doi: 10.1111/j.1467-9469.2006.00535.x

Non-parametric Analysis of Covariance – The Case of Inhomogeneous and Heteroscedastic Noise

AXEL MUNK

Institute for Mathematical Stochastics, Georg-August Universität Göttingen

NATALIE NEUMEYER

Department Mathematik, Schwerpunkt Stochastik, Universität Hamburg

ACHIM SCHOLZ

Institute for Mathematical Stochastics, Georg-August Universität Göttingen

ABSTRACT. The purpose of this paper was to propose a procedure for testing the equality of several regression curves f_i in non-parametric regression models when the noise is inhomogeneous and heteroscedastic, i.e. when the variances depend on the regressor and may vary between groups. The presented approach is very natural because it transfers the maximum likelihood statistic from a heteroscedastic one-way analysis of variance to the context of non-parametric regression. The maximum likelihood estimators will be replaced by kernel estimators of the regression functions f_i. It is shown that the asymptotic distribution of the obtained test-statistic is nuisance parameter free. Asymptotic efficiency is compared with a test of Dette & Neumeyer [Annals of Statistics (2001) Vol. 29, 1361–1400] and it is shown that the new test is asymptotically uniformly more powerful. For practical purposes, a bootstrap variant is suggested. In a simulation study, level and power of this test will be briefly investigated and compared with other procedures. In summary, our theoretical findings are supported by this study. Finally, a crop yield experiment is reanalysed.

Key words: ANOVA, efficacy, goodness-of-fit, heteroscedasticity, non-parametric regression, wild bootstrap, Wilks phenomenon

1. Introduction

A classical theme of statistical analysis is the comparison of two (or more) groups, which are measured under different experimental conditions. As an example consider, for instance, the comparison of wage functions in different groups defined by gender or location (see Lavergne, 2001, for more examples). To simplify notation we will restrict, for the moment, to the case of two groups, the extension to three and more groups will be presented later on. In the context of regression, one observes independent real-valued data Y_{ij}, which follow the model

$$Y_{ij} = f_i(t_{ij}) + \sigma_i(t_{ij})\varepsilon_{ij}, \quad j = 1, \ldots, n_i, \;\; i = 1, 2, \tag{1}$$

where t_{ij} are fixed locations of measurements, f_i denote the unknown regression functions, $f_i(t_{ij}) = E[Y_{ij}]$, and σ_i^2 the unknown variance functions, $\sigma_i^2(t_{ij}) = \text{var}(Y_{ij})$ of the ith group $(i = 1, 2)$. The errors ε_{ij} are assumed to be independent random variables with mean 0 and variance 1. Our aim is to test the equality of the regression functions f_1 and f_2.

Under a parametric assumption on the error ε_{ij} and the functions f_i and σ_i^2 this leads to the analysis of covariance (see Scheffé, 1959 or Chow, 1960). Without these assumptions, in particular, when the functional form of f_i is not specified, this is denoted as the non-parametric analysis of covariance (Young & Bowman, 1995) and has received much attention (see Hall & Hart, 1990; Delgado, 1993; Kulasekera, 1995; Munk & Dette, 1998; Yatchew, 1999, among many others). As pointed out by Gørgens (2002) many tests in the literature for

$$H_0: f_1 = f_2 \text{ versus } H_1: f_1 \neq f_2 \tag{2}$$

cannot be applied in the general model (1) because often it is assumed that sample sizes are equal, the regressors follow the same distribution between populations, or that there is a homoscedastic error, i.e. the variances σ_i^2 are independent of the regressor t. For the general setting (1), there are only a very few tests available, see Cabus (1998), Dette & Neumeyer (2001), Lavergne (2001), Gørgens (2002) and Neumeyer & Dette (2003). Whereas Lavergne (2001) and Gørgens (2002) consider a stochastic regressor, Cabus (1998) and Neumeyer & Dette (2003) use test-statistics, which are based on the associated marked empirical process. Recently, Pardo-Fernández *et al.* (2006) proposed a test based on a comparison of the empirical distributions of estimated errors in the two models.

The presented method is related to Dette & Neumeyer's (2001) test and Fan *et al.*'s (2001) test in the case of a one-dimensional predictor. Dette & Neumeyer (2001) compared, theoretically as well as using Monte Carlo study, their test with various tests from the literature and came to the conclusion that their test outperforms their competitors in terms of power. In this paper, we present a test, which will be shown to be superior to Dette & Neumeyer's (2001) test with respect to power.

More specifically, our test is based on the idea to compare a weighted 'least squares' estimator under the assumption of equal regression curves with an estimator, which is based on non-parametric estimators $\hat{f}_i$ for f_i, exactly as in a parametric analysis of covariance. To motivate the procedure assume for the moment the regression functions to be constant $f_i(t) \equiv \mu_i$, the variance functions to be constant and known $\sigma_i^2(t) \equiv \sigma_i^2$ and the errors ε_{ij} to be normally distributed. In other words, consider testing the equality of the means $H_0: \mu_1 = \mu_2$ in two samples

$$Y_{ij} \overset{\text{i.i.d.}}{\sim} N(\mu_i, \sigma_i^2), \quad j = 1, \ldots, n_i, \quad i = 1, 2.$$

The maximum likelihood method leads to the estimates $\hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$ in the individual samples ($i = 1, 2$), and

$$\hat{\mu} = a\hat{\mu}_1 + (1 - a)\hat{\mu}_2, \quad \text{where } a = \frac{\sigma_1^{-2} n_1}{\sigma_1^{-2} n_1 + \sigma_2^{-2} n_2},$$

in the pooled sample (under H_0). The logarithm of the likelihood ratio has the form

$$\frac{1}{N} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu})^2 \sigma_i^{-2} - \frac{1}{N} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_i)^2 \sigma_i^{-2}, \tag{3}$$

where $N = n_1 + n_2$ denotes the total sample size. Now, we transfer this statistic to a non-parametric set-up and consider in the non-parametric regression model (1) the class of pooled estimators

$$\tilde{f}(x) = a(x)\hat{f}_1(x) + (1 - a(x))\hat{f}_2(x), \tag{4}$$

where $\hat{f}_i$ denote kernel-based estimators of the regression functions f_i ($i = 1, 2$). In this class, minimization of the asymptotic mean-squared error

$$\text{AMSE}[\tilde{f}] = a^2(x) \int K^2(u)\,\mathrm{d}u \frac{\sigma_1^2(x)}{n_1 h r_1(x)} + (1 - a(x))^2 \int K^2(u)\,\mathrm{d}u \frac{\sigma_2^2(x)}{n_2 h r_2(x)},$$

where h denotes a smoothing parameter that fulfils conditions (14) stated in section 2, and K denotes a proper kernel function, gives the weight

$$a(x) = \frac{\sigma_1^{-2}(x) n_1 r_1(x)}{\sigma_1^{-2}(x) n_1 r_1(x) + \sigma_2^{-2}(x) n_2 r_2(x)}, \tag{5}$$

where r_i denotes the design density in the ith sample. Now, we replace σ_i^2 and r_i by appropriate kernel-based estimators $\hat{\sigma}_i^2$, $\hat{r}_i$ $(i = 1, 2)$ and denote by $\hat{f}$ the resulting pooled estimator $\tilde{f}$ as in (4). Hence, as a test-statistic for hypotheses (2), we consider in analogy of (3),

$$T_N = \frac{1}{N} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (Y_{ij} - \hat{f}(t_{ij}))^2 \hat{\sigma}_i^{-2}(t_{ij}) - \frac{1}{N} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (Y_{ij} - \hat{f}_i(t_{ij}))^2 \hat{\sigma}_i^{-2}(t_{ij}). \tag{6}$$

Note that the motivation of our procedure is similar to the method of generalized likelihood ratio statistics introduced by Fan *et al.* (2001), confer Remark 1. We will show that under the null hypothesis the standardized test-statistic

$$N\sqrt{h}\left(T_N - \frac{C}{Nh}\right)$$

is asymptotically centred normal, where the asymptotic variance as well as C only depend on the kernel function K. This feature has been phrased by Fan *et al.* (2001) as the new Wilks phenomenon and might be particularly appealing for practical purposes because asymptotically the resulting test does not depend on any nuisance parameter, such as f_i, σ_i^2 or on the distribution of the ε_{ij}, in contrast to most procedures suggested in the literature (a notable exception is Gørgens, 2002).

The rest of the paper is organized as follows. In section 2, we present the required theory. The asymptotic behaviour under fixed and local alternatives is discussed and it is shown that the test of Dette & Neumeyer (2001) is outperformed in general. Only in special cases asymptotically these tests achieve the same power. We show that in particular, when the variances are inhomogeneous, i.e. unequal in both groups, or when they are heteroscedastic, i.e. dependent on the regressor, the new test gains significantly in power. We mention that from a practical point of view the case of inhomogeneous variances is very common in applications. For analysis of variance (ANOVA) models, this is well known as the celebrated Behrens–Fisher problem (see, for example, Weerahandi, 1987); in our context of non-parametric analysis of covariance, we refer to Gørgens (2002) for an econometric example. Hence, our method may be regarded as an approach, which adapts automatically to inhomogeneous and heteroscedastic variability. In section 3, we address the selection of smoothing parameters from a theoretical (section 3.1) and practical point of view (section 3.2). In particular, we investigate the power and level of the proposed test numerically, and we find the test to be superior to Dette & Neumeyer's (2001) test with respect to power. A sensitivity analysis of the bandwidth required in the estimators $\hat{f}_i$ and $\hat{f}$ in (6) is performed and we find that the actual level of our test is robust for a large scale of bandwidths but quite sensitive with respect to power. Finally, in section 4, we illustrate the performance of our method by an example where we investigate whether there is a difference in the onion yield as a function of plant density in two locations in South Australia. Section 5 contains some concluding remarks. Proofs are postponed to Appendix to keep the paper more readable. We mention that because of the additional estimation of the optimal weighting function a in (5) the proofs are technically much more involved as it is the case for the statistic considered in Dette & Neumeyer (2001).

2. Asymptotic theory

2.1. Notation and main results

In this section, we will start with extending T_N in (6) to the case of k samples, i.e. we are concerned with the model

$$Y_{ij} = f_i(t_{ij}) + \sigma_i(t_{ij})\varepsilon_{ij}, \quad j = 1, \ldots, n_i, \; i = 1, \ldots, k, \tag{7}$$

and the testing problem is

$$H_0 : f_1 = \cdots = f_k \text{ versus } H_1 : f_i \neq f_j \text{ for some } i \neq j. \tag{8}$$

Further, assume for the sample sizes that

$$\frac{n_i}{N} = \kappa_i + O\left(\frac{1}{N}\right), \quad i = 1, \ldots, k, \tag{9}$$

where $\kappa_i \in (0, 1)$ and $N = \sum_{j=1}^{k} n_j$ denotes the total sample size. The fixed design points t_{ij} can be modelled by a so-called design density r_i on $[0, 1]$ such that

$$\int_0^{t_{ij}} r_i(t)\,\mathrm{d}t = \frac{j}{n_i}, \quad j = 1, \ldots, n_i, \; i = 1, \ldots, k, \tag{10}$$

see Sacks & Ylvisaker (1970). We further assume the densities r_i and the variance functions σ_i^2 to be bounded away from zero, i.e.

$$\inf_{t \in [0,\,1]} r_i(t) > 0, \quad \inf_{t \in [0,\,1]} \sigma_i^2(t) > 0, \; i = 1, \ldots, k. \tag{11}$$

The densities, regression and variance functions are assumed to be d-times continuously differentiable, i.e.

$$r_i, f_i, \sigma_i \in C^d(0, 1), \quad i = 1, \ldots, k, \tag{12}$$

where $d \geq 2$. As mentioned in section 1 our approach is based on kernel estimators of f_i and σ_i^2. To this end, we require a symmetrical kernel $K : \mathbb{R} \to \mathbb{R}$, which is compactly supported and of order d (cf. Gasser *et al.*, 1985), i.e.

$$\frac{(-1)^j}{j!} \int K(u) u^j \,\mathrm{d}u = \begin{cases} 1, & j = 0 \\ 0, & 1 \leq j \leq d-1, \\ k_d \neq 0, & j = d \end{cases} \quad \int K^2(u)\,\mathrm{d}u < \infty. \tag{13}$$

Let $h = h_N$ denote a sequence of bandwidths, such that

$$Nh^{2d} \to 0 \quad \text{and} \quad Nh^2/(\log h)^2 \to \infty \text{ for } N \to \infty. \tag{14}$$

In the following, we require various estimators for r_i, f_i and σ_i^2. To be concise, the theory will be presented for Nadaraya–Watson-type estimators. However, we mention that local polynomial estimators of higher order will work as well, of course, and because of their better performance at the boundary of the regressor space even better performance is to be expected (Fan & Gijbels, 1996). However, because the suggested test-statistic is an integrated quantity of these function estimators, the boundary behaviour will be of minor importance in the present context. To estimate the design densities r_i, we use

$$\hat{r}_i(x) = \frac{1}{n_i h} \sum_{j=1}^{n_i} K\left(\frac{x - t_{ij}}{h}\right), \tag{15}$$

which yields an estimator for f_i,

$$\hat{f}_i(x) = \frac{1}{n_i h} \sum_{j=1}^{n_i} K\left(\frac{x - t_{ij}}{h}\right) Y_{ij} \frac{1}{\hat{r}_i(x)}, \quad i = 1, \ldots, k. \tag{16}$$

Following the same idea as in section 1, we end up with a k-sample generalization of the ANOVA-Welch statistic (Welch, 1937)

$$T_N = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \hat{f}(t_{ij}))^2 \hat{\sigma}_i^{-2}(t_{ij}) - \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \hat{f}_i(t_{ij}))^2 \hat{\sigma}_i^{-2}(t_{ij}), \tag{17}$$

where a pooled estimator of f is obtained as (when $f_1 = f_2 = \cdots = f_k = f$),

$$\hat{f}(x) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} K(\frac{x-t_{ij}}{h}) Y_{ij} \hat{\sigma}_i^{-2}(t_{ij})}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} K(\frac{x-t_{ij}}{h}) \hat{\sigma}_i^{-2}(t_{ij})}, \tag{18}$$

and $\hat{f}_i$ was defined in (16), for $i = 1, \ldots, k$. Finally, the variances σ_i^2 have to be estimated by a non-parametric estimator, in general (see section 3.1 for a more detailed discussion). We propose an estimator which is similar in spirit to those estimators in Ruppert *et al.* (1997), Fan & Yao (1998) and Härdle & Tsybakov (1997). In the present context, we define

$$\hat{\sigma}_i^2(x) = \frac{1}{n_i h}\sum_{j=1}^{n_i} K\left(\frac{x - t_{ij}}{h}\right)(Y_{ij} - \hat{f}_i(t_{ij}))^2 \frac{1}{\hat{r}_i(x)}, \quad i = 1, \ldots, k. \tag{19}$$

Note that for $k = 2$, $\hat{f}$ equals $\tilde{f}$ defined in (4) using the weights (5) with estimators (15) and (16), that is

$$\hat{f}(x) = \hat{a}(x)\hat{f}_1(x) + (1 - \hat{a}(x))\hat{f}_2(x), \quad \text{where } \hat{a}(x) = \frac{\hat{\sigma}_1^{-2}(x) n_1 \hat{r}_1(x)}{\hat{\sigma}_1^{-2}(x) n_1 \hat{r}_1(x) + \hat{\sigma}_2^{-2}(x) n_2 \hat{r}_2(x)}.$$

Theorem 1 gives the asymptotic distribution of the test-statistic T_N.

Theorem 1

Assume model (7), where the ε_{ij} are independent centred random variables with variance $\mathrm{var}(\varepsilon_{ij}) = 1$ *and* $E[\varepsilon_{ij}^4] \le M < \infty \ \forall i, j$. *Then under assumptions (9)–(14) and* $H_0 : f_1 = \cdots = f_k = f$, *for T_N defined in (17) it holds that*

$$N\sqrt{h}\left(T_N - \frac{C}{Nh}\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, \tau^2),$$

where $\mathcal{N}(0, \tau^2)$ *denotes a centred normal random variable with variance*

$$\tau^2 = 2(k-1)\int (2K - K * K)^2(u)\,\mathrm{d}u,$$

where $*$ denotes convolution. The constant C is defined as $C = 2K(0) - \int K^2(u)\,\mathrm{d}u$.

Remark 1. Fan *et al.* (2001) introduced the method of generalized likelihood ratio test-statistics in the context of various non-parametric one-sample problems. This method has a similar motivation as is given in section 1 for our test-statistic. In spirit of these authors to show the coherence to parametric maximum likelihood ratio tests, we can write our asymptotic result as

$$\frac{r_K \lambda_N - b_N}{\sqrt{2b_N}} \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, 1),$$

where $\lambda_N = NT_N/2$,

$$r_K = \frac{K(0) - \frac{1}{2}\int K^2(u)\,\mathrm{d}u}{(k-1)\int (2K - K * K)^2(u)\,\mathrm{d}u},$$

$b_N = r_K[K(0) - \frac{1}{2}\int K^2(u)\,du]/h$ (compare Th. 5 of Fan *et al.*, 2001). We also observe what the aforementioned authors refer to as the new Wilks phenomenon, i.e. the asymptotic distribution does not depend on unknown parameters.

To test the hypotheses stated in (8), one rejects H_0 at a nominal level α, whenever

$$\frac{N\sqrt{h}\left(T_N - \frac{C}{Nh}\right)}{\tau} > u_{1-\alpha}, \tag{20}$$

where $u_{1-\alpha} = \Phi^{-1}(1-\alpha)$ denotes the $(1-\alpha)$-quantile of the standard normal distribution. Note, that C and τ are known constants. The consistency of the testing procedure (20) against any non-parametric alternative follows from the next result.

Theorem 2

Assume that $f_i \neq f_j$ on a set of positive Lebesgue measure for some i and j in $\{1,\ldots,k\}$. Under the assumptions of Theorem 1 we have

$$\sqrt{N}\left(T_N - \mu\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,\gamma^2),$$

where the constants are defined as

$$\mu = \sum_{j=1}^{k}\sum_{\substack{l=1\\ l<j}}^{k}\int (f_j - f_l)^2(x)\frac{\sigma_l^{-2}(x)\kappa_l r_l(x)\sigma_j^{-2}(x)\kappa_j r_j(x)}{\sum_{l=1}^{k}\sigma_l^{-2}(x)\kappa_l r_l(x)}\,dx \quad \textit{and} \quad \gamma^2 = 4\mu. \tag{21}$$

Theorem 2 can be utilized in various ways. First, a power approximation can be obtained via

$$P_{H_1}\left(\frac{N\sqrt{h}\left(T_N - \frac{C}{Nh}\right)}{\tau} > u_{1-\alpha}\right) = \Phi\left(\frac{\mu\sqrt{N}}{\gamma} - \frac{\tau u_{1-\alpha}}{\gamma\sqrt{Nh}} - \frac{C}{\gamma\sqrt{N}h}\right) + o(1)$$
$$= \Phi\left(\frac{\mu}{\gamma}\sqrt{N}\right) + o(1). \tag{22}$$

We will use this result in section 2.2 to compare the presented test with a procedure of Dette & Neumeyer (2001) in terms of power, see Lemma 1.

Second, a simple one-sided $(1-\alpha)$ confidence interval for the discrepancy measure μ in (21) between the functions f_i $(i=1,\ldots,k)$ is obtained as $(0<\alpha<1/2)$

$$CI_{1-\alpha} = \left[0, T_N + \sqrt{T_N c + \frac{c^2}{4}} + \frac{c}{2}\right] \tag{23}$$

where $c = 4u_{1-(\alpha/2)}^2/N$. The confidence interval (23) might be of some practical appeal because it gives a more accurate insight into *how much* the true regression functions $f_1,\ldots,f_k$ deviate from equality in terms of the discrepancy measure μ. In contrast, a simple decision based on (20) leaves the experimenter in the difficult situation whether rejection of H_0 is based on a *significantly* relevant difference between the f_i, or in the case of acceptance, whether there is really evidence in favour of $f_1 = \cdots = f_k$ or just a lack of power, e.g. because of too small sample sizes. For a careful discussion of these issues, cf. Munk & Dette (1998). Similarly, Theorem 2 allows one to test precise L^2-neighbourhoods

$$H_{\Delta_0}: \mu > \Delta_0 \text{ versus } K_{\Delta_0}: \mu \leq \Delta_0,$$

where Δ_0 is a preassigned discrepance the experimenter is willing to tolerate.

Finally, we mention that the test in (20) can detect local alternatives of the form

$$H_{1_N}: f_i = f + \frac{g_i}{(N\sqrt{h})^{1/2}} \quad \text{for } i=1,\ldots,k, \text{ where } g_i \neq g_j \text{ for some } i \neq j, \tag{24}$$

where $f, g_i \in C^d(0,1)$, that tend to the null hypothesis at a rate $1/(N\sqrt{h})^{1/2}$. Under the local alternatives H_{1_N}, the test-statistic

$$N\sqrt{h}\left(T_N - \frac{C}{Nh}\right)$$

converges in distribution to a normal distribution $\mathcal{N}(\Delta, \tau^2)$ with mean

$$\Delta = \sum_{j=1}^{k}\sum_{\substack{l=1\\ l<j}}^{k} \int (g_j - g_l)^2(x) \frac{\sigma_l^{-2}(x)\kappa_l r_l(x)\sigma_j^{-2}(x)\kappa_j r_j(x)}{\sum_{l=1}^{k}\sigma_l^{-2}(x)\kappa_l r_l(x)}\,\mathrm{d}x.$$

The constants C and τ^2 are defined in Theorem 1. Under (24) we obtain the following first-order approximation of the power,

$$P_{H_{1_N}}\left(N\sqrt{h}\left(T_N - \frac{C}{Nh}\right) > \tau u_{1-\alpha}\right) = \Phi\left(\frac{\Delta}{\tau} - u_{1-\alpha}\right) + o(1). \tag{25}$$

2.2. *Comparison with a procedure of Dette & Neumeyer (2001)*

To simplify the presentation, we restrict to the case $k=2$ in this section. The presented test-statistic T_N is an enhancement of Dette & Neumeyer's (2001) test-statistic

$$T_N^{(1)} = \frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(Y_{ij} - \tilde{f}(t_{ij}))^2 - \frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(Y_{ij} - \hat{f}_i(t_{ij}))^2, \tag{26}$$

where the pooled regression estimator is defined as

$$\tilde{f}(x) = \frac{\sum_{i=1}^{2}\sum_{j=1}^{n_i} K(\frac{x-t_{ij}}{h})Y_{ij}}{\sum_{i=1}^{2}\sum_{j=1}^{n_i} K(\frac{x-t_{ij}}{h})}. \tag{27}$$

$T_N^{(1)}$ does not take into account explicitly the potentially inhomogeneous or heteroscedastic variance functions in the two samples, albeit the test is consistent in these cases. The combined regression estimator $\tilde{f}$ and the test-statistic $T_N^{(1)}$ conform with the definitions of $\hat{f}$ in (18) and T_N in (6) but with replacing the variance estimates $\hat{\sigma}_i^2(\cdot)$ by the constant value 1 $(i=1,2)$. Under the assumptions of Theorems 1 and 2, the statistic $T_N^{(1)}$ has an asymptotic normal law, similar to T_N, but with different constants, i.e.

$$N\sqrt{h}\left(T_N^{(1)} - \frac{\tilde{C}}{Nh}\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, \tilde{\tau}^2) \quad \text{(under } H_0)$$
$$\sqrt{N}(T_N^{(1)} - \tilde{\mu}) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, \tilde{\gamma}^2) \quad \text{(under } H_1),$$

where

$$\tilde{C} = \left[2K(0) - \int K^2(u)\,\mathrm{d}u\right]\left(\int \sigma_1^2(x)\,\mathrm{d}x + \int \sigma_2^2(x)\,\mathrm{d}x - \int \frac{\sigma_1^2(x)\kappa_1 r_1(x) + \sigma_2^2(x)\kappa_2 r_2(x)}{\kappa_1 r_1(x) + \kappa_2 r_2(x)}\,\mathrm{d}x\right)$$

$$\tilde{\tau}^2 = 2\int (2K - K*K)^2(u)\,\mathrm{d}u \int \frac{(\sigma_2^2(x)\kappa_1 r_1(x) + \sigma_1^2(x)\kappa_2 r_2(x))^2}{(\kappa_1 r_1(x) + \kappa_2 r_2(x))^2}\,\mathrm{d}x$$

$$\tilde{\mu} = \int (f_1 - f_2)^2(x)\frac{\kappa_1 r_1(x)\kappa_2 r_2(x)}{\kappa_1 r_1(x) + \kappa_2 r_2(x)}\,\mathrm{d}x$$

$$\tilde{\gamma}^2 = 4\int (f_1 - f_2)^2(x)\frac{\kappa_1 r_1(x)\kappa_2 r_2(x)(\sigma_2^2(x)\kappa_1 r_1(x) + \sigma_1^2(x)\kappa_2 r_2(x))}{(\kappa_1 r_1(x) + \kappa_2 r_2(x))^2}\,\mathrm{d}x.$$

Note, that Dette & Neumeyer's (2001) result holds under slightly less restrictive assumptions on the bandwidth, however, at the price of an additional bias term. A similar generalization could be shown for our statistic as well.

The power approximation (22) (which is analogously valid for $T_N^{(1)}$) motivates that a large value of the ratio of the mean to the asymptotic standard deviation under the alternative yields large power. This gives us the possibility to compare the two competing procedures and leads to the following result.

Lemma 1

Under the assumptions of Theorem 2 (for $k=2$) we obtain for the asymptotic signal-to-noise ratio of T_N and $T_N^{(1)}$ that

$$\frac{\tilde{\mu}}{\tilde{\gamma}} \leq \frac{\mu}{\gamma}. \tag{28}$$

Proof. From Cauchy–Schwarz's inequality, we obtain

$$\begin{aligned}
\tilde{\mu} &= \int (f_1 - f_2)^2(x) \frac{\kappa_1 r_1(x) \kappa_2 r_2(x)}{\kappa_1 r_1(x) + \kappa_2 r_2(x)} \mathrm{d}x \\
&\leq \left(4 \int (f_1 - f_2)^2(x) \frac{\kappa_1 r_1(x) \kappa_2 r_2(x) (\sigma_2^2(x) \kappa_1 r_1(x) + \sigma_1^2(x) \kappa_2 r_2(x))}{(\kappa_1 r_1(x) + \kappa_2 r_2(x))^2} \mathrm{d}x \right)^{1/2} \\
&\quad \times \left(\frac{1}{4} \int (f_1 - f_2)^2(x) \frac{\kappa_1 r_1(x) \kappa_2 r_2(x)}{\sigma_2^2(x) \kappa_1 r_1(x) + \sigma_1^2(x) \kappa_2 r_2(x)} \mathrm{d}x \right)^{1/2} \\
&= \tilde{\gamma} \left(\frac{1}{2} \right) \mu^{1/2} = \tilde{\gamma} \frac{\mu}{\gamma},
\end{aligned}$$

which gives the assertion.

It follows from the Cauchy–Schwarz inequality that one obtains equality in (28) if and only if there exists a constant c such that a.e.

$$\frac{\sigma_2^2 \kappa_1 r_1 + \sigma_1^2 \kappa_2 r_2}{\kappa_1 r_1 + \kappa_2 r_2} \equiv c.$$

This holds in the case of homoscedastic and equal variances in the two samples or in the case of equal design densities and homoscedastic variances. However, from Lemma 1, we also see that Dette & Neumeyer's (2001) statistic becomes inefficient compared with that of our approach, when μ/γ is large compared with $\tilde{\mu}/\tilde{\gamma}$. As an example, assume that $\kappa_1 = \kappa_2 = \frac{1}{2}$ (equal sample sizes), $r_i \equiv 1$ (uniform designs) and let $f_1 - f_2 \equiv 1$. Then

$$\tilde{\mu} = \frac{1}{4}, \quad \tilde{\gamma} = \frac{1}{\sqrt{2}} \left\{ \int (\sigma_1^2(x) + \sigma_2^2(x)) \, \mathrm{d}x \right\}^{1/2}, \quad \mu = \frac{1}{2} \int (\sigma_1^2(x) + \sigma_2^2(x))^{-1} \, \mathrm{d}x, \quad \mu/\gamma = \frac{1}{2} \sqrt{\mu}.$$

Hence inequality (28) in Lemma 1 becomes equivalent to

$$\left(\int (\sigma_1^2(x) + \sigma_2^2(x)) \, \mathrm{d}x \right)^{-1/2} \leq \left(\int (\sigma_1^2(x) + \sigma_2^2(x))^{-1} \, \mathrm{d}x \right)^{1/2}.$$

For example, if $\sigma_1^2(x) + \sigma_2^2(x) = x$, the RHS is infinity, and it is expected that in this case our test outperforms the test of Dette & Neumeyer (2001) significantly. We will investigate this in more detail in section 3.2 where a simulation study is presented.

Remark 2. Under the local alternatives H_{1_N} considered in (24) (for $k=2$) the statistic $T_N^{(1)}$ of Dette & Neumeyer (2001) shows a similar behaviour like T_N but with asymptotic variance $\tilde{\tau}^2$ and mean

$$\tilde{\Delta} = \int (g_1 - g_2)^2(x) \frac{\kappa_1 r_1(x) \kappa_2 r_2(x)}{\kappa_1 r_1(x) + \kappa_2 r_2(x)} \mathrm{d}x.$$

Because of the power approximation in (25) an inequality of the form $\tilde{\Delta}/\tilde{\tau} \leq \Delta/\tau$ as in Lemma 1 for local alternatives would be desirable but is not valid in general.

Remark 3. In the random design case, the design points t_{ij} $(j=1,\dots,n_i)$ are i.i.d. realizations of a random variable X_i with design density r_i $(i=1,2)$. In this setting, the asymptotic distribution under the null hypothesis H_0 stated in Theorem 1 remains valid; but under the fixed alternative H_1, the asymptotic variance changes to

$$\gamma^2 + \sum_{i=1}^{2} \kappa_i \mathrm{var}\left((f_1-f_2)^2(X_i)\frac{\kappa_{3-i}^2 r_{3-i}^2(X_i)\sigma_i^4(X_i) + 2\kappa_i r_i(X_i)\kappa_{3-i} r_{3-i}(X_i)\sigma_{3-i}^4(X_i)}{(\kappa_1 r_1(X_i)\sigma_2^2(X_i) + \kappa_2 r_2(X_i)\sigma_1^2(X_i))^2}\right),$$

where γ^2 is defined in Theorem 2 $(k=2)$.

3. Bandwidth selection and additional prior information on the variances

3.1. Theoretical considerations

All results can be generalized to the use of different bandwidths in the three regression estimates, i.e. bandwidths h_i in $\hat{f}_i(\cdot)$ defined in (16), $i=1,\dots,k$, and a bandwidth h in the pooled estimator $\hat{f}(\cdot)$ defined in (18), cf. Remark 2.7 in Dette & Neumeyer (2001). However, we will not pursue this here. Note that the bandwidth conditions (14) required here are slightly more restrictive than the bandwidth conditions used by Dette & Neumeyer (2001), who also obtained slightly different asymptotics. This is because of the appearance of an additional bias that originates from the variance estimation (19). An advantage of our statistic is that it can be modified in various ways because of prior knowledge on the variances to weaken these bandwidth conditions and to simplify the required estimators. On the one hand, if homoscedasticity of the variances can be assumed, i.e. $\sigma_i^2(\cdot) \equiv \sigma_i^2$, $i=1,\dots,k$, then for the estimation of the constant variance within the ith sample every estimator that satisfies

$$\hat{\sigma}_i^2 - \sigma_i^2 = O_p\left(\frac{1}{\sqrt{N}}\right), \quad i=1,\dots,k$$

can be used, see, for example, Rice (1984) and Hall & Marron (1990) for an estimator which is asymptotically efficient. The bandwidth conditions (14) can then be weakened to the conditions used by Dette & Neumeyer (2001),

$$h = O(N^{-2/(4d+1)}) \quad \text{and} \quad Nh^2 \to \infty \text{ for } N \to \infty \tag{29}$$

and under these conditions we obtain the following limit distributions. Under the null hypothesis H_0 of equal regression functions we have

$$N\sqrt{h}\left(T_N - Bh^{2d} - \frac{C}{Nh}\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,\tau^2),$$

where the constant B is defined by

$$B = k_d^2\left(\int \frac{\{\sum_{i=1}^k \sigma_i^{-2}\kappa_i(f_1 r_i^{(d)} - (f_1 r_i)^{(d)})(x)\}^2}{\sum_{i=1}^k \sigma_i^{-2}\kappa_i r_i(x)}\,\mathrm{d}x - \sum_{i=1}^k \kappa_i \int \{(f_1 r_i)^{(d)}(x) - (f_1 r_i^{(d)})(x)\}^2 \frac{1}{\sigma_i^2 r_i(x)}\,\mathrm{d}x\right),$$

k_d is defined in (13) and C and τ^2 are defined in Theorem 1.

Under the fixed alternative H_1, the same limit distribution as in Theorem 2 holds. If additionally, equality of the variances $\sigma_i^2 = \sigma_0^2$, $i=1,\dots,k$, can be assumed, σ_0^2 could be estimated from the pooled sample, of course. However, in this case weighting by the variances is not necessary at all and our test-statistic essentially reduces to the statistic by Dette & Neumeyer (2001).

On the other hand, the less restrictive bandwidth conditions (29) can also be sufficient in the case where we have extra information about the smoothness of the variance functions. We consider the following setting. Condition (12) is replaced by the assumption

$$r_i, f_i \in C^d(0,1), \quad \sigma_i^2 \in C^s(0,1), \quad i=1,\dots,k,$$

where $s>d$. Moreover, instead of K and h we use a kernel $\tilde{K}$ of order s and a bandwidth $b=b_N$ in the definition (19) of the variance estimate. In place of the bandwidth conditions (14), we assume

$$Nb^{2s}\to 0, \quad Nb^2\to\infty, \quad h^{2d+1/2}=o(b) \quad \text{and} \quad \frac{b}{\sqrt{h}}=O(1) \text{ as } N\to\infty$$

for bandwidth b and the conditions (29) for the bandwidth h used for the regression estimators. Under these assumptions, the same limit distributions for T_N under H_0 and H_1 as stated above for the homoscedastic case hold. In section 3.2, we will investigate numerically the difficult issue of a proper bandwidth selection and we mention that in general a different choice of bandwidth for the estimation of σ_i^2 and f_i might be appropriate, provided it is expected that they differ much in smoothness. Further, it is recommended that for the estimation of the residuals $Y_{ij}-\hat{f}_i(t_{ij})$ required in $\hat{\sigma}_i^2$ in (19) undersmoothing may be advisable, to control the bias of $\hat{\sigma}_i^2$ (see Munk & Ruymgaart, 2002, for an explanation).

3.2. Wild bootstrap and finite sample properties

Although the testing procedure (20) is distribution free and therefore applicable directly without any estimation of nuisance parameters, our simulations indicated that for small and moderate sample sizes the performance of the test can be improved by the bootstrap technique. Hence, in this section, we present the finite sample behaviour of a wild bootstrap version of the proposed testing procedure. We compare it in terms of power with the procedure of Dette & Neumeyer (2001). These authors already compared their test to various procedures and we will show that the new test outperforms the testing procedure of the aforementioned authors in almost all cases. The gain in power may be substantial and has been observed up to twice. For the sake of brevity, we do not present level simulations, but our simulations show that the new procedure keeps the level just as well as Dette & Neumeyer's (2001) test. Moreover, we investigate how sensitive the performance of the procedure is with respect to the choice of the smoothing parameter h. Finally, at the end of this section, we compare the new test to a test by Neumeyer & Dette (2003). According to our simulations, we restrict the following presentation to the comparison of two regression functions, $k=2$. The general case is analogue. We consider the following wild bootstrap approach based on the residuals

$$\hat{\varepsilon}_{ij}=Y_{ij}-\hat{f}(t_{ij}), \quad j=1,\dots,n_i, \quad i=1,2,$$

where $\hat{f}$ is the pooled regression estimator defined in (18). Let V_{ij} denote i.i.d. random variables, independent of the sample $\{Y_{ij}\}$, with masses $(\sqrt{5}+1)/(2\sqrt{5})$ and $(\sqrt{5}-1)/(2\sqrt{5})$ at the points $\frac{1}{2}(1-\sqrt{5})$ and $\frac{1}{2}(1+\sqrt{5})$ respectively. We define bootstrap observations

$$Y_{ij}^*=\hat{f}(t_{ij})+V_{ij}\hat{\varepsilon}_{ij}, \quad j=1,\dots,n_i, \quad i=1,2,$$

and denote by T_N^* the test-statistic defined in (6) but based on the bootstrap sample $\{Y_{ij}^*\}$. A test of asymptotic level α rejects the null hypothesis whenever the statistic T_N (based on the original sample $\{Y_{ij}\}$) is larger than the $(1-\alpha)$-quantile of the distribution of T_N^* conditioned on the sample $\{Y_{ij}\}$. The consistency of this bootstrap procedure can be shown in the same spirit as in the proof of Dette & Neumeyer (2001, section 4.4). In each of the 1000 simulations we resampled $B=200$ times and estimated the bootstrap quantile by $T^*_{N(\lfloor B(1-\alpha)\rfloor)}$, where $T^*_{N(\ell)}$ denotes the ℓth order statistic of the bootstrap sample $T^*_{N,1},\ldots,T^*_{N,B}$.

For all kernel-based estimators we used the Epanechnikov kernel. The bandwidths are chosen according to the 'rule of thumb' (cf. Dette & Neumeyer, 2001), $h_i=(\hat{s}_i^2/n_i)^{0.3}$ in the estimators $\hat{f}_i$ and $\hat{\sigma}_i^2$ $(i=1,2)$ and $h=[(n_1\hat{s}_1^2+n_2\hat{s}_2^2)/N^2]^{0.3}$ in the pooled regression estimator $\hat{f}$. Here $\hat{s}_i^2$ denotes Rice's (1984) estimator

$$\hat{s}_i^2=\frac{1}{2(n_i-1)}\sum_{j=1}^{n_i-1}(Y_{ij+1}-Y_{ij})^2$$

of the integrated variance $s_i^2=\int\sigma_i^2(t)r_i(t)\,\mathrm{d}t$ in the ith sample $(i=1,2)$.

The analogous bootstrap procedure was also simulated for Dette & Neumeyer's (2001) test-statistic $T_N^{(1)}$ defined in (26). We restrict in the following our presentation to normal errors $\varepsilon_{ij}\sim\mathcal{N}(0,1)$ (various other settings have been simulated and yielded similar results) and present the results for different combinations of sample sizes (n_1,n_2) and nominal levels α. First, we consider the case of equidistant design points (i.e. $r_i\equiv 1$, $i=1,2$) in three settings corresponding to the cases of equal homoscedastic, equal heteroscedastic and inhomogeneous heteroscedastic variances. The results for the following regression functions and equal homoscedastic variances,

$$f_1(x)=\exp(x),\quad f_2(x)=\exp(x)+\sin(4\pi x),\quad \sigma_i^2\equiv 0.5,\ i=1,2, \tag{30}$$

can be depicted in Table 1 for the new test-statistic T_N and for Dette & Neumeyer's (2001) procedure for the sake of comparison. The new procedure turns out to be uniformly more powerful in this case.

Table 1. *Simulated power of the wild bootstrap version of the new test-statistic T_N defined in (6) (left panel) and $T_N^{(1)}$ defined in (26) (right panel), according to setting (30)*

	T_N			$T_N^{(1)}$		
(n_1,n_2)	$\alpha=2.5\%$	$\alpha=5\%$	$\alpha=10\%$	$\alpha=2.5\%$	$\alpha=5\%$	$\alpha=10\%$
(10,10)	0.020	0.043	0.083	0.017	0.028	0.054
(10,20)	0.106	0.158	0.232	0.046	0.077	0.126
(10,30)	0.166	0.237	0.340	0.075	0.132	0.211
(10,40)	0.201	0.291	0.402	0.119	0.194	0.301
(10,50)	0.295	0.373	0.510	0.157	0.247	0.357
(20,20)	0.109	0.165	0.262	0.058	0.109	0.162
(20,30)	0.197	0.285	0.399	0.138	0.210	0.298
(20,40)	0.344	0.427	0.545	0.278	0.354	0.459
(20,50)	0.433	0.533	0.645	0.349	0.447	0.543
(30,30)	0.272	0.364	0.484	0.189	0.267	0.377
(30,40)	0.416	0.501	0.624	0.326	0.419	0.530
(30,50)	0.532	0.639	0.739	0.465	0.550	0.644
(40,40)	0.458	0.564	0.663	0.370	0.470	0.567
(40,50)	0.607	0.708	0.797	0.525	0.633	0.728
(50,50)	0.663	0.750	0.822	0.592	0.664	0.755
(100,100)	0.989	0.997	0.997	0.984	0.989	0.993

Table 2. *Simulated power of the wild bootstrap version of the new test-statistic T_N defined in (6) (left panel) and $T_N^{(1)}$ defined in (26) (right panel), according to setting (31)*

	T_N			$T_N^{(1)}$		
(n_1, n_2)	$\alpha = 2.5\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 2.5\%$	$\alpha = 5\%$	$\alpha = 10\%$
(10,10)	0.030	0.054	0.084	0.019	0.033	0.059
(10,20)	0.102	0.149	0.214	0.033	0.059	0.100
(10,30)	0.175	0.234	0.322	0.071	0.132	0.198
(10,40)	0.248	0.328	0.430	0.112	0.183	0.282
(10,50)	0.296	0.379	0.493	0.162	0.239	0.341
(20,20)	0.099	0.152	0.225	0.061	0.089	0.131
(20,30)	0.197	0.267	0.366	0.130	0.181	0.279
(20,40)	0.292	0.401	0.518	0.214	0.301	0.383
(20,50)	0.355	0.449	0.573	0.288	0.376	0.476
(30,30)	0.252	0.328	0.430	0.190	0.257	0.340
(30,40)	0.373	0.473	0.579	0.313	0.407	0.503
(30,50)	0.461	0.590	0.717	0.418	0.522	0.626
(40,40)	0.401	0.513	0.620	0.336	0.416	0.528
(40,50)	0.521	0.648	0.754	0.475	0.563	0.673
(50,50)	0.651	0.734	0.832	0.567	0.662	0.751
(100,100)	0.991	0.995	1.000	0.984	0.990	0.996

The results for equal heteroscedastic variances according to the following setting,

$$f_1(x) = x^2, \quad f_2(x) = x^2 + \sin(4\pi x), \quad \sigma_i^2(x) = x, \quad i = 1, 2, \tag{31}$$

are presented in Table 2. In all cases, we observe a better power of the new test.

Results for the case of inhomogeneous and heteroscedastic variances,

$$f_1 \equiv 1, \quad f_2 \equiv 0, \quad \sigma_1^2(x) = x^2, \quad \sigma_2^2(x) = 5x - x^2 \tag{32}$$

are presented in Table 3. In this case, we observe slightly better power of Dette & Neumeyer's (2001) test for equal and nearly equal sample sizes, but the new procedure outperforms their test, when the sample sizes are rather different, e.g. when $n_1 = 10$, $n_2 = 50$. This phenomenon presumably originates from the interplay of sample size and variance in the weight $1 - a = \sigma_2^{-2} n_2 / (\sigma_1^{-2} n_1 + \sigma_2^{-2} n_2)$ from (5) that is assigned to the observations from the second sample in the pooled regression estimate in the definition of test-statistic T_N. In contrast, the corresponding weight used in test-statistic $T_N^{(1)}$ is $1 - \tilde{a} = n_2 / (n_1 + n_2)$.

Finally, we present simulations for the setting where both the design densities and the variances are different in the two samples,

$$r_1 \equiv 1, \quad r_2(x) = 0.5 + x, \quad f_1 \equiv 1, \quad f_2 \equiv 0, \quad \sigma_1^2 \equiv 2, \quad \sigma_2^2 \equiv 3. \tag{33}$$

The results are shown in Table 4, and the new test turns out to be uniformly more powerful in this case, where for equal sample sizes the gain in power is remarkable. This is perfectly in accordance with our theoretical findings in Lemma 1 and the explanations given in section 2.2.

To investigate how sensitive the performance of the test T_N is with respect to the choice of the smoothing parameters h_1, h_2 and h, we compare results for the choices

$$h_i = c(s_i^2 / n_i)^{0.3} \quad \text{and} \quad h = c((n_1 s_1^2 + n_2 s_2^2)/N^2)^{0.3}, \tag{34}$$

where in the definitions of the 'rule of thumb' bandwidths above we have replaced the estimators $\hat{s}_i^2$ by the true integrated variances s_i^2 and the constant c varies within $\{0.5, 0.75, 1,$

Table 3. *Simulated power of the wild bootstrap version of the new test-statistic T_N defined in (6) (left panel) and $T_N^{(1)}$ defined in (26) (right panel), according to setting (32)*

	T_N			$T_N^{(1)}$		
(n_1, n_2)	$\alpha=2.5\%$	$\alpha=5\%$	$\alpha=10\%$	$\alpha=2.5\%$	$\alpha=5\%$	$\alpha=10\%$
(10,10)	0.254	0.314	0.396	0.302	0.366	0.457
(10,20)	0.381	0.483	0.604	0.313	0.427	0.543
(10,30)	0.501	0.603	0.724	0.325	0.446	0.576
(10,40)	0.585	0.692	0.801	0.350	0.482	0.613
(10,50)	0.660	0.764	0.849	0.354	0.501	0.635
(20,20)	0.402	0.511	0.637	0.524	0.611	0.704
(20,30)	0.522	0.664	0.780	0.628	0.722	0.810
(20,40)	0.684	0.784	0.873	0.707	0.795	0.872
(20,50)	0.741	0.837	0.921	0.724	0.807	0.868
(30,30)	0.614	0.727	0.845	0.761	0.829	0.890
(30,40)	0.704	0.803	0.899	0.784	0.858	0.904
(30,50)	0.826	0.892	0.956	0.866	0.918	0.948
(40,40)	0.762	0.848	0.913	0.852	0.909	0.938
(40,50)	0.873	0.922	0.966	0.892	0.935	0.963
(50,50)	0.867	0.923	0.962	0.929	0.955	0.981
(100,100)	0.998	0.999	1.000	0.998	0.999	0.999

1.25, 1.5}. We display results for level $\alpha=5\%$ and sample sizes $(n_1, n_2)=(50, 50)$. We consider setting (30) as well as a setting corresponding to (30), but according to the null hypothesis, i.e.

$$f_1(x)=\exp(x)=f_2(x), \quad \sigma_i^2\equiv 0.5, \quad i=1, 2. \tag{35}$$

In the first two rows of Table 5 the simulation results are shown. We observe here that the choice of the constant c does not affect the level approximation very much. The power value we have obtained in Table 1 in this setting is 0.750, which is very similar to the value obtained when the estimators $\hat{s}_i^2$ are replaced by there true values, i.e. 0.763. Further, the results show that in this setting with smaller bandwidths even better detections of the alternative can be

Table 4. *Simulated power of the wild bootstrap version of the new test-statistic T_N defined in (6) (left panel) and $T_N^{(1)}$ defined in (26) (right panel), according to setting (33)*

	T_N			$T_N^{(1)}$		
(n_1, n_2)	$\alpha=2.5\%$	$\alpha=5\%$	$\alpha=10\%$	$\alpha=2.5\%$	$\alpha=5\%$	$\alpha=10\%$
(10,10)	0.071	0.109	0.175	0.005	0.008	0.029
(10,20)	0.175	0.234	0.347	0.134	0.225	0.310
(10,30)	0.217	0.287	0.410	0.205	0.282	0.371
(10,40)	0.281	0.355	0.466	0.244	0.341	0.446
(10,50)	0.259	0.346	0.452	0.226	0.323	0.451
(20,20)	0.082	0.139	0.220	0.010	0.021	0.056
(20,30)	0.244	0.311	0.398	0.182	0.239	0.314
(20,40)	0.315	0.421	0.532	0.302	0.386	0.494
(20,50)	0.391	0.496	0.615	0.384	0.477	0.584
(30,30)	0.103	0.157	0.246	0.026	0.051	0.083
(30,40)	0.278	0.366	0.472	0.191	0.257	0.346
(30,50)	0.393	0.500	0.611	0.337	0.432	0.540
(40,40)	0.125	0.195	0.288	0.030	0.050	0.099
(40,50)	0.286	0.378	0.476	0.188	0.266	0.352
(50,50)	0.131	0.193	0.287	0.035	0.067	0.113
(100,100)	0.162	0.243	0.348	0.062	0.111	0.173

Table 5. *Sensitivity of level and power approximations of the wild bootstrap version of T_N defined in (6) for nominal level $\alpha = 5\%$*

c	0.5	0.75	1.0	1.25	1.5
Setting (35) (level)	0.060	0.060	0.049	0.064	0.043
Setting (30) (power)	0.941	0.912	0.763	0.464	0.216
Setting (36) (level)	0.063	0.061	0.055	0.053	0.062
Setting (32) (power)	0.726	0.715	0.707	0.707	0.729

The constant c determining the bandwidth is defined in (34). The first two rows display results according to settings (30) and (35) for sample sizes $(n_1, n_2) = (50, 50)$, the last two rows correspond to settings (32) and (36) for sample sizes $(n_1, n_2) = (30, 30)$.

obtained, and as to be expected, for too large values of the bandwidth the power decreases. The two last rows of Table 5 show corresponding results for setting (32) and a similar model under the null hypothesis, i.e.

$$f_1 \equiv 1 \equiv f_2, \quad \sigma_1^2(x) = x^2, \quad \sigma_2^2(x) = 5x - x^2 \tag{36}$$

for sample sizes $(n_1, n_2) = (30, 30)$. In this model, the power and level results are not sensitive to the changes in the bandwidth.

As was motivated by a referee, finally, we compare the new testing procedure to a test proposed by Neumeyer & Dette (2003). These authors consider a Kolmogorov–Smirnov-type test $K_N^{(2)}$ based on a marked empirical process of estimated residuals. We simulated the performance of the new testing procedure T_N in settings (4.11) of Neumeyer & Dette (2003) with uniformly distributed random design and homoscedastic normally distributed errors with variances $\sigma_1^2 = 0.5$, $\sigma_2^2 = 0.25$ (compare Table 2 in the aforementioned paper). The performances of both tests are similar except for oscillating regression functions. For example, the simulated power with respect to 5% level in setting (v), i.e. $f_1(x) = \exp(x), f_2(x) = \exp(x) + x$ is 0.958 for T_N and 0.996 for $K_N^{(2)}$. For setting (vii) with oscillating alternative, i.e. $f_1(x) = 1, f_2(x) = 1 + \sin(2\pi x)$, we obtain a power of 0.996 for T_N and 0.574 for $K_N^{(2)}$. The approximated levels were very similar for both tests. It is well known, mainly from testing the goodness-of-fit of distribution functions, that tests based on kernel or orthogonal series estimators outperform tests based on empirical processes such as the Cramer von Mises or the Kolmogorov–Smirnov test for most alternatives (for a discussion see Neuhaus, 1976; Eubank & LaRiccia, 1992, among many others). However, for a very few alternatives, which are directed towards the principal eigenfunctions of the Karhunen–Loeve expansion of the underlying stochastic process, the opposite can happen. It is to be expected that a similar situation happens in our context and our simulations support this finding. Settings (30)–(33) give good examples. We display in Table 6 simulated power values of Neumeyer & Dette's (2003) test $K_N^{(2)}$ in these settings. For the sake of brevity, we only show results for sample sizes $(n_1, n_2) = (20, 20)$, $(20, 40)$ and $(50, 50)$. We observe in a comparison with the corresponding values in Tables 1–4 a far better performance of T_N compared with $K_N^{(2)}$ in settings (30) and (31) with oscillating regression functions, but the opposite performance in settings (32) and (33). However, here the differences in power are minor. We finally mention, that our results suggest that Neumeyer & Dette's (2003) tests [and Pardo-Fernández *et al.*'s (2006) test as well], which are based on the pooled regression estimator defined in (27) could be further improved when using the pooled regression estimator (18) instead, but these investigations are clearly beyond the scope of this paper.

4. A data example from crop yield

We will reanalyse an experiment provided by Ratkowsky (1983), later analysed in Bowman & Azzalini (1997) and Mammen *et al.* (2001). It was designed to investigate the relationship

Table 6. *Simulated power of Neumeyer & Dette's (2003) test $K_N^{(2)}$ according to settings (30)–(33)*

	Setting (30)			Setting (31)		
(n_1, n_2)	$\alpha=2.5\%$	$\alpha=5\%$	$\alpha=10\%$	$\alpha=2.5\%$	$\alpha=5\%$	$\alpha=10\%$
(20,20)	0.013	0.030	0.070	0.017	0.046	0.074
(20,40)	0.022	0.044	0.101	0.013	0.031	0.080
(50,50)	0.046	0.077	0.064	0.023	0.055	0.133
	Setting (32)			Setting (33)		
(20,20)	0.723	0.800	0.883	0.485	0.580	0.689
(20,40)	0.879	0.934	0.971	0.655	0.755	0.843
(50,50)	0.983	0.989	0.997	0.888	0.934	0.958

between the yield of onion plants and the density of planting. To this end measurements were taken at two different locations, Purnong Landing ($\times$) and Virginia ($\circ$), in South Australia. Figure 1(a) shows the data (plant density versus mean yield per plant) at these locations together with a local linear smoother and Figure 1(b) displays local linear fits after recentring both samples by its mean, respectively, to adjust for a possible constant location effect as described in Bowman & Azzalini (1997). Note, that asymptotically this correction does not affect the distribution of T_N, because the overall mean can be estimated $\sqrt{N}$-consistently. In Figure 1(c), a local linear fit of the standard deviation is displayed (cf. Ruppert *et al.*, 1997). From this we draw that the SDs are decreasing, hence heteroscedastic, further in both locations they differ significantly (i.e. they are inhomogenous), in particular when the plant density is smaller than 75 (cf. Figure 1(c) again). An application of our test yields $T_N = 0.35326$ and the following critical values were bootstraped for the levels $\alpha = 0.025, 0.05$ and 0.10, respectively ($B = 200$), $c_{0.025} = 0.082$, $c_{0.05} = 0.064$ and $c_{0.1} = 0.05$. Hence, T_N exceeds all these values and the assumption of equality is rejected at any of these levels. To investigate whether the result is sensitive with respect to the choice of the bandwidths, we repeated the same calculations for different bandwidths, but the hypothesis was always rejected. Bowman & Azzalini (1997) performed a similar analysis for their test, also based on smoothed curve estimators of the regression functions f_1 and f_2. Their test-statistic relies on the assumption of equal and constant variances in both groups and they found that for a smaller bandwidth a significant rejection of equality resulted, whereas for larger bandwidth this failed to hold. This slightly different finding may be because of a gain in power of our test when non-constant and unequal variances are present.

5. Conclusion

In this paper, we have suggested a new procedure for testing the equality of regression curves in different non-parametric regression models. The new test generalizes naturally the method of analysis of covariance to the setting of non-parametric regression. The asymptotic normal distribution of the proposed test-statistic under the null hypothesis of equal regression functions as well as under fixed and local alternatives is shown. Under the null hypothesis, the test turns out to be asymptotically distribution free. Our procedure is similar in spirit to a test based on a difference of variance estimators recommended by Dette & Neumeyer (2001). We have shown that the new test gains in power particularly in the case of inhomogeneous and heteroscedastic variances and for different sample sizes or design densities, respectively. We found that the actual level of the test is robust for a large scale of bandwidths, whereas the power may change significantly as to be expected because of the different resolution

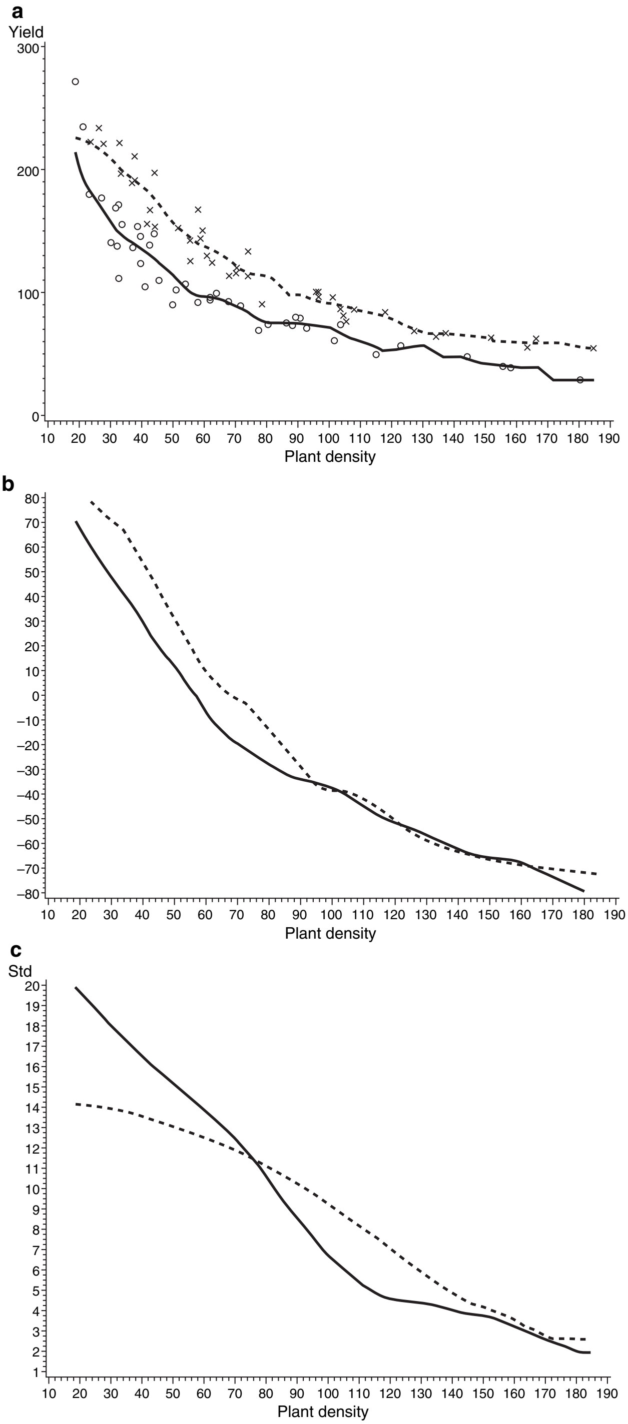

Fig. 1. (a) Scatter plot and local linear fit of the yield at Purnong Landing (×) and Virginia (○). (b) Local polynomial regression of the yield at Purnong Landing (dashed line) and Virginia (solid line) after adjustment. (c) Local polynomial regression of the standard deviation functions at Purnong Landing (dashed line) and Virginia (solid line).

levels. On a fine scale, smaller changes will be more likely to be detected leading to a larger rejection rate.

Acknowledgements

The financial support of the Deutsche Forschungsgemeinschaft is gratefully acknowledged. The authors are also grateful to two unknown referees for their constructive comments on an earlier version of this manuscript.

References

Akritas, M. & Van Keilegom, I. (2001). Nonparametric estimation of the residual distribution. *Scand. J. Stat.* **28**, 549–567.

Bowman, A. W. & Azzalini, A. (1997). *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*. Oxford University Press, Oxford.

Cabus, P. (1998). Un test de type Kolmogorov–Smirnov dans le cadre de comparaison de fonctions de régression. *Comptes Rendus des Séances de l'Académie des Sciences. Série I. Mathématique* **327**, 939–942.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica* **28**, 591–605.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Relat. Fields* **75**, 261–277.

Delgado, M. A. (1993). Testing the equality of nonparametric regression curves. *Stat. Probab. Lett.* **17**, 199–204.

Dette, H. & Neumeyer, N. (2001). Nonparametric analysis of covariance. *Ann. Stat.* **29**, 1361–1400.

Eubank, R. L. & LaRiccia, V. N. (1992). Asymptotic comparison of Cramer–von Mises and nonparametric function estimation techniques for testing goodness-of-fit. *Ann. Stat.* **20**, 2071–2086.

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

Fan, J. & Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85**, 645–660.

Fan, J., Zhang, C. & Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Stat.* **29**, 153–193.

Gasser, T., Müller, H.-G. & Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Stat. Ser. B* **47**, 238–252.

Gørgens, T. (2002). Nonparametric comparison of regression curves by local linear fitting. *Stat. Probab. Lett.* **60**, 81–89.

Hall, P. & Hart, J. W. (1990). Bootstrap test for difference between means in nonparametric regression. *J. Amer. Stat. Assoc.* **85**, 1039–1049.

Hall, P. & Marron, J. S. (1990). On variance estimation in nonparametric regression. *Biometrika* **77**, 415–419.

Härdle, W. & Tsybakov, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *J. Econometrics* **81**, 223–242.

Kulasekera, K. B. (1995). Comparison of regression curves using quasi-residuals. *J. Amer. Stat. Assoc.* **90**, 1085–1093.

Lavergne, P. (2001). An equality test across nonparametric regressions. *J. Econometrics* **103**, 307–344.

Mammen, E., Marron, J. S., Turlach, B. A. & Wand, M. P. (2001). A general projection framework for constrained smoothing. *Stat. Sci.* **16**, 232–248.

Müller, H. G. (1985). Kernel estimators of zeros and of location and size of extrema and regression functions. *Scand. J. Stat.* **12**, 221–232.

Munk, A. & Dette, H. (1998). Nonparametric comparison of several regression functions: exact and asymptotic theory. *Ann. Stat.* **26**, 2339–2368.

Munk, A. & Ruymgaart, F. (2002). Minimax rates for estimating the variance and its derivatives in nonparametric regression. *Aust. N. Z. J. Stat.* **44**, 479–488.

Neuhaus, G. (1976). Asymptotic power properties of the Cramer–von Mises test under contiguous alternatives. *J. Multivariate Anal.* **6**, 95–110.

Neumeyer, N. & Dette, H. (2003). Nonparametric comparison of regression curves – an empirical process approach. *Ann. Stat.* **31**, 880–920.

Pardo-Fernández, J. C., Van Keilegom, I. & González-Manteiga, W. (2006). Comparison of regression curves based on the estimation of the error distribution. *Statist. Sinica* (in press).
Ratkowsky, D. A. (1983). *Nonlinear regression modelling*. Dekker, New York.
Rice, J. A. (1984). Bandwidth choice for nonparametric regression. *Ann. Stat.* **12**, 1215–1230.
Ruppert, D., Wand, M. P., Holst, U. & Hössjer, O. (1997). Local polynomial variance-function estimation. *Technometrics* **39**, 262–273.
Sacks, J. & Ylvisaker, D. (1970). Designs for regression problems for correlated errors. *Ann. Math. Stat.* **41**, 2057–2074.
Scheffé, H. (1959). *The analysis of variance*. Wiley, New York.
Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivates. *Ann. Stat.* **6**, 177–184.
Weerahandi, S. (1987). Testing regression equality with unequal variances. *Econometrica* **55**, 1211–1215.
Welch, B. L. (1937). The significance of the difference between two means when the population variances are unequal. *Biometrika* **29**, 350–362.
Yatchew, A. (1999). An elementary nonparametric differencing test of equality of regression functions. *Econ. Lett.* **62**, 271–278.
Young, S. G. & Bowman, A. W. (1995). Non-parametric analysis of covariance. *Biometrics* **51**, 920–931.

Received April 2005, in final form July 2006

Axel Munk, Georg-August Universität, Maschmühlenweg 8–10, 37073 Göttingen, Germany.
E-mail: munk@math.uni-goettingen.de

Appendix A: Proofs

A.1. Proof of Theorems 1 and 2

The strategy of the proof is related to the proof of Theorem 2.1 of Dette & Neumeyer (2001). However, technically it becomes much more involved because of the additional variance estimators in T_N. For the sake of brevity, we will only state the main differences because of the additional variance estimation, and further assume $k=2$. With the definition of weights

$$w_{jk}^{(i)} = \frac{K\left(\frac{t_{ij}-t_{ik}}{h}\right)}{\sum_{l=1}^{n_i} K\left(\frac{t_{ij}-t_{il}}{h}\right)} \quad \text{and} \quad w_{lk,\,ij} = \frac{1}{Nh} K\left(\frac{t_{lk}-t_{ij}}{h}\right) \hat{\sigma}_{3-l}^2(t_{ij}) \frac{1}{\hat{R}(t_{ij})}, \tag{37}$$

where $\hat{R}(t) = [n_1 \hat{r}_1(t) \hat{\sigma}_2^2(t)/N] + [n_2 \hat{r}_2(t) \hat{\sigma}_1^2(t)/N]$ is an estimator for $R(t) = \kappa_1 r_1(t) \sigma_2^2(t) + \kappa_2 r_2(t) \sigma_1^2(t)$, the regression estimators defined in (16) and (18), respectively, are $\hat{f}_i(t_{ij}) = \sum_{k=1}^{n_i} w_{jk}^{(i)} Y_{ik}$ $(i=1,2)$ and $\hat{f}(t_{ij}) = \sum_{l=1}^{2} \sum_{k=1}^{n_l} w_{lk,\,ij} Y_{lk}$. Now with the notations $(j=1,\ldots,n_i$, $i=1,2)$

$$\Delta_{ij} = f_i(t_{ij}) - \sum_{l=1}^{2} \sum_{k=1}^{n_l} w_{lk,\,ij} f_l(t_{lk}) = \sum_{l=1}^{2} \sum_{k=1}^{n_l} w_{lk,\,ij} (f_i(t_{ij}) - f_l(t_{lk})) \tag{38}$$

$$\delta_{ij} = f_i(t_{ij}) - \sum_{k=1}^{n_i} w_{jk}^{(i)} f_i(t_{ik}) = \sum_{k=1}^{n_i} w_{jk}^{(i)} (f_i(t_{ij}) - f_i(t_{ik})) \tag{39}$$

we decompose T_N in (6) as

$$\begin{aligned} T_N = \frac{1}{N} \sum_{i=1}^{2} \sum_{j=1}^{n_i} \hat{\sigma}_i^{-2}(t_{ij}) \Bigg\{ & \Delta_{ij}^2 - \delta_{ij}^2 - 2\Delta_{ij} \sum_{l=1}^{2} \sum_{k=1}^{n_l} w_{lk,\,ij} \sigma_l(t_{lk}) \varepsilon_{lk} + 2\delta_{ij} \sum_{k=1}^{n_i} w_{jk}^{(i)} \sigma_i(t_{ik}) \varepsilon_{ik} \\ & + \left(\sum_{l=1}^{2} \sum_{k=1}^{n_l} w_{lk,\,ij} \sigma_l(t_{lk}) \varepsilon_{lk} \right)^2 - \left(\sum_{k=1}^{n_i} w_{jk}^{(i)} \sigma_i(t_{ik}) \varepsilon_{ik} \right)^2 + 2\sigma_i(t_{ij}) \varepsilon_{ij} (\Delta_{ij} - \delta_{ij}) \\ & - 2\sigma_i(t_{ij}) \varepsilon_{ij} \sum_{l=1}^{2} \sum_{k=1}^{n_l} w_{lk,\,ij} \sigma_l(t_{lk}) \varepsilon_{lk} + 2\sigma_i(t_{ij}) \varepsilon_{ij} \sum_{k=1}^{n_i} w_{jk}^{(i)} \sigma_i(t_{ik}) \varepsilon_{ik} \Bigg\}. \end{aligned} \tag{40}$$

Lemma 2

Under the assumptions of Theorem 1 we obtain an expansion of the test-statistic under the null hypothesis H_0, $T_N=\bar{T}_N+o_p(1/(N\sqrt{h}))$, *where* $E[\bar{T}_N]=C/(Nh)+o(1/(N\sqrt{h}))$. *Under the alternative* H_1, *we have* $T_N=\bar{\bar{T}}_N+o_p(1/\sqrt{N})$, *where* $E[\bar{\bar{T}}_N]=\mu+o(1/\sqrt{N})$ *and where the constants* C *and* μ *are defined in the Theorems 1 and 2.*

Proof. We use the above definitions and the decomposition (40) of the test-statistic T_N. A Taylor expansion together with (37) and (38) gives

$$\Delta_{ij}=\sum_{l=1}^{2}\frac{\hat{\sigma}^2_{3-l}(t_{ij})}{\hat{R}(t_{ij})}\frac{1}{Nh}\sum_{k=1}^{n_l}K\left(\frac{t_{ij}-t_{lk}}{h}\right)(f_i(t_{ij})-f_l(t_{lk}))$$

$$=\sum_{l=1}^{2}\frac{\hat{\sigma}^2_{3-l}(t_{ij})}{\hat{R}(t_{ij})}\left\{(f_i(t_{ij})-f_l(t_{ij}))\kappa_l r_l(t_{ij})+O(h^d)+O\left(\frac{1}{Nh}\right)\right\} \tag{41}$$

$$=\frac{\hat{\sigma}^2_i(t_{ij})}{\hat{R}(t_{ij})}(f_i(t_{ij})-f_{3-i}(t_{ij}))\kappa_{3-i}r_{3-i}(t_{ij})+\left\{O(h^d)+O\left(\frac{1}{Nh}\right)\right\}O_p(1) \tag{42}$$

where the last line only holds under the alternative H_1. For the sake of brevity, we explain our argumentation in detail only for the first term on the RHS in (40). For this, we obtain under the alternative H_1,

$$\sum_{i=1}^{2}\sum_{j=1}^{n_i}\hat{\sigma}_i^{-2}(t_{ij})\Delta_{ij}^2/N=A_N+B_N+C_N+o_p(1/\sqrt{N}),$$

where

$$A_N=\frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\frac{\sigma_i^2(t_{ij})}{R^2(t_{ij})}(f_i(t_{ij})-f_{3-i}(t_{ij}))^2\kappa_{3-i}^2r_{3-i}^2(t_{ij})$$

$$B_N=\frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\frac{\hat{\sigma}_i^2(t_{ij})-\sigma_i^2(t_{ij})}{R^2(t_{ij})}(f_i(t_{ij})-f_{3-i}(t_{ij}))^2\kappa_{3-i}^2r_{3-i}^2(t_{ij})$$

$$C_N=\frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\hat{\sigma}_i^2(t_{ij})\Big(\frac{1}{\hat{R}^2(t_{ij})}-\frac{1}{R^2(t_{ij})}\Big)(f_i(t_{ij})-f_{3-i}(t_{ij}))^2\kappa_{3-i}^2r_{3-i}^2(t_{ij}).$$

For the (non-random) A_N, we have by a Riemann-sum approximation

$$A_N=\sum_{i=1}^{2}\int\frac{\sigma_i^2(t)}{R^2(t)}(f_i(t)-f_{3-i}(t))^2\kappa_{3-i}^2r_{3-i}^2(t)\kappa_i r_i(t)\,\mathrm{d}t+o\left(\frac{1}{\sqrt{N}}\right)=\mu+o\left(\frac{1}{\sqrt{N}}\right).$$

With an application of Proposition 1 in Section A.2 and some tedious calculations of expectations and variances we obtain $B_N=o_p(1/\sqrt{N})$. To show $C_N=o_p(1/\sqrt{N})$, we use the decomposition

$$C_N=\frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\Big[\left(\hat{\sigma}_i^2(t_{ij})-\sigma_i^2(t_{ij})\right)\left(R(t_{ij})-\hat{R}(t_{ij})\right)\Big(\frac{1}{\hat{R}(t_{ij})}+\frac{1}{R(t_{ij})}\Big)\frac{1}{\hat{R}(t_{ij})R(t_{ij})}$$
$$+\left(R(t_{ij})-\hat{R}(t_{ij})\right)^2\frac{\sigma_i^2(t_{ij})}{R^2(t_{ij})\hat{R}(t_{ij})}\Big(\frac{1}{\hat{R}(t_{ij})}+\frac{2}{R(t_{ij})}\Big)\Big](f_i(t_{ij})-f_{3-i}(t_{ij}))^2\kappa_{3-i}^2r_{3-i}^2(t_{ij})$$
$$+\frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(R(t_{ij})-\hat{R}(t_{ij})\right)\frac{2\sigma_i^2(t_{ij})}{R^3(t_{ij})}(f_i(t_{ij})-f_{3-i}(t_{ij}))^2\kappa_{3-i}^2r_{3-i}^2(t_{ij}).$$

We have uniform almost sure convergence of $\hat{R}$ to R and $\hat{\sigma}_i^2$ to σ_i^2 with rates $O((\log h^{-1}/(Nh))^{1/2})$, see, for instance, Silverman (1978), Müller (1985) and Akritas & Van Keilegom (2001). Therefore, the first sum in the decomposition of C_N is of order $O_p(\log h^{-1}/(Nh)) = o_p(1/\sqrt{N})$ by assumption (14). By definition of $\hat{R}$, we have

$$R(t) - \hat{R}(t) = -\sum_{i=1}^{2} \left[\hat{r}_{3-i}(t)\ (\hat{\sigma}_i^2(t) - \sigma_i^2(t)) + \sigma_i^2(t)\ (\hat{r}_{3-i}(t) - r_{3-i}(t)) \right],$$

where $\hat{r}_{3-i}(t)$ is deterministic and converges to $r_{3-i}(t)$ with rate $o(1/\sqrt{N})$. Now, the expansion of $\hat{\sigma}_i^2(t) - \sigma_i^2(t)$ from Proposition 1 (see Section A.2) can be used to deduce that the second sum in the decomposition of C_N has an expectation of order $O(h^d) + O(1/(Nh)) = o(1/\sqrt{N})$.

Under the null hypothesis H_0, we directly obtain from (41)

$$\frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i} \hat{\sigma}_i^{-2}(t_{ij})\Delta_{ij}^2 = O_p(1)\left\{ O(h^d) + O\left(\frac{1}{Nh}\right)\right\}^2 = o_p\left(\frac{1}{N\sqrt{h}}\right).$$

With similar considerations as above, we obtain for the terms

$$\frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i} \hat{\sigma}_i^{-2}(t_{ij})\delta_{ij}^2, \quad \frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i} \hat{\sigma}_i^{-2}(t_{ij})\Delta_{ij}\sum_{l=1}^{2}\sum_{k=1}^{n_l} w_{lk,\,ij}\sigma_l(t_{lk})\varepsilon_{lk},$$

$$\text{and} \quad \frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i} \hat{\sigma}_i^{-2}(t_{ij})\delta_{ij}\sum_{k=1}^{n_i} w_{jk}^{(i)}\sigma_i(t_{ik})\varepsilon_{ik}$$

respective decompositions of the form $D_N + \tilde{D}_N$ where $E[D_N] = o(1/\sqrt{N})$, $\tilde{D}_N = o_p(1/\sqrt{N})$ under H_1 and $E[D_N] = o(1/(N\sqrt{h}))$, $\tilde{D}_N = o_p(1/(N\sqrt{h}))$ under H_0. With (37) and Proposition 1, we further obtain

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i} \hat{\sigma}_i^{-2}(t_{ij})\left(\sum_{l=1}^{2}\sum_{k=1}^{n_l} w_{lk,\,ij}\sigma_l(t_{lk})\varepsilon_{lk}\right)^2 \\
&= \frac{1}{N^3h^2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\sum_{l=1}^{2}\sum_{k=1}^{n_l} \frac{\hat{\sigma}_i^{-2}(t_{ij})}{\hat{R}^2(t_{ij})}\hat{\sigma}_{3-l}^4(t_{ij})K^2\left(\frac{t_{ij}-t_{lk}}{h}\right)\sigma_l^2(t_{lk})\varepsilon_{lk}^2 \\
&\quad + \frac{1}{N^3h^2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\sum_{l=1}^{2}\sum_{k=1}^{n_l}\sum_{l'=1}^{2}\sum_{\substack{k'=1\\(l,k)\neq(l',k')}}^{n_{l'}} \frac{\hat{\sigma}_i^{-2}(t_{ij})}{\hat{R}^2(t_{ij})}\hat{\sigma}_{3-l}^2(t_{ij})\hat{\sigma}_{3-l'}^2(t_{ij}) \\
&\quad \times K\left(\frac{t_{ij}-t_{lk}}{h}\right)K\left(\frac{t_{ij}-t_{l'k'}}{h}\right)\sigma_l(t_{lk})\sigma_{l'}(t_{l'k'})\varepsilon_{lk}\varepsilon_{l'k'}.
\end{aligned}$$

The expectation of the dominating term is

$$\begin{aligned}
&\frac{1}{Nh^2}\sum_{i=1}^{2}\sum_{l=1}^{2}\int\int \frac{\sigma_i^{-2}(t)}{R^2(t)}\sigma_{3-l}^4(t)K^2\left(\frac{t-x}{h}\right)\sigma_l^2(x)\kappa_i r_i(t)\kappa_l r_l(t)\,\mathrm{d}t\,\mathrm{d}x \\
&= \frac{1}{Nh}\int K^2(u)\,\mathrm{d}u + o\left(\frac{1}{N\sqrt{h}}\right).
\end{aligned}$$

An analogous calculation yields for the dominating term of

$$-\frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i} \hat{\sigma}_i^{-2}(t_{ij})\left(\sum_{k=1}^{n_i} w_{jk}^{(i)}\sigma_i(t_{ik})\varepsilon_{ik}\right)^2$$

the expectation $[-2/(Nh)]\int K^2(u)du + o(1/(N\sqrt{h}))$. Similarly, we obtain for

$$-\frac{2}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\hat{\sigma}_i^{-2}(t_{ij})\sigma_i(t_{ij})\varepsilon_{ij}\sum_{l=1}^{2}\sum_{k=1}^{n_l}w_{lk,\,ij}\sigma_l(t_{lk})\varepsilon_{lk}$$

an expectation $-2K(0)/(Nh)$ of the dominating term. Further, we have a decomposition

$$\frac{2}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\hat{\sigma}_i^{-2}(t_{ij})\sigma_i(t_{ij})\varepsilon_{ij}\sum_{k=1}^{n_i}w_{jk}^{(i)}\sigma_i(t_{ik})\varepsilon_{ik}=D_N+\tilde{D}_N$$

where $\tilde{D}_N=o_p(1/(N\sqrt{h}))$ and $E[D_N]=4K(0)/(Nh)+o(1/(N\sqrt{h}))$. Analogous to the previous calculations, we obtain that

$$\sum_{i=1}^{2}\sum_{j=1}^{n_i}\hat{\sigma}_i^{-2}(t_{ij})\sigma_i(t_{ij})\varepsilon_{ij}(\Delta_{ij}-\delta_{ij})/N$$

is of order $O_p(1/(Nh))=o_p(1/\sqrt{N})$ under H_1 and of order

$$O_p(1/(Nh))\ (O(h^d)+O(1/(Nh)))=o_p(1/(N\sqrt{h}))$$

under H_0. From the decomposition (40) of T_N and the above calculation the assertion follows.

A.1.1. Proof of Theorem 2

Analogous to the proof of Theorem 2.1, Dette & Neumeyer (2001), the following expansion of the test-statistic holds under the alternative H_1: $T_N-E[T_N]=T_N^{(1)}+T_N^{(2)}+o_p(1/\sqrt{N})$, where

$$T_N^{(i)}=\frac{1}{N}\sum_{j=1}^{n_i}\alpha_{ij}\varepsilon_{ij},\quad i=1,2$$

and the coefficients are defined by $\alpha_{ij}=2\Delta_{ij}\sigma_i(t_{ij})/\hat{\sigma}_i{}^2(t_{ij})$, $j=1,\ldots,n_i$ $(i=1,2)$.

Lemma 3

Under the assumptions of Theorem 1 and under the alternative H_1, it holds that $T_N^{(i)}=\bar{T}_N^{(i)}+o_p(1/N)$ where

$$\mathrm{var}(\bar{T}_N^{(i)})=\frac{4}{N}\int(f_1-f_2)^2(x)\frac{\kappa_i r_i(x)\kappa_{3-i}^2 r_{3-i}^2(x)\sigma_i^2(x)}{(\kappa_1 r_1(x)\sigma_2^2(x)+\kappa_2 r_2(x)\sigma_1^2(x))^2}\,\mathrm{d}x,\quad i=1,2.$$

Proof. We only consider the case $i=1$. With Δ_{1j} from (42) we obtain

$$T_N^{(1)}=\frac{2}{N}\sum_{j=1}^{n_1}(f_1(t_{1j})-f_2(t_{1j}))\kappa_2 r_2(t_{1j})\frac{\sigma_1(t_{1j})}{\hat{R}(t_{1j})}\varepsilon_{1j}+o_p\left(\frac{1}{\sqrt{N}}\right).$$

Now for calculating the variance of the dominating term we can substitute $\hat{R}(t)$ by $R(t)$. The remainder of the expansion of $1/\hat{R}(t)$ around $1/R(t)$ is equal to

$$\frac{R(t)-\hat{R}(t)}{\hat{R}(t)R(t)}=-\frac{1}{R^2(t)}\sum_{i=1}^{2}\Big\{\hat{r}_{3-i}(t)\ (\hat{\sigma}_i^2(t)-\sigma_i^2(t))+\sigma_i^2(t)\ (\hat{r}_{3-i}(t)-r_{3-i}(t))\Big\}(1+o_p(1)).$$

This yields the rest of the terms $T_N^{(1,i)}$ $(i=1,2)$ in the expansion $T_N^{(1)} = \bar{T}_N^{(1)} + T_N^{(1,1)} + T_N^{(1,2)} + o_p(1/\sqrt{N})$, where

$$\bar{T}_N^{(1)} = \frac{2}{N}\sum_{j=1}^{n_1}(f_1(t_{1j}) - f_2(t_{1j}))\ \kappa_2 r_2(t_{1j})\frac{\sigma_1(t_{1j})}{R(t_{1j})}\varepsilon_{1j}$$

and the remainders are of the form

$$T_N^{(1,i)} = \frac{1}{N}\sum_{j=1}^{n_1}\Delta(t_{1j})\varepsilon_{1j}\Big\{\big(r_{3-i}(t_{1j}) + o(1)\big)\,(\hat{\sigma}_i^2(t_{1j}) - \sigma_i^2(t_{1j})) + o(1)\Big\} = o_p\left(\frac{1}{\sqrt{N}}\right).$$

The last equality can be obtained by inserting the decomposition of the variance estimator $\hat{\sigma}_i^2(t)$ from Proposition 1 (see Section A.2) and a tedious calculation of the variance $\mathrm{var}(T_N^{(1,i)}) = o(1/N)$ $(i=1,2)$. We obtain for the variance of $\bar{T}_N^{(1)}$,

$$\mathrm{var}(\bar{T}_N^{(1)}) = \frac{4}{N}\int (f_1 - f_2)^2(x)\frac{\kappa_1 r_1(x)\kappa_2^2 r_2^2(x)\sigma_1^2(x)}{R^2(x)}\,\mathrm{d}x$$

and this completes the proof of Lemma 3.

From the proof of Lemma 2, we additionally obtain under the alternative H_1:

$$\sqrt{N}(T_N - E[T_N]) = \frac{1}{\sqrt{N}}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\varepsilon_{ij}(f_i(t_{ij}) - f_{3-i}(t_{ij}))\kappa_{3-i}r_{3-i}(t_{ij})\frac{\sigma_i(t_{ij})}{R(t_{ij})} + o_p(1)$$

with the asymptotic variance (of the dominating term)

$$4\int (f_1 - f_2)^2(x)\frac{\kappa_1 r_1(x)\kappa_2^2 r_2^2(x)\sigma_1^2(x)}{R^2(x)}\,\mathrm{d}x + 4\int (f_1 - f_2)^2(x)\frac{\kappa_2 r_2(x)\kappa_1^2 r_1^2(x)\sigma_2^2(x)}{R^2(x)}\,\mathrm{d}x = \gamma^2.$$

An application of the central limit theorem using Lyapunov's condition yields the asymptotic normality and completes the proof of Theorem 2.

A.1.2. Proof of Theorem 1

Under the hypothesis H_0 of equal regression functions in the two models, we obtain similar to the proof of Theorem 2.1 of Dette & Neumeyer (2001) the decomposition $T_N - E[T_N] = \sum_{j=3}^{5} T_N^{(j)} + o_p(1/(N\sqrt{h}))$, where

$$T_N^{(2+k)} = \frac{1}{N}\sum_{i=1}^{n_k}\sum_{\substack{j=1\\ j\neq i}}^{n_k}\beta_{ij}^{(k)}\varepsilon_{ki}\varepsilon_{kj},\quad k=1,2,\qquad T_N^{(5)} = \frac{1}{N}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\gamma_{ij}\varepsilon_{1i}\varepsilon_{2j}$$

and the coefficients are defined by

$$\beta_{ij}^{(1)} = \left\{\sum_{l=1}^{2}\sum_{k=1}^{n_l}\frac{w_{1i,lk}w_{1j,lk}}{\hat{\sigma}_l^2(t_{lk})} - \frac{2w_{1j,1i}}{\hat{\sigma}_1^2(t_{1i})} - \sum_{k=1}^{n_1}\frac{w_{ki}^{(1)}w_{kj}^{(1)}}{\hat{\sigma}_1^2(t_{1k})} + \frac{2w_{ij}^{(1)}}{\hat{\sigma}_1^2(t_{1i})}\right\}\sigma_1(t_{1i})\sigma_1(t_{1j})$$

$$\beta_{ij}^{(2)} = \left\{\sum_{l=1}^{2}\sum_{k=1}^{n_l}\frac{w_{2i,lk}w_{2j,lk}}{\hat{\sigma}_l^2(t_{lk})} - \frac{2w_{2j,2i}}{\hat{\sigma}_2^2(t_{2i})} - \sum_{k=1}^{n_2}\frac{w_{ki}^{(2)}w_{kj}^{(2)}}{\hat{\sigma}_2^2(t_{2k})} + \frac{2w_{ij}^{(2)}}{\hat{\sigma}_2^2(t_{2i})}\right\}\sigma_2(t_{2i})\sigma_2(t_{2j})$$

$$\gamma_{ij} = \left\{2\sum_{l=1}^{2}\sum_{k=1}^{n_l}\frac{w_{1i,lk}w_{2j,lk}}{\hat{\sigma}_l^2(t_{lk})} - \frac{2w_{2j,1i}}{\hat{\sigma}_1^2(t_{1i})} - \frac{2w_{1i,2j}}{\hat{\sigma}_2^2(t_{2j})}\right\}\sigma_1(t_{1i})\sigma_2(t_{2j}).$$

Lemma 4

Under the assumptions of Theorem 1 and under the null hypothesis H_0, it holds that $T_N^{(2+k)} = \bar{T}_N^{(2+k)} + o_p(1/(N\sqrt{h}))\,(k=1,2,3)$ where

$$\operatorname{var}(\bar{T}_N^{(2+k)}) = \frac{2}{N^2h}\int (2K - K*K)^2(u)\,\mathrm{d}u$$
$$\times\left[1+\int_0^1 \frac{\kappa_k^2 r_k^2(x)\sigma_{3-k}^4(x)}{R^2(x)}\mathrm{d}x - 2\int_0^1 \frac{\kappa_k r_k(x)\sigma_{3-k}^2(x)}{R(x)}\mathrm{d}x\right] + o\left(\frac{1}{N^2h}\right),\quad k=1,2,$$
$$\operatorname{var}(\bar{T}_N^{(5)}) = \frac{4}{N^2h}\int (2K-K*K)^2(u)\,\mathrm{d}u\int_0^1 \frac{\sigma_1^2(x)\sigma_2^2(x)\kappa_1 r_1(x)\kappa_2 r_2(x)}{R^2(x)}\mathrm{d}x + o\left(\frac{1}{N^2h}\right).$$

Proof. For simplicity, we only consider $T_N^{(5)}$, the other two terms are treated similarly. By the definition of the weights in (37), the coefficients γ_{ij} can be rewritten as $\gamma_{ij} = \tilde{\gamma}_{ij} + \tilde{\tilde{\gamma}}_{ij}$, where

$$\tilde{\gamma}_{ij} = \left\{\frac{2}{N^2h^2}\sum_{l=1}^{2}\sum_{k=1}^{n_l} K\left(\frac{t_{1i}-t_{lk}}{h}\right)K\left(\frac{t_{2j}-t_{lk}}{h}\right)\frac{1}{\hat{R}^2(t_{lk})}\sigma_{3-l}^2(t_{lk})\right.$$
$$\left. - \frac{2}{Nh}K\left(\frac{t_{2j}-t_{1i}}{h}\right)\frac{1}{\hat{R}(t_{1i})} - \frac{2}{Nh}K\left(\frac{t_{2j}-t_{1i}}{h}\right)\frac{1}{\hat{R}(t_{2j})}\right\}\sigma_1(t_{1i})\sigma_2(t_{2j})$$
$$\tilde{\tilde{\gamma}}_{ij} = \frac{2}{N^2h^2}\sum_{l=1}^{2}\sum_{k=1}^{n_l} K\left(\frac{t_{1i}-t_{lk}}{h}\right)K\left(\frac{t_{2j}-t_{lk}}{h}\right)\frac{1}{\hat{R}^2(t_{lk})}\sigma_1(t_{1i})\sigma_2(t_{2j})\left(\hat{\sigma}_{3-l}^2(t_{lk}) - \sigma_{3-l}^2(t_{lk})\right).$$

First, we consider the term $\tilde{T}_N^{(5)}$ defined as $T_N^{(5)}$, but with γ_{ij} replaced by $\tilde{\gamma}_{ij}$. Using the same argument as in the proof of Lemma 3, we find that asymptotically the estimator $\hat{R}(t)$ can be replaced by the true $R(t)$ with a remainder term that is negligible in probability to calculate the variance of the dominating term. We then obtain $\tilde{T}_N^{(5)} = \bar{T}_N^{(5)} + o_p(1/(N\sqrt{h}))$ with

$$\operatorname{var}(\bar{T}_N^{(5)}) = \frac{1}{N^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left\{\frac{2}{N^2h^2}\sum_{l=1}^{2}\sum_{k=1}^{n_l} K\left(\frac{t_{1i}-t_{lk}}{h}\right)K\left(\frac{t_{2j}-t_{lk}}{h}\right)\frac{1}{R^2(t_{lk})}\sigma_{3-l}^2(t_{lk})\right.$$
$$\left. - \frac{2}{Nh}K\left(\frac{t_{2j}-t_{1i}}{h}\right)\frac{1}{R(t_{1i})} - \frac{2}{Nh}K\left(\frac{t_{2j}-t_{1i}}{h}\right)\frac{1}{R(t_{2j})}\right\}^2 \sigma_1^2(t_{1i})\sigma_2^2(t_{2j})$$
$$= \frac{4}{N^2h}\int (2K-K*K)^2(u)\,\mathrm{d}u\int_0^1 \frac{\sigma_1^2(x)\sigma_2^2(x)\kappa_1 r_1(x)\kappa_2 r_2(x)}{R^2(x)}\mathrm{d}x + o\left(\frac{1}{N^2h}\right).$$

Finally, the asymptotic negligibility of the second term (defined as $T_N^{(5)}$, but with γ_{ij} replaced by $\tilde{\tilde{\gamma}}_{ij}$) can be shown by some tedious calculations of expectations and variances. This completes the proof of Lemma 4.

With similar calculations as in the proof of Lemma 4, we can rewrite $\bar{T}_N^{(5)}$ as

$$\bar{T}_N^{(5)} = \frac{2}{N^3}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\varepsilon_{1i}\varepsilon_{2j}\sigma_1(t_{1i})\sigma_2(t_{2j})\left\{\frac{1}{h^2}\int K\left(\frac{t_{1i}-z}{h}\right)K\left(\frac{t_{2j}-z}{h}\right)\frac{1}{R(z)}\mathrm{d}z\right.$$
$$\left. - \frac{1}{h}K\left(\frac{t_{2j}-t_{1i}}{h}\right)\frac{1}{R(t_{1i})} - \frac{1}{h}K\left(\frac{t_{2j}-t_{1i}}{h}\right)\frac{1}{R(t_{2j})}\right\} + o_p\left(\frac{1}{N\sqrt{h}}\right).$$

Applying the same arguments to the terms $\bar{T}_N^{(3)}$ and $\bar{T}_N^{(4)}$ we obtain

$$N\sqrt{h}(T_N - E[T_N]) = N\sqrt{h}\left(\bar{T}_N^{(3)} + \bar{T}_N^{(4)} + \bar{T}_N^{(5)}\right) + o_p(1),$$

where the dominating part can be written as a quadratic form $W_N = \varepsilon_N^{\mathrm{T}} A_N \varepsilon_N$ of the random variable $\varepsilon_N = (\varepsilon_{11}, \dots, \varepsilon_{1n_1}, \varepsilon_{21}, \dots, \varepsilon_{2n_2})^{\mathrm{T}}$ with a symmetric matrix A_N with vanishing diagonal elements. From Lemma 4, we obtain for the asymptotic variance $\mathrm{var}(W_N) = \tau^2 + o(1)$. Asymptotic normality of W_N can be proved by an application of Theorem 5.2 of de Jong (1987) and this gives the conclusion of Theorem 1.

A.2. Auxiliary result

Proposition 1

Assume model (1) where the ε_{ij} are independent centred random variables with variance 1, such that assumptions (9)–(14) hold. For the heteroscedastic variance estimators defined in (19), we obtain the expansion ($i = 1, 2$)

$$\hat{\sigma}_i^2(t) - \sigma_i^2(t) = \sum_{k=1}^{6} S_{n_i}^{(k)}(t)$$

where

$$S_{n_i}^{(1)}(t) = \frac{1}{n_i h} \frac{1}{\hat{r}_i(t)} \sum_{l=1}^{n_i} K\left(\frac{t - t_{il}}{h}\right) \sigma_i^2(t_{il})(\varepsilon_{il}^2 - 1) = O_p\left(\frac{1}{\sqrt{n_i h}}\right)$$

$$S_{n_i}^{(2)}(t) = \frac{1}{n_i h} \frac{1}{\hat{r}_i(t)} \sum_{l=1}^{n_i} K\left(\frac{t - t_{il}}{h}\right) (\sigma_i^2(t_{il}) - \sigma_i^2(t)) = O(h^d) + O\left(\frac{1}{n_i h}\right)$$

$$S_{n_i}^{(3)}(t) = \frac{2}{n_i h} \frac{1}{\hat{r}_i(t)} \sum_{l=1}^{n_i} K\left(\frac{t - t_{il}}{h}\right) \sigma_i(t_{il}) \varepsilon_{il} \left(\frac{1}{n_i h} \sum_{k=1}^{n_i} K\left(\frac{t_{il} - t_{ik}}{h}\right) \frac{f_i(t_{il}) - f_i(t_{ik})}{\hat{r}_i(t_{il})}\right)$$
$$= O_p\left(\frac{h^d}{\sqrt{n_i h}}\right) + O_p\left(\frac{1}{(n_i h)^{3/2}}\right)$$

$$S_{n_i}^{(4)}(t) = -\frac{2}{n_i h} \frac{1}{\hat{r}_i(t)} \sum_{l=1}^{n_i} K\left(\frac{t - t_{il}}{h}\right) \sigma_i(t_{il}) \varepsilon_{il} \left(\frac{1}{n_i h} \sum_{\substack{k=1 \\ k \neq l}}^{n_i} K\left(\frac{t_{il} - t_{ik}}{h}\right) \frac{\sigma_i(t_{ik}) \varepsilon_{ik}}{\hat{r}_i(t_{il})}\right) = O_p\left(\frac{1}{n_i h}\right)$$

$$S_{n_i}^{(5)}(t) = -\frac{2}{(n_i h)^2} \frac{1}{\hat{r}_i(t)} \sum_{l=1}^{n_i} K\left(\frac{t - t_{il}}{h}\right) \frac{K(0)}{\hat{r}_i(t_{il})} \sigma_i^2(t_{il}) \varepsilon_{il}^2 = O_p\left(\frac{1}{n_i h}\right)$$

$$S_{n_i}^{(6)}(t) = \frac{1}{n_i h} \frac{1}{\hat{r}_i(t)} \sum_{l=1}^{n_i} K\left(\frac{t - t_{il}}{h}\right) (f_i(t_{il}) - \hat{f}_i(t_{il}))^2 = O_p\left(\frac{1}{n_i h}\right) + O_p(h^{2d}).$$

The dominating part in the expansion is $S_{n_i}^{(1)}(t)$ and $S_{n_i}^{(2)}(t)$ is deterministic. For the expectations we have $E[S_{n_i}^{(1)}(t)] = E[S_{n_i}^{(3)}(t)] = E[S_{n_i}^{(4)}(t)] = 0$, $E[S_{n_i}^{(5)}(t)] = O(1/(n_i h))$ and $E|S_{n_i}^{(6)}(t)| = O(1/(n_i h)) + O(h^{2d})$.