consistency of the residual based kernel density estimator as well as results about the bootstrap based regression estimator are given, which are of independent interest. The proofs of the main results are deferred to section 7.

2 The residual-based empirical process

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. Durbin (1973) considered empirical distribution functions when parameters are estimated. Residual-based empirical processes and residual-based kernel density estimators in the context of time series were considered by Boldin (1982), Koul and Levental (1989), Koul and Sen (1992), Liebscher (1999), Schick and Wefelmeyer (2002) and Bai (1994), among others.

In this section we give some results about empirical processes based on nonparametrically estimated residuals from regression models. Akritas and Van Keilegom (2001) considered the residual-based empirical distribution function estimated in a nonparametric regression setting and showed weak convergence of the corresponding empirical process, whereas analogous results for a smooth version of the residual-based empirical distribution function were proved by Müller, Schick and Wefelmeyer (2004). Cheng (2002) showed uniform consistency of a histogram density estimator and the empirical distribution function of nonparametric residuals, whereas in Cheng (2004) similar results for a residual-based kernel density estimator were given. Analogous histogram density estimators for $m(n)$–dependent nonparametric regression errors were considered by Qin (1996) and Qin, Sunjuan and Genxiang (1996).

We assume $(X_1, Y_1), \dots, (X_n, Y_n)$ to be an independent and identically distributed sample from a homoscedastic univariate regression model, i. e.

$$Y_i \;=\; m(X_i) + \varepsilon_i, \quad i = 1, \dots, n, \tag{2.1}$$

with unkown regression function m and not observable errors $\varepsilon_1, \dots, \varepsilon_n$ under the following assumptions.

- A.1 The univariate design points $X_1, \dots, X_n$ 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 $(0, 1)$.

- A.2 The errors $\varepsilon_1, \dots, \varepsilon_n$ 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 $\sup_{y\in I\!R} f(y) < \infty$ and $\sup_{y\in I\!R} |f'(y)| < \infty$.

In order to estimate the distribution F of the unobserved errors, one builds nonparametric residuals

$$\hat\varepsilon_i \;=\; Y_i - \hat m(X_i), \quad i = 1, \dots, n, \tag{2.2}$$

where $\hat{m}(x)$ denotes the Nadaraya-Watson kernel regression estimator [Nadaraya (1964), Watson (1964)] for $m(x)$, that is

$$\hat{m}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}K\Big(\frac{X_i - x}{h_n}\Big)Y_i\frac{1}{\hat{f}_X(x)} \tag{2.3}$$

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

$$\hat{f}_X(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}K\Big(\frac{X_i - x}{h_n}\Big). \tag{2.4}$$

Here, K denotes a kernel function and h_n a sequence of bandwidths. We need the following assumptions.

- A.3 Let K denote a twice continuously differentiable symmetric density with compact support $[-1, 1]$, say, such that $\int uK(u)\,du = 0$ and $K(-1) = K(1) = 0$.
- A.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 A.4 can be chosen arbitrarily small. It has only relevance according to technics of proofs. Note that Akritas and Van Keilegom's (2001) bandwidth condition (A1)(i),

$$\frac{nh_n^{3+2\delta}}{\log(h_n^{-1})} \to \infty, \tag{2.5}$$

follows from our assumption A.4 for $\delta = \alpha/4$, for instance. This bandwidth condition will be used frequently during the proof of the bootstrap results.
We consider the residual-based empirical distribution function

$$\hat{F}_n(y) = \frac{1}{n}\sum_{i=1}^{n} I\{\hat{\varepsilon}_i \le y\}, \quad y \in I\!R, \tag{2.6}$$

with the following limit distribution. Here and throughout $I\{\cdot\}$ denotes the indicator function and $y \wedge z$ denotes the minimum of y and z.

Theorem 2.1 *Under model (2.1) with assumptions A.1–4 the process $R_n(y) = \sqrt{n}(\hat{F}_n(y) - F(y))$, $y \in I\!R$, 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 $U(y) = E[\varepsilon_1 I\{\varepsilon_1 \le y\}]$.

Note that the model considered by Akritas and Van Keilegom (2001) is different from model (2.1). First of all, their model is heteroscedastic, further the regression and variance functions are defined as certain L-functionals depending on a non-constant score function J in $[0, 1]$. Our model corresponds to a homoscedastic version with constant score function $J \equiv I_{[0,1]}$. Slight

modifications of Akritas and Van Keilegom's (2001) arguments show the validity of Theorem 2.1 under assumptions A.1–4, which are less restrictive than in the aforementioned publication [also see the proof of Theorem 3.1 by Van Keilegom et al. (2004)].
The covariance in Theorem 2.1 depends heavily on the unknown error distribution F and its density. For the calculation of confidence regions or for testing problems the use of the limit distribution is difficult because the asymptotic covariance has to be estimated by inserting nonparametric estimators for F, f, σ^2 and U. The rate of convergence of some of these estimators is usually rather slow and it is recommended to use a resampling procedure instead of the asymptotic distribution. In section 3 we will discuss the validity of a smooth residual bootstrap version of the stochastic process considered in Theorem 2.1.

Remark 2.2 *In theory throughout the present work we use the kernel regression estimator (2.3) for the regression function. This is mainly because of its simplicity in theory. However, all the same methods based on the residual empirical process can be used when other nonparametric function estimators are preferred, for instance local polynomial estimators [see Fan and Gijbels (1996), among many others] or estimators based on wavelets [see, for instance, Hall and Patil (1996)] et cetera. Theoretical properties which the function estimator $\hat{m}$ should fulfil to derive analogous asymptotic expansions and limit distributions for the residual-based empirical process as described in section 2 for the original sample and in section 3 for bootstrap samples, are uniform consistency, some regularity assumptions for the derivative and a linear expansion for $\int(\hat{m}-m)(x)f_X(x)\,dx$ with remainder of order $o_p(n^{-1/2})$.*

For the remainder of this section we turn our focus to a heteroscedastic regression model

$$(2.7) \qquad Y_i \;=\; m(X_i) + \sigma(X_i)\varepsilon_i, \quad i = 1, \ldots, n,$$

under the following assumptions.

- A.5 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 $\inf_{x\in[0,1]} \sigma^2(x) > 0$.

- A.6 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 $\sup_{y\in I\!R} |yf(y)| < \infty$ and $\sup_{y\in I\!R} |y^2 f'(y)| < \infty$.

The kernel regression estimator $\hat{m}$ defined in (2.3) is used. Further we estimate the variance function by a similar kernel approach, i. e.

$$(2.8) \qquad \hat{\sigma}^2(x) \;=\; \frac{1}{nh_n}\sum_{i=1}^{n} K\Big(\frac{X_i - x}{h_n}\Big)(Y_i - \hat{m}(x))^2 \frac{1}{\hat{f}_X(x)}$$

with $\hat{f}_X$ defined in (2.4). Now the standardized residuals from the nonparametric fit are defined by

$$\hat{\varepsilon}_i = \frac{Y_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)}, \quad i = 1, \dots, n. \tag{2.9}$$

The residual-based empirical distribution function is defined as $\hat{F}_n$ in (2.6) based on the new residuals. Due to the estimation of the variance function the asymptotic distribution changes compared to Theorem 2.1. We state the weak convergence result in the next theorem [compare Akritas and Van Keilegom (2001)].

Theorem 2.3 *Under the assumptions A.3–6 the process $\sqrt{n}(\hat{F}_n(y) - F(y))$, $y \in I\!R$, converges weakly to a centered Gaussian process G with covariance*

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

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

3 Smooth residual bootstrap

Our aim is to construct a bootstrap version of the residual-based empirical process that shares the limiting process of Theorem 2.1 in terms of conditional weak convergence given the initial sample $\mathcal{Y}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ from the homoscedastic model (2.1). To this end we generate bootstrap errors ε_i^* according to the equality

$$\varepsilon_i^* = \tilde{\varepsilon}_i^* + a_n Z_i. \tag{3.1}$$

Here a_n denotes a positive smoothing parameter, $Z_1, \dots, Z_n$ are independent, centered random variables (independent of $\mathcal{Y}_n$) with density k and $\tilde{\varepsilon}_1^*, \dots, \tilde{\varepsilon}_n^*$ are randomly drawn with replacement from centered residuals $\{\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n\}$ defined by

$$\tilde{\varepsilon}_i = \hat{\varepsilon}_i - \frac{1}{n}\sum_{j=1}^{n} \hat{\varepsilon}_j, \tag{3.2}$$

with $\hat{\varepsilon}_i$ from (2.2). We have for the conditional expectation $E[\varepsilon_i^* \mid \mathcal{Y}_n] = 0$ because of $E[Z_i] = 0$ and the centering of the residuals. Conditional on the original sample $\mathcal{Y}_n$ the random variables $\varepsilon_1^*, \dots, \varepsilon_n^*$ defined in (3.1) are independent and identically distributed with distribution function $\tilde{F}_n$ corresponding to the density

$$\tilde{f}_n(y) = \frac{1}{na_n}\sum_{i=1}^{n} k\Big(\frac{y - \tilde{\varepsilon}_i}{a_n}\Big). \tag{3.3}$$

Note that the smoothness of the distribution of the bootstrap residuals $\varepsilon_1^*, \dots, \varepsilon_n^*$ is a crucial point in the asymptotic theory. Next, we build new bootstrap observations,

$$Y_i^* \;=\; \hat{m}(X_i) + \varepsilon_i^*, \quad i = 1, \dots, n, \tag{3.4}$$

where the regression estimator $\hat{m}$ was defined in (2.3). From the bootstrap sample $(X_1, Y_1^*), \dots,$ (X_n, Y_n^*) we estimate the regression function by analogy with (2.3) and denote the estimator by $\hat{m}^*$. Note that, for the sake of simplicity, we use the same bandwidth h_n as for the estimation of $\hat{m}$. Residuals from the bootstrap observations are defined by

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

and the empirical distribution function of $\hat{\varepsilon}_1^*, \dots, \hat{\varepsilon}_n^*$ is denoted by $\hat{F}_n^*$. The smooth residual bootstrap version of the residual empirical process R_n defined in Theorem 2.1 is

$$R_n^*(y) \;=\; \sqrt{n}(\hat{F}_n^*(y) - \tilde{F}_n(y)) \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^n (I\{\hat{\varepsilon}_i^* \le y\} - \tilde{F}_n(y)),\; y \in \mathbb{R}. \tag{3.5}$$

Sections 6.2 and 6.3 in the appendix give some results about the error density estimator $\tilde{f}_n$ and the bootstrap regression estimator $\hat{m}^*$, respectively, some of which are of independent interest. For the following main theorem (Theorem 3.2) we need some more assumptions.

A.7 Let k denote a symmetric probability density function either with support $\mathbb{R}$ or compact support such that $\int k(z)z\,dz = 0$, $\int k(z)z^2\,dz < \infty$, and k has $\kappa \ge 2$ derivatives inside the support, such that the first $\kappa - 1$ derivatives are of bounded variation with $\int (k^{(\nu)}(z))^2\,dz < \infty$, $\nu = 1, \dots, \kappa - 1$, and all κ derivatives are bounded.

A.8 Let a_n denote a positive sequence of bandwidths such that for $n \to \infty$, $a_n = o(1)$,

$$\frac{nh_n a_n^{2+8\delta/3}}{\log h_n^{-1}} \to \infty, \qquad \frac{nh_n a_n^{2+\frac{2}{\kappa}(1+4\delta/3)}}{\log h_n^{-1}} \to \infty, \qquad h_n = o\Big(a_n^{1+\frac{4\delta}{3(\kappa+1)}}\Big),$$

where h_n is defined in assumption A.4 and $\delta \in (0, \frac{1}{2})$ denotes the constant defined in (2.5) (which can be chosen arbitrarily small).

A.9 Let the first derivative of the kernel K defined in assumption A.3 be of bounded variation.

A.10 Let $E[|\varepsilon_1|^\upsilon] < \infty$, where $\upsilon > 1 + \frac{2}{\delta}$ is some constant such that $\int k(u)u^\upsilon\,du < \infty$, and δ is defined in (2.5).

Remark 6.4 below shows a case where the bandwidth conditions A.8 can be relaxed, whereas Remark 6.9 explains how assumption A.10 can be omitted.
First we formulate a result about asymptotic expansions of the bootstrap version of the empirical process. The proof is given in section 7.

Lemma 3.1 *Assume model (2.1) is valid under assumptions A.1–4 and A.7–10.*
(i) *Then,* $\sup_{y\in\mathbb{R}} |V_n^*(y)| = o_p(1)$*, where*

$$\begin{aligned} V_n^*(y) &= R_n^*(y) - \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{\varepsilon_i^* \le y\} - \tilde{F}_n(y)\Big) \\ &\quad - \sqrt{n}\int \Big(\tilde{F}_n(y + \hat{m}^*(x) - \hat{m}(x)) - \tilde{F}_n(y)\Big) f_X(x)\,dx. \end{aligned}$$

(ii) *Then it holds that*

$$\sup_{y\in\mathbb{R}}\left|\sqrt{n}\int \Big(\tilde{F}_n(y + \hat{m}^*(x) - \hat{m}(x)) - \tilde{F}_n(y)\Big) f_X(x)\,dx - \tilde{f}_n(y)\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i^*\right| = o_p(1).$$

Now we state the main asymptotic result, from which consistency of smooth residual bootstrap versions of procedures based on the empirical process of residuals follows as explained below. The proof is given in section 7.

Theorem 3.2 *Assume model (2.1) is valid under assumptions A.1–4 and A.7–10. Then, conditional on the sample $\mathcal{Y}_n$, the process $R_n^*(y)$, $y \in \mathbb{R}$, defined in (3.5) converges weakly to a centered Gaussian process G with covariance* $\mathrm{Cov}(G(y), G(z))$ *defined in Theorem 2.1, in probability.*

Validity of the bootstrap procedure can be deduced as follows. Let T_n and its bootstrap version T_n^* be continuous functionals ψ of the stochastic processes R_n and R_n^*, respectively, i. e. $T_n = \psi(R_n)$, $T_n^* = \psi(R_n^*)$. Assume that R_n converges weakly to R, such that T_n converges in distribution to $T = \psi(R)$. Weak convergence in $\ell^\infty(\mathbb{R})$ of a process $R_n^*(y)$, $y \in \mathbb{R}$, conditional on the sequence $\mathcal{Y} = \{(X_1, Y_1), (X_2, Y_2), \ldots\}$, to R, in probability, is defined as follows. The conditional expectation $E[H(R_n^*) \mid \mathcal{Y}]$, given the sequence $\mathcal{Y}$, converges in probability to $E[H(R)]$ for all bounded and uniformly continuous functionals $H : \ell^\infty(\mathbb{R}) \to \mathbb{R}$; for this or similar definitions see van der Vaart and Wellner (1996), Sweeting (1989), Bickel and Freedman (1981), Singh (1981) or Giné and Zinn (1990). Note that in our context the observations in the original sample are independent and, hence, $E[H(R_n^*) \mid \mathcal{Y}] = E[H(R_n^*) \mid \mathcal{Y}_n]$, and we say that R_n^*, conditional on $\mathcal{Y}_n$, converges to R in probability (as in the theorem). Then, conditional on the sequence $\mathcal{Y}$, T_n^* converges in distribution to T, in probability. Define $c_{n,\alpha}^*$ by $P(T_n^* \le c_{n,\alpha}^* \mid \mathcal{Y}_n) = \alpha$. Then we have the following result [see Beran, LeCam and Millar (1987), Delgado and González Manteiga (2001)].

Corollary 3.3 *Under the assumptions of Theorem 3.2 with the above definitions it holds that* $P(T_n \le c_{n,\alpha}^*) = \alpha + o(1)$.

Remark 3.4 *For the bootstrap version of the process it is possible to use bandwidths that converge to zero with a slower rate than required in assumption A.4, because the bias can be* *reduced so that the assumption $nh_n^5 = O(1)$ instead of $nh_n^4 \to 0$ is sufficient. To see this, note that the bias originates from the term (compare the proof of Lemma 3.1 (ii))*

$$\tilde{f}_n(y)\sqrt{n}\int(\hat{m}^*(x) - \hat{m}(x))f_X(x)\,dx \;=\; \tilde{f}_n(y)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{m}^*(X_i) - \hat{m}(X_i)) + o_p(1).$$

The bias due to this term cancels, when we substract the term

$$\tilde{f}_n(y)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{nh_n}\sum_{j=1}^{n}\frac{K(\frac{X_i-X_j}{h_n})}{\hat{f}_X(X_i)}(\hat{m}(X_j) - \hat{m}(X_i)),$$

which is completely known, from the considered process $R_n^*(y) = \sqrt{n}(\hat{F}_n^*(y) - \tilde{F}_n(y))$.

Now we consider how the bootstrap approach should be adapted for the heteroscedastic model (2.7). We define the density $\tilde{f}_n$ [see (3.3)] as a kernel density estimate of $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$ with kernel k and bandwidth a_n, where now the standardised residuals are defined as

$$\tilde{\varepsilon}_i \;=\; \frac{\hat{\varepsilon}_i - \frac{1}{n}\sum_{j=1}^{n}\hat{\varepsilon}_j}{\sqrt{\frac{1}{n}\sum_{k=1}^{n}(\hat{\varepsilon}_k - \frac{1}{n}\sum_{j=1}^{n}\hat{\varepsilon}_j)^2}}. \tag{3.6}$$

When $\tilde{\varepsilon}_1^*, \ldots, \tilde{\varepsilon}_n^*$ are randomly drawn with replacement from $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$ we generate bootstrap errors $\varepsilon_1^*, \ldots, \varepsilon_n^*$ as defined in (3.1). Those have distribution $\tilde{F}_n$ corresponding to the density $\tilde{f}_n$, are centered and have the conditional variance $\mathrm{Var}(\varepsilon_i^* \mid \mathcal{Y}_n) = 1 + a_n^2\int k(u)u^2\,du$. Bootstrap observations are now by analogy with (3.4) defined by

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

where the variance function estimator $\hat{\sigma}^2$ is defined in (2.8). From the new bootstrap sample $(X_1, Y_1^*), \ldots, (X_n, Y_n^*)$ by analogy with (2.3) and (2.8) we define estimators $\hat{m}^*$ and $\hat{\sigma}^{*2}$ and the residuals obtained from the bootstrap sample are [compare (2.9)],

$$\hat{\varepsilon}_i^* \;=\; \frac{Y_i^* - \hat{m}^*(X_i)}{\hat{\sigma}^*(X_i)} \tag{3.7}$$

The empirical distribution function of $\hat{\varepsilon}_1^*, \ldots, \hat{\varepsilon}_n^*$ is denoted by $\hat{F}_n^*$ and the bootstrap version of the empirical process is $R_n^* = \sqrt{n}(\hat{F}_n^* - \tilde{F}_n)$. We impose an additional assumption.

A.11 Let $E[|\varepsilon_1|^{2\kappa}] < \infty$, where κ is defined in assumption A.7. The functions $y \mapsto yk^{(\nu)}(y)$ are assumed to be bounded and of bounded variation, $\nu = 0, \ldots, \kappa - 1$. For the bandwidths we further assume that $na_n^4 \to 0$ and $a_n^2/h_n^{1+\delta} = o(1)$.

Theorem 3.5 *Assume model (2.7) is valid under assumptions A.3–11. Then, conditional on the sample $\mathcal{Y}_n$, the process $R_n^*(y)$, $y \in \mathbb{R}$, converges weakly to a centered Gaussian process G with covariance* $\mathrm{Cov}(G(y), G(z))$ *defined in Theorem 2.3, in probability.*

The proof is similar to the proof of Theorem 3.2 as given in section 6, but requires some more technical effort. It is omitted for the sake of brevity, but details can be found in Neumeyer (2006).

4 Application to testing problems

In this section test statistics based on residual empirical processes for several testing problems in nonparametric regression models are considered and the bootstrap results of section 3 are applied in various contexts. The general idea is the following. 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 some empirical process of residuals R_n, i. e. $T_n = \psi(R_n)$. 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. Let t_n denote the realisation of the test statistic T_n based on the sample $\mathcal{Y}_n$. The bootstrap procedure is based on a "new sample" $\mathcal{Y}_n^*$, where now in contrast to section 3 it is important that the bootstrap sample fulfils the null hypothesis H_0. 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 $T_n^* = \psi(R_n^*)$. The $(1-\alpha)$–quantile of the distribution of T_n can now be approximated by the bootstrap quantile $c^*_{n,1-\alpha}$ defined by $P(T_n^* > c^*_{n,1-\alpha} \mid \mathcal{Y}_n) = \alpha$ and we obtain an asymptotic level α–test by rejecting the null hypothesis whenever $t_n > c^*_{n,1-\alpha}$ (compare Corollary 3.3). 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 $c^*_{n,1-\alpha}$ is usually estimated by the $\lfloor B(1-\alpha)\rfloor$-th order statistic of B replications $T_n^{*,1}, \dots, T_n^{*,B}$ of bootstrap experiments all based on the same sample $\mathcal{Y}_n$ (for some large B).

4.1 Goodness-of-fit tests for the regression function

Van Keilegom, González Manteiga and Sánchez Sellero (2004) use the theory of residual-based empirical processes to propose a new goodness-of-fit test in the nonparametric heteroscedastic regression model (2.7). For some parametric class of functions, $\mathcal{M} = \{m_\vartheta : [0,1] \to I\!R \mid \vartheta \in \Theta\}$, it is of interest in many applications whether the regression function m belongs to that class, and we define hypotheses as $H_0 : m \in \mathcal{M}$ versus $H_1 : m \notin \mathcal{M}$. Van Keilegom et al. (2004) suggest measuring the difference between null hypothesis and alternative by considering the difference between the empirical distribution functions of parametrically and nonparametrically estimated residuals, respectively. To be more specific, the estimator $\hat\vartheta$ of the parameter ϑ minimises the expression $Q_n(\vartheta) = \frac{1}{n}\sum_{i=1}^n (Y_i - m_\vartheta(X_i))^2$ over $\vartheta \in \Theta$. This gives an estimator for the value $\tilde\vartheta_0$ that minimises $E[(m(X_1) - m_\vartheta(X_1))^2]$ and coincides with the "true" parameter ϑ_0 under H_0. Theorem 2.1 in Van Keilegom et al. (2004) shows that the null hypothesis H_0 holds if and only if the error ε_1 has the same distribution as ε_0 defined by $\varepsilon_0 = (Y_1 - m_{\tilde\vartheta_0}(X_1))/\sigma(X_1)$. Therefore, a consistent test can be based on the difference of the residual-based empirical distribution function $\hat F_n$ and $\hat F_{0n}$, the empirical distribution function of parametric residuals $\hat\varepsilon_{01}, \dots, \hat\varepsilon_{0n}$

defined as $\hat{\varepsilon}_{0i} = (Y_i - \hat{m}_0(X_i))/\hat{\sigma}(X_i)$ with nonparametric variance estimator $\hat{\sigma}^2$ defined in (2.8), and where $\hat{m}_0$ denotes a smoothed version of $m_{\hat{\vartheta}}$ analogous to Härdle and Mammen (1993), i. e.

$$\hat{m}_0(x) \;=\; \frac{1}{\hat{f}_X(x)} \frac{1}{nh_n} \sum_{i=1}^{n} K\Big(\frac{x - X_i}{h_n}\Big) m_{\hat{\vartheta}}(X_i) \tag{4.1}$$

with $\hat{f}_X$ from (2.4). Under the heteroscedastic nonparametric regression model (2.7) with our assumptions A.3–6 and some additional regularity conditions Van Keilegom et al. (2004) prove that, under the null hypothesis H_0, the process $\sqrt{n}(\hat{F}_n(y) - \hat{F}_{n0}(y))$, $y \in I\!R$, converges weakly to $f(y)W$, where W is a centered normal random variable with variance

$$\operatorname{Var}(W) \;=\; E\Big[\Big(1 - \int \frac{f_X(x)}{\sigma(x)} \frac{\partial m_\vartheta(x)}{\partial \vartheta}\Big|_{\vartheta=\vartheta_0}^{T} dx\, \Omega^{-1} \frac{\partial m_\vartheta(X_1)}{\partial \vartheta}\Big|_{\vartheta=\vartheta_0} \sigma(X_1)\Big)^2\Big], \tag{4.2}$$

and

$$\Omega \;=\; E\Big[\frac{\partial m_\vartheta(X_1)}{\partial \vartheta}\Big|_{\vartheta=\tilde{\vartheta}_0} \frac{\partial m_\vartheta(X_1)}{\partial \vartheta}\Big|_{\vartheta=\tilde{\vartheta}_0}^{T}\Big].$$

As test statistics for testing for a parametric form of the regression function, a Kolmogorov–Smirnov and a Cramér–von Mises functional of $\sqrt{n}(\hat{F}_n - \hat{F}_{0n})$ are proposed. Because the limit distribution depends on unknown features, the aforementioned authors suggest using a smooth residual bootstrap explained below, and show in a simulation study that the new tests are at least competitive with the methods proposed by Härdle and Mammen (1993) and Stute (1997). However, a theoretical justification for consistency of the proposed resampling procedure in this context is not given, and we will now investigate how the theoretical results obtained in section 3 of the present work are applicable in this context.
We build bootstrap observations (X_i, Y_i^*), $i = 1, \ldots, n$, similarly to the heteroscedastic smooth residual bootstrap as explained in section 3. These observations should fulfil the null hypothesis of a parametric regression function m_ϑ for some $\vartheta \in \Theta$, therefore they are defined as

$$Y_i^* \;=\; m_{\hat{\vartheta}}(X_i) + \hat{\sigma}(X_i)\varepsilon_i^*,$$

where ε_i^* has density $\tilde{f}_n$ defined in (3.3), but based on standardized residuals defined in (3.6). Let $\hat{F}_n^*$ denote the empirical distribution function of the residuals $\hat{\varepsilon}_i^*$ obtained from the bootstrap sample and defined in (3.7) with the help of nonparametric regression and scale estimators $\hat{m}^*$ and $\hat{\sigma}^*$ analogous to (2.3) and (2.8) but based on the bootstrap sample (X_i, Y_i^*), $i = 1, \ldots, n$. To define the bootstrap analogue to $\hat{F}_{n0}$, let $\hat{\vartheta}^*$ be a minimiser of $Q_n^*(\vartheta) = \frac{1}{n}\sum_{i=1}^n (Y_i^* - m_\vartheta(X_i))^2$ over $\vartheta \in \Theta$. Analogous to (4.1) let $\hat{m}_0^*$ denote the smoothed version of $m_{\hat{\vartheta}^*}$ and let $\hat{F}_{n0}^*$ denote the empirical distribution function of $\hat{\varepsilon}_{0i}^* = (Y_i^* - \hat{m}_0^*(X_i))/\hat{\sigma}^*(X_i)$, $i = 1, \ldots, n$. With a very similar argumentation as given in sections 6 and 7, under the assumptions of Theorem 3.5 and the additional assumptions made by Van Keilegom et al. (2004) we can prove the next theorem. Details of the proof are omitted for the sake of brevity, but can be found in Neumeyer (2006).

Theorem 4.1 *Assume model (2.7) is valid under assumptions of Theorem 3.5 and the additional assumptions made by Van Keilegom et al. (2004). Then, conditional on the sample $\mathcal{Y}_n$, the process $\sqrt{n}(\hat{F}_n^*(y) - \hat{F}_{n0}^*(y))$, $y \in I\!R$, converges weakly to $f(y)W$, in probability, where W is a centered normal random variable with variance defined in (4.2).*

Note that Theorem 4.1 is valid both under the null hypothesis and under fixed alternatives, which shows that the bootstrap procedure suggested by Van Keilegom et al. (2004) to test for a parametric form of a regression function is consistent.

4.2 Testing for a parametric variance function

In the present work we have considered homoscedastic as well as heteroscedastic regression models. Testing for heteroscedasticity has attracted some attention in nonparametric regression literature, and recently, Dette and Van Keilegom (2005) proposed a test for a parametric form of the variance function. To be more specific, let $\mathcal{M} = \{\sigma_\vartheta^2 : [0,1] \to I\!R^+ \mid \vartheta \in \Theta\}$ denote a class of variance functions and consider the hypotheses $H_0 : \sigma^2 \in \mathcal{M}$ versus $H_1 : \sigma^2 \notin \mathcal{M}$ in the nonparametric heteroscedastic regression model (2.7). Testing for heteroscedasticity is a special case, setting $\sigma_\vartheta^2 = \vartheta$. Dette and Van Keilegom (2005) consider Kolmogorov–Smirnov and Cramér–von Mises test statistics based on the difference of the empirical distribution function $\hat{F}_n$ of nonparametrically estimated residuals $\hat\varepsilon_1, \ldots, \hat\varepsilon_n$ defined in (2.9) and the empirical distribution function $\hat{F}_{0n}$ of residuals $\hat\varepsilon_{01}, \ldots, \hat\varepsilon_{0n}$ estimated under the null hypothesis, i.e. $\hat\varepsilon_{0i} = (Y_i - \hat m(X_i))/\hat\sigma_0(X_i)$. Here, $\hat\sigma_0 = \sqrt{\hat\sigma_0^2}$ is defined by

$$\hat\sigma_0^2(x) \;=\; \frac{1}{\hat f_X(x)} \frac{1}{nh_n} \sum_{i=1}^n K\Big(\frac{x - X_i}{h_n}\Big) \sigma_{\hat\vartheta}^2(X_i),$$

where $\hat\vartheta$ is a minimiser of $Q_n(\vartheta) = \frac{1}{n}\sum_{i=1}^n [(Y_i - \hat m(X_i))^2 - \sigma_\vartheta^2(X_i)]^2$ over $\vartheta \in \Theta$. Under assumptions A.3–6 and some additional regularity assumptions, Dette and Van Keilegom (2005) show that, under H_0, the process $\sqrt{n}(\hat F_n(y) - \hat F_{0n}(y))$, $y \in I\!R$, converges weakly to $yf(y)W$, where W is a centered normal random variable with variance

$$\text{(4.3)} \qquad \begin{aligned} &\mathrm{Var}(W) \\ &= \frac{1}{4} E[(\varepsilon_1^2 - 1)^2] E\Big[\Big(1 - \int \frac{\sigma_{\vartheta_0}^2(X_1)}{\sigma_{\vartheta_0}^2(x)} \frac{\partial \sigma_\vartheta^2(x)}{\partial\vartheta}\Big|_{\vartheta=\vartheta_0}^T f_X(x)\,dx\,\Omega^{-1} \frac{\partial\sigma_\vartheta^2(X_1)}{\partial\vartheta}\Big|_{\vartheta=\vartheta_0}\Big)^2\Big], \end{aligned}$$

where

$$\Omega \;=\; E\Big[\frac{\partial\sigma_\vartheta^2(X_1)}{\partial\vartheta}\Big|_{\vartheta=\tilde\vartheta_0} \frac{\partial\sigma_\vartheta^2(X_1)}{\partial\vartheta}\Big|_{\vartheta=\tilde\vartheta_0}^T\Big].$$

In a simulation study Dette and Van Keilegom (2005) consider a Cramér–von Mises test statistic based on the process $\sqrt{n}(\hat F_n - \hat F_{0n})$ and implement a smooth residual bootstrap as follows. Let $\varepsilon_1^*, \ldots, \varepsilon_n^*$ be a sample from $\tilde f_n$ defined in (3.3) [where Dette and Van Keilegom (2005) use a standard normal kernel k] and define bootstrap observations as

$$Y_i^* \;=\; \hat m(X_i) + \sigma_{\hat\vartheta}(X_i)\varepsilon_i^*$$

such that these fulfil the null hypothesis of a parametric variance function. Nonparametric residuals as well as parametric residuals are built from the bootstrap sample (X_i, Y_i^*), $i = 1, \ldots, n$, to obtain the bootstrap version $\sqrt{n}(\hat F_n^* - \hat F_{0n}^*)$ of the process $\sqrt{n}(\hat F_n - \hat F_{0n})$. The procedure is compared to tests by Dette and Munk (1998), Zhu, Fujikoshi and Naito (2001), Dette (2002) and Francisco-Fernández and Vilar-Fernández (2005) in a simulation study, and in many cases the new test yields substantially larger power in the problem of testing homoscedasticity. However, a theoretical justification for this bootstrap procedure is not given in the aforementioned reference. Yet with a proof similar to the proof of Theorem 3.2 using methods of sections 6 and 7 as well as the methods employed by Dette and Van Keilegom (2005) we can show the following Theorem.

Theorem 4.2 *Assume model (2.7) is valid under the assumptions of Theorem 3.5 as well as the assumptions made by Dette and Van Keilegom (2005). Then, conditional on the sample $\mathcal{Y}_n$, the process $\sqrt{n}(\hat{F}_n^*(y) - \hat{F}_{0n}^*(y))$, $y \in I\!R$, converges weakly to $yf(y)W$, in probability, where W is a centered normal random variable with variance defined in (4.3).*

From Theorem 4.2 follows that Dette and Van Keilegom's (2005) test is consistent.

4.3 Two-sample problems

In this section suppose we have two independent nonparametric regression models of type (2.7),

$$\begin{aligned} Y_{1j} &= m_1(X_{1j}) + \sigma_1(X_{1j})\varepsilon_{1j}, \quad j = 1, \ldots, n_1, \\ Y_{2j} &= m_2(X_{2j}) + \sigma_2(X_{2j})\varepsilon_{2j}, \quad j = 1, \ldots, n_2, \end{aligned}$$

which are to be compared. Tests for equality of the two regression functions m_1 and m_2 with hypothesis $H_0 : m_1 = m_2$ versus $H_1 : m_1 \neq m_2$ have attracted a lot of attention during the last couple of decades. Recently Pardo Fernández, Van Keilegom and González Manteiga's (2004) proposed a test based on the difference of empirical distribution functions of residuals estimated under the null hypothesis [compare (4.4)] and under the alternative, respectively. The weak convergence of the process is shown and local alternatives are discussed. In a simulation study these authors use a smooth residual bootstrap. It is important to note that the new bootstrap observations have to fulfil the null hypothesis, therefore a pooled estimator

$$(4.4) \qquad \hat{m}_0(x) = \frac{\sum_{i=1}^2 \sum_{j=1}^{n_i} K(\frac{X_{ij}-x}{h_N})Y_{ij}}{\sum_{i=1}^2 \sum_{j=1}^{n_i} K(\frac{X_{ij}-x}{h_N})}$$

of the regression function is defined and bootstrap samples are defined by

$$\begin{aligned} Y_{1j}^* &= \hat{m}_0(X_{1j}) + \hat{\sigma}_1(X_{1j})\varepsilon_{1j}^*, \quad j = 1, \ldots, n_1 \\ Y_{2j}^* &= \hat{m}_0(X_{2j}) + \hat{\sigma}_2(X_{2j})\varepsilon_{2j}^*, \quad j = 1, \ldots, n_2, \end{aligned}$$

where for each $i \in \{1, 2\}$, $\varepsilon_{i1}^*, \ldots, \varepsilon_{in_i}^*$ denotes a random sample generated from a kernel density estimator based on residuals $\tilde{\varepsilon}_{i1}, \ldots, \tilde{\varepsilon}_{in_i}$ [standardised versions of $\hat{\varepsilon}_{ij} = (Y_{ij} - \hat{m}_i(X_{ij}))/\hat{\sigma}_i(X_{ij})$, analogous to (3.6)]. From the new samples (X_{ij}, Y_{ij}^*) test statistics are built in the same way as from the initial samples. With methods of the presented work we can give the theoretical justification for the consistency of Pardo Fernández et al.'s (2004) procedure.
Speckman, Chiu, Hewett and Bertelson (2003) suggested a test for equality of regression functions against one-sided alternatives $H_1 : m_1 > m_2$ based on ranking the residuals that was

further developed by Neumeyer and Dette (2005). The latter authors give the asymptotic distribution of the test statistic and also prove the consistency of a symmetric wild bootstrap under the additional restrictive assumption of a symmetric error distribution. Stronger results which show that the smooth residual bootstrap is consistent without the symmetry assumption on the error distribution can be deduced similarly to the methods in the present work.

A test for equality of error distributions in two regression models can be based on Kolmogorov-Smirnov or Cramér-von Mises tests based on the differences of empirical distribution functions of residuals in the two samples, respectively. A smooth residual bootstrap version of such a test was investigated by Mora and Neumeyer (2005) for parametric regression models. Here the bootstrap observations are built under the null hypothesis of equal error distributions and are generated from a smooth distribution estimated using all residuals from both samples. For nonparametric regression models similar results can be obtained from the methods explained in the present work.

5 Finite sample performance

In this section we present some simulation results. In particular we investigate how sensitive the smooth residual bootstrap procedures are with respect to the choice of the bandwidth a_n. We approximate the $(1-\alpha)$–quantile of the distribution of the statistic $T_n = \psi(\sqrt{n}(\hat{F}_n - F))$ for the functionals

$$\psi(Z) = \sup_{y \in I\!R} |Z(y)| \quad \text{and} \quad \psi(Z) = \int Z^2 d\hat{F}_n$$

corresponding to Kolmogorov–Smirnov and Cramér–von Mises statistics, respectively, by $\hat{c}^*_{n,1-\alpha}$, the $\lfloor (1-\alpha)B \rfloor$–th order statistic of bootstrap statistics $T_n^{*,1}, \dots, T_n^{*,B}$. Here the bootstrap statistic is

$$T_n^* = \psi(\sqrt{n}(\hat{F}_n^* - \tilde{F}_n)).$$

It is counted how often the original value T_n is larger than $\hat{c}^*_{n,1-\alpha}$, which should be approximately α times. We display results for $\alpha \in \{0.025, 0.05, 0.1\}$. All simulations are done using R [R Development Core Team (2005)]. For the estimation of the regression function we use a 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 A.4. Here, $\hat{s}$ denotes a simple estimator for the variance (or, in the heteroscedastic model, the integrated variance function), see Rice (1984). Note that we have employed a local linear estimator with bandwidth choice by plug-in methodology as described by Ruppert, Sheather and Wand (1995). Those results are not displayed, because we observed that the values obtained from the two methods of estimating m are quite similar.

For the bandwidth a_n used for smoothing the bootstrap errors we consider different choices, namely $a_n \in \{\frac{1}{2}n^{-1/4}, n^{-1/3}, n^{-1/4}, 2n^{-1/3}, 2n^{-1/4}\}$. We estimate the quantiles from $B = 200$ bootstrap replications in each of 1000 simulation runs.

We consider various settings corresponding to models (2.1) and (2.7) with constant or non-constant variance, with different regression functions, error distributions and sample sizes. The first three settings are according to model (2.1) with uniform design, normally distributed errors and a rather small sample size, i. e. $n = 50$, where we consider different regression functions. The values for bandwidth a_n for $n = 50$ are approximately 0.188, 0.271, 0.376, 0.543 and 0.752, respectively. To be more precise, the settings are the following,

(i) model (2.1), uniform design, $n = 50$, $N(0, 0.5)$ distributed errors, $m(x) = 1 - 2x$

(ii) model (2.1), uniform design, $n = 50$, $N(0, 0.5)$ distributed errors, $m(x) = e^x$

(iii) model (2.1), uniform design, $n = 50$, $N(0, 0.5)$ distributed errors, $m(x) = 2x^2$.

The results are displayed in Table 1 for the Cramér–von Mises statistic and Kolmogorov–Smirnov statistic.

model	(i)			(ii)			(iii)		
α	0.025	0.05	0.1	0.025	0.05	0.1	0.025	0.05	0.1
(a) $a_n = \frac{1}{2}n^{-1/4}$	0.022	0.051	0.101	0.017	0.054	0.095	0.025	0.054	0.108
$a_n = n^{-1/3}$	0.026	0.048	0.099	0.033	0.050	0.099	0.025	0.045	0.093
$a_n = n^{-1/4}$	0.025	0.046	0.101	0.025	0.057	0.105	0.027	0.057	0.111
$a_n = 2n^{-1/3}$	0.023	0.048	0.114	0.030	0.062	0.123	0.024	0.050	0.094
$a_n = 2n^{-1/4}$	0.034	0.053	0.099	0.027	0.058	0.100	0.035	0.061	0.105
(b) $a_n = \frac{1}{2}n^{-1/4}$	0.021	0.040	0.091	0.014	0.029	0.066	0.016	0.037	0.070
$a_n = n^{-1/3}$	0.024	0.048	0.106	0.019	0.038	0.076	0.026	0.046	0.096
$a_n = n^{-1/4}$	0.032	0.049	0.112	0.022	0.043	0.105	0.018	0.039	0.093
$a_n = 2n^{-1/3}$	0.028	0.064	0.105	0.020	0.046	0.095	0.021	0.039	0.075
$a_n = 2n^{-1/4}$	0.020	0.045	0.100	0.026	0.049	0.095	0.031	0.050	0.102

Table 1: (a) Cramér–von Mises, (b) Kolmogorov–Smirnov, homoscedastic models (i)–(iii)

We have also implemented the classical residual bootstrap, i. e. with smoothing parameter $a_n = 0$, to demonstrate that this method should not be applied in the context of residual-based empirical distribution functions. Those results are not diplayed in their entirety, but, for example, in the very simple setting (i) for $\alpha = 0.025$, 0.05 and 0.1, respectively, we obtained the approximations 0.002, 0.006 and 0.016 for the Kolmogorov–Smirnov statistic and 0.002, 0.007 and 0.028 for the Cramér–von Mises statistic.

In settings (iv)–(vi) we consider different error distributions, i. e. Student's t–distribution, double exponential distribution and skew-normal distribution, which was generated using the R-package *sn* [Azzalini (2005)]. The skew-normal distribution with scale parameter 1 and shape parameter β has the density $f(y) = 2\phi(y)\Phi(\beta y)$, where ϕ and Φ denote the density and cumulative distribution function of the standard normal distribution, respectively. The double exponential distribution has density

$$f(x) = \frac{1}{\gamma} \exp\Big(- \frac{x-q}{\gamma} \Big) \times \exp\Big(- \exp\Big(- \frac{x-q}{\gamma} \Big) \Big),$$

where we choose $\gamma = 3$ and $q = -\gamma c$ such that the expectation $\gamma c + q$ vanishes, where $c = 0.5772$ approximately. The values are scaled such that the variance is 0.5. The sample size is $n = 100$ and the corresponding values for the smoothing parameter a_n are approximately 0.158, 0.215, 0.316, 0.431 and 0.632, respectively. The different settings are

(iv) model (2.1), uniform design, $n = 100$, Student's t–distributed errors with three degrees of freedom, $m(x) = x$

(v) model (2.1), uniform design, $n = 100$, errors with centered double exponential distribution with variance 0.5, $m(x) = x$

(vi) model (2.1), uniform design, $n = 100$, errors with centered skew-normal distribution with scale parameter 1 and shape parameter 1.5, $m(x) = 2x^2$.

In Table 2 we display the results for settings (iv)–(vi) for the Cramér–von Mises statistic. Here as well as in Table 1 we observe that the Kolmogorov–Smirnov statistic reacts a bit more sensitively to too small choices of the bootstrap smoothing parameter a_n than the Cramér–von Mises statistic. All in all we observe good approximations of the quantiles in settings (i)–(iv) and (vi), whereas for the double exponential distribution (v) the quantiles are overestimated when applying the Cramér–von Mises statistic. Under model (iv) the quantiles are overestimated for too large bandwidths, i. e. $a_n = 2n^{-1/4}$.
Moreover we give a few results for the heteroscedastic model (2.7) according to the settings (vii)–(ix) defined below. Note that here the variances are larger than in the homoscedastic settings. We have observed that the heteroscedastic procedure reacts more sensitively to the choice of the smoothing parameter a_n, and smaller values of a_n lead to more accurate results. We display the performance for $a_n = \frac{1}{2}n^{-1/4}$, $\frac{1}{2}n^{-1/5}$, $n^{-1/3}$ and $n^{-1/4}$ in Table 3. The models chosen are

(vii) procedures according to the heteroscedastic model (2.7) model, but constant variance function $\sigma^2 \equiv 1$, $n = 50$, $m(x) = 2x^2$, standard normal errors

(viii) model (2.7) with variance function $\sigma^2(x) = e^x/2$, $n = 100$, $m(x) = 2x^2$, standard normal errors

(ix) model (2.7) with variance function $\sigma^2(x) = (1 + x)^2/2$, $n = 100$, $m(x) = 2x^2$, standard normal errors.

Overall, for the considered cases (i)–(ix) we can draw the conclusion that the performance of the smooth residual bootstrap in the context of residual-based empirical processes does not depend on the choice of a_n very heavily.

model	(iv)			(v)			(vi)		
α	0.025	0.05	0.1	0.025	0.05	0.1	0.025	0.05	0.1
(a) $a_n = \frac{1}{2}n^{-1/4}$	0.025	0.054	0.108	0.047	0.078	0.123	0.026	0.059	0.100
$a_n = n^{-1/3}$	0.020	0.048	0.111	0.034	0.081	0.140	0.034	0.057	0.094
$a_n = n^{-1/4}$	0.020	0.049	0.096	0.051	0.094	0.149	0.026	0.050	0.108
$a_n = 2n^{-1/3}$	0.029	0.063	0.130	0.062	0.099	0.162	0.035	0.066	0.110
$a_n = 2n^{-1/4}$	0.047	0.098	0.158	0.075	0.107	0.176	0.025	0.057	0.102
(b) $a_n = \frac{1}{2}n^{-1/4}$	0.008	0.033	0.073	0.027	0.051	0.087	0.016	0.030	0.065
$a_n = n^{-1/3}$	0.017	0.048	0.101	0.030	0.054	0.104	0.014	0.030	0.069
$a_n = n^{-1/4}$	0.021	0.056	0.127	0.036	0.059	0.113	0.022	0.040	0.090
$a_n = 2n^{-1/3}$	0.034	0.066	0.114	0.042	0.074	0.119	0.030	0.055	0.106
$a_n = 2n^{-1/4}$	0.042	0.078	0.143	0.046	0.086	0.142	0.025	0.047	0.098

Table 2: (a) Cramér–von Mises, (b) Kolmogorov–Smirnov, homoscedastic models (iv)–(vi)

model	(vii)			(viii)			(ix)		
α	0.025	0.05	0.1	0.025	0.05	0.1	0.025	0.05	0.1
(a) $a_n = \frac{1}{2}n^{-1/4}$	0.016	0.043	0.104	0.021	0.047	0.089	0.020	0.044	0.085
$a_n = \frac{1}{2}n^{-1/5}$	0.024	0.044	0.091	0.028	0.044	0.080	0.017	0.041	0.092
$a_n = n^{-1/3}$	0.020	0.038	0.090	0.020	0.041	0.083	0.022	0.048	0.106
$a_n = n^{-1/4}$	0.024	0.051	0.086	0.030	0.052	0.109	0.032	0.060	0.099
(b) $a_n = \frac{1}{2}n^{-1/4}$	0.028	0.043	0.074	0.025	0.049	0.093	0.021	0.051	0.101
$a_n = \frac{1}{2}n^{-1/5}$	0.013	0.033	0.073	0.021	0.046	0.088	0.017	0.045	0.093
$a_n = n^{-1/3}$	0.024	0.045	0.083	0.024	0.047	0.086	0.030	0.049	0.102
$a_n = n^{-1/4}$	0.020	0.045	0.104	0.043	0.071	0.113	0.033	0.058	0.101

Table 3: (a) Cramér–von Mises, (b) Kolmogorov–Smirnov, heteroscedastic models (vii)–(ix)

For simulations of the finite sample properties of the smooth residual bootstrap for testing goodness-of-fit of the regression and variance functions and testing the equality of regression functions as explained in section 4 we refer to Van Keilegom et al. (2004), Dette and Van Keilegom (2005), and Pardo Fernández et al. (2004). In all these publications good approximations of the test level and good detection of alternatives were observed. Simulations for tests for equality of error distributions and for symmetry of error distributions can be found in Neumeyer (2006) but are not displayed her for the sake of brevity.

6 Auxiliary results

6.1 Regression estimation

This section gives some results about uniform convergence of kernel density and regression estimators based on the original sample as well as based on the bootstrap observations. Recall the definitions of $\hat{m}$ and $\hat{f}_X$ in (2.3) and (2.4), respectively. Let m_n and $\tilde{m}_n$ be defined as

$$m_n(x) = \frac{1}{h_n}\int K\Big(\frac{x-y}{h_n}\Big)m(y)f_X(y)\,dy\,\frac{1}{f_{X,n}(x)} \tag{6.1}$$

$$\tilde{m}_n(x) = \frac{1}{h_n}\int K\Big(\frac{x-y}{h_n}\Big)m_n(y)f_X(y)\,dy\,\frac{1}{f_{X,n}(x)}, \tag{6.2}$$

where

$$f_{X,n}(x) = \frac{1}{h_n}\int K\Big(\frac{x-y}{h_n}\Big)f_X(y)\,dy. \tag{6.3}$$

Note that m_n and $\tilde{m}_n$ are deterministic functions with

$$\sup_{x\in[0,1]}|m_n(x)-m(x)| = O(h_n) \quad\text{and}\quad \sup_{x\in[0,1]}|\tilde{m}_n(x)-m(x)| = O(h_n). \tag{6.4}$$

For the kernel density estimator we have the following result.

Proposition 6.1 *Assume model (2.1) with assumptions A.1–4. Then it holds that*

$$\begin{aligned}
\sup_{x\in[0,1]}|\hat{f}_X(x)-f_{X,n}(x)| &= o(\beta_n) \quad \textit{almost surely}\\
\sup_{x\in[0,1]}|\hat{f}'_X(x)-f'_{X,n}(x)| &= \frac{1}{h_n}o(\beta_n) \quad \textit{almost surely}\\
\sup_{x,t\in[0,1]}\frac{|\hat{f}_X(x)-\hat{f}_X(t)|}{|x-t|^{\delta/2}} &= O_p(h_n^{-\delta})\\
\sup_{x,t\in[0,1]}\frac{|\hat{f}'_X(x)-\hat{f}'_X(t)|}{|x-t|^{\delta/2}} &= O_p(h_n^{-1-\delta})
\end{aligned}$$

for all sequences β_n of positive real numbers such that $nh_n\beta_n\to\infty$ and $\log n/(nh_n\beta_n^2)=o(1)$ and for δ defined in (2.5).

Proof. The proof of the first assertion is given in Example 38 of Pollard (1984, p. 35) and carried out by an application of Theorem 37 by Pollard (1984). The second assertion follows by the same steps. The proof of the third assertion is omitted because it is very similar to the proof of the last, which is shown as follows. By inserting the factor $1 = I\{|x-t|>c_n\}+I\{|x-t|\le c_n\}$ we obtain the estimate

$$h_n^{1+\delta}\sup_{x,t\in[0,1]}\frac{|\hat{f}'_X(x)-\hat{f}'_X(t)|}{|x-t|^{\delta/2}}$$

$$\leq\; h_n^{1+\delta} \sup_{x,t\in[0,1]} \frac{|\hat{f}'_X(x) - f'_X(x) - \hat{f}'_X(t) + f'_X(t)|}{|x-t|^{\delta/2}} + h_n^{1+\delta} \sup_{x,t\in[0,1]} \frac{|f'_X(x) - f'_X(t)|}{|x-t|^{\delta/2}}$$

$$\leq\; 2\frac{h_n^{1+\delta}}{c_n^{\delta/2}} \sup_{x\in[0,1]} |\hat{f}'_X(x) - f'_X(x)| + 2h_n^{1+\delta} \sup_{x,t\in[0,1]} \frac{|x-t|}{|x-t|^{\delta/2}} \sup_{x\in[0,1]} |f''_X(x)|$$

$$+ \sup_{|x-t|\leq c_n} \frac{h_n^{1+\delta}}{|x-t|^{\delta/2}} \Big| \frac{1}{nh_n^2} \sum_{i=1}^n \Big(K'(\frac{x-X_i}{h_n}) - K'(\frac{t-X_i}{h_n}) \Big) \Big|$$

$$\leq\; \frac{h_n^{1+\delta}}{c_n^{\delta/2}} O_p(1) + O(h_n^{1+\delta}) + \sup_{|x-t|\leq c_n} \frac{h_n^{1+\delta}}{|x-t|^{\delta/2}} \frac{1}{h_n^2} \Big| \frac{x-t}{h_n} \Big| \sup_{x\in[0,1]} |K''(x)|.$$

For the choice $c_n = h_n^{1+2/\delta}$ we obtain the rate $O_p(h_n^{\delta/2})$, which is even better than the desired $O_p(1)$. □

By bandwidth conditions A.4, for instance, the sequence $\beta_n = h_n^{1+\delta}$ or $\beta_n = c_n(\log(h_n^{-1})/(nh_n))^{1/2}$ fulfils the conditions of Proposition 6.1 for each sequence c_n with $\lim_{n\to\infty} c_n = \infty$. From this

$$\sup_{x\in[0,1]} |\hat{f}_X(x) - f_{X,n}(x)| \;=\; O_p\Big(\sqrt{\frac{\log(h_n^{-1})}{nh_n}}\Big)$$

can be deduced. Further one has

$$\sup_{x\in[0,1]} |\hat{f}_X(x) - f_X(x)| \;=\; o(\beta_n) + O(h_n) \;=\; o(1) \tag{6.5}$$

almost surely. Similar assertions hold for the kernel estimators $\hat{m}$.

Proposition 6.2 *Under model (2.1) with assumptions A.1–4 for $\hat{m}$ and m_n defined in (2.3) and (6.1), respectively, it holds that*

$$\sup_{x\in[0,1]} |\hat{m}(x) - m_n(x)| \;=\; o(\beta_n) \quad \textit{almost surely}$$

for all sequences β_n of positive real numbers such that $nh_n\beta_n \to \infty$ and $\log n/(nh_n\beta_n^2) = o(1)$. Further it holds that

$$\sup_{x\in[0,1]} |\hat{m}'(x) - m'_n(x)| \;=\; o_p(1)$$

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

where $\delta > 0$ denotes the constant in (2.5).

Proof. Writing $\hat{r} = \hat{m}\hat{f}_X$ and $r_n = m_n f_{X,n}$, we use the decomposition

$$|\hat{m} - m_n| \;\leq\; \frac{|\hat{r} - r_n|}{\hat{f}_X} + |r_n| \frac{|\hat{f}_X - f_{X,n}|}{\hat{f}_X f_{X,n}}.$$

Because r_n is bounded and f_X is bounded from zero, by an application of Proposition 6.1 and (6.5) for the first assertion it suffices to show

$$\sup_{x\in[0,1]} |\hat{r}(x) - r_n(x)| \;=\; o(\beta_n) \tag{6.6}$$

almost surely. Note that $E[\hat{r}(x)] = r_n(x)$ and assertion (6.6) can be written as

$$\frac{1}{h_n}\sup_{\varphi\in\mathcal{F}_n}\Big|\frac{1}{n}\sum_{i=1}^{n}(\varphi(X_i,Y_i) - E[\varphi(X_1,Y_1)])\Big| \;=\; o(\beta_n) \tag{6.7}$$

almost surely, where the function class $\mathcal{F}_n$ is defined as the subset of

$$\mathcal{F} \;=\; \Big\{\varphi : [0,1]\times I\!R \to I\!R, \varphi(X,Y) = K\Big(\frac{x-X}{h_n}\Big)Y \;\Big|\; x\in[0,1], h\in(0,c)\Big\}$$

given by $h = h_n$. Here c denotes some constant, and we can apply a generalization of Theorem 37 by Pollard (1984) for function classes with non-constant envelope for sequences $\delta_n^2 = Ch_n$ for some constant C and $\alpha_n = \beta_n$ as in the assertion to prove (6.7). A rigorous generalisation of the Theorem for that case can be found in Neumeyer (2001, p. 50–53); see also Neumeyer (2006).
For the second assertion we have for the derivatives

$$\hat{m}' = \frac{\hat{r}'}{\hat{f}_X} - \hat{r}\frac{\hat{f}_X'}{\hat{f}_X^2} \quad\text{and}\quad m_n' = \frac{r_n'}{f_{X,n}} - r_n\frac{f_{X,n}'}{f_{X,n}^2}$$

and obtain a decomposition

$$|\hat{m}' - m_n'| \;\le\; \frac{1}{\hat{f}_X}|\hat{r}' - r_n'| + |r_n'|\frac{|\hat{f}_X - f_{X,n}|}{\hat{f}_X f_{X,n}} + |\hat{r} - r_n|\frac{|f_{X,n}'|}{f_{X,n}^2} + |\hat{r}|\Big|\frac{\hat{f}_X'}{\hat{f}_X^2} - \frac{f_{X,n}'}{f_{X,n}^2}\Big|.$$

Because $h_n r_n'$ and $h_n f_{X,n}'$ are bounded, by applications of Proposition 6.1 for $\beta_n = h_n$, Proposition 6.2 and (6.5) is suffices to prove that

$$\sup_{x\in[0,1]} |\hat{r}'(x) - r_n'(x)| \;=\; o_p(1)$$

is valid. Noting that $E[\hat{r}'(x)] = r_n'(x)$ this can be shown with the same method as in the proof of the first assertion.
The proof of the last assertion can be given similarly to the proof of Lemma 6.5 (iii) below, but it is less complicated. □

From (6.4) and the proposition we have

$$\sup_{x\in[0,1]} |\hat{m}(x) - m(x)| \;=\; o(\beta_n) + O(h_n) \;=\; o(1) \tag{6.8}$$

almost surely, which is needed in some of the proofs of the main results. For these proofs we further need rates of convergence for $|\hat{m}(X_i) - m(X_i)|$. We obtain

$$\begin{aligned}|\hat{m}(X_i) - m(X_i)| \;\le&\; \sup_{x\in[0,1]}|\hat{m}(x) - m_n(x)| + \sup_{x\in[h_n,1-h_n]}|m_n(x) - m(x)|\\ &+ (I\{X_i < h_n\} + I\{X_i > 1-h_n\})\sup_{x\in[0,1]\setminus[h_n,1-h_n]}|m_n(x) - m(x)|\\ \le&\; t_n + C_1h_n^2 + C_2t_{ni}h_n \end{aligned} \tag{6.9}$$

for some constants C, C_1, C_2, some random variable $t_n = o(\beta_n)$, where β_n fulfils the conditions of 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 $l > 0$.

6.2 Residual-based density estimation

In this section we give results about uniform convergence of the kernel density estimator $\tilde{f}_n$ defined in (3.3) and its distribution function $\tilde{F}_n$ applied in the proof of Theorem 3.2. Under bandwidth conditions less restrictive than A.8, namely $a_n \to 0$, $nh_n a_n^2/\log n \to \infty$ (but only for kernels with bounded support), Cheng (2004) showed uniform almost sure consistency of $\tilde{f}_n$. We give a different proof that allows obtaining a rate of convergence.

Lemma 6.3 *Under model (2.1) with assumptions A.1–4 and A.7–8, we have almost surely* **(i)** $\sup_{y\in\mathbb{R}} |\tilde{f}_n(y) - f(y)| = o(a_n^{4\delta/3})$, **(ii)** $\sup_{y,z\in\mathbb{R}} \frac{|\tilde{f}_n(y) - f(y) - \tilde{f}_n(z) + f(z)|}{|y-z|^{\delta/2}} = o(1)$, *and* **(iii)** $\sup_{y\in\mathbb{R}} |\tilde{F}_n(y) - F(y)| = o(1)$. *Here the constant δ is defined in (2.5).*

Proof. We have for the residual-based density estimator $\tilde{f}_n$ defined in (3.3) by Taylor's expansion,

$$\tilde{f}_n(y) = \hat{f}_n(y) + \sum_{\nu=1}^{\kappa-1} U_n^{(\nu)}(y) + V_n(y),$$

where $U_n^{(\nu)}$ and V_n will be defined in a moment and

$$\hat{f}_n(y) = \frac{1}{na_n}\sum_{i=1}^n k\Big(\frac{y-\varepsilon_i}{a_n}\Big)$$

is the kernel density estimator based on the true errors and fulfils the assertion by the following considerations. We obtain for the expectations,

$$\sup_{y\in\mathbb{R}} |E[\hat{f}_n(y)] - f(y)| = \sup_{y\in\mathbb{R}} \Big|\frac{1}{a_n}\int k\Big(\frac{y-z}{a_n}\Big)(f(z)-f(y))\,dz\Big|.$$

By a change of variable $(y-z)/a_n = t$ and Taylor's expasion of $f(y - a_n t) - f(y)$ around y, using the assumptions, this gives the rate $O(a_n) = o(a_n^{4\delta/3})$. Next we write

$$\hat{f}_n(y) - E[\hat{f}_n(y)] = \frac{1}{a_n} W_n(\varphi_{y,a_n}) \tag{6.10}$$

for the stochastic process

$$W_n(\varphi) = \frac{1}{n}\sum_{i=1}^n (\varphi(\varepsilon_i) - E[\varphi(\varepsilon_i)])$$

indexed by a class $\mathcal{F}_n$ of functions, where each $\mathcal{F}_n$ is the subset of

$$\mathcal{F} = \Big\{\varphi_{y,b} : \mathbb{R} \to \mathbb{R} \;\Big|\; \varphi_{y,b}(\varepsilon) = k\Big(\frac{y-\varepsilon}{b}\Big), y \in \mathbb{R}, b \in (0,c)\Big\}$$

given by fixing $b = a_n$. The class $\mathcal{F}$ fulfills the assumptions of Theorem 37 by Pollard (1984, p. 34) and this theorem can be applied with sequences $\alpha_n = a_n^{4\delta/3}$ and $\delta_n^2 = O(a_n)$. For the uniform convergence of (6.10) to zero we obtain the almost sure rate $\frac{1}{a_n} o(\delta_n^2 \alpha_n) = o(a_n^{4\delta/3})$. Further we have

$$U_n^{(\nu)}(y) \;=\; \frac{1}{n a_n^{\nu+1}} \sum_{i=1}^n k^{(\nu)}\Big(\frac{y-\varepsilon_i}{a_n}\Big)(\tilde{\varepsilon}_i - \varepsilon_i)^{\nu} \frac{(-1)^{\nu}}{\nu!} \tag{6.11}$$

$$V_n(y) \;=\; \frac{1}{n a_n^{\kappa+1}} \sum_{i=1}^n k^{(\kappa)}(\xi_{y,i,n})(\tilde{\varepsilon}_i - \varepsilon_i)^{\kappa} \frac{(-1)^{\kappa}}{\kappa!}. \tag{6.12}$$

From the definition of $\tilde{\varepsilon}_i$ in (3.2) and $\hat{\varepsilon}_i$ in (2.2) it follows that

$$\begin{aligned}
\tilde{\varepsilon}_i - \varepsilon_i \;&=\; \hat{\varepsilon}_i - \varepsilon_i - \frac{1}{n}\sum_{j=1}^n \hat{\varepsilon}_j \\
&=\; -\frac{1}{n}\sum_{j=1}^n \varepsilon_j + m(X_i) - \hat{m}(X_i) - \frac{1}{n}\sum_{j=1}^n (m(X_j) - \hat{m}(X_j)) \qquad (6.13)\\
&=\; -\frac{1}{n}\sum_{j=1}^n \varepsilon_j + z_{ni} - \frac{1}{n}\sum_{j=1}^n z_{nj} + h_n C_{ni} t_{ni} - h_n \frac{1}{n}\sum_{j=1}^n C_{nj} t_{nj},
\end{aligned}$$

where the last line follows from (6.9) with random variables z_{ni} that are uniformly in i of order $o(\beta_n) + O(h_n^2)$ almost surely for a sequence β_n that fulfils the conditions of Proposition 6.1 and we choose $\beta_n = c_n(\log h_n^{-1}/(nh_n))^{1/2}$ for some sequence $c_n \to \infty$ that will be specified later. Further C_{ni} fulfils $\max_{i=1,\ldots,n} |C_{ni}| = O(1)$ almost surely. For the independent, non-negative random variables $t_{ni} = I\{X_i < h_n\} + I\{X_i > 1 - h_n\}$ we have $E[t_{ni}] = O(h_n)$ and $\mathrm{Var}(t_{ni}) = O(h_n)$. By the strong law of large numbers it holds that

$$h_n \frac{1}{n}\sum_{j=1}^n C_{nj} t_{nj} \;=\; O(h_n^2) \tag{6.14}$$

almost surely. Further note that $((\log\log n)/n)^{1/2} = o(\beta_n)$ and $h_n^2 = o(\beta_n)$. Hence, by the law of iterated logarithm for $\nu = 1, \ldots, \kappa - 1$, we have

$$\begin{aligned}
\sup_{y\in I\!R} |U_n^{(\nu)}(y)| \;\le\; & o(\beta_n^{\nu}) \left(\frac{1}{a_n^{\nu+1}} \sup_{y\in I\!R} |W_n^{(\nu)}(y)| + \frac{1}{a_n^{\nu+1}} \sup_{y\in I\!R} E\Big[\Big|k^{(\nu)}\Big(\frac{y-\varepsilon_1}{a_n}\Big)\Big|\Big] \right) \\
& + O(h_n^{\nu}) \left(\frac{1}{a_n^{\nu+1}} \sup_{y\in I\!R} |\tilde{W}_n^{(\nu)}(y)| + \frac{1}{a_n^{\nu+1}} \sup_{y\in I\!R} \Big| E\Big[t_{n1} k^{(\nu)}\Big(\frac{y-\varepsilon_1}{a_n}\Big)\Big]\Big| \right),
\end{aligned}$$

where

$$\begin{aligned}
W_n^{(\nu)}(y) \;&=\; \frac{1}{n}\sum_{i=1}^n \Big(\Big|k^{(\nu)}\Big(\frac{y-\varepsilon_i}{a_n}\Big)\Big| - E\Big[\Big|k^{(\nu)}\Big(\frac{y-\varepsilon_i}{a_n}\Big)\Big|\Big]\Big) \\
\tilde{W}_n^{(\nu)}(y) \;&=\; \frac{1}{n}\sum_{i=1}^n \Big(t_{ni} k^{(\nu)}\Big(\frac{y-\varepsilon_i}{a_n}\Big) - E\Big[t_{n1} k^{(\nu)}\Big(\frac{y-\varepsilon_i}{a_n}\Big)\Big]\Big).
\end{aligned}$$

Further we have

$$\sup_{y\in I\!R} E\Big[\Big|k^{(\nu)}\Big(\frac{y-\varepsilon_1}{a_n}\Big)\Big|\Big] \;=\; O(a_n), \tag{6.15}$$

$E[t_{n1}] = O(h_n)$, and $a_n^{-\nu-1}E[k^{(\nu)}((y-\varepsilon_i)/a_n)]$ converges to $f^{(\nu)}(y)$ and is uniformly of order $O(1)$. Similar to the considerations for W_n before we apply Theorem 37 by Pollard (1984, p. 34) for sequences of real numbers $\delta_n^2 = Ca_n$, $\alpha_n = a_n^{\nu+4\delta/3}/\beta_n^{\nu}$ to obtain the uniform almost sure convergence of $\sup_{y\in I\!R}|W_n^{(\nu)}(y)|$ to zero with rate $o(\delta_n^2\alpha_n)$. Similarly one shows that $\sup_{y\in I\!R}|\tilde W_n^{(\nu)}(y)| = o(\delta_n^2\alpha_n)$ for sequences $\delta_n^2 = O(h_n a_n)$ and $\alpha_n = a_n^{\nu+4\delta/3}/h_n^{\nu+1}$. Therefore from those results, (6.11) and (6.15) it follows that

$$\begin{aligned}\sup_{y\in I\!R}|U_n^{(\nu)}(y)| \;&=\; o(\beta_n^{\nu})\Big\{\frac{1}{a_n^{\nu+1}}o\Big(a_n^{\nu+1+4\delta/3}\frac{1}{\beta_n^{\nu}}\Big)+O(\frac{1}{a_n^{\nu}})\Big\}\\ &\quad+O\Big(\frac{h_n^{\nu}}{a_n^{\nu+1}}\Big)o\Big(\frac{a_n^{\nu+1+4\delta/3}}{h_n^{\nu}}\Big)+O(h_n^{\nu+1}) \;=\; o(a_n^{4\delta/3})\end{aligned}$$

$(\nu = 1,\ldots,\kappa-1)$ almost surely by the bandwidth conditions A.8.

Estimation of the remainder term V_n defined in (6.12) is straightforward using (6.13)–(6.14),

$$\sup_{y\in I\!R}|V_n(y)| \;\le\; \frac{1}{a_n^{\kappa+1}}\Big(o(\beta_n^{\kappa})+O(h_n^{\kappa+1})\Big) \;=\; o(a_n^{4\delta/3})$$

almost surely for

$$c_n \;=\; \sqrt{\frac{nh_na_n^{2+2/\kappa+8\delta/(3\kappa)}}{\log h_n^{-1}}} \;\longrightarrow\; \infty$$

by our bandwidth conditions. Hence, assertion (i) follows.

To prove (ii) we consider the cases $|y-z| > a_n^{8/3}$ and $|y-z|\le a_n^{8/3}$ separately. First we obtain

$$\sup_{|y-z|>a_n^{8/3}}\frac{|\tilde f_n(y)-f(y)-\tilde f_n(z)+f(z)|}{|y-z|^{\delta/2}} \;\le\; 2\sup_{y\in I\!R}|\tilde f_n(y)-f(y)|\frac{1}{a_n^{4\delta/3}} \;=\; o(1)$$

by (i). Secondly, we have

$$\begin{aligned}\sup_{|y-z|\le a_n^{8/3}}\frac{|\tilde f_n(y)-\tilde f_n(z)|}{|y-z|^{\delta/2}} \;&=\; \sup_{|y-z|\le a_n^{8/3}}\frac{1}{|y-z|^{\delta/2}}\frac{1}{na_n}\sum_{i=1}^n\Big|k\Big(\frac{y-\tilde\varepsilon_i}{a_n}\Big)-k\Big(\frac{z-\tilde\varepsilon_i}{a_n}\Big)\Big|\\ &\le\; \frac{1}{a_n^2}\sup_{|y-z|\le a_n^{8/3}}|y-z|^{1-\delta/2}\sup_{u\in I\!R}|k'(u)|\end{aligned}$$

by a first order Taylor expansion. By assumption A.7 this gives the rate $O(a_n^{\frac{2}{3}(1-2\delta)}) = o(1)$ because $\delta\in(0,\frac12)$.

Assertion (iii) follows from (i) with help of the Dominated Convergence Theorem. Details are omitted for the sake of brevity. □

Remark 6.4 *Note that from the proof it follows that the last two bandwidth conditions in A.8 are redundant when $k^{(\kappa)}\equiv 0$ and can be replaced by the simpler condition $h_n^{2\nu+2} = O(a_n^{2\nu-1})$ for $\nu = 1,\ldots,\kappa-1$.*

6.3 Bootstrap based regression estimation

We give results about convergence in probability of the bootstrap version of the regression estimator $\hat{m}^*$ and its derivative to the corresponding true functions. These findings will be used in the proof of Theorem 3.2. The statements presented here are different from results in literature concerning the regression estimators based on classical residual bootstrap or other bootstrap methods because they are not only valid conditional on the original sample.
Recall the definition of $\tilde{m}_n$ in (6.2). For the proof of our main theoretical results the convergence of $\hat{m}^* - \tilde{m}_n$ in probability is sufficient. However, with slightly greater effort but exactly following the lines of the proof of Lemma 6.5 it is possible to prove the rate of convergence, i. e. $O_p((\log n/(nh_n))^{1/2})$ under the very same assumptions.

Lemma 6.5 *Under model (2.1) with assumptions A.1–4 and A.7–9 it holds that*
(i) $\sup_{x\in[0,1]} |\hat{m}^*(x) - \tilde{m}_n(x)| = o_p(1)$, **(ii)** $\sup_{x\in[0,1]} |\hat{m}^{*\prime}(x) - \tilde{m}_n'(x)| = o_p(1)$, *and*
(iii) $\sup_{x,t\in[0,1]} \dfrac{|\hat{m}^{*\prime}(x) - \tilde{m}_n'(x) - \hat{m}^{*\prime}(t) + \tilde{m}_n'(t)|}{|x-t|^{\delta/2}} = o_p(1)$. *Here δ denotes the constant from (2.5).*

Note that for the validity of Lemma 6.5 instead of assumptions A.7 and A.8 it is sufficient to assume that $Z_1, Z_2, \ldots$ are centered independent random variables with existing second moment and $a_n = o(1)$. The results presented in the Lemma also hold conditional on the original sample. For instance, for all $\epsilon > 0$ we have

$$P\Big(\sup_{x\in[0,1]} |\hat{m}^*(x) - \tilde{m}_n(x)| > \epsilon \,\Big|\, \mathcal{Y}_n\Big) \;=\; o_p(1)$$

where $\mathcal{Y}_n = \{(X_i, Y_i) \mid i = 1, \ldots, n\}$, which can be deduced from Markov's inequality. Using the methods of the proof of Lemma 6.5, in particular decomposition (6.16), similarly to the results by Härdle and Bowman (1988) it can be shown that a centered version of $\sqrt{nh_n}(\hat{m}^*(x) - \hat{m}(x))$ converges, conditional on the sample $\mathcal{Y}_n$, in distribution to a centered normal distribution.

Proof. To prove (i) we use the definition (3.1) of the bootstrap residuals. The bootstrap observations (3.4) have the representation $Y_i^* = \hat{m}(X_i) + \tilde{\varepsilon}_i^* + a_n Z_i$ and therefore

$$\hat{m}^*(x) - \tilde{m}_n(x) \;=\; A_n(x) + B_n(x) + C_n(x), \tag{6.16}$$

where

$$A_n(x) \;=\; \frac{1}{\hat{f}_X(x)}\frac{1}{nh_n}\sum_{i=1}^n K\Big(\frac{x-X_i}{h_n}\Big)(\hat{m}(X_i) - m_n(X_i)) \tag{6.17}$$

$$B_n(x) \;=\; \frac{1}{\hat{f}_X(x)}\frac{1}{nh_n}\sum_{i=1}^n K\Big(\frac{x-X_i}{h_n}\Big)(m_n(X_i) + a_n Z_i) - \tilde{m}_n(x) \tag{6.18}$$

$$C_n(x) \;=\; \frac{1}{\hat{f}_X(x)}\frac{1}{nh_n}\sum_{i=1}^n K\Big(\frac{x-X_i}{h_n}\Big)\tilde{\varepsilon}_i^*. \tag{6.19}$$

The first term can be estimated as follows,

$$\sup_{x\in[0,1]} |A_n(x)| \ \le\ \sup_{x\in[0,1]} |\hat m(x) - m_n(x)| \ =\ o(h_n^{1+\delta}) \ =\ o(1) \tag{6.20}$$

almost surely by Proposition 6.2. The second term, $B_n(x)$, is the difference between the Nadaraya-Watson estimator from a regression model with regression function m_n and (centered) errors $a_n Z_i$ and $\tilde m_n$ and therefore also converges to zero uniformly almost surely analogous to Proposition 6.2.

The last term, $C_n(x)$ defined in (6.19), is more difficult to handle. Let $\Delta(1),\dots,\Delta(n)$ be independent random variables with Laplace distribution on $\{1,\dots,n\}$ corresponding to the drawing of $\tilde\varepsilon_1^*,\dots,\tilde\varepsilon_n^*$ with replacement from $\tilde\varepsilon_1,\dots,\tilde\varepsilon_n$, such that

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

This yields

$$C_n(x) \ =\ \frac{1}{\hat f_X(x)}\frac{1}{nh_n}\sum_{i=1}^n K\Big(\frac{x-X_i}{h_n}\Big)\Big(\sum_{l=1}^n I\{\Delta(i)=l\}\hat\varepsilon_l - \frac{1}{n}\sum_{j=1}^n \hat\varepsilon_j\Big) \tag{6.21}$$

and we have

$$\begin{aligned}
&\sup_{x\in[0,1]} |C_n(x)| \\
&\le \sup_{x\in[0,1]} \Big|\frac{1}{\hat f_X(x)}\frac{1}{nh_n}\sum_{i=1}^n K\Big(\frac{x-X_i}{h_n}\Big)\Big(\sum_{l=1}^n I\{\Delta(i)=l\}(m(X_l)-\hat m(X_l)) \qquad (6.22)\\
&\qquad - \frac{1}{n}\sum_{j=1}^n (m(X_j)-\hat m(X_j))\Big)\Big| \\
&\quad + \sup_{x\in[0,1]} \Big|\frac{1}{\hat f_X(x)}\frac{1}{nh_n}\sum_{i=1}^n K\Big(\frac{x-X_i}{h_n}\Big)\Big(\sum_{l=1}^n I\{\Delta(i)=l\}\varepsilon_l - \frac{1}{n}\sum_{j=1}^n \varepsilon_j\Big)\Big|, \qquad (6.23)
\end{aligned}$$

where $\sum_{l=1}^n I\{\Delta(i)=l\}=1$ for all $i=1,\dots,n$. By the definition of $\hat f_X$ in (2.4) term (6.22) can be bounded by $2\sup_{x\in[0,1]}|\hat m(x)-m(x)| = o(1)$ almost surely. For (6.23) we only consider the numerator because the denominator converges uniformly almost surely to the design density, which is by assumption bounded away from zero. With the notations

$$\eta_{ni} = \sum_{l=1}^n I\{\Delta(i)=l\}\varepsilon_l \quad \text{and} \quad \overline\varepsilon_n = \frac{1}{n}\sum_{i=1}^n \varepsilon_i \tag{6.24}$$

we have to prove that for all $\epsilon>0$

$$P\Big(\sup_{x\in[0,1]}\Big|\frac{1}{nh_n}\sum_{i=1}^n K\Big(\frac{x-X_i}{h_n}\Big)(\eta_{ni}-\overline\varepsilon_n)\Big| > \epsilon\Big) \ =\ o(1),$$

where, conditional on the sample $\mathcal{Y}_n$, the η_{ni} are randomly drawn with replacement from $\{\varepsilon_1, \ldots, \varepsilon_n\}$. Note that $\max_{i=1,\ldots,n} |\varepsilon_i| = o_p(n^{1/4})$ and by the strong law of large numbers there exists a constant A such that $P(A_n) = o(1)$, where A_n denotes the event that

$$\max_{i=1,\ldots,n} |\varepsilon_i| > An^{1/4} \quad \text{or} \quad \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 > A.$$

Furthermore we assume without restriction that the constant A is a bound for the kernel K and its derivative K' and fulfils

$$\frac{1}{h_n}\int K^2\Big(\frac{x-y}{h_n}\Big) f_X(y)\,dy \;\le\; A$$

for all $x \in [0,1]$. We assume in the following that A_n^c occurs. For a fixed $x \in [0,1]$ define $\xi_i = \frac{1}{h_n}K(\frac{x-X_i}{h_n})(\eta_{ni} - \overline{\varepsilon}_n)$, $V_n = nA^2/h_n$ and $M_n = 2A^2 n^{1/4}/h_n$. Given $\varepsilon_1, \ldots, \varepsilon_n$, the univariate random variables $\xi_1, \ldots, \xi_n$ are independent such that $E[\xi_i \mid \varepsilon_1, \ldots, \varepsilon_n] = 0$, $|\xi_i| \le M_n$, $\sum_{i=1}^n \operatorname{Var}(\xi_i \mid \varepsilon_1, \ldots, \varepsilon_n) \le V_n$. Applying Bernstein's inequality we obtain

$$P\Big(\Big|\frac{1}{nh_n}\sum_{i=1}^n K\Big(\frac{x-X_i}{h_n}\Big)(\eta_{ni} - \overline{\varepsilon}_n)\Big| > \epsilon \;\Big|\; \varepsilon_1, \ldots, \varepsilon_n\Big) \;\le\; 2\exp(-\log n\, w_n)$$

where $w_n^{-1} = 2(V_n + M_n\epsilon n/3)\log n/(n\epsilon)^2 = o(1)$ for all $\epsilon > 0$ by the bandwidth assumption A.4. The sequence w_n is independent of x and not random.

Next we cover $[0,1]$ by n intervals of length $\frac{1}{n}$ with centers x_k. We have

$$\begin{aligned}
&\sup_{x\in[0,1]}\Big|\frac{1}{nh_n}\sum_{i=1}^n K\Big(\frac{x-X_i}{h_n}\Big)(\eta_{ni} - \overline{\varepsilon}_n)\Big| \\
\le\;& \max_{k=1,\ldots,n}\Big|\frac{1}{nh_n}\sum_{i=1}^n K\Big(\frac{x_k-X_i}{h_n}\Big)(\eta_{ni} - \overline{\varepsilon}_n)\Big| \\
&+ \sup_{x,t\in[0,1],|x-t|\le n^{-1}}\Big|\frac{1}{nh_n}\sum_{i=1}^n \Big[K\Big(\frac{x-X_i}{h_n}\Big) - K\Big(\frac{t-X_i}{h_n}\Big)\Big](\eta_{ni} - \overline{\varepsilon}_n)\Big|
\end{aligned}$$

and from the above considerations,

$$P\Big(\max_{k=1,\ldots,n}\Big|\frac{1}{nh_n}\sum_{i=1}^n K\Big(\frac{x_k-X_i}{h_n}\Big)(\eta_{ni} - \overline{\varepsilon}_n)\Big| > \epsilon\Big) \;\le\; 2n\exp(-\log n\, w_n) \;=\; o(1)$$

for all $\epsilon > 0$. Now we estimate using the Mean Value Theorem (still assuming that A_n^c occurs)

$$\begin{aligned}
&\sup_{x,t\in[0,1],|x-t|\le n^{-1}}\Big|\frac{1}{nh_n}\sum_{i=1}^n \Big[K\Big(\frac{x-X_i}{h_n}\Big) - K\Big(\frac{t-X_i}{h_n}\Big)\Big](\eta_{ni} - \overline{\varepsilon}_n)\Big| \\
\le\;& 2A^2 n^{1/4}\frac{1}{h_n^2}\sup_{x,t\in[0,1],|x-t|\le n^{-1}}|x-t| \;=\; O\Big(\frac{n^{1/4}}{h_n^2 n}\Big) \;=\; o(1)
\end{aligned}$$

by assumptions A.3 and A.4. Altogether we can deduce that (6.23) converges to zero almost surely and (i) follows.

The proof of (ii) is similar and is therefore omitted.
For the proof of assertion (iii) using the representation $\hat{m}^{*\prime}(x) - \tilde{m}_n'(x) = A_n'(x) + B_n'(x) + C_n'(x)$ from (6.16) with derivatives of the functions A_n, B_n, C_n defined in (6.17)–(6.19) we have to show that

$$\sup_{x,t\in[0,1]} \frac{|D_n'(x) - D_n'(t)|}{|x-t|^{\delta/2}} \;=\; o_p(1) \tag{6.25}$$

for $D_n = A_n$, $D_n = B_n$ and $D_n = C_n$, respectively. We first concentrate on a proof of (6.25) for $D_n = A_n$. To this end we rewrite $A_n'(x)$ as

$$A_n'(x) \;=\; \frac{1}{\hat{f}_X(x)}\tilde{A}_n'(x) - \frac{\hat{f}_X'(x)}{\hat{f}_X^2(x)}\tilde{A}_n(x)$$

where

$$\tilde{A}_n(x) \;=\; A_n(x)\hat{f}_X(x) \;=\; \frac{1}{nh_n}\sum_{i=1}^n K\Big(\frac{x-X_i}{h_n}\Big)(\hat{m}(X_i) - m_n(X_i)) \tag{6.26}$$

and $\tilde{A}_n'(x)$ denotes the derivative of $\tilde{A}_n(x)$. Assertion (6.25) follows from the estimates

$$\sup_{x,t\in[0,1]} \frac{|\tilde{A}_n(x) - \tilde{A}_n(t)|}{|x-t|^{\delta/2}} \;=\; o_p(1) \tag{6.27}$$

$$\sup_{x,t\in[0,1]} \frac{|\tilde{A}_n'(x) - \tilde{A}_n'(t)|}{|x-t|^{\delta/2}} \;=\; o_p(1) \tag{6.28}$$

$$\sup_{x,t\in[0,1]} \frac{|1/\hat{f}_X(t) - 1/\hat{f}_X(t)|}{|x-t|^{\delta/2}} \;=\; O_p(\frac{1}{h_n^\delta}) \tag{6.29}$$

$$\sup_{x,t\in[0,1]} \frac{|\hat{f}_X'(x)/\hat{f}_X^2(x) - \hat{f}_X'(t)/\hat{f}_X^2(t)|}{|x-t|^{\delta/2}} \;=\; O_p(\frac{1}{h_n^{1+\delta}}) \tag{6.30}$$

$$\sup_{x\in[0,1]} |\tilde{A}_n(x)| \;=\; o_p(h_n^{1+\delta}) \tag{6.31}$$

$$\sup_{x\in[0,1]} |\tilde{A}_n'(x)| \;=\; o_p(h_n^{\delta}), \tag{6.32}$$

where (6.31) and (6.32) can be obtained from the arguments given in the proof of Lemma 6.5 [consider the rates of convergence in (6.20)]. Assertion (6.29) follows from the uniform almost sure convergence of $\hat{f}_X$ to f_X, see (6.5), (where f_X is bounded away from zero) and Proposition 6.1. Similarly, (6.30) also follows from Proposition 6.1. In the following we prove (6.28) in detail. The proof of (6.27) is very similar. Inserting $1 = I\{|x-t| > h_n^2\} + I\{|x-t| \le h_n^2\}$ we have from (6.26)

$$\sup_{x,t\in[0,1]} \frac{|\tilde{A}_n'(x) - \tilde{A}_n'(t)|}{|x-t|^{\delta/2}} \;\le\; \frac{2}{h_n^\delta}\sup_{x\in[0,1]}|\tilde{A}_n'(x)| + \sup_{x\in[0,1]}|\hat{m}(x) - m_n(x)| \sup_{|x-t|\le h_n^2} \frac{A|x-t|^{1-\delta/2}}{h_n^3}$$

by Taylor's expansion, where A denotes a bound for K''. From this we obtain with Proposition 6.2 for the left hand side of (6.28) the rate $o_p(1) + o_p(\beta_n/h_n^{1+\delta}) = o_p(1)$, where $\beta_n = h_n^{1+\delta}$ fulfils

the conditions in Proposition 6.1 by the bandwidth condition (2.5). Hence, assertion (6.25) for $D_n = A_n$ follows.
The proof of (6.25) for $D_n = B_n$ is similar and omitted for the sake of brevity.
For the remaining proof of (6.25) for $D_n = C_n$ we note the decomposition

$$C_n'(x) \;=\; \frac{1}{\hat{f}_X(x)}\tilde{C}_n'(x) - \frac{\hat{f}_X'(x)}{\hat{f}_X^2(x)}\tilde{C}_n(x),$$

where $\tilde{C}_n(x) = C_n(x)\hat{f}_X(x)$ and $\sup_{x\in[0,1]}|\tilde{C}_n(x)| = o_p(h_n^{1+\delta})$, $\sup_{x\in[0,1]}|\tilde{C}_n'(x)| = o_p(h_n^{\delta})$ can be shown following the lines of the proof of (i). Because of (6.29) and (6.30) in the following we concentrate on the proof of

$$\sup_{x,t\in[0,1]}\frac{|\tilde{C}_n'(x)-\tilde{C}_n'(t)|}{|x-t|^{\delta/2}} \;=\; o_p(1)$$

(the analogous statement for $\tilde{C}_n$ is shown in a similar, but simpler manner), where

$$\tilde{C}_n'(x) \;=\; \frac{1}{nh_n^2}\sum_{i=1}^n K'\Big(\frac{x-X_i}{h_n}\Big)(\eta_{ni}-\overline{\varepsilon}_n) + r_n(x)$$

with notations η_{ni}, $\overline{\varepsilon}_n$ from (6.24) and remainder term

$$\begin{aligned} r_n(x) \;&=\; \frac{1}{nh_n^2}\sum_{i=1}^n K'\Big(\frac{x-X_i}{h_n}\Big)\Big(\sum_{l=1}^n I\{\Delta(i)=l\}(m(X_l)-\hat{m}(X_l)) - \frac{1}{n}\sum_{j=1}^n(m(X_j)-\hat{m}(X_j))\Big) \\ &=\; \frac{1}{nh_n^2}\sum_{i=1}^n K'\Big(\frac{x-X_i}{h_n}\Big)\Big(z_{ni} + h_n\Big(\sum_{l=1}^n I\{\Delta(i)=l\}C_{nl}t_{nl} - \frac{1}{n}\sum_{j=1}^n C_{nj}t_{nj}\Big)\Big). \end{aligned}$$

Here we have applied a similar decomposition as in (6.9) with an almost surely uniformly in $i=1,\ldots,n$ bounded sequence C_{ni}, $t_{ni} = I\{X_i < h_n\} + I\{X_i > 1-h_n\}$ and where z_{ni} has uniformly in i the rate $o(\beta_n)+O(h_n^2)$. The sequence β_n fulfils the assumptions in Lemma 6.1 and we set $\beta_n = h_n^{1+\delta}$. For the remainder term we estimate by $1 = I\{|x-t|>h_n^2\} + I\{|x-t|\le h_n^2\}$, applying considerations similar to the proof of Proposition 6.1 to the first arising term and a Taylor expansion to the second term,

$$\sup_{x,t\in[0,1]}\frac{|r_n(x)-r_n(t)|}{|x-t|^{\delta/2}} \;\le\; \frac{1}{h_n^\delta}O_p\big(\frac{1}{h_n}\big)\Big(o_p(\beta_n)+O_p(h_n^2)+O_p(h_n^2)\Big) \;=\; o_p(1).$$

Next, we have to prove that for all $\epsilon>0$

$$P\Big(\sup_{x,t\in[0,1]}\Big|\frac{1}{nh_n^2}\sum_{i=1}^n\Big(K'\Big(\frac{x-X_i}{h_n}\Big)-K'\Big(\frac{t-X_i}{h_n}\Big)\Big)\frac{(\eta_{ni}-\overline{\varepsilon}_n)}{|x-t|^{\delta/2}}\Big| > \epsilon\Big) \;=\; o(1).$$

We concentrate all estimations on the event A_n^c that was defined before, where additionally f_X is bounded by A. For any fixed pair $(x,t)\in[0,1]^2$ we apply Bernstein's inequality with definitions $\xi_i = \frac{1}{h_n^2}(K'(\frac{x-X_i}{h_n}) - K'(\frac{t-X_i}{h_n}))(\eta_{ni}-\overline{\varepsilon}_n)/|x-t|^{\delta/2}$, $V_n = 4nA^3/h_n^{3+2\delta}$, $M_n = 2A^2n^{1/4}/h_n^{2+\delta/2}$, such that $|\xi_i| \le M_n$ and $\sum_{i=1}^n \mathrm{Var}(\xi_i \mid \varepsilon_1, \dots, \varepsilon_n) \le V_n$. Then conditional on $\varepsilon_1, \dots, \varepsilon_n$ applying Bernstein's inequality we obtain

$$P\Big(\Big|\frac{1}{nh_n^2}\sum_{i=1}^n \Big[K'\Big(\frac{x-X_i}{h_n}\Big) - K'\Big(\frac{t-X_i}{h_n}\Big)\Big]\frac{(\eta_{ni}-\overline{\varepsilon}_n)}{|x-t|^{\delta/2}}\Big| > \epsilon \;\Big|\; \varepsilon_1, \dots, \varepsilon_n\Big) \;\le\; 2\exp(-\log n\, w_n)$$

where $w_n^{-1} = O(\log n/(nh_n^{3+2\delta})) + O(\log n/(n^{3/4}h_n^{2+\delta/2})) = o(1)$ by assumption (2.5) on the bandwidth, for all $\epsilon > 0$. The sequence w_n is independent of (x,t) and not random.
Next we cover $[0,1]^2$ by n^{2d} squares with edge length $\frac{1}{n^d}$ with centers (x_k, t_l), where $d > 0$ will be further specified in a moment. We have

$$\begin{aligned}
&\sup_{x,t\in[0,1]}\Big|\frac{1}{nh_n^2}\sum_{i=1}^n \Big[K'\Big(\frac{x-X_i}{h_n}\Big) - K'\Big(\frac{t-X_i}{h_n}\Big)\Big]\frac{(\eta_{ni}-\overline{\varepsilon}_n)}{|x-t|^{\delta/2}}\Big| \\
\le\;& \max_{k=1,\dots,n^d, l=1,\dots,n^d}\Big|\frac{1}{nh_n^2}\sum_{i=1}^n \Big[K'\Big(\frac{x_k-X_i}{h_n}\Big) - K'\Big(\frac{t_l-X_i}{h_n}\Big)\Big]\frac{(\eta_{ni}-\overline{\varepsilon}_n)}{|x-t|^{\delta/2}}\Big| \\
&+ \sup_{\substack{x,t,y,z\in[0,1]\\ |x-y|\le n^{-d}, |t-z|\le n^{-d}}}\Big|\frac{1}{nh_n^2}\sum_{i=1}^n \Big(\frac{K'(\frac{x-X_i}{h_n}) - K'(\frac{t-X_i}{h_n})}{|x-t|^{\delta/2}} - \frac{K'(\frac{y-X_i}{h_n}) - K'(\frac{z-X_i}{h_n})}{|y-z|^{\delta/2}}\Big)(\eta_{ni}-\overline{\varepsilon}_n)\Big|
\end{aligned}$$

and from the above considerations the first term is of order $o_p(1)$ for all $d > 0$ because $n^{2d}\exp(-\log n\, w_n) = o(1)$. The derivation of a bound for the second term is a bit cumbersome. Still assuming that A_n^c occurs we can bound it by

$$\begin{aligned}
&\frac{2An^{1/4}}{nh_n^2}\sum_{i=1}^n\Big\{\sup_{\substack{|x-t|\le|y-z|\\ |x-y|\le n^{-d}, |t-z|\le n^{-d}}}\Big|\frac{1}{|x-t|^{\delta/2}} - \frac{1}{|y-z|^{\delta/2}}\Big|\Big|K'\Big(\frac{x-X_i}{h_n}\Big) - K'\Big(\frac{t-X_i}{h_n}\Big)\Big| \\
&+ \sup_{\substack{|x-t|\le|y-z|\le n^{-d}\\ |x-y|\le n^{-d}, |t-z|\le n^{-d}}}\frac{1}{|y-z|^{\delta/2}}\Big|K'\Big(\frac{x-X_i}{h_n}\Big) - K'\Big(\frac{t-X_i}{h_n}\Big) - K'\Big(\frac{y-X_i}{h_n}\Big) + K'\Big(\frac{z-X_i}{h_n}\Big)\Big| \\
&+ \sup_{\substack{|x-t|\le|y-z|, |y-z|>n^{-d}\\ |x-y|\le n^{-d}, |t-z|\le n^{-d}}}\frac{1}{|y-z|^{\delta/2}}\Big|K'\Big(\frac{x-X_i}{h_n}\Big) - K'\Big(\frac{t-X_i}{h_n}\Big) - K'\Big(\frac{y-X_i}{h_n}\Big) + K'\Big(\frac{z-X_i}{h_n}\Big)\Big|\Big\} \\
\le\;& \frac{2A^2n^{1/4}}{h_n^3}\Big\{\sup_{\substack{|x-t|\le|y-z|\\ |x-y|\le n^{-d}, |t-z|\le n^{-d}}}|x-t-y+z|^{\delta/2}\frac{|x-t|}{|x-t|^{\delta/2}|y-z|^{\delta/2}} \\
&+ \sup_{\substack{|x-t|\le|y-z|\le n^{-d}\\ |x-y|\le n^{-d}, |t-z|\le n^{-d}}}\frac{1}{|y-z|^{\delta/2}}\Big(|x-t| + |y-z|\Big) \\
&+ \sup_{\substack{|x-t|\le|y-z|, |y-z|>n^{-d}\\ |x-y|\le n^{-d}, |t-z|\le n^{-d}}}\frac{1}{|y-z|^{\delta/2}}\Big(|x-y| + |t-z|\Big)\Big\} \;\le\; 2A^2n^{1/4}h_n^{-3}(n^{-d\delta/2} + 4n^{-d(1-\delta/2)})
\end{aligned}$$

(note that $0 < \delta < 1$ by assumption A.4). This sequence is of order $o(1)$ for a sufficiently large d by the bandwidth assumptions. □

6.4 More auxiliary results

For a constant $L > 0$ let $C_L^{1+\beta}[0,1]$ be defined as the set of differentiable functions $g : [0,1] \to I\!R$ with derivatives g' such that

$$\max\Big\{ \sup_{x\in[0,1]} |g(x)|, \sup_{x\in[0,1]} |g'(x)| \Big\} + \sup_{x,t\in[0,1]} \frac{|g'(x) - g'(t)|}{|x-t|^{\alpha}} \;\le\; L\,,$$

see van der Vaart and Wellner (1996, p. 154). See this reference also for defintions of covering and bracketing numbers.

Lemma 6.6 *Let $\mathcal{H}$ denote the class of continuously differentiable, increasing functions $H : I\!R_0^+ \to [0,1]$ with uniformly bounded derivative h, such that*

$$\sup_{z\in I\!R^+} |h(z)| + \sup_{z,z'\in I\!R^+} \frac{|h(z) - h(z')|}{|z-z'|^{\beta}} \le L$$

and $|1 - H(x)| \le d/x^{\upsilon}$ for all $x \in I\!R^+$, where $\upsilon > 1 + \frac{1}{\beta}$ and the constant d is independent of $H \in \mathcal{H}$. Then, for the covering number with respect to the supremum norm,

$$\log N(\epsilon, \mathcal{H}, ||\cdot||_\infty) = O(\epsilon^{\frac{1}{1+a}})$$

is valid, where $a = (\beta(\upsilon - 1) - 1)/(1 + \beta + \upsilon) > 0$.

Proof. Let $\epsilon > 0$ and $j_0 = (d/\epsilon)^{1/\upsilon}$. Further let $I_j = [j, j+1)$ such that $I\!R^+ = \cup_{j=0}^{\infty} I_j$. For $j = 0, \ldots, j_0 - 1$ the class $\mathcal{H}_j = \{H|_{\overline{I_j}} \mid H \in \mathcal{H}\}$ is a subset of $C_1^{1+\beta}(\overline{I_j})$, where $\overline{I_j} = [j, j+1]$, and therefore, there exists an ϵ-covering with respect to the supremum-norm with centers $H_{j1}, \ldots, H_{jp_j}$ where $\log p_j \le K\epsilon^{-1/(1+\beta)}$ with constant K independent of j (and $p_j = 1$ for $\epsilon > 1$). For $j \ge j_0$ we choose $p_j = 1$ and $H_j \equiv I_{I_j}$ (indicator function on interval I_j). Any $H \in \mathcal{H}$ can be written as the sum of HI_j, $j \in I\!N_0$, and for $HI_{\overline{I_j}} \in \mathcal{H}_j$ $(j = 0, \ldots, j_0 - 1)$ there exists a H_{ji_j} such that $\sup_{x\in I_j} |H(x) - H_{ji_j}(x)| \le \epsilon$. For $j \ge j_0$ set $H_{ji_j} = H_j = I_{I_j}$. Then, we have

$$\begin{aligned} \sup_{x\in I\!R^+} \Big| H(x) - \sum_{j=0}^{\infty} H_{ji_j}(x) \Big| \;&\le\; \max\Big\{ \sup_{x\in I\!R^+} \sum_{j=0}^{j_0-1} |H(x) - H_{ji_j}(x)| I\{x \in I_j\}\;,\; \sup_{x\ge j_0} |1 - H(x)| \Big\} \\ &\le\; \max\{\epsilon, d/j_0^{\upsilon}\} \;=\; \epsilon. \end{aligned}$$

Thus, we have found an ϵ-net to cover $\mathcal{H}$. It contains $\prod_{j=0}^{\infty} p_j$ functions, where

$$\log \prod_{j=0}^{\infty} p_j \;=\; \sum_{j=0}^{j_0} \log p_j \;\le\; K j_0 \epsilon^{-1/(1+\beta)} \;=\; O(\epsilon^{-(\frac{1}{1+\beta} + \frac{1}{\upsilon})}) \;=\; O(\epsilon^{\frac{1}{1+a}}). \qquad \square$$

We define a sequence of function classes by

$$\begin{aligned} \mathcal{G}_n \;=\; &\Big\{ [0,1]^2 \to I\!R, (x,u) \mapsto I\{u \le H(y + g(x) + \kappa_n(x))\} \;\Big|\; \\ & y \in I\!R, g \in C_1^{1+\beta}[0,1], H \in \mathcal{H} \Big\}. \end{aligned} \tag{6.33}$$

For some constant L, $\mathcal{H}$ denotes the class of continuously differentiable distribution functions $H : I\!R \to [0,1]$ with uniformly bounded derivative h, such that

$$\sup_{z \in I\!R} |h(z)| + \sup_{z,z' \in I\!R} \frac{|h(z) - h(z')|}{|z - z'|^\beta} \leq L$$

and the tail condition

$$|1 - H(x)| \leq d/x^\upsilon \text{ for all } x \in I\!R^+, \ |H(x)| \leq d/|x|^\upsilon \text{ for all } x \in I\!R^- \tag{6.34}$$

is satisfied for some $\upsilon > 1 + \frac{1}{\beta}$, where the constant d is independent of $H \in \mathcal{H}$. The function κ_n is deterministic and converges to zero uniformly. Later we will set $\beta = \delta/2$ with δ from (2.5) and

$$L \ = \ 2\max\{1, \sup_{y \in I\!R} |f(y)| + \sup_{y \in I\!R} |f'(y)|\}. \tag{6.35}$$

Proposition 6.7 *The function classes $\mathcal{G}_n$ defined in (6.33) fulfils the conditions*

$$\int_0^{\delta_n} \sqrt{\log N_{[]}(\epsilon, \mathcal{G}_n, L_2(P))}\, d\epsilon \ \longrightarrow \ 0 \tag{6.36}$$

for every real sequence $\delta_n \to 0$. Here P denotes the probability distribution of (X_1, U_1), where X_1 has distribution F_X, U_1 is uniformly distributed in $[0,1]$ and they are independent. Further we have

$$\sup_{\delta_n} E\Big[\Big(\varphi_{n,H,y,g}(U_1, X_1) - \varphi_{n,\tilde{H},\tilde{y},\tilde{g}}(U_1, X_1)\Big)^2\Big] \ \longrightarrow \ 0 \tag{6.37}$$

for every real sequence $\delta_n \to 0$, where $\varphi_{n,H,y,g}$ denotes the element of $\mathcal{G}_n$ determined by H, y, g and the supremum is taken over $\rho((H,y,g),(\tilde{H},\tilde{y},\tilde{g})) < \delta_n$ for some suitably chosen semimetric ρ. By Theorem 2.11.23, van der Vaart and Wellner (1996, p. 221) the process

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(\varphi_{n,H,y,g}(U_i, X_i) - E[\varphi_{n,H,y,g}(U_1, X_1)]\Big), y \in I\!R, g \in C_1^{1+\beta}[0,1], H \in \mathcal{H}$$

converges to a centered Gaussian process.

Proof. The proof of (6.37) is similar to Lemma B.1 by Akritas and Van Keilegom (2001), but in their proof, $\mathcal{H}$ has only one element and κ_n is zero. We give a bound for the bracketing number $N_{[]}(\gamma, \mathcal{G}_n, L_2(P))$. For $\gamma \geq 1$ one bracket for $\mathcal{G}_n$ is sufficient. Now, let $\gamma \in (0,1)$ and let $\epsilon = \gamma/(L+3)^{1/2}$. From Lemma 6.6 (and an analogous version for $I\!R^-$) we have an $\epsilon^2/2$–net for $\mathcal{H}$ so that $c = N(\epsilon^2/2, \mathcal{H}, ||\cdot||_\infty) \leq O(\exp(\epsilon^{2/(1+a)}))$ for some $a > 0$ [here we need the assumption that $\upsilon > 1 + \frac{1}{\beta}$ in tail condition (6.34)]. Denote the centers by $\tilde{H}_j$, $j = 1, \ldots, c$. Brackets for $\mathcal{H}$ of length ϵ^2 are then given by

$$\Big[H_j^L, H_j^U\Big] \ = \ \Big[\tilde{H}_j - \frac{\epsilon^2}{2}, \tilde{H}_j + \frac{\epsilon^2}{2}\Big], \quad j = 1, \ldots, c.$$

Here and in the following the indices L and U stand for "lower" and "upper". We further have brackets $[g_{1k}^L, g_{1k}^U]$ for the function class $C_1^{1+\beta}[0,1]$, where $k = 1, \ldots, K_1 = O(\exp(\epsilon^{2/(1+\beta)}))$, [see Corollary 2.7.2, van der Vaart and Wellner (1996, p. 157)]. All brackets introduced so far are with respect to the supremum norm. Let us consider $y \in I\!R^+$ for the sake of simplicity and assume $\kappa_n \geq 0$. Because the functions $H \in \mathcal{H}$ are increasing we assume the same for H_j^L, H_j^U and have for $H_j^L \leq H \leq H_j^U$, $g_{1k}^L \leq g_1 \leq g_{1k}^U$,

$$\begin{aligned} I\Big\{u \leq H_j^L(y + g_{1k}^L(x) + \kappa_n(x))\Big\} \;&\leq\; I\Big\{u \leq H(y + g_1(x) + \kappa_n(x))\Big\} \\ &\leq\; I\Big\{u \leq H_j^U(y + g_{1k}^U(x) + \kappa_n(x))\Big\} \end{aligned}$$

for all $u, x \in [0,1]$, $y \in I\!R^+$. For each pair (j,k) we partition $I\!R^+$ in segments $[y_{jk\varrho}, y_{jk,\varrho+1}] = [y_{jk\varrho}^L, y_{jk\varrho}^U]$, $\varrho = 1, \ldots, R = O(\epsilon^{-2})$ with probability

$$\int \Big(H_j^L(y_{jk\varrho}^U + g_{1k}^L(x) + \kappa_n(x)) - H_j^L(y_{jk\varrho}^L + g_{1k}^L(x) + \kappa_n(x))\Big) f_X(x)\,dx \;\leq\; \epsilon^2. \tag{6.38}$$

Note that these segments may depend on n, but the number of segments does not. We have at most

$$cK_1R \;=\; O(\epsilon^{-2}\exp(\epsilon^{2/(1+\beta)})\exp(\epsilon^{2/(1+a)}))$$

brackets of the form

$$\Big[I\Big\{u \leq H_j^L(y_{jk\varrho}^L + g_{1k}^L(x) + \kappa_n(x))\Big\}, I\Big\{u \leq H_j^U(y_{jk\varrho}^U + g_{1k}^U(x) + \kappa_n(x))\Big\}\Big]$$

that cover $\mathcal{G}$ and are of $L_2(P)$–length

$$\Big(\int \Big|H_j^U(y_{jk\varrho}^U + g_{1k}^U(x) + \kappa_n(x)) - H_j^L(y_{jk\varrho}^L + g_{1k}^L(x) + \kappa_n(x))\Big| f_X(x)\,dx\Big)^{1/2}$$

$$\leq \Big(\int \Big|H_j^U(y_{jk\varrho}^U + g_{1k}^U(x) + \kappa_n(x)) - H(y_{jk\varrho}^U + g_{1k}^U(x) + \kappa_n(x))\Big| f_X(x)\,dx \tag{6.39}$$

$$+ \int \Big|H(y_{jk\varrho}^U + g_{1k}^U(x) + \kappa_n(x)) - H(y_{jk\varrho}^U + g_{1k}^L(x) + \kappa_n(x))\Big| f_X(x)\,dx \tag{6.40}$$

$$+ \int \Big|H(y_{jk\varrho}^U + g_{1k}^L(x) + \kappa_n(x)) - H_j^L(y_{jk\varrho}^U + g_{1k}^L(x) + \kappa_n(x))\Big| f_X(x)\,dx \tag{6.41}$$

$$+ \int \Big|H_j^L(y_{jk\varrho}^U + g_{1k}^L(x) + \kappa_n(x)) - H_j^L(y_{jk\varrho}^L + g_{1k}^L(x) + \kappa_n(x))\Big| f_X(x)\,dx\Big)^{1/2}, \tag{6.42}$$

where $H \in \mathcal{H}$ is any function within the bracket $[H_j^L, H_j^U]$ with derivative h. We consider terms (6.39)–(6.42) inside the square root. We can bound (6.39) and (6.41) by $\sup_{z \in I\!R} |H_j^U(z) - H_j^L(z)| \leq \epsilon^2$. Term (6.40) can be bounded by

$$\sup_{z \in I\!R} |h(z)| \sup_{x \in [0,1]} |g_{1k}^U(x) - g_{1k}^L(x)| \;\leq\; L\epsilon^2.$$

Term (6.42) is less than or equal to ϵ^2 by (6.38). Collecting the bounds together, the brackets have $L_2(P)$–length less than or equal to $(L+3)^{1/2}\epsilon = \gamma$. We have shown,

$$\begin{aligned} \log N_{[]}(\gamma, \mathcal{G}_n, L_2(P)) \;&=\; O(\log(\epsilon^{-2}\exp(2\epsilon^{2/(1+\beta)})\exp(\epsilon^{2/(1+a)}))) \\ &=\; O((\log\gamma) + \gamma^{2/(1+\beta)} + \gamma^{2/(1+a)}), \end{aligned}$$

where this number does not depend on n, and, hence, the bracketing condition (6.36) is fulfilled. To show (6.37) define $T = \{(H, y, g) \mid H \in \mathcal{H}, y \in I\!R, g \in C_1^{1+\beta}[0,1]\}$ and the semimetric ρ as

$$\rho((H, y, g), (\tilde{H}, \tilde{y}, \tilde{g})) \;=\; \int |H(y + g(x)) - \tilde{H}(\tilde{y} + \tilde{g}(x))| f_X(x)\, dx.$$

Then the bracketing number $N_{[]}(\epsilon, T, \rho)$ is finite for all $\epsilon > 0$, which can be shown very similarly to the proof of (6.36). Hence, the semimetric space (T, ρ) is totally bounded; see van der Vaart and Wellner (1996, p. 84). Now the supremum in (6.37) can be bounded by

$$2 \sup_{\delta_n} \int |H(y + g(x) + \kappa_n(x)) - \tilde{H}(\tilde{y} + \tilde{g}(x) + \kappa_n(x))| f_X(x)\, dx.$$

By the Mean Value Theorem applied to H and $\tilde{H}$ we obtain the bound

$$2\Big(\sup_{\delta_n} \rho((H, y, g), (\tilde{H}, \tilde{y}, \tilde{g})) + 2 \sup_{H \in \mathcal{H}} \sup_{z \in I\!R} |h(z)| \int |\kappa_n(x)| f_X(x)\, dx \Big)$$

which converges to zero. □

Proposition 6.8 *Under model (2.1) with assumptions A.1–4 and A.7–9 the distribution function $\tilde{F}_n$ of the residual-based density estimator $\tilde{f}_n$ satisfies the tail condition, namely there exists a constant d such that $|1 - \tilde{F}_n(x)| \le d/x^{\upsilon}$ for all $x \in I\!R^+$, $|\tilde{F}_n(x)| \le d/|x|^{\upsilon}$ for all $x \in I\!R^-$, with probability converging to 1.*

Proof. We have for $x > 0$ (analogous considerations for $x < 0$)

$$\begin{aligned} 0 \;&\le\; x^{\upsilon}(1 - \tilde{F}_n(x)) \;\le\; x^{\upsilon} \int_x^{\infty} \tilde{f}_n(t)\, dt \;\le\; \int_x^{\infty} t^{\upsilon} \tilde{f}_n(t)\, dt \\ &= \frac{1}{n} \sum_{j=1}^{n} \int_x^{\infty} t^{\upsilon} \frac{1}{a_n} k\Big(\frac{t - \tilde{\varepsilon}_j}{a_n}\Big)\, dt \;\le\; \frac{1}{n} \sum_{j=1}^{n} \int k(u) |a_n u + \tilde{\varepsilon}_j|^{\upsilon}\, du. \end{aligned}$$

The assertion follows from (6.13) and assumption A.10. □

Remark 6.9 *For the proof of Theorem 3.2 we need the probability that $\tilde{F}_n$ belongs to some function class $\mathcal{H}$ with certain properties to converge to one. This was shown in Proposition 6.8 for the function class $\mathcal{H}$ defined below (6.33). Different function classes $\mathcal{H}$ could be chosen provided that $N_{[]}(\epsilon, \mathcal{H}, ||\cdot||_{\infty}) = O(\exp(\epsilon^{1/(1+a)}))$ for some $a > 0$ as in Lemma 6.6 is valid. In assumption A.10 we give simple conditions so that Proposition 6.8 is valid and the function class $\mathcal{H}$ defined below (6.33) and used in Proposition 6.7 fulfils the above bracketing number condition. However, the moment assumption $E|\varepsilon_1|^{\upsilon} < \infty$ for $\upsilon > 1 + 2/\delta$ is quite restrictive, because then the bandwidth constant δ from (2.5) cannot be chosen arbitrarily small. However, some weakening of the assumption is possible. First of all the simple proof of Proposition 6.8 contains a rather crude estimation, and more complicated proofs of the same assertion under less restrictive assumptions might be possible. Further, different definitions of the function class $\mathcal{H}$ might lead to the destination as well provided that $\tilde{F}_n$ belongs to that class asymptotically*

and it fulfils the required bracketing condition. In assumption A.10 we give simple sufficient, albeit restrictive, conditions, but these can very probably be weakened.
There is another method to dispose of the moment assumption A.10 completely. When a local linear estimator $\hat{m}$ or boundary kernels are used such that

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

(compare Proposition 6.2) is valid, and the analogous estimator is used for the bootstrap sample, then in the proof of Lemma 3.1 (i) the decomposition of $\hat{m}^* - \hat{m}$ in $\hat{g}_n = \hat{m}^* - \tilde{m}_n - (\hat{m} - m_n)$ and $\kappa_n = \tilde{m}_n - m_n$ is not necessary. Instead we set $\hat{g}_n = \hat{m}^* - \hat{m}$. The function class $\mathcal{G}$ to be considered is then independent of n and defined as $\mathcal{G}_n$ in (6.33), but setting $\kappa_n = 0$. Then no tail condition for the function class $\mathcal{H}$ is required and, hence, Proposition 6.8 as well as assumption A.10 are dispensable; compare Neumeyer (2006) for details.

7 Proofs of the main results

Different presentations of the process defined in (3.5) like

$$\begin{aligned} R_n^*(y) \;&=\; \frac{1}{\sqrt{n}}\sum_{i=1}^n (I\{\varepsilon_i^* \le y + \hat{m}^*(X_i) - \hat{m}(X_i)\} - \tilde{F}_n(y)) \\ &=\; \frac{1}{\sqrt{n}}\sum_{i=1}^n (I\{U_i \le \tilde{F}_n(y + \hat{m}^*(X_i) - \hat{m}(X_i))\} - \tilde{F}_n(y)) \end{aligned} \tag{7.1}$$

will turn out to be useful in the proof of Theorem 3.2. Here we use the representation $\varepsilon_i^* = \tilde{F}_n^{-1}(U_i)$, where $U_1, \ldots, U_n \sim U[0,1]$ are independent and uniformly distributed random variables in $[0,1]$ independent of the sample $\mathcal{Y}_n$. It should be mentioned that by definition $U_i = U_{n,i} = \tilde{F}_n(\varepsilon_i^*)$ the $U_{n,1}, \ldots, U_{n,n}$ depend on $\mathcal{Y}_n$ and form a triangular array, because with each n the whole tupel changes. However, it is easy to see that for each n the tupel $(U_{n,1}, \ldots, U_{n,n})$ has the same distribution as $(U_1, \ldots, U_n)$, where $U_1, U_2, \ldots$ denotes a sequence of independent and identically $U[0,1]$ distributed random variables. In particular the distribution does not depend on $\mathcal{Y}_n$. In results and proofs where we are only interested in convergence in distribution we can therefore use the representation (7.1).

Proof of Lemma 3.1 (i). From the presentation (7.1) one obtains

$$\begin{aligned} V_n^*(y) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(& I\{U_i \le \tilde{F}_n(y + \hat{m}^*(X_i) - \hat{m}(X_i))\} - I\{U_i \le \tilde{F}_n(y)\} \\ & - \int \tilde{F}_n(y + \hat{m}^*(x) - \hat{m}(x)) f_X(x)\,dx + \tilde{F}_n(y)\Big). \end{aligned}$$

We further introduce the notations $\hat{m}^* - \hat{m} = \hat{g}_n + \kappa_n$, where $\hat{g}_n = \hat{m}^* - \tilde{m}_n - (\hat{m} - m_n)$, $\kappa_n = \tilde{m}_n - m_n$, and m_n and $\tilde{m}_n$ are defined in (6.1) and (6.2), respectively. Note that κ_n is a deterministic function converging to zero uniformly by (6.4). Let $\delta > 0$ denote the constant from bandwidth condition (2.5). For any $\epsilon > 0$ we have

$$P\Big(\sup_{y\in I\!R}|V_n^*(y)| > \epsilon\Big) \;\le\; P\Big(\sup_{y\in I\!R}|V_n^*(y)| > \epsilon, \tilde F_n \in \mathcal{H}, \hat g_n \in C_1^{1+\delta/2}[0,1]\Big) \\ + P\Big(\tilde F_n \notin \mathcal{H}\Big) + P\Big(\hat g_n \notin C_1^{1+\delta/2}[0,1]\Big)$$

where $\mathcal{H}$ for $\beta = \delta/2$ and the Donsker class of smooth functions $C_1^{1+\alpha}[0,1]$ are defined in section 6.4. The last two probabilities converge to zero according to Propositions 6.8 and 6.2 and Lemmata 6.3 and 6.5; note that $C_1^{1+\delta}[0,1] \subseteq C_1^{1+\delta/2}[0,1]$ and that the error distribution F fulfils $\sup_{y\in I\!R} F(y) \le 1$ and $\sup_{z,z'\in I\!R} |f(z) - f(z')|/|z - z'|^\beta \le L$; compare the definition of $\mathcal{H}$ and (6.35). Therefore in what follows we assume that $\tilde F_n \in \mathcal{H}$ and $\hat g_n \in C_1^{1+\delta/2}[0,1]$. We consider the following empirical process,

$$V_n(y,H,g) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{U_i \le H(y + g(X_i) + \kappa_n(X_i))\} - I\{U_i \le H(y)\} \\ - \int H(y + g(x) + \kappa_n(x)) f_X(x)\, dx + H(y)\Big)$$

indexed by $y \in I\!R$, $g \in C_1^{1+\delta/2}[0,1]$ and $H \in \mathcal{H}$. The random variables $(X_1, U_1), (X_2, U_2), \ldots$ are independent and identically distributed, where X_i has distribution function F_X and U_i is uniformly distributed in $[0,1]$ (independent of X_i). We have

$$E\Big[I\{U_i \le H(y + g(X_i) + \kappa_n(X_i))\} - I\{U_i \le H(y)\}\Big] \\ = \int H(y + g(x) + \kappa_n(x)) f_X(x)\, dx - H(y)$$

and according to Proposition 6.7 the process $V_n(y,F,g)$ converges weakly to a Gaussian process. Applying a uniform version of Lemma 19.24 by van der Vaart (1998, p. 280) [compare the proof of Theorem 19.26 in that reference], we insert the random functions $g = \hat g_n$ and $H = \tilde F_n$. To this end we calculate (remember that $\hat g_n + \kappa_n = \hat m^* - \hat m$)

$$\sup_{y\in I\!R}\int_0^1\int_0^1 \Big(I\{u \le \tilde F_n(y + \hat m^*(z) - \hat m(z))\} - I\{u \le \tilde F_n(y)\} \\ - \int_0^1 \tilde F_n(y + \hat m^*(x) - \hat m(x)) f_X(x)\, dx + \tilde F_n(y)\Big)^2 f_X(z)\, dz\, du \\ \le \sup_{y\in I\!R}\int_0^1\int_0^1 \Big(I\{u \le \tilde F_n(y + \hat m^*(z) - \hat m(z))\} - I\{u \le \tilde F_n(y)\}\Big)^2 f_X(z)\, dz\, du \\ = \sup_{y\in I\!R}\int_0^1 \Big|\tilde F_n(y + \hat m^*(z) - \hat m(z)) - \tilde F_n(y)\Big| f_X(z)\, dz.$$

A Taylor expansion of order one yields the following bound for the last line,

$$\sup_{y\in I\!R}|\tilde f_n(y)| \int_0^1 (|\hat m^*(z) - m(z)| + |\hat m(z) - m(z)|) f_X(z)\, dz$$

and this converges to zero in probability by Lemmata 6.3 and 6.5 and (6.4). It can be shown now applying the mentioned results by van der Vaart (1998) that $V_n^*(y) = V_n(y, \tilde{F}_n, \hat{g}_n)$ converges to zero in probability, uniformly with respect to $y \in I\!R$. □

Proof of Lemma 3.1 (ii). We start the proof by Taylor's expansion of $\tilde{F}_n$ to obtain

$$\sqrt{n}\int \Big(\tilde{F}_n(y+\hat{m}^*(x)-\hat{m}(x))-\tilde{F}_n(y)\Big) f_X(x)\,dx$$

$$= \tilde{f}_n(y)\sqrt{n}\int(\hat{m}^*(x)-\hat{m}(x))f_X(x)\,dx \tag{7.2}$$

$$+\sqrt{n}\int \tilde{f}_n'(\xi_{y,x,n})(\hat{m}^*(x)-\hat{m}(x))^2 f_X(x)\,dx \tag{7.3}$$

for some $\xi_{y,x,n}$ between y and $y+\hat{m}^*(x)-\hat{m}(x)$. The absolute value of (7.3) can be bounded as follows, uniformly in $y \in I\!R$,

$$\Big(\sup_{y\in I\!R}|f'(y)| + \sup_{y\in I\!R}|\tilde{f}_n'(y)-f'(y)|\Big)\sqrt{n}\int(\hat{m}^*(x)-\hat{m}(x))^2 f_X(x)\,dx.$$

The term in parantheses converges almost surely to $\sup_{y\in I\!R}|f'(y)|$ by Lemma 6.3. From (6.16) we obtain

$$\begin{aligned}\sqrt{n}\int(\hat{m}^*(x)-\hat{m}(x))^2 f_X(x)\,dx \;&\le\; 2\sqrt{n}\int(\tilde{m}_n(x)-m_n(x))^2 f_X(x)\,dx\\ &\quad+4\sqrt{n}\int(\hat{m}(x)-m_n(x))^2 f_X(x)\,dx\\ &\quad+16\sqrt{n}\int(A_n^2(x)+B_n^2(x)+C_n^2(x))f_X(x)\,dx\end{aligned}$$

with definitions of A_n, B_n, C_n from (6.17)–(6.19), m_n from (6.1) and $\tilde{m}_n$ from (6.2). The first integral on the right-hand side is of order $O_p(\sqrt{n}h_n^2) = o_p(1)$ by (6.4). We can further bound the second integral on the right-hand side as well as $\sqrt{n}\int A_n^2(x)f_X(x)\,dx$ by $\sqrt{n}(\sup_{x\in[0,1]}|\hat{m}(x)-m_n(x)|)^2 = \sqrt{n}o(\beta_n^2) = o(1)$ almost surely by Lemma 6.2 with $\beta_n = c_n(\log h_n/(nh_n))^{1/2}$, where $c_n^2 = \sqrt{n}h_n/\log(h_n^{-1})$ tends to infinity by assumption A.4. Similar estimations apply for the integral $\sqrt{n}\int B_n^2(x)f_X(x)\,dx$ because B_n is the term corresponding to A_n but built from a regression model with regression function m_n and errors a_nZ_i. Further we have from (6.21)

$$\begin{aligned}C_n(x) \;&=\; \frac{1}{\hat{f}_X(x)}\frac{1}{nh_n}\sum_{i=1}^n K\Big(\frac{X_i-x}{h_n}\Big)\eta_{ni} - \frac{1}{n}\sum_{i=1}^n\varepsilon_i + \frac{1}{n}\sum_{i=1}^n(\hat{m}(X_i)-m(X_i))\\ &\quad-\frac{1}{\hat{f}_X(x)}\frac{1}{nh_n}\sum_{i=1}^n K\Big(\frac{X_i-x}{h_n}\Big)\sum_{l=1}^n I(\Delta(i)=l)(\hat{m}(X_l)-m(X_l)),\end{aligned}$$

where $\eta_{ni} = \sum_{l=1}^n I(\Delta(i)=l)\varepsilon_l$ is drawn randomly with replacement from the true errors $\varepsilon_1,\ldots,\varepsilon_n$ [compare (6.24)]. From $\sum_{l=1}^n I(\Delta(i)=l)=1$ we can obtain the bound

$$\begin{aligned}|C_n(x)| \;&\le\; \Big|\frac{1}{\hat{f}_X(x)}\frac{1}{nh_n}\sum_{i=1}^n K\Big(\frac{X_i-x}{h_n}\Big)\eta_{ni}\Big| + \Big|\frac{1}{n}\sum_{i=1}^n\varepsilon_i\Big| + 2\sup_{z\in[0,1]}|\hat{m}(z)-m(z)|\\ &=\; \Big|\frac{1}{\hat{f}_X(x)}\frac{1}{nh_n}\sum_{i=1}^n K\Big(\frac{X_i-x}{h_n}\Big)\eta_{ni}\Big| + O\Big(\sqrt{\frac{\log\log n}{n}}\Big) + o(\beta_n) + O(h_n)\end{aligned}$$

uniformly in $x \in [0,1]$ almost surely with β_n as above by the law of iterated logarithm and (6.8). For the estimation of $\sqrt{n}\int C_n^2(x) f_X(x)\,dx$ the last three terms are negligible because they are of rates $\log\log n/\sqrt{n} = o(1)$, $\sqrt{n}o(\beta_n^2) = o(1)$ and $\sqrt{n}O(h_n^2) = o(1)$ almost surely. To conclude the proof of $\sqrt{n}\int C_n^2(x) f_X(x)\,dx = o_p(1)$ it is sufficient to verify that the following expectation converges to zero,

$$
\begin{aligned}
& E\Big[\sqrt{n}\int\Big(\frac{1}{\hat{f}_X(x)}\frac{1}{nh_n}\sum_{i=1}^n K\Big(\frac{X_i-x}{h_n}\Big)\eta_{ni}\Big)^2 f_X(x)\,dx\Big] \\
&= \sqrt{n}\int\sum_{i=1}^n \frac{1}{(nh_n)^2} E\Big[K^2\Big(\frac{x-X_i}{h_n}\Big)\frac{1}{n}\sum_{k=1}^n \varepsilon_k^2\Big]\frac{1}{f_X(x)}\,dx(1+o(1)) \;=\; O\Big(\frac{1}{\sqrt{n}h_n}\Big) \;=\; o(1)
\end{aligned}
$$

because conditional on the sample $\mathcal{Y}_n = \{(X_i,Y_i) \mid i=1,\dots,n\}$, the random variable η_{ni} has distribution function $F_n(y) = \frac{1}{n}\sum_{k=1}^n I\{\varepsilon_k \le y\}$. Altogether we obtain the rate $o_p(1)$ for (7.3) uniformly in $y \in I\!R$.

It is therefore sufficient to consider the term on the right-hand side of (7.2) in the following. To this end we write

$$
\begin{aligned}
\sqrt{n}\int(\hat{m}^*(x)-\hat{m}(x))f_X(x)\,dx &= \sqrt{n}\int(\hat{m}^*(x)-m(x))f_X(x)\,dx \\
&\quad - \sqrt{n}\int(\hat{m}(x)-m(x))f_X(x)\,dx.
\end{aligned} \tag{7.4}
$$

The second term is the same as appears in the original process $R_n = \sqrt{n}(\hat{F}_n - F)$ and we have

$$
-\sqrt{n}\int(\hat{m}(x)-m(x))f_X(x)\,dx \;=\; -\frac{1}{\sqrt{n}}\sum_{j=1}^n \varepsilon_j + o_p(1). \tag{7.5}
$$

Similar approximations apply for the first term. From the definition of $\hat{m}^*$ [compare the definition of $\hat{m}$ in (2.3)] we have

$$
\begin{aligned}
\sqrt{n}\int(\hat{m}^*(x)-m(x))f_X(x)\,dx &= \frac{\sqrt{n}}{nh_n}\sum_{i=1}^n\int K\Big(\frac{X_i-x}{h_n}\Big)(\varepsilon_i^* + \hat{m}(X_i)-m(x))\frac{f_X(x)}{\hat{f}_X(x)}\,dx \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i^*\int\frac{1}{h_n}K\Big(\frac{X_i-x}{h_n}\Big)\,dx \\
&\quad + \frac{1}{\sqrt{n}}\sum_{j=1}^n \varepsilon_j \frac{1}{nh_n}\sum_{i=1}^n\int K\Big(\frac{X_i-x}{h_n}\Big)\frac{dx}{f_X(X_i)}K\Big(\frac{X_i-X_j}{h_n}\Big) \\
&\quad + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i^* + \frac{1}{\sqrt{n}}\sum_{j=1}^n \varepsilon_j\int\frac{1}{h_n}K\Big(\frac{t-X_j}{h_n}\Big)\,dt + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i^* + \frac{1}{\sqrt{n}}\sum_{j=1}^n \varepsilon_j + o_p(1) \qquad (7.6)
\end{aligned}
$$

and the assertion of the lemma follows from (7.2), (7.4), (7.5), (7.6). □

Proof of Theorem 3.2. From Lemma 3.1 we have the following expansion of the bootstrap process defined in (3.5), $R_n^*(y) = \tilde{R}_n^*(y) + o_p(1)$ uniformly with respect to $y \in I\!R$, where

$$\tilde{R}_n^*(y) \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(I\{\varepsilon_i^* \le y\} - \tilde{F}_n(y) + \tilde{f}_n(y)\varepsilon_i^* \Big). \tag{7.7}$$

Note that the process is centered with respect to the conditional expectation, given the sample $\mathcal{Y}_n$. To prove conditional weak convergence of the process R_n^* in probability, it is sufficient to consider $\tilde{R}_n^*$. We now show the stronger result that for almost all sequences $\mathcal{Y} = \{(X_1, Y_1), (X_2, Y_2), \ldots\}$ we have conditional weak convergence. This follows from stochastic asymptotic equicontinuity and convergence of the finite dimensional distributions. Note that

$$\begin{aligned}
&P\Big(\sup_{y,z \in I\!R, |y-z|<\delta} \Big| \tilde{R}_n^*(y) - \tilde{R}_n^*(z) \Big| > \epsilon \;\Big|\; \mathcal{Y} \Big) \\
&\le P\Big(\sup_{y,z \in I\!R, 0 \le y-z<\delta} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(I\{z < \varepsilon_i^* \le y\} - \tilde{F}_n(y) + \tilde{F}_n(z) \Big) \Big| > \frac{\epsilon}{2} \;\Big|\; \mathcal{Y} \Big) && (7.8) \\
&\quad + P\Big(\sup_{y,z \in I\!R, |y-z|<\delta} |\tilde{f}_n(y) - \tilde{f}_n(z)| \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i^* \Big| > \frac{\epsilon}{2} \;\Big|\; \mathcal{Y} \Big) && (7.9)
\end{aligned}$$

and we have to show that the limit $\lim_{\delta \to 0} \limsup_{n \to \infty}$ of these probabilities is zero for almost all random sequences $\mathcal{Y}$. To this end let C denote a constant such that $\sup_{y \in I\!R} f(y) \le C/2$. The probability (7.8) can be bounded by

$$\begin{aligned}
&I\{\sup_{y \in I\!R} |\tilde{f}_n(y)| \le C\}\, P\Big(\sup_{y,z \in I\!R, 0 \le y-z<\delta} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(I\{z < \varepsilon_i^* \le y\} - \tilde{F}_n(y) + \tilde{F}_n(z) \Big) \Big| > \frac{\epsilon}{2} \;\Big|\; \mathcal{Y} \Big) \\
&+ I\{\sup_{y \in I\!R} |f(y) - \tilde{f}_n(y)| > C/2\}.
\end{aligned}$$

The second indicator function converges to zero almost surely due to Lemma 6.3 (i). We therefore assume that $\sup_{y \in I\!R} |\tilde{f}_n(y)| \le C$. From the Mean Value Theorem we have that

$$|\tilde{F}_n(y) - \tilde{F}_n(z)| \;\le\; \sup_{\xi \in I\!R} |\tilde{f}_n(\xi)||y-z| \;\le\; C|y-z|$$

and obtain for (7.8) the bound

$$P\Big(\sup_{s,t \in [0,1], 0 \le t-s < C\delta} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(I\{s \le U_i \le t\} - t + s \Big) \Big| > \frac{\epsilon}{2} \Big).$$

Because the $U_1, \ldots, U_n$ are independent and uniformly distributed in $[0,1]$ and the uniform empirical process is asymptotically equicontinuous, the limit $\lim_{\delta \to 0} \limsup_{n \to \infty}$ of this probability is zero.

Applying Markov's inequality, the probability (7.9) can be bounded by

$$4\Big(\sup_{y,z \in I\!R, |y-z|<\delta} |\tilde{f}_n(y) - \tilde{f}_n(z)| \Big)^2 \frac{1}{\epsilon^2} \mathrm{Var}(\varepsilon_i^* \mid \mathcal{Y}_n).$$

The supremum inside parantheses can further be bounded by

$$r(\delta, n) \;=\; \sup_{y,z\in I\!R, |y-z|<\delta} |f(y) - f(z)| + 2 \sup_{y\in I\!R} |\tilde{f}_n(y) - f(y)|$$

and we obtain $\lim_{\delta\to 0} \limsup_{n\to\infty} r(\delta, n) = 0$ almost surely by assumption A.2 and Lemma 6.3 (i). For the conditional variance we have from (3.1)

$$\mathrm{Var}(\varepsilon_i^* \mid \mathcal{Y}_n) \;=\; \frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_i^2 + O(a_n^2) \;=\; \sigma^2 + o(1) \tag{7.10}$$

almost surely. Hence, (7.9) converges to zero almost surely.

We use Lindeberg's condition to prove convergence of the finite dimensional distributions, conditional on the sequence $\mathcal{Y}$. Let $a_1, \ldots, a_k,\ y_1, \ldots, y_k \in I\!R$ and consider for $\tilde{R}_n^*$ defined in (7.7) the linear combination

$$\sum_{j=1}^k a_j \tilde{R}_n^*(y_j) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^n (A_{ni} + B_{ni}\varepsilon_i^*)$$

with the definitions

$$\begin{aligned} A_{ni} \;&=\; \sum_{j=1}^k a_j(I\{\varepsilon_i^* \le y_j\} - \tilde{F}_n(y_j)), \quad |A_{ni}| \le \sum_{j=1}^k |a_j| =: A, \quad i = 1, \ldots, n, \\ B_{ni} \;&=\; \sum_{j=1}^k a_j \tilde{f}_n(y_j), \quad |B_{ni}| \le \sum_{j=1}^k |a_j| C =: B, \quad i = 1, \ldots, n, \end{aligned}$$

where, as above, we only consider the event $\sup_{y\in I\!R} |\tilde{f}_n(y)| \le C$. We have further

$$\begin{aligned} L_n(\delta) \;&=\; \frac{1}{n}\sum_{i=1}^n E\Big[\Big(A_{ni} + B_{ni}\varepsilon_i^*\Big)^2 I\Big\{\Big|A_{ni} + B_{ni}\varepsilon_i^*\Big| > n^{1/2}\delta\Big\} \;\Big|\; \mathcal{Y}_n\Big] \\ &\le\; 2\int (A^2 + B^2 z^2) I\{|z| > (n^{1/2}\delta - A)/B\} \tilde{f}_n(z)\, dz. \end{aligned}$$

This integral converges almost surely to zero applying Lemma 3.2 by Koul and Lahiri (1994), our Lemma 6.3 (iii), and $\int z^2 \tilde{f}_n(z)\, dz = \mathrm{Var}(\varepsilon_i^* \mid \mathcal{Y}_n)$ from (7.10).

We conclude the proof by calculating the conditional covariances,

$$\begin{aligned} &E\Big[\tilde{R}_n^*(y)\tilde{R}_n^*(z) \;\Big|\; \mathcal{Y}_n\Big] \\ &= \frac{1}{n}\sum_{i=1}^n E\Big[\Big(I\{\varepsilon_i^* \le y\} - \tilde{F}_n(y) + \tilde{f}_n(y)\varepsilon_i^*\Big)\Big(I\{\varepsilon_i^* \le z\} - \tilde{F}_n(z) + \tilde{f}_n(z)\varepsilon_i^*\Big) \;\Big|\; \mathcal{Y}_n\Big] \\ &= \tilde{F}_n(y \wedge z) - \tilde{F}_n(y)\tilde{F}_n(z) + \tilde{f}_n(y)\tilde{f}_n(z)\mathrm{Var}(\varepsilon_i^* \mid \mathcal{Y}_n) \\ &\quad + \tilde{f}_n(y) E[\varepsilon_i^* I\{\varepsilon_i^* \le z\} \mid \mathcal{Y}_n] + \tilde{f}_n(z) E[\varepsilon_i^* I\{\varepsilon_i^* \le y\} \mid \mathcal{Y}_n]. \end{aligned} \tag{7.11}$$

Further we have

$$E[\varepsilon_i^* I\{\varepsilon_i^* \le z\} \mid \mathcal{Y}_n] \;=\; \int_{-\infty}^{z} y\tilde{f}_n(y)\,dy \;=\; \frac{1}{n}\sum_{i=1}^{n}\int yI\{y\le z\}\frac{1}{a_n}k\Big(\frac{y-\tilde{\varepsilon}_i}{a_n}\Big)\,dy$$

$$= \frac{1}{n}\sum_{i=1}^{n}\int (\tilde{\varepsilon}_i + ta_n)I\{\tilde{\varepsilon}_i + ta_n \le z\}k(t)\,dt$$

$$= \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\int I\{\varepsilon_i \le z - ta_n + z_{ni}\}k(t)\,dt + o(1)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i I\{\varepsilon_i \le z\} + r_n + o(1) \tag{7.12}$$

almost surely, where for

$$z_{ni} \;=\; \hat{m}(X_i) - m(X_i) + \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j + \frac{1}{n}\sum_{j=1}^{n}(m(X_j) - \hat{m}(X_j)),$$

it follows that $\max_{i=1,\dots,n}|z_{ni}| = o(1)$ almost surely. The remainder is bounded by

$$|r_n| \;\le\; \frac{1}{n}\sum_{i=1}^{n}\int I\{z < \varepsilon_i \le z - ta_n + z_{ni}\}k(t)\,dt$$

$$+\frac{1}{n}\sum_{i=1}^{n}\int I\{z - ta_n + z_{ni} < \varepsilon_i \le z\}k(t)\,dt.$$

Both summands are treated very similarly and, hence, we concentrate on the first one that is bounded by

$$\frac{1}{n}\sum_{i=1}^{n}\int I\{z<\varepsilon_i \le z - ta_n + c_n\}k(t)\,dt + I\{\max_{i=1,\dots,n}|z_{ni}| > c_n\} \tag{7.13}$$

where the (deterministic) sequence $c_n = o(1)$ is chosen such that the second indicator converges to zero almost surely. By the strong law of large numbers the first term in (7.13) is bounded by

$$\int |F(z - ta_n + c_n) - F(z)|k(t)\,dt + o(1) \;\le\; \sup_{y\in I\!R}|f(y)|\int (|t|a_n + c_n)k(t)\,dt + o(1) \;=\; o(1)$$

almost surely. Hence, by the strong law of large numbers we obtain from (7.12), $E[\varepsilon_i^* I\{\varepsilon_i^* \le z\} \mid \mathcal{Y}_n] = U(z) + o(1)$ almost surely. From this calculation together with (7.10) and Lemma 6.3 it follows that the conditional covariances (7.11) converge almost surely to the asserted covariance from Theorem 3.2. □

Acknowledgements

This paper contains results of the author's Habiliationsschrift [Neumeyer (2006)]. She would like to thank Holger Dette for his great support and encouragement over the years. A major part of the Habilitationsschrift was written while she visited the Australian National University supported by a research grant of the Deutsche Forschungsgemeinschaft (NE 1189/1-1) and she is very grateful to the members of the Mathematical Sciences Institute, especially Peter Hall, for their great hospitality.

References

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

A. Azzalini (2005). *R package 'sn': The skew-normal and skew-t distributions.* (version 0.33). http://azzalini.stat.unipd.it/SN

J. Bai (1994). *Weak convergence of the sequential empirical processes of residuals in ARMA models.* Ann. Statist. 22, 2051–2061.

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.

P. J. Bickel and D. A. Freedman (1981). *Some asymptotic theory for the bootstrap.* Ann. Statist. 9, 1196–1217.

M. V. Boldin (1982). *Estimation of the distribution of noise in an autoregression scheme.* Theory Probab. Appl. 27, 866–871.

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.

D. De Angelis, P. Hall and G. A. Young (1993). *Analytical and bootstrap approximations to estimator distributions in L^1 regression.* J. Amer. Statist. Assoc. 88, 1310–1316.

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

H. Dette (2002). *A consistent test for heteroscedasticity in nonparametric regression based on the kernel method.* J. Statist. Plann. Inference 103, 311–329.

H. Dette and A. Munk (1998). *Testing heteroscedasticity in nonparametric regression.* J. Roy. Statist. Soc. B 60, 693–708.

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

J. Durbin (1973). *Weak convergence of the sample distribution function when parameters are estimated.* Ann. Statist. 1, 279–290.

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

M. Francisco-Fernández and J. M. Fernández (2005). *Two tests for heteroscedasticity in nonparametric regression.* J. Amer. Statist. Assoc. 85, 1039–1049.

E. Giné and J. Zinn (1990). *Bootstrapping general empirical measures.* Ann. Probab. 18, 851–869.

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

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

P. Hall and P. Patil (1996). *On the choice of smoothing parameter, threshold and truncation in nonparametric regression by non-linear wavelet methods.* J. Roy. Statist. Soc. Ser. B 58, 361–377.

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 S. N. Lahiri (1994). *On Bootstrapping M-Estimated Residual Processes in Multiple Linear Regression Models.* J. Multivariate Anal. 49, 255–265.

H. L. Koul and S. Levental (1989). *Weak convergence of the residual empirical process in explosive autoregression.* Ann. Statist. 17, 1784–1794.

H. L. Koul and P. K. Sen (1992). *Weak convergence of a weighted residual empirical process in autoregression.* Statist. Decisions 9, 235–262.

E. Liebscher (1999). *Estimating the Density of the Residuals in Autoregressive Models.* Statistical Inference for Stochastic Processes 2, 105–117.

H. Liero (2003). *Testing homoscedasticity in nonparametric regression.* J. Nonparametr. Statist. 15, 31–51.

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

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

N. Neumeyer (2001). *Vergleich nichtparametrischer Regressionsfunktionen unter Verwendung stochastischer Prozesse.* PhD thesis, Ruhr-Universität Bochum (in German). http://www-brs.ub.rub.de/netahtml/HSS/Diss/NeumeyerNatalie/diss.pdf

N. Neumeyer (2006). *Bootstrap procedures for empirical processes of nonparametric residuals.* Habilitationsschrift, Ruhr-Universität Bochum.

N. Neumeyer and H. Dette (2005). *A note on one-sided nonparametric analysis of covariance by ranking residuals.* Math. Methods Statist. 14, 80–104.

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, to appear. http://www.stat.ucl.ac.be/IAP/publication_tr.html

D. Pollard (1984). *Convergence of Stochastic Processes.* Springer, New York.

S. Portnoy (1986). *Asymptotic behavior of the empiric distribution of M-estimated residuals from a regression model with many parameters.* Ann. Statist. 14, 1152–1170.

G. Qin (1996). *Consistent Residuals Density Estimation in Nonparametric Regression Under m(n)-Dependent Sample.* Journal of Mathematical Research and Exposition 16, 505–516.

G. Qin, S. Shi and G. Chai (1996) *Asymptotics of the residual density estimation in nonparametric regression under m(n)-dependence.* Applied Mathematics. A Journal of Chinese Universities Ser. B 11, 59–76.

R Development Core Team (2005). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

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

D. Ruppert, S. J. Sheather and M. P. Wand (1995). *An effective bandwidth selector for local least squares regression.* J. Amer. Statist. Assoc. 90, 1257–1270.

A. Schick and W. Wefelmeyer (2002). *Estimating the innovation distribution in nonlinear autoregressive models.* Ann. Inst. Statist. Math. 54, 245–260.

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

R. S. Singh (1981). *On the asymptotic accuracy of Efron's bootstrap.* Ann. Statist. 9, 1187–1195.

P. L. Speckman, J.-E. Chiu, J. E. Hewett and S. E. Bertelson (2003). *A one-sided test adjusting for covariates by ranking residuals following smoothing.* Biometrics, to appear. http://www.stat.missouri.edu/∼speckman/pub.html

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

T. J. Sweeting (1989). *On Conditional Weak Convergence.* J. Theoret. Probab. 2, 461–474.

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

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

C.–F. J. Wu (1986). *Jackknife, bootstrap and other resampling methods in regression analysis.* Ann. Statist. 14, 1261–1350.

L.–X. Zhu, Y. Fujikoshi and K. Naito (2001). *Heteroscedasticity checks for regression models.* Sci. China Ser. A 44, 1236–1252.