Testing for symmetric error distribution in nonparametric regression models

Natalie Neumeyer and Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
natalie.neumeyer@ruhr-uni-bochum.de
holger.dette@ruhr-uni-bochum.de

June 4, 2003

Abstract

For the problem of testing symmetry of the error distribution in a nonparametric regression model we propose as a test statistic the difference between the two empirical distribution functions of estimated residuals and their counterparts with opposite signs. The weak convergence of the difference process to a Gaussian process is shown. The covariance structure of this process depends heavily on the density of the error distribution, and for this reason the performance of a symmetric wild bootstrap procedure is discussed in asymptotic theory and by means of a simulation study. In contrast to the available procedures the new test is also applicable under heteroscedasticity.

AMS Classification: 62G10, 60F17
Keywords and Phrases: empirical process of residuals, testing for symmetry, nonparametric regression

1 Introduction

Consider the nonparametric heteroscedastic regression model

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

with unknown regression and variance functions $m(\cdot)$ and $\sigma^2(\cdot)$, respectively, where $X_1, \dots, X_n$ are independent identically distributed. The unknown errors $\varepsilon_1, \dots, \varepsilon_n$ are assumed to be independent of the design points, centered and independent identically distributed with absolutely continuous distribution function F_ε and density f_ε. Hence $-\varepsilon_i$ has density $f_{-\varepsilon}(t) = f_\varepsilon(-t)$ and cumulative distribution function $F_{-\varepsilon}(t) = 1 - F_\varepsilon(-t)$. In this paper we are interested in testing the symmetry of the error distribution, that is:

$$H_0 \ : \ F_\varepsilon(t) = 1 - F_\varepsilon(-t) \text{ for all } t \in I\!R \quad \text{versus} \quad H_1 \ : \ F_\varepsilon(t) \neq 1 - F_\varepsilon(-t) \text{ for some } t \in I\!R$$

or equivalently

$$H_0 \; : \; f_\varepsilon(t) = f_\varepsilon(-t) \text{ for all } t \in I\!\!R \quad \text{versus} \quad H_1 \; : \; f_\varepsilon(t) \neq f_\varepsilon(-t) \text{ for some } t \in I\!\!R.$$

The problem of testing symmetry of the unknown distribution of the residuals in regression models has been considered by numerous authors in the literature for various special cases of the nonparametric regression model (1). Most of the literature concentrates on the problem of testing the symmetry of the distribution of an i.i.d. sample about an unknown mean [see for example Bhattacheraya, Gastwirth and Wright (1982), Aki (1981), Koziol (1985), Schuster and Berker (1987) or Psaradakis (2003) among many others]. Ahmad and Li (1997) transferred an approach of Rosenblatt (1975) for testing independence to the problem of testing symmetry in a linear model with homoscedastic errors. Ahmad and Li's test was generalized to the nonparametric regression model (1) with homoscedastic errors in the fixed design case by Dette, Kusi–Appiah and Neumeyer (2002). These approaches are based on estimates for the L^2–distance $\int (f_\varepsilon(t) - f_\varepsilon(-t))^2 \, dt$ of the densities f_ε and $f_{-\varepsilon}$. A similar test was proposed recently by Hyndman and Yao (2002) in the context of testing the symmetry of the conditional density of a stationary process.
In this paper we propose an alternative approach for testing symmetry in nonparametric regression models. Our interest in this problem stems from two facts. On the one hand we are looking for a test which is applicable to observations with a heteroscedastic error structure. On the other hand the available procedures for the nonparametric regression model with homoscedastic errors are only consistent against alternatives which converge to the null hypothesis of symmetry at a rate $(n\sqrt{h})^{-1}$, where h denotes a smoothing parameter of a kernel estimator. It is the second purpose of this paper to construct a test for the symmetry of the error distribution in model (1), which can detect local alternatives at a rate $n^{-1/2}$. To explain the basic idea consider for a moment the regression function $m \equiv 0$ and variance $\sigma \equiv 1$ in model (1) which leads to the well investigated problem of testing the symmetry of the common distribution of an i. i. d. sample $\varepsilon_1, \ldots, \varepsilon_n$ [see e.g. Hušková (1984), Hollander (1988) for reviews]. One possible approach is to compare the empirical distribution functions of ε_i and $-\varepsilon_i$ [see, for example, Shorack and Wellner (1986, p. 743)] using the empirical process

$$S_n(t) = \frac{1}{n} \sum_{i=1}^{n} \Big(I\{\varepsilon_i \leq t\} - I\{-\varepsilon_i \leq t\} \Big) = F_{n,\varepsilon}(t) - F_{n,-\varepsilon}(t), \tag{2}$$

where $I\{\cdot\}$ denotes the indicator function, $F_{n,\varepsilon}$ is the empirical distribution function of $\varepsilon_1, \ldots, \varepsilon_n$ and $F_{n,-\varepsilon}$ is the empirical distribution function of $-\varepsilon_1, \ldots, -\varepsilon_n$, that is

$$F_{n,-\varepsilon}(t) \; = \; 1 - F_{n,\varepsilon}(-t-).$$

Throughout this paper we will call the process $S_n(t)$ (and any process of the same form) empirical symmetry process. Under the hypothesis of symmetry $F_\varepsilon = F_{-\varepsilon}$ the process $\sqrt{n}S_n$ converges weakly to the process $S = B(F_\varepsilon) + B(1 - F_\varepsilon)$, where B denotes a Brownian bridge. The covariance of this limit process is given by

$$\begin{aligned} \operatorname{Cov}(S(s), S(t)) \; &= \; F_\varepsilon(s \wedge t) - F_\varepsilon(s)F_\varepsilon(t) + F_\varepsilon((-s) \wedge t) - F_\varepsilon(-s)F_\varepsilon(t) \\ &\quad + F_\varepsilon(s \wedge (-t)) - F_\varepsilon(s)F_\varepsilon(-t) + F_\varepsilon((-s) \wedge (-t)) - F_\varepsilon(-s)F_\varepsilon(-t) \\ &= \; 2F_\varepsilon(-(|s| \vee |t|)), \end{aligned} \tag{3}$$

and a suitable asymptotic distribution free test statistic is then obtained by

$$T_n = n \int_0^\infty S_n^2(t)\, dH_n(t),$$

where $H_n = F_{n,\varepsilon} + F_{n,-\varepsilon} - 1$ denotes the empirical distribution function of the absolute values $|\varepsilon_1|, \ldots, |\varepsilon_n|$. The test statistic T_n converges in distribution to the random variable $\int_0^1 R^2(t)\, dt$, where $\{R(1-t)\}_{t\in[0,1]}$ is a Brownian motion. The null hypothesis of a symmetric error distribution is rejected for large values of this test statistic and the resulting test is consistent with respect to local alternatives converging to the null at a rate $n^{-1/2}$. A generalization of the empirical symmetry process (2) for the unknown residuals $\varepsilon_1, \ldots, \varepsilon_n$ in linear models with homoscedastic error structure [that is a regression function $m(X_i) = h^T(X_i)\beta$ in model (1) with a known function h, finite dimensional parameter β, constant variance function $\sigma(X_i) \equiv \sigma$ and a fixed design] can be found in Koul (2002, p. 258).

In the present paper we propose to generalize this approach to the problem of testing the hypothesis of a symmetric error distribution in a nonparametric regression model with heteroscedastic error structure. The empirical symmetry process defined in (2) is modified by replacing the unknown errors ε_i by estimated residuals $\widehat{\varepsilon}_i = (Y_i - \widehat{m}(X_i))/\widehat{\sigma}(X_i)$ $(i = 1, \ldots, n)$ where $\widehat{m}(\cdot)$ and $\widehat{\sigma}(\cdot)$ denote kernel based nonparametric estimators for the regression and variance function, respectively. This yields the process

$$\widehat{S}_n(t) = \frac{1}{n} \sum_{i=1}^n \Big(I\{\widehat{\varepsilon}_i \le t\} - I\{-\widehat{\varepsilon}_i \le t\} \Big),$$

and allows us also to consider heteroscedastic nonparametric regression models. In Section 3 we prove weak convergence of a centered version of this empirical process to a Gaussian process under the null hypothesis of a symmetric error distribution and any fixed alternative. The covariance structure of the limiting process depends in a complicated way on the unknown distribution of the error and as a consequence an asymptotically distribution free test statistic cannot be found. For this reason we propose a modification of the wild bootstrap approach to compute critical values. The consistency of this bootstrap procedure is discussed in asymptotic theory and by means of a simulation study in Section 4 and Section 5, respectively. The numerical results indicate that the new bootstrap test is applicable for sample sizes larger than 20 and is more powerful than the existing procedures, which were derived under the additional assumption of homoscedasticity.

2 Technical assumptions

In this section we state some basic technical assumptions which are required for the statement of the main results in Section 3 and 4. We assume that the distribution function of the explanatory variables X_i, say F_X, has support $[0, 1]$ and is twice continuously differentiable with density f_X bounded away from zero. We also assume that the error distribution has a finite fourth moment, that is

$$E[\varepsilon^4] = \int t^4 f_\varepsilon(t)\, dt < \infty.$$

Further suppose that the conditional distribution $P^{Y_i|X_i=x}$ of Y_i given $X_i = x$ has distribution function

$$F(y|x) = F_\varepsilon\Big(\frac{y-m(x)}{\sigma(x)}\Big)$$

and density

$$f(y|x) = \frac{1}{\sigma(x)} f_\varepsilon\Big(\frac{y-m(x)}{\sigma(x)}\Big)$$

such that

$$\inf_{x\in[0,1]} \inf_{y\in[0,1]} f\left(F^{-1}(y|x)|x\right) > 0 \quad \text{and} \quad \sup_{x,y} |yf(y|x)| < \infty,$$

where $F(y|x)$ and $f(y|x)$ are continuous in (x, y), the partial derivative $\frac{\partial}{\partial y} f(y|x)$ exists and is continuous in (x, y) such that

$$\sup_{x,y} \left| y^2 \frac{\partial f(y|x)}{\partial y} \right| < \infty.$$

In addition, we also require that the derivatives $\frac{\partial}{\partial x} F(y|x)$ and $\frac{\partial^2}{\partial x^2} F(y|x)$ exist and are continuous in (x, y) such that

$$\sup_{x,y} \left| y \frac{\partial F(y|x)}{\partial x} \right| < \infty\,, \quad \sup_{x,y} \left| y^2 \frac{\partial^2 F(y|x)}{\partial x^2} \right| < \infty.$$

The regression and variance functions m and σ^2 are assumed to be twice continuously differentiable such that $\min_{x\in[0,1]} \sigma^2(x) \geq c > 0$ for some constant c.
Throughout this paper let K be a symmetric twice continuously differentiable density with compact support and vanishing first moment $\int uK(u)\,du = 0$ and $h = h_n$ denote a sequence of bandwidths converging to zero for an increasing sample size $n \to \infty$ such that $nh^4 = O(1)$ and $nh^{3+\delta}/\log(1/h) \to \infty$ for some $\delta > 0$.

3 Weak convergence of the empirical symmetry process

We explained in the introduction that the basic idea of the proposed procedure for testing symmetry is to replace the unknown random variables ε_i by estimated residuals $\widehat{\varepsilon}_i$ $(i = 1, \ldots, n)$ in the definition (2) of the empirical symmetry process. For the estimation of the residuals we define nonparametric kernel estimators for the unknown regression function $m(\cdot)$ and variance function $\sigma^2(\cdot)$ in model (1) by

$$\widehat{m}(x) = \frac{\sum_{i=1}^n K(\frac{X_i-x}{h})Y_i}{\sum_{j=1}^n K(\frac{X_j-x}{h})} \tag{4}$$

$$\widehat{\sigma}^2(x) = \frac{\sum_{i=1}^n K(\frac{X_i-x}{h})(Y_i-\widehat{m}(x))^2}{\sum_{j=1}^n K(\frac{X_j-x}{h})}. \tag{5}$$

Note, that $\widehat{m}(\cdot)$ is the usual Nadaraya–Watson estimator which is considered here for the sake of simplicity, but the following results are also correct for local polynomial estimators [see Fan and Gijbels (1996)], where the kernel K has to be replaced by its asymptotically equivalent kernel [see Wand and Jones (1995)]. Now the standardized residuals from the nonparametric fit are defined by

$$\hat{\varepsilon}_i = \frac{Y_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)} \quad (i = 1, \ldots, n). \tag{6}$$

The estimated empirical symmetry process is based on the residuals (6) and given by

$$\hat{S}_n(t) = \hat{F}_{n,\varepsilon}(t) - \hat{F}_{n,-\varepsilon}(t) = \frac{1}{n}\sum_{i=1}^{n}\Big(I\{\hat{\varepsilon}_i \le t\} - I\{-\hat{\varepsilon}_i \le t\}\Big). \tag{7}$$

Our first result states the asymptotic behaviour of this process.

Theorem 3.1 *Under the assumptions stated in Section 2 the process $\{R_n(t)\}_{t\in\mathbb{R}}$ defined by*

$$R_n(t) = \sqrt{n}\Big(\hat{S}_n(t) - F_\varepsilon(t) + (1 - F_\varepsilon(-t)) - h^2 B(t)\Big)$$

converges weakly to a centered Gaussian process $\{R(t)\}_{t\in\mathbb{R}}$ with covariance structure

$$\begin{aligned}
G(s,t) &= \operatorname{Cov}(R(s), R(t)) \\
&= F_\varepsilon(s \wedge t) - F_\varepsilon(s)F_\varepsilon(t) + F_\varepsilon((-s)\wedge t) - F_\varepsilon(-s)F_\varepsilon(t) \\
&\quad + F_\varepsilon(s \wedge (-t)) - F_\varepsilon(s)F_\varepsilon(-t) + F_\varepsilon((-s)\wedge(-t)) - F_\varepsilon(-s)F_\varepsilon(-t) \\
&\quad + (f_\varepsilon(t) + f_\varepsilon(-t))(f_\varepsilon(s) + f_\varepsilon(-s)) \\
&\quad + (f_\varepsilon(s) + f_\varepsilon(-s))\int_{-\infty}^{t} x(f_\varepsilon(x) + f_\varepsilon(-x))\,dx \\
&\quad + (f_\varepsilon(t) + f_\varepsilon(-t))\int_{-\infty}^{s} x(f_\varepsilon(x) + f_\varepsilon(-x))\,dx \\
&\quad + \frac{s}{2}(f_\varepsilon(s) - f_\varepsilon(-s))\int_{-\infty}^{t} (x^2-1)(f_\varepsilon(x) + f_\varepsilon(-x))\,dx \\
&\quad + \frac{t}{2}(f_\varepsilon(t) - f_\varepsilon(-t))\int_{-\infty}^{s} (x^2-1)(f_\varepsilon(x) + f_\varepsilon(-x))\,dx \\
&\quad + \frac{s}{2}(f_\varepsilon(s) - f_\varepsilon(-s))(f_\varepsilon(t) + f_\varepsilon(-t))E[\varepsilon_1^3] \\
&\quad + \frac{t}{2}(f_\varepsilon(t) - f_\varepsilon(-t))(f_\varepsilon(s) + f_\varepsilon(-s))E[\varepsilon_1^3] \\
&\quad + \frac{st}{4}(f_\varepsilon(s) - f_\varepsilon(-s))(f_\varepsilon(t) - f_\varepsilon(-t))\operatorname{Var}(\varepsilon_1^2),
\end{aligned}$$

where the bias $B(t)$ is defined by

$$\begin{aligned}
B(t) &= \frac{1}{2}\int K(u)u^2\,du\,\Big((f_\varepsilon(t) + f_\varepsilon(-t))\int \frac{1}{\sigma(x)}\left((mf_X)''(x) - (mf_X'')(x)\right)dx \\
&\quad + t(f_\varepsilon(t) - f_\varepsilon(-t))\int \frac{1}{2\sigma^2(x)}\left((\sigma^2 f_X)''(x) - (\sigma^2 f_X'')(x) + 2(m'(x))^2 f_X(x)\right)dx\Big).
\end{aligned}$$

Note that the first two lines in the definition of the asymptotic covariance can be rewritten as follows,

$$\begin{aligned}
& F_\varepsilon(s\wedge t) - F_\varepsilon(s)F_\varepsilon(t) + F_\varepsilon((-s)\wedge t) - F_\varepsilon(-s)F_\varepsilon(t) + F_\varepsilon(s\wedge(-t)) - F_\varepsilon(s)F_\varepsilon(-t)\\
& \quad + F_\varepsilon((-s)\wedge(-t)) - F_\varepsilon(-s)F_\varepsilon(-t)\\
= \; & F_\varepsilon(s\wedge t) + 1 - F_\varepsilon(-(s\wedge t)) - (F_\varepsilon(t) + 1 - F_\varepsilon(-t)) + F_\varepsilon((-s)\wedge t) + 1 - F_\varepsilon(-(-s)\wedge t)\\
& \quad + (F_\varepsilon(t) - 1 + F_\varepsilon(-t))(F_\varepsilon(s) - 1 + F_\varepsilon(-s)),
\end{aligned}$$

and under the hypothesis $H_0 : F_\varepsilon(t) = 1 - F_\varepsilon(-t)$ this expression reduces to

$$2F_\varepsilon(s\wedge t) - 2F_\varepsilon(t) + 2F_\varepsilon((-s)\wedge t) \; = \; 2F_\varepsilon(-(|s|\vee|t|)),$$

which coincides with the covariance (3) of the limit of the classical empirical symmetry process (2) based on an i.i.d. sample. Additionally, under H_0 we deduce for the bias in Theorem 3.1

$$B(t) \; = \; \int K(u)u^2\,du\, f_\varepsilon(t) \int \frac{1}{\sigma(x)}\left((mf_X)''(x) - (mf_X'')(x)\right)\,dx.$$

Corollary 3.2 *If the assumptions of Theorem 3.1 and the null hypothesis H_0 of a symmetric error distribution are satisfied, the process $\{\sqrt{n}(\widehat{S}_n(t) - h^2B(t))\}_{t\in\mathbb{R}}$ defined in (7) converges weakly to a centered Gaussian process $\{S(t)\}_{t\in\mathbb{R}}$ with covariance*

$$\begin{aligned}
H(s,t) \; &= \; \operatorname{Cov}(S(s),S(t))\\
&= \; 2F_\varepsilon(-(|s|\vee|t|)) + 4f_\varepsilon(s)f_\varepsilon(t) + 4f_\varepsilon(s)\int_{-\infty}^{t} xf_\varepsilon(x)\,dx + 4f_\varepsilon(t)\int_{-\infty}^{s} xf_\varepsilon(x)\,dx.
\end{aligned}$$

Comparing the covariance kernel H with the expression (3) we see that there appear three additional terms depending on the density of the error distribution. This complication is caused by the estimation of the variance and regression function in our procedure. We finally note that the bias $h^2B(t)$ in Theorem 3.1 and Corollary 3.2 can be omitted if $h^4n = o(1)$.

Proof of Theorem 3.1:

From Theorem 1 in Akritas and Keilegom (2001) we obtain the following expansion of the estimated empirical distribution function,

$$\begin{aligned}
\widehat{F}_{n,\varepsilon}(t) \; &= \; \frac{1}{n}\sum_{i=1}^{n} I\{\widehat{\varepsilon}_i \le t\}\\
&= \; \frac{1}{n}\sum_{i=1}^{n} I\{\varepsilon_i \le t\} + \frac{1}{n}\sum_{i=1}^{n}\varphi(X_i,Y_i,t) + \beta_n(t) + r_n(t)
\end{aligned}$$

where, uniformly in $t\in\mathbb{R}$,

$$r_n(t) \; = \; o_p(\frac{1}{\sqrt{n}}) + o_p(h^2) \; = \; o_p(\frac{1}{\sqrt{n}})$$

and

$$
\begin{aligned}
\varphi(x,z,t) &= -\frac{f_\varepsilon(t)}{\sigma(x)}\int (I\{z\le v\}-F(v|x))\Big(1+t\frac{v-m(x)}{\sigma(x)}\Big)\,dv \\
&= -\frac{f_\varepsilon(t)}{\sigma(x)}\Big(1-\frac{tm(x)}{\sigma(x)}\Big)\Big(\int_z^\infty (1-F(v|x))\,dv-\int_{-\infty}^z F(v|x)\,dv\Big) \\
&\quad -\frac{f_\varepsilon(t)}{\sigma(x)}\frac{t}{\sigma(x)}\Big(\int_z^\infty v(1-F(v|x))\,dv-\int_{-\infty}^z vF(v|x)\,dv\Big) \\
&= -\frac{f_\varepsilon(t)}{\sigma(x)}\Big(1-\frac{tm(x)}{\sigma(x)}\Big)(m(x)-z)-\frac{f_\varepsilon(t)t}{\sigma^2(x)}\left(\frac{1}{2}(\sigma^2(x)+m^2(x))-\frac{z^2}{2}\right) \\
&= -\frac{f_\varepsilon(t)}{\sigma^2(x)}\Big(\sigma(x)(m(x)-z)-tm^2(x)+tm(x)z+\frac{1}{2}\sigma^2(x)t+\frac{1}{2}m^2(x)t-\frac{1}{2}tz^2\Big).
\end{aligned}
$$

This gives for $z=m(x)+\sigma(x)\varepsilon$:

$$
\varphi(x,z,t) \;=\; \varphi(x,m(x)+\sigma(x)\varepsilon,t) \;=\; f_\varepsilon(t)\left(\varepsilon+\frac{t}{2}(\varepsilon^2-1)\right).
$$

From the proof of Theorem 1 in Akritas and Keilegom (2001), p. 555, we also have for the bias term

$$
\begin{aligned}
\beta_n(t) &= E\Big[f_\varepsilon(t)\int\frac{\widehat{m}(x)-m(x)}{\sigma(x)}\,dF_X(x)+tf_\varepsilon(t)\int\frac{\widehat{\sigma}(x)-\sigma(x)}{\sigma(x)}\,dF_X(x)\Big] \\
&= \frac{h^2}{2}\int K(u)u^2\,du\,\Big(f_\varepsilon(t)\int\frac{1}{\sigma(x)}\left((mf_X)''(x)-(mf_X'')(x)\right)dx \\
&\quad + tf_\varepsilon(t)\int\frac{1}{2\sigma^2(x)}\left((\sigma^2 f_X)''(x)-(\sigma^2 f_X'')(x)+2(m'(x))^2 f_X(x)\right)dx\Big)+o(h^2)+o(\frac{1}{\sqrt{n}}).
\end{aligned}
$$

An analogous expansion for the estimated empirical distribution function $\widehat{F}_{n,-\varepsilon}(t)$ of the signed residuals now yields

$$
\begin{aligned}
\widehat{S}_n(t) &= \widehat{F}_{n,\varepsilon}(t)-\widehat{F}_{n,-\varepsilon}(t) \\
&= \frac{1}{n}\sum_{i=1}^n\Big(I\{\widehat{\varepsilon}_i\le t\}-I\{-\widehat{\varepsilon}_i\le t\}\Big) \\
&= \frac{1}{n}\sum_{i=1}^n\Big(I\{\varepsilon_i\le t\}-I\{-\varepsilon_i\le t\}+\varepsilon_i(f_\varepsilon(t)+f_\varepsilon(-t))+(\varepsilon_i^2-1)\frac{t}{2}(f_\varepsilon(t)-f_\varepsilon(-t))\Big) \qquad (8)\\
&\quad + h^2B(t)+o_p(\frac{1}{\sqrt{n}})
\end{aligned}
$$

uniformly with respect to $t\in I\!R$, where $B(t)=(\beta_n(t)+\beta_n(-t))/h^2+o(1)$ is defined in Theorem 3.1. Note, that under the null hypothesis the quadratic term in ε_i in (8), which is due to the estimation of the variance function, vanishes. From the above expansion we obtain

$$
\begin{aligned}
R_n(t) &= \sqrt{n}\Big(\widehat{S}_n(t)-F_\varepsilon(t)+(1-F_\varepsilon(-t))-h^2B(t)\Big) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(I\{\varepsilon_i\le t\}-F_\varepsilon(t)-I\{-\varepsilon_i\le t\}+(1-F_\varepsilon(-t))
\end{aligned}
$$

$$
\begin{aligned}
&\quad + \varepsilon_i(f_\varepsilon(t) + f_\varepsilon(-t)) + (\varepsilon_i^2 - 1)\frac{t}{2}(f_\varepsilon(t) - f_\varepsilon(-t))\Big) + o_p(1) \\
&= \widetilde{R}_n(t) + o_p(1)
\end{aligned}
$$

uniformly with respect to $t \in \mathbb{R}$, where the last line defines the process $\widetilde{R}_n$. Now a straightforward calculation of the covariances gives:

$$
\begin{aligned}
\operatorname{Cov}(\widetilde{R}_n(s), \widetilde{R}_n(t)) &= E\Big[\Big(I\{\varepsilon_1 \le s\} - F_\varepsilon(s) - I\{-\varepsilon_1 \le s\} + F_{-\varepsilon}(s)) + \varepsilon_1(f_\varepsilon(s) + f_\varepsilon(-s)) \\
&\quad + (\varepsilon_1^2 - 1)\frac{s}{2}(f_\varepsilon(s) - f_\varepsilon(-s))\Big)\Big(I\{\varepsilon_1 \le t\} - F_\varepsilon(t) - I\{-\varepsilon_1 \le t\} \\
&\quad + F_{-\varepsilon}(t)) + \varepsilon_1(f_\varepsilon(t) + f_\varepsilon(-t)) + (\varepsilon_1^2 - 1)\frac{t}{2}(f_\varepsilon(t) - f_\varepsilon(-t))\Big)\Big] + o(1) \\
&= G(s,t),
\end{aligned}
$$

where $G(s,t)$ is defined in Theorem 3.1. To prove weak convergence of the process $\{R_n(t)\}_{t\in\mathbb{R}}$ we prove weak convergence of $\{\widetilde{R}_n(t)\}_{t\in\mathbb{R}}$ and write

$$
\widetilde{R}_n(t) = \sqrt{n}(P_n h_t - P h_t),
$$

where P_n denotes the empirical measure based on $\varepsilon_1, \dots, \varepsilon_n$, that is $P_n h_t = \frac{1}{n}\sum_{i=1}^n h_t(\varepsilon_i)$, Ph_t denotes the expectation $E[h_t(\varepsilon_i)]$ and

$$
\mathcal{H} = \{h_t \mid t \in \mathbb{R}\}
$$

is the class of functions of the form

$$
h_t(\varepsilon) = I\{\varepsilon \le t\} - I\{-\varepsilon \le t\} + \varepsilon(f_\varepsilon(t) + f_\varepsilon(-t)) + (\varepsilon^2 - 1)\frac{t}{2}(f_\varepsilon(t) - f_\varepsilon(-t)).
$$

To conclude the proof of weak convergence in $\ell^\infty(\mathcal{H})$ we show that the class $\mathcal{H}$ is Donsker. Applying Theorem 2.6.8 (and the remark in the corresponding proof) of van der Vaart and Wellner (1996, p. 142) we have to verify that $\mathcal{H}$ is pointwise separable, is a VC–class and has an envelope with finite second moment.

Using the assumptions made in Section 2 we have $\sup_{t\in\mathbb{R}} |f_\varepsilon(t)| < \infty$, $\sup_{t\in\mathbb{R}} |tf_\varepsilon(t)| < \infty$ and due to this the class $\mathcal{H}$ has an envelope of the form

$$
H(\varepsilon) = c_1 + \varepsilon c_2 + (\varepsilon^2 - 1)c_3,
$$

where c_1, c_2, c_3 are constants. This envelope has obviously a finite second moment.

The function class $\mathcal{G} = \{h_t \mid t \in \mathbb{Q}\}$ is a countable subclass of $\mathcal{H}$. For each $\varepsilon \in \mathbb{R}$ the function $t \mapsto h_t(\varepsilon)$ is right continuous. Hence for a sequence $t_m \in \mathbb{Q}$ with $t_m \searrow t$ as $m \to \infty$ we have pointwise convergence $g_m(\varepsilon) = h_{t_m}(\varepsilon) \to h_t(\varepsilon)$ for $m \to \infty$. The convergence is also valid in the L^2–sense:

$$
\begin{aligned}
P((g_m - h_t)^2) &\le 6\Big(F_\varepsilon(t) - F_\varepsilon(t_m) + F_{-\varepsilon}(t) - F_{-\varepsilon}(t_m) + E[\varepsilon^2](f_\varepsilon(t) - f_\varepsilon(t_m))^2 \\
&\quad + E[\varepsilon^2](f_\varepsilon(-t) - f_\varepsilon(-t_m))^2 + E[(\varepsilon^2 - 1)^2]\frac{1}{4}(tf_\varepsilon(t) - t_m f_\varepsilon(t_m))^2 \\
&\quad + E[(\varepsilon^2 - 1)^2]\frac{1}{4}(tf_\varepsilon(-t) - t_m f_\varepsilon(-t_m))^2\Big) \\
&\longrightarrow 0 \quad \text{for} \quad m \to \infty.
\end{aligned}
$$

This proves pointwise seperability of $\mathcal{H}$ [see van der Vaart and Wellner (1996, p. 116)]. Sums of VC–classes of functions are VC–classes again [see van der Vaart and Wellner (1996, p. 147)]. The classes $\{\varepsilon \mapsto I\{\varepsilon \leq t\} \mid t \in I\!R\}$ and $\{\varepsilon \mapsto I\{-\varepsilon \leq t\} \mid t \in I\!R\}$ are obviously VC. Finally, the function class

$$\{\varepsilon \mapsto \varepsilon(f_\varepsilon(t) + f_\varepsilon(-t)) + (\varepsilon^2 - 1)\frac{t}{2}(f_\varepsilon(t) - f_\varepsilon(-t)) \mid t \in I\!R\}$$

is a subclass of the VC–class $\{\varepsilon \mapsto a\varepsilon + b\varepsilon^2 \mid a, b \in I\!R\}$. This yields the VC–property of $\mathcal{H}$ and concludes the proof of the weak convergence of the process $\{R_n(t)\}_{t \in I\!R}$. □

4 Symmetric wild bootstrap

Suitable test statistics for testing symmetry of the error distribution F_ε are, for example, Kolmogorov–Smirnov or Cramer–von–Mises type test statistics,

$$\sup_{t \in I\!R} |\widehat{S}_n(t)| \quad \text{and} \quad \int \widehat{S}_n^2(t)\, d\widehat{H}_n(t), \tag{9}$$

where $\widehat{H}_n$ is the empirical distribution function of $|\widehat{\varepsilon}_1|, \ldots, |\widehat{\varepsilon}_n|$ and the null hypothesis of symmetry is rejected for large values of these statistics. The asymptotic distribution of the test statistics can be obtained from Theorem 3.1, an application of the Continuous Mapping Theorem and (in the latter case) the uniform convergence of $\widehat{H}_n$,

$$\sup_{t \in I\!R} |\widehat{H}_n(t) - H(t)| \;=\; o_p(1),$$

where H denotes the distribution function of $|\varepsilon_1|$. A standard argument on contiguity [see e. g. Witting, Müller–Funk (1995), Theorem 6.113, 6.124 and 6.138 or van der Vaart (1998), Section 6] now shows that the resulting tests are consistent with respect to local alternatives converging to the null at a rate $n^{-1/2}$. However, because of the complicated dependence of the asymptotic null distribution of the process $\widehat{S}_n(t)$ on the unknown distribution function these test statistics are not asymptotically distribution free. Thus the critical values cannot be computed without estimating the unknown features of the error distribution of the data generating process. To avoid the problem of estimating the distribution and density function F_ε, f_ε we propose a modification of the wild bootstrap approach, which is adapted to the specific problem of testing symmetry.

For this let $v_1, \ldots, v_n$ be Rademacher variables, which are independent identically distributed such that $P(v_i = 1) = P(v_i = -1) = 1/2$, independent of the sample (X_j, Y_j), $j = 1, \ldots, n$. Note that wether the underlying error distribution F_ε is symmetric or not the distribution of the random variable $v_i\varepsilon_i$ is symmetric with density g_ε and distribution function G_ε defined by

$$g_\varepsilon(t) \;=\; \frac{1}{2}(f_\varepsilon(t) + f_\varepsilon(-t)), \quad G_\varepsilon(t) \;=\; \frac{1}{2}(F_\varepsilon(t) + 1 - F_\varepsilon(-t)), \tag{10}$$

respectively. Define bootstrap residuals as follows,

$$\varepsilon_i^* \;=\; v_i(Y_i - \widehat{m}(X_i)) \;=\; v_i\widehat{\sigma}(X_i)\widehat{\varepsilon}_i \quad (i = 1, \ldots, n)$$

where $\widehat{\varepsilon}_i$ is given in (6). Now we build new bootstrap observations ($i = 1, \ldots, n$)

$$\begin{aligned} Y_i^* &= \widehat{m}(X_i) + \varepsilon_i^* \\ &= v_i\sigma(X_i)\varepsilon_i + \widehat{m}(X_i) + v_i(m(X_i) - \widehat{m}(X_i)) \end{aligned}$$

and estimated residuals from the bootstrap sample,

$$\widehat{\varepsilon}_i^* = \frac{Y_i^* - \widehat{m}^*(X_i)}{\widehat{\sigma}^*(X_i)}, \tag{11}$$

where the regression and variance estimates $\widehat{m}^*$ and $\widehat{\sigma}^{*2}$ are defined analogous to $\widehat{m}$ and $\widehat{\sigma}^2$ in (4) and (5) but are based on the bootstrap sample (X_i, Y_i^*), $i = 1, \ldots, n$. In generalization of definition (7) the bootstrap version of the empirical symmetry process is now defined as

$$\widehat{S}_n^*(t) = \widehat{F}_{n,\varepsilon}^*(t) - \widehat{F}_{n,-\varepsilon}^*(t) = \frac{1}{n}\sum_{i=1}^n \Big(I\{\widehat{\varepsilon}_i^* \le t\} - I\{-\widehat{\varepsilon}_i^* \le t\}\Big).$$

The asymptotic behaviour of the bootstrap process conditioned on the initial sample is stated in the following theorem. Note that the result is valid under the hypothesis of symmetry $f_\varepsilon = f_{-\varepsilon}$ and under the alternative of a non-symmetric error distribution.

Theorem 4.1 *Under the assumptions of Theorem 3.1 the bootstrap process*

$$\{\sqrt{n}(\widehat{S}_n^*(t) - h^2B(t))\}_{t\in \mathbb{R}},$$

conditioned on the sample $\mathcal{Y}_n = \{(X_i, Y_i) \mid i = 1, \ldots, n\}$, converges weakly to a centered Gaussian process $\{S(t)\}_{t\in\mathbb{R}}$ with covariance

$$\mathrm{Cov}(S(s), S(t)) = 2G_\varepsilon(-(|s| \vee |t|)) + 4g_\varepsilon(s)g_\varepsilon(t) + 4g_\varepsilon(s)\int_{-\infty}^t xg_\varepsilon(x)\,dx + 4g_\varepsilon(t)\int_{-\infty}^s xg_\varepsilon(x)\,dx$$

in probability, where the bias term is defined by

$$B(t) = \int K(u)u^2\,du\, g_\varepsilon(t)\int \frac{1}{\sigma(x)}\left((mf_X)''(x) - (mf_X'')(x)\right)\,dx.$$

Here g_ε and G_ε are given by (10) and under the null hypothesis of symmetry we have $g_\varepsilon = f_\varepsilon$, $G_\varepsilon = F_\varepsilon$ and $\mathrm{Cov}(S(s), S(t)) = H(s,t)$, where the kernel $H(s,t)$ is defined in Corollary 3.2.

The proof of Theorem 4.1 is deferred to the Appendix.
From the theorem the consistency of a test for symmetry based on the wild bootstrap procedure can be deduced as follows. Let T_n denote the test statistic based on a continuous functional of the process $\widehat{S}_n$ and let T_n^* denote the corresponding bootstrap statistic based on $\widehat{S}_n^*$. If t_n is the realization of the test statistic T_n based on the sample $\mathcal{Y}_n$ then a level α–test is obtained by rejecting symmetry whenever $t_n > c_{1-\alpha}$, where $P_{H_0}(T_n > c_{1-\alpha}) = \alpha$. The quantile $c_{1-\alpha}$ can now be approximated by the bootstrap quantile $c_{1-\alpha}^*$ defined by

$$P(T_n^* > c_{1-\alpha}^* \mid \mathcal{Y}_n) = \alpha. \tag{12}$$

From Theorem 4.1 and the Continuous Mapping Theorem we obtain a consistent asymptotic level α–test by rejecting the null hypothesis if $t_n > c_{1-\alpha}^*$. We will illustrate this approach in a finite sample study in Section 5.

5 Finite sample properties

In this section we investigate the finite sample properties of the bootstrap procedure proposed in Section 4 by means of a simulation study. Exemplarily we consider the statistic

$$T_n = \int \widehat{S}_n^2(t) d\widehat{H}_n(t), \tag{13}$$

where

$$\widehat{H}_n(t) = \frac{1}{n} \sum_{i=1}^n I\{|\widehat{\varepsilon}_i| \leq t\}$$

denotes the empirical distribution function of the absolute residuals $|\widehat{\varepsilon}_1|, \ldots, |\widehat{\varepsilon}_n|$. If

$$T_n^* = \int (\widehat{S}_n^*)^2(t)\, d\widehat{H}_n^*(t)$$

is the bootstrap version of T_n, where $\widehat{H}_n^* = \widehat{F}_{n,\varepsilon}^* + \widehat{F}_{n,-\varepsilon}^* - 1$ denotes the empirical distribution function of $|\widehat{\varepsilon}_1^*|, \ldots, |\widehat{\varepsilon}_n^*|$, the consistency of the bootstrap procedure follows from Theorem 4.1, the Continuous Mapping Theorem and the fact that for all $\delta > 0$ we have

$$P\Big(\sup_{t \in I\!R} |\widehat{H}_n^*(t) - H(t)| > \delta \,\Big|\, \mathcal{Y}_n \Big) \;=\; o_p(1).$$

For the bandwidth in the regression and variance estimator defined by (4) and (5), respectively, we used

$$h = \Big(\frac{\widehat{\sigma}^2}{n} \Big)^{3/10}, \tag{14}$$

where

$$\widehat{\sigma}^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (Y_{[i+1]} - Y_{[i]})^2 \tag{15}$$

is an estimator of the integrated variance function $\int_0^1 \sigma^2(t) f_X(t) dt$ and $Y_{[1]}, \ldots, Y_{[n]}$ denotes the ordered sample of $Y_1, \ldots, Y_n$ according to the X values [see Rice (1984)]. The same bandwidth was used in the bootstrap step for the calculation of $\varepsilon_1^*, \ldots, \varepsilon_n^*$ and the corresponding estimators $\widehat{m}^*, \widehat{\sigma}^*$.

$B = 200$ bootstrap replications based on one sample $\mathcal{Y}_n = \{(X_i, Y_i) \mid i = 1, \ldots, n\}$ were performed for each simulation, where 1000 runs were used to calculate the rejection probabilities. The quantile estimate $c_{1-\alpha}^*$ defined in (12) from the bootstrap sample $T_n^{*,1}, \ldots, T_n^{*,B}$ was defined by

$$\widehat{c}_{1-\alpha}^* \;=\; T_n^{*,(\lfloor B(1-\alpha) \rfloor)},$$

where $T_n^{*,(i)}$ denotes the ith order statistic of $T_n^{*,1}, \ldots, T_n^{*,B}$. The null hypothesis H_0 of a symmetric error distribution was rejected if the original test statistic T_n based on the sample $\mathcal{Y}_n$ exceeded $\widehat{c}_{1-\alpha}^*$.

The model under consideration was

$$Y_i = \sin(2\pi X_i) + \sigma(X_i)\varepsilon_i\ , \qquad i = 1, \ldots, n, \tag{16}$$

for the sake of comparison with the results of Dette, Kusi-Appiah and Neumeyer (2002), who proposed a test for symmetry in a nonparametric homoscedastic regression model with a fixed design. Table 5.1 shows the approximation of the nominal level for the uniform design on the interval $[0, 1]$. The error distribution is a normal distribution, a convolution of two uniform distributions and a logistic distribution standardized such that $E[\varepsilon] = 0, E[\varepsilon^2] = 1$, while the variance function is constant i.e. $\sigma(x) \equiv 1$. We observe an accurate approximation of the nominal level for sample sizes $n \geq 20$.
The performance of the new test under alternatives is illustrated in Table 5.2, where a standardized chi-square distribution with $k = 1, 2, 3$ degrees of freedom is considered. The non-symmetry is detected in all cases with high probability, where the power increases with the sample size and decreases with increasing degrees of freedom. The cases $k = 1, 2$ should be compared with the simulation results in Dette, Kusi-Appiah and Neumeyer (2002), where the same situation for a fixed design has been considered. We observe notable improvements with respect to the probabilities of rejection in all considered cases. We note again that the procedure of these authors requires a homoscedastic error, while the bootstrap test proposed in Section 4 is also applicable under heteroscedasticity.
In order to investigate the impact of heteroscedasticity on the approximation of the level and the probability of rejection under the alternative we conducted a small simulation study for the case $m(x) = \sin(2\pi x), \sigma(x) = e^{-x}\sqrt{2}(1 - e^{-2})^{-1/2}$, a normal distribution and chi-squared distribution with $k = 1, 2, 3$ degrees of freedom standardized such that $E[\varepsilon] = 0, E[\varepsilon^2] = 1$. The explanatory variable is again uniformly distributed on the interval $[0, 1]$. Note that the variance function was normalized such that $\int_0^1 \sigma^2(x)dx = 1$ in order to make the results comparable with the scenario displayed in Table 5.1 and 5.2. The results are presented in Table 5.3. We observe no substantial differences with respect to the approximation of the nominal level (compare the first case in Table 5.1 and 5.3) and a slight loss with respect to power, which is caused by the heteroscedasticity (compare the cases df_1, df_2 and df_3 in Table 5.3 with Table 5.2). The results indicate that our procedure has a good performance under heteroscedasticity.

	α	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$
	0.025	0.029	0.033	0.029	0.029	0.027
	0.05	0.057	0.060	0.051	0.057	0.052
df_1	0.10	0.109	0.111	0.107	0.107	0.104
	0.20	0.214	0.216	0.215	0.193	0.209
	0.025	0.035	0.032	0.024	0.032	0.029
	0.05	0.062	0.055	0.051	0.068	0.057
df_2	0.10	0.113	0.111	0.101	0.113	0.108
	0.20	0.215	0.209	0.204	0.210	0.193
	0.025	0.031	0.030	0.028	0.028	0.030
	0.05	0.055	0.051	0.061	0.049	0.067
df_3	0.10	0.108	0.101	0.112	0.102	0.105
	0.20	0.199	0.204	0.202	0.197	0.192

Table 5.1: *Simulated level of the wild bootstrap test of symmetry in the nonparametric regression model (16) with $\sigma(x) \equiv 1$. The error distribution is a normal distribution (df_1), a*

logistic distribution (df_2) and a sum of two uniforms (df_3) standardized such that $E[\varepsilon] = 0$ and $E[\varepsilon^2] = 1$.

k	α	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$
	0.025	0.358	0.654	0.849	0.957	1.000
	0.05	0.484	0.764	0.912	0.981	1.000
1	0.10	0.584	0.847	0.959	0.991	1.000
	0.20	0.716	0.914	0.983	0.998	1.000
	0.025	0.239	0.458	0.698	0.817	0.998
	0.05	0.342	0.570	0.805	0.896	1.000
2	0.10	0.442	0.681	0.865	0.936	1.000
	0.20	0.594	0.794	0.934	0.976	1.000
	0.025	0.208	0.436	0.604	0.750	0.982
	0.05	0.303	0.565	0.710	0.833	0.995
3	0.10	0.414	0.667	0.812	0.895	0.998
	0.20	0.551	0.790	0.886	0.939	0.999

Table 5.2: *Simulated power of the wild bootstrap test of symmetry in the nonparametric regression model (16) with $\sigma(x) \equiv 1$. The error distribution is a chi-square distribution with k degrees of freedom standardized such that $E[\varepsilon] = 0$ and $E[\varepsilon^2] = 1$.*

	α	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$
	0.025	0.030	0.033	0.034	0.031	0.032
	0.05	0.056	0.061	0.063	0.062	0.050
df_0	0.10	0.094	0.113	0.107	0.101	0.106
	0.20	0.185	0.211	0.191	0.211	0.202
	0.025	0.308	0.610	0.838	0.941	1.000
	0.05	0.419	0.715	0.902	0.969	1.000
df_1	0.10	0.551	0.814	0.947	0.987	1.000
	0.20	0.693	0.898	0.975	0.993	1.000
	0.025	0.218	0.413	0.639	0.796	0.995
	0.05	0.314	0.541	0.737	0.870	0.997
df_2	0.10	0.425	0.674	0.835	0.925	0.999
	0.20	0.570	0.791	0.920	0.966	0.999
	0.025	0.197	0.377	0.559	0.710	0.985
	0.05	0.291	0.485	0.676	0.814	0.992
df_3	0.10	0.407	0.618	0.776	0.881	0.997
	0.20	0.539	0.766	0.884	0.941	1.000

Table 5.3: *Simulated level and power of the wild bootstrap test of symmetry in the nonparametric regression model (16) with $\sigma(x) = \sqrt{2}e^{-x}(1 - e^{-2})^{-1/2}$. The error distribution is a standard normal distribution (df_0) and chi-square distribution with k degrees of freedom ($df_k, k = 1, 2, 3$) standardized such that $E[\varepsilon] = 0, E[\varepsilon^2] = 1$.*

A Appendix: Proof of Theorem 4.1

We decompose the residuals ε_i^* defined in (11) in the following way,

$$\widehat{\varepsilon}_i^* \;=\; v_i \frac{\sigma(X_i)}{\widehat{\sigma}^*(X_i)}\varepsilon_i + v_i \frac{m(X_i) - \widehat{m}(X_i)}{\widehat{\sigma}^*(X_i)} + \frac{\widehat{m}(X_i) - \widehat{m}^*(X_i)}{\widehat{\sigma}^*(X_i)}.$$

Hence for $t \in I\!R$ the inequality $\widehat{\varepsilon}_i^* \leq t$ is equivalent to

$$v_i \varepsilon_i \;\leq\; t d_{n2}^*(X_i) + v_i d_{n1}(X_i) + d_{n1}^*(X_i)$$

and $v_i \widehat{\varepsilon}_i \leq t$ is equivalent to

$$v_i \varepsilon_i \;\leq\; t d_{n2}(X_i) + v_i d_{n1}(X_i),$$

where we introduced the definitions

$$\begin{aligned} d_{n1}(x) &= \frac{\widehat{m}(x) - m(x)}{\sigma(x)}, \quad d_{n2}(x) \;=\; \frac{\widehat{\sigma}(x)}{\sigma(x)}, \\ d_{n1}^*(x) &= \frac{\widehat{m}^*(x) - \widehat{m}(x)}{\sigma(x)}, \quad d_{n2}^*(x) \;=\; \frac{\widehat{\sigma}^*(x)}{\sigma(x)}. \end{aligned}$$

In the following we need four auxiliary results which are listed in Proposition 4.2–4.5 and can be proved by similar arguments as given in Abritas and Keilegom (2001). For the sake of brevity we will only sketch a proof of Proposition A.1 at the end of the general proof. The verification of Proposition A.2 follows from a Taylor expansion as in the proof of Theorem 1 of Akritas and Keilegom (2001) while the proof of Proposition A.3 follows exactly the lines of the proof of Lemma 1, Appendix B, in this reference. The proof of Proposition A.4 is done by some straightforward calculations of expectations and variances and is therefore omitted.

Proposition A.1 *Under the assumptions of Theorem 3.1 we have*

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} \Big(& I\{\widehat{\varepsilon}_i^* \leq t\} - P(v\varepsilon \leq t d_{n2}^*(X) + v d_{n1}(X) + d_{n1}^*(X) \mid \mathcal{Y}_n) \\ & - I\{v_i \widehat{\varepsilon}_i \leq t\} + P(v\varepsilon \leq t d_{n2}(X) + v d_{n1}(X) \mid \mathcal{Y}_n) \Big) \;=\; o_p(\frac{1}{\sqrt{n}}) \end{aligned}$$

uniformly in $t \in I\!R$.

Proposition A.2 *Under the assumptions of Theorem 3.1 we have*

$$\begin{aligned} & P(v\varepsilon \leq t d_{n2}^*(X) + v d_{n1}(X) + d_{n1}^*(X) \mid \mathcal{Y}_n) - P(v\varepsilon \leq t d_{n2}(X) + v d_{n1}(X) \mid \mathcal{Y}_n) \\ & - P(-v\varepsilon \leq t d_{n2}^*(X) - v d_{n1}(X) - d_{n1}^*(X) \mid \mathcal{Y}_n) + P(-v\varepsilon \leq t d_{n2}(X) - v d_{n1}(X) \mid \mathcal{Y}_n) \\ = \;& 2 g_\varepsilon(t) \int \frac{\widehat{m}^*(x) - \widehat{m}(x)}{\sigma(x)} \, dF_X(x) + o_p(\frac{1}{\sqrt{n}}) \end{aligned}$$

uniformly in $t \in I\!R$, where g_ε is defined in (10).

Proposition A.3 *Under the assumptions of Theorem 3.1 we have*

$$\frac{1}{n}\sum_{i=1}^{n}\Big(I\{v_i\widehat{\varepsilon}_i\le t\}-I\{v_i\varepsilon_i\le t\}-P(v\varepsilon\le td_{n2}(X)+vd_{n1}(X)\mid\mathcal{Y}_n)+P(v\varepsilon\le t)\Big)\;=\;o_p(\frac{1}{\sqrt{n}})$$

uniformly in $t\in I\!R$.

Proposition A.4 *Under the assumptions of Theorem 3.1 we have*

$$\int\frac{\widehat{m}^*(x)-\widehat{m}(x)}{\sigma(x)}\,dF_X(x)\;=\;h^2B(t)/(2g_\varepsilon(t))+\frac{1}{n}\sum_{j=1}^{n}\varepsilon_jv_j+o_p(\frac{1}{\sqrt{n}})$$

where $B(t)$ is defined in Theorem 4.1.

From Proposition A.1, an analogous result for the empirical distribution function $\widehat{F}^*_{n,-\varepsilon}(t)=\frac{1}{n}\sum_{i=1}^{n}I\{-\widehat{\varepsilon}^*_i\le t\}$, and Proposition A.2 we have uniformly with respect to $t\in I\!R$ [see also the identity (8) in the proof of Theorem 3.1 and note that g_ε is symmetric]

$$\begin{aligned}\widehat{S}^*_n(t)-h^2B(t)\;&=\;\frac{1}{n}\sum_{i=1}^{n}\Big(I\{\widehat{\varepsilon}^*_i\le t\}-I\{-\widehat{\varepsilon}^*_i\le t\}\Big)-h^2B(t)\\&=\;\frac{1}{n}\sum_{i=1}^{n}\Big(I\{v_i\widehat{\varepsilon}_i\le t\}-I\{-v_i\widehat{\varepsilon}_i\le t\}\Big)+2g_\varepsilon(t)\int\frac{\widehat{m}^*(x)-\widehat{m}(x)}{\sigma(x)}\,dF_X(x)\\&\quad-h^2B(t)+o_p(\frac{1}{\sqrt{n}}).\end{aligned}$$

Now an application of Proposition A.3, an analogous result for $\widehat{F}^*_{n,-\varepsilon}(t)$, and Proposition A.4 yields

$$\begin{aligned}\widehat{S}^*_n(t)-h^2B(t)\;&=\;\frac{1}{n}\sum_{i=1}^{n}\Big(I\{v_i\varepsilon_i\le t\}-I\{-v_i\varepsilon_i\le t\}\Big)+P(v\varepsilon\le td_{n2}(X)+vd_{n1}(X)\mid\mathcal{Y}_n)\\&\quad-P(v\varepsilon\le t)-P(-v\varepsilon\le td_{n2}(X)-vd_{n1}(X)\mid\mathcal{Y}_n)+P(-v\varepsilon\le t)\\&\quad+2g_\varepsilon(t)\frac{1}{n}\sum_{j=1}^{n}\varepsilon_jv_j+o_p(\frac{1}{\sqrt{n}})\\&=\;\frac{1}{n}\sum_{i=1}^{n}\Big(I\{v_i\varepsilon_i\le t\}-I\{-v_i\varepsilon_i\le t\}+2g_\varepsilon(t)\varepsilon_iv_i\Big)+o_p(\frac{1}{\sqrt{n}})\\&=\;\frac{1}{n}\sum_{i=1}^{n}v_i\Big(I\{\varepsilon_i\le t\}-I\{-\varepsilon_i\le t\}+2g_\varepsilon(t)\varepsilon_i\Big)+o_p(\frac{1}{\sqrt{n}}),\end{aligned}$$

where in the last two equalities we have used $P(v_i=1)=P(v_i=-1)=1/2$. By an application of Markov's inequality we obtain, conditional on $\mathcal{Y}_n$, that the processes $\sqrt{n}(\widehat{S}^*_n(t)-h^2B(t))$ and

$$R^*_n(t)\;=\;\frac{1}{\sqrt{n}}\sum_{i=1}^{n}v_i\Big(I\{\varepsilon_i\le t\}-I\{-\varepsilon_i\le t\}+2g_\varepsilon(t)\varepsilon_i\Big),$$

are asymptotically equivalent with respect to weak convergence, that is, for all $\delta > 0$

$$P\Big(\sup_{t\in I\!R} \Big| \sqrt{n}\, \Big(\widehat{S}_n^*(t) - h^2 B(t) \Big) - R_n^*(t) \Big| > \delta \;\Big|\; \mathcal{Y}_n \Big) \;=\; o_p(1).$$

To prove weak convergence we rewrite the process R_n^* as follows,

$$R_n^*(t) \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^n v_i h_t(\varepsilon_i),$$

where the function class

$$\mathcal{F} = \{ h_t(\varepsilon) = I\{\varepsilon \le t\} - I\{-\varepsilon \le t\} + 2g_\varepsilon(t)\varepsilon \mid t \in I\!R \}$$

is Donsker. Conditionally on the sample $\mathcal{Y}_n$, the finite dimensional distributions converge to normal distributed random vectors on account of Lindeberg's central limit theorem, for almost every sequence $\varepsilon_1, \varepsilon_2 \ldots$ [compare van der Vaart and Wellner (1996, Lemma 2.9.5, p. 181)]. To evaluate the covariance of the limit process, at first we calculate the conditional covariance,

$$\begin{aligned}
\mathrm{Cov}(R_n^*(s), R_n^*(t) \mid \mathcal{Y}_n) \;&=\; E[R_n^*(s) R_n^*(t) \mid \mathcal{Y}_n] \;=\; \frac{1}{n} \sum_{i=1}^n h_s(\varepsilon_i) h_t(\varepsilon_i) \\
&= \frac{1}{n} \sum_{i=1}^n \Big(I\{\varepsilon_i \le s \wedge t\} + I\{-\varepsilon_i \le s \wedge t\} - I\{\varepsilon_i \le t, -\varepsilon_i \le s\} \\
&\qquad - I\{-\varepsilon_i \le t, \varepsilon_i \le s\} + 4g_\varepsilon(s) g_\varepsilon(t) \varepsilon_i^2 + 2g_\varepsilon(s)\varepsilon_i (I\{\varepsilon_i \le t\} \\
&\qquad - I\{-\varepsilon_i \le t\}) + 2g_\varepsilon(t)\varepsilon_i (I\{\varepsilon_i \le s\} - I\{-\varepsilon_i \le s\}) \Big).
\end{aligned}$$

This conditional covariance converges almost surely to

$$\begin{aligned}
& F_\varepsilon(s \wedge t) + F_{-\varepsilon}(s \wedge t) + F_\varepsilon((-s) \wedge t) + F_{-\varepsilon}((-s) \wedge t) - F_\varepsilon(t) - F_{-\varepsilon}(t) + 4g_\varepsilon(s) g_\varepsilon(t) \\
& + 2g_\varepsilon(s) \Big(\int_{-\infty}^t x f_\varepsilon(x)\, dx - \int_{-t}^\infty x f_\varepsilon(x)\, dx \Big) + 2g_\varepsilon(t) \Big(\int_{-\infty}^s x f_\varepsilon(x)\, dx - \int_{-s}^\infty x f_\varepsilon(x)\, dx \Big) \\
=\;& \mathrm{Cov}(S(s), S(t)),
\end{aligned}$$

which is the covariance kernel claimed in Theorem 4.1. Conditional weak convergence of the process R_n^* to a Gaussian process with the above covariance structure, in probability, now follows from an imitation of the proof of the Conditional Multiplier Central Limit Theorem of van der Vaart and Wellner (1996, Th. 2.9.6, p. 182). Note that under the null hypothesis of symmetry the aforementioned theorem is directly applicable because in this case we have $E[h_t(\varepsilon_1)] = 0$. $\square$

Proof of Proposition A.1.

The proof of Proposition A.1 is similar to the proof of Lemma 1, Appendix B, of Akritas and Keilegom (2001) and we only explain the main differences. The idea is to consider the process quoted in Proposition A.1 as an empirical process of the form

$$\frac{1}{n} \sum_{i=1}^n \Big(f(v_i, \varepsilon_i, X_i) - E[f(v_i, \varepsilon_i, X_i)] \Big) \tag{17}$$

indexed by a class of functions $\mathcal{F}$ given by

$$\mathcal{F} = \Big\{ f(v,\varepsilon,x) = I\{v\varepsilon \le td_2^*(x) + vd_1(x) + d_1^*(x)\} - I\{v\varepsilon \le td_2(x) + vd_1(x)\} \ \Big| \ t \in I\!R,\ d_1, d_1^* \in C_1,\ d_2, d_2^* \in C_2 \Big\}.$$

The function classes C_1 and C_2 are defined by

$$C_1 = \Big\{ d : [0,1] \to I\!R \ \Big| \ d \text{ differentiable },\ ||d||_{1+\delta} \le 1 \Big\}$$
$$C_2 = \Big\{ d : [0,1] \to I\!R \ \Big| \ d \text{ differentiable },\ ||d||_{1+\delta} \le 2,\ \inf_{x\in[0,1]} d(x) \ge 1/2 \Big\},$$

where δ is defined in Section 2 and

$$||d||_{1+\delta} = \max\Big\{ \sup_{x\in[0,1]} |d(x)|, \sup_{x\in[0,1]} |d'(x)| \Big\} + \sup_{x,y\in[0,1]} \frac{|d'(x) - d'(y)|}{|x-y|^{\delta/2}}.$$

The function class $\mathcal{F}$ is Donsker and we have

$$\lim_{n\to\infty} P\Big(d_{n1} \in C_1 \text{ and } d_{n2} \in C_2 \Big) = 1$$

[see the proof of Lemma 1, Appendix B, Akritas and Keilegom (2001)]. Similarly, in order to show

$$\lim_{n\to\infty} P\Big(d_{n1}^* \in C_1 \text{ and } d_{n2}^* \in C_2 \Big) = 1$$

we have to verify the following conditions,

$$\sup_{x\in[0,1]} |\widehat{m}^*(x) - \widehat{m}(x)| = o(1) \quad \text{a.s.} \tag{18}$$

$$\sup_{x\in[0,1]} |\widehat{\sigma}^*(x) - \widehat{\sigma}(x)| = o(1) \quad \text{a.s.} \tag{19}$$

$$\sup_{x\in[0,1]} |(\widehat{m}^*)'(x) - \widehat{m}'(x)| = o(1) \quad \text{a.s.} \tag{20}$$

$$\sup_{x\in[0,1]} |(\widehat{\sigma}^*)'(x) - \widehat{\sigma}'(x)| = o(1) \quad \text{a.s.} \tag{21}$$

$$\sup_{x,y\in[0,1]} |(\widehat{m}^*)'(x) - \widehat{m}'(x) - (\widehat{m}^*)'(y) + \widehat{m}'(y)| \,/\, |x-y|^{\delta/2} = o(1) \quad \text{a.s.} \tag{22}$$

$$\sup_{x,y\in[0,1]} |(\widehat{\sigma}^*)'(x) - \widehat{\sigma}'(x) - (\widehat{\sigma}^*)'(y) + \widehat{\sigma}'(y)| \,/\, |x-y|^{\delta/2} = o(1) \quad \text{a.s.} \tag{23}$$

The condition (18) is valid due to the following decomposition

$$\widehat{m}^*(x) - \widehat{m}(x) = \frac{1}{\widehat{f}_X(x)} \frac{1}{nh} \sum_{i=1}^n K\Big(\frac{x - X_i}{h}\Big)(Y_i^* - \widehat{m}(x)) = M_1(x) + \ldots + M_4(x)$$

where $\widehat{f}_X(x) = \frac{1}{nh}\sum_{i=1}^n K(\frac{x-X_i}{h})$ is the kernel estimator for the design density f_X and

$$M_1(x) = m(x) - \widehat{m}(x)$$

$$\begin{aligned} M_2(x) &= \frac{1}{\widehat{f}_X(x)} \frac{1}{nh} \sum_{i=1}^n K\Big(\frac{x-X_i}{h}\Big)(m(X_i)+\sigma(X_i)v_i\varepsilon_i - m(x)) \\ M_3(x) &= \frac{1}{\widehat{f}_X(x)} \frac{1}{nh} \sum_{i=1}^n K\Big(\frac{x-X_i}{h}\Big)(\widehat{m}(X_i)-m(X_i)) \\ M_4(x) &= \frac{1}{\widehat{f}_X(x)} \frac{1}{nh} \sum_{i=1}^n K\Big(\frac{x-X_i}{h}\Big)v_i(\widehat{m}(X_i)-m(X_i)). \end{aligned}$$

We directly obtain $\sup_{x\in[0,1]}|M_1(x)| = o(1)$ and $\sup_{x\in[0,1]}|M_2(x)| = o(1)$, almost surely, because $M_2(x)$ is equivalent to $\widetilde{m}(x) - m(x)$ where $\widetilde{m}(x)$ is the Nadaraya–Watson estimator for $m(x)$ in the regression model $\widetilde{Y}_i = m(X_i) + \sigma(X_i)v_i\varepsilon_i$. We straightforward estimate

$$\sup_{x\in[0,1]}|M_3(x)| \leq \sup_{x\in[0,1]}|\widehat{m}(x) - m(x)| = o(1)$$

almost surely and the same estimation is valid for $\sup_{x\in[0,1]}|M_4(x)|$. Conditions (19)–(23) can be shown in a similar manner.
Now the result of Proposition A.1 follows analogous to the proof of Lemma 1 of Akritas and Keilegom (2001) with an application of Corollary 2.3.12 of van der Vaart and Wellner (1996, p. 115) and $\mathrm{Var}(f(v_i,\varepsilon_i,X_i)) \to 0$ for

$$f(v_i,\varepsilon_i,X_i) = I\{v_i\varepsilon_i \leq td^*_{n2}(X_i) + v_i d_{n1}(X_i) + d^*_{n1}(X_i)\} - I\{v_i\varepsilon_i \leq td_{n2}(X_i) + v_i d^*_{n1}(X_i)\}.$$

□

Acknowledgements. The authors are grateful to Isolde Gottschlich, who typed parts of this paper with considerable technical expertise. The work of the authors was supported by the Deutsche Forschungsgemeinschaft (SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen). Parts of this paper were written during a visit of the second author at Purdue University and this author would like to thank the Department of Statistics for its hospitality.

References

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

S. Aki (1981). *Asymptotic distribution of a Cramér-von-Mises type statistic for testing symmetry when the center ist estimated.* Ann. Inst. Statist. Math. 33, 1–14.

M. Akritas and I. van Keilegom (2000). *Nonparametric estimation of the residual distribution.* Scand. J. Statist. 28, 549–567.

Bhattacharya, Gastwirth, Wright (1982). *Two modified Wilcoxon tests for symmetry about an unknown location parameter.* Biometrika 69, 377–382.

H. Dette, S. Kusi–Appiah and N. Neumeyer (2002). *Testing symmetry in nonparametric regression models.* Nonparam. Statistics 14, 477–494.

J. Fan, I. Gijbels (1996). *Local polynomial modelling and its applications.* Chapman and Hall, London.

M. Hollander (1988). *Testing for symmetry.* Encycl. Statist. Sci 9 (eds. N.L. Johnson, S. Kotz), 211–216.

M. Huškova (1984). *Hypothesis of symmetry.* In: Handbook of Statistics, Vol. 4, Nonparametric methods, 63–78. North Holland, Amsterdam.

R. J. Hyndman and Q. Yao (2002). *Nonparametric estimation and symmetry tests for conditional density functions.* Nonparam. Statistics 14, 259–278.

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

J. A. Koziol (1985). *A note on testing symmetry with estimated parameters.* Statist. Prob. Letters 3, 227–230.

Z. Psaradakis (2003). *A bootstrap test for symmetry of dependent data based on a Komogorov-Smirnov type statistic.* Comm. Statist., Ser. B, Simul. 8 Comput. 32, 113–126.

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

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

E. F. Schuster, R. G. Berger (1987). *Using the bootstrap in testing symmetry versus asymmetry.* Comm. Statist., Ser. B, Simul. & Comput. 16, 69–84.

G. R. Shorack and J. A. Wellner (1986). *Empirical Processes with Applications ot Statistics.* Wiley.

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

A. W. van der Vaart (1998). *Asymptotic Statistics.* Cambridge University Press.

M. P. Wand, M. C. Jones (1995). *Kernel smoothing.* Chapman and Hall, London.

H. Witting, U. Müller–Funk (1995). *Mathematische Statistik II.* B. G. Teubner Stuttgart.