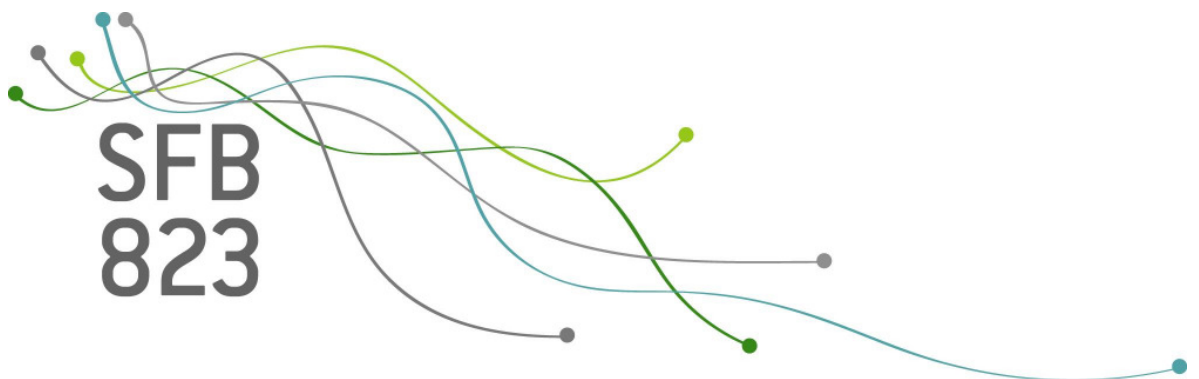# Testing symmetry of a nonparametric bivariate regression function

Melanie Birke, Holger Dette,
Kristin Stahljans

SFB 823

Discussion Paper

SFB
823

SFB
823

Testing symmetry of a nonparametric bivariate regression function

Melanie Birke[1], Holger Dette and Kristin Stahljans

Fakultät für Mathematik

Ruhr-Universität Bochum, Germany

July 10, 2009

Abstract

We propose a test for symmetry of a regression function with a bivariate predictor based on the L_2 distance between the original function and its reflection. This distance is estimated by kernel methods and it is shown that under the null hypothesis as well as under the alternative the test statistic is asymptotically normally distributed. The finite sample properties of a bootstrap version of this test are investigated by means of a simulation study and a possible application in detecting asymmetries in gray-scale images is discussed.

1 Introduction

Symmetric structures play an important role in many contexts in science, art and the real world (see e.g. Weil, 1952, Rosen, 1995 or Conway et al., 2008). Most parts of the human body are nearly symmetric, for example, the right hand is more or less symmetric to the left hand. By modeling this mathematically the problems reduce to symmetry of functions. For example, a symmetric grey-scale image is a symmetric function of the pixel location taking values between 0 and 1 or a distribution of some characteristics in a population modeled as random variables is symmetric if the underlying density is symmetric. So, most models of this kind can be reduced to symmetry of regression functions or densities and the problem of testing for symmetry is of particular interest for these functions. There exists already some work on tests for symmetry of densities. Hušková (1984) proposes a test for symmetry of the common distribution of two p-dimensional random vectors. Hollander (1988) gives an overview of several tests for symmetry of distribution or density functions. In Ahmad and Li (1997) a test for symmetry of a density function based on kernel methods is proposed. A test based on transformed empirical processes is given in Cabaña and

---

[1]Adress for correspondance: Melanie Birke, Ruhr-Universität Bochum, Fakultät für Mathematik, Universitätsstraße 150, 44780 Bochum, Germany, e-mail: melanie.birke@rub.de,Fon: +49/234/32–23286, Fax: +49/551/32–14559

Cabaña (2000), while Dette, Kusi-Appiah and Neumeyer (2002) develop a test for symmetry of the error distribution in regression models.

In the present paper we discuss the problem of testing for symmetry of a regression function, which is of similar importance as the problem of testing for symmetry of density functions. Our work is motivated by some applications in thermographic image analysis, where abnormalities such as malignancies, inflammation and infection cause localized increases in temperature which show as asymmetrical patterns in an infrared thermogram [see for example Gratt et. al (1995), Jones (1998), Kruganti and Qi (2002) among many others]. In such cases – besides the visual inspection of the images – statistical tests for the symmetry of the regression function are of interest to detect asymmetries at a controlled significance level. A typical application will be given in Section 4.2. Moreover, such tests can also be used generally in image analysis to test, whether an underlying object is symmetric with respect to some (known) axis or has some symmetric components. Additionally, the knowledge of symmetry is for example of advantage in image compression or reconstruction.

There already exist various methods in the literature addressing the issue of symmetry of objects or functions. They can generally be classified in two groups. One group consists of methods for determining a symmetry axis in images or functions or, more general, for fixing regions of (local) symmetries in two-dimensional objects. We mention here the work of Atallah (1985), Friedberg (1986), Marola (1989), Kiryathi and Gofman (1998), Loy and Zelinsky (2003) or Xiao et al. (2005) among others. In most of these articles it is assumed, that the regression function or image is observed without any stochastic errors. The second group is concerned with the question if, given an image or data, an object or regression function is symmetric with respect to some known or unknown axis. Recently, Bissantz, Holzmann and Pawlak (2009) propose tests for reflection symmetry and rotational invariance of two-dimensional regression functions or images based on Zernike polynomials. On the other hand statistical tests based on kernel regression estimates do not exist so far, especially in multivariate settings. On the one hand such methods are attractive because of their simplicity - they are usually based on a direct comparison of the functions estimating the images. On the other hand statistical tests based on kernel regression estimates also provide, beside the result of a statistical test, an estimate of a simply interpretable measure for the deviation from the null hypothesis. For example, if symmetry with respect to the y-axis has to be tested, the test proposed in this paper will be based on a kernel estimate of the quantity

$$I = \int_A \left(m(x,y) - m(-x,y) \right)^2 d(x,y) \,, \tag{1.1}$$

and the null hypothesis of symmetry is rejected for large values of the corresponding estimate. The L^2-distance I can be considered as a measure for asymmetry and further statistical inference could be performed using estimates of this quantity. Moreover, as pointed out by Rosenblatt (1975) or

2

Ghosh and Huang (1991) kernel based tests are usually powerful with respect to local alternatives of the form

$$m(x, y) = m(-x, y) + \beta_n \omega \left(\frac{x - c_1}{\eta_n}, \frac{y - c_2}{\eta_n} \right) \tag{1.2}$$

where $\beta_n, \eta_n \to 0$, $c_1, c_2 \in \mathbb{R}$ (see Remark 1 in Section 3).

In the present paper we propose a test for symmetry of a regression function with a two dimensional predictor based on kernel regression estimates. Although there are various characterizations of symmetry of a function, we investigate the fundamental one, which is motivated by several applications in thermographic image analysis, i.e.

$$H_0: \ m(-x, y) = m(x, y) \ \text{ for all } x, y \ . \tag{1.3}$$

Note that we assume that the axis of symmetry is known. If this is not the case, it has to be estimated first. Then, the general case can be easily traced back to the case investigated below. The remaining part of this paper is organized as follows. In Section 2 we define the test statistic and give a heuristic motivation. In Section 3 the asymptotic properties of the test statistic are investigated under the nullhypothesis and the alternative, and we also show the consistency of the test which rejects the null hypothesis (1.3) for large values of the estimate for the L^2-distance defined in (1.1). The applicability of the test is demonstrated in Section 4 where we show some simulation results and apply the procedure to an example from image analysis. The proofs are all deferred to the Appendix.

2 The test statistic and basic assumptions

We consider the nonparametric regression model with a bivariate predictor

$$Z_i = m(X_i, Y_i) + \sigma(X_i, Y_i)\varepsilon_i \qquad i = 1, \ldots, n \ . \tag{2.1}$$

Throughout this paper we restrict ourselves to the situation of symmetry with respect to the y-axis and assume that observations are available on the set $A = [-1, 1]^2$. The hypotheses are

H_0 : $m(x, y) = m(-x, y)$ for almost all $(x, y)^T \in [-1, 1]^2$,

H_1 : there exists a set $B \subset [-1, 1]^2$ with positive Lebesgue measure such that $m(x, y) \neq m(-x, y)$ for all $(x, y)^T \in B$.

All methods can be easily modified for testing symmetry with respect to the x axis or even with respect to any other axis. If the true regression function m is symmetric in the first component,

3

then this should also hold approximately for any consistent estimator $\hat{m}$ of m. Therefore we propose to estimate the distance I in (1.1) by

$$\hat{I}_n = \int_A [\hat{m}(x,y) - \hat{m}(-x,y)]^2 d(x,y) \tag{2.2}$$

and to reject the null hypothesis for large values of this test statistic, where $\hat{m}$ is an estimator of the regression function. If the null hypothesis of symmetry is true, the statistic $\hat{I}_n$ converges in probability to 0 as long as $\hat{m}$ is uniformly consistent, which intuitively explains why this procedure should yield a consistent test. In principle any consistent estimate could be used in this procedure but for the sake of brevity, we restrict ourselves to the Nadaraya-Watson and the local linear estimate in the following discussion. The Nadaraya-Watson estimate is defined by

$$\hat{m}^{NW}(x,y) = \frac{1}{na_1 a_2 \hat{f}_n(x,y)} k\Big(\frac{X_i - x}{a_1}, \frac{Y_i - y}{a_2}\Big) Z_i$$

with

$$\hat{f}_n(x,y) = \frac{1}{na_1 a_2} \sum_{i=1}^{n} k\Big(\frac{X_i - x}{a_1}, \frac{Y_i - y}{a_2}\Big),$$

and if this estimate is used in (2.2), the resulting statistic will be denoted by $\hat{I}_n^{NW}$ in the following discussion. Similarly the two-dimensional local linear estimator for a kernel k and a bandwidth $(a_1, a_2)^T$ is defined by

$$\hat{m}_{0,1}(x,y) = \Big[(S_{2,0}(x,y)S_{0,2}(x,y) - S_{0,2}^2(x,y))T_{0,0}(x,y)$$

$$+(S_{1,0}(x,y)S_{1,1}(x,y) - S_{0,1}(x,y)S_{2,0}(x,y))T_{0,1}(x,y)(S_{0,1}(x,y)S_{1,1}(x,y) - S_{0,2}(x,y)S_{1,0}(x,y))T_{1,0}(x,y) \Big]$$

$$\times \Big[2S_{0,1}(x,y)S_{1,0}(x,y)S_{1,1}(x,y) - S_{0,2}(x,y)S_{1,0}^2(x,y) - S_{0,0}(x,y)S_{1,1}^2(x,y) - S_{1,0}^2(x,y)S_{2,0}(x,y)$$

$$+S_{0,0}(x,y)S_{0,2}(x,y)S_{2,0}(x,y) \Big]^{-1},$$

where

$$S_{j,k}(x,y) = \sum_{i=1}^{n} k\Big(\frac{X_i - x}{a_1}, \frac{Y_i - y}{a_2}\Big)(X_i - x)^j (Y_i - y)^k$$

$$T_{j,k}(x,y) = \sum_{i=1}^{n} k\Big(\frac{X_i - x}{a_1}, \frac{Y_i - y}{a_2}\Big)(X_i - x)^j (Y_i - y)^k Z_i.$$

and the resulting statistic in (2.2) is denoted by

$$\hat{I}_{n,0,1} = \int_A (\hat{m}_{0,1}(x,y) - \hat{m}_{0,1}(-x,y))d(x,y)$$

4

throughout this paper. For the sake of transparency, we assume for the asymptotic considerations in the following that $a_1 = a_2 = h$ and $k(u, v) = k_1(u)k_2(v)$ where k_1 and k_2 are univariate kernels. However, all results presented here remain valid in the general case with an additional amount of notation. Obviously critical values can only be obtained by asymptotic considerations, which require knowledge of the asymptotic distribution of the statistic I_n under the null hypothesis and under the alternative. Roughly speaking, it will be shown in the next section that an appropriately scaled version of the statistic I_n is asymptotically normally distributed under the null hypothesis and the alternative, where different rates appear in both cases. We conclude this section with a statement of the necessary assumptions for these asymptotic results.

Assumption 1

1. The random variables (X_i, Y_i), $i = 1, \dots, n$ are independent and identically distributed with density f whose support contains the set

$$A^\delta = \{z \in \mathbb{R}^2 | \inf_{a \in A} ||z - a|| \leq \delta\}.$$

2. The error variables ε_i, $i = 1, \dots, n$ are independent and identically distributed with $\mathrm{E}[\varepsilon_i] = 0$, $\mathrm{E}[\varepsilon_i^2] = 1$ and existing finite fourth moment $\mu_4 = \mathrm{E}[\varepsilon_i]$ and independent of the sequence $\{(X_i, Y_i)\}_{i=1}^n$.

Assumption 2

1. The regression function m and the density f are two times continuously differentiable with uniformly continuous second derivative.

2. The function $\sigma : A^\delta \to \mathbb{R}^+$ of the standard deviation of Z_i is continuous.

Assumption 3

1. The symmetric kernels k_1 and k_2 are of order 2 and have compact support, say $[-1, 1]$.

2. The bandwidth h is proportional to the optimal bandwidth for two-dimensional regression estimation, that is $h = Cn^{-1/6}$.

3 Weak convergence under the null hypothesis and alternative

For the construction of an asymptotic level α test based on the statistic $\hat{I}_n$ in (2.2) we need the $(1 - \alpha)$-quantile of the asymptotic distribution of the test statistic $\hat{I}_n$ under the null hypothesis.

Our first theorem states that in this case the test statistic converges weakly with rate nh to a normal distribution. Throughout this paper $g_1 * g_2$ denotes the convolution of the functions g_1 and g_2, and for a function g the symbol $\partial_j g$ denotes the partial derivative of g with respect to the j-th component.

Theorem 1 *Let the assumptions 1 - 3 stated in Section 2 be fulfilled. Then*

$$\tau_n^{-1}\left(\hat{I}_n^{NW} - \mu_n\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{3.1}$$

where $\mu_n = \mathrm{E}[S_n]$,

$$\tau_n = \left(\frac{4}{n^2 h^2}\alpha_1 + \frac{16h^4}{n}\alpha_2\right)^{1/2}$$

$$S_n = \int_A \frac{[\hat{m}^{NW}(x,y) - \hat{m}^{NW}(-x,y)]^2 \hat{f}_n^2(x,y)\hat{f}_n^2(-x,y)}{\mathrm{E}[\hat{f}_n(x,y)]^2\,\mathrm{E}[\hat{f}_n(-x,y)]^2}d(x,y)$$

$$\alpha_1 = 2\int[(k*k)(u,v)]^2\,d(u,v)\int_A \sigma^2(x,y)\left[\frac{\sigma^2(x,y)}{f^2(x,y)} + \frac{\sigma^2(-x,y)}{f(x,y)f(-x,y)}\right]d(x,y),$$

$$\alpha_2 = \int_A \frac{\sigma^2(x,y)\gamma^2(x,y)}{f^3(x,y)f^2(-x,y)}d(x,y)$$

and

$$\gamma(x,y) = \int_{-1}^1 k_1(u)u^2 du[f(-x,y)\partial_1 f(x,y) - f(x,y)\partial_1 f(-x,y)]\partial_1 m(x,y)$$

$$+ \int_{-1}^1 k_2(v)v^2 dv[f(-x,y)\partial_2 f(x,y) - f(x,y)\partial_2 f(-x,y)]\partial_2 m(x,y),$$

The proof of this result is complicated and therefore deferred to the Appendix. If m and f are at least three times continuously differentiable, the bias term simplifies substantially, that is μ_n in (3.1) can be replaced by

$$\tilde{\mu}_n = \frac{1}{nh^2}\int_A\left(\frac{\sigma^2(x,y)}{f(x,y)} + \frac{\sigma^2(-x,y)}{f(-x,y)}\right)d(x,y)\int k^2(u,v)d(u,v) + h^4\int_A\left(\frac{\gamma(x,y)}{f(x,y)f(-x,y)}\right)^2 d(x,y).$$

If the alternative $H_1 : I \neq 0$ holds, then we observe a different asymptotic behavior of the test statistic T_n. For a statement of the precise result we introduce the following notation

$$T_n = \frac{1}{h^2}\int_A \frac{[Z_i - m(x,y)]k_1\left(\frac{X_i-x}{h}\right)k_2\left(\frac{Y_i-y}{h}\right)[m(x,y) - m(-x,y)]}{\mathrm{E}[\hat{f}_n(x,y)]}d(x,y)$$

$$\tau^2 = \int_A \frac{\sigma^2(u,v)(m(u,v) - m(-u,v))^2}{f(u,v)}d(u,v)$$

$$\tag{3.2}$$

6

Theorem 2 *Let the assumptions of Theorem 1 be fulfilled. If the regression function m is not symmetric then*

$$\sqrt{n}(\hat{I}_n^{NW} - I - 4b_n) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2), \tag{3.3}$$

where $b_n = \mathrm{E}[T_n]$.

If the regression and density function are at least three times continuously differentiable, the bias term b_n in (3.3) can be replaced by

$$
\begin{aligned}
\tilde{b}_n &= h^2 \Big(\kappa_2(k_1)(\partial_1^2 m(x,y)f(x,y) - \partial_1 m(x,y)\partial_1 f(x,y)) + \kappa_2(k_2)(\partial_2^2 m(x,y)f(x,y) - \partial_2 m(x,y)\partial_2 f(x,y)) \Big) \\
&\quad \times \frac{m(x,y) - m(-x,y)}{f(x,y)}.
\end{aligned}
$$

From Theorem 1 and 2 the asymptotic distribution of the test statistic under the null hypothesis and the alternative the following result can easily be obtained.

Corollary 1 *Under the assumptions of Theorem 2 the test which rejects the null hypothesis of symmetry for large values of the statistic $\hat{I}_n^{NW}$ is consistent.*

Remark 1 It can be shown by similar arguments as given in the Appendix, that for alternatives of the form (1.2) with a square integrable twice continuously differentiable function ω and

$$\beta_n = \frac{1}{\sqrt{nh^{1+2\delta}}} \qquad \eta_n = h^\delta$$

$(\delta > 1/2)$ the statistic on the left hand side of (3.1) is also asymptotically normally distributed with variance 1 and mean proportional to $\int w^2(x)dx$. Because the size of the integral

$$\int_{-\infty}^{c_1} \int_{-\infty}^{c_2} \beta_n w((x - c_1)\eta_n^{-1}, (y - c_2)\eta_n^{-1})dydx$$

is of order

$$\beta_n \eta_n^2 = n^{-1/2} h^{(1+2\delta)/2} h^{2\delta} = n^{-1/2} h^{\delta-1/2} = n^{-1/2} n^{-(\delta-1/2)/6},$$

the test proposed in this paper has greater power against such local alternatives than tests based on empirical processes [see Rosenblatt (1975) or Ghosh and Huang (1991) for more details].

Remark 2
(A) By Theorem 1 an asymptotic and consistent level α test for the hypothesis of symmetry is obtained by rejecting the null hypothesis if $\hat{I}_n^{NW} - \hat{\mu}_n > \hat{\tau}_n z_{1-\alpha}$, where $\hat{\mu}_n, \hat{\tau}_n$ are appropriate

(consistent) estimates of μ_n and τ_n, respectively, and $z_{1-\alpha}$ denotes the $(1-\alpha)$ quantile of the standard normal distribution. It now follows from Theorem 2 that the power of this test can be approximated by

$$P(H_0 \text{ rejected} \mid H_1 \text{ is true}) \approx 1 - \Phi\left(-\sqrt{n}(I + 4b_n)/\tau + \sqrt{n}(\mu_n + \tau_n z_{1-\alpha})/\tau\right). \qquad (3.4)$$

where μ_n, τ_n and b_n are defined in Theorem 1 and 2, respectively. This formula provides also information about the behavior of the power function, which depends (asymptotically) on the particular alternative only through the three quantities b_n, I and τ^2.

(B) Note that the quantity I defined in (1.1) can be interpreted as a measure of symmetry of the regression function. From Theorem 2 we obtain under the alternative

$$\hat{I}_n^{NW} - 4b_n + \frac{\tilde{\tau}_n z_{1-\alpha}}{\sqrt{n}}$$

as an upper (asymptotic) $(1-\alpha)$ confidence bound for the parameter I, where $\tilde{\tau}_n^2$ is an appropriate (consistent) estimator of the asymptotic variance given in Theorem 2.

(C) A further important application of Theorem 2 arises from the fact that in practice - in particular when analyzing thermal images - perfect symmetry will usually never be observed. The more realistic question in this context is, if there exists approximately symmetry. Therefore we propose to investigate the so called *precise hypotheses* [see Berger and Delampady (1987)]

$$H_0 : I > \varepsilon \quad \text{versus} \quad H_1 : I \leq \varepsilon, \qquad (3.5)$$

where I is the measure defined by (1.1) and $\varepsilon > 0$ is a prespecified constant for which the experimenter agrees to accept the image as symmetric. An asymptotic α-level test for the hypothesis (3.5) is obtained by rejecting the null hypothesis, whenever

$$\sqrt{n}(\hat{I}_n^{NW} - \varepsilon - 4b_n) < \hat{\tau}_n z_{1-\alpha}.$$

We conclude this section with the corresponding statements for the local linear case. Note that the discussion presented in Remark 1 and 2 is also valid for this type of estimate.

Theorem 3 *If the assumptions of Theorem 1 are satisfied, we have under the null hypothesis*

$$nh\alpha_{0,1}^{-1/2}(\hat{I}_{n,0,1} - \mathbb{E}_{\mathcal{X}}[\hat{I}_{n,0,1}]) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \qquad (3.6)$$

where $\mathbb{E}_{\mathcal{X}}$ denotes the conditional expectation with respect to $\mathcal{X} = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ and

$$\alpha_{0,1} = 2 \int (k * k)^2(u,v)d(u,v) \int_A \sigma^2(x,y)\left[\frac{\sigma^2(x,y)}{f^2(x,y)} + \frac{\sigma^2(-x,y)}{f(x,y)f(-x,y)}\right]d(x,y).$$

8

Under the alternative $I > 0$ it follows that

$$\sqrt{n}\tau^{-1/2}(\hat{I}_{n,0,1} - \mathbb{E}_{\mathcal{X}}[\hat{I}_{n,0,1}]) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{3.7}$$

where τ is defined in Theorem 2. If m is at least three times continuously differentiable, the centering term $\mathbb{E}_{\mathcal{X}}[\hat{I}_{n,0,1}]$ in (3.6) and (3.7) can be replaced by

$$\frac{1}{nh^2}\int_A \left(\frac{\sigma^2(x,y)}{f(x,y)} + \frac{\sigma^2(-x,y)}{f(-x,y)}\right)d(x,y)\int k^2(u,v)d(u,v)$$

and

$$I + h^2\int_A (\kappa_2(k_1)\partial_1^2 m(x,y) + \kappa_2(k_2)\partial_2^2 m(x,y))(m(x,y) - m(-x,y))d(x,y),$$

respectively.

4   Finite sample properties and a data example

In this section we study the finite sample behavior of the test based on the L^2-distance by means of a simulation study and an application.

4.1   Simulated examples

Although Theorem 1 and 3 specify the asymptotic behavior of the statistics $\hat{I}_n^{NW}$ and $\hat{I}_{n,0,1}$, we have so far no information about the size and power of the test in finite samples. One typical problem in the application of tests based on an L^2-distance between nonparametric regression estimates is, that the asymptotic distribution still depends on certain parameters of the data generating process, which are often difficult to estimate. Moreover, even if the asymptotic variance and bias in Theorem 1 can be estimated the approximation of the nominal level is usually rather poor [see e.g. Fan and Linton (2003)]. Hence, we propose to use a wild bootstrap version of the test which is often used in heteroscedastic regression models [see e.g. Wu (1986) or Härdle and Mammen (1993)]. For this purpose we form a bootstrap sample $(X_1, Y_1, Z_1^*), \ldots, (X_n, Y_n, Z_n^*)$ from $(X_1, Y_1, Z_1) \ldots, (X_n, Y_n, Z_n)$ with

$$Z_i^* = \hat{m}_S(X_i, Y_i) + C_i\hat{\varepsilon}_i, \ i = 1, \ldots, n.$$

In this definition we use the symmetrized regression estimator

$$\hat{m}_S(x,y) = \frac{\hat{m}(x,y) + \hat{m}(-x,y)}{2},$$

the estimated residuals

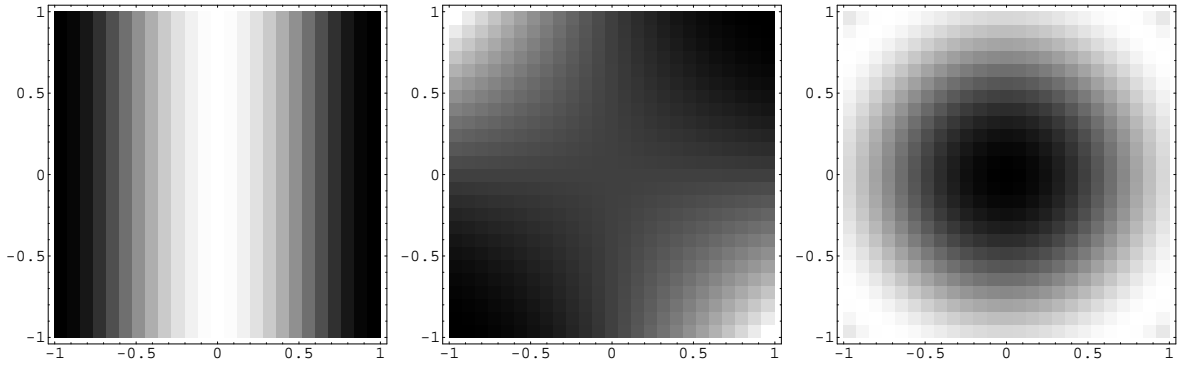$$\hat{\varepsilon}_i = Z_i - \hat{m}(X_i, Y_i), \ i = 1, \ldots, n$$

Figure 1: *Contour plots of the regression functions m_1, m_2 and m_3 with $a = 0$ (from left to right).*

and independent identically distributed random variables C_i which are independent of the sample $(X_1, Y_1, Z_1), \ldots, (X_n, Y_n, Z_n)$ and are drawn from the distribution which puts mass $(\sqrt{5} + 1)/2\sqrt{5}$ at the point $(1 - \sqrt{5})/2$ and mass $(\sqrt{5} - 1)/2\sqrt{5}$ at the point $(1 + \sqrt{5})/2$. From this bootstrap sample we calculate the bootstrap version $\hat{I}_n^*$ of the test statistic. After B repetitions of the bootstrap procedure we obtain a sample

$$\hat{I}_{n,1}^*, \ldots, \hat{I}_{n,B}^*$$

of the test statistics, which can be used to estimate the $(1-\alpha)$-quantile of the distribution under the null hypothesis by the $\lfloor (1 - \alpha)B \rfloor$-th order statistic, say $\hat{I}_{n,(\lfloor (1-\alpha)B \rfloor)}^*$. Therefore the null hypothesis of symmetry is rejected if

$$\hat{I}_n > \hat{I}_{n,(\lfloor (1-\alpha)B \rfloor)}^*. \tag{4.1}$$

The consistency of this bootstrap procedure follows by similar arguments as presented in the proof of Theorem 1 (see the Appendix) along the lines given in Härdle and Mammen (1993) or Dette and Neumeyer (2001), who prove consistency of the wild bootstrap in the context of testing parametric assumptions and comparing regression curves.

We restrict the investigations of the finite sample properties to the local linear estimator because of its advantages with respect to boundary effects. We use least squares cross validation to determine the bandwidth for the estimation of the regression function in the test statistic $\hat{I}_{n,0,1}$ and the errors $\hat{\varepsilon}_i$. For the symmetric estimate $\hat{m}_s$ required in the construction of the bootstrap sample we use a slightly larger bandwidth to ensure a correct estimation of the bias.

For our simulation study we consider the following regression functions

$$
\begin{aligned}
m_1(x, y) &= (1 - x^2)^2 \\
m_2(x, y) &= (1 - xy)^2 \\
m_{3,a}(x, y) &= \sin((x - a)^2 + y^2), \ a = 0, 0.005, 0.0075, 0.01.
\end{aligned}
$$

10

Table 1: *Size and power of the test (4.1) for the symmetry of the regression functions m_1 and m_2 estimated from 500 simulation runs for $\alpha = 0.05$ and $\alpha = 0.1$.*

	(a) Size for m_1					(b) Power for m_2			
		$n = 50$		$n = 100$			$n = 50$		$n = 100$
α	0.05	0.1	0.05	0.1	α	0.05	0.1	0.05	0.1
σ					σ				
0.05	0.073	0.107	0.05	0.1	0.05	1	1	1	1
0.1	0.056	0.11	0.054	0.094	0.1	1	1	1	1

The first regression function is symmetric and the second one is not symmetric. The third one is symmetric with respect to the y axis for $a = 0$. For $a = 0.005$, 0.0075 and 0.01, it is shifted to the right and therefore not symmetric with respect to the y axis. Contour plots of these functions are presented in Figure 1. We consider a uniform design on $[-1, 1]^2$, normally distributed errors with different standard deviations σ and different α. The sample sizes are $n = 50$ and 100 while the number of bootstrap replications in each simulation run is B=200. The size and power of the test is estimated from 500 simulation runs. The results for the three different regression functions are presented in Tables 1 and 2 for the regression functions m_1 and m_2, respectively. We observe that for the regression function m_1 the nominal level is approximated very well in both cases $\sigma = 0.05$ and $\sigma = 0.1$. For the function m_2, which is not symmetric with respect to the y-axis, the simulated power is 100% in all cases. For a more detailed investigation of the behaviour of the test under alternatives we consider the function m_3 for different values of a. The simulation results are presented in Table 2. We see that a small shift ($a = 0.005$) is harder to detect than a larger shift ($a = 0.0075$ or $a = 0.01$), especially for larger standard deviations. But in all cases where the regression function m_3 is not symmetric to the y-axis, the simulated power is considerably higher than the size of the test. The simulations also show (as expected), that the power is smaller for a higher standard deviation of the errors.

4.2   Testing gray-scale images for symmetry

Gray-scale images can be understood as a two dimensional regression problem. Every pixel is defined by its horizontal position x_i and its vertical position y_i in the picture and has a value $Z_i \in [0, 1]$ defining its scale of gray. Without loss of generality we may assume, that the coordinates (x_i, y_i), $i = 1, \ldots, n$ of every pixel are located in the set $[-1, 1]^2$ so that the origin of the coordinate system

Table 2: *Size and power of the test (4.1) for the symmetry of the regression function $m_{3,a}$ estimated from 500 simulation runs for $\alpha = 0.05$ and $\alpha = 0.1$.*

	α	$n=50$		$n=100$			α	$n=50$		$n=100$	
		0.05	0.1	0.05	0.1			0.05	0.1	0.05	0.1
σ							σ				
		$a=0$						$a=0.0075$			
0.05		0.066	0.1	0.05	0.1		0.05	0.952	0.976	1	1
0.1		0.052	0.08	0.042	0.078		0.1	0.4	0.538	0.958	0.976
		$a=0.005$						$a=0.01$			
0.05		0.602	0.708	0.994	0.998		0.05	0.998	1	1	1
0.1		0.16	0.25	0.616	0.732		0.1	0.68	0.78	1	1

is located in the center of the image. In many cases the images are disturbed by error terms ε_i. Therefore the image of a symmetric object usually does not show perfect symmetry. Only knowing the image of an object, this evokes the question if the real object can be symmetric with respect to the vertical or horizontal axis.

We now illustrate the application of the new test in the context of analyzing thermographic images, which are used in medicine to detect diseases, especially inflammations in parts of the human body [see for example Gratt et. al. (1995), Jones (1998) or Kuruganti and Qi (2002)]. While most parts of the human body are nearly symmetric concerning the thermographical characteristics (e.g. left and right hand/arm/leg etc.) an inflammation e.g. of a finger violates this symmetry between the left and right hand. In Figure 2 we show two thermographic images of a right hand and its reflection. The image in Figure 2(a) exhibits symmetry. The image in 2(b) has a small modification which violates the symmetry. We now use the new bootstrap test to check the symmetry with respect to the centered vertical axis in both cases. For the image in Figure 2(a), the p-value of the test is 0.232 and therefore the null hypothesis is not rejected at significance level $\alpha = 0.05$. On the other hand for the image in Figure 2(b) the p-value is 0 and therefore the test rejects the null hypothesis. Thus the asymmetry is clearly detected.
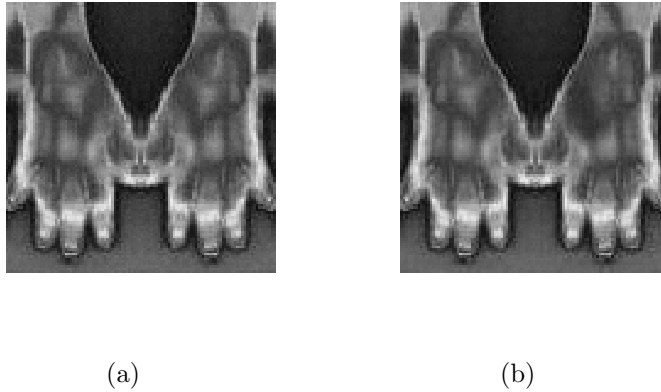
(a) (b)

Figure 2: *Thermographic image of a hand: (a) duplicated and mirrowed image of a right hand (perfect symmetry) disturbed with errors. (b) duplicated and mirrowed image of a right hand with modification in the right part of the image (small deviation from symmetry) disturbed with errors.*

5   Concluding Remarks

In this paper we have developed a consistent test for symmetry of a bivariate regression function with respect to a known axis using kernel methods. As natural application we showed an example from image analysis with a more or less continuous change of grey scales. This assumption is not always plausible for all images because there are often edges where colours change abruptly. Then the standard kernel estimates do not provide useful estimates and modifications like e.g. in Gijbels, Lambert and Qiu (2007), Hall, Qiu and Rau (2008) or wavelet estimators might provide better results.

Image analysis can always be seen as inverse problem and testing for symmetry of the true regression function is then of special interest if the point spread function is not symmetric. Topics of current research are therefore testing for symmetry with respect to an unknown axis and in inverse problems.

References

I.A. Ahmad (1982). *Nonparametric estimation of the location and scale parameters based on density estimation.* Annals Inst. Statist. Math. 34, 39-53.

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

M.J. Atallah (1985). *On symmetry detection.* IEEE Trans. Comput. C-34, 663-666.

L. Baringhaus and N. Henze (1992). *A characterization of and new consistent tests of symmetry.* Commun. Statist. 21, 1111-1125.

J.O. Berger and M. Delampady (1987). Testing precise hypotheses. *Statist. Sci.*, 2, 317-352.

M. Birke (2008). *Central limit theorems for the integrated squared error of derivative estimators.* Statist. Probab. Lett. 78, 1903–1913.

Bissantz, Holzmann and Pawlak (2009). *Testing for Image Symmetries – with Application to Confocal Microscopy.* IEEE Trans. Information Theory, to appear.

A. Cabaña and M. Cabaña (2000). *Tests of symmetry based on transformed empirical processes.* Canad. J. Statist. 28, 829-839.

H. Dette and N. Neumeyer (2001). *Nonparametric analysis of covariance.* Annals of Statistics 29, 1361-1400.

H. Dette, S. Kusi-Appiah and N. Neumeyer (2002). *Testing symmetry in nonparametric regression models.* Nonparam. Statist. 14 (5), 477-494.

G.K. Eagleson (1975). *Martingale Convergence to mixtures of infinitely divisible laws.* Ann. Probab. 3, 557-662.

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

Y. Fan, O. Linton (2003). Some higher order theory for a consistent nonparametric model specification test. Journal of Statistical Planning and Inference 109, 125-154.

S.A. Friedberg (1986). *Finding Axes of a skewed symmetry.* Comput. Vision Graphics Image Process 32, pp 138-155.

B.K. Ghosh and W.-M. Huang (1991). *The Power and Optimal Kernel of the Bickel-Rosenblatt Test for Goodness of Fit,* Annals of Statistics 19, 999-1009.

I. Gijbels, A. Lambert and P. Qiu (2007). *Jump-preserving regression and smoothing using local linear fitting: A compromise.* Ann. Inst. Stat. Math. 59, 235-272

E. Giné and A. Guillou (2002). *Rates of strong uniform consistency for multivariate kernel density estimators.* Ann. Inst. Poincaré 6, 907-921.

B.M. Gratt, V. Shetty, M. Saiar, E.A. Sickles (1995) *Electronic thermography for the assessment of inferior alveolar nerve deficit.* Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod 80, 153-160.

W. Härdle and E. Mammen (1993). *Testing parametric versus nonparametric regression.* Annals of Statistics 21, 1926-1947.

P. Hall (1984). *Integrated square error properties of kernel estimators of regression functions.* Ann. Statist. 12, No.1, 214-260.

P. Hall, P. Qiu and C. Rau (2008). *Tracking edges, corners and vertices in an image.* Scand. J. Statist. 35, 1-17.

P. Hall and C.C Heyde (1980). *Martingale Limit Theory and its Application.* Academic. New York, pp. 52-53.

M. Hollander (1988). *Testing for symmetry.* N.L. Johnson and S. Kotz (Eds.), Encyl. Statist. Sci. 9, 211-216.

M. Hušková (1984). *Hypothesis of Symmetry. In: Handbook of Statistics.* Nonparametric Methods, 4, North Holland, Amsterdam, pp. 63-78.

B.F. Jones (1998). *A reappraisal of the use of infrared thermal image analysis in medicine.* IEEE Transactions on Medical Imaging 17, 1019-1027.

N. Kiryathi and Y. Gofman (1998). *Detecting symmtery in grey level images: the global optimization approach.* Int. J. of Computer Vision 29, pp. 29-45.

P.T. Kuruganti and H. Qi (2002). *Asymmetry analysis in breast cancer detection using thermal infrared images.* Proc. of the 2nd Joint EMBS-BMES Conference, Vol. 2, 1129-1130, Houston.

G. Loy and A. Zelinsky (2003). *Fast radial symmetry for detecting points of interest.* IEEE Trans. Pattern Anal. Mach. Intell. 25, 595-973.

G. Marola (1989). *On the detection of the axes of symmtery of symmetric and almost symmetric planar images.* IEEE Trans. Pattern Anal. Mach. Intell., 11, 104-108.

E.A. Nadaraya (1964). *On estimating regression.* Theory Probab. Appl. 9, 141-142.

M. Rosenblatt (1975). *A Quadratic Measure of Deviation of Two-Dimensional Density Estimates and A Test of Independence.* Annals of Statistics 3, 1-14.

D. Ruppert and M.P. Wand (1994). *Multivariate locally weighted least squares regression,* Ann. Statist. 22, 1346-1370.

K. Stahljans (2007). *Symmetrietests in multivariaten nichtparametrischen Modellen,* Diploma thesis, Ruhr-Universität Bochum (in German)

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

G.S. Watson (1964). *Smooth regression analysis.* Sankhyā Ser. A 26, 359-372.

C.F.J. Wu (1986). *Jackknife, bootstrap and other resampling methods in regression analysis.* Ann. Statist. 14, 1261-1295.

Z. Xiao, Z. Hou, C. Miao & J. Wang (2005). *Using phase information for symmetry detection.* Pattern Recognition Lett. 26, 1985-1994.

A   Appendix: Proofs

For the sake of brevity we restrict ourselves to a proof of Theorem 1 and 2. The results in Theorem 3 can be derived by similar arguments. In the following discussion we use the notation $\hat{I}_n = \hat{I}_n^{NW}$.

A.1   Proof of Theorem 1

The proof of this theorem is similar to that of Theorem 4 in Hall (1984), who investigated the asymptotic properties of the L^2-distance between the Nadaraya-Watson estimate and the true regression function. In a first step we decompose $\hat{I}_n$ into

$$\hat{I}_n = I_n + \tilde{I}_n \tag{4.2}$$

with

$$I_n = \int_A \frac{[\kappa_n(x,y)f(-x,y) - \kappa_n(-x,y)f(x,y)]^2}{\hat{f}_n^2(x,y)\hat{f}_n^2(-x,y)} d(x,y), \tag{4.3}$$

$$\tilde{I}_n = 2\int_A \frac{1}{\hat{f}_n^2(x,y)\hat{f}_n^2(-x,y)} \Big\{ \kappa_n^2(x,y)[\hat{f}_n^2(-x,y) - f^2(-x,y)]$$

$$- \kappa_n(x,y)\kappa_n(-x,y)[\hat{f}_n(x,y)\hat{f}_n(-x,y) - f(x,y)f(-x,y)] \Big\} d(x,y), \tag{4.4}$$

where $\kappa_n(x, y) = \hat{m}(x, y) - m(x, y)\hat{f}_n(x, y)$ and show, that $\tilde{I}_n$ gives the asymptotic distribution. In a second step we prove that T_n is asymptotically negligible. The details are stated in the following two theorems.

Theorem 4 *Under the assumptions of Theorem 1 we have*

$$\left(\frac{4}{n^2 h^2}\alpha_1 + \frac{16h^4}{n}\alpha_2\right)^{-1/2} \left(I_n - \mathrm{E}[S_n^{[I]}]\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

where

$$S_n^{[I]} = \int_A \frac{\{\kappa_n(x, y)f(-x, y) - \kappa_n(-x, y)f(x, y)\}^2}{\mathrm{E}[\hat{f}_n(x, y)]^2 \, \mathrm{E}[\hat{f}_n(-x, y)]^2} d(x, y).$$

Theorem 5 *Under the assumptions of Theorem 1 we have*

$$\left(\frac{4}{n^2 h^2}\alpha_1 + \frac{16h^4}{n}\alpha_2\right)^{-1/2} (\tilde{I}_n - \mathrm{E}[\tilde{S}_n]) = o_p(1),$$

where

$$\tilde{S}_n = 2\int_A \frac{\kappa_n^2(x, y)[\hat{f}_n^2(-x, y) - f^2(-x, y)] - \kappa_n(x, y)\kappa_n(-x, y)[\hat{f}_n(x, y)\hat{f}_n(-x, y) - f(x, y)f(-x, y)]}{\mathrm{E}[\hat{f}_n(x, y)]^2 \, \mathrm{E}[\hat{f}_n^2(-x, y)]^2} d(x, y).$$

Proof of Theorem 4 Let us first consider a slightly modified L_2-distance

$$\bar{I}_n := \bar{I}_n(v_n) := \int_A [\kappa_n(x, y)f(-x, y) - \kappa_n(-x, y)f(x, y)]^2 v_n(x, y)d(x, y), \tag{4.5}$$

where v_n is a stochastic weight function which is symmetric and positive and converges in probability to a bounded deterministic function v (finally we choose $v_n(x) = 1/\hat{f}_n^2(x, y)\hat{f}_n^2(-x, y)$). We introduce the notation

$$\kappa_n(x, y)f(-x, y) - \kappa_n(-x, y)f(x, y) = \frac{1}{nh^2}\sum_{i=1}^n H(X_i, Y_i, x, y)[Z_i - m(x, y)]$$

with

$$H(X_i, Y_i, x, y) = \left\{k_1\left(\frac{X_i - x}{h}\right)f(-x, y) - k_1\left(\frac{X_i + x}{h}\right)f(x, y)\right\}k_2\left(\frac{Y_i - y}{h}\right).$$

Substituting this in the definition of $\bar{I}_n$ yields the decomposition

$$\bar{I}_n = I_{n1} + I_{n2} + I_{n3} + I_{n4}$$

where

$$I_{n1} = \frac{1}{n^2 h^4} \sum_{i=1}^{n} [Z_i - m(X_i, Y_i)]^2 \int_A H^2(X_i, Y_i, x, y) v_n(x, y) d(x, y)$$

$$I_{n2} = \frac{2}{n^2 h^4} \sum_{1 \le i < j \le n} [Z_i - m(X_i, Y_i)][Z_j - m(X_j, Y_j)] \int_A H(X_i, Y_i, x, y) H(X_j, Y_j, x, y) v_n(x, y) d(x, y)$$

$$I_{n3} = \frac{2}{nh^2} \sum_{i=1}^{n} [Z_i - m(X_i, Y_i)] \int_A H(X_i, Y_i, x, y) g_n(x, y) v_n(x, y) dx dy$$

$$I_{n4} = \int_A g_n^2(x, y) v_n(x, y) dx dy$$

and

$$g_n(x, y) = \frac{1}{nh^2} \sum_{i=1}^{n} [m(X_i, Y_i) - m(x, y)] H(X_i, Y_i, x, y).$$

In this decomposition, the function H takes the role of the kernel K in Hall (1984). The main difference in the estimation of I_{n1} to I_{n4} here, compared to Hall (1984) is, that H is not necessarily positive. In what follows, we will show that I_{n2} and I_{n3} give the asymptotic normal distribution while I_{n1} and I_{n4} form the bias of the test statistic. We begin with the consideration of I_{n1}. Similar steps as in Birke (2008) lead to

$$\left(\frac{nh^2}{2} \right)^4 \mathbb{E}_{\mathcal{X}} \left[\{I_{n1} - \mathbb{E}_{\mathcal{X}}[I_{n1}]\}^2 \right] = O_p(nh^4)$$

and applying the Markov inequality gives for any sequence $\lambda_n \to \infty$ for $n \to \infty$

$$P(|I_{n1} - \mathbb{E}_{\mathcal{X}}[I_{n1}]| > \frac{\lambda_n}{n^{3/2}h^2} | X_1, Y_1, \ldots, X_n, Y_n) \le \frac{1}{\lambda_n^2} O_p \left(\frac{n^3 h^4 nh^4}{n^4 h^8} \right) \xrightarrow[n \to \infty]{} 0.$$

Therefore we have

$$I_{n1} = \mathbb{E}_{\mathcal{X}}[I_{n1}] + O_p \left(\frac{1}{n^{3/2}h^2} \right) = \mathbb{E}_{\mathcal{X}}[I_{n1}] + o_p \left(\frac{1}{nh} \right) \tag{4.6}$$

with

$$\mathbb{E}_{\mathcal{X}}[I_{n1}] = \frac{1}{n^2 h^4} \sum_{i=1}^{n} \sigma^2(X_i, Y_i) \int_A H^2(X_i, Y_i, x, y) v_n(x, y) d(x, y). \tag{4.7}$$

Let us now turn to the random variable I_{n2}. We introduce the notations

$$W_{nij} = \int_A H(X_i, Y_i, x, y) H(X_j, Y_j, x, y) v_n(x, y) d(x, y)$$

$$= 2 \int_A \tilde{H}(X_i, Y_i, x, y) v_n(x, y) d(x, y), \tag{4.8}$$

$$\tilde{W}_{nij} = 2 \int_A \tilde{H}(X_i, Y_i, x, y) v(x, y) d(x, y), \tag{4.9}$$

18

where

$$\tilde{H}(X_i, Y_i, x, y) = \left[k_1\left(\frac{X_i - x}{h}\right) f(-x, y) - k_1\left(\frac{X_i + x}{h}\right) f(x, y) \right]$$
$$\times k_2\left(\frac{Y_i - y}{h}\right) k_1\left(\frac{X_j - x}{h}\right) f(-x, y) k_2\left(\frac{Y_j - y}{h}\right)$$

and

$$Z_{ni} = \{Z_i - m(X_i, Y_i)\} \sum_{j=1}^{i-1} \{Z_j - m(X_j, Y_j)\} W_{nij}$$

and apply the central limit theorem for martingale difference arrays (see Hall and Heyde, 1980) to the statistic

$$\frac{n^2 h^4}{2} I_{n2} = \sum_{i=2}^{n} Z_{ni}.$$

The conditional variance yields

$$V_n^2 = \sum_{i=2}^{n} \sigma^2(X_i, Y_i) \sum_{j=1}^{i-1} \{Z_j - m(X_j, Y_j)\}^2 \tilde{W}_{nij}^2 (1 + o_p(1)),$$

which converges to

$$\bar{\alpha}_1 = 2 \int [(k * k)(u, v)]^2 \, d(u, v) \int_A (\sigma^2 f^2 v^2)(x, y) f^2(-x, y) [\sigma^2(x, y) f(-x, y) + \sigma^2(-x, y) f(x, y)] d(x, y)$$

and a standard but tedious calculation shows that the Lindeberg condition

$$\frac{1}{n^2 h^6} \sum_{i=2}^{n} \mathbb{E}_{\mathcal{X}}[Z_{ni}^2 1\{|Z_{ni}| > \varepsilon n h^3\}] \xrightarrow{P} 0$$

is fulfilled (see Stahljans, 2007, for details). Now the central limit theorem yields

$$\frac{nh}{2} \bar{\alpha}_1^{-1/2} I_{n2} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

In addition it can be shown, that I_{n2} is asymptotically independent of any sequence of events A_n contained in the σ-algebra $\mathcal{F}_{n0}$ generated by $X_1, \ldots, X_n$.

The term I_{n3} can be decomposed into

$$I_{n3} = 4J_{n1} + 4J_{n2} \tag{4.10}$$

with

$$J_{nj} = \frac{1}{nh^2} \sum_{i=1}^{n} \{Z_i - m(X_i, Y_i)\} \int_A k_1\left(\frac{X_i - x}{h}\right) k_2\left(\frac{Y_i - y}{h}\right) a_j(x, y) v_n(x, y) f(-x, y) d(x, y)$$

19

and $a_1(x, y) = \gamma_n(x, y) = \mathrm{E}[g_n(x, y)]$, $a_2(x, y) = g_n(x, y) - \gamma_n(x, y)$. We will show that J_{n1} is asymptotically normal and that J_{n2} is asymptotically negligible. To this end we define

$$Z_{ni} := \{Z_i - m(X_i, Y_i)\}\tilde{Z}_{ni},$$

where

$$\tilde{Z}_{ni} = \int_A k_1\left(\frac{X_i - x}{h}\right) k_2\left(\frac{Y_i - y}{h}\right) f(-x, y)\gamma_n(x, y)v_n(x, y)d(x, y),$$

and write

$$J_{n1} = \frac{1}{nh^2}\sum_{i=1}^{n} Z_{ni}. \tag{4.11}$$

Therefore, conditionally on $\mathcal{F}_{n0}$, J_{n1} is a centered sum of independent random variables and its asymptotic normality follows from the following lemma.

Lemma 1 *Under the assumptions of Theorem 1 we have*

$$\frac{1}{nh^8}\mathbb{E}_{\mathcal{X}}[\{\sum_{i=1}^{n} Z_{ni}\}^2] \xrightarrow{P} \bar{\alpha}_2,$$

where

$$\bar{\alpha}_2 = \int_A (\sigma^2\gamma^2 v^2 f)(x, y)f^2(-x, y)d(x, y).$$

Moreover, for all $\varepsilon > 0$ the Lindeberg condition

$$\frac{1}{nh^8}\sum_{i=1}^{n} \mathbb{E}_{\mathcal{X}}[Z_{ni}^2 1\{|Z_{ni}| > \varepsilon\sqrt{n}h^4\}] \xrightarrow{P} 0$$

is fulfilled.

The proof of this Lemma follows the structure of the corresponding one in Hall (1984) where we replace the kernel K by the function H defined above. For details see again Stahljans (2007). Again the limiting random variable in (4.11) is independent of $\mathcal{F}_{n0}$. This yields, together with the results for I_{n2}, that

$$\left(\frac{4}{n^2h^2}\bar{\alpha}_1 + \frac{16h^4}{n}\bar{\alpha}_2\right)^{-1/2} (I_{n2} + J_{n1}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Similar methods as in Birke (2008) yield for the term J_{n2} in (4.10).

$$J_{n2} = o_p\left(\frac{1}{nh}\right).$$

So far, we have used the more general weight function $v_n(x, y)$. We now substitute it by $v_n(x, y) = 1/[\hat{f}_n^2(x, y)\hat{f}_n^2(-x, y)]$ and (4.5) yields the statistic $\tilde{I}_n$. This weight function is positive, symmetric

and converges in probability to $1/[f^2(x,y)f^2(-x,y)]$. Below, we give a representation of the remaining stochastic terms $\mathbb{E}_{\mathcal{X}}[I_{n1}]$ and I_{n4} whose dominating parts form the bias of $\tilde{I}_n$. To overcome the problem of the stochastic part in the denominator of v_n we use the following two Taylor expansions of $1/x^2$ in a neighbourhood of the point $\mathrm{E}[\hat{f}_n(x,y)]\,\mathrm{E}[\hat{f}_n(-x,y)]$, that is

$$
\begin{aligned}
(\hat{f}_n(x,y)\hat{f}_n(-x,y))^{-2} &= (\mathrm{E}[\hat{f}_n(x,y)]\,\mathrm{E}[\hat{f}_n(-x,y)])^{-2} \\
&\quad + O_P(1)|\hat{f}_n(x,y)\hat{f}_n(-x,y) - \mathrm{E}[\hat{f}_n(x,y)]\,\mathrm{E}[\hat{f}_n(-x,y)]| \quad (4.12) \\
(\hat{f}_n(x,y)\hat{f}_n(-x,y))^{-2} &= (\mathrm{E}[\hat{f}_n(x,y)]\,\mathrm{E}[\hat{f}_n(-x,y)])^{-2} \\
&\quad -2\frac{\hat{f}_n(x,y)\hat{f}_n(-x,y) - \mathrm{E}[\hat{f}_n(x,y)]\,\mathrm{E}[\hat{f}_n(-x,y)]}{(\mathrm{E}[\hat{f}_n(x,y)]\,\mathrm{E}[\hat{f}_n(-x,y)])^3} \\
&\quad + O_P(1)(\hat{f}_n(x,y)\hat{f}_n(-x,y) - \mathrm{E}[\hat{f}_n(x,y)]\,\mathrm{E}[\hat{f}_n(-x,y)])^2. \quad (4.13)
\end{aligned}
$$

The order of the remainders in both expansions can be determined by using the uniform convergence rates for multivariate density estimators given in Giné and Guillou (2002),

$$
\sup|\hat{f}_n(x,y) - \mathrm{E}[\hat{f}_n(x,y)]| = O_p\left(\frac{\log n}{nh^2}\right)^{1/2}, \quad (4.14)
$$

and the decomposition

$$
\begin{aligned}
\left|\hat{f}_n(x,y)\hat{f}_n(-x,y) - \mathrm{E}[\hat{f}_n(x,y)]\,\mathrm{E}[\hat{f}_n(-x,y)]\right| &\leq |\hat{f}_n(x,y) - \mathrm{E}[\hat{f}_n(x,y)]||\hat{f}_n(-x,y)| \\
&\quad + |\mathrm{E}[\hat{f}_n(x,y)]||\hat{f}_n(-x,y) - \mathrm{E}[\hat{f}_n(-x,y)]| \\
&= O_p\left(\frac{\log n}{nh^2}\right)^{1/2}.
\end{aligned}
$$

This gives for $\mathbb{E}_{\mathcal{X}}[I_{n1}]$ with the Taylor expansion (4.13)

$$
\mathbb{E}_{\mathcal{X}}[I_{n1}] = \frac{1}{n^2h^4}\sum_{i=1}^n \sigma^2(X_i,Y_i)\int_A \frac{H^2(X_i,Y_i,x,y)}{(\mathrm{E}[\hat{f}_n(x,y)]\,\mathrm{E}[\hat{f}_n(-x,y)])^2}d(x,y) + o_p\left(\frac{1}{nh} + \frac{h^2}{\sqrt{n}}\right).
$$

Since the variance of the first term is of order $O(1/n^3h^4) = o(1/n^2h^2)$ we can substitute this term by its expectation which yields

$$
\mathbb{E}_{\mathcal{X}}[I_{n1}] = \frac{1}{n^2h^4}\sum_{i=1}^n \int_A \frac{\mathrm{E}[\sigma^2(X_i,Y_i)H^2(X_i,Y_i,x,y)]}{(\mathrm{E}[\hat{f}_n(x,y)]\,\mathrm{E}[\hat{f}_n(-x,y)])^2}d(x,y) + o_p\left(\frac{1}{nh} + \frac{h^2}{\sqrt{n}}\right).
$$

Now it remains to find a representation for I_{n4}. This time we use the Taylor expansion (4.12) and obtain

$$
\begin{aligned}
I_{n4}(\tilde{v}_n) &= \int_A \frac{g_n^2(x,y)}{(\mathrm{E}[\hat{f}(x,y)]\,\mathrm{E}[\hat{f}(-x,y)])^2}d(x,y) \\
&\quad + O_p(1)\int_A g_n^2(x,y)|\hat{f}_n(x,y)\hat{f}_n(-x,y) - \mathrm{E}[\hat{f}_n(x,y)]\,\mathrm{E}[\hat{f}_n(-x,y)]|d(x,y) \\
&= I_{n4}^{[1]} + I_{n4}^{[2]}.
\end{aligned}
$$

21

The variance of $I_{n4}^{[1]}$ is of order $o(1/n^2h^2)$ and therefore $I_{n4}^{[1]}$ can be replaced by its expectation. An application of the Cauchy-Schwarz inequality to $I_{n4}^{[2]}$ and a straight forward calculation yields

$$\int_A g_n^4(x,y)d(x,y) = O_P\left(\frac{1}{n^2} + h^8\right)$$

and by again using the uniform rates in (4.14) it follows that

$$\int_A \{\hat{f}_n(x,y)\hat{f}_n(-x,y) - \mathrm{E}[\hat{f}_n(x,y)]\,\mathrm{E}[\hat{f}_n(-x,y)]\}^2 d(x,y) = O_P\left(\frac{\log n}{nh^2}\right).$$

This means that

$$I_{n4}^{[2]} = O_P\left(\frac{1}{n^2} + h^8\right)^{1/2} O_P\left(\frac{\log n}{nh^2}\right)^{1/2} = o_P\left(\frac{1}{nh}\right),$$

and

$$I_{n4} = \int_A \frac{\mathrm{E}[g_n^2(x,y)]}{(\mathrm{E}[\hat{f}(x,y)]\,\mathrm{E}[\hat{f}(-x,y)])^2} d(x,y),$$

which completes the proof of Theorem 4. $\qquad\square$

Proof of Theorem 5   With the Taylor expansion in (4.12) it follows

$$\tilde{I}_n = \int_A \frac{\kappa_n^2(x,y)[\hat{f}_n(-x,y) - f(-x,y)][\hat{f}_n(-x,y) + f(-x,y)]}{(\mathrm{E}[\hat{f}_n(x,y)]\,\mathrm{E}[\hat{f}_n(-x,y)])^2} d(x,y)$$

$$- \int_A \frac{\kappa_n(x,y)\kappa_n(-x,y)[\hat{f}_n(-x,y) - f(-x,y)]\hat{f}_n(x,y)}{(\mathrm{E}[\hat{f}_n(x,y)]\,\mathrm{E}[\hat{f}_n(-x,y)])^2} d(x,y)$$

$$+ O_p(1)\int_A \kappa_n^2(x,y)[\hat{f}_n(-x,y) - f(-x,y)][\hat{f}_n(-x,y) + f(-x,y)]$$

$$\times |\hat{f}_n(x,y)\hat{f}_n(-x,y) - \mathrm{E}[\hat{f}_n(x,y)]\,\mathrm{E}[\hat{f}_n(-x,y)]| d(x,y)$$

$$- O_p(1)\int_A \kappa_n(x,y)\kappa_n(-x,y)[\hat{f}_n(-x,y) - f(-x,y)]$$

$$\times |\hat{f}_n(x,y)\hat{f}_n(-x,y) - \mathrm{E}[\hat{f}_n(x,y)]\,\mathrm{E}[\hat{f}_n(-x,y)]| d(x,y)$$

$$= O_p\left(\frac{\log n}{nh^2}\right)^{1/2}\int_A \kappa_n^2(x,y)d(x,y) - O_p\left(\frac{\log n}{nh^2}\right)^{1/2}\int_A \kappa_n(x,y)\kappa_n(-x,y)d(x,y)$$

$$+ O_p\left(\frac{\log n}{nh^2}\right)\int_A \kappa_n^2(x,y)d(x,y) - O_p\left(\frac{\log n}{nh^2}\right)\int_A \kappa_n(x,y)\kappa(-x,y)d(x,y)$$

$$= O_p\left(\frac{\log n}{nh^2}\right)^{1/2}\int_A \kappa_n^2(x,y)d(x,y) - O_p\left(\frac{\log n}{nh^2}\right)^{1/2}\int_A \kappa_n(x,y)\kappa_n(-x,y)d(x,y).$$

Jensen's inequality yields

$$\mathrm{Var}\left(\int_A \kappa_n^2(x,y)d(x,y)\right) \le \int_A \mathrm{E}[\kappa_n^4(x,y)]d(x,y) = o\left(\frac{1}{n\log n}\right),$$

22

and therefore we have

$$
\begin{aligned}
\frac{1}{2}|T_n| \;=\; & O_P\left(\frac{\log n}{nh^2}\right)^{1/2} \int_A \mathrm{E}[\kappa_n^2(x,y)]d(x,y) \\
& -O_P\left(\frac{\log n}{nh^2}\right)^{1/2} \int_A \mathrm{E}[|\kappa_n(x,y)\kappa_n(-x,y)|]d(x,y) + o_P\left(\frac{1}{nh}\right).
\end{aligned}
$$

For the expectations we obtain analogously $\int_A \mathrm{E}[\kappa_n^2(x,y)]d(x,y) = o((n\log n)^{-1/2})$, which results in $\tilde{I}_n = o_P((nh)^{-1})$. The calculations for the term $\mathrm{E}[\tilde{S}_n]$ are very similar and therefore omitted at this place. For details we refer to Stahljans (2007). $\qquad \square$

A.2 Proof of Theorem 2

The test statistic can be decomposed into

$$
\begin{aligned}
\hat{I}_n \;=\; & \int_A [\hat{m}(x,y) - m(x,y)]^2 d(x,y) + \int_A [\hat{m}(-x,y) - m(-x,y)]^2 d(x,y) \\
& -2\int_A [\hat{m}(x,y) - m(x,y)][\hat{m}(-x,y) - m(-x,y)]d(x,y) \\
& +2\int_A [\hat{m}(x,y) - m(x,y)][m(x,y) - m(-x,y)]d(x,y) \\
& -2\int_A [\hat{m}(-x,y) - m(-x,y)][m(x,y) - m(-x,y)]d(x,y) + \int_A [m(x,y) - m(-x,y)]^2 d(x,y) \\
\;=\; & 2\int_A [\hat{m}(x,y) - m(x,y)]^2 d(x,y) - 2\int_A [\hat{m}(x,y) - m(x,y)][\hat{m}(-x,y) - m(-x,y)]d(x,y) \\
& +4\int_A [\hat{m}(x,y) - m(x,y)][m(x,y) - m(-x,y)]d(x,y) + I
\end{aligned}
$$

The first two terms are, due to Theorem 4 in Hall (1984), of order $O_P(1/nh) = o_P(1/\sqrt{n})$. For the estimation of the second term we also use the Cauchy-Schwarz inequality. The deterministic part I is an additional bias and is subtracted for the asymptotic consideration. Therefore, it remains to show that the third term is asymptotically normal with rate $\sqrt{n}$. Note that

$$
\int_A [\hat{m}(x,y) - m(x,y)][m(x,y) - m(-x,y)]d(x,y) = \sum_{i=1}^n \mu_i + \sum_{i=1}^n (\tilde{\mu}_i - \mu_i), \qquad (4.15)
$$

where μ_i is defined in (3.2) and

$$
\bar{\mu}_i = \frac{1}{nh^2} \int_A \frac{[Z_i - m(x,y)]k_1\left(\frac{X_i - x}{h}\right)k_2\left(\frac{Y_i - y}{h}\right)[m(x,y) - m(-x,y)]}{\hat{f}_n(x,y)} d(x,y).
$$

We get for the variance of $\sum_{i=1}^{n} \mu_i$

$$\operatorname{Var}\left(\sum_{i=1}^{n} \mu_i\right) = \frac{1}{nh^4} \operatorname{E}\left[\left(\int_A \frac{\sigma(X_1, Y_1)\varepsilon_1 k_1\left(\frac{X_1-x}{h}\right) k_2\left(\frac{Y_1-y}{h}\right)(m(x,y) - m(-x,y))}{\operatorname{E}[\hat{f}(x,y)]} d(x,y)\right)^2\right] + o\left(\frac{1}{n}\right)$$

$$= \frac{1}{n} \int_A \frac{\sigma^2(u,v)(m(u,v) - m(-u,v))^2}{f(u,v)} d(u,v) + o\left(\frac{1}{n}\right)$$

Since the random variables μ_i, $i = 1, \ldots, n$ are independent, we can prove asymptotic normality showing Ljapunov's condition. For this purpose note that

$$n^2 \operatorname{E}[(\mu_i - \operatorname{E}[\mu_i])^4] \leq 8n^2 \operatorname{E}[\mu_i^4] + 8n^2(\operatorname{E}[\mu_i])^4.$$

A straight forward calculation yields

$$\operatorname{E}[\mu_i] = \frac{1}{n} \int_A \frac{[m(x,y) - m(-x,y)]}{\operatorname{E}[\hat{f}_n(x,y)]}$$

$$\times \int_L [m(x+hu, y+hv) - m(x,y)]k_1(u)k_2(v)f(x+hu, y+hv)d(u,v)d(x,y)$$

$$= O\left(\frac{h^2}{n}\right)$$

and

$$\operatorname{E}[\mu_i^4] = \frac{1}{n^4 h^8} \operatorname{E}\left[\left(\int_A \frac{[Z_i - m(x,y)]k_1\left(\frac{X_i-x}{h}\right) k_2\left(\frac{Y_i-y}{h}\right)[m(x,y) - m(-x,y)]}{\operatorname{E}[\hat{f}_n(x,y)]} d(x,y)\right)^4\right]$$

$$= O\left(\frac{1}{n^4}\right).$$

This results in

$$n^2 \sum_{i=1}^{n} \operatorname{E}[(\mu_i - \operatorname{E}[\mu_i])^4] = O\left(\frac{h^8}{n} + \frac{1}{n}\right) = o(1).$$

Therefore the Ljapunov condition is satisfied and we have with the notation of (3.2) the asymptotic normality

$$\sqrt{n} \sum_{i=1}^{n} (\mu_i - \operatorname{E}[\mu_i]) = \sqrt{n}(T_n - \operatorname{E}[T_n]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2).$$

We conclude the proof by showing that the term $\sum_{i=1}^{n} (\tilde{\mu}_i - \mu_i)$ in (4.15) is asymptotically negligible. A Taylor expansion of $g(\hat{f}_n(x,y))$ with $g(z) = 1/z$ in a neighbourhood of the point $\operatorname{E}[\hat{f}_n(x,y)]$ and an application of the uniform convergence rates for density estimates gives

$$\sqrt{n} \sum_{i=1}^{n} (\bar{\mu}_i - \mu_i) = O_P\left(\frac{(\log n)^{1/2}}{nh^3}\right) \sum_{i=1}^{n} M_{i,n}$$

24

with
$$M_{i,n} = \int_A [Z_i - m(x,y)][m(x,y) - m(-x,y)]k_1\Big(\frac{X_i - x}{h}\Big)k_2\Big(\frac{Y_i - y}{h}\Big)d(x,y).$$

A straight forward calculation now yields

$$\sum_{i=1}^n M_{i,n} = \sum_{i=1}^n \mathrm{E}[M_{i,n}] + O_P(n^{1/2}h^2) = O(nh^4) + O_P(n^{1/2}h^2) = o_P\Big(\frac{nh^3}{(\log n)^{1/2}}\Big)$$

and therefore the proof of Theorem 2 is complete. $\qquad\square$