A martingale-transform goodness-of-fit test for the form of the conditional variance

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: holger.dette@ruhr-uni-bochum.de

Benjamin Hetzler
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
email: benjamin.hetzler@ruhr-uni-bochum.de

April 9, 2008

Abstract

In the common nonparametric regression model the problem of testing for a specific parametric form of the variance function is considered. Recently Dette and Hetzler (2008) proposed a test statistic, which is based on an empirical process of pseudo residuals. The process converges weakly to a Gaussian process with a complicated covariance kernel depending on the data generating process. In the present paper we consider a standardized version of this process and propose a martingale transform to obtain asymptotically distribution free tests for the corresponding Kolmogorov-Smirnov and Cramér-von-Mises functionals. The finite sample properties of the proposed tests are investigated by means of a simulation study.

Keywords and Phrases: nonparametric regression, goodness-of-fit test, martingale transform, conditional variance
AMS Subject Classification: 62G05

1 Introduction

We consider the common nonparametric regression model

$$Y_{i,n} = m(t_{i,n}) + \sigma(t_{i,n})\varepsilon(t_{i,n}), \quad i = 1, \ldots, n, \tag{1.1}$$

where $\varepsilon_{1,1}, \ldots, \varepsilon_{n,n}$ with $\varepsilon_{i,n} := \varepsilon(t_{i,n})$ are assumed to form a triangular array of rowwise independent random variables with mean 0 and variance 1 and m and σ^2 denote the unknown regression and variance function, respectively. In the regression model (1.1) the quantities $0 \leq t_{1,n} < t_{2,n} < \ldots < t_{n,n} \leq 1$ denote the explanatory variables satisfying

$$\frac{i}{n+1} = \int_0^{t_{i,n}} f(t)\, dt, \quad i = 1, \ldots, n, \tag{1.2}$$

where f denotes a positive density on the interval $[0,1]$ [see Sacks and Ylvisacker (1970)]. Because additional information on the variance function such as homoscedasticity can improve the efficiency of the statistical inference, several authors have considered the problem of testing the hypothesis

$$H_0 : \sigma^2(t) = \sigma^2(t, \theta); \quad \forall\, t \in [0,1], \tag{1.3}$$

in the nonparametric regression model (1.1), where $\{\sigma^2(\cdot, \theta) \mid \theta \in \Theta\}$ is a given parametric class of variance functions and $\Theta \subset \mathbb{R}^d$ denotes a finite dimensional parameter space. Most authors consider linear regression models [see e.g. Bickel (1978), Breusch and Pagan (1979), Cook and Weisberg (1983) among others or Pagan and Pak (1993) for a review]. In the nonparametric regression model (1.1) there exist several papers discussing the problem of testing homoscedasticity [see Dette and Munk (1998), Zhu, Fujikoshi and Naito (2001), Dette (2002) or Liero (2003)]. Recently Dette, van Keilegom and Neumeyer (2007) proposed a test for the parametric hypothesis (1.3), which is based on the difference of two empirical processes of standardized nonparametric residuals under the null hypothesis and alternative. Weak convergence of the resulting process is shown and – because the limit distribution is complicated and depends on certain features of the data generating process – the consistency of a smoothed bootstrap procedure is established. Moreover, although the resulting test has nice theoretical and finite sample properties (in particular, it can detect local alternatives converging to the null hypothesis at a rate $n^{-1/2}$) the approach requires rather strong assumptions regarding the differentiability of the variance and regression function. Dette and Hetzler (2008) suggested a procedure, which is, on the one hand, able to detect local alternatives at a rate $n^{-1/2}$ and requires, on the other hand, minimal assumptions regarding the smoothness of the regression and variance function. These authors proposed to estimate the process

$$S_t(w) = \int_0^t \Big(\sigma^2(x) - \sigma^2(x, \theta^*)\Big) \sqrt{w(x)} f(x)\, dx \tag{1.4}$$

using pseudo residuals [see Gasser, Sroka and Jennen-Steinmetz (1986) or Hall, Kay and Titterington (1990)], where

$$\theta^* = \arg\min_{\theta \in \Theta} \int_0^1 \Big(\sigma^2(x) - \sigma^2(x, \theta)\Big)^2 \sqrt{w(x)} f(x)\, dx \tag{1.5}$$

is the parameter corresponding to the best approximation of the function σ^2 by the parametric class $\{\sigma^2(\cdot, \theta \mid \theta \in \Theta\}$ and w denotes a weight function [which was actually chosen as $w \equiv 1$ by Dette and Hetzler (2008)]. Under very weak smoothness assumptions on the regression and variance function they proved weak convergence of the estimated process, say $(\hat{S}_t(w))_{t\in[0,1]}$, to a Gaussian process. The Kolmogorov-Smirnov and Cramér-von-Mises statistic based on $(\hat{S}_t(w))_{t\in[0,1]}$ were proposed for testing the hypothesis (1.3). Because the covariance kernel of the limiting process depends on the data generating process in a complicated way, a bootstrap procedure was applied to obtain the critical values.

It is the purpose of the present paper to construct an asymptotically distribution free test for the parametric form of the variance function which is on the one hand able to detect local alternatives converging to the null hypotheses at a rate $n^{-1/2}$ and on the other hand requires minimal smoothness assumptions. For this purpose we consider a standardized version of the process discussed by Dette and Hetzler (2008), where the weight function is estimated from the data. We apply the martingale transform proposed by Khmaladze (1981, 1989) in order to obtain a distribution free limiting process. This transformation has been used successfully by several authors in goodness-of-fit testing problems for hypotheses regarding the regression function [see Stute, Thies and Zhu (1998), Khmaladze and Koul (2004) or Koul (2006) among others], but to our best knowledge, it has not been studied in the context of testing hypotheses regarding the variance function. In Section 2 we briefly review the main features of the empirical process proposed by Dette and Hetzler (2008) and introduce a standardized version of this process which will be the basis for our test statistic. In Section 3 and 4 we consider the martingale transform and show that the transformed (and standardized) empirical process is asymptotically distribution free. In Section 5 we discuss several examples and investigate the finite sample properties of a Cramér-von-Mises test based on the martingale transformation, while some of the more technical details are deferred to an appendix.

2 The basic process based on pseudo residuals

We assume that the regression function m, the variance function σ^2 in (1.1), the design density f and the weight function w in (1.4) are Lipschitz continuous of order $\gamma > \frac{1}{2}$ and that the moments of order 8 of the errors $\varepsilon_{i,n}$ exist and are uniformly bounded. In general, the moments of order $j \geq 3$ of the errors may depend on the explanatory variables $t_{i,n}$, that is

$$m_j(t_{i,n}) = E\,[\varepsilon_{i,n}^j], \qquad j = 3, \ldots, 8,$$

and the functions m_3 and m_4 are also assumed to be Lipschitz continuous of order $\gamma > \frac{1}{2}$. For the sake of a transparent presentation we consider at the moment linear hypotheses of the form

$$H_0 : \sigma^2(t) = \sum_{j=1}^{d} \theta_j \sigma_j^2(t)\,, \qquad \text{for all } t \in [0,1]\,, \tag{2.1}$$

where $\theta_1, \ldots, \theta_d \in \mathbb{R}$ are unknown parameters and $\sigma_1^2, \ldots, \sigma_d^2$ are given linearly independent functions satisfying

$$\sigma_j^2 \in \operatorname{Lip}_\gamma[0,1], \quad j = 1, \ldots, d. \tag{2.2}$$

The general case of testing hypotheses of the form (1.3) will be briefly discussed at the end of this section. It is is shown in Dette and Hetzler (2008) that the process defined in (1.4) can be consistently estimated by

$$\hat{S}_t(w) = \hat{B}_t^0(w) - \hat{B}_t^T \hat{A}^{-1} \hat{C}(w), \tag{2.3}$$

where the elements of the matrix $\hat{A} = (\hat{a}_{ij})_{1 \le i,j \le d}$ and the vector $\hat{C}(w) = (\hat{c}_1(w), \ldots, \hat{c}_d(w))^T$ are defined by

$$\hat{a}_{ij} = \frac{1}{n} \sum_{k=1}^{n} \sigma_i^2(t_{k,n}) \sigma_j^2(t_{k,n}), \quad 1 \le i,j \le d, \tag{2.4}$$

$$\hat{c}_i(w) = \frac{1}{n-r} \sum_{k=r+1}^{n} R_{k,n}^2 \sqrt{w(t_{k,n})} \sigma_i^2(t_{k,n}), \quad 1 \le i \le d, \tag{2.5}$$

respectively,

$$\hat{B}_t^0(w) = \frac{1}{n-r} \sum_{j=r+1}^{n} 1_{\{t_{j,n} \le t\}} \sqrt{w(t_{j,n})} R_{j,n}^2 \tag{2.6}$$

and $\hat{B}_t = (\hat{B}_t^1, \ldots, \hat{B}_t^d)^T$ with

$$\hat{B}_t^i = \frac{1}{n} \sum_{j=1}^{n} 1_{\{t_{j,n} \le t\}} \sigma_i^2(t_{j,n}), \quad i = 1, \ldots, d\,. \tag{2.7}$$

In (2.5) and (2.6) the quantities $R_{j,n}$ denote pseudo residuals defined by

$$R_{j,n} = \sum_{i=0}^{r} d_i Y_{j-i,n}, \quad j = r+1, \ldots, n, \tag{2.8}$$

where the vector $(d_0, \ldots, d_r)^T \in I\!R^{r+1}$ satisfies

$$\sum_{i=0}^{r} d_i = 0, \quad \sum_{i=0}^{r} d_i^2 = 1 \tag{2.9}$$

and is called difference sequence of order r [see Gasser, Sroka and Jennen-Steinmetz (1986) or Hall, Kay and Titterington (1990) among others]. The following result was proved in Dette and Hetzler (2008) and provides the asymptotic properties of the process $\hat{S}_t(w)$ for an increasing sample size.

Theorem 2.1. *If the conditions stated at the beginning of this section are satisfied, then the process* $\{\sqrt{n}(\hat{S}_t(w) - S_t(w))\}_{t\in[0,1]}$ *converges weakly in* $D[0,1]$ *to a centered Gaussian process with covariance kernel* $k(t_1,t_2)$ *given by the non-diagonal elements of the matrix* $V_2\Sigma_{t_1,t_2}V_2^T \in \mathbb{R}^{2\times 2}$*, where the matrices* $\Sigma_{t_1,t_2} \in \mathbb{R}^{(d+2)\times(d+2)}$ *and* $V_2 \in \mathbb{R}^{2\times(d+2)}$ *are defined by*

$$\Sigma_{t_1,t_2} = \begin{pmatrix} v_{11} & v_{12} & w_{11} & \cdots & w_{1d} \\ v_{21} & v_{22} & w_{21} & \cdots & w_{2d} \\ w_{11} & w_{21} & z_{11} & \cdots & z_{1d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{1d} & w_{2d} & z_{d1} & \cdots & z_{dd} \end{pmatrix}, \tag{2.10}$$

$$V_2 = (I_2|U), \qquad U = -\begin{pmatrix} B_{t_1}^T A^{-1} \\ B_{t_2}^T A^{-1} \end{pmatrix}, \tag{2.11}$$

respectively. The vector B_t^T *is defined by*

$$B_t^T = \left(\int_0^t \sigma_1^2(x) f(x)\, dx, \ldots, \int_0^t \sigma_d^2(x) f(x)\, dx\right), \tag{2.12}$$

the elements of the matrix $A = (a_{ij})_{1\le i,j\le d}$ *are given by*

$$a_{ij} = \int_0^1 \sigma_i^2(x)\sigma_j^2(x) f(x)\, dx, \qquad 1 \le i,j \le d, \tag{2.13}$$

the elements of the matrix in (2.10) are defined by

$$\begin{aligned} v_{ij} &= \int_0^1 \tau_r(s)\sigma^4(s)\, 1_{[0,t_i\wedge t_j)}(s)\, w(s) f(s)\, ds, \qquad 1 \le i,j \le 2, \\ w_{ij} &= \int_0^1 \tau_r(s)\sigma^4(s)\, \sigma_j^2(s)\, 1_{[0,t_i)}(s)\, w(s) f(s)\, ds, \qquad 1 \le i \le 2, 1 \le j \le d, \\ z_{ij} &= \int_0^1 \tau_r(s)\sigma^4(s)\, \sigma_i^2(s)\, \sigma_j^2(s)\, w(s) f(s)\, ds, \qquad 1 \le i,j \le d \end{aligned}$$

with $\tau_r(s) = m_4(s) - 1 + 4\delta_r$*, and the quantity* δ_r *is given by*

$$\delta_r = \sum_{m=1}^{r} \Big(\sum_{j=0}^{r-m} d_j d_{j+m}\Big)^2. \tag{2.14}$$

Note that the null hypothesis (2.1) (or more generally the hypothesis (1.3)) is equivalent to $S_t(w) \equiv 0 \ \ \forall\, t \in [0,1]$, and consequently rejecting (2.1) for large values of the Kolmogorov-Smirnov or Cramér-von-Mises statistic

$$K_n = \sqrt{n} \sup_{t\in[0,1]} |\hat{S}_t(w)|\ , \quad G_n = n \int_0^1 |\hat{S}_t(w)|^2 dF_n(t)$$

yields a consistent test. Here $F_n(t) = \frac{1}{n}\sum_{i=1}^n 1_{\{t_{i,n}\le t\}}$ is the empirical distribution function of the design points. Moreover, it is demonstrated by Dette and Hetzler (2008) that this test can detect alternatives which converge to the null hypothesis with a rate $n^{-1/2}$. Because the limiting distribution depends on certain features of the data generating process, these authors proposed a bootstrap procedure to calculate the critical values.

If $(A\,(t,w))_{t\in[0,1]}$ denotes the limiting process in Theorem 2.1 it follows from the Continuous Mapping Theorem [see Pollard (1984)] that

$$K_n \xrightarrow{\mathcal{D}} \sup_{t\in[0,1]} |A\,(t,w)\,|\;, \quad G_n \xrightarrow{\mathcal{D}} \int_0^1 |A(t,w)|^2 dF(t).$$

Using the Lipschitz continuity of the regression and variance function, it was shown in the proof of Theorem 2.1 that the process $A_n(t,w) = \sqrt{n}(\hat{S}_t(w) - S_t(w))$ exhibits the same asymptotic behaviour as the process

$$\bar{A}_n(t,w) \;=\; C_n(t,w) - D_n(t,w), \tag{2.15}$$

where

$$C_n(t,w) \;=\; \frac{\sqrt{n}}{n-r}\sum_{i=r+1}^{n} 1_{\{t_{i,n}\le t\}}\sqrt{w(t_{i,n})}Z_{i,n}, \tag{2.16}$$

$$D_n(t,w) \;=\; B_t^T A^{-1}\Big(\frac{\sqrt{n}}{n-r}\sum_{i=r+1}^{n} Z_{i,n}\sqrt{w(t_{i,n})}\;\sigma_j^2(t_{i,n})\Big)_{j=1}^{d}, \tag{2.17}$$

the vector $B_t^T = (B_t^1,\dots,B_t^d)$ and the matrix $A = (a_{ij})_{1\le i,j\le d}$ are defined in (2.12) and (2.13), respectively, and the random variables $Z_{i,n}$ are given by $Z_{i,n} = L_{i,n}^2 - \mathrm{E}[L_{i,n}^2]$, with

$$L_{i,n} = \sum_{j=0}^{r} d_j\sigma(t_{i-j,n})\varepsilon_{i-j,n}. \tag{2.18}$$

Because $\{Z_{i,n} \mid i = 1,\dots,n,\; n\in\mathbb{N}\}$ is a triangular array of r-dependent random variables, it follows observing

$$\mathrm{E}\left[Z_{j,n}^2\right] + 2\sum_{m=1}^{r}\mathrm{E}\left[Z_{j,n}Z_{j+m,n}\right] = \left(m_4\left(t_{j,n}\right) - 1 + 4\delta_r\right)\sigma^4\left(t_{j,n}\right) + O\left(n^{-\gamma}\right)$$

[see Dette and Hetzler (2008)] that the process $\{C_n(t,w)\}_{t\in[0,1]}$ converges weakly in D[0,1] to the process $W\circ\psi$, where W denotes a Brownian motion and the function ψ is defined by

$$\psi(t) = \int_0^t \beta(x)w(x)f(x)\,dx \tag{2.19}$$

with

$$\beta(x) = (m_4(x) - 1 + 4\delta_r)\sigma^4(x).$$

Note that the transformation ψ depends on the unknown function β which is not known, because it contains the variance and the fourth moments of the innovations $\varepsilon_{i,n}$. In the following we will use the specific weight function $w(x) = 1/\beta(x)$ for which the function ψ reduces to $\psi(t) = F(t) = \int_0^t f(x)dx$, and for this choice the process $\{C_n(t, 1/\beta)\}_{t\in[0,1]}$ converges weakly to a Brownian motion $W \circ F$. We assume in a first step that the function β is known and investigate the martingale transformation of the standardized process

$$A_n^0(t) = C_n^0(t) - D_n^0(t) \,, \tag{2.20}$$

where $A_n^0(t) = \bar{A}_n(t, 1/\beta)$,

$$C_n^0(t) \;=\; C_n(t, 1/\beta) = \frac{\sqrt{n}}{n-r} \sum_{i=r+1}^{n} 1_{\{t_{i,n}\le t\}} Z_{i,n} \beta^{-1/2}(t_{i,n}), \tag{2.21}$$

$$D_n^0(t) \;=\; D_n(t, 1/\beta) = B_t^T A^{-1} \frac{\sqrt{n}}{n-r} \sum_{i=r+1}^{n} Z_{i,n} g(t_{i,n}) \beta^{-1/2}(t_{i,n}), \tag{2.22}$$

$g(x) = (\sigma_1^2(x), \ldots, \sigma_d^2(x))^T$. In a second step we will estimate the function β nonparametrically and consider the corresponding processes standardized by this estimate. More precisely, we will show that the corresponding martingale transform of the process

$$\hat{\Gamma}_t = \hat{S}_t(1/\hat{\beta}) = \hat{B}_t^0(1/\hat{\beta}) - \hat{B}_t^T \hat{A}^{-1} \hat{C}(1/\hat{\beta}) \tag{2.23}$$

leads to an asymptotically distribution free test. Here $\hat{\beta}$ is an appropriate estimate of the function β and $\hat{C}(w)$, $\hat{B}_t^0(w)$ and $\hat{B}_t$ are defined in (2.5), (2.6) and (2.7), respectively.

Remark 2.2. For the problem of testing a general nonlinear hypothesis of the form (1.3) we propose to consider the process

$$\hat{S}_t(w) = \hat{B}_t^0(w) - \frac{1}{n} \sum_{i=1}^{n} 1_{\{t_{i,n}\le t\}} \sigma^2(t_{i,n}, \hat{\theta}) \sqrt{w(t_{i,n})},$$

where

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \frac{1}{n-r} \sum_{i=r+1}^{n} \Big(R_{i,n}^2 - \sigma^2(t_{i,n}, \theta) \Big)^2 \sqrt{w(t_{i,n})}$$

is the least squares estimate of the parameter θ^* defined by (1.5). In this case it was shown by Dette and Hetzler (2008) that under assumptions of regularity the process $\{\sqrt{n}(\hat{S}_t(w) - S_t(w))\}_{t\in[0,1]}$ exhibits the same asymptotic behaviour as described in Theorem 2.1 for the linear case, where the functions σ_j^2 have to be replaced by

$$\sigma_j^2(t) = \frac{\partial}{\partial \theta_j} \sigma^2(t, \theta)\Big|_{\theta=\theta_0}, \quad j = 1, \ldots, d.$$

Thus all results presented in the following section can be transferred to the nonlinear case using this identification.

3 The martingale transform of the process A_n^0

It follows by similar arguments as given in Dette and Hetzler (2008) that the process $\{A_n^0(t)\}_{t\in[0,1]}$ defined by (2.20) converges weakly in $D[0,1]$, that is

$$A_n^0 \xrightarrow{\mathcal{D}} W \circ F - B_t^T A^{-1} V_0 = \tilde{R}_\infty^0, \tag{3.1}$$

where W is a Brownian motion and V_0 denotes a centered normal random variable with mean 0 and covariance matrix

$$L = \int_0^1 g(x) g^T(x) f(x) dx.$$

Because the distribution of the process $\tilde{R}_\infty^0$ is complicated, we consider in the following section an operator, which transforms the process $\tilde{R}_\infty^0$ on the martingale part in its corresponding Doob-Meyer decomposition. Following Khmaladze and Koul (2004) we define a linear operator T such that

$$TR_\infty^0 \stackrel{\mathcal{D}}{=} R_\infty^0 , \tag{3.2}$$

$$T(B_t^T A^{-1} V_0) \equiv 0 , \tag{3.3}$$

where the symbol $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution and the process R_∞^0 is given by $R_\infty^0 = W \circ F$. For this purpose we consider the matrix

$$H(t) = \int_t^1 g(u) g^T(u) f(u)\, du \tag{3.4}$$

and define for a function η its transformation $T\eta$ by

$$(T\eta)(t) = \eta(t) - \int_0^t g^T(y) H^{-1}(y) \int_y^1 g(z) \eta(dz) F(dy), \tag{3.5}$$

where only functions are considered such that the integral on the right hand side of (3.5) exists. Note that the matrix $H(x)$ is non-singular for all $x \in [0,1)$ because the functions $\sigma_1^2, \dots, \sigma_d^2$ are linearly independent; see Achieser (1956). If η is a stochastic process on the interval [0,1], the corresponding integral in (3.5) is interpreted as an Ito-integral [see Øksendal (2003)]. A straightforward calculation shows that

$$T(B_t^T A^{-1} V_0) = 0 ,$$
$$\operatorname{Cov}(TR_\infty(r), TR_\infty(s)) = F(r \wedge s) ,$$

which yields for the process defined on the right hand side of (3.1)

$$T\tilde{R}_\infty^0 \stackrel{\mathcal{D}}{=} TR_\infty^0 \stackrel{\mathcal{D}}{=} R_\infty^0 \stackrel{\mathcal{D}}{=} W \circ F \tag{3.6}$$

(note that $\tilde{R}^0_\infty$ is a Gaussian process and that the operator T is linear). The following theorem shows that a similar property holds in an asymptotic sense for the process $\{A^0_n(t)\}_{t\in[0,1]}$.

Theorem 3.1. *If the assumptions stated in Section 2 are satisfied, then the transformed process* $\{TA^0_n(t)\}_{t\in[0,1]}$ *converges weakly in* $D[0,1]$ *to a Brownian motion in time* F*, that is*

$$\{TA^0_n(t)\}_{t\in[0,1]} \xrightarrow{\mathcal{D}} \{W\circ F(t)\}_{t\in[0,1]}.$$

Proof. The assertion of the theorem follows from the statements

$$TA^0_n \;=\; TC^0_n, \tag{3.7}$$

$$\{TC^0_n(t)\}_{t\in[0,1]} \;\xrightarrow{\mathcal{D}}\; \{W\circ F(t)\}_{t\in[0,1]}. \tag{3.8}$$

For a proof of (3.7) we recall the notation $D^0_n = A^0_n - C^0_n$ in (2.22) and obtain by a straightforward calculation from the definitions (3.5) and (2.22)

$$\begin{aligned}
TD^0_n(t) &= TA^0_n(t) - TC^0_n(t) \\
&= D^0_n(t) - \int_0^t g^T(y)H^{-1}(y)\int_y^1 g(z)g^T(z)F(dz)F(dy)A^{-1} \\
&\quad \times \Big(\frac{\sqrt{n}}{n-r}\sum_{k=r+1}^n Z_{k,n}\beta^{-1/2}(t_{k,n})g(t_{k,n})\Big) \\
&= D^0_n(t) - \int_0^t g^T(y)H^{-1}(y)H(y)F(dy)A^{-1}\Big(\frac{\sqrt{n}}{n-r}\sum_{k=r+1}^n Z_{k,n}\beta^{-1/2}(t_{k,n})g(t_{k,n})\Big) \\
&= D^0_n(t) - B_t^TA^{-1}\Big(\frac{\sqrt{n}}{n-r}\sum_{k=r+1}^n Z_{k,n}\beta^{-1/2}(t_{k,n})g(t_{k,n})\Big) \;=\; 0.
\end{aligned} \tag{3.9}$$

The process TC^0_n is a sum of r-dependent random variables. Therefore, weak convergence of the finite dimensional distributions and tightness can be shown using similar arguments as in Dette and Hetzler (2008). Thus the assertion follows showing that the covariance kernel of the limiting process is given by $F(s\wedge t)$. For the calculation of the asymptotic covariances we use the representation

$$TC^0_n(t) \;=\; \frac{\sqrt{n}}{n-r}\sum_{i=r+1}^n C_{i,n}(t), \tag{3.10}$$

where

$$\begin{aligned}
C_{i,n}(t) &= 1_{\{t_{i,n}\le t\}}Z_{i,n}\beta^{-1/2}(t_{i,n}) - \int_0^t g^T(y)H^{-1}(y)1_{\{t_{i,n}\ge y\}}g(t_{i,n})Z_{i,n}\beta^{-1/2}(t_{i,n})F(dy) \\
&= C^{(1)}_{i,n}(t) - C^{(2)}_{i,n}(t)
\end{aligned} \tag{3.11}$$

and the last line defines the random variables $C_{i,n}^{(1)}(t)$ and $C_{i,n}^{(2)}(t)$ in an obvious manner. Observing that

$$\mathrm{E}[Z_{i,n}^2] + 2\sum_{m=1}^{r} \mathrm{E}[Z_{i,n}Z_{i+m,n}] = \beta(t_{i,n}) + O(n^{-\gamma}) \tag{3.12}$$

[see Dette and Hetzler (2008)], it follows for $r \le s$

$$\begin{aligned} &\mathrm{E}[C_{i,n}^{(1)}(r)C_{i,n}^{(1)}(s)] + 2\sum_{m=1}^{r} \mathrm{E}[C_{i,n}^{(1)}(r)C_{i+m,n}^{(1)}(s)] \\ &= 1_{\{t_{i,n}\le r\}} E[Z_{i,n}^2]\beta^{-1}(t_{i,n}) + 2\sum_{m=1}^{r} 1_{\{t_{i,n}\le r\}}\,\mathrm{E}[Z_{i,n}Z_{i+m,n}]\beta^{-1}(t_{i,n}) + o(1) = 1_{\{t_{i,n}\le r\}} + o(1)\,. \end{aligned}$$

This implies

$$\frac{n}{(n-r)^2}\sum_{i=r+1}^{n-r} \mathrm{E}[C_{i,n}^{(1)}(r)C_{i,n}^{(1)}(s)] + 2\sum_{m=1}^{r} \mathrm{E}[C_{i,n}^{(1)}(r)C_{i+m,n}^{(1)}(s)] = F(r) + o(1)\,,$$

and similar arguments show

$$\begin{aligned} &\frac{n}{(n-r)^2}\sum_{i=r+1}^{n-r}\Big(\mathrm{E}[C_{i,n}^{(1)}(r)C_{i,n}^{(2)}(s)] + 2\sum_{m=1}^{r}\mathrm{E}[C_{i,n}^{(1)}(r)C_{i+m,n}^{(2)}(s)]\Big) \\ &= \int_0^s g^T(y)H^{-1}(y)\int_y^r g(x)F(dx)F(dy) + o(1), \end{aligned}$$

$$\begin{aligned} &\frac{n}{(n-r)^2}\sum_{i=r+1}^{n-r}\Big(\mathrm{E}[C_{i,n}^{(2)}(r)C_{i,n}^{(2)}(s)] + 2\sum_{m=1}^{r}\mathrm{E}[C_{i,n}^{(2)}(r)C_{i+m,n}^{(2)}(s)]\Big) \\ &= \int_0^r\int_0^s g^T(y_1)H^{-1}(y_1)H(y_1\vee y_2)H^{-1}(y_2)g(y_2)F(dy_2)F(dy_1) + o(1)\,. \end{aligned}$$

A combination of these results and an application of Fubini's theorem yield

$$\begin{aligned} \mathrm{E}[TC_n^0(r)TC_n^0(s)] &= \frac{n}{(n-r)^2}\sum_{i=r+1}^{n-r}\Big(\mathrm{E}[C_{i,n}(r)C_{i,n}(s)] + 2\sum_{m=1}^{r}\mathrm{E}[C_{i,n}(r)C_{i+m,n}(s)]\Big) + o(1) \\ &= F(r) + \int_0^s g^T(y)H^{-1}(y)\int_y^r g(x)F(dx)F(dy) \\ &\quad + \int_0^r g^T(y)H^{-1}(y)\int_y^s g(x)F(dx)F(dy) \\ &\quad + \int_0^r\int_0^s g^T(y_1)H^{-1}(y_1)H(y_1\vee y_2)H^{-1}(y_2)g(y_2)F(dy_2)F(dy_1) + o(1)\,, \\ &= F(r) + o(1)\,, \end{aligned}$$

which implies the assertion of the theorem. □

4 The martingale transform of the process $\{\hat{\Gamma}_t\}_{t\in[0,1]}$

As pointed out in Section 2, the process $\{\sqrt{n}(\hat{S}_t(1/\beta) - S_t(1/\beta))\}_{t\in[0,1]}$ (or its asymptotically equivalent counterpart $\{A_n^0(t)\}_{t\in[0,1]}$) depends on the unknown function β (more precisely on the (unknown) functions $\sigma^2(\cdot)$ and $m_4(\cdot)$). Similarly, the operator T defined by (3.5) is not completely known and has to be estimated from the data. In this section we propose an empirical process, where the unknown quantities have been replaced by estimates and study the application of an empirical version of the martingale transform. For this purpose we first have to specify the estimate in the process $\{\hat{\Gamma}_t\}_{t\in[0,1]}$ defined in (2.23). We consider the Nadaraya-Watson weights

$$(4.1) \qquad w_{ij} = \frac{K\left(\frac{t_{j,n}-t_{i,n}}{h}\right)}{\sum_{l=1}^{n} K\left(\frac{t_{l,n}-t_{i,n}}{h}\right)}, \qquad i,j = 1,\ldots,n,$$

at the points $t_{i,n}$ $(i = 1,\ldots,n)$ where K denotes a symmetric kernel function and h defines a bandwidth converging to 0 with increasing sample size. The estimate of the function $\beta(\cdot)$ is now defined by

$$\begin{aligned}\hat{\beta}(t_{i,n}) \;&=\; \sum_{j=1}^{n} w_{ij}(Y_{j,n} - \hat{m}_h(t_{j,n}))^4 \\ (4.2) \qquad &+(4\delta_r - 1)\sum_{j=1}^{n-r-1} w_{ij}(Y_{j,n} - \hat{m}_h(t_{j,n}))^2(Y_{j+r+1,n} - \hat{m}_h(t_{j+r+1,n}))^2,\end{aligned}$$

where $\hat{m}_h(t_{i,n}) = \sum_{j=1}^{n} w_{ij}Y_{j,n}$ denotes the Nadaraya-Watson estimate at the point $t_{i,n}$ $(i = 1,\ldots,n)$. Throughout this paper we assume that

(H) The bandwidth h satisfies $h = h_n = O(n^{-\frac{1}{2\gamma+1}})$, where $\gamma > \frac{1}{2}$ denotes the Lipschitz constant defined in Section 2.

(K) The kernel K is symmetric, nonnegative, supported on the interval $[-1,1]$ and satisfies $K(u) \le 1$ for all $u \in [-1,1]$ and $K(u) \ge \kappa$ for all $\|u\| \le 1/2$, where $\kappa > 0$.

It will be proved in the appendix that under these additional assumptions

$$(4.3) \qquad \sup_{t\in[0,t_0]} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} 1_{\{t_{i,n}\le t\}} Z_{i,n}\{\hat{\beta}(t_{i,n}) - \beta(t_{i,n})\} \Big| = o_p(1)\,,$$

and similar arguments as given in Dette and Hetzler (2008) show that

$$(4.4) \qquad \Lambda_n(t) = \sqrt{n}\left(\hat{\Gamma}_t - S_t(1/\beta)\right) = A_n^1(t) + o_p(1)$$

uniformly with respect to $t \in [0,1]$. In this representation, $\{A_n^1(t)\}_{t\in[0,1]}$ denotes the process obtained from $\{A_n^0(t)\}_{t\in[0,1]}$ by replacing $\beta(t)$ by its estimate $\hat{\beta}(t)$ defined in (4.2) and the vector B_t and the matrix A by their estimates $\hat{B}_t$ and $\hat{A}$ defined in (2.4) and (2.7), that is

$$A_n^1(t) = C_n^1(t) - D_n^1(t), \tag{4.5}$$

where

$$\begin{aligned} C_n^1(t) &= \frac{\sqrt{n}}{n-r} \sum_{i=r+1}^{n} 1_{\{t_{i,n}\le t\}} Z_{i,n} \hat{\beta}^{-1/2}(t_{i,n}) \,, \\ D_n^1(t) &= \hat{B}_t^T \hat{A}^{-1} \frac{\sqrt{n}}{n-r} \sum_{i=r+1}^{n} Z_{i,n} g(t_{i,n}) \hat{\beta}^{-1/2}(t_{i,n}) \,. \end{aligned} \tag{4.6}$$

Similarly, we replace the operator T by its empirical version defined by

$$(T_n\eta)(t) = \eta(t) - \int_0^t g^T(y) H_n^{-1}(y) \int_y^1 g(z)\eta(dz) F_n(dy), \tag{4.7}$$

where the matrix $H_n(x)$ is given by

$$H_n(x) = \int_x^1 g(u) g^T(u) F_n(du) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{t_{i,n}\ge x\}} g(t_{i,n}) g^T(t_{i,n}) \tag{4.8}$$

and $F_n(t) = \frac{1}{n}\sum_{i=1}^n 1_{\{t_{i,n}\le t\}}$ denotes the empirical distribution function of the design points. Note that the matrix $H(x)$ used in the transformation (3.5) is singular at the point $x=1$, and as a consequence, the matrices $H_n^{-1}(x)$ are unbounded on the whole interval $[0,1]$. To circumvent this difficulty, we restrict the process $T_nA_n^1$ to the interval $[0,t_0]$ with a fixed $0 < t_0 < 1$. This approach was also suggested by Khmaladze (1989) and Stute, Thies and Zhu (1998) among others.
The following results show that the asymptotic properties of the processes $\{TA_n^0(t)\}_{t\in[0,t_0]}$ and $\{T_nA_n^1(t)\}_{t\in[0,t_0]}$ coincide, and as a consequence we obtain weak convergence of the martingale transform of the process defined on the left hand side of (4.4).

Theorem 4.1. *If the assumptions stated at the beginning of Section 2 and the assumptions (H) and (K) are satisfied, then for any* $0 < t_0 < 1$ *the process* $\{T_nA_n^1(t)\}_{t\in[0,t_0]}$ *converges weakly on* $D[0,t_0]$ *to a Brownian motion in time* F*, that is*

$$\{T_nA_n^1(t)\}_{t\in[0,t_0]} \xrightarrow{\mathcal{D}} \{W\circ F(t)\}_{t\in[0,t_0]}$$

Corollary 4.2. *If the assumptions of Theorem 3.2 are satisfied, then for any* $0 < t_0 < 1$ *the process* $\{T_n\Lambda_n(t)\}_{t\in[0,t_0]}$ *converges weakly on* $D[0,t_0]$ *to a Brownian motion in time* F*, that is*

$$\{T_n\Lambda_n(t)\}_{t\in[0,t_0]} \xrightarrow{\mathcal{D}} \{W\circ F(t)\}_{t\in[0,t_0]}.$$

Proof of Theorem 4.1. Obviously the assertion follows from the statement

$$\sup_{t\in[0,t_0]} |TA_n^0(t) - T_nA_n^1(t)| = o_p(1)\ . \tag{4.9}$$

In order to prove the estimate (4.9) we note that (using the notation $D_n^1 = A_n^1 - C_n^1$)

$$\begin{aligned}
(T_nA_n^1 - T_nC_n^1)(t) &= D_n^1(t) - \int_0^t g^T(y)H_n^{-1}(y)\int_y^1 g(z)D_n^1(dz)F_n(dy) \\
&= D_n^1(t) - \int_0^t g^T(y)F_n(dy)\hat{A}^{-1}\Big(\frac{\sqrt{n}}{n-r}\sum_{i=r+1}^n Z_{i,n}g(t_{i,n})\hat{\beta}^{-1/2}(t_{i,n})\Big) \\
&= D_n^1(t) - \hat{B}_t\hat{A}^{-1}\frac{\sqrt{n}}{n-r}\sum_{i=r+1}^n Z_{i,n}g(t_{i,n})\hat{\beta}^{-1/2}(t_{i,n}) = 0.
\end{aligned}$$

Consequently (observing the corresponding result for $TA_n^0 - TC_n^0$ in (3.9)), the assertion follows if the statement

$$\sup_{t\in[0,t_0]} |TC_n^0(t) - T_nC_n^1(t)| = o_p(1) \tag{4.10}$$

can be proved, where

$$\begin{aligned}
TC_n^0(t) &= C_n^0(t) - \int_0^t g^T(y)H^{-1}(y)\int_y^1 g(z)C_n^0(dz)F(dy) = C_n^0(t) - B_n^0(t), \\
T_nC_n^1(t) &= C_n^1(t) - \int_0^t g^T(y)H_n^{-1}(y)\int_y^1 g(z)C_n^1(dz)F_n(dy) = C_n^1(t) - B_n^1(t),
\end{aligned}$$

C_n^0 and C_n^1 are defined in (2.21) and (4.6), respectively, and the equalities define the processes B_n^0 and B_n^1 in an obvious manner. It follows by a Taylor expansion, by the estimate (4.3) and the estimate

$$\sup_{t\in[0,t_0]} |\hat{m}_h(t) - m(t)| = O_p\left(n^{-\frac{\gamma}{2\gamma+1}}\sqrt{\log n}\right) \tag{4.11}$$

[see Mack and Silverman (1982)] that

$$\sup_{t\in[0,t_0]} |C_n^1(t) - C_n^0(t)| = o_p(1). \tag{4.12}$$

We now consider the remaining difference

$$\begin{aligned}
B_n^1(t) - B_n^0(t) &= \int_0^t g^T(y)(H_n^{-1}(y) - H^{-1}(y))(\hat{U}_n(y) - U_n(y))F_n(dy) \\
&\quad + \int_0^t g^T(y)(H_n^{-1}(y) - H^{-1}(y))U_n(y)F_n(dy) \\
&\quad + \int_0^t g^T(y)H^{-1}(y)(\hat{U}_n(y) - U_n(y))F_n(dy) \\
&\quad + \int_0^t g^T(y)H^{-1}(y)U_n(y)F_n(dy) - \int_0^t g^T(y)H^{-1}(y)U_n(y)F(dy)\ ,
\end{aligned} \tag{4.13}$$

where we have used the notation

$$\hat{U}_n(y) = \int_y^1 g(z) C_n^1(dz) \, , \qquad U_n(y) = \int_y^1 g(z) C_n^0(dz) \, .$$

The five terms in this expression are estimated separately. The second term is bounded by

$$\sup_{y \in [0,t_0]} \|H_n^{-1}(y) - H^{-1}(y)\| T_{n1},$$

where

$$\begin{aligned} T_{n1} &:= \int_0^t \|g^T(y)\| \|U_n(y)\| F_n(dy) \\ &\le T_{n11} := \frac{1}{n} \sum_{i=1}^n \Big[\|g^T(t_{i,n})\| \| \frac{\sqrt{n}}{n-r} \sum_{j=r+1}^n 1_{\{t_{j,n} \ge t_{i,n}\}} g(t_{j,n}) Z_{j,n} \beta^{-1/2}(t_{j,n}) \| \Big], \end{aligned}$$

$\| \cdot \|$ denotes the euclidean norm on $\mathbb{R}^d$ and its induced matrix norm on $\mathbb{R}^{d \times d}$ simultaneously, and we have used the definition of $U_n(y)$. A straightforward application of the Cauchy-Schwarz inequality shows

$$\mathrm{E}\, T_{n11} \le \frac{1}{n} \sum_{i=1}^n \|g^T(t_{i,n})\| \Big(\mathrm{E}\, \| \frac{\sqrt{n}}{n-r} \sum_{j=r+1}^n 1_{\{t_{j,n} \ge t_{i,n}\}} g(t_{j,n}) Z_{j,n} \beta^{-1/2}(t_{j,n}) \|^2 \Big)^{1/2} = O(1)$$

uniformly with respect to $t \in [0, t_0]$, which implies $T_{n1} = O_p(1)$ uniformly on the interval $[0, t_0]$. From the assumption (1.2) on the design it follows that

$$\sup_{y \in [0,t_0]} \|H_n^{-1}(y) - H^{-1}(y)\| = o(1), \tag{4.14}$$

and we obtain that the second term in (4.13) is of order $o_p(1)$. Using similar arguments if follows that the first and third term on the right hand side of (4.13) are also of order $o_p(1)$. For the estimate of the remaining difference we show the estimate

$$\sup_{t \in [0,t_0]} \Big| \int_0^t g^T(y) H^{-1}(y) U_n(y) F_n(dy) - \int_0^t g^T(y) H^{-1}(y) U_n(y) F(dy) \Big| = o_p(1) \tag{4.15}$$

using Lemma 6.6.4 in Koul (2002). Note that for the application of this result one has to show the tightness of the process $\{U_n(x)\}_{x \in [0,t_0]}$. For this purpose we consider the components of U_n separately, that is

$$U_n^{(p)}(x) = \frac{\sqrt{n}}{n-r} \sum_{i=r+1}^n 1_{\{t_{i,n} \ge x\}} \sigma_p^2(t_{i,n}) \beta^{-1/2}(t_{i,n}) Z_{i,n}, \qquad p = 1, \ldots, d,$$

and introduce the notation

$$\nu_p(x) = 1_{\{y_1 \le x \le y_2\}} \sigma_p^2(x) \beta^{-1/2}(x).$$

Now a similar calculation as in Dette and Hetzler (2008) yields

$$\mathrm{E}[(U_n^{(p)}(y_2) - U_n^{(p)}(y_1))^4] \;=\; \frac{n^2}{(n-r)^4}\,\mathrm{E}\Big[\Big(\sum_{i=r+1}^{n} \nu_p(t_{i,n}) Z_{i,n}\Big)^4\Big] \le C(y_2 - y_1)^2$$

for some constant $C > 0$ and $0 \le y_1 \le y_2 \le t_0$. This implies tightness of each component $U_n^{(p)}$ [see Billingsley (1999)] and as a consequence tightness of the process U_n [see Billingsley (1979)]. □

Remark 4.3. Theorem 4.1 and Corollary 4.2 remain correct if the Nadaraya-Watson weights in the estimate $\hat{\beta}$ defined in (4.2) are replaced by local linear weights. This follows by a careful inspection of the proof of the estimate (4.3) in the appendix. In practical applications the use of local linear weights is strictly recommended because of the better performance of the local linear estimate at the boundary of the design space.

5 Finite sample properties

In this section we investigate the finite sample properties of the new test by means of a simulation study. We have generated data according to the model

(5.1) $$Y_{i,n} = 1 + t_{i,n} + \sigma(t_{i,n})\varepsilon_{i,n}, \quad i = 1, \dots, n,$$

where $t_{i,n} = i/(n+1), \quad i = 1, \dots, n$, and simulated the power of the test for the hypothesis

(5.2) $$H_0: \ \sigma^2(t) = 1 + \theta t^2$$

and the variance functions

(5.3) $$\sigma^2(t) \;=\; 1 + 3t^2 + 2.5c\sin(2\pi t),$$

(5.4) $$\sigma^2(t) \;=\; 1 + 3t^2 + 2ce^{2t},$$

(5.5) $$\sigma^2(t) \;=\; 1 + 3t^2 + 4c\sqrt{t}.$$

Note that the choice $c = 0$ in (5.3) - (5.5) corresponds to the null hypothesis of a quadratic variance function. The errors $\varepsilon_{i,n}$ are standard normal distributed and we use a difference sequence of order $r = 1$ for the calculation of the pseudo residuals $R_{i,n}$, which determines the weights as $d_0 = -d_1 = 1/\sqrt{2}$ and yields $\beta(x) = m_4(x)\sigma^4(x)$. In order to apply the test we have to calculate the transformation

$$T_n\Lambda_n(t) = T_n(\sqrt{n}(\hat{S}_t(1/\hat{\beta}) - S_t(1/\beta)))$$

for the process $\Lambda_n(t)$ given in (4.4). Under the null hypothesis (5.2) we have $S_t(1/\beta) = 0$ for all $t \in [0, 1]$, and the process $\Lambda_n(t)$ can be written as

$$\Lambda_n(t) = \sqrt{n}\,\hat{S}_t(1/\hat{\beta}) = \hat{C}_n(t) - \hat{D}_n(t)$$

with

$$\hat{C}_n(t) = \frac{\sqrt{n}}{n-r} \sum_{i=r+1}^{n} 1_{\{t_{i,n} \le t\}} \hat{\beta}^{-1/2}(t_{i,n}) R_{i,n}^2,$$

$$\hat{D}_n(t) = \hat{B}_t^T \hat{A}^{-1} \frac{\sqrt{n}}{n-r} \sum_{i=r+1}^{n} \hat{\beta}^{-1/2}(t_{i,n}) R_{i,n}^2 g(t_{i,n}).$$

By a similar argument as given in the proof of Theorem 4.1 it can be shown that $T_n \hat{D}_n(t) = 0$ for all $t \in [0,1]$, and as a consequence it is sufficient to calculate the transformation $T_n \hat{C}_n$. We use the Cramér-von-Mises statistic $G_n = \int_0^1 (T_n \Lambda_n)^2(t) dF_n(t)$, and from Corollary 4.2 and the Continuous Mapping Theorem it follows that

$$G_n = \int_0^1 (T_n \Lambda_n)^2(t)\, dF_n(t) \xrightarrow{\mathcal{D}} \int_0^1 W^2(F(t))\, dF(t) = \int_0^1 W^2(t)\, dt, \tag{5.6}$$

where W denotes a standard Brownian motion. If w_α denotes the $1-\alpha$ quantile of the distribution of the random variable $\int_0^1 W^2(t)dt$, then the test, which rejects the null hypothesis (5.2) if

$$G_n \ge w_\alpha \tag{5.7}$$

has asymptotically level α and is consistent against local alternatives converging to the null hypothesis at a rate $n^{-1/2}$. As an estimator of the function $\beta(x) = m_4(x)\sigma^4(x)$ we use the estimator (4.2), where $\hat{m}_h(\cdot)$ is the local linear estimator of the regression function. The bandwidth for the calculation of the local linear estimate was determined by least squares cross validation. If h_{CV} is the bandwidth obtained by this procedure, the bandwidth in the estimator (4.2) was chosen as $h_{CV}/2$.

1000 simulation runs were performed in each scenario to calculate the rejection probabilities, which are shown in Table 5.1. For the sake of comparison, the table also contains the corresponding rejection probabilities of the bootstrap test proposed by Dette and Hetzler (2008), which are displayed in brackets. We observe a rather precise approximation of the nominal level in all cases, even for the sample size $n = 50$. Under the alternatives the behaviour of the two tests is different. In model (5.3) the bootstrap test yields a substantially larger power than the test based on the martingale transformation. In model (5.5) the situation is similar for the sample size $n = 50$, but the differences between the rejection probabilities of the two steps are smaller. In model (5.4) the bootstrap test is more powerful for the alternative corresponding to $c = 0.5$, while for $c = 1.0$ the test proposed in this paper yields better rejection probabilities. On the other hand, if the sample size is $n = 100$ or $n = 200$ the test based on the martingale transformation always yields a larger power than the bootstrap test.

		$n = 50$			$n = 100$			$n = 200$		
	c	.025	.05	.10	.025	.05	.10	.025	.05	.10
(5.3)	0	.028	.040	.085	.014	.037	.086	.021	.038	.090
		(.025)	(.049)	(.105)	(.024)	(.038)	(.092)	(.024)	(.055)	(.107)
	0.5	.063	.132	.214	.171	.264	.415	.420	.553	.724
		(.316)	(.382)	(.481)	(.458)	(.528)	(.613)	(.661)	(.732)	(.811)
	1	.101	.194	.319	.277	.404	.582	.680	.823	.935
		(.655)	(.726)	(.777)	(.855)	(.893)	(.928)	(.975)	(.984)	(.996)
(5.4)	0	.026	.046	.083	.020	.033	.070	.021	.049	.094
		(.029)	(.049)	(.099)	(.024)	(.043)	(.103)	(.023)	(.049)	(.088)
	0.5	.110	.194	.331	.343	.507	.676	.780	.894	.958
		(.182)	(.253)	(.331)	(.243)	(.331)	(.434)	(.371)	(.450)	(.560)
	1	.226	.349	.525	.582	.732	.862	.878	.947	.987
		(.242)	(.331)	(.413)	(.286)	(.388)	(.491)	(.491)	(.610)	(.729)
(5.5)	0	.036	.054	.096	.017	.032	.085	.018	.053	.104
		(.035)	(.059)	(.099)	(.030)	(.054)	(.111)	(.020)	(.034)	(.073)
	0.5	.079	.141	.242	.224	.334	.500	.508	.649	.788
		(.189)	(.249)	(.340)	(.259)	(.335)	(.439)	(.402)	(.493)	(.622)
	1	.144	.268	.412	.415	.568	.732	.859	.931	.973
		(.316)	(.393)	(.508)	(.487)	(.596)	(.678)	(.713)	(.788)	(.856)

Table 5.1. *Rejection probabilities of the Cramér-von-Mises test (5.7) for the hypothesis (5.2) in the regression model (5.1). The corresponding rejection probabilities of the bootstrap test proposed by Dette and Hetzler (2008) are displayed in brackets.*

Acknowledgements. The work of the authors was supported by the Deutsche Forschungsgemeinschaft (SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen) and in part by a NIH grant award IR01GM072876:01A1. The authors are also grateful to Martina Stein, who typed parts of this paper with considerable expertise.

6 Appendix: Proof of (4.3)

Throughout this section we omit the index n; in particular we write t_j and Z_j instead of $t_{j,n}$ and $Z_{j,n}$, respectively. For the sake of brevity we only indicate the main steps of the proof, details can be found in Hetzler (2008). Furthermore we restrict ourselves to the case $\sigma \equiv 1$ and $r = 1$ and note that the general case is proved exactly in the same way with some additional notation. This simplification yields for the random variables Z_i

$$Z_i = d_0\sigma(t_i)\varepsilon_i + d_1\sigma(t_{i-1})\varepsilon_{i-1} = \frac{\varepsilon_i - \varepsilon_{i-1}}{\sqrt{2}}.$$

A straightforward calculation gives

$$A(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} 1_{\{t_i \le t\}} Z_i \{\hat{\beta}(t_i) - \beta(t_i)\} = \sum_{j=1}^{5} A_j(t), \tag{6.1}$$

where

$$\begin{aligned}
A_1(t) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} 1_{\{t_i \le t\}} Z_i \sum_{j=1}^{n} w_{ij}(\varepsilon_j^4 - m_4(t_i)), \\
A_2(t) &= \frac{4}{\sqrt{n}} \sum_{i=1}^{n} 1_{\{t_i \le t\}} Z_i \sum_{j=1}^{n} w_{ij}\varepsilon_j^3(m(t_j) - \hat{m}_h(t_j)), \\
A_3(t) &= \frac{6}{\sqrt{n}} \sum_{i=1}^{n} 1_{\{t_i \le t\}} Z_i \sum_{j=1}^{n} w_{ij}\varepsilon_j^2(m(t_j) - \hat{m}_h(t_j))^2, \\
A_4(t) &= \frac{4}{\sqrt{n}} \sum_{i=1}^{n} 1_{\{t_i \le t\}} Z_i \sum_{j=1}^{n} w_{ij}\varepsilon_j(m(t_j) - \hat{m}_h(t_j))^3, \\
A_5(t) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} 1_{\{t_i \le t\}} Z_i \sum_{j=1}^{n} w_{ij}(m(t_j) - \hat{m}_h(t_j))^4.
\end{aligned}$$

We rewrite $m(t_j) - \hat{m}_h(t_j) = \rho_j - \sum_{k=1}^{n} w_{jk}\varepsilon_k$ with

$$\rho_j := m(t_j) - \sum_{k=1}^{n} w_{jk} m(t_k) = \sum_{k=1}^{n} w_{jk}(m(t_j) - m(t_k)) \tag{6.2}$$

and first consider the term A_1. For its expectation we have

$$\mathrm{E}\, A_1(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} 1_{\{t_i \le t\}}\, \mathrm{E}[Z_i \sum_{j=1}^{n} w_{ij} h_{ij}],$$

where we used the notation $h_{ij} := \varepsilon_j^4 - m_4(t_i)$. Note that $|\,\mathrm{E}\, h_{ij}| = |m_4(t_j) - m_4(t_i)| \le Lh^\gamma$ whenever $|t_j - t_i| \le h$ (recall the Hölder continuity for the function m_4) and that it follows from the assumption on the design and the kernel

$$\frac{K_h(t_j - t_i)}{C_2 n} \le w_{ij} \le \frac{K_h(t_j - t_i)}{\kappa C_1 n/2} \tag{6.3}$$

where $K_h(x) = K(x/h)/h$ and C_1 and C_2 denote positive constants. This yields

$$\mathrm{E}[Z_i \sum_{j=1}^{n} w_{ij} h_{ij}] = \mathrm{E}[Z_i(w_{ii}h_{ii} + w_{i,i-1}h_{i,i-1})] = O\left(\frac{1}{nh}\right)$$

uniformly with respect to $i = r+1, \dots, n$, and it follows that $\mathrm{E}\, A_1(t) = O(\frac{1}{h\sqrt{n}}) = o(1)$. For the estimation of the second moment we decompose $A_1^2(t)$ as follows

$$A_1^2(t) \;=\; D_1(t) + D_2(t) + D_3(t),$$

with

$$\begin{aligned}
D_1(t) &= \frac{1}{n} \sum_{i=1}^{n} 1_{\{t_i \le t\}} Z_i^2 (\sum_{j=1}^{n} w_{ij} h_{ij})^2, \\
D_2(t) &= \frac{2}{n} \sum_{i=1}^{n-1} 1_{\{t_i \le t\}} 1_{\{t_{i+1} \le t\}} Z_i Z_{i+1} \sum_{j=1}^{n} w_{ij} h_{ij} \sum_{k=1}^{n} w_{i+1,k} h_{i+1,k}, \\
D_3(t) &= \frac{1}{n} \sum_{|i-l| \ge 2} 1_{\{t_i \le t\}} 1_{\{t_l \le t\}} Z_i Z_l \sum_{j=1}^{n} w_{ij} h_{ij} \sum_{k=1}^{n} w_{lk} h_{lk}.
\end{aligned}$$

Observing (6.3) it follows for the set $A_i := \{i-1, i\}$

$$\begin{aligned}
\mathrm{E}[Z_i^2 \sum_{j,k=1}^{n} w_{ij} w_{ik} h_{ij} h_{ik}] &= \mathrm{E}[Z_i^2 \sum_{j,k \notin A_i} w_{ij} w_{ik} h_{ij} h_{ik}] + 2\,\mathrm{E}[Z_i^2 (w_{ii} h_{ii} + w_{i,i-1} h_{i,i-1}) \sum_{k \notin A_i} w_{ik} h_{ik}] \\
&\quad + \mathrm{E}[Z_i^2 (w_{ii} h_{ii} + w_{i,i-1} h_{i,i-1})^2] \\
&= \mathrm{E}[Z_i^2 \sum_{j,k \notin A_i} w_{ij} w_{ik} h_{ij} h_{ik}] + O\left(n^{-1} h^{-1}\right) + O\left(n^{-2} h^{-2}\right) \\
&= O(h^{2\gamma}) + O\left(n^{-1} h^{-1}\right) + O\left(n^{-2} h^{-2}\right) \\
&= O(h^{2\gamma}).
\end{aligned}$$

A similar calculation shows $ED_2(t) = O(h^{2\gamma})$. For the remaining estimate for the term $D_3(t)$ we consider the set $A_{i,l} = \{i-1, i, l-1, l\}$ and obtain

$$\begin{aligned}
\mathrm{E}[Z_i Z_l \sum_{j=1}^{n} w_{ij} h_{ij} \sum_{k=1}^{n} w_{lk} h_{lk}] &= \mathrm{E}[Z_i Z_l \sum_{j,k \notin A_{i,l}} w_{ij} w_{lk} h_{ij} h_{lk}] \\
&\quad + \mathrm{E}[Z_i Z_l (w_{ii} h_{ii} + w_{i,i-1} h_{i,i-1} + w_{il} h_{il} + w_{i,l-1} h_{i,l-1}) \sum_{k \notin A_{i,l}} w_{lk} h_{lk}] \\
&\quad + \mathrm{E}[Z_i Z_l (w_{li} h_{li} + w_{l,i-1} h_{l,i-1} + w_{ll} h_{ll} + w_{l,l-1} h_{l,l-1}) \sum_{j \notin A_{i,l}} w_{ij} h_{ij}] \\
&\quad + \mathrm{E}[Z_i Z_l \sum_{j,k \in A_{i,l}} w_{ij} w_{lk} h_{ij} h_{lk}].
\end{aligned}$$

Note that the random variables Z_i and Z_l are independent whenever $|l-i| \ge 2$ and consequently the first three terms in the above expression vanish. The remaining fourth term can be decomposed in a sum of 16, which are all of the form

$$\mathrm{E}[Z_i Z_l w_{ii} w_{ll} h_{ii} h_{ll}] = O\left(\frac{1}{n^2 h^2}\right).$$

This yields

$$\mathrm{E}\, D_3(t) = O\left(\frac{1}{nh^2}\right)$$

and as a consequence $\mathrm{E}\, A_1^2(t) = O\left(\frac{1}{nh^2}\right) = o(1)$. Thus we obtain

$$A_1(t) = o_p(1) \tag{6.4}$$

uniformly with respect to $t \in [0, t_0]$. In order to derive a corresponding estimate for the term A_2 we use the decomposition

$$A_2(t) \;=\; A_{21}(t) - A_{22}(t)$$

with

$$\begin{aligned} A_{21}(t) \;&=\; \frac{4}{\sqrt{n}} \sum_{i=1}^{n} 1_{\{t_i \le t\}} Z_i \sum_{j=1}^{n} w_{ij} \varepsilon_j^3 \rho_j, \\ A_{22}(t) \;&=\; \frac{4}{\sqrt{n}} \sum_{i=1}^{n} 1_{\{t_i \le t\}} Z_i \sum_{j=1}^{n} w_{ij} \varepsilon_j^3 \sum_{k=1}^{n} w_{jk} \varepsilon_k. \end{aligned}$$

Now the Hölder continuity of the regression function implies $|\rho_j| = |\sum_{k=1}^n w_{jk}(m(t_j) - m(t_k))| \le Lh^\gamma$ for some positive constant L and a straightforward calculation shows (note that the random variables Z_i depend only on ε_i and ε_{i-1})

$$\mathrm{E}[Z_i \sum_{j=1}^{n} w_{ij} \varepsilon_j^3 \rho_j] = O\left(\frac{h^{\gamma-1}}{n}\right),$$

which implies

$$\mathrm{E}[A_{21}(t)] = O\left(\frac{h^{\gamma-1}}{\sqrt{n}}\right). \tag{6.5}$$

By a similar calculation it follows that $\mathrm{E}[A_{22}(t)] = O\left(\frac{1}{h\sqrt{n}}\right)$ and a combination of this estimate with (6.5) gives

$$\mathrm{E}[A_2(t)] = O\left(\frac{1}{h\sqrt{n}}\right). \tag{6.6}$$

The estimation of the second moments of $A_{21}(t)$ and $A_{22}(t)$ is more complicated and we indicate the calculations for the term $A_{21}(t)$, which can be decomposed as

$$A_{21}^2(t) \;=\; B_1(t) + B_2(t) + B_3(t),$$

where

$$B_1(t) \;=\; \frac{16}{n} \sum_{i=1}^{n} 1_{\{t_i \le t\}} Z_i^2 (\sum_{j=1}^{n} w_{ij} \varepsilon_j^3 \rho_j)^2,$$

$$B_2(t) = \frac{32}{n} \sum_{i=1}^{n-1} 1_{\{t_i \le t\}} 1_{\{t_{i+1} \le t\}} Z_i Z_{i+1} \sum_{j=1}^{n} w_{ij} \varepsilon_j^3 \rho_j \sum_{l=1}^{n} w_{i+1,l} \varepsilon_l^3 \rho_l,$$

$$B_3(t) = \frac{16}{n} \sum_{|l-i| \ge 2} 1_{\{t_i \le t\}} 1_{\{t_l \le t\}} Z_i Z_l \sum_{j=1}^{n} w_{ij} \varepsilon_j^3 \rho_j \sum_{r=1}^{n} w_{lr} \varepsilon_r^3 \rho_r.$$

Using the estimates (6.3) we obtain

$$E[B_1(t)] = \frac{16}{n} \sum_{i=1}^{n} E\Big[1_{\{t_i \le t\}} Z_i^2 \sum_{j=1}^{n} w_{ij}^2 \varepsilon_j^6 \rho_j^2\Big] + \frac{16}{n} \sum_{i=1}^{n} E\Big[1_{\{t_i \le t\}} Z_i^2 \sum_{j \ne l} w_{ij} w_{il} \varepsilon_j^3 \varepsilon_l^3 \rho_j \rho_l\Big]$$
$$= O\left(\frac{h^{2\gamma-1}}{n}\right) + O(h^{2\gamma}) = O(h^{2\gamma}).$$

A similar calculation shows $\mathrm{E}\, B_2(t) = O(h^{2\gamma})$ and

$$\mathrm{E}[B_3(t)] = O\left(\frac{h^{2\gamma-2}}{n}\right),$$

which implies

(6.7) $$\mathrm{E}[A_{21}^2(t)] = O(h^{2\gamma}).$$

Similarly we obtain

$$\mathrm{E}\, A_{22}^2(t) = O\left(\frac{1}{nh^2}\right),$$

and a combination with (6.7) gives

$$\mathrm{E}\, A_2^2(t) = O\left(\frac{1}{nh^2}\right) = O\left(n^{\frac{1-2\gamma}{2\gamma+1}}\right) = o(1).$$

On the other hand we have from (6.6) the estimate $EA_2(t) = O\left(\frac{1}{h\sqrt{n}}\right) = O\left(n^{\frac{1-2\gamma}{4\gamma+2}}\right) = o(1)$ and it follows that

(6.8) $$A_2(t) = o_p(1)$$

uniformly on the interval $t \in [0, t_0]$. The term $A_3(t)$ can be treated by similar arguments, which are omitted for the sake of brevity [see Hetzler (2008) for more details]. Tedious calculations yield

(6.9) $$A_3(t) = o_p(1)$$

uniformly with respect to $t \in [0, t_0]$. Finally we use the estimate (4.11) and obtain the remaining terms in (6.1)

$$|A_4(t)| \le \frac{4}{\sqrt{n}} \sum_{i=1}^{n} |Z_i \sum_{j=1}^{n} w_{ij} \varepsilon_j (m(t_j) - \hat{m}_h(t_j))^3|$$
$$\le \sup_{t \in [0,t_0]} |\hat{m}_h(t) - m(t)|^3 \cdot \frac{4}{\sqrt{n}} \sum_{i=1}^{n} |Z_i| \sum_{j=1}^{n} w_{ij} |\varepsilon_j|$$
$$= O_p\left(n^{-\frac{3\gamma}{2\gamma+1}} (\log n)^{3/2} n^{\frac{2\gamma+2}{4\gamma+2}}\right) = O_p\left(n^{\frac{2-4\gamma}{4\gamma+2}} \log n\right) = o_p(1)$$

and

$$\begin{aligned}|A_5(t)| &\leq \sup_{t\in[0,t_0]} |\hat{m}_h(t)-m(t)|^4 \, \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |Z_i| \\ &= O_p\left(n^{-\frac{4\gamma}{2\gamma+1}} \sqrt{n} (\log n)^2\right) \;=\; O_p\left(n^{\frac{1-6\gamma}{4\gamma+2}} \log n\right) \;=\; o_p(1)\end{aligned}$$

uniformly in $t \in [0, t_0]$. Combining these estimates with (6.4), (6.8) and (6.9) it follows that $A(t) = o_p(1)$ holds uniformly with respect to $t \in [0, t_0]$, which proves (4.3).

References

N. I. Achieser (1956). Theory of Approximation. Frederik Ungar Publishing Co., N.Y.

P. L. Bickel (1978). Using residuals robustly I: Tests for heteroscedasticity and nonlinearity. Annals of Statistics 6, 266-291.

P. Billingsley (1979). Probability and measure. Wiley Series in Probability and Statistics, N.Y.

P. Billingsley (1999). Convergence of probability measures, 2nd ed. Wiley Series in Probability and Statistics, N.Y.

T. S. Breusch and A. R. Pagan (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrica 47, 1287-1294.

R. D. Cook and S. Weisberg (1983). Diagnostics for heteroscedasticity in regression. Biometrika 70, 1-10.

H. Dette (2002). A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. Jour. Statist. Plan. and Infer. 103, 311-330.

Dette and Hetzler (2008). A simple test for the parametric form of the variance function in nonparametric regression. To appear in: Annals of the Institute of Statistical Mathematics. Online: http://www.springerlink.com/content/102845/?Content+Status=Accepted

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

H. Dette, I. van Keilegom and N. Neumeyer (2007). A new test for the parametric form of the variance function in nonparametric regression. J. Roy. Statist. Soc., Ser. B 69, 903-917.

T. Gasser, L. Sroka and G. Jennen-Steinmetz (1986). Residual variance and residual pattern in nonlinear regression. Biometrika 73, 626-633.

P. Hall, J.W. Kay and D.M. Titterington (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. Biometrika 77, 521-528.

B. Hetzler (2008). Tests auf die funktionale Form der Varianzfunktion in der nichtparametrischen Regression. PhD Thesis, Fakultät für Mathematik, Ruhr Universität Bochum (in German).

E.V. Khmaladze (1981). A martingale approach in the theory of goodness-of-fit tests. Theory Probab. Appl. 26, 240-257.

E.V. Khmaladze (1989). Goodness of fit problem and scanning innovation martingales. Ann. Statist. 21, 798-829.

E.V. Khmaladze and H.L. Koul (2004). Martingale transforms goodness-of-fit tests in regression models. Ann. Statist. 32, 995-1034.

H.L. Koul (2002). Weighted empirical processes in dynamic nonlinear models. Lecture Notes in Statistics, 166. Springer, N.Y.

H.L. Koul (2006). Model diagnostics via martingale transforms: a brief review. In: Frontiers in Statistics (ed. J. Fan, H.L. Koul), 183-206. Hackensack, NJ. World Scientific.

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

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

B. Øksendal (2003). Stochastic Differential Equations: An Introduction with Applications. Springer-Verlag.

A. R. Pagan and Y. Pak (1993). Testing for Heteroscedasticity in: G. S. Maddala, C. R. Rao and H. D. Vinod, eds., Handbook of Statistics, Vol. 11, Elsevier Science Publishers B.V.

D. Pollard (1984). Convergence of Stochastic Processes. Springer, N.Y.

J. Sacks and D. Ylvisacker (1970). Design for regression problems with correlated errors III. Ann. Math. Statist. 41, 2057-2074.

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

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