Universität Hamburg

Testing independence in nonparametric regression

Natalie Neumeyer

Preprint–No. 2007-06
Dezember 2007

DEPARTMENT MATHEMATIK

SCHWERPUNKT MATHEMATISCHE STATISTIK
UND STOCHASTISCHE PROZESSE

BUNDESSTR. 55, D – 20146 HAMBURG

Testing independence in nonparametric regression

Natalie Neumeyer

University of Hamburg, Department of Mathematics

Bundesstrasse 55, 20146 Hamburg, Germany

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

December 20, 2007

Abstract

We propose a new test for independence of error and covariate in a nonparametric regression model. The test statistic is based on a kernel estimator for the L_2-distance of the conditional and unconditional distribution of the covariates. In contrast to tests so far available in literature, the test can be applied in the important case of multivariate covariates. It can also be adjusted for models with heteroscedastic variance. Asymptotic normality of the test statistic is shown. Simulation results and a real data example are presented.

AMS Classification: 62G10, 62G08, 62G09

Keywords and Phrases: bootstrap, goodness-of-fit, kernel estimator, nonparametric regression, test for independence

1 Introduction

We consider independent and identically distributed data $(X_1, Y_1), \ldots, (X_n, Y_n)$, where X_i is d-dimensional and Y_i one-dimensional. Our purpose is to test whether the data follow a homoscedastic regression model

$$Y_i = m(X_i) + \varepsilon_i \tag{1.1}$$

with regression function $m(x) = E[Y_i \mid X_i = x]$, where the error $\varepsilon_i = Y_i - E[Y_i \mid X_i]$ is independent of the covariate X_i. As a new test for the hypothesis

$$H_0 : X_i \text{ and } \varepsilon_i \text{ are independent} \tag{1.2}$$

a simple kernel based test statistic is suggested.

Although the independence of error and covariate is a common assumption in nonparametric regression [compare Fan and Gijbels (1995), Koul and Schick (1997), Akritas and Van Keilegom (2001), Müller, Schick and Wefelmeyer (2004, 2006), Cheng (2005), Neumeyer, Dette and Nagel (2006), Pardo–Fernández, Van Keilegom and González–Manteiga (2007), ...], to the present author's knowledge so far there are only two tests for hypothesis (1.2) available in literature. In the homoscedastic model with univariate covariate Einmahl and Van Keilegom (2007a) consider a stochastic process based on differences of the observations Y_i, which converges weakly to a bivariate Gaussian process. In the heteroscedastic model [defined in (1.3) below] with univariate covariate Einmahl and Van Keilegom (2007b) propose tests based on the difference of the empirical distribution function of $(X_i, \hat{\varepsilon}_i)$ and the product of the empirical distribution functions of the covariates X_i and residuals $\hat{\varepsilon}_i$, respectively. The considered process converges weakly to a bivariate Gaussian process. Both tests have been shown to be superior to tests for simple heteroscedasticity in cases, where the dependence of error and covariate is complicated. However, both procedures cannot easily be extended to the important case of multivariate covariates. In contrast the test considered in the present paper is valid for multivariate covariates and theory is less complicated. In Einmahl and Van Keilegom (2007a) the reason for only considering one-dimensional covariates lies within the special structure of the processes which uses ordering of the covariates. The test proposed in Einmahl and Van Keilegom (2007b) depends on residual based empirical processes, for which (in a nonparametric setting) asymptotic theory is so far not available in the case of multivariate covariates [see Akritas and Van Keilegom (2001) for the case of one-dimensional covariates]. Hence, in the multivariate design case tests cannot be based on estimators for the joint and marginal distributions of (X_i, ε_i) (such as a Kolmogorov-Smirnov type test). Further tests based on density estimates such as estimators for the L_2-distance of the joint and marginal densities of (X_i, ε_i) should not be recommended, because those require the choice of 4 smoothing parameters (for the estimation of the regression function, the joint density and both marginal densities, respectively). Nonparametric tests for independence of components of paired iid-data include those by Hoeffding (1948), Blum, Kiefer and Rosenblatt (1961), Rosenblatt (1975) and Deheuvels (1980), among others.

We follow an idea by Zheng (1997), who proposed a procedure for testing independence of components U_i and V_i in an iid sample of paired observations $(U_1, V_1), \ldots, (U_n, V_n)$ and showed asymptotic normality under the null hypothesis of independence and under local alternatives. This test statistic was further investigated by Dette and Neumeyer (2000) who proved asymptotic normality under fixed alternatives. Please note that in our model the errors ε_i are not observable. Hence, when applying procedures constructed for testing independence of components U_i and V_i in paired observations $(U_1, V_1), \ldots, (U_n, V_n)$ for testing independence of $U_i = X_i$ and ε_i the errors are nonparametrically estimated by residuals $\hat{\varepsilon}_i = Y_i - \hat{m}(X_i)$. Here $\hat{m}$ denotes, for instance, a local polynomial estimator for the regression function. This complicates asymptotic theory as now the random variables $V_i = \hat{\varepsilon}_i$ are dependent. In general it depends crucially on the structure of the test statistic whether estimating the regression function has influence on the asymptotic distribution of the test or not, compare Ahmad and Li (1997), where there is no influence or Loynes (1980), where replacing true (unobservable) errors by residuals changes the asymptotic distribution. Zheng's (1997) test is based on an estimator for an L_2-distance between the conditional distribution of V_i and U_i, denoted by $F_{V|U}$, and the unconditional distribution, F_V. Zheng's test is not symmetric in interchanging the roles of U_i and V_i, which may not be desirable, when paired random variables are observed and one is interested in detecting dependence of the two components. However, this asymmetry is essential for our purpose. Our key idea is to apply Zheng's test to X_i and $\hat{\varepsilon}_i$ instead of U_i and V_i. A natural way seems to be to set $V_i = X_i$ and $U_i = \hat{\varepsilon}_i$ to investigate whether the (measurement) error ε_i at some measurement point X_i actually depends on X_i, i. e. whether the conditional distribution $F_{\varepsilon|X}$ depends on X. Yet, doing so we run into the same problems as Einmahl and Van Keilegom (2007b) and asymptotic distribution theory is only available in the one-dimensional covariate case. Moreover the estimation of the residuals here results in unwanted bias and a rather complicated asymptotic distribution (as will be explained in section 4 of the present paper). This is not the case when interchanging the roles of X_i and $\hat{\varepsilon}_i$, applying Zheng's (1997) test to $U_i = X_i$, $V_i = \hat{\varepsilon}_i$, and this is the route we will take in this paper. I. e. we base our test on an estimated L_2-distance between the conditional distribution of the covariate given the error, $F_{X|\varepsilon}$, and the unconditional covariate distribution, F_X. This approach lacks the above measurement error interpretation but then the test statistic is applicable in the multivariate covariate case. Under regularity assumptions asymptotic normality of the test statistic can be shown both under the null hypothesis of independence and under alternatives.

Because the asymptotic null distribution depends on unknown features of the data-generating process, we recommend to apply resampling procedures. We prove consistency of the classical residual bootstrap as introduced by Härdle and Bowman (1988) in our context and discuss the small sample performance of this procedure.

The test can also be adjusted to justify a regression model with heteroscedastic variance, i. e.

$$Y_i = m(X_i) + \sigma(X_i)\varepsilon_i, \tag{1.3}$$

where the covariates X_i are independent of the errors $\varepsilon_i = (Y_i - E[Y_i \mid X_i])/(\mathrm{Var}(Y_i \mid X_i))^{1/2}$. We should mention some of the vast literature on model tests in nonparametric regression models such as goodness-of-fit tests for the regression function [see Härdle and Mammen (1993), Zheng (1996), Stute (1997), Stute, Thies and Zhu (1998), Alcalá, Cristóbal and

González Manteiga (1999), Dette (1999) and Van Keilegom, González Manteiga and Sánchez Sellero (2007), among others] and tests for heteroscedasticity [for instance, Eubank and Thomas (1993), Dette and Munk (1998), Dette (2002), Zhu, Fujikoshi and Naito (2001) and Liero (2003)].

The remainder of the paper is organized as follows. In section 2 we motivate and define the test statistic and list the model assumptions. Section 3 states the asymptotic distributions under H_0 as well as under fixed alternatives, and discusses bootstrap theory. In section 4 interchanging the roles of residuals and covariates is discussed for the case of univariate covariates. Section 5 explains the extension to heteroscedastic models and compares related literature, whereas in section 6 the small sample performance is investigated in a simulation study, and a real data example is presented. Section 6 concludes the paper and all proofs are given in an appendix.

2 Assumptions and test statistic

First we consider the homoscedastic model (1.1) where our aim is to test whether the covariate X_i is independent of the error $\varepsilon_i = Y_i - m(X_i) = Y_i - E[Y_i \mid X_i]$ [see hypothesis (1.2)]. We impose the following model assumptions. The d-dimensional design points $X_1, \ldots, X_n$ are independent and identically distributed with distribution function F_X on compact support, say $[0,1]^d$. F_X has a twice continuously differentiable density f_X such that $\inf_{x\in[0,1]^d} f_X(x) > 0$. The regression function m is twice continuously differentiable in $(0,1)^d$. The errors $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed with bounded density f_ε which has one bounded continuous derivative. The errors are centered, i. e. $E[\varepsilon_i] = 0$ (by definition), with existing fourth moment, $E[\varepsilon_i^4] < \infty$.

The error ε_i is estimated nonparametrically by the residual

$$\hat{\varepsilon}_i \;=\; Y_i - \hat{m}(X_i), \quad i = 1, \ldots, n, \tag{2.1}$$

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

$$\hat{m}(x) \;=\; \frac{1}{nb_n^d} \sum_{i=1}^{n} k\Big(\frac{X_i - x}{b_n}\Big) Y_i \frac{1}{\hat{f}_X(x)} \tag{2.2}$$

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

$$\hat{f}_X(x) \;=\; \frac{1}{nb_n^d} \sum_{i=1}^{n} k\Big(\frac{X_i - x}{b_n}\Big). \tag{2.3}$$

Here b_n denotes a sequence of positive bandwidths such that $b_n \to 0$, $nb_n^{2d} \to \infty$, $nb_n^{4d} \to 0$ for $n \to \infty$. Let further κ denote a twice continuously differentiable symmetric density with compact support $[-1,1]$, say, such that $\int u\kappa(u)\,du = 0$, and let k denote the product kernel $k(u_1,\dots,u_d) = \kappa(u_1)\cdots\kappa(u_d)$. We assume a modification of the estimator $\hat{m}$ at the boundaries to obtain uniform rates of convergence, compare Müller (1984) or Härdle (1989, p. 130). Note, that other uniformly consistent nonparametric function estimators such as local polynomial estimators [see, e. g. Fan and Gijbels (1996)] can be applied as well and very similar results can be obtained.

To test hypothesis (1.2) we suggest the simple kernel based test statistic (compare Zheng, 1997)

$$T_n = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{1}{h_n}K\Big(\frac{\hat\varepsilon_i-\hat\varepsilon_j}{h_n}\Big)\int (I\{X_i\le x\}-F_{X,n}(x))(I\{X_j\le x\}-F_{X,n}(x))w(x)\,dx, \tag{2.4}$$

where $I\{\cdot\}$ denotes the indicator function, w a positive integrable weight function, and $F_{X,n}$ is the empirical distribution function of the covariates $X_1,\dots,X_n$. Further, h_n is a sequence of positive bandwidths such that $h_n\to 0$, $nh_n\to\infty$ and $h_n/b_n^{d-2}\to 0$ for $n\to\infty$. K denotes a bounded symmetric density function such that $\int u^2K(u)\,du<\infty$. We assume that K is $\lambda\ge 2$ times continuously differentiable in the inside of its support with bounded derivatives. These assumptions are sufficient when the λ-th derivative of K vanishes [when K is chosen to be a (truncated) polynomial such as the Epanechnikov kernel]. Otherwise we have to impose the following additional technical bandwidth condition which is less restrictive for very smooth kernels (for instance, the Gaussian kernel where λ is arbitrarily large). We assume that there exists some $\rho\in(0,\frac{1}{2})$ such that

$$nb_n^{\frac{d\rho\lambda}{\rho\lambda-1}}h_n^{\frac{\lambda-\frac{1}{2}}{\lambda-2}}\to\infty.$$

For example, consider the limit case $\rho=\frac{1}{2}$ with $\lambda=8$, which gives the condition: $nb_n^{4d/3}h_n^{2.5}\to\infty$.

Please note that in contrast to Zheng's (1997) setting the pairs $(X_i,\hat\varepsilon_i)$ are dependent for different $i\in\{1,\dots,n\}$. We will investigate whether this changes the asymptotic theory. T_n is an estimator for the expectation

$$\begin{aligned} B_{h_n} &= E\Big[\frac{1}{h_n}K\Big(\frac{\varepsilon_1-\varepsilon_2}{h_n}\Big)\int (I\{X_1\le x\}-F_X(x))(I\{X_2\le x\}-F_X(x))w(x)\,dx\Big] \\ &= \int\!\!\int\!\!\int \frac{1}{h_n}K\Big(\frac{y_1-y_2}{h_n}\Big)(F_{X|\varepsilon}(x|y_1)-F_X(x))(F_{X|\varepsilon}(x|y_2)-F_X(x))f_\varepsilon(y_1)f_\varepsilon(y_2) \\ &\quad w(x)\,dy_1\,dy_2\,dx, \end{aligned} \tag{2.5}$$

where $F_{X|\varepsilon}(\cdot|y)=P(X_1\le\cdot\mid\varepsilon_1=y)$ denotes the conditional distribution of X_i given $\varepsilon_i=y$. Under the null hypothesis of independence B_{h_n} is zero, whereas under the alternative it converges for $h_n \to 0$ to a positive L_2-distance of the conditional and unconditional distribution of the covariates, i. e.

$$\int\int (F_{X|\varepsilon}(x|y) - F_X(x))^2 f_\varepsilon^2(y) w(x)\, dy\, dx,$$

Under the alternative H_1 we assume that $F_{X|\varepsilon}(\cdot|y)$ has a uniformly bounded continuous derivatives with respect to y and the corresponding density $f_{X|\varepsilon}(x|y)$ is continuously differentiable with bounded partial derivatives with respect to y and all components of x. For the asymptotic theory under H_1 we additionally assume for the bandwidths $nh_n^2 b_n^d \to \infty$ and $nh_n^4 \to \infty$.

Throughout we will use the notations

$$\Delta(x|y) \;=\; F_{X|\varepsilon}(x|y) - F_X(x), \quad v(x|y) \;=\; \Delta(x|y) f_\varepsilon(y) - E[\Delta(x|\varepsilon_1) f_\varepsilon(\varepsilon_1)]. \qquad (2.6)$$

3 Asymptotic results and bootstrap

For the test statistic T_n defined in (2.4) under the assumptions stated in section 2 we obtain the following limiting distributions.

Theorem 3.1 (a) *Under the null hypothesis (1.2) of independence $nh_n^{1/2} T_n$ converges to a mean zero normal distribution with variance*

$$\tau^2 = 2\int K^2(u)\, du \int f_\varepsilon^2(y)\, dy \int\int (F_X(x \wedge t) - F_X(x)F_X(t))^2 w(x) w(t)\, dx\, dt,$$

where $x \wedge t$ denotes the componentwise minimum of the vectors x and t.
(b) *Under the fixed alternative of dependence of errors and covariates, $\sqrt{n}(T_n - B_{h_n})$ converges to a mean zero normal distribution with variance*

$$\omega^2 \;=\; \tilde{\omega}^2 + 4\mathrm{Var}(Z_1) + 8\mathrm{Cov}\Big(Z_1\,,\, \int (I\{X_1 \le x\} - F_X(x)) v(x|\varepsilon_1) w(x)\, dx\Big),$$

where

$$Z_1 \;=\; \varepsilon_1 \int (I\{X_1 \le x\} - F_X(x)) \int\int\int \frac{f_{X|\varepsilon}(X_1|z)}{f_X(X_1)} \frac{\partial f_{X|\varepsilon}(s|z) f_\varepsilon(z)}{\partial z}$$
$$\times (I\{s \le x\} - F_X(x)) f_\varepsilon(z) w(x)\, dx\, ds\, dz \int K'(u) u\, du$$

$$\tilde{\omega}^2 \;=\; 4\mathrm{Var}\Big(\int \Delta(x|\varepsilon_1) v(x|\varepsilon_1)\, w(x)\, dx\Big)$$
$$+ 4E\Big[\int\int (F_{X|\varepsilon}(x \wedge t|\varepsilon_1) - F_{X|\varepsilon}(x|\varepsilon_1) F_{X|\varepsilon}(t|\varepsilon_1)) \Delta(x|\varepsilon_1) v(x|\varepsilon_1) v(t|\varepsilon_1) w(x) w(t)\, dx\, dt\Big]$$

with notations (2.6) and B_{h_n} defined in (2.5).

The proof of Theorem 3.1 is given in the appendix.

Remark 3.2 Note that for $\tilde{T}_n$ defined by

$$\tilde{T}_n = \tag{3.1}$$
$$\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{1}{h_n} K\Big(\frac{\varepsilon_i - \varepsilon_j}{h_n}\Big) \int (I\{X_i \le x\} - F_{X,n}(x))(I\{X_j \le x\} - F_{X,n}(x)) w(x)\, dx,$$

we have from results by Zheng (1999) and Dette and Neumeyer (2000) that $nh_n^{1/2}\tilde{T}_n$ and $\sqrt{n}(\tilde{T}_n - B_{h_n})$ converge to mean zero normal distributions with variance τ^2 and $\tilde{\omega}^2$ (both defined in Theorem 3.1), respectively. Hence, replacing the true (unobservable) errors by residuals does not change the asymptotic distribution under the null hypothesis of independence. However, it changes the asymptotic distribution under the alternative. ■

Remark 3.3 Local alternatives of convergence rate $n^{1/2}h_n^{1/4}$ can be detected with the proposed test. Under $H_{1n} : F_{X|\varepsilon}(x|y) = F_X(x) + d(x,y)(nh_n^{1/2})^{-1/2}$ with continously differentiable d with bounded partial derivatives, $nh_n^{1/2}T_n$ converges to a normal distribution with expectation $\int\int d^2(x,y) f_\varepsilon^2(y) w(x)\, dy\, dx$ and variance τ^2 from Theorem 3.1. As was described by Zheng (1997) optimal choices of the weight function w with respect to maximizing power under local alternatives would now depend on $F_{X|\varepsilon} - F_X$.

Considering asymptotics for the test statistic T_n with a fixed bandwidth $h_n \equiv h$ would result in a test which can detect local alternatives of convergence rate $n^{-1/2}$, compare Hall and Hart (1990) or Hart (1997, chapter 6). ■

A consistent asymptotic level α test could be obtained from Theorem 3.1 by applying a standardized test statistic $nh_n^{1/2}T_n/\hat{\tau}$ with asymptotic standard normal law (under H_0), where $\hat{\tau}^2$ consistently estimates τ^2. However, as is quite common in nonparametric regression model tests, more accurate critical values can be obtained by resampling procedures. In the following we discuss the applicability of the classical residual bootstrap as introduced by Härdle and Bowman (1988) for nonparametric regression models.

Let $\tilde{\varepsilon}_i = \hat{\varepsilon}_i - n^{-1}\sum_{l=1}^n \hat{\varepsilon}_l$ denote the centered residuals and let $\tilde{F}_{\varepsilon,n}$ denote the empirical distribution function of $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$. Given the original sample $\mathcal{Y}_n = \{(X_1,Y_1), \ldots, (X_n,Y_n)\}$ let the bootstrap errors $\varepsilon_1^*, \ldots, \varepsilon_n^*$ be independent with distribution function $\tilde{F}_{\varepsilon,n}$. Then we define as bootstrap observations

$$Y_i^* = \hat{m}(X_i) + \varepsilon_i^*. \tag{3.2}$$

Let T_n^* be defined as T_n in (2.4), but based on the bootstrap sample $(X_1,Y_1^*), \ldots, (X_n,Y_n^*)$. Then we have the following asymptotic result.

Theorem 3.4 *Conditionally on $\mathcal{Y}_n$ the bootstrap test statistic $nh_n^{1/2}T_n^*$ converges in distribution to a mean zero normal distribution with variance τ^2 defined in Theorem 3.1, in probability.*

A sketch of the proof is given in the appendix. Critical values for an asymptotic level α test can then be approximated by repeating the bootstrap procedure B times, resulting in ordered values of the bootstrap test statistics $T_n^{*(1)}, \ldots, T_n^{*(B)}$. The null hypothesis is rejected if $T_n > T_n^{*([B(1-\alpha)])}$.

Remark 3.5 Please note that even in the heteroscedastic model wild bootstrap is not applicable in our context. Wild bootstrap asymptotically changes the error distribution and, hence, in general changes the asymptotic variance τ^2. More importantly, bootstrap observations should be generated under the null hypothesis, i. e. independence of covariates and bootstrap error. However, wild bootstrap is usually applied in heteroscedastic models because it preserves dependence of errors and covariates [see for example, Stute, González Manteiga and Presedo Quindimil (1998)], which is clearly not desired in our test. ■

4 The univariate design case

A critical point of discussion of Zheng's (1997) test statistic is the asymmetry in the roles of X_i and ε_i, which is not desirable in the context of testing independence of components in an iid-sample (X_i, ε_i), $i = 1, \ldots, n$. However, in our case, where ε_i has to be estimated before performing the test, this asymmetry turns out to be an advantage because it allows us to consider multivariate design in contrast to the so far available procedures by Einmahl and Van Keilegom (2007a, 2007b). In the case $d = 1$ we discuss in this section an alternative test statistic interchanging the roles of X_i and $\hat{\varepsilon}_i$. Of course, then also symmetrized versions of the test could be considered.

We assume during this section that ε_i and X_i are independent, i. e. H_0 is valid. It turns out that even then replacing residuals by true errors changes the asymptotic behaviour of the test statistic seriously. Let us consider model (1.1) for $d = 1$. We define the test statistic

$$S_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h_n} K\Big(\frac{X_i - X_j}{h_n}\Big) \int (I\{\hat{\varepsilon}_i \leq y\} - \hat{F}_{\varepsilon,n}(y))(I\{\hat{\varepsilon}_j \leq y\} - \hat{F}_{\varepsilon,n}(y)) w(y)\, dy,$$

where $\hat{F}_{\varepsilon,n}$ denotes the empirical distribution function of residuals $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$. Akritas and Van Keilegom (2001) prove weak convergence of the process $\sqrt{n}(\hat{F}_{\varepsilon,n} - F_\varepsilon)$ (for one-dimensional covariates) under some regularity conditions. Please note that analogous results for multivariate design have not been proven so far and Akritas and Van Keilegom's (2001) proof is not transferable to that case.

Denote by $\tilde{S}_n$ the (not available) statistic defined analogously to S_n, but replacing all $\hat{\varepsilon}_i$ by the true errors ε_i and $\hat{F}_{\varepsilon,n}$ by $F_{\varepsilon,n}$, the empirical distribution function of errors $\varepsilon_1, \ldots, \varepsilon_n$. Zheng (1997) showed under H_0 that $nh_n^{1/2}\tilde{S}_n$ converges to a mean zero normal distribution with variance

$$\tilde{\gamma}^2 = 2\int K^2(u)\,du \int f_X^2(x)\,dx \int\int (F_\varepsilon(y \wedge z) - F_\varepsilon(y)F_\varepsilon(z))^2 w(y)w(z)\,dy\,dz.$$

We introduce the notations

$$\begin{aligned} U(y) &= E[\varepsilon_1 I\{\varepsilon_1 \le y\}], \quad V(y) \;=\; \int (I\{r \le y\} - F_\varepsilon(y)) f_\varepsilon'(r)\,dr \\ C_n &= \frac{K(0)}{nh_n}\Big(2\int U(y)V(y)w(y)\,dy + \sigma^2 \int V^2(y)w(y)\,dy\Big). \end{aligned}$$

It is interesting that here even under H_0 we obtain a bias term C_n in contrast to the results in section 3 (see Theorem 3.1). Further the asymptotic distribution is much more complicated. Under suitable regularity conditions it can be shown that under the null hypothesis (1.2) of independence $nh_n^{1/2}(S_n - C_n)$ converges to a mean zero normal distribution with variance

$$\begin{aligned} \gamma^2 \;=\; & \tilde{\gamma}^2 + 2\int K^2(u)\,du \int f_X^2(x)\,dx \Big[\sigma^4 \Big(\int V^2(y)w(y)\,dy\Big)^2 \\ & + 4\sigma^2 \int\int (F_\varepsilon(y \wedge z) - F_\varepsilon(y)F_\varepsilon(z))V(y)V(z)w(y)w(z)\,dy\,dz \\ & + 4\Big(\int U(y)V(y)w(y)\,dy\Big)^2 + 8\int\int U(y)(F_\varepsilon(y \wedge z) - F_\varepsilon(y)F_\varepsilon(z))V(z)w(y)w(z)\,dy\,dz \\ & + 8\int U^2(y)w(y)\,dy \int V^2(y)w(y)\,dy + 8\sigma^2 \int U(y)V(y)w(y)\,dy \int V^2(y)w(y)\,dy\Big]. \end{aligned}$$

Because the proof uses results by Akritas and Van Keilegom (2001) we run into the same problems as Einmahl and Van Keilegom (2007b) in terms of generalization to the multivariate design case. Further for test statistics based on the empirical distribution function of residuals theory relies substantially on the smoothness of the error distribution. Hence a smooth residual bootstrap [as was discussed by Neumeyer (2006)] should be applied instead of the classical residual bootstrap. For this kind of bootstrap an additional smoothing parameter has to be chosen, such that for the bootstrap version of S_n one needs three different bandwidths. Because of these disadvantages we do not recommend to apply test statistic S_n, but rather T_n as defined in section 2. Hence, details of the derivations of the above asymptotic result are omitted.

5 The heteroscedastic model

In the following we consider a modification of the test statistic to detect dependence of the standardized error from the covariate in a heteroscedastic model, i. e. we are testing hypothesis (1.2) in model (1.3). To this end the errors $\varepsilon_i = (Y_i - E[Y_i \mid X_i])/(\mathrm{Var}(Y_i \mid X_i))^{1/2}$ are estimated by residuals

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

where

$$\hat{\sigma}^2(x) = \frac{1}{nb_n^d} \sum_{i=1}^n k\Big(\frac{X_i - x}{b_n}\Big)(Y_i - \hat{m}(X_i))^2 \frac{1}{\hat{f}_X(x)}$$

and $\hat{m}$, $\hat{f}_X$ are defined in (2.2) and (2.3), respectively. With this different definition of residuals the same test statistic as defined in (2.4) can be applied for testing (1.2). Additional to the assumptions in section 2 we now assume that σ^2 is twice continuously differentiable and bounded away from zero. Then under the null hypothesis of independence, $nh_n^{1/2}T_n$ has the same limit distribution as given in Theorem 3.1. For the bootstrap version of the test, similar to (3.2) bootstrap observations are defined as

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

where now ε_i^* is generated from the empirical distribution function of the standardized residuals

$$\tilde{\varepsilon}_i = \frac{\hat{\varepsilon}_i - n^{-1}\sum_{l=1}^n \hat{\varepsilon}_l}{(\sum_{k=1}^n (\hat{\varepsilon}_k - \sum_{l=1}^n \hat{\varepsilon}_l)^2)^{1/2}}, \quad i = 1, \dots, n.$$

6 Finite sample performance

First we compare our procedure to the so far availabe procedures by Einmahl and Van Keilegom (2007a, 2007b) in the univariate case, which have been shown to be superior to tests for heteroscedasticity in cases where higher error moments depend on the design.

In our experiments the design variable X_1 is uniformly distributed in $[0, 1]$ and the regression function is $m(x) = x - x^2/2$. Sample sizes are $n = 100$ and $n = 200$. Under the null hypothesis, ε_1 is centered normally distributed with standard deviation $\sigma = 0.1$. We use R for the simulations [R Development Core Team (2006)]. For the regression estimation we apply procedure *sm.regression* in R package *sm* [Bowman and Azzalini (1997)] and procedure *h.select* for a choice of the bandwidth b_n via cross validation. The bandwidth h_n is chosen by a rule of thumb: $h_n^* = (\hat{s}_n^2/n)^{1/5}$, where $\hat{s}_n^2$ denotes an estimator for the variance, i. e. the variance estimator by Rice (1984) applied to the residuals. To investigate how sensitive the results are with respect to the choice of this bandwidth we also display simulation results for

$h_n \in \{h_n^*, h_n^*/2, 2h_n^*\}$. The simulation results are based on 500 simulations with 250 bootstrap repetitions in each simulation. The bootstrap samples are generated according to the classical residual bootstrap as explained in section 3. The nominal level is $\alpha = 0.05$. We consider the following alternatives investigated by Einmahl and Van Keilegom (2007a, 2007b): The conditional distributions of the error ε_1 given that $X_1 = x$, are normal, chisquared and t-distributions defined by

$$
\begin{aligned}
H_{1,A} &: \varepsilon_1 | X_1 = x \sim N(0, \frac{1+ax}{100}) \\
H_{1,B} &: \varepsilon_1 | X_1 = x \stackrel{\mathcal{D}}{=} \frac{W_x - r_x}{10\sqrt{2r_x}}, \text{ where } W_x \sim \chi^2_{r_x}, r_x = 1/(bx) \\
H_{1,C} &: \varepsilon_1 | X_1 = x \stackrel{\mathcal{D}}{=} \frac{1}{10}\sqrt{(1-cx)^{1/4}} T_x, \text{ where } T_x \sim t_{2/(cx)^{1/4}}
\end{aligned}
$$

with parameters $a, b > 0$, $c \in (0, 1]$ that control variance, skewness and kurtosis, respectively.

$H_{1,A}$	$n = 100$			$n = 200$		
$a \setminus h_n$	h_n^*	$2h_n^*$	$h_n^*/2$	h_n^*	$2h_n^*$	$h_n^*/2$
0	0.054	0.032	0.048	0.047	0.044	0.054
1	0.222	0.190	0.148	0.352	0.390	0.254
2.5	0.476	0.506	0.358	0.748	0.860	0.624
5	0.706	0.734	0.508	0.932	0.976	0.848
10	0.822	0.884	0.678	0.976	0.972	0.998

Table 1: Rejection probabilities for model $H_{1,A}$

$H_{1,B}$	$n = 100$			$n = 200$		
$b \setminus h_n$	h_n^*	$2h_n^*$	$h_n^*/2$	h_n^*	$2h_n^*$	$h_n^*/2$
0	0.054	0.032	0.048	0.047	0.044	0.054
1	0.122	0.084	0.212	0.224	0.108	0.416
2.5	0.140	0.104	0.224	0.228	0.106	0.436
5	0.162	0.134	0.272	0.252	0.152	0.400
10	0.146	0.116	0.225	0.240	0.126	0.360

Table 2: Rejection probabilities for model $H_{1,B}$

$H_{1,C}$	$n = 100$			$n = 200$		
$c \setminus h_n$	h_n^*	$2h_n^*$	$h_n^*/2$	h_n^*	$2h_n^*$	$h_n^*/2$
0	0.054	0.032	0.048	0.047	0.044	0.054
0.2	0.084	0.050	0.096	0.114	0.050	0.156
0.4	0.090	0.106	0.120	0.192	0.154	0.216
0.6	0.174	0.138	0.224	0.304	0.214	0.412
0.8	0.308	0.272	0.420	0.536	0.372	0.688
1	0.540	0.438	0.676	0.830	0.648	0.904

Table 3: Rejection probabilities for model $H_{1,C}$

The results are given in tables 1–3. For model $H_{1,A}$ they show a behaviour very similar to the best simulation results by Einmahl and Van Keilegom (2007b) in the same setting. In model $H_{1,B}$ Einmahl and Van Keilegom's (2007b) test has larger rejection probabilities than the new test, but for model $H_{1,C}$ results are again very similar

We consider additionally the alternatives $H_{1,D}$, where $(X_1, \varepsilon_1) = (X, U - E[U|X])$, where (X, U) are generated with the Farlie-Gumbel-Morgenstern copula as distribution function with parameter $a \in \{0, 1, 2, 3, 5\}$ [here $E[U|X] = \frac{1}{2} - \frac{a}{6}(2X - 1)$]. Further the heteroscedastic model $H_{1,D,h}$ is defined by model (1.3) with variance function $\sigma^2(x) = (2 + x)^2/100$ and $\varepsilon_1 | X_1 = x$ distributed as in $H_{1,D}$. The rejection probabilities for $H_{1,D}$ and $H_{1,D,h}$ are displayed in table 4 and show slightly underestimated levels, but good power properties.

$H_{1,D}$	$n = 100$	$n = 200$	$H_{1,D,h}$	$n = 100$	$n = 200$
$b \setminus h_n$	h_n^*	h_n^*	$c \setminus h_n$	h_n^*	h_n^*
0	0.024	0.034	0	0.036	0.035
1	0.172	0.620	1	0.278	0.630
2	0.284	0.926	2	0.388	0.774
3	0.452	0.998	3	0.190	0.402
5	0.662	1.000	5	0.156	0.268

Table 4: Rejection probabilities for model $H_{1,D}$ and heteroscedastic model $H_{1,D,h}$

Finally, we give a simulation example in a bivariate design case. Here the covariates X_i are distributed with the Farlie-Gumbel-Morgenstern copula with parameter 3 as distribution function (and uniformly distributed marginals). The regression function is $m(x_1, x_2) =$

$x_1 - x_2^2/2$ and we consider the following model,

$$H_{1,A(2)} \; : \; \varepsilon_1 | X_1 = (x_1, x_2) \sim N(0, (1 + a(x_1 + x_2))\frac{1}{100})$$

The results are given in table 5 and show very good detection of the alternatives, but levels are overestimated in the case $n = 100$. Please note that there exist no competing procedures for our test statistic in this setting.

$H_{1,A(2)}$	$n = 100$			$n = 200$		
$a \setminus h_n$	h_n^*	$2h_n^*$	$h_n^*/2$	h_n^*	$2h_n^*$	$h_n^*/2$
0	0.104	0.100	0.086	0.076	0.068	0.054
1	0.294	0.172	0.272	0.574	0.574	0.482
2.5	0.620	0.532	0.484	0.938	0.940	0.818
5	0.802	0.778	0.646	0.988	0.994	0.912
10	0.878	0.904	0.694	1.000	1.000	0.950

Table 5: Rejection probabilities for model $H_{1,A(2)}$ with bivariate covariates

To conclude this section we consider the application of the method to a real data example. We used $n = 200$ observations of the aircraft-based LIDAR (Light Detection and Ranging) data set as available in the R-library MBA (the sample size is chosen to be 200 to give a similar setting as in the simulations; for the data structure to remain intact we chose every 50th observation in the original data set). The covariates are bivariate (longitude and latitude), measured is the ground surface elevation. Figure 1 shows an estimator of the regression function after transformation of the covariates into the unit interval.

We tested for the homoscedastic regression model (1.1) with $B = 1000$ bootstrap replications. The bandwidth selection is as described for the simulations. Here we have $h_n^* = 0.439$, for which we obtain 0.003 as p-value, so clearly the hypothesis (1.2) for model (1.1) is rejected. This result is not sensitive to the choice of the bandwidth; we obtain p-values of 0.009, 0.005, 0.003, 0.001, 0.001 for bandwidhts $h_n = 0.2$, 0.3, 0.4, 0.5, and 0.6, respectively. Please note that heteroscedasticity in LIDAR data was observed, for instance, by Lindström, Holst and Weibring (2005). We then tested hypothesis (1.2) for the heteroscedastic model (1.3) for the same data set and setting as above with $h_n^* = 0.339$. For bandwidths $h_n = 0.1$, 0.2, 0.3, 0.4, 0.6 and h_n^* we have p-values of 0.000, 0.005, 0.007, 0.013, 0.014, and 0.009, respectively. Hence, this model is also rejected and methods applied to analyse this data set should take into account an error distribution dependend of the covariate.

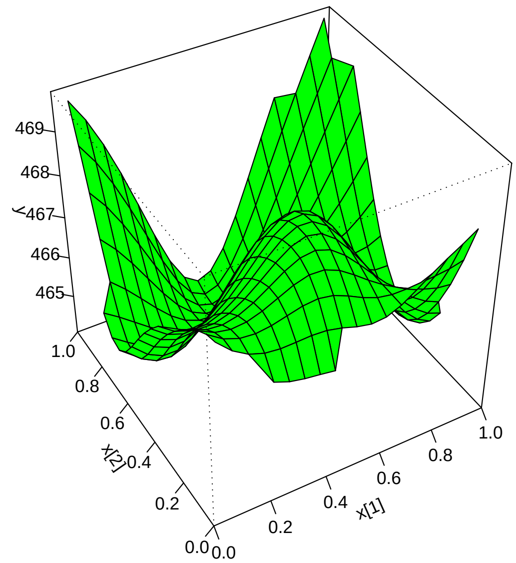

Figure 1: Estimated regression curve in the LIDAR data set

7 Conclusion

This paper proposes a new test for independence of error and covariate in the nonparametric regression model. The simple kernel based test statistic has an asymptotic normal law. It can be applied to models with multivariate covariates, which is very important for applications, for example in econometrics. It can also be adjusted to test for independence of a standardized error and the covariate in heteroscedastic regression models. We suggest to apply a residual bootstrap version of the test and we investigate its behaviour in theory as well as in simulations.

So far in literature there are only two tests available for the same testing problem. Einmahl and Van Keilegom (2007a) propose a very innovative procedure that is based on a stochastic process of differences of the observations, which converges weakly to a bivariate Gaussian process. Generalizations to multivariate covariates or heteroscedasticity as well as universal consistency are not clear for this test.

Einmahl and Van Keilegom's (2007b) test is valid for the homoscedastic as well as heteroscedastic model and is based on differences of empirical distribution functions. This test is asymptotically distribution free, which is a very nice theoretical result. The authors nevertheless suggest the application of a bootstrap version of the test. This procedures needs a bandwidth b_n for the estimation of the regression and variance function, but avoids the choice of a second bandwidth h_n as we need for our test. It is not clear whether Einmahl and Van Keilegom's (2007b) test can be generalized to higher dimensional covariates because so far no weak convergence results for the empirical process of residuals in nonparametric regression models with multivariate covariates are available in literature.

Our test is a version of Zheng's (1997) test, but applied to covariates and (dependent) residuals instead of an independent sample of paired observations. Zheng (1997) and Dette and Neumeyer (2000) gave results on asymptotic distributions of the latter test, but no bootstrap version has been considered so far. Zheng's (1997) test could be criticized for its asymmetry in interchanging the components of the paired observations. However, in our context this asymmetry is desired and essential. When we interchange the roles of residuals and covariates in the test statistic, asymptotic theory so far is only available for the univariate design case (and is much more complicated). This situation is comparable to Einmahl and Van Keilegom (2007b).

The simulation results show comparable results to Einmahl and Van Keilegom (2007a, 2007b) in most (univariate) settings and good power properties in a multivariate setting, where no competing procedures exist in literature so far.

A Proofs

A.1 Proof of Theorem 3.1

From the definitions of T_n in (2.4) and $\tilde{T}_n$ in (3.1) it follows that

$$\begin{aligned} T_n - \tilde{T}_n \;&=\; \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h_n} \Big(K\Big(\frac{\hat{\varepsilon}_i - \hat{\varepsilon}_j}{h_n}\Big) - K\Big(\frac{\varepsilon_i - \varepsilon_j}{h_n}\Big)\Big) \\ &\qquad \times \int (I\{X_i \leq x\} - F_{X,n}(x))(I\{X_j \leq x\} - F_{X,n}(x)) w(x)\, dx. \end{aligned}$$

By a Taylor expansion we obtain $T_n - \tilde{T}_n = \sum_{\ell=1}^{\lambda} \frac{1}{\ell!} V_n^{(\ell)}$, where

$$\begin{aligned} V_n^{(\ell)} \;&=\; \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h_n^{\ell+1}} K^{(\ell)}\Big(\frac{\varepsilon_i - \varepsilon_j}{h_n}\Big)(m(X_i) - \hat{m}(X_i) - m(X_j) + \hat{m}(X_j))^{\ell} H_n(X_i, X_j) \\ &\qquad (\ell = 1, \ldots, \lambda - 1) \\ V_n^{(\lambda)} \;&=\; \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h_n^{\lambda+1}} K^{(\lambda)}(\xi_{i,j,n})(m(X_i) - \hat{m}(X_i) - m(X_j) + \hat{m}(X_j))^{\lambda} H_n(X_i, X_j) \end{aligned}$$

with $\xi_{i,j,n}$ between $(\varepsilon_i - \varepsilon_j)/h_n$ and $(\hat{\varepsilon}_i - \hat{\varepsilon}_j)/h_n$, and

$$H_n(X_i, X_j) \;=\; \int (I\{X_i \leq x\} - F_{X,n}(x))(I\{X_j \leq x\} - F_{X,n}(x)) w(x)\, dx$$

$$= \frac{1}{n^2} \sum_{\substack{k=1\\k\neq i}}^n \sum_{\substack{l=1\\l\neq j}}^n \int \eta_{i,k}(x)\eta_{j,l}(x)w(x)\,dx.$$

Here the $\eta_{i,k}(x) = I\{X_i \le x\} - I\{X_k \le x\}$ are centered and bounded. The following lemma gives the asymptotic behaviour of V_n^ℓ, $\ell = 1, \dots, \lambda$.

Lemma A.1 *Under the assumptions of section 2, under the null hypothesis H_0 we have $V_n^{(\ell)} = o_p(1/(nh_n^{1/2}))$ for $\ell = 1, \dots, \lambda$, whereas under the alternative H_1, $V_n^{(\ell)} = o_p(1/\sqrt{n})$ for $\ell = 2, \dots, \lambda$, and $V_n^{(1)} = \tilde{T}_n^{(1)} + o_p(1/\sqrt{n})$, where $\tilde{T}_n^{(1)} = \frac{2}{n}\sum_{i=1}^n w_{ni}$ and*

$$\begin{aligned} w_{ni} &= \varepsilon_i \int\int\int\int \frac{1}{b_n^d} k\Big(\frac{X_i - t}{b_n}\Big) \frac{f_{X|\varepsilon}(t|z)}{f_X(t)} \frac{\partial f_{X|\varepsilon}(s|z) f_\varepsilon(z)}{\partial z} (I\{t \le x\} - F_X(x)) \\ &\quad \times (I\{s \le x\} - F_X(x)) f_\varepsilon(z) w(x)\,dx\,dt\,dz\,ds \int K'(u)u\,du. \end{aligned}$$

Proof of Lemma A.1. Note that using symmetry of K we obtain

$$V_n^{(1)} = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1\\j\neq i}}^n \frac{1}{h_n^2} K'\Big(\frac{\varepsilon_i - \varepsilon_j}{h_n}\Big)(\hat{m}(X_j) - m(X_j)) H_n(X_i, X_j).$$

Applying further the decomposition $\hat{m}(X_j) - m(X_j) = \mu_{j,n}^{(1)} + \mu_{j,n}^{(2)} + \mu_{j,n}^{(3)} + \mu_{j,n}^{(4)}$, where

$$\begin{aligned} \mu_{j,n}^{(1)} &= \frac{1}{nb_n^d} \sum_{\nu=1}^n k\Big(\frac{X_\nu - X_j}{b_n}\Big)\varepsilon_\nu \frac{1}{f_X(X_j)} \\ \mu_{j,n}^{(2)} &= \frac{1}{nb_n^d} \sum_{\nu=1}^n k\Big(\frac{X_\nu - X_j}{b_n}\Big)(m(X_\nu) - m(X_j)) \frac{1}{f_X(X_j)} \\ \mu_{j,n}^{(3)} &= \frac{1}{nb_n^d} \sum_{\nu=1}^n k\Big(\frac{X_\nu - X_j}{b_n}\Big)\varepsilon_\nu \Big(\frac{1}{\hat{f}_X(X_j)} - \frac{1}{f_X(X_j)}\Big) \\ \mu_{j,n}^{(2)} &= \frac{1}{nb_n^d} \sum_{\nu=1}^n k\Big(\frac{X_\nu - X_j}{b_n}\Big)(m(X_\nu) - m(X_j)) \Big(\frac{1}{\hat{f}_X(X_j)} - \frac{1}{f_X(X_j)}\Big) \end{aligned}$$

we obtain a decomposition $V_n^{(1)} = U_n^{(1)} + U_n^{(2)} + U_n^{(3)} + U_n^{(4)}$, where $(k = 1, \dots, 4)$

$$U_n^{(k)} = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1\\j\neq i}}^n \frac{1}{h_n^2} K'\Big(\frac{\varepsilon_i - \varepsilon_j}{h_n}\Big)\mu_{j,n}^{(k)}.$$

To begin with we consider

$$\begin{aligned} U_n^{(1)} &= \frac{2}{n^4(n-1)h_n^2 b_n^d} \sum_{i=1}^n \sum_{\substack{j=1\\j\neq i}}^n \sum_{\substack{k=1\\k\neq i}}^n \sum_{\substack{l=1\\l\neq j}}^n \sum_{\nu=1}^n K'\Big(\frac{\varepsilon_i - \varepsilon_j}{h_n}\Big)\varepsilon_\nu \\ &\quad \times k\Big(\frac{X_\nu - X_j}{b_n}\Big) \frac{1}{f_X(X_j)} \int \eta_{i,k}(x)\eta_{j,l}(x)w(x)\,dx. \end{aligned}$$

Because by definition $E[\varepsilon_\nu \mid X_1, \ldots, X_n] = 0$, for the expectation $E[U_n^{(1)}]$, only terms where $\nu \in \{i, j\}$ can have a nonvanishing impact on the sum. For example for $\nu = i$ (all other indices different) one obtains

$$
\begin{aligned}
& E\Big[K'\Big(\frac{\varepsilon_1 - \varepsilon_2}{h_n}\Big)\varepsilon_1 k\Big(\frac{X_1 - X_2}{b_n}\Big)\frac{1}{f_X(X_2)}\int \eta_{1,3}(x)\eta_{2,4}(x)w(x)\,dx\Big] \\
= & \int\!\!\int\!\!\int\!\!\int K'\Big(\frac{y-z}{h_n}\Big)yk\Big(\frac{s-t}{b_n}\Big)\frac{1}{f_X(t)}f_\varepsilon(y)f_\varepsilon(z)f_{X|\varepsilon}(s|y)f_{X|\varepsilon}(t|z)f_X(v)f_X(w) \\
& \times \int (I\{s \le x\} - I\{v \le x\})(I\{t \le x\} - I\{w \le x\})w(x)\,dx\,d(s,y)\,d(t,z)\,dv\,dw \\
= & \int\!\!\int\!\!\int\!\!\int\!\!\int\!\!\int K'(u)yk(r)\frac{1}{f_X(t)}f_\varepsilon(y)f_\varepsilon(y - h_n u)f_{X|\varepsilon}(t + b_n r|y)f_{X|\varepsilon}(t|y - h_n u)f_X(v)f_X(w) \\
& \times \int (I\{t + b_n r \le x\} - I\{v \le x\})(I\{t \le x\} - I\{w \le x\})w(x)\,dx\,du\,dt\,dr\,dy\,dv\,dw\,h_n b_n^d \\
= & O(h_n^2 b_n^d),
\end{aligned}
$$

where the last equality is obtained by a Taylor expansion of $f_\varepsilon(y-h_n u)f_{X|\varepsilon}(t|y-h_n u)$ around y and we use the assumptions, especially $\int K'(u)\,du = 0$. Similar considerations for $\nu = j$ (all other indices different) yield under the alternative the rate $O(h_n^2)$, because one still can use $\int K'(u)\,du = 0$, but no substitution to gain the factor b_n^d is possible. Under the null hypothesis of independence of ε_i and X_i however one can apply that $E[\eta_{i,k}(x)] = 0$ and, hence, the summand in the expectation does not vanish only when $\{i, k\} \cap \{j, l\} \neq \emptyset$. This yields altogether

$$
E[U_n^{(1)}] \;=\; \frac{2}{n^4(n-1)h_n^2 b_n^d}\Big[O(n^4)O(h_n^2 b_n^d) + O(n^3)O(h_n^2)\Big] = O(\frac{1}{n}) + O(\frac{1}{n^2 b_n^d}) = o(\frac{1}{n h_n^{1/2}})
$$

under the null hypothesis H_0, and

$$
E[U_n^{(1)}] \;=\; \frac{2}{n^4(n-1)h_n^2 b_n^d}\Big[O(n^4)O(h_n^2 b_n^d) + O(n^4)O(h_n^2)\Big] = O(\frac{1}{n}) + O(\frac{1}{n b_n^d}) = o(\frac{1}{n^{1/2}})
$$

under fixed alternatives H_1.

Next, we show that $E[(U_n^{(1)})^2] = o(1/(n^2 h_n))$ under the null hypothesis H_0 of independence of errors and covariates. To this end we calculate

$$
\begin{aligned}
E[(U_n^{(1)})^2] \;=\; & O(\frac{1}{n^{10}h_n^4 b_n^{2d}})\sum_{i=1}^n \sum_{\substack{j=1\\ j\neq i}}^n \sum_{\substack{k=1\\ k\neq i}}^n \sum_{\substack{l=1\\ l\neq j}}^n \sum_{\nu=1}^n \sum_{i'=1}^n \sum_{\substack{j'=1\\ j'\neq i'}}^n \sum_{\substack{k'=1\\ k'\neq i'}}^n \sum_{\substack{l'=1\\ l'\neq j'}}^n \sum_{\nu'=1}^n \\
& E\Big[K'\Big(\frac{\varepsilon_i - \varepsilon_j}{h_n}\Big)K'\Big(\frac{\varepsilon_{i'} - \varepsilon_{j'}}{h_n}\Big)\varepsilon_\nu\varepsilon_{\nu'}k\Big(\frac{X_\nu - X_j}{b_n}\Big)k\Big(\frac{X_{\nu'} - X_{j'}}{b_n}\Big)\frac{1}{f_X(X_j)f_X(X_{j'})} \\
& \times \int \eta_{i,k}(x)\eta_{j,l}(x)w(x)\,dx \int \eta_{i',k'}(x)\eta_{j',l'}(x)w(x)\,dx\Big].
\end{aligned}
$$

Here, the term

$$E\Big[K'\Big(\frac{\varepsilon_i-\varepsilon_j}{h_n}\Big)K'\Big(\frac{\varepsilon_{i'}-\varepsilon_{j'}}{h_n}\Big)\varepsilon_\nu\varepsilon_{\nu'}\Big]$$

is zero unless $\nu=\nu'$ or at least two pairs of indices are equal such as $\nu=j$ and $\nu'=i'$ or similarly. When the other indices are distinct, the term is of order $O(h_n^4)$ which follows, as before, from $\int K'(u)\,du=0$. Under the null hypothesis we further can exploit the fact that $E[\eta_{i,k}(x)]=0$ and obtain the rate

$$E[(U_n^{(1)})^2] \;=\; O(\frac{1}{n^{10}h_n^4b_n^{2d}})\Big[O(n^8h_n^4b_n^{2d})+O(n^7h_nb_n^d)\Big]=O(\frac{1}{n^2})+O(\frac{1}{n^3h_n^3b_n^d})=o(\frac{1}{n^2h_n})$$

by our bandwidth assumptions. We obtain by Markov's inequality that $U_n^{(1)}$ is of order $o_p(1/(nh_n^{1/2}))$ under H_0 and, hence, negligible.

However, under the alternative with similar calculations one only obtains $E[(U_n^{(1)})^2]=O(\frac{1}{n})$ and, hence, this term is not negligible under H_1. $U_n^{(1)}$ is approximately a U–statistic of order 5. By symmetrizing its kernel and applying a Hoeffding decomposition one can show that the dominating term (with remainder of order $o_p(1/n^{1/2})$) are the U–statistics of order 1 in the Hoeffding decomposition. Because ε_ν is centered those are

$$\begin{aligned}\tilde U_n^{(1)} \;=\; &\frac{2}{n}\sum_{\nu=1}^n\varepsilon_\nu E\Big[\frac{1}{h_n^2b_n^d}K'\Big(\frac{\varepsilon_i-\varepsilon_j}{h_n}\Big)k\Big(\frac{X_\nu-X_j}{b_n}\Big)\frac{1}{f_X(X_j)}\\&\times\int\eta_{i,k}(x)\eta_{j,l}(x)w(x)\,dx\;\Big|\;X_\nu\Big],\end{aligned}$$

where the conditional expectation is evaluated for $\nu\cap\{i,j,k,l\}=\emptyset$. This gives

$$\begin{aligned}\tilde U_n^{(1)} \;=\; &\frac{2}{n}\sum_{\nu=1}^n\varepsilon_\nu\int\int\frac{1}{h_n^2}K'\Big(\frac{y-z}{h_n}\Big)\int\frac{1}{b_n^d}k\Big(\frac{X_\nu-t}{b_n}\Big)\frac{1}{f_X(t)}\int(I\{s\le x\}-F_X(x))\\&\times(I\{t\le x\}-F_X(x))w(x)\,dx\,f_{X|\varepsilon}(t|z)f_{X|\varepsilon}(s|y)f_\varepsilon(y)f_\varepsilon(z)\,d(s,y)d(s,z).\end{aligned}$$

With similar methods as before one shows that $\tilde U_n^{(1)}=\tilde T_n^{(1)}+o_p(n^{-1/2})$.

Next, we show that the remaining terms, $U_n^{(k)}$ $(k=2,3,4)$ and V_n are negligible both under H_0 and H_1. To this end we first consider the expectation $E[U_n^{(2)}]$ for

$$\begin{aligned}U_n^{(2)} \;=\; &\frac{2}{n^4(n-1)h_n^2b_n^d}\sum_{i=1}^n\sum_{\substack{j=1\\j\ne i}}^n\sum_{\substack{k=1\\k\ne i}}^n\sum_{\substack{l=1\\l\ne j}}^n\sum_{\nu=1}^n K'\Big(\frac{\varepsilon_i-\varepsilon_j}{h_n}\Big)(m(X_\nu)-m(X_j))\\&\times k\Big(\frac{X_\nu-X_j}{b_n}\Big)\frac{1}{f_X(X_j)}\int\eta_{i,k}(x)\eta_{j,l}(x)w(x)\,dx.\end{aligned}$$

The argumentation is similar to that for $E[U_n^{(1)}]$. However we cannot use the fact of centered errors here. Instead, under the null hypothesis we use $E[\eta_{i,k}(x)]=0$ to explain that at least two indices have to be equal to gain nonvanishing terms in the sum. Further, $\int K'(u)\,du = 0$ can be exploited as before, and the factor b_n^d can always be gained by a substitution in the integral because the terms for $\nu = j$ vanish due to the factor $m(X_\nu) - m(X_j)$. We obtain the rate

$$E[U_n^{(2)}] \;=\; \frac{2}{n^4(n-1)h_n^2 b_n^d} O(n^4) O(h_n^2 b_n^d) = O(\frac{1}{n}) = o(\frac{1}{nh_n^{1/2}})$$

under H_0. Under H_1 however, even the term where all indices are distinct gives the dominating term, i. e.

$$\begin{aligned}
& E\Big[K'\Big(\frac{\varepsilon_1-\varepsilon_2}{h_n}\Big)(m(X_3)-m(X_2))k\Big(\frac{X_3-X_2}{b_n}\Big)\frac{1}{f_X(X_2)}\int \eta_{1,4}(x)\eta_{2,5}(x)w(x)\,dx\Big] \\
&= \int\int\int\int\int K'\Big(\frac{y-z}{h_n}\Big)(m(u)-m(t))k\Big(\frac{u-t}{b_n}\Big)\frac{f_X(u)f_X(v)f_X(w)}{f_X(t)} f_\varepsilon(y)f_\varepsilon(z)f_{X|\varepsilon}(s|y) \\
&\quad \times f_{X|\varepsilon}(t|z)\int (I\{s\le x\}-I\{v\le x\})(I\{t\le x\}-I\{w\le x\})w(x)\,dx\,d(s,y)\,d(t,z)\,du\,dv\,dw \\
&= \int\int\int\int\int\int \Big[K'(u)f_\varepsilon(z+h_n u)f_{X|\varepsilon}(s|z+h_n u)\,du\Big] f_X(v)f_X(w)f_\varepsilon(z)g(t|z) \\
&\quad \times \Big[\int k(r)(m(t+b_n r)-m(t))\frac{f_X(t+b_n r)}{f_X(t)}\,dr\Big] \\
&\quad \times \int (I\{s\le x\}-I\{v\le x\})(I\{t\le x\}-I\{w\le x\})w(x)\,dx\,ds\,dt\,dz\,dv\,dw\,h_n b_n^d.
\end{aligned}$$

Applying Taylor expansions and utilizing that $\int K'(u)\,du = 0$ and $\int \kappa(u)u\,du = 0$, we obtain the rate $O(h_n^2 b_n^{3d})$, which yields

$$E[U_n^{(2)}] \;=\; \frac{2}{n^4(n-1)h_n^2 b_n^d} O(n^5) O(h_n^2 b_n^{3d}) = O(b_n^{2d}) = o(\frac{1}{n^{1/2}})$$

under H_1, by the bandwidth assumptions.

With similar considerations as before one can show that $E[(U_n^{(2)})^2]$ has, under H_0 the following rate,

$$O(\frac{1}{n^{10}h_n^4 b_n^{2d}})\Big[O(n^9 h_n^4 b_n^{6d}) + O(n^8 h_n b_n^{6d}) + O(n^8 h_n^4 b_n^2)\Big] = o(\frac{1}{n^2 h_n}),$$

where we have used the bandwidth conditions $b_n^{2d}/h_n \to 0$ and $h_n/b_n^{d-2} \to 0$. From this we obtain $U_n^{(2)} = o_p(1/(nh_n^{1/2}))$ under H_0.

Under the alternative, applying the bandwidth conditions $nb_n^{4d} \to 0$, $b_n^{4d}/(nh_n^3) \to 0$, $nb_n^{d-2} \to \infty$, one can show $E[(U_n^{(2)})^2] = o_p(1/n)$, and obtains that $U_n^{(2)}$ is of order $o_p(1/n^{1/2})$ under H_1.

Similar argumentations to before show that $U_n^{(l)}$ for $l = 3, 4$ and $V_n^{(\ell)}$ for $\ell = 2, \ldots, \lambda - 1$ are of order $o_p(1/(nh_n^{1/2}))$ or $o_p(1/n^{1/2})$ under H_0 and H_1, respectively.

To conclude the proof we show that $V_n^{(\lambda)}$ is negligible. This term has to be treated differently because $\xi_{i,j,n}$ depends on all observations and, hence, to obtain rates of convergence for the expectation and variance of $V_n^{(\lambda)}$ is difficult. Instead we use the direct bound [compare Härdle and Mammen (1993, proof of Prop. 1) for the uniform convergence rate]

$$|V_n^{(\lambda)}| \;\leq\; O(\frac{1}{h_n^{\lambda+1}}) \sup_{x\in[0,1]^d} |m(x)-\hat{m}(x)|^\lambda \;=\; O_p\Big(\frac{1}{h_n^{\lambda+1}(nb_b^d)^{\lambda\rho}}\Big) \;=\; o_p(\frac{1}{n\sqrt{h_n}}) \;=\; o_p(\frac{1}{\sqrt{n}})$$

for $0<\rho<\frac{1}{2}$ by the bandwidth assumptions. □

From the decomposition $T_n = \tilde{T}_n + \sum_{\ell=1}^{\lambda}\frac{1}{\ell!}V_n^{(\ell)}$, Lemma A.1, and the results by Zheng (1997) (see Remark 3.2) directly follows part (a) of Theorem 3.1.

To conclude the proof of part (b) of Theorem 3.1 under the alternative we have to consider $\tilde{T}_n + \tilde{T}_n^{(1)}$ in the following to obtain the asymptotic distribution. From results by Neumeyer and Dette (2000) we have under H_1 with notations from (2.6) and $I_i(x) = I\{X_i \leq x\} - F_{X|\varepsilon}(x|\varepsilon_i)$,

$$\sqrt{n}(\tilde{T}_n - B_{h_n}) \;=\; \sqrt{n}(W_n^{(1)} - E[W_n^{(1)}] + 2W_n^{(2)} - 2W_n^{(3)}) + o_p(1), \qquad \text{(A.1)}$$

where

$$W_n^{(1)} \;=\; \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{\substack{j=1\\ j\neq i}}^n \frac{1}{h_n}K\Big(\frac{\varepsilon_i-\varepsilon_j}{h_n}\Big)\int \Delta(x|\varepsilon_i)\Delta(x|\varepsilon_j)w(x)\,dx$$

$$W_n^{(2)} \;=\; \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{\substack{j=1\\ j\neq i}}^n \frac{1}{h_n}K\Big(\frac{\varepsilon_i-\varepsilon_j}{h_n}\Big)\int I_i(x)\Delta(x|\varepsilon_j)w(x)\,dx$$

$$W_n^{(3)} \;=\; \frac{1}{n^2(n-1)}\sum_{i=1}^n\sum_{\substack{j=1\\ j\neq i}}^n\sum_{k=1}^n \frac{1}{h_n}K\Big(\frac{\varepsilon_i-\varepsilon_j}{h_n}\Big)\int \Delta(x|\varepsilon_i)(I\{X_k\leq x\}-F_X(x))w(x)\,dx.$$

Those statistics are U–statistics and one can deduce that the dominating terms (with remainder of order $o_p(n^{-1/2})$) are the U–statistics of order 1 in a Hoeffding decomposition. For instance, for $W_n^{(1)}$ we have

$$\begin{aligned} W_n^{(1)} \;&=\; \frac{1}{\binom{n}{2}}\sum_{i=1}^n\sum_{\substack{j=1\\ j<i}}^n \frac{1}{h_n}K\Big(\frac{\varepsilon_i-\varepsilon_j}{h_n}\Big)\int \Delta(x|\varepsilon_i)\Delta(x|\varepsilon_j)w(x)\,dx \\ &=\; E[W_n^{(1)}] + \frac{2}{n}\sum_{i=1}^n\Big(E\Big[\frac{1}{h_n}K\Big(\frac{\varepsilon_i-\varepsilon_j}{h_n}\Big)\int \Delta(x|\varepsilon_i)\Delta(x|\varepsilon_j)w(x)\,dx \;\Big|\; \varepsilon_i\Big] - E[W_n^{(1)}]\Big) \\ &\quad + o_p(\frac{1}{\sqrt{n}}), \end{aligned}$$

where the conditional expectation is evaluated for $i \neq j$, and with calculations as in the proof of Lemma A.1 one obtains

$$
\begin{aligned}
W_n^{(1)} &= E[W_n^{(1)}] + \frac{2}{n}\sum_{i=1}^n \Big(\int \Delta^2(x|\varepsilon_i) f_\varepsilon(\varepsilon_i) w(x)\,dx - E\Big[\int \Delta^2(x|\varepsilon_1) f_\varepsilon(\varepsilon_1) w(x)\,dx\Big]\Big) \\
&\quad + o_p(\frac{1}{\sqrt{n}}),
\end{aligned}
$$

and similarly

$$
\begin{aligned}
W_n^{(2)} &= \frac{1}{n}\sum_{i=1}^n \int I_i(x)\Delta(x|\varepsilon_i) f_\varepsilon(\varepsilon_i) w(x)\,dx + o_p(\frac{1}{\sqrt{n}}) \\
W_n^{(3)} &= \frac{1}{n}\sum_{i=1}^n \int (I\{X_i \le x\} - F_X(x)) E[\Delta(x|\varepsilon_1) f_\varepsilon(\varepsilon_1)] w(x)\,dx + o_p(\frac{1}{\sqrt{n}}).
\end{aligned}
$$

The dominating terms in the decomposition (A.1) can further be combined nicely because $I\{X_i \le x\} - F_X(x) = I_i(x) + \Delta(x|\varepsilon_i)$. By Lemma A.1 we have

$$
\sqrt{n}(\tilde{T}_n + \tilde{T}_n^{(1)} - B_{h_n}) = \frac{2}{\sqrt{n}}\sum_{i=1}^n (v_{ni} - E[v_{ni}] + w_{ni}) + o_p(1),
$$

where

$$
v_{ni} = \int (I\{X_i \le x\} - F_X(x)) v(x|\varepsilon_i) w(x)\,dx
$$

with v from (2.6), and w_{ni} is defined in Lemma A.1. Asymptotic normality follows from the central limit theorem and the asymptotic variance is derived straightforwardly from the formula

$$
\mathrm{Var}\Big(\frac{2}{\sqrt{n}}\sum_{i=1}^n v_{ni}\Big) = 4\Big(\mathrm{Var}(E[v_{ni} \mid \varepsilon_i]) + E[\mathrm{Var}(v_{ni} \mid \varepsilon_i)] + E[w_{ni}^2] + 2\mathrm{Cov}(v_{ni}, w_{ni})\Big).
$$

Hence, the assertion of Theorem 3.1 follows. $\square$

A.2 Proof of Theorem 3.4

From the bootstrap sample $(X_1, Y_1^*), \ldots, (X_n, Y_n^*)$ defined in (3.2) one calculates the Nadaraya-Watson estimator $\hat{m}^*$ as well as bootstrap residuals $\hat{\varepsilon}_i = Y_i^* - \hat{m}^*(X_i)$. With this notation the bootstrap test statistic is

$$
T_n^* = \frac{1}{n(n-1)}\sum_{i=1}^n \sum_{\substack{j=1\\ j\neq i}}^n \frac{1}{h_n} K\Big(\frac{\hat{\varepsilon}_i^* - \hat{\varepsilon}_j^*}{h_n}\Big) H_{ij},
$$

where, conditionally on $\mathcal{Y}_n$ the

$$H_{ij} \;=\; \int (I\{X_i \le x\} - F_{X,n}(x))(I\{X_j \le x\} - F_{X,n}(x))w(x)\,dx$$

are known (not random). Let P^* be the conditional distribution $P(\cdot \mid \mathcal{Y}_n)$ and denote the corresponding conditional expectation and variance by $E^*[\cdot]$, $\mathrm{Var}^*(\cdot)$.

Then let $\tilde{T}_n^*$ be defined as T_n^*, but replacing the bootstrap residuals $\hat{\varepsilon}_i^*$ by bootstrap errors ε_i^* $(i = 1, \dots, n)$. With calculations similar to the proof of Theorem 3.1 one can show that

$$T_n^* - \tilde{T}_n^* \;=\; o_p((n\sqrt{h_n})^{-1}). \tag{A.2}$$

From this follows that in terms of conditional weak convergence in probability, T_n^* and $\tilde{T}_n^*$ are equivalent. To be more specific, note that for all $\eta > 0$ it follows that $P^*(n\sqrt{h_n}|T_n^* - \tilde{T}_n^*| > \eta)$ converges to zero in probability, which follows from Markov's inequality and (A.2).

Further we have $E^*[\tilde{T}_n^*] = O_p(n^{-1}) = o_p((n\sqrt{h_n})^{-1})$ and it remains to show the assertion for $n\sqrt{h_n}(\tilde{T}_n^* - E^*[\tilde{T}_n^*])$. Though with respect to P^* the statistic $\tilde{T}_n^*$ is not a U-statistic, we mimic Hoeffding's decomposition to obtain

$$n\sqrt{h_n}(\tilde{T}_n^* - E^*[\tilde{T}_n^*]) \;=\; \tilde{T}_n^{*,1} + \tilde{T}_n^{*,2},$$

where

$$\begin{aligned}
\tilde{T}_n^{*,1} \;=\;& \frac{\sqrt{h_n}}{(n-1)} \sum_{i=1}^n \sum_{\substack{j=1\\ j\neq i}}^n H_{ij} \Big\{ \frac{1}{h_n} K\Big(\frac{\varepsilon_i^* - \varepsilon_j^*}{h_n}\Big) - E^*\Big[\frac{1}{h_n}K\Big(\frac{\varepsilon_i^* - \varepsilon_j^*}{h_n}\Big) \;\Big|\; \varepsilon_i^*\Big] \\
&- E^*\Big[\frac{1}{h_n}K\Big(\frac{\varepsilon_i^* - \varepsilon_j^*}{h_n}\Big) \;\Big|\; \varepsilon_j^*\Big] + E^*\Big[\frac{1}{h_n}K\Big(\frac{\varepsilon_i^* - \varepsilon_j^*}{h_n}\Big)\Big]\Big\} \\
\tilde{T}_n^{*,2} \;=\;& \frac{2\sqrt{h_n}}{(n-1)} \sum_{i=1}^n \sum_{\substack{j=1\\ j\neq i}}^n H_{ij} \Big\{ E^*\Big[\frac{1}{h_n}K\Big(\frac{\varepsilon_i^* - \varepsilon_j^*}{h_n}\Big) \;\Big|\; \varepsilon_i^*\Big] - E^*\Big[\frac{1}{h_n}K\Big(\frac{\varepsilon_i^* - \varepsilon_j^*}{h_n}\Big)\Big]\Big\}.
\end{aligned}$$

By the definition of $F_{X,n}$ we have $\sum_{j=1}^n H_{ij} = 0$ for all $i = 1, \dots, n$, and hence

$$\begin{aligned}
\tilde{T}_n^{*,2} \;&=\; \frac{2\sqrt{h_n}}{(n-1)} \sum_{i=1}^n H_{ii} \Big\{ \int \frac{1}{h_n} K\Big(\frac{\varepsilon_i^* - y}{h_n}\Big)\, d\tilde{F}_{\varepsilon,n}(y) - \int\int \frac{1}{h_n} K\Big(\frac{z - y}{h_n}\Big)\, d\tilde{F}_{\varepsilon,n}(y)\, d\tilde{F}_{\varepsilon,n}(z)\Big\} \\
&=\; o_p(1).
\end{aligned}$$

What remains to show now is conditional weak convergence of $\tilde{T}_n^{*,1}$ to a centered normal distributed random variable T with variance τ^2, in probability. In the original proof by Zheng (1997) (without bootstrap) a central limit theorem for U-statistics with n-dependent kernel by Hall (1984, Th. 1) is applied. In our context we do not have the U-statistic structure, but similar methods can be applied. Theorem 1 by Hall (1984) is proven with an application of Brown's (1971) central limit theorem for Martingales [see Hall and Heyde (1980, p. 58)], which itself is proven by establishing convergence of characteristic functions. In our context we need to show that the characteristic function of $\tilde{T}_n^{*,1}$ with respect to the conditional probability measure P^*, i.e. $E^*[\exp(it\tilde{T}_n^{*,1})]$, converges (for every $t \in \mathbb{R}$) in probability to the characteristic function $E[\exp(itT)]$, compare Shao and Tu (1995, p. 78).

To this end note that $\tilde{T}_n^{*,1} = S_{n,n}^*$, where $S_{k,n}^* = \sum_{i=2}^k Z_{ni}^*$, $2 \le k \le n$, is a Martingale with respect to P^* with filtration $\mathcal{F}_{ni} = \sigma(\varepsilon_1^*, \dots, \varepsilon_i^*)$. Here we use the definition

$$Z_{ni}^* = \sum_{j=1}^{i-1} H_{ij} H_n(\varepsilon_i^*, \varepsilon_j^*) \frac{2\sqrt{h_n}}{n-1},$$

where

$$\begin{aligned} H_n(\varepsilon_i^*, \varepsilon_j^*) &= \frac{1}{h_n} K\Big(\frac{\varepsilon_i^* - \varepsilon_j^*}{h_n}\Big) - E^*\Big[\frac{1}{h_n} K\Big(\frac{\varepsilon_i^* - \varepsilon_j^*}{h_n}\Big) \;\Big|\; \varepsilon_i^*\Big] \\ &\quad - E^*\Big[\frac{1}{h_n} K\Big(\frac{\varepsilon_i^* - \varepsilon_j^*}{h_n}\Big) \;\Big|\; \varepsilon_j^*\Big] + E^*\Big[\frac{1}{h_n} K\Big(\frac{\varepsilon_i^* - \varepsilon_j^*}{h_n}\Big)\Big]. \end{aligned}$$

Now the rest of the proof follows along the lines of the proofs of Theorem 1 by Hall (1984) and Corollary 3.1 by Hall and Heyde (1980, p. 58). Here for the convergence of the sequence of characteristic functions we only need to establish the conditions

$$\sum_{i=2}^n E^*[(Z_{ni}^*)^2] \xrightarrow{P} \tau^2, \quad \sum_{i=2}^n E^*[(Z_{ni}^*)^4] \xrightarrow{P} 0$$
$$\sum_{i=2}^n E^*[(Z_{ni}^*)^2 \mid \varepsilon_1^*, \dots, \varepsilon_{i-1}^*] \xrightarrow{P} \tau^2.$$

As an example we prove validity of the first condition, the others follow similarly. For $i \neq j \neq k \neq i$ we have $E^*[H_n(\varepsilon_i^*, \varepsilon_j^*) H_n(\varepsilon_i^*, \varepsilon_k^*)] = 0$, and hence

$$\sum_{i=2}^n E^*[(Z_{ni}^*)^2] = \frac{4h_n}{(n-1)^2} \sum_{i=2}^n \sum_{j=1}^{i-1} H_{ij}^2 E^*[(H_n(\varepsilon_i^*, \varepsilon_j^*)^2].$$

Inserting the definition of H_n the dominating term is clearly

$$\begin{aligned} &\frac{4h_n}{(n-1)^2} \sum_{i=2}^n \sum_{j=1}^{i-1} H_{ij}^2 E^*\Big[\frac{1}{h_n^2} K^2\Big(\frac{\varepsilon_i^* - \varepsilon_j^*}{h_n}\Big)\Big] \\ &= \frac{2}{(n-1)^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ i\neq j}}^n H_{ij}^2 \frac{1}{n^2} \sum_{l=1}^n \sum_{k=1}^n \frac{1}{h_n} K^2\Big(\frac{\tilde{\varepsilon}_l - \tilde{\varepsilon}_k}{h_n}\Big). \end{aligned}$$

It is not difficult to show that this term converges in probability to the limit (for $h_n \to 0$) of

$$E\Big[\Big(\int (I\{X_1 \le x\} - F_X(x))(I\{X_2 \le x\} - F_X(x)) w(x)\, dx\Big)^2\Big] \frac{2}{h_n} \int K^2\Big(\frac{y-z}{h_n}\Big)\, dF_\varepsilon(y)\, dF_\varepsilon(z)$$

which equals τ^2. $\square$

Acknowledgements. The author would like to thank Peter Hall and Nils Hjort for helpful suggestions during an inspiring workshop at Oberwolfach.

B References

I. A. Ahmad and Q. Li (1997). *Testing symmetry of an unknown density function by kernel method.* J. Nonpar. Statist. 7, 279–293.

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

J. T. Alcalá, J. A. Cristóbal and W. González Manteiga (1999). *Goodness-of-fit test for linear models based on local polynomials.* Statist. Probab. Lett. 42, 39–46.

J. R. Blum, J. Kiefer and M. Rosenblatt (1961). *Distribution free tests of independence based on the sample distribution function.* Ann. Math. Statist. 32, 485–498.

A. W. Bowman and A. Azzalini (1997). *Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations.* Oxford University Press, Oxford.

B. M. Brown (1971). *Martingale central limit theorems.* Ann. Math. Statist. 42, 59–66.

F. Cheng (2005). *Asymptotic distributions of error density and distribution function estimators in nonparametric regression.* J. Statist. Plan. and Inf. 128, 327–349.

P. Deheuvels (1980). *Nonparametric test of independence.* Nonparametric asymptotic statistics (Proc. Conf., Rouen, 1979) (French), pp. 95–107, Lecture Notes in Math., 821, Springer, Berlin.

H. Dette (1999). *A consistent test for the functional form of a regression based on a difference of variance estimators.* Ann. Statist. 27, 1012–1040.

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

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

H. Dette and N. Neumeyer (2000). *A note on a specification test of independence.* Metrika 51, 133–144.

J. Einmahl and I. Van Keilegom (2007a). *Tests for independence in nonparametric regression.* Statist. Sinica (to appear).
http://www.stat.ucl.ac.be/ISpersonnel/vankeile/pub.html

J. Einmahl and I. Van Keilegom (2007b). *Specification tests in nonparametric regression.* J. Econometrics (to appear).
http://www.stat.ucl.ac.be/ISpersonnel/vankeile/pub.html

R. L. Eubank and W. Thomas (1993). *Detecting heteroscedasticity in nonparametric regression.* J. Roy. Statist. Soc. Ser. B 55, 145–155.

J. Fan and I. Gijbels (1996). *Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation.* J. Roy. Statist. Soc. Ser. B 57, 371–394.

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

P. Hall (1984). *Central limit theorem of integrated square error of multivariate nonparametric density etimators.* J. Multiv. Anal. 14, 1–16.

P. Hall and J. W. Hart (1990). *Bootstrap test for difference between means in nonparametric regression.* J. Amer. Statist. Assoc. 85, 1039–1049.

P. Hall and C. C. Heyde (1980). *Martingale limit theory and its application.* Probability and Mathematical Statistics. Academic Press, Inc., New York-London.

W. Härdle (1989). *Applied nonparametric regression.* Cambridge University Press, New York.

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.

J. Hart (1997). *Nonparametric smoothing and lack-of-fit tests.* Springer.

W. Hoeffding (1948). *A nonparametric test of independence.* Ann. Math. Statist. 19 546–557.

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

T. Lindström, U. Holst and P. Weibring (2005). *Analysis of Lidar Fields using local polynomial regression.* Environmetrics 16, 619–634.

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

H. L. Koul and A. Schick (1997). *Testing for the equality of two nonparametric regression curves.* J. Statist. Plann. Inf. 65, 293–314.

H.–G. Müller (1984). *Boundary Effects in Nonparametric Curve Estimation.* Compstat 1984, Physica Verlag, Heidelberg, 84–89.

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

U. U. Müller, A. Schick and W. Wefelmeyer (2006). *Estimating the error distribution function in semiparametric regression.* To appear in Statist. Decisions. 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 (2006). Smooth residual bootstrap for empirical processes of nonparametric regression residuals. preprint. http://www.math.uni-hamburg.de/research/ims.html

N. Neumeyer, H. Dette and E.–R. Nagel (2005). *Bootstrap tests for the error distribution in linear and nonparametric regression models.* Australian and New Zealand Journal of Statistics 48, 129–156.

J. C. Pardo–Fernández, I. Van Keilegom and W. González–Manteiga (2007). *Comparison of regression curves based on the estimation of the error distribution.* Statistica Sinica 17, 1115–1137.

R Development Core Team (2006). *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.

M. Rosenblatt (1975). *A quadratic measure of deviation of two-dimensional density estimates and a test of independence.* Ann. Statist. 3, 1–14.

J. Shao and D. Tu (1995). *The Jackknife and bootstrap.* Springer.

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

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

W. Stute, S. Thies and L.–X. Zhu (1998). *Model checks for regression: an innovation process approach.* Ann. Statist. 26, 1916–1934.

I. Van Keilegom, W. González–Manteiga and C. Sánchez Sellero (2007). *Goodness-of-fit tests in parametric regression based on the estimation of the error distribution.* TEST, to appear. http://www.stat.ucl.ac.be/IAP/publication_tr.html

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

J. X. Zheng (1996). *A consistent test of functional form via nonparametric estimation techniques.* J. Econometrics 75, 263–289.

J. X. Zheng (1997). *A consistent specification test of independence.* J. Nonparametr. Statist. 7, 297–306.

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