Testing strict monotonicity in nonparametric regression

Melanie Birke
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: melanie.birke@rub.de

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: holger.dette@rub.de
FAX: +49 234 3214 559

December 7, 2006

Abstract

A new test for strict monotonicity of the regression function is proposed which is based on a composition of an estimate of the inverse of the regression function with a common regression estimate. This composition is equal to the identity if and only if the "true" regression function is strictly monotone, and a test based on an L^2-distance is investigated. The asymptotic normality of the corresponding test statistic is established under the null hypothesis of strict monotonicity.

AMS Subject Classification: 62G10

Keywords and Phrases: nonparametric regression, strictly monotone regression, goodness-of-fit test

1 Introduction

Consider the common nonparametric regression model

$$Y_i = m(X_i) + \sigma(X_i)\varepsilon_i, \quad i = 1, \dots, n \tag{1.1}$$

where $(X_i, Y_i)_{i=1,\dots,n}$ is a sample of bivariate observations and $E[\varepsilon_i] = 0$. In nonparametric regression models one typically assumes that $m(\cdot)$ is continuously differentiable of a certain order and estimates this function by some smoothing procedure. In many practical applications additional qualitative information regarding the unknown regression function $m(\cdot)$ is available. A typical information of this type is that of strict monotonicity, which is often motivated by biological, economic or physical reasoning. If this assumption is justified it can be incorporated in the estimation procedure and there exists a vast amount of literature on the estimation of a regression function under the monotonicity constraint [see e.g. Brunk (1955), Friedman and Tibshirani (1984), Mukerjee (1988), Mammen (1991), Ramsay (1998), Hall and Huang (2001) or Dette, Neumeyer and Pilz (2006) among many others]. Although a goodness-of-fit test for monotonicity is important to justify this assumption, the literature on this subject is not so rich and the problem of testing for monotonicity has only found recently attention in the literature. Schlee (1982) proposed a test for this hypothesis, which is based on estimates of the derivative of the regression function. Bowman Jones and Gijbels (1998) used Silverman's (1981) "critical bandwidth" approach to construct a bootstrap test for monotonicity while Gijbels, Hall, Jones and Koch (2000) considered the length of runs for that purpose. More recent work on testing monotonicity can be found in Hall and Heckman (2001), Goshal, Sen and Van der Vaart (2000), Durot (2003), Baraud, Huet and Laurent (2003) and Domínguez-Menchero, González-Rodríguez and López -Palomo (2005).

In the present paper we propose an alternative procedure for testing monotonicity. In contrast to the literature cited in the previous paragraph we consider the null hypothesis of strict monotonicity, which has - to our knowledge - not been considered before. We propose to consider the composition of an estimate proposed by Dette et al. (2006) for the inverse regression function with an unconstrained estimate of the regression function. Under the null hypothesis of strict monotonicity this composition equals the identity and an L^2-distance between the composition and the identity is proposed as test statistic. We prove consistency and asymptotic normality of this statistic under the null hypothesis. For the sake of brevity we restrict ourselves to the hypothesis

$$H_0 : m \quad \text{is strictly isotone} \tag{1.2}$$

but the transformation to the strictly antitone case is rather obvious and indicated in Remark 2.3. The paper is organized as follows. Our idea for constructing the test statistic is carefully described in Section 2, while Section 3 contains the main results and gives some further discussion. Auxiliary results needed in the proof of our main theorem are deferred to the Appendix.

2 Testing for a strictly isotone regression

Recall the definition of the nonparametric regression model in (1.1), assume that X_i has a density, say f, with compact support $[0, 1]$, and that the random errors $\varepsilon_1, \ldots, \varepsilon_n$ are centered with mean 0 and variance 1. In order to motivate the test statistic, we briefly recall the definition of an estimate of the "inverse" of the regression function $m(\cdot)$, which was recently proposed by Dette et al. (2006). For this purpose let

$$\hat{f}_n(x) = \frac{1}{nh_r} \sum_{i=1}^{n} K_r\Big(\frac{x - X_i}{h_r}\Big) \tag{2.1}$$

denote the common density estimate and define

$$\hat{m}(x) = \frac{1}{nh_r} \sum_{i=1}^{n} K_r\Big(\frac{x - X_i}{h_r}\Big) Y_i / \hat{f}_n(x) \tag{2.2}$$

as the Nadaraya-Watson estimate. Dette et al. (2006) proposed

$$\hat{\phi}_{h_d}(t) = \frac{1}{h_d} \int_0^1 \int_{-\infty}^{t} K_d\Big(\frac{\hat{m}(v) - u}{h_d}\Big) du dv. \tag{2.3}$$

as an estimate of the "inverse" of the regression function m, where K_d is a symmetric kernel with compact support, say $[-1, 1]$, and h_d is a bandwidth converging to 0 with increasing sample size. Intuitively, if $h_d \to 0$, the statistic $\hat{\phi}_{h_d}(t)$ approaches

$$\hat{\phi}(t) \quad = \quad \int_0^1 I\{\hat{m}(v) \leq t\} dv \approx \int_0^1 I\{m(v) \leq t\} dv =: \phi(t) \tag{2.4}$$

where the approximation is justified for an increasing sample size using the uniform consistency of the Nadaraya-Watson estimate [see e.g. Mack and Silverman (1982)]. Note that the right hand side of (2.4) is equal to $m^{-1}(t)$ if the null hypothesis (1.2) is satisfied. In this case $\hat{\phi} \circ \hat{m}$ would converge to the identity and therefore we propose

$$T_n = \int_0^1 (\hat{\phi}_{h_d}(\hat{m}(x)) - x)^2 dx \tag{2.5}$$

as test statistic for the hypothesis of a strictly increasing regression function in model (1.1). Our first result specifies the limit of (2.5), if the estimate $\hat{m}$ converges uniformly to the true regression function [for sufficient assumptions for this property see e.g. Mack and Silverman (1982).

Lemma 2.1. *Assume that the assumptions stated at the beginning of this section are satisfied and that the estimate $\hat{m}$ converges uniformly to m. If $n \to \infty, h_d \to 0$ we have $T_n \xrightarrow{P} T$, where the quantity T is defined by*

$$T = \int_0^1 \Big(\int_0^1 I\{m(v) \leq m(x)\} dv - x\Big)^2 dx \tag{2.6}$$

Proof. The difference between the statistic T_n and the "parameter" T can be written as

$$T_n - T \quad = \quad \int_0^1 \Big((\hat{\phi}_{h_d}(\hat{m}(x)) - x)^2 - (\phi(m(x)) - x)^2 \Big) dx$$

$$
\begin{aligned}
&= \int_0^1 \Big(\hat{\phi}^2_{h_d}(\hat{m}(x)) - \phi^2(m(x)) - 2x(\hat{\phi}_{h_d}(\hat{m}(x)) - \phi(m(x)))\Big) \\
&= \int_0^1 (\hat{\phi}_{h_d}(\hat{m}(x)) + \phi(m(x)) - 2x)(\hat{\phi}_{h_d}(\hat{m}(x)) - \phi(m(x)))dx \\
&= O_P(1) \int_0^1 (\hat{\phi}_{h_d}(\hat{m}(x)) - \phi(m(x)))dx
\end{aligned}
$$

by using the boundedness of $\hat{\phi}_{h_d}(\hat{m}(x))$ and $\phi(m(x))$. Therefore it suffices to show that the difference $\hat{\phi}_{h_d}(\hat{m}(x)) - \phi(m(x))$ converges uniformly to 0. Using the definition of the statistic $\hat{\phi}_{h_d}(\hat{m}(x))$ yields

$$
\begin{aligned}
\hat{\phi}_{h_d}(\hat{m}(x)) &= \frac{1}{h_d}\int_0^1 \int_{-\infty}^{\hat{m}(x)} K_d\Big(\frac{\hat{m}(v)-u}{h_d}\Big)dudv \\
&= \frac{1}{h_d}\int_0^1 I\{\hat{m}(v) \le \hat{m}(x) + h_d\} \int_{\hat{m}(v)-h_d}^{\hat{m}(x)} K_d\Big(\frac{\hat{m}(v)-u}{h_d}\Big)dudv \\
&= \int_0^1 I\{\hat{m}(v) \le \hat{m}(x) + h_d\} \int_{\frac{\hat{m}(v)-\hat{m}(x)}{h_d}}^{1} K_d(u)dudv \\
&= \int_0^1 I\{\hat{m}(v) \le \hat{m}(x) - h_d\}dv \\
&\quad + \int_0^1 I\{\hat{m}(x) - h_d \le \hat{m}(v) \le \hat{m}(x) + h_d\} \int_{\frac{\hat{m}(v)-\hat{m}(x)}{h_d}}^{1} K_d(u)dudv.
\end{aligned}
$$

The first term converges to $\phi(m(x))$ because of the uniform consistency of the estimate $\hat{m}$. The second term is smaller than

$$
\int_0^1 I\{\hat{m}(x) - h_d \le \hat{m}(v) \le \hat{m}(x) + h_d\}dv
$$

which converges to 0 by again using the uniform consistency of the estimate $\hat{m}$. This proofs Lemma 2.1. □

Obviously, if the regression function m is strictly increasing the parameter T vanishes and the following result shows that this is a necessary and sufficient condition for strict monotonicity.

Proposition 2.2. *Assume that the regression function m is continuous. The parameter T defined by (2.6) is equal to 0 if and only if the regression function m is strictly increasing on the interval $[0,1]$.*

Proof of Proposition 2.2. Obviously the result follows if we can prove that the assertion

$$
\int_0^1 I\{m(v) \le m(x)\}dv = x \text{ for almost all } x \in [0,1] \tag{2.7}
$$

holds if and only if the regression function m is strictly increasing. If the latter case is satisfied, then (2.7) is obviously true for all $x \in [0, 1]$, and it remains to prove the necessary part.

For this purpose we assume that (2.7) holds and distinguish three cases

(a) m is increasing on the interval $[0, 1]$ but not strictly increasing

(b) m is decreasing on the interval $[0, 1]$

(c) m is neither increasing nor decreasing on the interval $[0, 1]$

(a) In this case there exist disjoint intervals A_i, $i \in I$, where m is constant and intervals B_j, $j \in J$, where m is strictly increasing with

$$\Big(\bigcup_{i\in I} A_i\Big) \cup \Big(\bigcup_{j\in J} B_j\Big) = [0, 1].$$

This decomposition implies the representation

$$m(x) = \sum_{i\in I} m_i I_{A_i}(x) + \sum_{j\in J} m_j^{\nearrow}(x) \tag{2.8}$$

for some constants $m_i \in \mathbb{R}$ $(i \in I)$ and strictly increasing functions $m_j^{\nearrow} = m_{|B_j}$ $j \in J$. Note that

$$\phi(t) = \int_0^1 I\{m(v) \le t\}dv = \sup\{v \in [0, 1] | m(v) \le t\}$$

if m is increasing and $t \in \mathrm{Im}(m)$. Consequently, if $x \in \mathrm{Int}(A_i)$ for some $i \in I$ we have $\phi(m(x)) > x$, which implies $\phi(m(x)) - x > 0$ on a set with positive Lebesgue measure which contradicts assumption (2.7). Note that this argument also covers the case, where the regression function m is constant on the interval $[0, 1]$.

(b) If the regression function m is decreasing but not constant on the interval $[0, 1]$ there exist intervals A_i, $i \in I$, where m is constant and intervals B_j, $j \in J$, where m is strictly decreasing. As in case (a) we have a decomposition of the form (2.8) with constants $m_i \in \mathbb{R}$ $(i \in I)$ and strictly decreasing functions $m_j^{\searrow} = m_{|B_j}$ $(j \in J)$, that is

$$m(x) = \sum_{i\in I} m_i I_{A_i}(x) + \sum_{j\in J} m_j^{\searrow}(x).$$

In this case it follows

$$\phi(m(x)) = \int_0^1 I\{m(v) \le m(x)\}dv = 1 - \inf\{v \in [0, 1] \,|\, m(v) \le m(x)\}$$

Because $J \ne \emptyset$ we have $\phi(m(x)) = 1 - x \ne x$ on $\cup_{j\in J} B_j$. This is a set of positive Lebesgue measure, which contradicts assumption (2.7).

(c) This follows by combining similar arguments as given in (a) and (b).

□

Remark 2.3. For a test of the hypothesis of a strictly antitone regression function a strictly antitone inverse regression estimate instead of the isotone inverse regression estimate is used in the definition of the test statistic. An antitone inverse regression estimate is defined by

$$\hat{\varphi}(t) = \int_0^1 I\{\hat{m}(v) \geq t\} dv,$$

and the smoothed version is given by

$$\hat{\varphi}_{h_d}(t) = \frac{1}{h_d} \int_0^1 \int_t^\infty K_d\Big(\frac{\hat{m}(v) - u}{h_d}\Big) du dv.$$

We now obtain a test statistic for the null hypothesis

$$\tilde{H}_0 : m \quad \text{is strictly antitone}$$

as

$$\tilde{T}_n = \int_0^1 (\hat{\varphi}_{h_d}(\hat{m}(x)) - x)^2 dx.$$

It can be shown by similar methods as above that $\tilde{T}_n$ converges to the quantity

$$T_A = \int_0^1 \Big(\int_0^1 I\{m(v) \geq m(x)\} dv - x \Big)^2 dx,$$

which vanishes if and only if m is strictly decreasing.

In the following section we derive the asymptotic distribution of the test statistic under the null hypothesis. We restrict ourselves to the case of testing strict isotonicity but a similar result for testing the hypothesis of a strictly antitone regression function can be obtained in a similar way.

3 Main result

In this section we investigate the weak convergence of the statistic defined in (2.5). For this purpose we require several regularity assumptions on the kernels K_d, K_r and the bandwidths h_d, h_r in the estimate of the inverse regression function:

(K1) The kernel K_r is of order 2 and three times continuously differentiable with compact support $[-1, 1]$ such that $K_r(\pm 1) = K_r'(\pm 1) = 0$

(K2) The kernel K_d is of order 2, positive and twice continuously differentiable with compact support $[-1, 1]$ and $K_d(\pm 1) = K_d'(\pm 1) = 0$

(B) If $n \to \infty$ the bandwidths h_d and h_r have to satisfy

$$\begin{aligned}
&h_r, h_d \to 0 \\
&nh_r, nh_d \to \infty \\
&h_r = O(n^{-1/5}) \\
&h_d^2 \log h_r^{-1} / h_r^{5/2} \to 0 \\
&h_r^{1/2} (\log h_r^{-1})^2 / nh_d^4 = O(1).
\end{aligned}$$

If the bandwidth h_r is chosen asymptotically optimal as $h_r = \gamma_r n^{-1/5}$ for a constant $\gamma_r > 0$, then the last two conditions simplify to $\sqrt{nh_d^4} \log n \to 0$ and $(\log n)^2 / n^{11/10} h_d^4 = O(1)$. The second bandwidth can then, for example, be chosen as $h_d = \gamma_d n^{-a}$ with $1/4 < a < 11/40$ and $\gamma_d > 0$.

Theorem 3.1. *Assume that the regression function m in model (1.1) is four times continuously differentiable with $m'(x) > 0$ for all $x \in [0,1]$, f is three times continuously differentiable and positive and σ^2 is continuously differentiable on the interval $[0,1]$. If $\mathrm{E}[\mu_4(X_1)] < \infty$ with $\mu_4(X_1) = \mathrm{E}[(Y_1 - m(X_1))^4 | X_1]$ and conditions (K1), (K2) and (B) are satisfied, we have as $n \to \infty$*

$$\frac{nh_r^{9/2}}{h_d^4} \Big(T_n - h_d^4 \kappa_2^2(K_d)(B_n^{[1]} + B_n^{[2]}) \Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V),$$

where the asymptotic bias and variance are given by

$$\begin{aligned}
B_n^{[1]} &= \frac{1}{nh_r^5} \int_0^1 \frac{\sigma^2(x)}{f(x)(m'(x))^6} dx \int_{-1}^1 K_r''^2(y) dy \\
B_n^{[2]} &= \int_0^1 \frac{(m''(x))^2}{(m'(x))^6} dx
\end{aligned}$$

and

$$V = 4\kappa_2^4(K_d) \Big(\int_0^1 \sigma^2(y) f^2(y) (m'(y))^{-12} dy \Big) \Big(\int_0^1 \Big(\int_0^1 K_r''(x) K_r''(x+z) dx \Big)^2 dz \Big).$$

Proof of Theorem 3.1. Let $\mathcal{C}(A)$ denote the set of all continuous functions on $A \subset \mathbb{R}$. We consider the test statistic T_n as functional on $\mathcal{C}(\mathbb{R}) \times \mathcal{C}(\mathbb{R})$, i.e. $T_n = \Psi(\hat{\phi}_{hd}, \hat{m})$, where

$$\Psi(f, g) = \int_0^1 (f(g(x)) - x)^2 dx.$$

For sufficiently smooth f, g the functional ψ is Gateáux differentiable and we obtain by a Taylor expansion [see Serfling (1980) pp. 314-315] the stochastic expansion

$$T_n = \int_0^1 \Big\{(\hat{m}-m)(x)(m^{-1})'(m(x)) + (\hat{\phi}_{h_d} - m^{-1})(m(x))\Big\}^2 dx + \frac{1}{6}P^{(3)}(\lambda^*), \tag{3.1}$$

where $\lambda^* \in [0,1]$ and the remainder $P^{(3)}$ is defined by

$$\begin{aligned} P^{(3)}(\lambda) &= 6\int_0^1 \Big\{\tilde{g}(x)[f^{(1)} + \lambda \tilde{f}^{(1)}]([g+\lambda\tilde{g}](x)) + \tilde{f}([g+\lambda\tilde{g}](x))\Big\} \\ &\quad \times\Big\{\tilde{g}^2(x)[f^{(2)} + \lambda\tilde{f}^{(2)}]([g+\lambda\tilde{g}](x)) + 3\tilde{g}(x)\tilde{f}^{(1)}([g+\lambda\tilde{g}](x))\Big\}dx \\ &\quad +2\int_0^1 \Big\{[f+\lambda\tilde{f}]([g+\lambda\tilde{g}](x)) - x\Big\} \\ &\quad \times\Big\{\tilde{g}^3(x)[f^{(3)} + \lambda\tilde{f}^{(3)}]([g+\lambda\tilde{g}](x)) + 2\tilde{g}^2(x)\tilde{f}^{(2)}([g+\lambda\tilde{g}](x))\Big\}dx. \end{aligned} \tag{3.2}$$

A similar calculation shows

$$\hat{\phi}_{h_d}(m(x)) - m^{-1}(m(x)) = A_{h_d}(m(x)) + \Delta_n^{(1)}(m(x)) + \frac{1}{2}\Delta_n^{(2)}(m(x))(1+o_P(1)), \tag{3.3}$$

where the quantities $A_{h_d}, \Delta_n^{(1)}$ and $\Delta_n^{(2)}$ are given by

$$A_{h_d}(m(x)) = \phi_{h_d}(m)(m(x)) - m^{-1}(m(x)) \tag{3.4}$$

$$\begin{aligned} \Delta_n^{(1)}(m(x)) &= -\int_0^1 K_d(v)(m^{-1})'(m(x)+h_dv)(\hat{m}-m)(m^{-1}(m(x)+h_dv))dv \\ &= -(m^{-1})'(m(x))(\hat{m}-m)(x) - h_d^2\kappa_2(K_d)[(m^{-1})'(m(x))]^3(\hat{m}-m)''(x) \\ &\quad -R_n(x) \end{aligned} \tag{3.5}$$

$$\Delta_n^{(2)}(m(x)) = -\frac{1}{h_d}\int_0^1 K_d'(v)(m^{-1})'(m(x)+h_dv)(\hat{m}-m)^2(m^{-1}(m(x)+h_dv))dv \tag{3.6}$$

and the remainder in (3.5) is defined by

$$\begin{aligned} R_n(x) &= h_d^2\kappa_2(K_d)[(m^{-1})^{(3)}(m(x))(\hat{m}-m) + 3(m^{-1})''(m(x))(m^{-1})'(m(x))(\hat{m}-m)'(x)] \\ &\quad +\frac{h_d^3}{6}[(m^{-1})'(\hat{m}-m)\circ m^{-1}]^{(3)}(\xi_n(x)). \end{aligned} \tag{3.7}$$

A combination of these estimates yields for the test statistic the representation

$$T_n = h_d^4\kappa_2^2(K_d)\int_0^1 [m'(x)]^{-6}(\hat{m}^{(2)}(x) - m^{(2)}(x))^2dx + \int_0^1 A_{h_d}^2(m(x))dx + Q_n, \tag{3.8}$$

where the remainder term Q_n is given by

$$\begin{aligned} Q_n &= \int_0^1 R_n^2(x)dx + \frac{1}{4}\int_0^1 (\Delta_n^{(2)}(m(x)))^2dx \\ &\quad +2\Big\{-h_d^2\kappa_2(K_d)\int_0^1 [m'(x)]^{-3}(\hat{m}^{(2)}(x) - m^{(2)}(x))A_{h_d}(x)dx \end{aligned}$$

$$
\begin{aligned}
&-h_d^2\kappa_2(K_d)\int_0^1 [m'(x)]^{-3}(\hat m^{(2)}(x)-m^{(2)}(x))R_n(x)dx\\
&-h_d^2\kappa_2(K_d)\int_0^1 [m'(x)]^{-3}(\hat m^{(2)}(x)-m^{(2)}(x))\Delta_n^{(2)}(m(x))dx\\
&+\int_0^1 A_{h_d}(x)R_n(x)dx+\frac{1}{2}\int_0^1 A_{h_d}(x)\Delta_n^{(2)}(m(x))dx+\frac{1}{2}\int_0^1 R_n(x)\Delta_n^{(2)}(m(x))dx\Big\}\\
&+\frac{1}{6}P^{(3)}(\lambda^*)
\end{aligned}
$$

It follows from Theorem A.1 in the Appendix that the first term in (3.8) converges weakly with a normal limit, that is

$$
(3.9)\qquad \frac{nh_r^{9/2}}{h_d^4}\cdot h_d^4\kappa_2^2(K_d)\Big(\int_0^1 [m'(x)]^{-6}(\hat m^{(2)}(x)-m^{(2)}(x))^2dx-B_n^{[1]}\Big)\xrightarrow{\mathcal{D}}\mathcal{N}(0,V).
$$

For the second term we have by a straightforward calculation

$$
(3.10)\qquad \int_0^1 A_{h_d}^2(m(x))dx = h_d^4\kappa_2^2(K_d)B_n^{[2]}+o(h_d^6).
$$

(note that the remainder term is of order $o(h_d^4/nh_r^{9/2})$. The assertion is now a consequence of the estimate

$$
(3.11)\qquad Q_n = o_p(h_d^4/nh_r^{9/2}),
$$

which will be proved in several steps.

First note that a standard argument yields

$$
\begin{aligned}
\int_0^1 R_n^2(x)dx &\le Ch_d^4\Big[\int_0^1 w_1(x)d^2(x)dx+\int_0^1 w_2(x)(d'(x))^2dx\\
&\quad +h_d^2\int_0^1\Big([(m^{-1})'d\circ m^{-1}]^{(3)}(\xi(x))\Big)^2dx\Big]\\
&= O_P\Big(\frac{h_d^4}{nh_r^{5/2}}\Big)+O_P\Big(\frac{h_d^6\log h_r^{-1}}{nh_r^7}\Big)=o_P\Big(\frac{h_d^4}{nh_r^{9/2}}\Big),
\end{aligned}
$$

where $w_1(x)=[(m^{-1})^{(3)}(m(x))]^2$, $w_2(x)=[(m^{-1})^{(2)}(m(x))(m^{-1})'(m(x))]^2$, $d(x)=\hat m(x)-m(x)$, and the second inequality follows from the fact that the integrand $(\hat m^{(3)}-m^{(3)})^2$ is of order $O_p(\log h_r^{-1}/nh_r^7)$ uniformly with respect to x [this can be derived by similar methods as in Mack and Silverman (1982)] . Similarly, we obtain for the second and third term in the decomposition of Q_n

$$
\int_0^1(\Delta_n^{(2)}(m(x)))^2dx \;=\; \int_0^1\Big(\int_{-1}^1 K_d(v)[(m^{-1})''(m(x)+h_dv)d^2(m^{-1}(m(x)+h_dv))
$$

$$
\begin{aligned}
&+ \quad 2[(m^{-1})'(m(x)+h_dv)]^2 d'(m^{-1}(m(x)+h_dv))d^2(m^{-1}(m(x)+h_dv))]dv\Big)^2 dx \\
&= \quad O_P\Big(\frac{(\log h_r^{-1})^2}{n^2h_r^4}\Big) = O_P\Big(\frac{h_d^4}{nh_r^{9/2}}\frac{(\log h_r^{-1})^2}{nh_d^4}h_r^{1/2}\Big) = o_P\Big(\frac{h_d^4}{nh_r^{9/2}}\Big),
\end{aligned}
$$

$$
\begin{aligned}
& \Big|h_d^2\int_0^1 [m'(x)]^{-3}(\hat{m}^{(2)}(x)-m^{(2)}(x))A_{h_d}(m(x))dx\Big| \\
\leq & \Big|h_d^2[m'(x)]^{-3}(\hat{m}'(x)-m'(x))A_{h_d}(m(x))\Big|_0^1\Big| \\
+ & \Big|h_d^2\int_0^1 (\hat{m}'(x)-m'(x))[(m'(x))^{-3}A_{h_d}(m(x))]'dx\Big| \\
= & O_P\Big(h_d^3\Big(\frac{\log h_r^{-1}}{nh_r^3}\Big)^{1/2}\Big) \;=\; o_P\Big(\frac{h_d^4}{nh_r^{9/2}}\Big),
\end{aligned}
$$

where we have used integration by parts and the assumption that the kernel K_d vanishes at the boundary of its support. The remaining five terms of Q_n are estimated by means of the Cauchy-Schwarz inequality and are all of order $o_p(h_d^4/(nh_r^{9/2}))$. Consequently, the assertion (3.11) (and from this estimate the assertion of the theorem) now follows if the estimate

$$
P^{(3)}(\lambda^*) = o_p\Big(\frac{h_d^4}{nh_r^{9/2}}\Big) \tag{3.12}
$$

for the random variable defined in (3.2) can be established. For this estimate we introduce the notation $d(x)=\hat{m}(x)-m(x)$ and $d_{I,-1}(y)=\hat{\phi}_{h_d}(y)-m^{-1}(y)$, and obtain the representation

$$
\begin{aligned}
P^{(3)}(\lambda) \;=\;& 6\int_0^1 \Big\{d(x)[(m^{-1})^{(1)}+\lambda d_{I,-1}^{(1)}]([m+\lambda d](x))+d_{I,-1}([m+\lambda d](x))\Big\} \\
&\times\Big\{d^2(x)[(m^{-1})^{(2)}+\lambda d_{I,-1}^{(2)}]([m+\lambda d](x))+2d(x)d_{I,-1}^{(1)}([m+\lambda d](x))\Big\}dx \\
&+2\int_0^1 \Big\{d(x)(m^{-1})'(\hat{\xi}(x))+\lambda d_{I,-1}([m+\lambda d](x))\Big\} \\
&\times\Big\{d^3(x)[(m^{-1})^{(3)}+\lambda d_{I,-1}^{(3)}]([m+\lambda d](x))+3d^2(x)d_{I,-1}^{(2)}([m+\lambda d](x))\Big\}dx
\end{aligned}
$$

for some $\hat{\xi}(x)$ with $|\hat{\xi}(x)-m(x)|\leq|\hat{m}(x)-m(x)|$. ¿From Mack and Silverman (1982) and Lemma B.2 in the Appendix it follows

$$
\begin{aligned}
d(x) &= O_P\Big(\frac{\log h_r^{-1}}{nh_r}\Big), \\
d_{I,-1}^{(k)}(y) &= O_P\Big(\frac{\log h_r^{-1}}{nh_r^{2k+1}}\Big)^{1/2}+O(h_d^2) \text{ for } \quad k=0,1,2, \\
d_{I,-1}^{(3)}(y) &= O_P\Big(\frac{\log h_r^{-1}}{nh_r^{7}}\Big)^{1/2}+o(h_d),
\end{aligned}
$$

which yields the estimate

$$
\begin{aligned}
P^{(3)}(\lambda) &= \Big\{O_P\Big(\frac{\log h_r^{-1}}{nh_r}\Big)^{1/2}\Big[O_P(1)+O_P\Big(\frac{\log h_r^{-1}}{nh_r^3}\Big)^{1/2}+O(h_d^2)\Big]+O(h_d^2)\Big\} \\
&\quad \times\Big\{O_P\Big(\frac{\log h_r^{-1}}{nh_r}\Big)\Big[O_P(1)+O_P\Big(\frac{\log h_r^{-1}}{nh_r^5}\Big)^{1/2}+O(h_d^2)\Big] \\
&\qquad +O_P\Big(\frac{\log h_r^{-1}}{nh_r}\Big)^{1/2}\Big[O_P\Big(\frac{\log h_r^{-1}}{nh_r^3}\Big)^{1/2}+O(h_d^2)\Big]\Big\} \\
&+ \Big\{O_P\Big(\frac{\log h_r^{-1}}{nh_r}\Big)^{1/2}+O(h_d^2)\Big\}\Big\{O_P\Big(\frac{\log h_r^{-1}}{nh_r}\Big)^{3/2}\Big[O_P(1)+O_P\Big(\frac{\log h_r^{-1}}{nh_r^7}\Big)^{1/2}+o(h_d)\Big] \\
&\qquad +O_P\Big(\frac{\log h_r^{-1}}{nh_r}\Big)\Big[O_P\Big(\frac{\log h_r^{-1}}{nh_r^5}\Big)^{1/2}+O(h_d^2)\Big]\Big\} \\
&= o_P\Big(\frac{h_d^4}{nh_r^{9/2}}\Big)
\end{aligned}
$$

by using the last two conditions on the bandwidths specified in (B). This proves assertion (3.12) and therefore the proof of Theorem 3.1 is completed. □

Remark 3.2 If a local polynomial estimate instead of the Nadaraya-Watson estimator is used Theorem 3.1 still holds with a different bias and variance. If we use the representation of the local polynomial estimate of order p

$$\hat{m}_p(x) = \frac{1}{nh\,f(x)}\sum_{i=1}^{n} K_r^*\Big(\frac{x-X_i}{h}\Big)Y_i(1+o_P(1))$$

with K_r^* denoting the corresponding equivalent kernel [see Fan and Gijbels (1997)], we get under the assumptions of Theorem 3.1

$$\frac{nh_r^{9/2}}{h_d^4}\Big(T_n - h_d^4\kappa_2^2(K_d)(\tilde{B}_n^{[1]}+B_n^{[2]})\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\tilde{V}),$$

where the asymptotic bias and variance are given by

$$
\begin{aligned}
\tilde{B}_n^{[1]} &= \frac{1}{nh_r^5}\int_0^1 \frac{\sigma^2(x)}{f(x)(m'(x))^6}dx\int_{-1}^1 (K_r^*)''^2(y)dy \\
B_n^{[2]} &= \int_0^1 \frac{(m''(x))^2}{(m'(x))^6}dx
\end{aligned}
$$

and

$$\tilde{V} = 4\kappa_2^4(K_d)\Big(\int_0^1 \sigma^2(y)f^2(y)(m'(y))^{-8}dy\Big)\Big(\int_0^1(\int_0^1 (K_r^*)''(x)(K_r^*)''(x+z)dx)^2dz\Big).$$

Appendix: Some auxiliary results

In this section we present several auxiliary results which are required for a proof of Theorem 3.1. The first one generalizes a result of Hall (1984), who proved asymptotic normality of the integrated squared error between the Nadaraya-Watson estimate and the unknown regression function. The proof is similar to the corresponding statement in Hall (1984) for the case $k = 0$ and therefore not presented here.

Theorem A.1. *Let $k \in \{0, 1, 2\}$ and denote by w a nonnegative weight function. Assume that $A \subset \mathbb{R}$ is compact and define*

$$A^\varepsilon := \{x \in I\!R | \inf_{a \in A} |x - a| < \varepsilon\}.$$

Suppose that the variance function σ^2 in model (1.1) is bounded and continuously differentiable on A^ε, w is bounded and continuous on A^ε, m is $(k+2)$-times continuously differentiable on A^ε and f is $(k+1)$-times continuously differentiable such that $f^{(k+1)}$ is uniformly continuous on A^ε. If $h_r \to 0; nh_r \to \infty, nh_r^{3/2+k} \to \infty, h_r = O(n^{-1/5})$ we have for $k = 0, 1, 2$

$$T_n^{(k)} := (n^{-1}h^{-4k-1}\alpha_{1,k} + nh^{2k-4}\alpha_{2,k})^{-1/2}\Big(\int_A (\hat{m}^{(k)}(x) - m^{(k)}(x))^2 w(x)dx - B_{n,k}\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

where the constants $\alpha_{j,k}$, γ_k and $B_{n,k}$ are given by

$$\begin{aligned}
\alpha_{1,k} &= 2\Big(\int_A \sigma^4(x)w^2(x)f^{-2}(x)dx\Big)\Big(\int \Big(\int K_r^{(k)}(x)K_r^{(k)}(x+y)dx\Big)^2 dy\Big), \quad k = 0, 1, 2, \\
\alpha_{2,k} &= \begin{cases} 4\int_A \sigma^2(x)\gamma_0^2(x)w^2(x)f^{-4}(x)dx & \text{if } k = 0 \\ 0 & \text{else} \end{cases} \\
\gamma_k(x) &= \kappa_2(K_r)\Big(m^{(k+2)}(x)f(x) + 2m^{(1)}(x)f^{(k+1)}(x) + \sum_{j=0}^{k-1}\binom{k}{j+1}\frac{k+2+j}{k-j}m^{(k+2-j)}(x)f^{(j)}(x)\Big) \\
B_{n,k} &= \begin{cases} \frac{1}{nh_r}\int_0^1 \frac{\sigma^2(x)w(x)}{f(x)}dx \int_{-1}^1 K_r^2(y)dy + h_r^4\kappa_2^2(K_r)\int_0^1 \frac{(m''(x)f(x) - m(x)f''(x))w(x)}{f^2(x)}dx & \text{if } k = 0 \\ \frac{1}{nh_r^{2k+1}}\int_0^1 \frac{\sigma^2(x)w(x)}{f(x)}dx \int_{-1}^1 K_r^{(k)^2}(y)dy & \text{if } k = 1, 2 \end{cases}
\end{aligned}$$

Theorem A.2. *Define $J := J(\delta) = [m(0) + \delta, m(1) - \delta]$, where $\delta := \delta(h_d) > 0$ is chosen such that for all $t \in J(\delta)$: $t + h_d v \in [m(0), m(1)]$, whenever $v \in [-1, 1]$. Assume that the assumptions of Theorem 3.1 are satisfied, then almost surely*

$$\begin{aligned}
\sup_t |(\hat{\phi}_{h_d})^{(s)}(t) - (m^{-1})^{(s)}(t)| &= O\Big(\frac{\log h_r^{-1}}{nh_r^{2s+1}}\Big)^{1/2} + O(h_d^2) \text{ for } \quad s = 0, 1, 2 \\
\sup_t |(\hat{\phi}_{h_d})^{(3)}(t) - (m^{-1})^{(3)}(t)| &= O\Big(\frac{\log h_r^{-1}}{nh_r^5}\Big)^{1/2} + o(h_d).
\end{aligned}$$

Proof. Note that the supremum can be decomposed into two stochastic parts and one deterministic part, i.e.

$$\sup_{t\in J}|\hat{\phi}^{(s)}_{h_d}(t)-(m^{-1})^{(s)}(t)| \le \sup_{t\in J}|\frac{\partial^s}{\partial t^s}A_{h_d}(t)| + \sup_{t\in J}|\frac{\partial^s}{\partial t^s}\Delta_n^{(1)}(t)| + \sup_{t\in J}|\frac{\partial^s}{\partial t^s}\Delta_n^{(2)}(t)|, \tag{A.1}$$

where A_{h_d}, $\Delta_n^{(1)}$ and $\Delta_n^{(2)}$ are defined in (3.4) - (3.6). From (3.5) we get the s-th derivative of $\Delta_n^{(1)}(t)$ as

$$\frac{\partial^s}{\partial t^s}\Delta_n^{(1)}(t) = \int_{-1}^{1} K_d(v)\Big\{\sum_{j=0}^{s}\frac{\partial^j}{\partial t^j}[d\circ m^{-1}](t+h_dv)\frac{\partial^{s-j}}{\partial t^{s-j}}(m^{-1})'(t+h_dv)\Big\}dv,$$

where we again define $d(x)=\hat{m}(x)-m(x)$. Observing that the supremum of the j-th derivative of d is almost surely of order $O(\log h_r^{-1}/nh_r^{2j+1})^{1/2}$ it follows

$$\sup_{t\in J}|\frac{\partial^s}{\partial t^s}\Delta_n^{(1)}(t)| \overset{f.s.}{=} O\Big(\frac{\log h_r^{-1}}{nh_r^{2s+1}}\Big)^{1/2}. \tag{A.2}$$

For the consideration of $\partial^s/\partial t^s\Delta_n^{(2)}(t)$ when $0\le s\le 2$ we use integration by parts in a first step and obtain the representation

$$\begin{aligned}\frac{\partial^s}{\partial t^s}\Delta_n^{(2)}(t) &= -\int_{-1}^{1}K_d(v)\frac{\partial^s}{\partial t^s}\{2d(m^{-1}(t_n))d^{(1)}(m^{-1}(t_n))(m^{-1})'^2(t_n)+d^2(m^{-1}(t_n))(m^{-1})^{(2)}(t_n)\}\\ &= O\Big(\frac{\log h_r^{-1}}{nh_r^{s+2}}\Big) = o\Big(\frac{\log h_r^{-1}}{nh_r^{2s+1}}\Big)^{1/2} \qquad \text{(A.3)}\end{aligned}$$

with $t_n=t+h_dv$. If $s=3$ a different representation is neccessary because m is only four times differentiable. In this case it follows by directly differentiating in representation (3.6)

$$\frac{\partial^3}{\partial t^3}\Delta_n^{(2)}(t) = O\Big(\frac{\log h_r^{-1}}{nh_r^4h_d}\Big) = o\Big(\frac{\log h_r^{-1}}{nh_r^7}\Big)^{1/2}. \tag{A.4}$$

A similar calculation as for (3.10) yields for the deterministic part

$$A_{h_d}(t) = h_d\int_{-1}^{1}vK_d(v)(m^{-1})'(t+h_dv) = h_d^2(m^{-1})^{(2)}(t+h_dv)\kappa_2(K_d)+o(h_d^2). \tag{A.5}$$

For $0\le s\le 2$ we get an estimate of the deterministic part by differentiating s times in (A.5). Therefore the order is

$$\sup_{t\in J}|\frac{\partial^s}{\partial t^s}A_{h_d}(t)| = O(h_d^2). \tag{A.6}$$

If $s=3$ differentiating in (A.5) yields

$$\sup_{t\in J}|\frac{\partial^3}{\partial t^3}A_{h_d}(t)| = o(h_d). \tag{A.7}$$

The assertion of Theorem A.2 finally follows by combining the results (A.1)-(A.4), (A.6) and (A.7). □

Acknowledgements. The authors are grateful to Isolde Gottschlich who typed numerous versions of this paper with considerable technical expertise. The work of the authors was supported by the Sonderforschungsbereich 475, Komplexitätsreduktion in multivariaten Datenstrukturen.

References

Y. Baraud, S. Huet, B. Laurent (2003). Adaptive tests of qualitative hypotheses. ESAIM, Probab. Stat. 7, 147-159

A.W. Bowman, M.C. Jones, I. Gijbels (1998). Testing monotonicity of regression. Journal of Computational and Graphical statistics 7, 489-500.

H.D. Brunk (1955). Maximum likelihood estimates of monotone parameters. Ann. Math. Statist. 26, 607-616.

H. Dette, N. Neumeyer, K.F. Pilz (2006). A simple nonparametric estimator of a monotone regression function. Bernoulli 12, 469-490.

J. Domínguez-Menchero, G. González-Rodríguez, M.J. López-Palomo (2005). An L_2 point of view in testing monotone regression. J. Nonparametric Stat. 17, 135-153.

C. Durot (2003). A Kolmogorov-type test for monotonicity of regression. Stat. Probab. Lett. 63, 425-433.

J. Fan, I. Gijbels (1997). Local polynomial modelling and its applications. Chapman and Hall, London

J. Friedman, R. Tibshirani (1984). The monotone smoothing of scatterplots. Technometrics 26, 243-250.

S. Ghosal, A. Sen, A. W. van der Vaart (2000). Testing monotonicity of regression. Ann. Stat. 28, 1054-1082

I. Gijbels,P. Hall, M.C. Jones, I. Koch (2000). Tests for monotonicity of a regression mean with guaranteed level. Biometrika 87, 663-673.

P. Hall (1984). Integrated square error properties of kernel estimators of regression functions. Ann. Statist. 12, 241-260

P. Hall, N.E. Heckman (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. Ann. Stat. 28, 20-39.

P. Hall, L.-S. Huang (2001). Nonparametric kernel regression subject to monotonicity constraints. Ann. Statist. 29, 624-647.

Y.P. Mack, B.W. Silverman (1982). Weak and strong uniform consistency of kernel regression estimates. Z. Wahrsch. Verw. Gebiete 61, 405-415.

E. Mammen (1991). Estimating a smooth monotone regression function. Ann. Statist. 19, 724-740.

H. Mukerjee (1988). Monotone nonparametric regression. Ann. Statist. 16, 741-750.

J.O. Ramsay (1998). Estimating smooth monotone functions. J. R. Stat. Soc., Ser. B, Stat. Methodol. 60, 365-375.

W. Schlee (1982). Nonparametric tests of the monotony and convexity of regression. Nonparametric statistical inference, Budapest 1980, Vol. II, Colloq. Math. Soc. Janos Bolyai 32, 823-836.

R. J. Serfling (1980). Approximation theorems of mathematical statistics. John Wiley & Sons.

B. W. Silverman (1981). Using kernel density estimates to investigate multimodality. J. Roy. Statist. Soc. Ser. B 43, 97–99.