A simple test for the parametric form of the variance function in nonparametric regression

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: holger.dette@ruhr-uni-bochum.de
FAX: +49 2 34 32 14 559

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

January 24, 2006

Abstract

In this paper a new test for the parametric form of the variance function in the common nonparametric regression model is proposed which is applicable under very weak smoothness assumptions. The new test is based on an empirical process formed from pseudo residuals, for which weak convergence to a Gaussian process can be established. In the special case of testing for homoscedasticity the limiting process is essentially a Brownian bridge, such that critical values are easily available. The new procedure has three main advantages. First, in contrast to many other methods proposed in the literature, it does not depend directly on a smoothing parameter. Secondly, it can detect local alternatives converging to the null hypothesis at a rate $n^{-1/2}$. Thirdly, in contrast to most of the currently available tests, it does not require strong smoothness assumptions regarding the regression and variance function. We also present a simulation study and compare the tests with the procedures that are currently available for this problem and require the same minimal assumptions.

Keywords and Phrases: Homoscedasticity, nonparametric regression, pseudo residuals, empirical process, goodness-of-fit testing, bootstrap.
AMS Subject Classificaton: 62G05

1 Introduction

Consider the common nonparametric regression model with a fixed design

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

where $0 \leq t_{1,n} < t_{2,n} < \ldots < t_{n,n} \leq 1$ denote the design points, m and σ^2 are the regression and variance function, respectively, and the errors $\varepsilon_{1,1}, \ldots, \varepsilon_{1,n}$ are independent identically distributed with expectation $E[\varepsilon_{i,n}] = 0$ and variance $V[\varepsilon_{i,n}] = 1$. Additional information on the variance function σ^2, such as homoscedasticity or a specific parametric form of σ^2 usually simplifies the analysis of the data substantially. Moreover, statistical inference incorporating such an additional knowledge is also more efficient. On the other hand - if the assumption on the variance function (i.e. homoscedasticity) is not satisfied - data analysis should address for heteroscedasticity in order to obtain reliable results [see e.g. Leedan and Meer (2000) or Sadray, Rezaee and Rezaklah (2003)]. For these reasons many authors point out that it is important to check an assumption on the parametric form of the variance function by means of a goodness-of-fit test [see for example Carroll and Ruppert (1988), Cook and Weisberg (1983) among others]. Most of the available literature for this problem concentrates on the problem of testing for homoscedasticity. Tests based on a parametrically specified regression and variance function and the assumption of a normal distribution for the errors have been studied by Davidan and Carroll (1987) and Carroll and Ruppert (1988) using likelihood methods. Bickel (1978) and Carroll and Ruppert (1981) propose a test for homoscedasticity which does not impose a normal distribution for the errors but the regression function is still assumed to be linear, while Diblasi and Bowman (1997) consider the nonparametric model (1.1) with a normal distributed error.

A test for homoscedasticity in a completely nonparametric regression model was first proposed by Dette and Munk (1998). This test has the nice property that it does not depend on the subjective choice of a smoothing parameter and requires rather weak assumptions regarding the smoothness of the regression function. A disadvantage of the method is that it can only detect local alternatives converging to the null hypothesis at a rate $n^{-1/4}$. More recently Zhu, Fujikoshi and Naito (2001) [see also Zhu (2005); Chapter 7], Dette (2002) and Liero (2003) suggested test procedures, which are based on residuals from a nonparametric fit. The two last named tests can detect local alternatives converging to the null hypothesis at a rate $(n\sqrt{h})^{-1/2}$, where h denotes a bandwidth, while the rate for the test of Zhu et al. (2001) is $n^{-1/2}$. A drawback of these methods consists in the fact that the corresponding tests depend on the subjective choice of a smoothing parameter, which can affect the results of the statistical analysis.

The present paper has three purposes. First, we are interested in a test which does not require the specification of a smoothing parameter. Secondly, the new procedure should be able to detect local alternatives converging to the null hypothesis at a rate $n^{-1/2}$. Thirdly, the new test should be applicable under minimal smoothness assumptions on the variance and regression function. Moreover, in contrast to most papers which concentrate on tests for homoscedasticity, we are also interested in a test for more general hypotheses for the parametric form of the variance function, i.e.

$$H_0 : \sigma^2(t) = \sigma^2(t, \theta); \quad \forall\, t \in [0, 1]. \tag{1.2}$$

Here the form of the function $\sigma^2(t, \theta) \in \Theta \subset \mathbb{R}^d$ is known except for the d-dimensional parameter $\theta = (\theta_1, \ldots, \theta_d)^T \in \Theta \subset \mathbb{R}^d$ (note that the hypothesis of homoscedasticity is obtained for $d = 1$ and $\sigma^2(t, \theta) = \theta$).

In Section 2 we consider linear parametric classes for the function $\sigma^2(\cdot, \theta)$ and propose a stochastic process which vanishes for all t if and only if the null hypothesis in (1.2) is satisfied. We prove weak convergence of this process to a Gaussian process, and as a consequence Kolmogorov-Smirnov or Crámer-von-Mises type statistics can be constructed. In the special case of testing for homoscedasticity the limit distribution is particularly simple and given by a scaled Brownian bridge. The test is able to detect Pitman alternatives converging to the null hypothesis at a rate $n^{-1/2}$. Moreover, the asymptotic theory is applicable if the regression and variance function are Lipschitz continuous of order $\gamma > 1/2$, while the alternative procedures of Zhu et al. (2001), Zhu (2005), Dette (2002) and Liero (2003) require Lipschitz continuity of order 1 or a two times continuously differentiable regression function, respectively. The extension of the procedure to general nonlinear hypotheses is briefly mentioned in Section 3. In Section 4 we present a small simulation study, compare the new test with the currently available procedures in the literature and a data example is analyzed in order to illustrate the application of the procedure. For the problem of testing homoscedasticity we use the approximation by a Brownian bridge to obtain critical values, while for the general hypothesis of a parametric form a bootstrap procedure is proposed. It is demonstrated by means of a simulation study that in many cases the new tests based on the Cramér-von-Mises statistic yield a substantial improvement with respect to power. The case of a random design is briefly discussed in Section 5, where we demonstrate that the corresponding process has a different limit behaviour than in the case of a fixed design. Finally, some of the technical arguments are deferred to an appendix.

2 An empirical process of pseudo residuals

Consider the nonparametric regression model (1.1) where the design points $t_{i,n}$ are defined by

$$\frac{i}{n+1} = \int_0^{t_{i,n}} f\,(t)\; dt, \quad i = 1, \ldots, n, \tag{2.1}$$

[see Sacks and Ylvisaker (1970)] and f is a positive density on the interval $[0, 1]$, which is Lipschitz continuous of order $\gamma > \frac{1}{2}$, i.e. $f \in \mathrm{Lip}_\gamma[0, 1]$. Throughout this paper define $m_j(t) = E[\varepsilon^j_{i,n}(t)],\ j = 3, 4$, assume that for some $\gamma > \frac{1}{2}$

$$f, \sigma, m_3, m_4 \in \mathrm{Lip}_\gamma[0, 1] \tag{2.2}$$

and that $E[\varepsilon_{i,n}^6(t)] \leq m_6 < \infty$ with a constant m_6, which does not depend on the variable t. For the sake of a transparent presentation we consider at the moment linear hypotheses of the form

$$H_0 : \sigma^2(t) = \sum_{j=1}^{d} \theta_j \sigma_j^2(t), \qquad \text{for all } t \in [0,1], \tag{2.3}$$

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

$$\sigma_j^2 \in \mathrm{Lip}_\gamma[0,1], \qquad j = 1, \ldots, d. \tag{2.4}$$

The general case of testing hypotheses of the form (1.2) will be briefly discussed in Section 3. In order to construct a test for hypothesis (2.3) we introduce the function

$$S_t = \int_0^t \Big(\sigma^2(x) - \sum_{j=1}^{d} \alpha_j \sigma_j^2(x) \Big) f(x)\, dx, \tag{2.5}$$

where $t \in [0,1]$ and the vector $\alpha = (\alpha_1, \ldots, \alpha_d)^T$ is defined by

$$\alpha = \arg \min_{\beta \in I\!R^d} \int_0^1 \Big(\sigma^2(x) - \sum_{j=1}^{d} \beta_j \sigma_j^2(x) \Big)^2 f(x)\, dx. \tag{2.6}$$

Note that the null hypothesis (2.3) is equivalent to $S_t = 0$ for all $t \in [0,1]$, and therefore an appropriate estimate of the process S_t will be the basic tool for the construction of the new test statistic. In order to obtain such an estimate we note that it follows from standard Hilbert space theory [see Achieser (1956)] that

$$\alpha = A^{-1} C, \tag{2.7}$$

where the elements of the matrix $A = (a_{ij})_{1 \leq i,j \leq d}$ and the vector $C = (c_1, \ldots, c_d)^T$ are defined by

$$\begin{aligned} a_{ij} &= \int_0^1 \sigma_i^2(x)\, \sigma_j^2(x)\, f(x)\, dx, \quad 1 \leq i,j \leq d, \\ c_i &= \int_0^1 \sigma^2(x)\, \sigma_i^2(x)\, f(x)\, dx, \quad 1 \leq i \leq d. \end{aligned} \tag{2.8}$$

With the notation

$$B_t^0 = \int_0^t \sigma^2(x)\, f(x)\, dx, \tag{2.9}$$

$$B_t = \Big(\int_0^t \sigma_1^2(x)\, f(x)\, dx, \ldots, \int_0^t \sigma_d^2(x)\, f(x)\, dx \Big)^T \tag{2.10}$$

we therefore obtain $S_t = B_t^0 - B_t^T \alpha = B_t^0 - B_t^T A^{-1} C$ for the process in (2.5). The quantities in this representation are now estimated as follows. Let $(d_0, \dots, d_r)^T$ denote a vector with real components satisfying

$$\sum_{i=0}^{r} d_i = 0, \quad \sum_{i=0}^{r} d_i^2 = 1. \tag{2.11}$$

Following Gasser, Sroka and Jennen-Steinmetz (1986) or Hall, Kay and Titterington (1990) we define pseudo residuals

$$R_j = \sum_{i=0}^{r} d_i Y_{j-i}, \quad j = r+1, \dots, n, \tag{2.12}$$

and an estimate of (2.7) by $\hat{\alpha} = \hat{A}^{-1} \hat{C}$, where $\hat{A} = (\hat{a}_{ij})_{1 \le i,j \le d}$, $\hat{C} = (\hat{c}_1, \dots, \hat{c}_d)^T$ and the elements in these matrices are given by

$$\hat{a}_{ij} = \frac{1}{n} \sum_{k=1}^{n} \sigma_i^2 (t_{k,n}) \sigma_j^2 (t_{k,n}), \quad \hat{c}_i = \frac{1}{n-r} \sum_{k=r+1}^{n} R_k^2 \sigma_i^2 (t_{k,n}). \tag{2.13}$$

Finally, the quantities in (2.9) and (2.10) are estimated by

$$\hat{B}_t^0 = \frac{1}{n-r} \sum_{j=r+1}^{\lfloor nt \rfloor} R_j^2, \quad \hat{B}_t^i = \frac{1}{n} \sum_{j=1}^{\lfloor nt \rfloor} \sigma_i^2 (t_{j,n}), \quad i = 1, \dots, d \tag{2.14}$$

(note that $\hat{a}_{ij}$ and $\hat{B}_t^i$ are not random), and the sample version of the process S_t is given by

$$\hat{S}_t = \hat{B}_t^0 - \hat{B}_t^T \hat{A}^{-1} \hat{C}, \tag{2.15}$$

where $\hat{B}_t = (\hat{B}_t^1, \dots, \hat{B}_t^d)^T$. The following result provides the asymptotic properties of the process $\hat{S}_t$ for an increasing sample size. The proof is complicated and therefore deferred to the Appendix.

Theorem 2.1. *If the conditions (2.1), (2.2) and (2.4) are satisfied, then the process* $\{\sqrt{n}(\hat{S}_t - S_t)\}_{t \in [0,1]}$ *converges weakly in* $D[0,1]$ *to a Gaussian process with covariance kernel*

$$k(t_1, t_2) = V_2 \Sigma_{t_1,t_2} V_2^T, \tag{2.16}$$

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*

$$\Sigma_{t_1,t_2} = \begin{pmatrix} v_{11} & v_{12} & w_{11} & \cdots & w_{1d} \\ v_{21} & v_{22} & w_{21} & \cdots & w_{2d} \\ w_{11} & w_{21} & z_{11} & \cdots & z_{1d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{1d} & w_{2d} & z_{d1} & \cdots & z_{dd} \end{pmatrix}, \tag{2.17}$$

$$V_2 = (I_2 | U), \qquad U = -\left(B_{t_1}^T A^{-1}, B_{t_2}^T A^{-1}\right)^T, \tag{2.18}$$

respectively, the elements of the matrix in (2.17) are given by

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

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

$$\delta_r = \sum_{m=1}^{r}\Big(\sum_{j=0}^{r-m} d_j d_{j+m}\Big)^2. \tag{2.19}$$

Remark 2.2. It is easy to see that the matrix Σ_{t_1,t_2} in (2.17) is given by $E[PP^T]$, where the $(d+2)$-dimensional random vector P is defined by

$$P = \sqrt{\tau_r(U)}\sigma^2(U)(I\{U\le t_1\}, I\{U\le t_2\}, \sigma_1^2(U),\ldots,\sigma_d^2(U))^T,$$

and the random variable U has density f.

Remark 2.3. The main idea of the proof of Theorem 2.1 is to use the Lipschitz continuity of the regression function to derive an asymptotically equivalent representation for the process $\{\sqrt{n}(\hat{S}_t - S_t)\}_{t\in[0,1]}$, i.e.

$$\sqrt{n}(\hat{S}_t - S_t) = \sqrt{n}\Big\{\frac{1}{n-r}\sum_{j=r+1}^{\lfloor nt\rfloor} Z_{j,n} - \frac{t}{n-r}\sum_{j=r+1}^{n} h(t_{j,n})Z_{j,n}\Big\} + o_p(1) \tag{2.20}$$

uniformly with respect to $t\in[0,1]$, where h is a deterministic function and the random variables $\{Z_{j,n} \mid j=1,\ldots,n; n\in\mathbb{N}\}$ form a triangular array of rowwise $(r+1)$-dependent centered random variables. For the process on the right hand side of (2.20) we then prove tightness and convergence of the finite dimensional distributions. The technical details can be found in the appendix.

Remark 2.4. As pointed out previously the null hypothesis (2.3) is equivalent to $S_t \equiv 0 \;\; \forall\, t \in [0,1]$ and consequently rejecting (2.3) for large values of the Kolmogorov-Smirnov or Cramer-von-Mises statistic

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

yields a consistent test. Here $F_n(t) = \frac{1}{n}\sum_{i=1}^n I\{t_{i,n}\le t\}$ is the empirical distribution function of the design points. If $(A(t))_{t\in[0,1]}$ denotes the limiting process in Theorem 2.1 it follows from the Continuous Mapping Theorem

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

Remark 2.5. Define the $(n-r) \times d$ matrix

$$X = \left(\sigma_j^2(t_{i,n})\right)_{i=1,\ldots,n-r}^{j=1,\ldots,d} \in \mathbb{R}^{(n-r)\times d} \tag{2.21}$$

and a vector $R = (R_{r+1}^2, \ldots, R_n^2)^T$ of squared pseudo residuals, then it follows that the estimate $\hat{\alpha}$ of (2.7) is essentially the least squares estimate in the linear model $E[R \mid t] = X\alpha$, that is

$$\hat{\alpha} = (X^T X)^{-1} X^T R + O_p(\frac{1}{n}). \tag{2.22}$$

Example 2.6. In general the covariance structure of the limiting process is very complicated as indicated by the following example, which considers the situation for $d = 1$. In this case the matrix A in (2.7) is given by the scalar $a_{11} = \int_0^1 \sigma_1^4(x) f(x) dx$. Defining

$$s_{t,1} = \frac{B_t}{a_{11}} = \frac{\int_0^t \sigma_1^2(x) f(x)\, dx}{\int_0^1 \sigma_1^4(x) f(x)\, dx},$$

it follows from Theorem 2.1 that the process $\{\sqrt{n}(\hat{S}_t - S_t)\}_{t\in[0,1]}$ converges weakly to a Gaussian process with covariance kernel

$$\begin{aligned} k(t_1, t_2) = & \int_0^{t_1 \wedge t_2} \tau_r(x)\sigma^4(x) f(x)\, dx + s_{t_1,1} s_{t_2,1} \int_0^1 \tau_r(x)\sigma^4(x) \sigma_1^4(x) f(x)\, dx \\ & - s_{t_2,1} \int_0^{t_1} \tau_r(x)\sigma^4(x) \sigma_1^2(x) f(x)\, dx - s_{t_1,1} \int_0^{t_2} \tau_r(x)\sigma^4(x)\sigma_1^2(x) f(x)\, dx. \end{aligned} \tag{2.23}$$

In the case of testing homoscedasticity (i.e. $\sigma_1^2(t) = 1$) we have $s_{t,1} = F(t)$, where F is the distribution of the design density, and (2.23) simplifies to

$$\begin{aligned} k(t_1, t_2) = & \int_0^{t_1 \wedge t_2} \tau_r(x)\sigma^4(x) f(x)\, dx + F(t_1)F(t_2) \int_0^1 \tau_r(x)\sigma^4(x) f(x)\, dx \\ & -F(t_2) \int_0^{t_1} \tau_r(x)\sigma^4(x) f(x)\, dx - F(t_1) \int_0^{t_2} \tau_r(x)\sigma^4(x) f(x)\, dx \end{aligned}$$

The following corollary is now obvious.

Corollary 2.7. *Assume that the hypothesis of homoscedasticity $H_0 : \sigma^2(t) = \theta_1$ has to be tested (i.e. $d = 1, \sigma_1^2(t) = 1$) and that additionally $m_4(t) \equiv m_4$ is constant. If condition (2.1) and (2.2) are satisfied, then under the null hypothesis of homoscedasticity the process $\{\sqrt{n}(\hat{S}_t - S_t)\}_{t\in[0,1]}$ converges weakly on $D[0,1]$ to a scaled Brownian bridge in time F, where F is the distribution function of the design density, i.e.*

$$\{\sqrt{n}(\hat{S}_t - S_t)\}_{t\in[0,1]} \Rightarrow \sqrt{(m_4 - 1 + 4\delta_r)\theta_1^2}\{B \circ F\}_{t\in[0,1]}.$$

3 General hypotheses and local alternatives

3.1 Nonlinear hypotheses for the variance function

In this paragraph we briefly explain how the results have to be adapted if a general nonlinear hypothesis of the form (1.2) has to be tested. For this purpose we assume that the parameter space Θ is compact and that the infimum

$$\inf_{\theta\in\Theta}\int_0^1 \{\sigma^2(t)-\sigma^2(t,\theta)\}^2 f(t)dt \tag{3.1}$$

is attained at a unique point, say $\theta_0=(\theta_1^{(0)},\dots,\theta_d^{(0)})^T$, in the interior of Θ. Observing the interpretation of the estimate $\hat\alpha$ in Remark 2.4, we define

$$\hat\theta = \arg\min_{\theta\in\Theta}\frac{1}{n-r}\sum_{i=r+1}^{n}\Big(R_{i,n}^2-\sigma^2(t_{i,n},\theta)\Big)^2 \tag{3.2}$$

as the nonlinear least squares estimate. Under some regularity assumptions [see Gallant (1987), Chapter 4, or Seber and Wild (1989), p. 572-574] the sum of squares in (3.2) can be approximated by

$$\frac{1}{n-r}H^T(I_{n-r}-X(X^TX)^{-1}X^T)H+O_p\Big(\frac{1}{n}\Big),$$

where I_{n-r} is the $(n-r)\times(n-r)$ identity matrix, the components of the vector $H=(H_{r+1,n},\dots,H_{n,n})^T$ are defined by

$$H_{j,n}=\Big(\sum_{i=0}^{r}d_i\sigma(t_{j-i,n})\varepsilon_{j-i,n}\Big)^2-\sigma^2(t_{j,n},\theta_0)\ ,\ j=r+1,\dots,n,$$

and the matrix X is given by (2.21) with $\sigma_j^2(t)=\frac{\partial}{\partial\theta_j}\sigma^2(t,\theta)|_{\theta=\theta_0}$ $(j=1,\dots,d)$. Similarly, the analogue of the process in (2.15) is given by

$$\begin{aligned}\hat S_t &= \hat B_t^0-\frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor}\sigma^2(t_{i,n},\hat\theta)\\ &= \frac{1}{n-r}\sum_{i=r+1}^{\lfloor nt\rfloor}\Big\{H_{i,n}-\frac{\partial}{\partial\theta}\sigma^2(t_{i,n},\theta)\mid_{\theta=\theta_0}(\hat\theta-\theta)\Big\}+o_p(n^{-1/2}).\end{aligned} \tag{3.3}$$

Roughly speaking this means that the nonlinear case can be treated as the linear case, where the variance function has to be replaced by $\sigma^2(x)-\sigma^2(x,\theta_0)$ and the functions σ_j^2 are given by $\frac{\partial}{\partial\theta_j}\sigma^2(x,\theta)\mid_{\theta=\theta_0}$ $(j=1,\dots,d)$. In particular, with the notation $S_t=\int_0^t(\sigma^2(x)-\sigma^2(x,\theta_0))dx$, we obtain the representation

$$\sqrt{n}(\hat S_t-S_t)=\frac{\sqrt{n}}{n-r}\sum_{i=r+1}^{\lfloor nt\rfloor}\Big\{H_{i,n}-E[H_{i,n}]-\sum_{j=1}^{d}\frac{\partial}{\partial\theta_j}\sigma^2(t_{i,n},\theta)\mid_{\theta=\theta_0}\alpha_j\Big\}+o_p(1), \tag{3.4}$$

where $\alpha_j = \hat{\theta}_j - \theta_j^{(0)}$ $(j = 1, \ldots, d)$ and the vector $\alpha = (\alpha_1, \ldots, \alpha_d)^T$ satisfies

$$\alpha = \hat{\theta} - \theta_0 = (X^T X)^{-1} X^T H.$$

From (2.21) and the condition

$$0 = \frac{\partial}{\partial \theta_j} \int_0^1 (\sigma^2(x) - \sigma^2(x, \theta))^2 f(x) dx \Big|_{\theta=\theta_0} \;=\; -2 \int_0^1 \sigma_j^2(x)(\sigma^2(x) - \sigma^2(x, \theta_0)) f(x) dx$$

it follows that

$$\frac{1}{n} X^T X - \hat{A} = O\Big(\frac{1}{n}\Big)$$
$$\frac{1}{n} X^T H - \frac{1}{n} \Big(\sum_{i=r+1}^{n} \frac{\partial}{\partial \theta_j} \sigma^2(t_{i,n}, \theta) \mid_{\theta=\theta_0} (H_{i,n} - E[H_{i,n}]) \Big)_{j=1}^{d} = O\Big(\frac{1}{n}\Big) .$$

Consequently the right hand side of (3.4) corresponds to the expression in (2.20) [see also the representation (A.13) in the proof of Theorem 2.1 in the Appendix]. This means that the process $\{\sqrt{n}(\hat{S}_t - S_t)\}_{t \in [0,1]}$ 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

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

3.2 Local alternatives

In this paragraph we briefly discuss the asymptotic behaviour of the process $\{\sqrt{n}(\hat{S}_t - S_t)\}_{t \in [0,1]}$ in the case of local alternatives

$$\sigma_n^2(t) = \sigma^2(t, \theta_0) + \frac{1}{\sqrt{n}} h(t)$$

for a fixed function $h : [0, 1] \to \mathbb{R}$, such that $\sigma_n^2(t)$ is nonnegative for all $t \in [0, 1]$. Denote $\{A(t)\}_{t \in [0,1]}$ as the limiting process in Theorem 2.1 and define

$$\gamma = (\gamma_1, \ldots, \gamma_d)^T = \arg \min_{\beta \in \mathbb{R}^d} \int_0^1 \Big(h^2(x) - \sum_{j=1}^{d} \beta_j \frac{\partial}{\partial \theta_j} \sigma^2(t, \theta) \Big|_{\theta=\theta_0} \Big)^2 f(x)\, dx,$$

then it follows from the arguments given in the Appendix that the process $\{\sqrt{n}(\hat{S}_t - S_t)\}_{t \in [0,1]}$ converges weakly to the process

$$\Big\{ A(t) + \int_0^t \Big(h(x) - \sum_{j=1}^{d} \gamma_j \frac{\partial}{\partial \theta_j} \sigma^2(x, \theta) \Big|_{\theta=\theta_0} \Big) f(x)\, dx \Big\}_{t \in [0,1]}.$$

This means that tests based on the process $\{\sqrt{n}(\hat{S}_t - S_t)\}_{t \in [0,1]}$ can detect local alternatives converging to the null hypothesis at a rate $n^{-1/2}$, whenever

$$h \notin \operatorname{span}\Big\{ \frac{\partial}{\partial \theta} \sigma^2(\cdot, \theta)|_{\theta=\theta_0}, \ldots, \frac{\partial}{\partial \theta} \sigma^2(\cdot, \theta)|_{\theta=\theta_0} \Big\}.$$

4 Finite sample properties and a data example

In this section we illustrate the finite sample properties of the new test by means of a simulation study and a data example. We first investigate the performance in the problem of testing for homoscedasticity and also compare the new procedure with alternative tests for this problem.

4.1 Testing for homoscedasticity

To our knowledge there exists only one test for the hypothesis of homoscedasticity which does not depend on the subjective choice of a smoothing parameter and requires the same minimal assumptions regarding the smoothness of the regression and variance functions. This test was proposed by Dette and Munk (1998) and is based on an estimate of the L^2-distance between the variance function under the null hypothesis and alternative. Following these authors we considered the problem of testing for homoscedasticity in the nonparametric regression model (1.1) with regression and variance function given by

$$
\begin{aligned}
&(4.1) \qquad & m(t) &= 1 + \sin(t); \quad & \sigma(t) &= \sigma \exp(ct), \\
&(4.2) \qquad & m(t) &= 1 + t; \quad & \sigma(t) &= \sigma[1 + c\sin(10t)]^2, \\
&(4.3) \qquad & m(t) &= 1 + t; \quad & \sigma(t) &= \sigma[1 + ct]^2,
\end{aligned}
$$

where $\sigma = 0.5, c = 0, 0.5, 1$ and the case $c = 0$ corresponds to the null hypothesis of homoscedasticity [i.e. $d = 1, \sigma_1^2(t) = \theta_1$]. The design is uniform (i.e. $f \equiv 1$) and the random variables $\varepsilon_{i,n}$ have a standard normal distribution. All rejection probabilities were calculated with 5000 simulation runs. As pointed out in Section 2 rejecting the null hypothesis of homoscedasticity for large values of the statistic $\int_0^1 \hat{S}_t^2 dF_n(t)$ yields a consistent test (recall that F_n denotes the empirical distribution function of the design points). It follows from Corollary 2.6 and the Continuous Mapping Theorem that under the null hypothesis of homoscedasticity

$$
C_n = n \int_0^1 \hat{S}_t^2 dF_n(t) \xrightarrow{\mathcal{D}} (m_4 - 1 + 4\delta_r)\theta_1^2 \int_0^1 B^2(F(t))dF(t) = (m_4 - 1 + 4\delta_r)\theta_1^2 \int_0^1 B^2(t)dt,
$$

where B denotes a standard Brownian bridge. If w_α denotes the $1-\alpha$ quantile of the distribution of the random variable $\int_0^1 B^2(t)dt$ and $\hat{m}_4$ is an estimate of the fourth moment, then the test, which rejects the hypothesis of homoscedasticity $H_0 : \sigma^2(t) = \theta_1$ if

$$
C_n = n \int_0^1 \hat{S}_t^2 dF_n(t) \geq w_\alpha(\hat{m}_4 - 1 + 4\delta_r)\theta_1^2, \tag{4.4}
$$

has asymptotically level α. Note that the estimate of m_4 depends on the choice of the difference sequence $d_0, \ldots, d_r$ for the calculation of the pseudo residuals $R_{i,n}$. For example, if $r = 1$ we have $d_0 = -d_1 = 1/\sqrt{2}$ and it is easy to see that

$$
\hat{m}_4 = \Big(\frac{1}{2(n-1)} \sum_{j=2}^{n} R_{j,n}^4\Big)\Big(\frac{1}{2(n-1)} \sum_{j=2}^{n} R_{j,n}^2\Big)^{-2} - 3
$$

is a consistent estimate of m_4. The corresponding estimates for other cases can be obtained similarly. We first briefly investigate the impact of the choice of the order of the difference scheme $d_0, \ldots, d_r$ for the calculation of the pseudo residuals. As pointed out by Dette, Munk and Wagner (1998), the sequence $(d_0, \ldots, d_r)$ could be chosen such that the bias $E[R_{i,n}^2] - \sigma^2(t_{i,n})$ is diminished or such that the variance of the estimate $\frac{1}{n-r}\sum_{i=r+1}^n R_{i,n}^2$ of the integrated variance $\int_0^1 \sigma^2(x)f(x)dx$ is minimal. The lastnamed choice corresponds to the minimization of δ_r with respect to the difference sequence $(d_0, \ldots, d_r)$ and the optimal weights for various values of r can be found in Hall et al. (1990). However, it turns out that the bias has a substantial impact on the approximation of the nominal level of the new test. As a consequence optimal difference sequences as proposed by Hall et al. (1990) cannot be recommended for our test procedure (for the sake of brevity these results are not presented). In Table 4.1 and 4.2 we display the level and power of the new test for the difference sequence

$$d_j = (-1)^j \frac{\binom{r}{j}}{\binom{2r}{r}^{1/2}}, \quad j = 0, \ldots, r, \tag{4.5}$$

with $r = 1$ and $r = 2$, respectively, which was recommended for a uniform design by Dette et al. (1998) in order to reduce the bias of a nonparametric variance estimator.

		$n = 50$			$n = 100$			$n = 200$		
$r = 1$	c	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
	0	0.050	0.080	0.131	0.033	0.061	0.115	0.029	0.057	0.104
(4.1)	0.5	0.171	0.245	0.357	0.256	0.361	0.490	0.504	0.628	0.743
	1	0.413	0.543	0.695	0.743	0.842	0.919	0.980	0.992	0.997
	0	0.050	0.078	0.130	0.036	0.061	0.114	0.025	0.051	0.106
(4.2)	0.5	0.132	0.184	0.271	0.181	0.267	0.419	0.330	0.515	0.748
	1	0.138	0.196	0.285	0.207	0.315	0.462	0.390	0.585	0.807
	0	0.051	0.077	0.128	0.032	0.062	0.115	0.025	0.051	0.105
(4.3)	0.5	0.313	0.423	0.564	0.561	0.691	0.804	0.897	0.943	0.975
	1	0.588	0.724	0.851	0.910	0.962	0.987	0.999	1.000	1.000

Table 4.1. *Simulated rejection probabilities of the test (4.4) with a difference sequence of the form (4.5) and $r = 1$. The case $c = 0$ corresponds to the null hypothesis of homoscedasticity.*

We observe that the theoretical level is well approximated for sample sizes larger than $n = 100$. If the sample size is smaller the approximation is less precise for difference sequences of order $r = 1$ [see Table 4.1 with $n = 50$] but reasonable accurate for the case $r = 2$ [see Table 4.2]. On the other hand an increase of the order yields to some loss in power in the case $r = 2$. This corresponds to the asymptotic theory, which indicates that a smaller value of δ_r yields a more powerful procedure. In particular, for $r = 1, 2$ the values corresponding to the sequence (4.5) are given by $\delta_1 = 1/4, \delta_2 = 17/36$, respectively. Based on an extensive study we recommend to use a difference sequence of order $r = 1$ (in order to increase the power) and to use the bootstrap (as described in the following section) for sample sizes smaller than 50 (in order to obtain a reasonable approximation of the nominal level.)

		$n = 50$			$n = 100$			$n = 200$		
$r = 2$	c	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
	0	0.029	0.058	0.113	0.025	0.051	0.106	0.030	0.057	0.111
(4.1)	0.5	0.133	0.219	0.336	0.199	0.297	0.433	0.406	0.523	0.653
	1	0.354	0.493	0.651	0.622	0.749	0.859	0.939	0.969	0.987
	0	0.027	0.058	0.110	0.024	0.053	0.101	0.024	0.050	0.099
(4.2)	0.5	0.066	0.109	0.190	0.106	0.180	0.311	0.197	0.344	0.584
	1	0.067	0.109	0.200	0.113	0.195	0.327	0.255	0.413	0.673
	0	0.032	0.061	0.115	0.027	0.053	0.104	0.028	0.052	0.102
(4.3)	0.5	0.242	0.365	0.531	0.457	0.595	0.726	0.795	0.880	0.937
	1	0.482	0.643	0.802	0.831	0.922	0.968	0.995	0.998	1.000

Table 4.2. *Simulated rejection probabilities of the test (4.4) with a difference sequence of the form (4.5) with $r = 2$. The case $c = 0$ corresponds to the null hypothesis of homoscedasticity.*

It is also of interest to compare these results with the corresponding rejection probabilities of the test suggested by Dette and Munk (1998) which requires the same minimal assumptions as the procedure proposed in this paper. The results in Table 4.1 are directly comparable with the results of Table 1 in this reference. We observe that for model (4.1) and (4.3) the new test yields substantially larger power than the test of Dette and Munk (1998). On the other hand, in model (4.2) the procedure of Dette and Munk (1998) based on the L^2-distance is substantially more powerful for the sample sizes $n = 50$ and $n = 100$, while both tests are comparable for the sample size $n = 200$ [see Table 4.1]. Recall once again that Dette and Munk's (1998) test can detect local alternatives converging to the null hypothesis at a rate $n^{-1/4}$ while the rate for the procedure proposed in this paper is $n^{-1/2}$. The reason for the difference between the asymptotic theory and the empirical results for small sample sizes in model (4.2) can be explained by the specific form of the function

$$S_t = \int_0^t (\sigma^2(x) - \theta_0)dx = \int_0^t \sigma^2(x)dx - t\int_0^1 \sigma^2(x)dx \tag{4.6}$$

which is depicted in Figure 4.1 for the case $c = 0.5$ and $c = 1$. We observe that it is difficult to distinguish these functions from the line $\bar{S}_t \equiv 0$. As a consequence the asymptotic advantages of the new test with respect to Pitman alternatives are only visible for a large sample size as $n = 200$. This effect is even more visible if the sample size is $n = 400$. For example if $c = 0.5$ the rejection probabilities of the test of Dette and Munk (1998) are 0.810, 0.887, 0.951 while the new test yields larger power, namely 0.898, 0.978, 0.997 at level 2.5%, 5% and 10%, respectively.

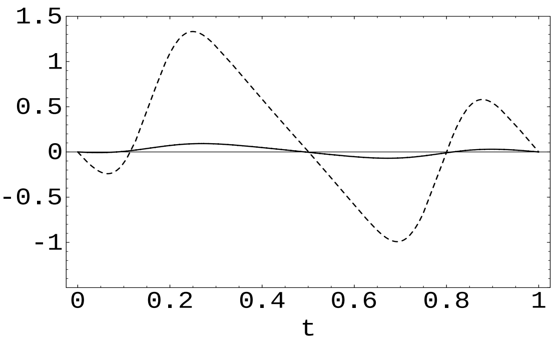

Figure 4.1: *The function S_t defined in (4.6) for $c = 0.5$ (solid line) and $c = 1$ (dotted line).*

4.2 Testing for a parametric hypothesis

In this paragraph we consider the general hypothesis (1.2). We begin with a linear parametric class of variance functions

$$H_0 : \sigma^2(t) = 1 + \theta t^2 \tag{4.7}$$

($\theta \in \mathbb{R}$). We simulated data according to the model

$$m(t) = 1 + t \quad , \quad \sigma^2(t) = 1 + 3t^2 + 2.5c \sin(2\pi t), \tag{4.8}$$

where the case $c = 0$ corresponds to the null hypothesis and the choices $c = 0.5, 1$ to two alternatives. The errors are again standard normal distributed and the design is uniform. Because the limit distribution provided by Theorem 2.1 is complicated we applied a bootstrap procedure to obtain the critical values. More precisely, we calculated nonparametric residuals $\hat{\varepsilon}_i = (Y_{i,n} - \hat{m}(t_{i,n}))\,/\hat{\sigma}(t_{i,n}), \quad i = 1, \ldots, n$, where

$$\hat{m}(t) = \sum_i W_i(t,h) Y_i, \quad \hat{\sigma}^2(t) = \sum_i W_i(t,h)(Y_i - \hat{m}(t_{i,n}))^2$$

and $W_i(t,h)$ are the local linear weights [see Fan and Gijbels (1996)]. The bandwidth h in these estimates was chosen by least squares cross validation. In a second step we defined $\varepsilon_1^*, \ldots, \varepsilon_n^*$ as a sample of i.i.d. observations with distribution function $\hat{F}_\varepsilon$ and generated bootstrap data according to the model

$$Y_i^* = \hat{m}(t_{i,n}) + \sigma(t_{i,n}, \hat{\theta})\varepsilon_i^*,$$

where $\sigma^2(\cdot, \hat{\theta})$ is the estimate of the variance function under the null hypothesis (4.7). Finally, the corresponding Cramér-von-Mises statistic, say C_n^*, was calculated from the bootstrap data. If B bootstrap replications have been performed and $C_n^{*(1)} < \ldots < C_n^{*(B)}$ denote the order statistics of the calculated bootstrap sample, the null hypothesis (4.7) was rejected if $C_n > C_n^{*(\lfloor B(1-\alpha) \rfloor)}$. $B = 100$ bootstrap replications were performed to calculate the rejection probabilities and 1000 simulation runs were used for each scenario. The results are depicted in the first part of Table 4.3.

We observe a rather precise approximation of the nominal level and a reasonable power under the alternatives.

	c	$n = 50$ 2.5%	5%	10%	$n = 100$ 2.5%	5%	10%	$n = 200$ 2.5%	5%	10%
(4.7)	0	0.023	0.060	0.102	0.023	0.057	0.116	0.023	0.051	0.114
	0.5	0.319	0.386	0.463	0.459	0.549	0.632	0.721	0.803	0.864
	1	0.659	0.718	0.774	0.888	0.922	0.948	0.988	0.993	0.997
(4.9)	0	0.032	0.065	0.115	0.023	0.056	0.116	0.025	0.057	0.105
	0.5	0.191	0.281	0.357	0.268	0.362	0.445	0.426	0.546	0.640
	1	0.403	0.511	0.603	0.504	0.608	0.711	0.892	0.939	0.968

Table 4.3. *Simulated rejection probabilities of the bootstrap test for the one-parametric hypotheses (4.7) and (4.9) in the regression models (4.8) and (4.10), respectively.*

We will conclude this section with an investigation of a nonlinear hypothesis for the variance function, i.e.

$$\sigma^2(t,\theta) = e^{\theta t} \tag{4.9}$$

($\theta \in \mathbb{R}$). We simulated data according to the model

$$m(t) = 1 + t \quad , \quad \sigma^2(t,\theta) = (1 + c\sin(2\pi t))e^{\theta t}, \tag{4.10}$$

where the case $c = 0$ corresponds to the null hypothesis and the choices $c = 0.5, 1$ to two alternatives. The errors are again standard normal distributed and the design is uniform. In the second part of Table 4.3 we display the corresponding rejection probabilities of the bootstrap test based on the procedure described in Section 3.1. We observe a precise approximation of the nominal level (similar as in the linear case). Moreover, the alternatives are detected with reasonable power.

4.3 Data example

In this section we briefly illustrate an application of the new test in an example of econometrics. For this purpose we reanalyze data of average weakly expenditure on food and average weakly income in Dollars [see Griffiths, Hill and Judge (1993), p. 182]. Table 5.2 in this reference shows the data for 40 households taken from a larger sample. Only households with at least three family members are investigated and the statistical model is used to estimate the influence of income on food expenditure. From the scattergram on page 183 in Griffiths, Hill and Judge (1993) it is fairly obvious that heteroscedasticity is present and the hypothesis of homoscedasticity is clearly rejected by our test with a p-value 0.01.

Based on a visual examination of least squares residuals Griffiths, Hill and Judge (1993) proposed a parametric model for the variance function, that is

$$\sigma^2(t) = \theta t$$

[see page 185 in this reference]. The application of our procedure for testing this hypothesis yields the p-value 0.272, which supports the assumption made by these authors. It might be also of interest to test further polynomial hypotheses, that is

$$H_0 : \sigma^2(t) = \theta t^k \tag{4.11}$$

for some fixed $k \geq 0$. The p-values of our test are listed in Table 4.4 for various values of k. These results indicate that the alternative parametrization $\sigma^2(t) = \theta t^2$ might be more appropriate, because the test for this hypothesis yields a substantially larger p-value.

k	0.0	0.2	0.4	0.6	0.8	1	2	3	4	5
p	0.010	0.020	0.062	0.098	0.142	0.272	0.99	0.348	0.068	0.012

Table 4.4. *p-values of the bootstrap test for hypotheses of the form (4.11).*

5 Random design

In this section we briefly discuss the behaviour of a corresponding stochastic process in the case of a regression model with a random design, which turns out to be different from the fixed design case. For this purpose consider the model

$$Y_i = m\left(X_i\right) + \sigma\left(X_i\right)\varepsilon_i, \quad i = 1, \ldots, n, \tag{5.1}$$

where $X_1, \ldots, X_n$ are i.i.d. with positive density f on the interval $[0,1]$ and the random errors $\varepsilon_1, \ldots \varepsilon_n$ have mean 0, variance 1 and are also i.i.d.. We denote by $m_j(x) = E[\varepsilon^j | X = x]$ the jth conditional moment of the errors and assume that $m_6(x)$ is bounded by some constant, say m_6. We consider the process $\{\hat{S}_t\}_{t\in[0,1]}$ defined in (2.15) with the following modifications. The elements of the matrix $\hat{A}$ are defined as in (2.13), where the fixed design points $t_{i,n}$ have been replaced by the random variables X_i. Additionally, the statistics $\hat{c}_i$, $\hat{B}_t^0$, $\hat{B}_t^i$ have been replaced by

$$\hat{c}_i = \frac{1}{n-r} \sum_{j=r+1}^{n} R_j^2 \sigma^2(X_{(j)}) \tag{5.2}$$

$$\hat{B}_t^0 = \frac{1}{n-r} \sum_{j=r+1}^{n} R_j^2 I\{X_{(j)} \leq t\}, \tag{5.3}$$

$$\hat{B}_t^i = \frac{1}{n} \sum_{j=1}^{n} \sigma_i^2\left(X_{(j)}\right) I\{X_{(j)} \leq t\}, \tag{5.4}$$

respectively, the pseudo residuals are defined by $R_j = \sum_{i=0}^{r} d_i Y_{A_{j-i}}, \quad j = r+1, \ldots, n$, $X_{(1)}, \ldots, X_{(n)}$ and $A_1, \ldots, A_n$ denote the order statistic and the antiranks of $X_1, \ldots, X_n$. It is easy to see that for a fixed design the corresponding estimates in (5.2), (5.3), (5.4) and in (2.13) and (2.14) differ only by a term of order $o_P(n^{-1/2})$, and as a consequence for a fixed design the process $\hat{S}_t$ with the estimates $\hat{c}_i$, $\hat{B}_t^0$ and $\hat{B}_t^i$ defined in (5.2), (5.3) and (5.4), respectively, exhibits the same asymptotic behaviour as described in Theorem 2.1. However, the following result shows that in the case of the random design the stochastic process has a different asymptotic behaviour.

Theorem 5.1. *Consider the nonparametric regression model (5.1) with a random design and the stochastic process $\hat{S}_t$ defined in (2.15), where $\hat{c}_i$, $\hat{B}_t^0$ and $\hat{B}_t^i$ are defined in (5.2), (5.3) and (5.4), respectively. If the conditions (2.1), (2.2), (2.4) and the conditions stated at the beginning of this section are satisfied, then the process $\{\sqrt{n}(\hat{S}_t - S_t)\}_{t\in[0,1]}$ converges weakly in $D[0,1]$ to a Gaussian process with covariance kernel*

$$k(t_1,t_2) \;=\; V_2\bar{\Sigma}_{t_1,t_2}V_2^T, \tag{5.5}$$

where the matrix $V_2 \in \mathbb{R}^{2\times(d+2)}$ is defined in (2.18), $\bar{\Sigma}_{t_1,t_2} = \Sigma_{t_1,t_2} + \Phi_{t_1,t_2}$, the matrix Σ_{t_1,t_2} is given in (2.17),

$$\Phi_{t_1,t_2} \;=\; \begin{pmatrix} \bar{v}_{11} & \bar{v}_{12} & \bar{w}_{11} & \cdots & \bar{w}_{1d} \\ \bar{v}_{21} & \bar{v}_{22} & \bar{w}_{21} & \cdots & \bar{w}_{2d} \\ \bar{w}_{11} & \bar{w}_{21} & \bar{z}_{11} & \cdots & \bar{z}_{1d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{w}_{1d} & \bar{w}_{1d} & \bar{z}_{d1} & \cdots & \bar{z}_{dd}. \end{pmatrix} \tag{5.6}$$

and the elements of the matrix Φ_{t_1,t_2} are defined by

$$\begin{aligned} \bar{v}_{ij} &= \int_0^1 \sigma^4(s)\, 1_{[0,t_i\wedge t_j)}(s)\, f(s)\; ds - B^0_{t_i}B^0_{t_j}, \qquad 1\le i,j\le 2, \\ \bar{w}_{ij} &= \int_0^1 \sigma^4(s)\,\sigma_j^2(s)\, 1_{[0,t_i)}(s)\, f(s)\; ds - B^0_{t_i}c_j, \qquad 1\le i\le 2, 1\le j\le d, \\ \bar{z}_{ij} &= \int_0^1 \sigma^4(s)\,\sigma_i^2(s)\,\sigma_j^2(s)\, f(s)\; ds - c_ic_j, \qquad 1\le i,j\le d. \end{aligned} \tag{5.7}$$

Remark 5.2. It is easy to see that the matrix Φ_{t_1,t_2} in (5.6) is the covariance matrix of the $(d+2)$-dimensional random vector $Q = \sigma^2(U)(I\{U\le t_1\}, I\{U\le t_2\}, \sigma_1^2(U),\ldots,\sigma_d^2(U))^T)$, where the random variable U has density f. Observing the definition of the vector P in Remark 2.2 we therefore obtain $\bar{\Sigma}_{t_1,t_2} = E[PP^T] + \mathrm{Var}[Q]$. Comparing Theorem 2.1 and 5.1 we observe that in the case of a random design there appears the additional term $V_2\Phi_{t_1,t_2}V_2^T$ in the covariance kernel of the limiting process. A similar phenomenon was observed by Munk (2002) in the context of testing for the parametric form of the regression function. However, our final result shows that in the context of testing for homoscedasticity the covariance kernel of the limiting process in the case of a random design differs only by a factor from the kernel obtained under the fixed design assumption.

Corollary 5.3. *Consider the nonparametric regression model (5.1) with a random design and the stochastic process $\hat{S}_t$ given in (2.15), where $\hat{c}_i$, $\hat{B}_t^0$ and $\hat{B}_t^i$ are defined in (5.2), (5.3) and (5.4), respectively. Assume that the hypothesis of homoscedasticity $H_0 : \sigma^2(t) = \theta_1$ has to be tested (i.e. $d = 1, \sigma_1^2(t) = 1$) and that additionally $m_4(t) \equiv m_4$ is constant. If the conditions (2.1), (2.2) and the conditions stated at the beginning of this section are satisfied, then under the null hypothesis of homoscedasticity the process*

$\{\sqrt{n}(\hat{S}_t - S_t)\}_{t\in[0,1]}$ *converges weakly on* $D[0,1]$ *to a scaled Brownian bridge in time* F*, where* F *is the distribution function of the random variables* X_i*, i.e.*

$$\{\sqrt{n}(\hat{S}_t - S_t)\}_{t\in[0,1]} \Rightarrow \sqrt{(m_4 + 4\delta_r)\,\theta_1^2}\{B \circ F\}_{t\in[0,1]}.$$

A Appendix

Proof of Theorem 2.1. For the sake of a transparent notation we omit the index n in this section, whenever the dependence on n will be clear from the context. In particular we write t_j and ε_j instead of $t_{j,n}$ and $\varepsilon_{j,n}$, respectively. We define the random variables

$$L_k = \sum_{j=0}^{r} d_j \sigma(t_{k-j})\varepsilon_{k-j}\ ,\ k = r+1, \dots, n, \tag{A.1}$$

and analogues of the estimates $\hat{B}_t^0$ and $\hat{c}_i$ by

$$\tilde{B}_t^0 = \frac{1}{n-r}\sum_{j=r+1}^{\lfloor nt \rfloor} L_j^2, \quad \tilde{c}_i = \frac{1}{n-r}\sum_{j=r+1}^{n} L_j^2 \sigma_i^2(t_j)\,. \tag{A.2}$$

With the notation $\tilde{C} = (\tilde{c}_1, \dots, \tilde{c}_d)^T$ we introduce the stochastic process

$$\tilde{S}_t = \tilde{B}_t^0 - \hat{B}_t^T \hat{A}^{-1}\tilde{C} = \hat{S}_t + o_p(n^{-1/2}) \tag{A.3}$$

uniformly with respect to $t \in [0,1]$, where we used the Lipschitz continuity of the regression function. Consequently, the processes $A_n(t) = \{\sqrt{n}(\hat{S}_t - S_t)\}_{t\in[0,1]}$ and $\{\tilde{A}_n(t)\}_{t\in[0,1]} = \{\sqrt{n}(\tilde{S}_t - S_t)\}_{t\in[0,1]}$ exhibit the same asymptotic behaviour, and the assertion of Theorem 2.1 follows if a corresponding statement for the process $\{\tilde{A}_n(t)\}_{t\in[0,1]}$ can be established.
For a proof of this property we introduce a further decomposition

$$\tilde{A}_n(t) = \sqrt{n}(\tilde{S}_t - \mathrm{E}[\tilde{S}_t]) + \sqrt{n}(\mathrm{E}[\tilde{S}_t] - S_t) = \bar{A}_n(t) + \bar{B}_n(t)\,, \tag{A.4}$$

where the last equality defines the processes $\bar{A}_n(t)$ and $\bar{B}_n(t)$. A simple calculation and the Lipschitz continuity of σ^2 show $\bar{B}_n(t) = o(1)$, uniformly with respect to $t \in [0,1]$, and therefore it is sufficient to consider the process $\bar{A}_n$ in the following discussion. Thus the assertion of Theorem 2.1 follows from the weak convergence

$$\{\bar{A}_n(t)\}_{t\in[0,1]} \Rightarrow \{A(t)\}_{t\in[0,1]}, \tag{A.5}$$

where $\{A(t)\}_{t\in[0,1]}$ is a Gaussian process with covariance kernel defined in (2.16). For a proof of this statement we first show convergence of the finite dimensional distributions, i.e.

$$\left(\bar{A}_n(t_1), \dots, \bar{A}_n(t_k)\right)^T \xrightarrow{\mathcal{D}} \left(A(t_1), \dots, A(t_k)\right)^T \tag{A.6}$$

for any vector $(t_1, \dots, t_k) \in [0,1]^k$. Secondly, we prove that there exists a constant, say C, such that for all $0 \le s < t \le 1$

$$\mathrm{E}\left[\left|\bar{A}_n(t) - \bar{A}_n(s)\right|^4\right] \le C\,(t-s)^2\,. \tag{A.7}$$

The assertion (A.5) then follows from Theorem 13.5 in Billingsley (1999).
For a proof of (A.6) we restrict ourselves to the case $k = 2$ (the general case follows exactly the same way with an additional amount of notation) and note that the process $\bar{A}_n$ can be represented as

$$\bar{A}_n(t) = \bar{B}_t^0 - \hat{B}_t \hat{A}^{-1} \bar{C}, \tag{A.8}$$

where $\bar{C} = (\bar{c}_1, \dots, \bar{c}_d)^T$,

$$\bar{B}_t^0 = \frac{1}{n-r} \sum_{j=r+1}^{\lfloor nt \rfloor} Z_j, \quad \bar{c}_i = \frac{1}{n-r} \sum_{j=r+1}^{n} Z_j \sigma_i^2(t_j), \tag{A.9}$$

and the random variables Z_j are defined by $Z_j = L_j^2 - \mathrm{E}[L_j^2]$. From the representation $\left(\bar{A}_n(t_1), \bar{A}_n(t_2)\right)^T = \hat{V}_2 X_n$, with $X_n = \sqrt{n}\left(\bar{B}_{t_1}^0, \bar{B}_{t_2}^0, \bar{c}_1, \dots, \bar{c}_d\right)^T$, $\hat{V}_2 = (I_2 | \hat{U})$, $\hat{U} = -(\hat{B}_{t_1}^T \hat{A}^{-1}, \hat{B}_{t_2}^T \hat{A}^{-1})^T$ and $V_2 = \hat{V}_2 + o(1)$ it follows that it is sufficient to establish the weak convergence

$$X_n \xrightarrow{\mathcal{D}} \mathcal{N}_{2+d}(0, \Sigma_{t_1,t_2}), \tag{A.10}$$

where the matrix Σ_{t_1,t_2} is defined in (2.17). For a proof of this statement we first calculate the asymptotic covariance matrix of the random vector X_n. Observing the identity

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

(uniformly with respect to $t_j, j = 1, \dots, n$) we obtain for $i = 1, 2$

$$\begin{aligned} n\,\mathrm{E}\left[\left(\bar{B}_{t_i}^0\right)^2\right] &= n\,\mathrm{E}\Big(\frac{1}{n-r}\sum_{j=r+1}^{\lfloor nt_i \rfloor} Z_j\Big)^2 = \frac{1}{n-r}\sum_{j=r+1}^{\lfloor nt_i \rfloor - r}\Big(\mathrm{E}\left[Z_j^2\right] + 2\sum_{m=1}^{r}\mathrm{E}\left[Z_j Z_{j+m}\right]\Big) + O\left(\frac{1}{n}\right) \\ &= \int_0^{t_i} \tau_r(x)\sigma^4(x) f(x)\, dx + O\left(n^{-\gamma}\right) \end{aligned}$$

(uniformly with respect to $t_i, i = 1, \dots, n$), where we have used the Lipschitz-continuity of the functions σ^2, σ_j^2, f. A similar calculation yields for $1 \le i, \ell \le 2; t_i \le t_\ell$

$$\begin{aligned} n\,\mathrm{E}\left[\bar{B}_{t_i}^0 \bar{B}_{t_\ell}^0\right] &= n\,\mathrm{E}\Big(\frac{1}{n-r}\sum_{j=r+1}^{\lfloor nt_i \rfloor} Z_j \cdot \frac{1}{n-r}\sum_{j=r+1}^{\lfloor nt_\ell \rfloor} Z_j\Big) \\ &= \frac{1}{n-r}\sum_{j=r+1}^{\lfloor nt_i \rfloor - r}\Big(\mathrm{E}\left[Z_j^2\right] + 2\sum_{m=1}^{r}\mathrm{E}\left[Z_j Z_{j+m}\right]\Big) + O\left(\frac{1}{n}\right) = \int_0^{t_i} \tau_r(x)\sigma^4(x) f(x) dx + O(n^{-\gamma}) \end{aligned}$$

[recall that $\tau_r(x) = m_4(x) - 1 + 4\delta_r$] and for $1 \le i \le 2; 1 \le \ell \le d$

$$\begin{aligned} n\,\mathrm{E}\left[\bar{B}_{t_i}^0 \bar{c}_\ell\right] &= \int_0^{t_i} \tau_r(x)\sigma^4(x)\sigma_\ell^2(x) f(x)\, dx + O\left(n^{-\gamma}\right), \\ n\,\mathrm{E}\left[\bar{c}_i \bar{c}_\ell\right] &= \int_0^{1} \tau_r(x)\sigma^4(x)\sigma_i^2(x)\sigma_\ell^2(x) f(x)\, dx + O\left(n^{-\gamma}\right). \end{aligned}$$

Therefore it follows $\mathrm{Var}(X_n) = \Sigma_{t_1,t_2} + O(n^{-\gamma})$, where the matrix Σ_{t_1,t_2} is defined in Theorem 2.1.
For a proof of the asymptotic normality we introduce the notation $c = (a_1, a_2, b_1, \ldots, b_d)^T$ and show with the aid of a central limit Theorem for α-mixing arrays in Liebscher (1996) that

$$T_n = \frac{c^T X_n}{\sigma} = \frac{\sqrt{n}}{\sigma}\Big(a_1 \bar{B}^0_{t_1} + a_2 \bar{B}^0_{t_2} + \sum_{i=1}^{d} b_i \bar{c}_i\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{A.11}$$

where $\sigma^2 = c^T \Sigma_{t_1,t_2} c$ denotes the asymptotic variance of $c^T X_n$. For this we assume $t_1 \leq t_2$ and note that the statistic T_n can be represented as $T_n = \sum_{j=r+1}^{n} C_{n,j}$, where

$$C_{n,j} = \frac{\sqrt{n}}{\sigma(n-r)} \begin{cases} (a_1 + a_2 + \sum_{i=1}^{d} b_i \sigma_i^2(t_j))Z_j & j \leq \lfloor nt_1 \rfloor \\ (a_2 + \sum_{i=1}^{d} b_i \sigma_i^2(t_j))Z_j & \lfloor nt_1 \rfloor < j \leq \lfloor nt_2 \rfloor \\ \sum_{i=1}^{d} b_i \sigma_i^2(t_j) Z_j & j > \lfloor nt_2 \rfloor \end{cases}.$$

Obviously, $\{C_{n,j} \mid j = r+1, \ldots, n; n \in \mathbb{N}\}$ is a triangular array of $(r+1)$-dependent random variables and

$$\mathrm{E}\left[|Z_j|^3\right] \leq \mathrm{E}\left[L_j^6\right] + 3\,\mathrm{E}\left[L_j^4\right]\mathrm{E}\left[L_j^2\right] + 4(\mathrm{E}\left[L_j^2\right])^3. \tag{A.12}$$

Now a straightforward calculation gives $E|Z_j|^3 = O(1)$ and $E|Z_j|^4 = O(1)$ uniformly with respect to $j = r+1, \ldots, n$. As a consequence we obtain $E|C_{n,j}^3| = O(n^{-3/2})$ and $E|C_{n,j}^4| = O(n^{-2})$ uniformly with respect to $j = r+1, \ldots, n$. From the calculation of the covariance matrix of X_n it follows that $\lim_{n\to\infty} \mathrm{E}\left[T_n^2\right] = 1$, and the assumptions in the central limit theorem of Liebscher (1996) hold with $q = 4$ and $p = 3$, respectively. This theorem now yields the assertion (A.11), and as a consequence we obtain

$$\sigma T_n = c^T X_n \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, c^T \Sigma_{t_1,t_2} c\right).$$

By the Cramér-Wold device the weak convergence of the finite dimensional distributions and the statement in (A.6) follows.
In order to prove the remaining assertion (A.7) we introduce a further decomposition

$$\begin{aligned} \bar{A}_n(t) &= \frac{1}{n-r}\sum_{j=r+1}^{\lfloor nt \rfloor} Z_j - \frac{1}{n-r}\sum_{j=r+1}^{n} Z_j\{\hat{s}_{t,1}\sigma_1^2(t_j) + \ldots + \hat{s}_{t,d}\sigma_d^2(t_j)\} \\ &= \bar{A}_n^{(1)}(t) - \bar{A}_n^{(2)}(t), \end{aligned} \tag{A.13}$$

where the last equality defines the processes $\bar{A}_n^{(1)}$ and $\bar{A}_n^{(2)}$, $\hat{s}_{t,j} = \sum_{k=1}^{d} \hat{b}_{kj} \hat{B}_t^k$, and $\hat{b}_{ij}$ denotes the element in the ith row and jth column of the inverse of the matrix $\hat{A}$. Obviously, the assertion (A.7) follows from

$$\mathrm{E}\left[n^2 |\bar{A}_n^{(i)}(t) - \bar{A}_n^{(i)}(s)|^4\right] \leq C(t-s)^2, \quad i = 1, 2 \tag{A.14}$$

for some positive constant. For a proof of this property in the case $i = 1$ we use the representation $\bar{A}_n^{(1)}(t) - \bar{A}_n^{(1)}(s) = \frac{1}{n-r}\sum_{j=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} Z_j$ and obtain by a straightforward but tedious calculation

$$\beta_n = \mathrm{E}\left[n^2 |\bar{A}_n^{(1)}(t) - \bar{A}_n^{(1)}(s)|^4\right]$$

$$
\begin{aligned}
&= 3\Big[\Big(\frac{1}{n}\sum_{i=\lfloor ns\rfloor+1}^{\lfloor nt\rfloor} \mathrm{E}\left[Z_i^2\right]\Big)^2 + \Big(\frac{2}{n}\sum_{i=\lfloor ns\rfloor+1}^{\lfloor nt\rfloor-r}\sum_{k=1}^{r}\mathrm{E}\left[Z_iZ_{i+k}\right]\Big)^2 \\
&\quad + 2\Big(\frac{1}{n}\sum_{i=\lfloor ns\rfloor+1}^{\lfloor nt\rfloor} \mathrm{E}\left[Z_i^2\right]\cdot\frac{2}{n}\sum_{i=\lfloor ns\rfloor+1}^{\lfloor nt\rfloor-r}\sum_{k=1}^{r}\mathrm{E}\left[Z_iZ_{i+k}\right]\Big)\Big] + O\left(\frac{1}{n}\right) \\
&= 3\Big\{\frac{1}{n}\sum_{i=\lfloor ns\rfloor+1}^{\lfloor nt\rfloor-r}\Big(\mathrm{E}\left[Z_i^2\right] + 2\sum_{k=1}^{r}\mathrm{E}\left[Z_iZ_{i+k}\right]\Big)\Big\}^2 + O\left(\frac{1}{n}\right) \\
&= 3\left(\int_s^t \tau_r(x)\sigma^4(x) f(x)\,dx\right)^2 + O\left(n^{-\gamma}\right)
\end{aligned}
$$

uniformly with respect to $s,t \in [0,1]$. The estimate (A.14) in the case $i = 1$ is now obvious from the mean value theorem.
In order to derive a similar estimate for the process $\bar{A}_n^{(2)}$ we note that $E[n^2|\bar{A}_n^{(2)}(t) - \tilde{A}_n^{(2)}(s)|^4] = o(1)$ uniformly with respect to $t \in [0,1]$, where the process $\tilde{A}_n^{(2)}$ is defined by

$$
\tilde{A}_n^{(2)}(t) = \frac{1}{n-r}\sum_{j=r+1}^{n} Z_j\{s_{t,1}\sigma_1^2(t_j) + \ldots + s_{t,d}\sigma_d^2(t_j)\}
$$

with $s_{t,j} = \sum_{k=1}^d b_{kj}B_t^k$ and b_{kj} denotes the element in the kth row and jth column of the inverse of the matrix A. Obviously, we have for some constants $C_1,\ldots,C_d$

$$
s_{t,j} - s_{s,j} = \sum_{k=1}^{d} b_{kj}\Big(\int_s^t \sigma_k^2(x) f(x)\,dx\Big) = (t-s)\sum_{k=1}^{d} b_{kj}C_k,
$$

and obtain

$$
\begin{aligned}
\tilde{A}_n^{(2)}(t) - \tilde{A}_n^{(2)}(s) &= \frac{1}{n-r}\sum_{j=r+1}^{n} Z_j\Big\{(t-s)\sum_{k=1}^{d} b_{k1}C_k\sigma_1^2(t_j) + \ldots + (t-s)\sum_{k=1}^{d} b_{kd}C_k\sigma_d^2(t_j)\Big\} \\
&= \frac{t-s}{n-r}\sum_{j=r+1}^{n}\mu_j Z_j,
\end{aligned}
$$

where the constants μ_j are defined by $\mu_j = \sum_{i=1}^d (\sum_{k=1}^d b_{ki}C_k)\sigma_i^2(t_j)$. A similar calculation as used in the proof of the tightness of the process $\bar{A}_n^{(1)}$ shows that the inequality (A.14) also holds in the case $i = 2$. This establishes the remaining condition (A.7) and the proof of Theorem 2.1 is completed. □

Proof of Theorem 5.1. For the sake of brevity we only sketch the main difference in the proof, which emerges in the different variance of the empirical process in the case of a random design. Let $\tilde{c}_i$ and $\tilde{B}_t^0$ be defined as in (5.2) and (5.3), where the random variables R_j are replaced by the variables $L_j = \sigma(X_{(j)})\sum_{i=0}^r d_i\varepsilon_{A_{j-i}}$. By the Lipschitz continuity of the regression function the limiting behaviour of the process $\hat{S}_t$ is not changed by this replacement. For the calculation of the asymptotic covariance we now use the random variables $\tilde{B}_{t_1}^0$ and $\tilde{B}_{t_2}^0$ (with $0 \le t_1 \le t_2 \le 1$) and the formula

$$
\text{(A.15)} \qquad \mathrm{Cov}(\tilde{B}_{t_1}^0, \tilde{B}_{t_2}^0) = \mathrm{Cov}(\mathrm{E}[\tilde{B}_{t_1}^0|\mathcal{F}_n], \mathrm{E}[\tilde{B}_{t_2}^0|\mathcal{F}_n]) + \mathrm{E}[\mathrm{Cov}(\tilde{B}_{t_1}^0, \tilde{B}_{t_2}^0|\mathcal{F}_n)],
$$

where $\mathcal{F}_n$ denotes the $\sigma-$field generated by the order statistics $X_{(1)}, \ldots, X_{(n)}$. For the conditional expectation we have

$$\mathrm{E}[\tilde{B}_t^0|\mathcal{F}_n] \;=\; \frac{1}{n-r}\sum_{j=r+1}^{n}\sigma^2(X_{(j)})I\{X_{(j)}\leq t\}\,\mathrm{E}\,[(\sum_{i=0}^{r}d_i\varepsilon_{A_{j-i}})^2|\mathcal{F}_n] = \frac{1}{n}\sum_{j=1}^{n}\sigma^2(X_j)I\{X_j\leq t\}+o(1),$$

and an easy calculation gives for $t_1 \leq t_2$

$$n\,\mathrm{Cov}\,(\,\mathrm{E}\,[\tilde{B}_{t_1}^0|\mathcal{F}_n],\mathrm{E}\,[\tilde{B}_{t_2}^0|\mathcal{F}_n]) = \int_0^{t_1}\sigma^4\,(x)\,f\,(x)\;dx - B_{t_1}^0B_{t_2}^0 + o(1). \tag{A.16}$$

For the second term in equation (A.15) we obtain

$$\begin{aligned}\mathrm{Cov}\,(\tilde{B}_{t_1}^0,\tilde{B}_{t_2}^0|\mathcal{F}_n) \;=\;& \frac{1}{(n-r)^2}\Big\{\sum_{j=r+1}^{n}\sigma^4(X_{(j)})I\{X_{(j)}\leq t_1\}\mathrm{Var}\Big((\sum_{i=0}^{r}d_i\varepsilon_{A_{j-i}})^2|\mathcal{F}_n\Big)\\ &+2\sum_{m=1}^{r}\sum_{j=r+1}^{n-r}\sigma^4(X_{(j)})I\{X_{(j)}\leq t_1\}\,\mathrm{Cov}\,\Big((\sum_{i=0}^{r}d_i\varepsilon_{A_{j-i}})^2,(\sum_{i=0}^{r}d_i\varepsilon_{A_{k-i}})^2|\mathcal{F}_n\Big)\Big\}\\ &+o_p(1).\end{aligned}$$

Observing that

$$\begin{aligned}\mathrm{Var}\Big((\sum_{i=0}^{r}d_i\varepsilon_{A_{j-i}})^2|\mathcal{F}_n\Big)+2\sum_{m=1}^{r}\mathrm{Cov}\,\Big((\sum_{i=0}^{r}d_i\varepsilon_{A_{j-i}})^2,(\sum_{i=0}^{r}d_i\varepsilon_{A_{k-i}})^2|\mathcal{F}_n\Big)&\\ = m_4\left(X_{(j)}\right)-1+4\delta_r+o_p(1),&\end{aligned}$$

it follows that

$$\begin{aligned}n\,\mathrm{Cov}\,((\tilde{B}_{t_1}^0,\tilde{B}_{t_2}^0)|\mathcal{F}_n) \;&=\; \frac{n}{(n-r)^2}\sum_{j=r+1}^{n}\sigma^4(X_{(j)})I\{X_{(j)}\leq t_1\}+o_p(1)\\ &=\; \frac{1}{n}\sum_{j=1}^{n}\sigma^4(X_j)(m_4(X_j)-1+4\delta_r)I\{X_j\leq t_1\}+o_p(1)\end{aligned}$$

and

$$\mathrm{E}\,[\,\mathrm{Cov}\,((\tilde{B}_{t_1}^0,\tilde{B}_{t_2}^0)|\mathcal{F}_n)] = \int_0^{t_1}\sigma^4(x)\tau_r(x)f(x)\,dx+o(1).$$

Note that this expression is exactly the same as the asymptotic covariance calculated in the fixed design case. From (A.16) we obtain the representation of $\bar{v}_{ij}$ in (5.7), and formula (A.15) yields the representation of the corresponding element in the matrix $\bar{\Sigma}_{t_1,t_2}$. The other elements in the matrix $\bar{\Sigma}_{t_1,t_2}$ are calculated exactly in the same way and the details are omitted for the sake of brevity. □

Acknowledgements. The work of the authors was supported by the Deutsche Forschungsgmeinschaft (SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen). The authors are also grateful to Isolde Gottschlich, who typed parts of this paper with considerable expertise.

References

N.J. Achieser (1956). Theory of Approximation. Dover Publications, Inc., N.Y.

P. Bickel (1978). Using residuals robustly I: Tests for heteroscedasticity. Ann. Statist. 6, 266-291.

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

R.J. Carroll, D. Ruppert (1981). On robust tests for heteroscedasticity. Ann. Statist. 9, 205 - 209.

R.J. Carroll, D. Ruppert (1988). Transformation and weighting in regression. Monographs on Statistics and Applied Probability. New York, Chapman and Hall.

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

M. Davidan, R. Caroll (1987). Variance function estimation. J. Amer. Statist. Assoc. 82, 1079-1091.

H. Dette (2002). A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. J. Statist. Planning Infer., 103, 311–329.

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

H. Dette, A. Munk, T. Wagner (1998). Estimating the variance in nonparametric regression - what is a reasonable choice? J. Roy. Statist. Soc., Ser. B, 60, 751-764.

A. Diblasi, A. Bowman (1997). Testing for constant variance in a linear model. Statist. Probab. Lett., 33, 95–103.

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

A.R. Gallant (1987). Nonlinear Statistical Models. Wiley, N.Y.

T., Gasser, L. Sroka, G. Jennen-Steinmetz (1986). Residual variance and residual pattern in nonlinear regression. Biometrika 73, 625-633.

W.E. Griffiths, R.C. Hill, G.G. Judge (1993). Learning and practicing econometrics. Wiley, N.Y.

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

Y. Leedan, P. Meer (2000). Heteroscedastic regression in computer vision: problems with bilinear constraint. Int. J. of Comput. Vision 37, 127-150.

E. Liebscher (1996). Central limit theorems for sums of α-mixing random variables. Stochastics and Stochastic Rep. 59, 241-258.

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

A. Munk (2002). Testing the goodness of fit of parametric regression models with random Toeplitz forms. Scand. J. Statist. 29, 501-533.

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

S. Sadray, S. Rezaee, S. Rezakhah (2003). Non-linear heteroscedastic regression model for determination of methrotrexate in human plasma by high-performance liquid chromatography. Journal of Chromatography B, 787, 293-302.

G.A.F. Seber, G.J. Wild (1989). Nonlinear Regression. Wiley, N.Y.

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