

TESTING FOR SYMMETRIC ERROR DISTRIBUTION IN NONPARAMETRIC REGRESSION MODELS

Natalie Neumeyer and Holger Dette

Universität Hamburg and Ruhr-Universität Bochum

Abstract: For the problem of testing symmetry of the error distribution in a nonparametric regression model, we investigate the asymptotic properties of 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 available procedures, the new test is also applicable under heteroscedasticity.

Key words and phrases: Empirical process of residuals, testing for symmetry, nonparametric regression.

1. Introduction

Consider the nonparametric heteroscedastic regression model

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

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:

$$H_0 : F_\varepsilon(t) = 1 - F_\varepsilon(-t) \text{ for all } t \in \mathbb{R}$$

versus

$$H_1 : F_\varepsilon(t) \neq 1 - F_\varepsilon(-t) \text{ for some } t \in \mathbb{R}$$

In the case where $\varepsilon_1, \dots, \varepsilon_n$ are directly observable Smirnov (1947) proposed to compare the empirical distribution functions of ε_i and $-\varepsilon_i$ using the empirical

process

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

where $I\{\cdot\}$ denotes the indicator function, $F_{n,\varepsilon}$ is the empirical distribution function of $\varepsilon_1, \dots, \varepsilon_n$ and $F_{n,-\varepsilon}$ is the empirical distribution function of $-\varepsilon_1, \dots, -\varepsilon_n$. Under the hypothesis of symmetry $F_\varepsilon = F_{-\varepsilon}$ and 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, and a suitable asymptotic distribution-free test statistic is then obtained by the Cramer-von-Mises functional. The resulting test is consistent with respect to local alternatives converging to the null at a rate $n^{-1/2}$ (see Koul and Staudte Jr. (1976) for some bounds on the power function of these tests).

The problem of testing symmetry of the unknown distribution of the residuals in special parametric regression models has been considered by numerous authors (see for example Bhattacharya, Gastwirth and Wright (1982), Aki (1981), Hušková (1984), Koziol (1985), Schuster and Berger (1987), Hollander (1988), Ahmad and Li (1997), Hyndman and Yao (2002) or Psaradakis (2003), among many others). A generalization of the process (2) for the unknown residuals $\varepsilon_1, \dots, \varepsilon_n$ in linear models with fixed design and homoscedastic error structure can be found in Koul (2002, p.258). In the present paper we transfer this approach to the problem of testing the hypothesis of a symmetric error distribution in a nonparametric regression model with heteroscedastic error structure. The process defined in (2) is modified by replacing the unknown errors ε_i by estimated residuals $\hat{\varepsilon}_i = (Y_i - \hat{m}(X_i))/\hat{\sigma}(X_i)$ ($i = 1, \dots, n$), where $\hat{m}(\cdot)$ and $\hat{\sigma}(\cdot)$ denote kernel based nonparametric estimators for the regression and variance function, respectively. This yields the process

$$\hat{S}_n(t) = \frac{1}{n} \sum_{i=1}^n \left(I\{\hat{\varepsilon}_i \leq t\} - I\{-\hat{\varepsilon}_i \leq t\} \right),$$

and this allows us to consider heteroscedastic 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 the smoothing parameter of a kernel estimator (see Ahmad and Li (1997) and Dette, Kusi-Appiah and Neumeier (2002)). It is the second purpose of this paper to construct a test for the symmetry of the error distribution in model (1) that can detect local alternatives at a rate $n^{-1/2}$.

In Section 3 we prove weak convergence of a centered version of the empirical process \widehat{S}_n to a Gaussian process under the null hypothesis of a symmetric error distribution, local alternatives, 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 asymptotically 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 existing procedures derived under the additional assumption of homoscedasticity.

2. Technical Assumptions

In this section we state some technical assumptions that are required for the statement of the main results in Sections 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. Further suppose that the conditional distribution $P^{Y_i|X_i=x}$ of Y_i given $X_i = x$ has distribution function

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

and density

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

such that $\sup_{x,y} |yf(y|x)| < \infty$, 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 with $\min_{x \in [0,1]} \sigma^2(x) \geq c > 0$ for some constant c .

Throughout, let K be a symmetric twice continuously differentiable density with compact support and vanishing first moment and let $h = h_n$ denote a sequence of bandwidths converging to zero as $n \rightarrow \infty$ such that $nh^4 = O(1)$ and $nh^{3+\delta}/\log(1/h) \rightarrow \infty$ for some $\delta > 0$.

3. Weak Convergence of the Empirical Symmetry Process

For the estimation of the residuals we define nonparametric kernel estimators of the unknown regression function $m(\cdot)$ and variance function $\sigma^2(\cdot)$ in model (1) by

$$\hat{m}(x) = \frac{\sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)Y_i}{\sum_{j=1}^n K\left(\frac{X_j-x}{h}\right)} \quad (3)$$

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)(Y_i - \hat{m}(x))^2}{\sum_{j=1}^n K\left(\frac{X_j-x}{h}\right)}. \quad (4)$$

Note, that $\hat{m}(\cdot)$ is the usual Nadaraya–Watson estimator (Nadaraya (1964) and Watson (1964)) 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

$$\hat{\varepsilon}_i = \frac{Y_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)} \quad (i = 1, \dots, n). \quad (5)$$

The considered empirical process is based on the residuals (5) and is given by

$$\hat{S}_n(t) = \hat{F}_{n,\varepsilon}(t) - \hat{F}_{n,-\varepsilon}(t) = \frac{1}{n} \sum_{i=1}^n \left(I\{\hat{\varepsilon}_i \leq t\} - I\{-\hat{\varepsilon}_i \leq t\} \right). \quad (6)$$

Throughout, we call the process $\hat{S}_n(t)$ (and any process of the same form) an empirical symmetry process. 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*

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

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

ture

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

where the bias is

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

The proof of Theorem 3.1 is deferred to the Appendix.

Note that in the asymptotic covariance, one can write

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

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

$$2F_\varepsilon(s \wedge t) - 2F_\varepsilon(t) + 2F_\varepsilon((-s) \wedge t) = 2F_\varepsilon(-(|s| \vee |t|)), \tag{7}$$

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

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

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 $\{\sqrt{n}(\widehat{S}_n(t) - h^2 B(t))\}_{t \in \mathbb{R}}$ defined in (6) converges weakly to a centered Gaussian process $\{S(t)\}_{t \in \mathbb{R}}$ with covariance*

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

Comparing the covariance kernel H with the expression (7) obtained by Smirnov (1947) we see that there are 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 also note that the bias $h^2 B(t)$ in Theorem 3.1 and Corollary 3.2 can be omitted if $h^4 n = o(1)$.

Remark 3.3. *Linear models.* We compare the results obtained in the nonparametric setting to results of Koul (2002). To this end consider a homoscedastic linear model $Y_{ni} = x_{ni}^T \beta + \varepsilon_{ni}$ with fixed design, where the errors $\varepsilon_{n1}, \dots, \varepsilon_{nn}$ are independent and identically distributed with distribution function F_ε and variance $\text{Var}(\varepsilon_{ni}) \equiv 1$, for simplicity. Koul (2002) considers M-estimators $\widehat{\beta}_n$ for the parameter β , but we concentrate here on the special case of the least squares estimator $\widehat{\beta}_n$. Then, under certain regularity assumptions, the process $\sqrt{n}\widehat{S}_n$ based on the parametric residuals $\widehat{\varepsilon}_{ni} = Y_{ni} - x_{ni}^T \widehat{\beta}_n$ converges weakly under the null hypothesis of a symmetric error distribution to a centered Gaussian process S with covariance

$$\begin{aligned} \text{Cov}(S(s), S(t)) &= 2F_\varepsilon(-(|s| \vee |t|)) + 4cf_\varepsilon(s)f_\varepsilon(t) \\ &\quad + 4cf_\varepsilon(s) \int_{-\infty}^t x f_\varepsilon(x) dx + 4cf_\varepsilon(t) \int_{-\infty}^s x f_\varepsilon(x) dx. \quad (8) \end{aligned}$$

Here the constant c is $\lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^n \sum_{i=1}^n x_{ni}^T (X_n^T X_n)^{-1} x_{nj}$, where x_{ni}^T denotes the i th row of the design matrix X_n . Thus the covariance structure in (8) is essentially the same as obtained in Corollary 3.2 for the nonparametric regression model.

Remark 3.4. *Local alternatives.* Asymptotic tests based on the empirical symmetry process can detect local alternatives converging to the null at a rate $n^{-1/2}$. To see this, consider the nonparametric regression model $Y_{ni} = m(X_i) + \sigma(X_i)\varepsilon_{ni}$, where the errors $\varepsilon_{n1}, \dots, \varepsilon_{nn}$ are independent with distribution function F_{ε_n} . We assume that

$$\lim_{n \rightarrow \infty} \sqrt{n}G_n(t) := \lim_{n \rightarrow \infty} \sqrt{n}[F_{\varepsilon_n}(t) - (1 - F_{\varepsilon_n}(-t))] = G(t)$$

uniformly in $t \in \mathbb{R}$, for some function G , and that the functions G_n are continuously differentiable such that $\lim_{n \rightarrow \infty} \sup_{t \in \mathbb{R}} |G'_n(t)| = 0$. The conditional distribution function $F(y|x)$, now n -dependent, should fulfill the assumptions stated in Section 2 uniformly in $n \in \mathbb{N}$. Then the process $\sqrt{n}(\widehat{S}_n(t) - h^2B(t))$, $t \in \mathbb{R}$, defined in (6), converges weakly to a Gaussian process with expectation $G(t)$ and covariance defined in Corollary 3.2. Essentially, the proof follows along the lines of the proof of Theorem 3.1, but Theorem 1 in Akritas and Van Keilegom (2001) is not directly applicable because now the error variables $\varepsilon_{n1}, \dots, \varepsilon_{nn}$ build a triangular array with n -dependent distribution. With some technical effort Akritas and Van Keilegom's (2001) proof can be generalized to this case under application of Theorem 2.11.9 of Van der Vaart and Wellner (1996, p.211) The details are omitted for the sake of brevity.

4. Symmetric Wild Bootstrap

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

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

where \widehat{H}_n is the empirical distribution function of $|\widehat{\varepsilon}_1|, \dots, |\widehat{\varepsilon}_n|$, 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 , $\sup_{t \in \mathbb{R}} |\widehat{H}_n(t) - H(t)| = o_p(1)$, where H denotes the distribution function of $|\varepsilon_1|$. 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 $F_\varepsilon, f_\varepsilon$ we propose a modification of the wild bootstrap approach, which is adapted to the specific problem of

testing symmetry (for wild bootstrap procedures, see Wu (1986), Liu (1988) and Härdle and Mammen (1993), among many others).

For this let v_1, \dots, v_n be Rademacher variables, independent identically distributed with $P(v_i = 1) = P(v_i = -1) = 1/2$, independent of the sample (X_j, Y_j) , $j = 1, \dots, n$. Note that whether the underlying error distribution F_ε is symmetric or not the distribution of the random variable $v_i \varepsilon_i$ is symmetric with density and distribution functions

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

respectively. Define bootstrap residuals as follows,

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

where $\hat{\varepsilon}_i$ is given in (5). Now we build new bootstrap observations ($i = 1, \dots, n$)

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

and estimated residuals

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

where the regression and variance estimates \hat{m}^* and $\hat{\sigma}^{*2}$ are defined analogously to \hat{m} and $\hat{\sigma}^2$ in (3) and (4), but are based on the bootstrap sample (X_i, Y_i^*) , $i = 1, \dots, n$. In generalization of (6) the bootstrap version of the empirical symmetry process is

$$\hat{S}_n^*(t) = \hat{F}_{n,\varepsilon}^*(t) - \hat{F}_{n,-\varepsilon}^*(t) = \frac{1}{n} \sum_{i=1}^n \left(I\{\hat{\varepsilon}_i^* \leq t\} - I\{-\hat{\varepsilon}_i^* \leq t\} \right).$$

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*

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

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

$$\begin{aligned} \text{Cov}(S(s), S(t)) &= 2G_\varepsilon(-(|s| \vee |t|)) + 4g_\varepsilon(s)g_\varepsilon(t) + 4g_\varepsilon(s) \int_{-\infty}^t x g_\varepsilon(x) dx \\ &\quad + 4g_\varepsilon(t) \int_{-\infty}^s x g_\varepsilon(x) dx \end{aligned}$$

in probability, where the bias term is

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

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 $Cov(S(s), S(t)) = H(s, t)$, 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

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

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 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. Consider the statistic

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

where $\widehat{H}_n(t) = (1/n) \sum_{i=1}^n I\{|\widehat{\varepsilon}_i| \leq t\}$ denotes the empirical distribution function of the absolute residuals $|\widehat{\varepsilon}_1|, \dots, |\widehat{\varepsilon}_n|$. 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^* denotes the empirical distribution function of $|\widehat{\varepsilon}_1^*|, \dots, |\widehat{\varepsilon}_n^*|$, the consistency of the bootstrap procedure follows from Theorem 4.1, the Continuous Mapping Theorem and the fact that for all $\tau > 0$ we have

$$P\left(\sup_{t \in \mathbb{R}} |\widehat{H}_n^*(t) - H(t)| > \tau \mid \mathcal{Y}_n\right) = o_p(1).$$

For the bandwidth in the regression and variance estimator defined by (3) and (4), respectively, we used

$$h = \left(\frac{\widehat{s}^2}{n}\right)^{\frac{3}{10}}, \quad (14)$$

where

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

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

$B = 200$ bootstrap replications based on one sample $\mathcal{Y}_n = \{(X_i, Y_i) \mid i = 1, \dots, n\}$ were performed for each simulation, where 1,000 runs were used to calculate the rejection probabilities. The quantile estimate $c_{1-\alpha}^*$ defined in (12) from the bootstrap sample $T_n^{*,1}, \dots, T_n^{*,B}$ was estimated by $\widehat{c}_{1-\alpha}^* = T_n^{*,(\lfloor B(1-\alpha) \rfloor)}$, where $T_n^{*,(i)}$ denotes the i th order statistic of $T_n^{*,1}, \dots, T_n^{*,B}$. 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

$$Y_i = \sin(2\pi X_i) + \sigma(X_i)\varepsilon_i, \quad i = 1, \dots, n, \quad (16)$$

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 $\sigma(x) \equiv 1$. 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, 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 a 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 5.1. 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 $E[\varepsilon^2] = 1$.

| | α | $n = 20$ | $n = 30$ | $n = 40$ | $n = 50$ | $n = 100$ |
|--------|----------|----------|----------|----------|----------|-----------|
| df_1 | 0.025 | 0.029 | 0.033 | 0.029 | 0.029 | 0.027 |
| | 0.05 | 0.057 | 0.060 | 0.051 | 0.057 | 0.052 |
| | 0.10 | 0.109 | 0.111 | 0.107 | 0.107 | 0.104 |
| | 0.20 | 0.214 | 0.216 | 0.215 | 0.193 | 0.209 |
| df_2 | 0.025 | 0.035 | 0.032 | 0.024 | 0.032 | 0.029 |
| | 0.05 | 0.062 | 0.055 | 0.051 | 0.068 | 0.057 |
| | 0.10 | 0.113 | 0.111 | 0.101 | 0.113 | 0.108 |
| | 0.20 | 0.215 | 0.209 | 0.204 | 0.210 | 0.193 |
| df_3 | 0.025 | 0.031 | 0.030 | 0.028 | 0.028 | 0.030 |
| | 0.05 | 0.055 | 0.051 | 0.061 | 0.049 | 0.067 |
| | 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 5.2. 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 $E[\varepsilon^2] = 1$.

| k | α | $n = 20$ | $n = 30$ | $n = 40$ | $n = 50$ | $n = 100$ |
|-----|----------|----------|----------|----------|----------|-----------|
| 1 | 0.025 | 0.358 | 0.654 | 0.849 | 0.957 | 1.000 |
| | 0.05 | 0.484 | 0.764 | 0.912 | 0.981 | 1.000 |
| | 0.10 | 0.584 | 0.847 | 0.959 | 0.991 | 1.000 |
| | 0.20 | 0.716 | 0.914 | 0.983 | 0.998 | 1.000 |
| 2 | 0.025 | 0.239 | 0.458 | 0.698 | 0.817 | 0.998 |
| | 0.05 | 0.342 | 0.570 | 0.805 | 0.896 | 1.000 |
| | 0.10 | 0.442 | 0.681 | 0.865 | 0.936 | 1.000 |
| | 0.20 | 0.594 | 0.794 | 0.934 | 0.976 | 1.000 |
| 3 | 0.025 | 0.208 | 0.436 | 0.604 | 0.750 | 0.982 |
| | 0.05 | 0.303 | 0.565 | 0.710 | 0.833 | 0.995 |
| | 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 5.3. 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 of 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.

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 a chi-square distribution with k degrees of freedom ($df_k, k = 1, 2, 3$) standardized such that $E[\varepsilon] = 0, E[\varepsilon^2] = 1$.

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

Appendix. Proofs

A.1. Proof of Theorem 3.1

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

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

where, uniformly in $t \in \mathbb{R}$, $r_n(t) = o_p(1/\sqrt{n}) + o_p(h^2) = o_p(1/\sqrt{n})$ and

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

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 Van Keilegom (2001, p.555), we also have for the bias term

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

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

$$\begin{aligned} \widehat{S}_n(t) &= \widehat{F}_{n,\varepsilon}(t) - \widehat{F}_{n,-\varepsilon}(t) \\ &= \frac{1}{n} \sum_{i=1}^n \left(I\{\widehat{\varepsilon}_i \leq t\} - I\{-\widehat{\varepsilon}_i \leq t\} \right) \\ &= \frac{1}{n} \sum_{i=1}^n \left(I\{\varepsilon_i \leq t\} - I\{-\varepsilon_i \leq t\} + \varepsilon_i (f_\varepsilon(t) + f_\varepsilon(-t)) + (\varepsilon_i^2 - 1) \frac{t}{2} (f_\varepsilon(t) - f_\varepsilon(-t)) \right) \\ &\quad + h^2 B(t) + o_p\left(\frac{1}{\sqrt{n}}\right) \end{aligned} \tag{17}$$

uniformly with respect to $t \in \mathbb{R}$, 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 (17), which is due to the estimation of the variance function, vanishes. From the above expansion we obtain

$$\begin{aligned} R_n(t) &= \sqrt{n} \left(\widehat{S}_n(t) - F_\varepsilon(t) + (1 - F_\varepsilon(-t)) - h^2 B(t) \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(I\{\varepsilon_i \leq t\} - F_\varepsilon(t) - I\{-\varepsilon_i \leq t\} + (1 - F_\varepsilon(-t)) \right. \\ &\quad \left. + \varepsilon_i (f_\varepsilon(t) + f_\varepsilon(-t)) + (\varepsilon_i^2 - 1) \frac{t}{2} (f_\varepsilon(t) - f_\varepsilon(-t)) \right) + o_p(1) \\ &= \widetilde{R}_n(t) + o_p(1), \end{aligned}$$

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

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

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 $P_n h_t = \frac{1}{n} \sum_{i=1}^n h_t(\varepsilon_i)$, $P h_t$ 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

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

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 $\sup_{t \in \mathbb{R}} |f_\varepsilon(t)| < \infty$, $\sup_{t \in \mathbb{R}} |t f_\varepsilon(t)| < \infty$ 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 a finite second moment by our assumptions.

The function class $\mathcal{G} = \{h_t \mid t \in \mathcal{Q}\}$ 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 $t_m \in \mathcal{Q}$

with $t_m \searrow t$ as $m \rightarrow \infty$, we have $g_m(\varepsilon) = h_{t_m}(\varepsilon) \rightarrow h_t(\varepsilon)$ for $m \rightarrow \infty$. The convergence is also valid in the L^2 -sense:

$$\begin{aligned} & P((g_m - h_t)^2) \\ & \leq 6 \left(F_\varepsilon(t) - F_\varepsilon(t_m) + F_{-\varepsilon}(t) - F_{-\varepsilon}(t_m) + E[\varepsilon_1^2](f_\varepsilon(t) - f_\varepsilon(t_m))^2 \right. \\ & \quad + E[\varepsilon_1^2](f_\varepsilon(-t) - f_\varepsilon(-t_m))^2 + E[(\varepsilon_1^2 - 1)^2] \frac{1}{4} (tf_\varepsilon(t) - t_m f_\varepsilon(t_m))^2 \\ & \quad \left. + E[(\varepsilon_1^2 - 1)^2] \frac{1}{4} (tf_\varepsilon(-t) - t_m f_\varepsilon(-t_m))^2 \right) \\ & \rightarrow 0 \quad \text{for } m \rightarrow \infty. \end{aligned}$$

This proves pointwise separability 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 $\{\varepsilon \mapsto I\{\varepsilon \leq t\} \mid t \in \mathbb{R}\}$ and $\{\varepsilon \mapsto I\{-\varepsilon \leq t\} \mid t \in \mathbb{R}\}$ are VC by standard results. Finally, the function class

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

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

A.2. Proof of Theorem 4.1

We decompose the residuals $\hat{\varepsilon}_i^*$ defined in (11) as

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

Hence for $t \in \mathbb{R}$, the inequality $\hat{\varepsilon}_i^* \leq t$ is equivalent to $v_i \varepsilon_i \leq td_{n2}^*(X_i) + v_i d_{n1}(X_i) + d_{n1}^*(X_i)$ and $v_i \hat{\varepsilon}_i \leq t$ is equivalent to $v_i \varepsilon_i \leq td_{n2}(X_i) + v_i d_{n1}(X_i)$, where we introduced the definitions

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

In the following we need four auxiliary results which are listed in Proposition A.1–A.4 and can be proved by similar arguments as given in Akritas and Van Keilegom (2001). For the sake of brevity we 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 Van Keilegom (2001), while the proof of Proposition A.3 follows exactly along 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.

Let (X, ε, v) denote a random variable with the same distribution as $(X_1, \varepsilon_1, v_1)$, but independent from \mathcal{Y}_n and v_1, \dots, v_n .

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

$$\frac{1}{n} \sum_{i=1}^n \left(I\{\widehat{\varepsilon}_i^* \leq t\} - P(v\varepsilon \leq td_{n2}^*(X) + vd_{n1}(X) + d_{n1}^*(X) \mid \mathcal{Y}_n) - I\{v_i \widehat{\varepsilon}_i \leq t\} + P(v\varepsilon \leq td_{n2}(X) + vd_{n1}(X) \mid \mathcal{Y}_n) \right) = o_p\left(\frac{1}{\sqrt{n}}\right)$$

uniformly in $t \in \mathbb{R}$.

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

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

uniformly in $t \in \mathbb{R}$, where g_ε is defined in (10).

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

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

uniformly in $t \in \mathbb{R}$.

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

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

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

From Proposition A.1, an analogous result for the empirical distribution function $\widehat{F}_{n,-\varepsilon}^*(t) = (1/n) \sum_{i=1}^n I\{-\widehat{\varepsilon}_i^* \leq t\}$, and Proposition A.2, we have uniformly with respect to $t \in \mathbb{R}$ (see (17) in the proof of Theorem 3.1 and note that

g_ε is symmetric),

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

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

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

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 $\sqrt{n}(\widehat{S}_n^*(t) - h^2 B(t))$ and

$$R_n^*(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n v_i \left(I\{\varepsilon_i \leq t\} - I\{-\varepsilon_i \leq t\} + 2g_\varepsilon(t) \varepsilon_i \right),$$

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

$$P\left(\sup_{t \in \mathbb{R}} \left| \sqrt{n} \left(\widehat{S}_n^*(t) - h^2 B(t) \right) - R_n^*(t) \right| > \tau \mid \mathcal{Y}_n \right) = o_p(1).$$

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 $\mathcal{F} = \{h_t(\varepsilon) = I\{\varepsilon \leq t\} - I\{-\varepsilon \leq t\} + 2g_\varepsilon(t)\varepsilon \mid t \in \mathbb{R}\}$ 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 $\varepsilon_1, \varepsilon_2 \dots$ (compare Van der Vaart and Wellner (1996, Lemma 2.9.5, p.181)). To evaluate the covariance of the limit process, calculate the conditional covariance

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

This conditional covariance converges almost surely to

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

which is the covariance kernel claimed in Theorem 4.1. Conditional weak convergence in probability of the process R_n^* to a Gaussian process with the above covariance structure now follows from an imitation of the proof of the Conditional Multiplier Central Limit Theorem of Van der Vaart and Wellner (1996, Thm. 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 $E[h_t(\varepsilon_1)] = 0$.

Proof of Proposition A.1

The proof of Proposition A.1 is similar to the proof of Lemma 1, Appendix B, of Akritas and Van 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

$$\frac{1}{n} \sum_{i=1}^n \left(f(v_i, \varepsilon_i, X_i) - E[f(v_i, \varepsilon_i, X_i)] \right), \quad (18)$$

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

$$\mathcal{F} = \left\{ f(v, \varepsilon, x) = I\{v\varepsilon \leq td_2^*(x) + vd_1(x) + d_1^*(x)\} - I\{v\varepsilon \leq td_2(x) + vd_1(x)\} \mid t \in \mathbb{R}, d_1, d_1^* \in C_1, d_2, d_2^* \in C_2 \right\}.$$

The function classes C_1 and C_2 are defined by

$$C_1 = \left\{ d : [0, 1] \rightarrow \mathbb{R} \mid d \text{ differentiable, } \|d\|_\delta \leq 1 \right\}$$

$$C_2 = \left\{ d : [0, 1] \rightarrow \mathbb{R} \mid d \text{ differentiable, } \|d\|_\delta \leq 2, \inf_{x \in [0, 1]} d(x) \geq \frac{1}{2} \right\},$$

where δ is defined in Section 2 and

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

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

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

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

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

we have to verify the conditions,

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

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

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

$$\sup_{x \in [0, 1]} |(\hat{\sigma}^*)'(x) - \hat{\sigma}'(x)| = o(1) \quad \text{a. s.}, \tag{22}$$

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

$$\sup_{x, y \in [0, 1]} \frac{|(\hat{\sigma}^*)'(x) - \hat{\sigma}'(x) - (\hat{\sigma}^*)'(y) + \hat{\sigma}'(y)|}{|x - y|^{\frac{\delta}{2}}} = o(1) \quad \text{a. s.} \tag{24}$$

The condition (19) is valid due to the following decomposition

$$\hat{m}^*(x) - \hat{m}(x) = \frac{1}{\hat{f}_X(x)} \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) (Y_i^* - \hat{m}(x)) = M_1(x) + \dots + M_4(x),$$

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

$$\begin{aligned} M_1(x) &= m(x) - \widehat{m}(x), \\ M_2(x) &= \frac{1}{\widehat{f}_X(x)} \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) (m(X_i) + \sigma(X_i)v_i\varepsilon_i - m(x)), \\ M_3(x) &= \frac{1}{\widehat{f}_X(x)} \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) (\widehat{m}(X_i) - m(X_i)), \\ M_4(x) &= \frac{1}{\widehat{f}_X(x)} \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) v_i(\widehat{m}(X_i) - m(X_i)). \end{aligned}$$

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 straightforwardly 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 (20)–(24) can be shown in a similar manner.

Now the result of Proposition A.1 follows analogously to the proof of Lemma 1 of Akritas and Van Keilegom (2001) with an application of Corollary 2.3.12 of Van der Vaart and Wellner (1996, p.115) and $\text{Var}(f(v_i, \varepsilon_i, X_i)) \rightarrow 0$ for

$$\begin{aligned} f(v_i, \varepsilon_i, X_i) &= I\{v_i\varepsilon_i \leq td_{n2}^*(X_i) + v_id_{n1}(X_i) + d_{n1}^*(X_i)\} \\ &\quad - I\{v_i\varepsilon_i \leq td_{n2}(X_i) + v_id_{n1}^*(X_i)\}. \end{aligned}$$

Acknowledgements

The authors are grateful to two referees for their helpful comments on an earlier version of this paper and 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

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

- Aki, S. (1981). Asymptotic distribution of a Cramér-von-Mises type statistic for testing symmetry when the center is estimated. *Ann. Inst. Statist. Math.* **33**, 1-14.
- Akritas, M. and Van Keilegom, I. (2001). Nonparametric estimation of the residual distribution. *Scan. J. Statist.* **28**, 549-567.
- Bhattacharya, P. K., Gastwirth, J. L. and Wright, A. L. (1982). Two modified Wilcoxon tests for symmetry about an unknown location parameter. *Biometrika* **69**, 377-382.
- Dette, H., Kusi-Appiah, S. and Neumeyer, N. (2002). Testing symmetry in nonparametric regression models. *Nonparametr. Statist.* **14**, 477-494.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.
- Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1926-1947.
- Hollander, M. (1988). Testing for symmetry. In *Encycl. Statist. Sci.* **9** (Edited by N. L. Johnson, S. Kotz), 211-216.
- Hušková, M. (1984). Hypothesis of symmetry. In *Handbook of Statistics*, **4**, Nonparametric methods, 63-78. North Holland, Amsterdam.
- Hyndman, R. J. and Yao, Q. (2002). Nonparametric estimation and symmetry tests for conditional density functions. *Nonparametr. Statist.* **14**, 259-278.
- Koul, H. L. and Staudte Jr., R. G. (1976). Power bounds for a Smirnov statistic in testing the hypothesis of symmetry. *Ann. Statist.* **4**, 924-935.
- Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*. Second edition. Springer, New York.
- Koziol, J. A. (1985). A note on testing symmetry with estimated parameters. *Statist. Probab. Lett.* **3**, 227-230.
- Liu, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **16**, 1696-1708.
- Nadaraya, É. A. (1964). On non-parametric estimates of density functions and regression curves. *J. Probab. Appl.* **10**, 186-190.
- Psaradakis, Z. (2003). A bootstrap test for symmetry of dependent data based on a Komogorov-Smirnov type statistic. *Comm. Statist., Ser. B, Simul. & Comput.* **32**, 113-126.
- Rice, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12**, 1215-1230.
- Schuster, E. F. and Berger, R. G. (1987). Using the bootstrap in testing symmetry versus asymmetry. *Comm. Statist., Ser. B, Simul. & Comput.* **16**, 69-84.
- Smirnov, N. V. (1947). Sur un critère de symétrie de la loi de distribution d'une variable aléatoire. *C. R. (Doklady) Acad. Sci. URSS (N.S.)* **56**, 11-14.
- Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence And Empirical Processes*. Springer, New York.
- Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman and Hall, London.
- Watson, G. S. (1964). Smooth Regression Analysis. *Sankhyā A* **26**, 359-372.
- Wu, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14**, 1261-1350.
- Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany.
E-mail: natalie.neumeyer@ruhr-uni-bochum.de
- Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany.
E-mail: holger.dette@ruhr-uni-bochum.de

(Received January 2005; accepted November 2005)