

# A test for the parametric form of the variance function in a partial linear regression model

Holger Dette

Ruhr-Universität Bochum  
Fakultät für Mathematik  
44780 Bochum, Germany

e-mail: holger.dette@ruhr-uni-bochum.de

Mareen Marchlewski

Ruhr-Universität Bochum  
Fakultät für Mathematik  
44780 Bochum, Germany

email: mareen.marchlewski@ruhr-uni-bochum.de

August 13, 2007

## Abstract

We consider the problem of testing for a parametric form of the variance function in a partial linear regression model. A new test is derived, which can detect local alternatives converging to the null hypothesis at a rate  $n^{-1/2}$  and is based on a stochastic process of the integrated variance function. We establish weak convergence to a Gaussian process under the null hypothesis, fixed and local alternatives. In the special case of testing for homoscedasticity the limiting process is a scaled Brownian bridge. We also compare the finite sample properties with a test based on an  $L^2$ -distance, which was recently proposed by You and Chen (2005).

## 1 Introduction

Partial linear regression models have found considerable interest in the recent literature, because they combine the attractive features of linear models (such as interpretability of the parameter estimates or well established theoretical properties) with the more flexible concept of nonparametric regression [see e.g. Green and Silverman (1994), Yatchew (1997), Härdle, Liang and Gao (2000) among many others]. Typically the model is defined as

$$(1.1) \quad Y_i = x_i^T \beta + m(t_i) + \sigma(t_i) \varepsilon_i, \quad i = 1, \dots, n,$$

where the  $Y_i$  are the responses,  $\beta = (\beta_1, \dots, \beta_p)^T$  is a vector of unknown parameters and  $m$  and  $\sigma$  are smooth functions. The vectors  $x_i^T = (x_{i1}, \dots, x_{ip})$  and the real numbers  $t_i$  ( $i = 1, \dots, n$ ) are fixed design points and  $\varepsilon_1, \dots, \varepsilon_n$  denote random variables with mean 0 and variance 1. Much effort has been spent on the problem of testing hypotheses regarding  $\beta$  or  $m$  [see e.g. Gao (1997), Fan and Huang (2001), González-Manteiga and Aneiros-Pérez (2003), Aneiros-Pérez, González-Manteiga and Vieu (2004), Bianco, Boente and Martínez (2006) among many others] but less literature is available on the problem of testing hypotheses regarding the variance function  $\sigma$ .

In the purely nonparametric model  $Y_i = m(t_i) + \sigma(t_i)\varepsilon_i$  several authors have emphasized the importance of detecting heteroscedasticity and have proposed various tests for heteroscedasticity [see e.g. Koenker and Basset (1981), Cook and Weisberg (1983), Diblasi and Bowman (1997), Dette and Munk (1998a), Liero (2003) among many others]. Recently, You and Chen (2005) proposed a test for homoscedasticity in the partial linear regression model (1.1), which is based on an estimate of the  $L^2$ -distance between the variance function  $\sigma^2(\cdot)$  and its best constant approximation. This test is - to the knowledge of the authors - the only procedure which has been proposed for testing the hypothesis of homoscedasticity in a partial linear regression model of the form (1.1). It can detect local hypotheses, which converge to the null hypothesis of homoscedasticity at a rate  $n^{-1/4}$ , where  $n$  denotes the sample size.

The present paper has two purposes. On the one hand we are interested in a test for a homoscedastic error structure which is more efficient with respect to Pitman alternatives, on the other hand we will also consider the more general problem of testing for a parametric form of the variance function, that is

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

where  $\sigma^2(t, \theta)$  is a known function,  $\theta = (\theta_1, \dots, \theta_d)^T \in \Theta \subset \mathbb{R}^d$  an unknown vector of parameters and the set  $\Theta$  is assumed to be compact. Note that the hypothesis of homoscedasticity is obtained for  $d = 1$  and  $\sigma^2(t, \theta) = \theta$ , but many other hypotheses are of interest in practical applications.

In Section 2 we introduce two stochastic processes, which will be used as the basis for constructing test statistics for the hypothesis (1.2). The basic idea is to compare estimates of the integrated variance  $\int_0^t \sigma^2(u)du$  under the null hypothesis and the alternative. Weak convergence of this process to a centered Gaussian process and the corresponding statistical applications are discussed in Section 3. In particular Kolmogorov-Smirnov and Cramér-von-Mises type tests are proposed and it is demonstrated that the new tests can detect local alternatives converging to the null hypothesis at a rate  $n^{-1/2}$ . We also discuss the asymptotic properties of the test in the case of a random design, which differ from the results obtained under the fixed design assumption. In Section 4 we present a small simulation study and compare the new tests with a test which has recently been suggested by You and Chen (2005). We also illustrate the application of the test by means of a data example. Finally, some technical details are given in an Appendix in Section 5.

## 2 Two new tests for the variance function in partial linear regression models

Recall the definition of the partial linear regression model (1.1). We consider a triangular array of random variables without mentioning this in our notation (that is - we write  $Y_i, t_i, x_i$  and  $\varepsilon_i$  instead of  $Y_{i,n}, t_{i,n}, x_{i,n}$  and  $\varepsilon_{i,n}$ , respectively). For the explanatory variables we consider a fixed design satisfying

$$(2.1) \quad \frac{i}{n+1} = \int_0^{t_i} f(t) dt, \quad i = 1, \dots, n,$$

for some positive density  $f$  on the interval  $[0, 1]$  [see Sacks and Ylvisacker (1970)] and

$$(2.2) \quad \|x_i\| \leq c, \quad i = 1, \dots, n,$$

for some constant  $c \in \mathbb{R}^+$  (here and throughout this paper  $\|\cdot\|$  denotes the euclidean norm). The case of a random design will briefly be discussed in Section 3.2. If  $m_j(t) = E[\varepsilon_i^j]$  ( $j = 3, 4$ ) denotes the third and fourth moment of the error (which may depend on  $t$ ) we further assume

$$(2.3) \quad f, \sigma, m, m_3, m_4 \in \text{Höl}_\gamma[0, 1]$$

for some  $\gamma > \frac{1}{2}$ , where  $\text{Höl}_\gamma[0, 1]$  denotes the class of all Hölder continuous functions of order  $\gamma$  defined on the interval  $[0, 1]$ . The basic idea for the construction of the testing procedure for the hypothesis (1.2) is to eliminate the effect of the linear regression component in the partial linear regression model (1.1) and for this purpose two methods are considered.

The first approach is based on an estimate for the parameter  $\beta$ , which was essentially suggested in a paper by Speckman (1988). This author proposed the estimate

$$(2.4) \quad \hat{\beta}_n = (\hat{X}^T \hat{X})^{-1} \hat{X}^T \hat{Y},$$

where

$$\begin{aligned} \hat{X} &= (I_n - \hat{K})X, \\ \hat{Y} &= (I_n - \hat{K})Y, \\ X &= (x_1, \dots, x_n)^T, \quad Y = (Y_1, \dots, Y_n)^T, \\ \hat{K} &= (W_i(t_j, h))_{1 \leq i, j \leq n}, \end{aligned}$$

and  $W_i(t_j, h)$  denote the weights of the local linear estimator at the points  $t_i$ , that is

$$(2.5) \quad W_i(t_j, h) = \frac{\hat{s}_2(t_j, h) - \hat{s}_1(t_j, h)(t_i - t_j)}{\hat{s}_2(t_j, h)\hat{s}_0(t_j, h) - \hat{s}_1^2(t_j, h)} K\left(\frac{t_i - t_j}{h}\right),$$

where

$$\hat{s}_l(t_j, h) = \sum_{k=1}^n K\left(\frac{t_k - t_j}{h}\right) (t_k - t_j)^l,$$

with a kernel  $K$  and a bandwidth  $h$  converging to 0 with increasing sample size.

Note that Speckman (1988) considered simpler weights and the homoscedastic partial linear regression models, but it can be shown by similar arguments that the statistic  $\hat{\beta}_n$  is  $\sqrt{n}$ -consistent, if the limit

$$(2.6) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \hat{X}^T \hat{X}$$

exists. We now define modified data by

$$(2.7) \quad Y_i^* = Y_i - x_i^T \hat{\beta}_n, \quad i = 1, \dots, n,$$

and consider the pseudo residuals

$$(2.8) \quad R_j^* = \sum_{i=0}^r d_i Y_{j-i}^*, \quad j = r+1, \dots, n,$$

where  $(d_0, \dots, d_r)$  is a difference sequence satisfying

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

[see Gasser, Sroka and Jennen-Steinmetz (1986) or Hall, Kay and Titterton (1990)]. Note that by the consistency of the estimate  $\hat{\beta}_n$  and by (2.3) and (2.9) it is intuitively clear that

$$(2.10) \quad R_j^* \approx \sum_{i=0}^r d_i (Y_{j-i} - x_{j-i}^T \beta) = \sum_{i=0}^r d_i m(t_{j-i}) + \sum_{i=0}^r d_i \sigma(t_{j-i}) \varepsilon_{j-i} \approx \sum_{i=0}^r d_i \sigma(t_{j-i}) \varepsilon_{j-i},$$

which implies

$$(2.11) \quad E[R_j^{*2}] \approx \sum_{i=0}^r d_i^2 \sigma^2(t_{j-i}) \approx \sigma^2(t_j).$$

Consequently analysis of the variance function can be based on the pseudo residuals  $R_j^*$ . For this we define

$$(2.12) \quad \hat{\theta}^* = \arg \min_{\theta \in \Theta} \sum_{i=r+1}^n \left( R_i^{*2} - \sigma^2(t_i, \theta) \right)^2$$

as the least squares estimate of the value  $\theta_0$ , which is defined as

$$(2.13) \quad \theta_0 = \arg \min_{\theta \in \Theta} \int_0^1 (\sigma^2(t) - \sigma^2(t, \theta))^2 f(t) dt.$$

Throughout this paper it is assumed that  $\theta_0$  exists, is unique and an interior point of the compact set  $\Theta$ . We also assume that all partial derivatives up to order three of  $\sigma^2(t, \theta)$  with respect to the components of  $\theta$  exist and are continuous in  $t$  and  $\theta$ . We now define for  $t \in [0, 1]$  the stochastic process

$$(2.14) \quad \hat{S}_t^* = \frac{1}{n-r} \sum_{i=r+1}^n 1_{\{t_i \leq t\}} R_i^{*2} - \frac{1}{n} \sum_{i=1}^n 1_{\{t_i \leq t\}} \sigma^2(t_i, \hat{\theta}^*).$$

It is heuristically clear that  $\hat{S}_t^*$  is an estimate of the (deterministic) process

$$(2.15) \quad S_t = \int_0^t (\sigma^2(u) - \sigma^2(u, \theta)) f(u) du,$$

which vanishes (a.e.) for all  $t \in [0, 1]$  if and only if the null hypothesis (1.2) is valid. Consequently, the hypothesis can be rejected for large values of the Cramér-von-Mises or Kolmogorov-Smirnov type statistics

$$(2.16) \quad C_n^* = n \int_0^1 |\hat{S}_t^*|^2 F_n(dt), \quad K_n^* = \sqrt{n} \sup_{t \in [0,1]} |\hat{S}_t^*|.$$

The asymptotic properties of these statistics will be discussed in Section 3.

Our second method for constructing test statistics for the hypothesis (1.2) in the partial linear regression model is based on the observation that

$$(2.17) \quad \check{Y}_i := Y_{i+1} - Y_i \approx (x_{i+1}^T - x_i^T)\beta + \check{\varepsilon}_i,$$

where  $\check{\varepsilon}_i = \sigma(t_{i+1})\varepsilon_{i+1} - \sigma(t_i)\varepsilon_i$ , and the approximation is motivated by the Hölder continuity of the function  $m$ . We introduce the notation  $\check{x}_i = x_{i+1} - x_i$  ( $i = 1, \dots, n-1$ ),  $\check{Y} = (\check{Y}_1, \dots, \check{Y}_{n-1})^T$ ,  $\check{X} = (\check{x}_1, \dots, \check{x}_{n-1})^T$ , then the estimate

$$(2.18) \quad \check{\beta}_n = (\check{X}^T \check{X})^{-1} \check{X}^T \check{Y}$$

is  $\sqrt{n}$  consistent, if the limit

$$(2.19) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \check{X}^T \check{X}$$

exists and is non-singular. In the same way as in the previous paragraph we now define pseudo residuals as

$$(2.20) \quad R_j^{**} = \sum_{i=0}^r d_i Y_{j-i}^{**}, \quad j = r+1, \dots, n,$$

where

$$(2.21) \quad Y_i^{**} = Y_i - x_i^T \check{\beta}_n, \quad i = 1, \dots, n.$$

This yields to the alternative stochastic process given by

$$(2.22) \quad \hat{S}_t^{**} = \frac{1}{n-r} \sum_{j=r+1}^n 1_{\{t_j, n \leq t\}} R_j^{**2} - \frac{1}{n} \sum_{i=1}^n 1_{\{t_i \leq t\}} \sigma^2(t_i, \hat{\theta}^{**}),$$

$t \in [0, 1]$ , where the value  $\hat{\theta}^{**}$  is given by

$$(2.23) \quad \hat{\theta}^{**} = \arg \min_{\theta \in \Theta} \sum_{i=r+1}^n (R_i^{**2} - \sigma^2(t_i, \theta))^2.$$

Again, by  $\sqrt{n}$ -consistency of the estimate  $\check{\beta}_n$ , it is intuitively clear that  $\{\hat{S}_t^{**}\}_{t \in [0,1]}$  is a consistent estimate of the stochastic process  $\{S_t\}_{t \in [0,1]}$  defined in (2.1). The asymptotic properties of the processes  $\{\hat{S}_t^*\}_{t \in [0,1]}$  and  $\{\hat{S}_t^{**}\}_{t \in [0,1]}$  will be investigated in the following section.

### 3 Asymptotic properties

In this section we present several results on the weak convergence of the stochastic processes  $\{\hat{S}_t^*\}_{t \in [0,1]}$  and  $\{\hat{S}_t^{**}\}_{t \in [0,1]}$  in the partial linear regression model and in several extensions of this model.

### 3.1 The partial linear regression model with fixed predictors

Our first result specifies the asymptotic properties in the situation described in Section 2. Before we give the precise result, we note that by the assumptions made in Section 2 we have

$$(3.1) \quad 0 = \frac{\partial}{\partial \theta_j} \int_0^1 (\sigma^2(x) - \sigma^2(x, \theta))^2 f(x) dx \Big|_{\theta=\theta_0} = -2 \int_0^1 \sigma_j^2(x) (\sigma^2(x) - \sigma^2(x, \theta_0)) f(x) dx,$$

where

$$(3.2) \quad \sigma_j^2(u) = \frac{\partial}{\partial \theta_j} \sigma^2(u, \theta) \Big|_{\theta=\theta_0}, \quad j = 1, \dots, d,$$

denote the partial derivatives of the variance function with respect to the parameters (at the point  $\theta_0$ ).

**Theorem 3.1.** *Assume that the assumptions (2.1) - (2.3), (2.6) or (2.19) are satisfied, and define the process  $\{\hat{S}_t\}_{t \in [0,1]}$  either as  $\{\hat{S}_t^*\}_{t \in [0,1]}$  [see (2.14)] or as  $\{\hat{S}_t^{**}\}_{t \in [0,1]}$  [see (2.22)], then the stochastic process  $\{\sqrt{n}(\hat{S}_t - S_t)\}_{t \in [0,1]}$  converges weakly in  $D[0,1]$  to a centered Gaussian process  $G$  with covariance kernel*

$$(3.3) \quad k(t_1, t_2) = (0, 1) V_2 \Sigma_{t_1, t_2} V_2^T (1, 0)^T$$

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

$$(3.4) \quad \Sigma_{t_1, t_2} = \begin{pmatrix} v_{11} & v_{12} & w_{11} & \cdots & w_{1d} \\ v_{21} & v_{22} & w_{21} & \cdots & w_{2d} \\ w_{11} & w_{21} & z_{11} & \cdots & z_{1d} \\ \vdots & \vdots & \vdots & & \vdots \\ w_{1d} & w_{2d} & z_{d1} & \cdots & z_{dd} \end{pmatrix},$$

$$(3.5) \quad V_2 = (I_2 \mid U), \quad U = -(B_{t_1}^T A^{-1}, B_{t_2}^T A^{-1})^T,$$

$$(3.6) \quad A = (a_{ij})_{1 \leq i, j \leq d}, \quad B_t^T = (B_t^1, \dots, B_t^d)$$

and

$$B_t^i = \int_0^t \sigma_i^2(s) f(s) ds, \quad 1 \leq i \leq d$$

$$a_{ij} = \int_0^1 \sigma_i^2(s) \sigma_j^2(s) f(s) ds, \quad 1 \leq i, j \leq d$$

$$v_{ij} = \int_0^1 \tau_r(s) \sigma^4(s) 1_{[0, t_i \wedge t_j]}(s) f(s) ds, \quad 1 \leq i, j \leq 2$$

$$w_{ij} = \int_0^1 \tau_r(s) \sigma^4(s) \sigma_j^2(s) 1_{[0, t_i]}(s) f(s) ds, \quad 1 \leq i \leq 2, 1 \leq j \leq d$$

$$z_{ij} = \int_0^1 \tau_r(s) \sigma^4(s) \sigma_i^2(s) \sigma_j^2(s) f(s) ds, \quad 1 \leq i, j \leq d$$

$$\tau_r(s) = m_4(s) - 1 + 4\delta_r, \quad \delta_r = \sum_{m=1}^r \left( \sum_{j=0}^{r-m} d_j d_{j+m} \right)^2.$$

We note that the processes  $\hat{S}_t^*$  and  $\hat{S}_t^{**}$  exhibit the same asymptotic behaviour as the corresponding process considered by Dette and Hetzler (2006) in the classical nonparametric regression model. Consequently, we obtain from Corollary 2.7 in this reference:

**Corollary 3.2.** *Assume that the hypothesis of homoscedasticity  $H_0 : \sigma^2(t) = \theta_1$  has to be tested (i.e.  $d = 1, \sigma_1^2(t) = 1$ ), that the assumptions of Theorem 3.1 are satisfied and that additionally  $m_4(t) \equiv m_4$  is constant. Let  $\{\hat{S}_t\}_{t \in [0,1]}$  denote either the process  $\{\hat{S}_t^*\}_{t \in [0,1]}$  defined in (2.14) or the process  $\{\hat{S}_t^{**}\}_{t \in [0,1]}$  defined in (2.22), then under the null hypothesis of homoscedasticity the process  $\{\sqrt{n}(\hat{S}_t - S_t)\}_{t \in [0,1]}$  converges weakly in  $D[0,1]$  to a scaled Brownian bridge in time  $F$ , where  $F$  is the distribution function of the design density, i.e.*

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

**Remark 3.3.** The test based on the process  $\{\sqrt{n}(\hat{S}_t - S_t)\}_{t \in [0,1]}$ , with  $\hat{S}_t$  either  $\hat{S}_t^*$  or  $\hat{S}_t^{**}$ , can detect alternatives of the form

$$\sigma^2(t) = \sigma^2(t, \theta_0) + n^{-1/2}h(t),$$

whenever

$$(3.7) \quad h \notin \text{span} \left\{ \frac{\partial}{\partial \theta_1} \sigma^2(\cdot, \theta) \Big|_{\theta=\theta_0}, \dots, \frac{\partial}{\partial \theta_d} \sigma^2(\cdot, \theta) \Big|_{\theta=\theta_0} \right\}.$$

Here  $h : [0, 1] \rightarrow \mathbb{R}$  denotes a fixed function, such that the variance function  $\sigma^2(t)$  is nonnegative for all  $t \in [0, 1]$ . Condition (3.7) results from the weak convergence of the process  $\{\sqrt{n}(\hat{S}_t - S_t)\}_{t \in [0,1]}$  to the process

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

where  $\{G(t)\}_{t \in [0,1]}$  denotes the limiting process defined in Theorem 3.1 and the coefficients  $\varphi_j$  are defined by

$$(\varphi_1, \dots, \varphi_d)^T = \arg \min_{\phi \in \mathbb{R}^d} \int_0^1 \left( h(x) - \sum_{j=1}^d \phi_j \frac{\partial}{\partial \theta_j} \sigma^2(x, \theta) \Big|_{\theta=\theta_0} \right)^2 f(x) dx.$$

## 3.2 Random predictors

As observed in Dette and Munk (1998b) the limit distribution of test statistics for goodness of fit tests in nonparametric regression models may be different for fixed and random predictors. For this reason we demonstrate in this subsection the effect of random predictors on the asymptotic

properties of the stochastic processes  $\{\hat{S}_t^*\}_{t \in [0,1]}$  and  $\{\hat{S}_t^{**}\}_{t \in [0,1]}$ . We have to distinguish several cases, corresponding to random and nonrandom  $x_i$  and  $t_i$ . We concentrate on the case, where the points  $t_i$  are random and  $x_i$  are fixed design points. The other cases are briefly discussed in Remark 3.6.

To be precise we consider the partial linear regression model

$$(3.8) \quad Y_i = x_i^T \beta + m(T_i) + \sigma(T_i) \varepsilon_i, \quad i = 1, \dots, n,$$

where  $x_1, \dots, x_n$  are fixed explanatory variables satisfying assumption (2.2) and  $T_1, \dots, T_n$  are i.i.d. with positive density  $f$  on the interval  $[0, 1]$ . We denote by  $m_j(t) = E[\varepsilon^j | T = t]$  the  $j$ th conditional moment of the errors and assume that  $m_6(t)$  is bounded by some constant, say  $m_6$ . We consider the processes  $\{\hat{S}_t^*\}_{t \in [0,1]}$  and  $\{\hat{S}_t^{**}\}_{t \in [0,1]}$  defined in Section 2 and 3 where the fixed design points  $t_i$  have been replaced by the random variables  $T_{(i)}$ , and  $T_{(1)} \leq \dots \leq T_{(n)}$  denotes the order statistic of  $T_1, \dots, T_n$ . The pseudo residuals are defined by

$$(3.9) \quad \hat{R}_j = \sum_{i=0}^r d_i \hat{Y}_{A_{j-i}}, \quad j = r+1, \dots, n,$$

with  $A_1, \dots, A_n$  denoting the antiranks of  $T_1, \dots, T_n$  and  $\hat{Y}_j$  is either  $Y_j^*$  or  $Y_j^{**}$  corresponding to the two cases considered in Section 2.

The following result shows that in the case of the random design the stochastic processes  $\{\hat{S}_t^*\}_{t \in [0,1]}$  and  $\{\hat{S}_t^{**}\}_{t \in [0,1]}$  have a different asymptotic behaviour. The proof follows from the fact that the random design assumption regarding the explanatory variables  $T_1, \dots, T_n$  does not change the asymptotic properties of the estimates  $\hat{\beta}_n$  and  $\check{\beta}_n$  defined in Section 2. As a consequence the same arguments as given in Section 3.1 show that the asymptotic behaviour of the processes is the same as that of the corresponding processes obtained in the nonparametric regression model  $Y_i = m(T_i) + \sigma(T_i) \varepsilon_i$ , which was established in Dette und Hetzler (2006).

**Theorem 3.4.** *Consider the partial linear regression model (3.8), assume that the assumptions (2.1) - (2.3), (2.6) or (2.19) are satisfied, and define the process  $\hat{S}_t$  either as  $\hat{S}_t^*$  or as  $\hat{S}_t^{**}$  (with the obvious modification for the random design assumption), then the stochastic process  $\{\sqrt{n}(\hat{S}_t - S_t)\}_{t \in [0,1]}$  converges weakly in  $D[0, 1]$  to a centered Gaussian process  $\bar{G}$  with covariance kernel  $\bar{k}_{t_1, t_2} = (0, 1) V_2 \bar{\Sigma}_{t_1, t_2} V_2^T (1, 0)^T \in \mathbb{R}^{2 \times 2}$ , where  $\bar{\Sigma}_{t_1, t_2} = \Sigma_{t_1, t_2} + \Phi_{t_1, t_2}$ , the matrix  $\Sigma_{t_1, t_2}$  is given in (3.4),*

$$(3.10) \quad \Phi_{t_1, t_2} = \begin{pmatrix} \bar{v}_{11} & \bar{v}_{12} & \bar{w}_{11} & \cdots & \bar{w}_{1d} \\ \bar{v}_{21} & \bar{v}_{22} & \bar{w}_{21} & \cdots & \bar{w}_{2d} \\ \bar{w}_{11} & \bar{w}_{21} & \bar{z}_{11} & \cdots & \bar{z}_{1d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{w}_{1d} & \bar{w}_{2d} & \bar{z}_{d1} & \cdots & \bar{z}_{dd} \end{pmatrix}$$

and the elements of the matrix  $\Phi_{t_1, t_2}$  are defined by

$$(3.11) \quad \bar{v}_{ij} = \int_0^{t_i \wedge t_j} \sigma^4(s) f(s) ds - \int_0^{t_i} \sigma^2(s) f(s) ds \int_0^{t_j} \sigma^2(s) f(s) ds,$$



$$\begin{aligned}\bar{w}_{ij} &= \int_0^{t_i} \sigma^4(s) \sigma_j^2(s) f(s) ds - \int_0^{t_i} \sigma^2(s) f(s) ds \int_0^1 \sigma^2(s) \sigma_j^2(s) f(s) ds, \\ \bar{z}_{ij} &= \int_0^1 \sigma^4(s) \sigma_i^2(s) \sigma_j^2(s) f(s) ds - \int_0^1 \sigma^2(s) \sigma_i^2(s) f(s) ds \int_0^1 \sigma^2(s) \sigma_j^2(s) f(s) ds.\end{aligned}$$

Note that it follows from Theorem 3.1 and 3.4 that the weak limit of the process  $\{\hat{S}_t\}_{t \in [0,1]}$  is different for the random and fixed design assumption for the explanatory variables  $t_i$ , which was also observed by Dette and Munk (1998b), who considered an  $L^2$ -type test for the parametric form of the regression function. However, in the case considered by these authors there is no difference between the two cases in the limit distribution under the null hypothesis. Only under the fixed alternative different distributions are observable. For the processes considered here the limit distributions are even different under the null hypothesis. Consider for example the problem of testing for homoscedasticity  $H_0 : \sigma^2(t) = \theta_1$ . For the fixed design the limit distribution is specified in Corollary 3.2, while Theorem 3.4 gives the following result for random predictors.

**Corollary 3.5.** *Assume that the hypothesis of homoscedasticity  $H_0 : \sigma^2(t) = \theta_1$  has to be tested (i.e.  $d = 1, \sigma_1^2(t) = 1$ ), that the assumptions of Theorem 3.4. are satisfied and that additionally  $m_4(t) \equiv m_4$  is constant. Let  $\{\hat{S}_t\}_{t \in [0,1]}$  denote either the process  $\{\hat{S}_t^*\}_{t \in [0,1]}$  defined in (2.14) or the process  $\{\hat{S}_t^{**}\}_{t \in [0,1]}$  defined in (2.22) (with the obvious modification for the random design assumption), then under the null hypothesis of homoscedasticity the process  $\{\sqrt{n}(\hat{S}_t - S_t)\}_{t \in [0,1]}$  converges weakly in  $D[0,1]$  to a scaled Brownian bridge in time  $F$ , where  $F$  is the distribution function of the design density, i.e.*

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

**Remark 3.6.** The case of random predictors  $X_i$  can be considered in a similar manner and it can be shown that the assumption regarding the randomness of the parametric part has no effect on the asymptotic distribution of the stochastic processes  $\{\hat{S}_t^*\}_{t \in [0,1]}$  and  $\{\hat{S}_t^{**}\}_{t \in [0,1]}$ . More precisely consider the model

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

where  $T_1, \dots, T_n$  are i.i.d. with positive density  $f$  on the interval  $[0, 1]$  and  $X_1, \dots, X_n$  are i.i.d with density  $g$  having compact support such that the analogs of (2.6) and (2.19) hold in probability. In this case Theorem 3.4 remains valid without any changes. Similarly, if the  $X_i$  are random variables but the  $T_i$  are fixed design points satisfying (2.1) the corresponding stochastic processes exhibit exactly the same asymptotic behaviour as described in Theorem 3.1.

## 4 Finite sample properties

In this section we investigate the finite sample properties of two Cramér-von-Mises tests derived from the two stochastic processes  $\{S_n^*\}_{t \in [0,1]}$  and  $\{S_n^{**}\}_{t \in [0,1]}$  and perform a comparison with the test based on the  $L^2$ -distance, which has recently been proposed by You and Chen (2005). We also analyse a data example to illustrate the application of the new procedure.

## 4.1 Testing the hypothesis of homoscedasticity

We first concentrate on the problem of testing for homoscedasticity, where it follows from Corollary 3.2 and the continuous mapping theorem that under the null hypothesis  $H_0 : \sigma^2(t) = \theta_1$  (for some  $\theta_1 \in \mathbb{R}^+$ ) the Cramér-von-Mises statistic converges weakly, i.e.

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

where  $\hat{C}_n$  is either the statistic  $\hat{C}_n^*$  or  $\hat{C}_n^{**}$  corresponding to the cases  $\hat{S}_t = \hat{S}_t^*$  or  $\hat{S}_t = \hat{S}_t^{**}$ , respectively. Consequently the hypothesis of homoscedasticity is rejected if

$$\hat{C}_n \geq \omega_{1-\alpha}(\hat{m}_4 - 1 + 4\delta_r)\hat{\theta}^2.$$

Here  $\omega_{1-\alpha}$  is the  $(1 - \alpha)$  quantile of the distribution of the random variable  $\int_0^1 B^2(t) dt$  and  $\hat{\theta}$  is either  $\hat{\theta}^*$  or  $\hat{\theta}^{**}$  corresponding to the least squares estimate

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{i=r+1}^n \left( \hat{R}_i^2 - \sigma^2(t_i, \theta) \right)^2$$

obtained from the pseudo residuals  $\hat{R}_i = R_i^*$  or  $\hat{R}_i = R_i^{**}$ , respectively. The estimate  $\hat{m}_4$  for the fourth moment depends on the used difference sequence and in order to reduce the bias we used  $r = 2$  and the sequence

$$(4.1) \quad d_0 = d_2 = 1/\sqrt{6}, \quad d_1 = -2/\sqrt{6}$$

[see Gasser et al. (1986) or Dette, Munk and Wagner (1998)]. For this sequence a consistent estimate of  $m_4$  is given by

$$(4.2) \quad \hat{m}_4 = \left( \frac{1}{18(n-2)} \sum_{j=3}^n \hat{R}_j^4 - 3 \frac{1}{36(n-5)} \sum_{k=3}^{n-3} \hat{R}_k^2 \hat{R}_{k+3}^2 \right) \left( \frac{1}{6(n-2)} \sum_{j=3}^n \hat{R}_j^2 \right)^{-2},$$

which can be proved by similar arguments as in Dette and Munk (1998a).

The design considered in our study is a uniform design on the interval  $[0, 1]$  given by  $t_i = (i-0.5)/n$ ,  $i = 1, \dots, n$  and two models are investigated. The first model is given by

$$(4.3) \quad Y_i = 3.5x_i + \cos(2\pi t_i) + \sigma(t_i)\varepsilon_i \quad \text{with} \quad x_i = 5t_i^2 + 0.5\eta_i,$$

while the second model is defined by

$$(4.4) \quad Y_i = x_i + t_i/(t_i^2 + 1) + \sigma(t_i)\varepsilon_i \quad \text{with} \quad x_i = t_i^3(1-t_i)^3 + \sqrt{0.1}\eta_i,$$

where in all models the random variable  $\eta_i$  are i.i.d.  $\sim U(-\sqrt{3}, \sqrt{3})$  and the errors  $\varepsilon_i$  are also i.i.d.  $\sim U(-\sqrt{3}, \sqrt{3})$ . For the variance function three cases are considered, namely

$$(4.5) \quad \begin{aligned} \text{(I)} \quad & \sigma(t) = \sigma \exp(ct) \\ \text{(II)} \quad & \sigma(t) = \sigma [1 + c \sin(10t)]^2 \\ \text{(III)} \quad & \sigma(t) = \sigma (1 + ct)^2 \end{aligned}$$

where the choice  $c = 0$  always corresponds to the null hypothesis of homoscedasticity and  $\sigma = 0.5$  [see Dette and Munk (1998a)]. For the calculation of the statistics  $\hat{C}_n^*$  we use Speckman's (1988) estimate with local linear weights, which requires the specification of a bandwidth and a kernel  $K$ . For the latter we used the Epanechnikov kernel, while the bandwidth was chosen according to the rule of Fan and Gijbels (1995). In Table 4.1 and 4.2 we display the corresponding rejection probabilities for sample size  $n = 50, 100$  and  $200$  based on 1000 simulation runs. The corresponding results for the Cramér-von-Mises test obtained from the process  $\{\hat{S}_i^{**}\}_{t \in [0,1]}$  are presented in Table 4.3 and 4.4 corresponding to example (4.3) and (4.4), respectively. A comparison of the tests based on  $\hat{C}_n^*$  and  $\hat{C}_n^{**}$  shows that there are only minor differences between the two tests. For sample size  $n = 50$  the test based on the statistic  $\hat{C}_n^{**}$  yields to a better approximation of the nominal level, while usually the test based on the statistic  $\hat{C}_n^*$  yields slightly larger rejection probabilities under the alternative.

The results are also directly comparable with simulated rejection probabilities in You and Chen (2005) [see Table 1 and 2 in this reference]. We observe that for the variance functions (I) and (III) in (4.5) the new tests yield substantially larger rejection probabilities. However, for the variance function (II) the test proposed by You and Chen (2005) is more powerful. Note that on a first glance this contradicts asymptotic theory because the test of You and Chen can only detect local alternatives converging to the null hypothesis at a rate  $n^{-1/4}$ , while the rate for the new procedures is  $n^{-1/2}$ . The reason for the difference between the asymptotic theory and the empirical results for small sample sizes in model (II) can be explained by the specific form of the function  $S_t = \int_0^t (\sigma^2(x) - \theta_0) dx = \int_0^t \sigma^2(x) dx - t \int_0^1 \sigma^2(x) dx$  which has a maximal absolute value of 0.0924102 in the case  $c = 0.5$  and 0.283507 in the case  $c = 1$ . Therefore it is difficult to distinguish these functions from the line  $\bar{S}_t \equiv 0$  and the asymptotic advantages of the new tests will only become visible for very large sample sizes. For example, in model (4.4) with variance structure (II) and  $n = 1500$  observations the rejection probabilities of the test based on the statistic  $\hat{C}_n^*$  for  $c = 0.1$  are 0.525, 0.750 and 0.918, while they are 0.441, 0.532 and 0.647 for the test proposed by You and Chen (2005), which reflect the asymptotic superiority of the new procedure with respect to Pitman alternatives.

## 4.2 Bootstrap and testing for a parametric hypothesis

The purpose of this paragraph is twofold. First we explain how the bootstrap can be used in order to improve the finite sample properties of the test procedure. Secondly we investigate the performance of the new procedure for testing for the parametric form of the variance function. For the sake of brevity we restrict ourselves to the process  $\{\hat{S}_t^*\}_{t \in [0,1]}$ . For the application of the bootstrap we calculated the residuals

$$(4.6) \quad \hat{\varepsilon}_i = \frac{(Y_i - x_i^T \hat{\beta}_n - \hat{m}(t_i))}{\hat{\sigma}(t_i)}, \quad i = 1, \dots, n.$$

Here  $\hat{\beta}_n$  is the estimate of Speckman (1988) (with local linear weights) and  $\hat{m}(t)$  and  $\hat{\sigma}^2(t)$  are nonparametric estimates of the variance function defined by

$$\hat{m}(t_i) = \sum_{j=1}^n W_j(t_i, h)(Y_j - x_j \hat{\beta}_n),$$

	$c$	$n = 50$			$n = 100$			$n = 200$		
		2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
I	0	0.053	0.080	0.128	0.029	0.044	0.090	0.029	0.053	0.102
	0.5	0.172	0.261	0.372	0.312	0.418	0.553	0.562	0.666	0.760
	1	0.430	0.572	0.720	0.795	0.864	0.928	0.983	0.995	0.997
II	0	0.042	0.066	0.120	0.027	0.050	0.106	0.018	0.036	0.086
	0.5	0.140	0.198	0.277	0.173	0.268	0.441	0.466	0.658	0.860
	1	0.149	0.198	0.270	0.253	0.354	0.523	0.585	0.766	0.923
III	0	0.046	0.070	0.127	0.035	0.063	0.103	0.033	0.062	0.115
	0.5	0.296	0.409	0.568	0.591	0.711	0.831	0.900	0.935	0.970
	1	0.574	0.719	0.839	0.929	0.968	0.989	1.000	1.000	1.000

Table 4.1: *Rejection probabilities of the test (4.1) with  $\hat{C}_n = \hat{C}_n^*$  in model (4.3) with a difference sequence of order  $r = 2$ . The null hypothesis of homoscedasticity corresponds to the case  $c = 0$ .*

$$\hat{\sigma}^2(t_i) = \sum_{j=1}^n W_j(t_i, h) (Y_j - x_j \hat{\beta}_n - \hat{m}(t_j))^2,$$

where the weights  $W_j(t_i, h)$  are given in (2.5). The bandwidth  $h$  has again been chosen according to the rule of Fan and Gijbels (1995). If  $\hat{F}_{\hat{\varepsilon}}$  denotes the empirical distribution function of the residuals  $\hat{\varepsilon}_i$  we generated i.i.d. data  $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n \sim \hat{F}_{\hat{\varepsilon}}$  and the bootstrap sample

$$\tilde{Y}_i = \hat{m}(t_i) + \sigma(t_i, \hat{\theta}^*) \tilde{\varepsilon}_i, \quad i = 1, \dots, n.$$

where  $\hat{\theta}^*$  is defined in (2.12). Finally, the corresponding Cramér-von-Mises statistic, say  $\hat{C}_n^*$ , is calculated from the bootstrap data. If  $B$  bootstrap replications have been performed and  $\tilde{C}_n^{(1)} < \dots < \tilde{C}_n^{(B)}$  denote the order statistics of the calculated bootstrap sample, the null hypothesis is rejected if

$$(4.7) \quad \hat{C}_n^* > \tilde{C}_n^{(\lfloor B(1-\alpha) \rfloor)}.$$

$B = 100$  bootstrap replications were performed to calculate the rejection probabilities and 1000 simulation runs were used for each scenario. In Table 4.5 we display the rejection probabilities of this bootstrap procedure for the problem of testing for homoscedasticity in model (4.3). The results are comparable with Table 4.1. It is remarkable that by the bootstrap procedure the approximation of the nominal level is improved substantially, even for sample size  $n = 50$ . Moreover, for the variance function (II) the bootstrap procedure yields distinctly larger rejection probabilities under the alternative. A comparison with the results of You and Chen (2005) shows that the bootstrap version of the new tests performs better than the test based on the  $L^2$ -distance in nearly all cases. Only for the variance functions (II) the test of You and Chen (2005) yields a substantially larger power, provided that the sample size is small.

Finally, we consider the problem of testing a nonlinear parametric structure for the variance function, that is

$$(4.8) \quad H_0 : \sigma^2(t) = \exp(\theta t),$$

$r=2$	$c$	$n = 50$			$n = 100$			$n = 200$		
		2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
I	0	0.042	0.068	0.118	0.036	0.066	0.122	0.027	0.053	0.105
	0.5	0.165	0.238	0.356	0.273	0.365	0.503	0.542	0.640	0.745
	1	0.413	0.529	0.696	0.785	0.875	0.942	0.990	0.996	0.999
II	0	0.043	0.069	0.121	0.035	0.055	0.111	0.022	0.051	0.111
	0.5	0.137	0.193	0.273	0.217	0.307	0.465	0.449	0.649	0.876
	1	0.114	0.172	0.272	0.256	0.382	0.547	0.563	0.754	0.925
III	0	0.054	0.078	0.130	0.031	0.055	0.102	0.029	0.043	0.086
	0.5	0.310	0.418	0.556	0.602	0.721	0.827	0.931	0.957	0.985
	1	0.577	0.721	0.856	0.931	0.972	0.989	1.000	1.000	1.000

Table 4.2: Rejection probabilities of the test (4.1) with  $\hat{C}_n = \hat{C}_n^*$  in model (4.4) with a difference sequence of order  $r = 2$ . The null hypothesis of homoscedasticity corresponds to the case  $c = 0$ .

	$c$	$n = 50$			$n = 100$			$n = 200$		
		2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
I	0	0.036	0.066	0.123	0.025	0.042	0.093	0.026	0.051	0.092
	0.5	0.169	0.245	0.349	0.285	0.284	0.508	0.558	0.675	0.773
	1	0.421	0.531	0.686	0.783	0.862	0.932	0.985	0.992	0.999
II	0	0.036	0.063	0.108	0.024	0.049	0.094	0.036	0.061	0.117
	0.5	0.108	0.151	0.249	0.191	0.285	0.448	0.429	0.628	0.850
	1	0.124	0.183	0.282	0.245	0.345	0.541	0.540	0.726	0.907
III	0	0.036	0.060	0.111	0.024	0.057	0.116	0.031	0.055	0.101
	0.5	0.292	0.409	0.548	0.599	0.718	0.829	0.905	0.958	0.977
	1	0.558	0.724	0.845	0.936	0.972	0.986	0.999	1.000	1.000

Table 4.3: Rejection probabilities of the test (4.1) with  $\hat{C}_n = \hat{C}_n^{**}$  in model (4.3) with a difference sequence of order  $r = 2$ . The null hypothesis of homoscedasticity corresponds to the case  $c = 0$ .

	$c$	$n = 50$			$n = 100$			$n = 200$		
		2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
I	0	0.033	0.056	0.099	0.034	0.057	0.106	0.028	0.057	0.101
	0.5	0.189	0.274	0.393	0.270	0.367	0.486	0.537	0.661	0.767
	1	0.427	0.534	0.699	0.735	0.836	0.909	0.981	0.992	1.000
II	0	0.042	0.062	0.111	0.027	0.057	0.113	0.025	0.053	0.105
	0.5	0.116	0.178	0.268	0.185	0.295	0.440	0.427	0.629	0.867
	1	0.134	0.185	0.300	0.247	0.348	0.508	0.568	0.756	0.914
III	0	0.048	0.079	0.138	0.035	0.059	0.103	0.028	0.057	0.110
	0.5	0.288	0.400	0.542	0.569	0.694	0.808	0.903	0.952	0.979
	1	0.566	0.698	0.820	0.913	0.956	0.984	1.000	1.000	1.000

Table 4.4: Rejection probabilities of the test (4.1) with  $\hat{C}_n = \hat{C}_n^{**}$  in model (4.4) with a difference sequence of order  $r = 2$ . The null hypothesis of homoscedasticity corresponds to the case  $c = 0$ .

	$c$	$n = 50$			$n = 100$			$n = 200$		
		2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
I	0	0.024	0.054	0.106	0.020	0.044	0.103	0.017	0.045	0.094
	0.5	0.158	0.269	0.376	0.287	0.414	0.545	0.574	0.699	0.812
	1	0.512	0.667	0.780	0.812	0.901	0.957	0.994	0.997	1.000
II	0	0.021	0.057	0.121	0.028	0.053	0.098	0.019	0.055	0.114
	0.5	0.214	0.324	0.491	0.456	0.655	0.863	0.926	0.994	0.999
	1	0.364	0.529	0.731	0.820	0.957	0.993	0.998	1.000	1.000
III	0	0.017	0.048	0.111	0.021	0.052	0.105	0.030	0.069	0.123
	0.5	0.353	0.535	0.653	0.633	0.757	0.852	0.920	0.959	0.980
	1	0.709	0.819	0.904	0.963	0.989	0.996	1.000	1.000	1.000

Table 4.5: Rejection probabilities of the bootstrap test (4.7) with  $\hat{C}_n = \hat{C}_n^*$  in model (4.3) with a difference sequence of order  $r = 2$ . The null hypothesis of homoscedasticity corresponds to the case  $c = 0$ .

where the regression model is given by (4.3) and

$$(4.9) \quad \sigma^2(t) = (1 + c \sin(2\pi t)) \exp(t),$$

with the case  $c = 0$  corresponding to the null hypothesis. The errors  $\varepsilon_i$  are standard normal distributed and the design is uniform. The corresponding rejection probabilities of the bootstrap test are depicted in Table 4.6 for a difference sequence of order  $r = 1$  and the second order sequence defined in (4.1). In most cases we observe a reasonable approximation of the nominal level for sample sizes larger than 100 and alternatives are detected with rather large power.

	$c$	$n = 50$			$n = 100$			$n = 200$		
		2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
$r = 1$	0	0.054	0.082	0.151	0.038	0.077	0.144	0.020	0.054	0.106
	0.5	0.255	0.323	0.407	0.333	0.433	0.523	0.481	0.571	0.654
	1	0.614	0.675	0.726	0.838	0.884	0.921	0.968	0.977	0.990
$r = 2$	0	0.045	0.077	0.139	0.035	0.061	0.136	0.038	0.066	0.120
	0.5	0.205	0.269	0.336	0.270	0.358	0.430	0.390	0.491	0.595
	1	0.520	0.596	0.658	0.718	0.768	0.822	0.924	0.952	0.972

Table 4.6: *Rejection probabilities of the bootstrap test (4.7) in model (4.3) with a nonlinear variance function of the form (4.9) and two difference sequences. The null hypothesis (4.8) corresponds to the case  $c = 0$ .*

### 4.3 Data example

Daniel and Wood (1999) and Bianco et al. (2006) studied a data set of 82 observations obtained in a process variable study of a refinery unit. The response variable  $Y$ , which is depicted in the upper left panel of Figure 4.1, is the octane number of the final product. The first three covariates represent the feed compositions and the fourth the logarithm of a combination of process conditions. For this data we have performed the test for the hypothesis of homoscedasticity, i.e.  $H_0 : \sigma^2(t) = \theta$ , where we have used Speckman's estimate (1988) with local linear weights to estimate the parameter  $\beta$ , which is defined in (2.4). We have used the Epanechnikov kernel with bandwidth  $h = 0.1$ , which was chosen according to the rule of Fan and Gijbels (1995). Daniel and Wood (1999) and Bianco et al. (2006) both discussed the presence of three observations (75 – 77), which extend the range of the variables  $Y$  and  $x_1$ . The data is depicted in the upper left panel of Figure 4.1. While the proposed test for homoscedasticity is robust against these three outliers we found that observation 39 has a strong influence on the result of the test. This observation corresponds to an outlier in the data  $\{Y_i - x_i^T \hat{\beta}_n\}_{i=1}^{82}$  and yields to a very large pseudo residual. Therefore we did not use this point in our data analysis. The corresponding plots of the residuals with and without the 39th observation are given in the upper right and lower left panel of Figure 4.1, while the process  $\{\hat{S}_t^*\}_t$  is shown in the lower right panel. The resulting value of the Cramér-von-Mises statistic  $\hat{C}_n^*$  is 0.0877812, which yields to a  $p$ -value of 0.065. Therefore the test rejects the hypothesis of homoscedasticity at level 0.1.

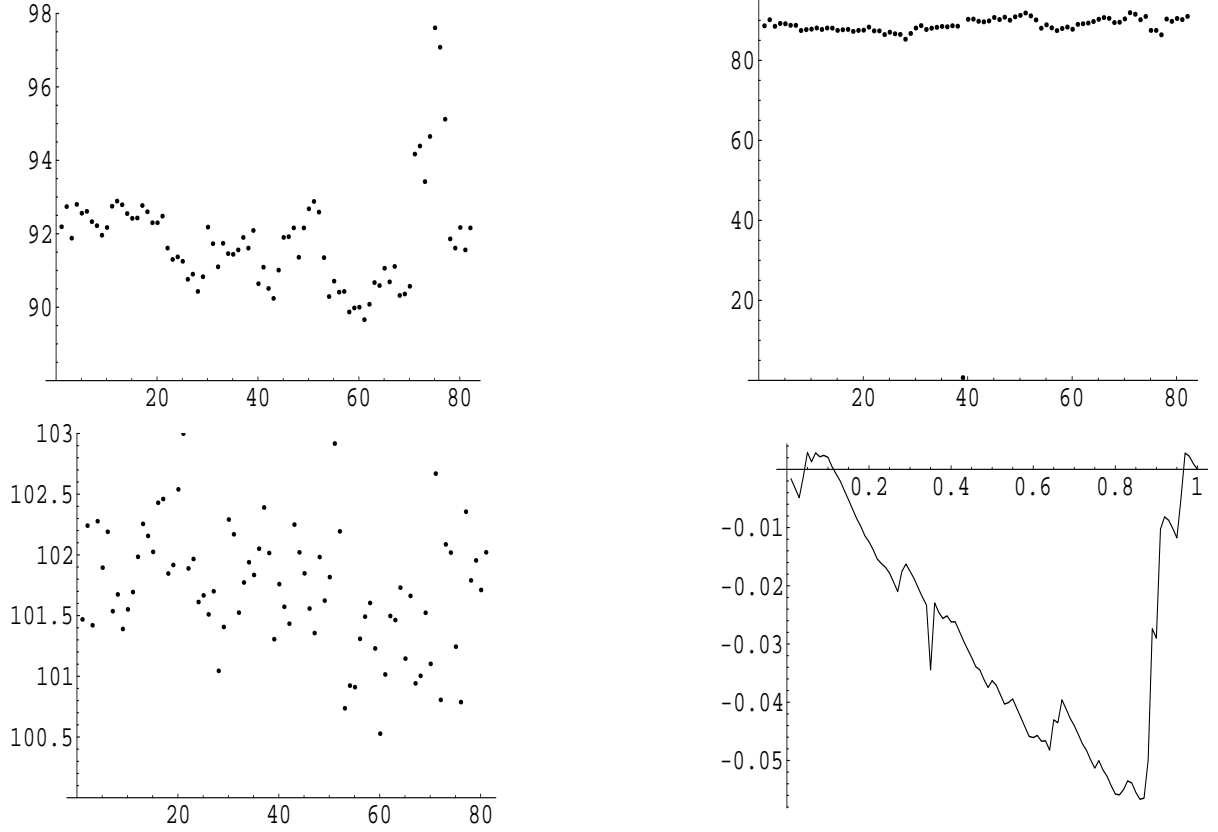


Figure 4.1: The refinery data discussed in Daniel and Wood (1999) and Bianco et al. (2006) (left upper panel), the residuals  $\{Y_i - x_i^T \hat{\beta}_n\}_{i=1}^{82}$  (right upper panel). The left lower panel shows the corresponding residuals after deleting the 39th observation and the right lower panel the empirical process  $\{\hat{S}_t^*\}_t$  based on this data.

## 5 Appendix

### 5.1 Proof of Theorem 3.1.

The proof of the theorem has to be given separately for the two processes  $\{S_t^*\}_{t \in [0,1]}$  and  $\{\hat{S}_t^{**}\}_{t \in [0,1]}$  considered in the theorem.

#### (a) Weak convergence of $\{\sqrt{n}(\hat{S}_t^* - S_t)\}_{t \in [0,1]}$

Recall the definition of the first term  $\hat{B}_t^{0*} := \frac{1}{n-r} \sum_{i=r+1}^n 1_{\{t_i \leq t\}} R_i^{*2}$  in (2.14). At the end of this proof we will show the approximation

$$(5.1) \quad D_n := \hat{B}_t^{0*} - \tilde{B}_t^0 = \frac{1}{n-r} \sum_{i=r+1}^n 1_{\{t_i \leq t\}} (R_i^{*2} - L_i^2) = o_p(n^{-1/2}),$$



where

$$(5.2) \quad \tilde{B}_t^0 = \frac{1}{n-r} \sum_{i=r+1}^n 1_{\{t_i \leq t\}} L_i^2,$$

$$(5.3) \quad L_j = \sum_{i=0}^r d_i \sigma(t_{j-i}) \varepsilon_{j-i}, \quad j = r+1, \dots, n.$$

This result and a Taylor expansion now yield

$$\hat{S}_t^* = \frac{1}{n-r} \sum_{i=r+1}^n 1_{\{t_i \leq t\}} \left( H_i - \frac{\partial}{\partial \theta} \sigma^2(t_i, \theta) \Big|_{\theta=\theta_0} (\hat{\theta}^* - \theta_0) \right) + o_p\left(\frac{1}{\sqrt{n}}\right),$$

where the random variables  $H_j$  are defined by

$$(5.4) \quad H_j = \left( \sum_{i=0}^r d_i \sigma(t_{j-n}) \varepsilon_{j-n} \right)^2 - \sigma^2(t_j, \theta_0), \quad j = r+1, \dots, n.$$

The Hölder continuity of the function  $\sigma$  therefore implies

$$(5.5) \quad \sqrt{n}(\hat{S}_t^* - S_t) = \frac{\sqrt{n}}{n-r} \sum_{i=r+1}^n 1_{\{t_i \leq t\}} \left( Z_i - \sum_{j=1}^d \sigma_j^2(t_i) \vartheta_j^* \right) + o_p(1),$$

where the random variables  $Z_i$  are defined by  $Z_i = H_i - E[H_i]$  and  $\vartheta_j^*$  denotes the  $j$ th component of the vector  $\hat{\theta}^* - \theta_0$ . Recall the definition (3.2) and define  $\Sigma = (\sigma_j^2(t_i))_{i=1, \dots, n-r}^{j=1, \dots, d} \in \mathbb{R}^{(n-r) \times d}$ , then it follows from standard results about nonlinear regression [see Seber and Wild (1989), p. 572-574] that

$$(5.6) \quad \hat{\theta}^* - \theta_0 = (\Sigma^T \Sigma)^{-1} \Sigma^T R^* - \theta_0 + O_p\left(\frac{1}{n}\right) = (\Sigma^T \Sigma)^{-1} \Sigma^T H + O_p\left(\frac{1}{n^\gamma}\right),$$

where  $R^* = ((R_{r+1}^*)^2, \dots, (R_n^*)^2)^T$ ,  $H = (H_{r+1}, \dots, H_n)^T$  and we use the fact that  $\theta_0$  corresponds to the best approximation of the function  $\sigma^2(\cdot)$  by functions of the form  $\sigma^2(\cdot, \theta)$ ;  $\theta \in \Theta$ . Observing (3.1) and the definition of the matrix  $\Sigma$  it therefore follows that

$$\begin{aligned} \frac{1}{n} \Sigma^T \Sigma - \hat{A} &= O(n^{-1}), \\ \frac{1}{n} \Sigma^T H - \frac{1}{n} \left( \sum_{i=r+1}^n \sigma_j^2(t_i) (H_i - E[H_i]) \right)_{j=1}^d &= O_p(n^{-1}), \end{aligned}$$

where

$$\hat{A} = (\hat{a}_{ij})_{1 \leq i, j \leq d}, \quad \hat{a}_{ij} = \frac{1}{n} \sum_{k=1}^n \sigma_i^2(t_k) \sigma_j^2(t_k).$$

With the notation  $Z = (Z_{r+1}, \dots, Z_n)^T$  and

$$\begin{aligned} \hat{B}_t &= (\hat{B}_t^1, \dots, \hat{B}_t^d), & \hat{B}_t^k &= \frac{1}{n} \sum_{j=1}^n 1_{\{t_j \leq t\}} \sigma_k^2(t_j), \\ \bar{C} &= (\bar{c}_1, \dots, \bar{c}_d)^T, & \bar{c}_i &= \frac{1}{n-r} \sum_{j=r+1}^n Z_j \sigma_i^2(t_j) \quad \text{and} \\ \bar{B}_t^0 &= \frac{1}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} Z_j, \end{aligned}$$

we obtain the stochastic expansion

$$(5.7) \quad \sqrt{n}(\hat{S}_t^* - S_t) = \sqrt{n}(\bar{B}_t^0 - \hat{B}_t \hat{A}^{-1} \bar{C}) + o_p(1).$$

The weak convergence of the stochastic process defined on the right hand side of (5.7) has been established by Dette and Hetzler (2006) and the assertion of the theorem follows if the estimate (5.1) can be shown.

For this we introduce the notation

$$(5.8) \quad \Delta m_j = \sum_{i=0}^r d_i m(t_{j-i}), \quad \Delta x_j = \sum_{i=0}^r d_i x_{j-i}^T$$

and obtain the decomposition,

$$R_j^* = L_j + \Delta m_j + \Delta x_j(\beta - \hat{\beta}_n),$$

which yields

$$(5.9) \quad D_n = D_{n1} + \dots + D_{n5}$$

where

$$(5.10) \quad D_{n1} = \frac{2}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} L_j \Delta m_j,$$

$$(5.11) \quad D_{n2} = \frac{2}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} L_j \Delta x_j(\beta - \hat{\beta}_n),$$

$$(5.12) \quad D_{n3} = \frac{1}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} (\Delta m_j)^2,$$

$$(5.13) \quad D_{n4} = \frac{2}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} \Delta m_j \Delta x_j(\beta - \hat{\beta}_n),$$

$$(5.14) \quad D_{n5} = \frac{1}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} (\Delta x_j(\beta - \hat{\beta}_n))^2.$$

The Hölder continuity of the regression function implies

$$(5.15) \quad D_{n3} = \frac{1}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} (\Delta m_j)^2 \leq \frac{(\max_j \Delta m_j)^2}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} = o(n^{-1/2}).$$

Because  $\hat{\beta}_n$  is a  $\sqrt{n}$ -consistent estimate of the parameter  $\beta$  we have from the Cauchy-Schwarz inequality

$$\begin{aligned}
(5.16) \quad D_{n5} &= \frac{1}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} \left( \sum_{i=0}^r d_i x_{j-i}^T (\beta - \hat{\beta}_n) \right)^2 \\
&\leq \frac{1}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} \left( \sum_{i=0}^r d_i^2 \cdot \sum_{i=0}^r (x_{j-i}^T (\beta - \hat{\beta}_n))^2 \right) \\
&\leq \frac{1}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} \left( \sum_{i=0}^r \|x_{j-i}\|^2 \|\beta - \hat{\beta}_n\|^2 \right) = o_p(n^{-1/2}),
\end{aligned}$$

and by a similar argument

$$(5.17) \quad D_{n4} = \frac{2}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} \left( \sum_{i=0}^r d_i x_{j-i}^T (\beta - \hat{\beta}_n) \right) \left( \sum_{i=0}^r d_i m(t_{j-i}) \right) = o_p(n^{-1/2}).$$

For a treatment of the term  $D_{n1}$  we note that a standard calculation (using again the Hölder continuity of the regression function) shows that  $\text{Var}(D_{n1}) = o(n^{-1})$ , and Markov's inequality gives

$$(5.18) \quad D_{n1} = o_p(n^{-1/2}).$$

The term  $D_{n2}$  is treated exactly in the same way and it follows from (5.15) - (5.17) that  $D_n = o_p(n^{-1/2})$ , which establishes (5.1) and completes the proof of Theorem 3.1 in the case  $\hat{S}_t = \hat{S}_t^*$ .

**(b) Weak convergence of  $\{\sqrt{n}(\hat{S}_t^{**} - S_t)\}_{t \in [0,1]}$**

We begin the proof showing that the statistic  $\check{\beta}_n$  is in fact a  $\sqrt{n}$ -consistent estimate for the parameter  $\beta$ . For this we note that the Hölder continuity of the regression function yields

$$E[\check{\beta}_n] = \beta + O(n^{-1}).$$

Similarly, it follows from the fact that the random variables  $\check{\varepsilon}_i$  are 1-dependent and assumption (2.19) that

$$\text{Var}(\check{\beta}_n) = O(n^{-1}),$$

which implies  $\check{\beta}_n - \beta = O_p(n^{-1/2})$  by Markov's inequality. For the remaining part of the proof we establish the estimate

$$(5.19) \quad \hat{S}_t^{**} = \frac{1}{n-r} \sum_{i=r+1}^n 1_{\{t_i \leq t\}} \left( H_i - \frac{\partial}{\partial \theta} \sigma^2(t_i, \theta) \Big|_{\theta=\theta_0} (\hat{\theta}^{**} - \theta_0) \right) + o_p\left(\frac{1}{\sqrt{n}}\right).$$

The assertion then follows along the lines of the proof given in part (a). For a proof of (5.19) we note that it is again sufficient to show

$$(5.20) \quad \check{D}_n := \frac{1}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} (R_j^{**2} - L_j^2) = o_p(n^{-1/2}).$$

With the notation (5.8) we obtain the decomposition

$$(5.21) \quad R_j^{**} = L_j + \Delta m_j + \Delta x_j(\beta - \check{\beta}_n).$$

This yields the representation

$$(5.22) \quad \check{D}_n := D_{n1} + \check{D}_{n2} + D_{n3} + \check{D}_{n4} + \check{D}_{n5},$$

where  $\check{D}_{n2}$ ,  $\check{D}_{n4}$  and  $\check{D}_{n5}$  are defined as  $D_{n2}$ ,  $D_{n4}$  and  $D_{n5}$  in (5.11), (5.13) and (5.14) with  $\hat{\beta}_n$  replacing  $\check{\beta}_n$ , and  $D_{n1}$  and  $D_{n3}$  are given in (5.10) and (5.12).

Using the Cauchy-Schwarz inequality, the boundedness of the vectors  $x_j$  and the  $\sqrt{n}$ -consistency of the estimate of the parameter yields

$$(5.23) \quad \begin{aligned} \check{D}_{n5} &= \frac{1}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} \left( \sum_{i=0}^r d_i x_{j-i}^T (\beta - \check{\beta}_n) \right)^2 \\ &\leq \frac{1}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} \left( \sum_{i=0}^r d_i^2 \cdot \sum_{i=0}^r (x_{j-i}^T (\beta - \check{\beta}_n))^2 \right) \\ &\leq \frac{1}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} \left( \sum_{i=0}^r \|x_{j-i}\|^2 \|\beta - \check{\beta}_n\|^2 \right) = o_p(n^{-1/2}). \end{aligned}$$

By a similar argument it follows from the Hölder continuity of the regression function that

$$(5.24) \quad \check{D}_{n4} = \frac{2}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} \Delta m_j \Delta x_j (\beta - \check{\beta}_n) = o_p(n^{-1/2}).$$

Markov's inequality gives

$$(5.25) \quad \check{D}_{n2} = \frac{1}{n-r} \sum_{j=r+1}^n 1_{\{t_j \leq t\}} 2L_j \Delta x_j (\beta - \check{\beta}_n) = o_p(n^{-1/2}).$$

and it follows from (5.15), (5.18) and (5.23)-(5.25) that  $\check{D}_n = o_p(n^{-1/2})$ . This establishes (5.20) and completes the proof of Theorem 3.1 in the case  $\hat{S}_t = \hat{S}_t^{**}$ .

**Acknowledgements.** The authors are grateful to Isolde Gottschlich and Martina Stein who typed numerous versions of this paper with considerable technical expertise. The work of the authors was supported in part by the Sonderforschungsbereich 475, Komplexitätsreduktion in multivariaten Datenstrukturen (Teilprojekt B1) and by a NIH grant award IR01GM072876:01A1.

## References

Aneiros-Pérez, G., González-Manteiga, W., Vieu, P. (2004). Estimation and testing in a partial linear regression model under long-memory dependence. *Bernoulli*, 10, 49-78.

- Bianco, A., Boente, G., Martínez, E. (2006). Robust tests in semiparametric partly linear models. *Scand. J. Statist.*, 33, 435-450.
- Cook, R.D., Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*, 70, 1-10.
- Daniel, C., Wood, F.S. (1999). *Fitting equations to data: computer analysis of multifactor data*. Wiley, New York.
- Dette, H., Hetzler, B. (2006). A simple test for the parametric form of the variance function in nonparametric regression. Preprint, Ruhr-Universität-Bochum.
- Dette, H., Munk, A. (1998a). Testing heteroscedasticity in nonparametric regression. *J. R. Statist. Soc., Ser. B*, 60, 693-708.
- Dette, H., Munk, A. (1998b). A simple goodness-of-fit test for linear models under a random design assumption. *Ann. Inst. Statist. Math.*, 50, 253-275.
- Dette, H., Munk, A., Wagner, T. (1998). Estimating the variances in nonparametric regression - what is a reasonable choice? *J. R. Statist. Soc., Ser. B*, 60, 751-764.
- Dibiasi, A., Bowman, A. (1997). Testing for constant variance in a linear model. *Statist. Probab. Lett.*, 33, 95-103.
- Fan, J., Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaption. *J. R. Statist. Soc. B*, 57, pp. 371-394.
- Fan, J., Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman & Hall, London.
- Fan, J., Huang, L.S. (2001). Goodness-of-fit tests for parametric regression models. *J. Am. Statist. Ass.*, 96, 640-652.
- Gao, J. (1997). Adaptive parametric test in a semiparametric regression model. *Comm. Statist., Theory and Methods*, 26, 787-800.
- Gasser, T., Sroka, L., Jennen-Steinmetz, G. (1986). Residual variance and pattern in nonlinear regression. *Biometrika*, 73, 625-633.
- González-Manteiga, W., Aneiros-Pérez, G. (2003). Testing in partial linear regression models with dependent errors. *J. Nonpar. Statist.*, 15, 93-111.
- Green, P. J., Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: A roughness penalty approach*. Chapman & Hall, London.
- Härdle, W., Ling, H., Gao, J. (2000). *Partially linear models*. Physica-Verlag, Heidelberg.
- Hall, P., Kay, J.W., Titterton, D.M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, 77, 521-528.
- Koenker, R., Bassett, G. (1981). Robust tests for heteroscedasticity based on regression quantiles. *Econometrika*, 50, 43-61.
- Liero, H. (2003). Testing homoscedasticity in nonparametric regression. *J. Nonpar. Statist.*, 15, 31-51.
- Sacks, J., Ylvisacker, D. (1970). Designs for regression problems with correlated errors III. *Ann. Math. Statist.*, 41, 2057-2074.
- Seber, G.A.F., Wild, G.J. (1989). *Nonlinear Regression*. Wiley, N.Y.

- Speckman, P. (1988). Kernel smoothing in partial linear models. *J. R. Statist. Soc., Ser. B*, 50, 413-436.
- Yatchew, A. (1997). An elementary estimator for the partial linear model. *Econom. Lett.* 57, 135-143.
- You, J., Chen, G. (2005). Testing heteroscedasticity in partially linear regression models. *Statist. Probab. Lett.*, 73, 61-70.