# Nonparametric comparison of regression curves — an empirical process approach


Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: holger.dette@ruhr-uni-bochum.de
FAX: +49 2 34 70 94 559

Natalie Neumeyer
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: natalie.neumeyer@ruhr-uni-bochum.de


November 9, 2000


**Abstract**

We propose a new test for the comparison of two regression curves, which is based on a difference of two marked empirical processes based on residuals. The large sample behaviour of the corresponding statistic is studied to provide a full nonparametric comparison of regression curves. In contrast to most procedures suggested in the literature the new procedure is applicable in the case of different design points and heteroscedasticity. Moreover, it is demonstrated that the proposed test detects continuous alternatives converging to the null at a rate $N^{-1/2}$. In the case of equal design points the fundamental statistic reduces to a test statistic proposed by Delgado (1993) and therefore resembles in spirit classical goodness-of-fit tests. As a by-product we explain the problems of a related test proposed by Kulasekera (1995) and Kulasekera and Wang (1997) with respect to accuracy in the approximation of the level. These difficulties mainly originate from the comparison with the quantiles of an inappropriate limit distribution.

A simulation study is conducted to investigate the finite sample properties of a wild bootstrap version of the new tests.



AMS Classification: Primary 62G05, Secondary 60F15, 60F17
Keywords and Phrases: comparison of regression curves, goodness-of-fit, marked empirical process, VC-classes, U-processes


## 1 Introduction

The comparison of two regression curves is a fundamental problem in applied regression analysis. In many cases of practical interest (after rescaling the covariable into the unit interval) we end up with a sample of $N = n_1 + n_2$ observations

$$Y_{ij} = f_i(X_{ij}) + \sigma_i(X_{ij})\varepsilon_{ij}\,, \qquad j = 1\,, \ldots, n_i; \quad i = 1\,, 2, \tag{1.1}$$

where $X_{ij}$ ($j = 1\,, \ldots, n_i$) are independent observations with positive density $r_i$ on the interval $[0,1]$ ($i = 1\,, 2$) and $\varepsilon_{ij}$ are independent identically distributed random variables with mean 0 and variance 1. In equation (1.1) $f_i$ and $\sigma_i$ denote the regression and variance function in the $i$-th sample ($i = 1\,, 2$). In this paper we are interested in the problem of testing the equality of the mean functions, i.e.

$$H_0 : f_1 = f_2 \quad \text{versus} \quad H_1 : f_1 \neq f_2. \tag{1.2}$$

Much effort has been devoted to this problem in the recent literature [see e.g. Härdle and Marron (1990), King, Hart and Wehrly (1991), Hall and Hart (1990), Delgado (1993), Young and Bowman (1995), Hall, Huber and Speckman (1997), Dette and Munk (1998) or Dette and Neumeyer (1999)]. Most authors concentrate on equal design points and a homoscedastic error [see e.g. Härdle and Marron (1990), Hall and Hart (1990), King, Hart and Wehrly (1991), Delgado (1993)]. Kulasekera (1995) and Kulasekera and Wang (1997) proposed a test for the hypothesis (1.2) which is applicable under the assumption of different designs in both groups, but requires homoscedasticity in the individual groups. In principle this test can detect alternatives which converge to the null at a rate $N^{-1/2}$ (here $N = n_1 + n_2$ denotes the total sample size), but in the same papers these authors mention some practical problems with the performance of their procedure, especially with respect to the accuracy of the approximation of the nominal level.
To our knowledge the problem of testing the equality of two regression curves in the general heteroscedastic model (1.1) with unequal design points was firstly considered by Dette and Munk (1998) who considered the fixed design and proposed a consistent test which can detect alternatives converging to the null at a rate $N^{-1/4}$ under very mild conditions for the regression and variance function (i.e. differentiability is not required). Recently Dette and Neumeyer (1999) proposed several tests for the hypothesis (1.2) which are based on kernel smoothing methods and applicable in the general model (1.1). These methods can detect alternatives converging to the null at a rate $(N\sqrt{h})^{-1/2}$, where $h$ is a bandwidth (converging to 0) required for the estimation of nonparametric residuals.
It is the purpose of the present paper to suggest a new test for the equality of the two regression curves $f_1$ and $f_2$ which can detect alternatives converging to the null at a rate $N^{-1/2}$ and is applicable in the general model (1.2) with unequal design points and heteroscedastic errors. The test statistic is based on a difference of two marked empirical processes based on residuals obtained under the assumption of equal regression curves. We prove weak convergence of the underlying empirical process to a Gaussian process generalizing recent results on $U$-processes of Nolan and Pollard (1987, 1988) to two-sample $U$-statistics. The asymptotic null distribution of the test statistic depends on certain features of the data and the finite sample performance of a wild bootstrap version is investigated by means of a simulation study.
We finally note that marked empirical processes have already been applied by Delgado (1993) and Kulasekera (1995) and Kulasekera and Wang (1997) for testing the equality of two regression functions. However, Delgado's (1993) approach sensitively relies on the assumption of equal design points and homoscedastic errors because the marked empirical process is based on the differences of the observations at the joint design points. The method proposed in this paper

uses two marked empirical processes of the residuals for both samples, where the residuals are obtained from a nonparametric estimate of the (under $H_0$) joint regression function from the total sample. Moreover, in the case of equal design points the basic statistic considered here essentially reduces to the test statistic considered by Delgado (1993). On the other hand the methods proposed by Kulasekera (1995) and Kulasekera and Wang (1997) require a homoscedastic error distribution. Moreover, these authors mention some practical problems because the performance of their procedure depends sensitively on the chosen smoothing parameters for the estimation of the regression curves and larger noises yield levels substantially different from the nominal level. As a by-product of this paper we will prove that the problem with the accuracy of the approximation of the nominal level is partially caused by a substantial mistake in the proof of Theorem 2.1 and 2.2 in Kulasekera (1995), because this author ignores the variablitiy caused by the nonparametric estimation of the regression function in the application of Donsker's invariance principle.
The present paper is organized as follows. Section 2 introduces the marked empirical processes, the corresponding test statistics and gives their asymptotic behaviour. Some comments regarding the test of Kulasekera (1995) and a clarification of its asymptotic properties are given in Section 3. The finite sample behaviour of a wild bootstrap version of the discussed procedures is studied in Section 4 which also gives a result regarding the consistency of a wild bootstrap. Finally, all proofs are deferred to the appendix.

# 2 A marked empirical process and its weak convergence

Recall the formulation of the general two sample problem (1.1). We assume that the explanatory variables $X_{ij}$ $(j = 1, \dots, n_i)$ are i.i.d. with positive density $r_i$ on the interval $[0, 1]$ $(i = 1, 2)$. The regression functions $f_1, f_2$ and the densities $r_1, r_2$ are supposed to be $r$ $(\geq 2)$ times continuously differentiable, i.e.

(2.1) $$r_i, f_i \in C^r([0, 1]), \quad i = 1, 2.$$

Throughout this paper let

(2.2) $$\hat{r}(x) = \frac{1}{Nh} \sum_{i=1}^{2} \sum_{j=1}^{n_i} K\Big(\frac{x - X_{ij}}{h}\Big)$$

denote the density estimator from the combined sample $X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2}$ where $h$ denotes a bandwidth satisfying

(2.3) $$h \to 0, \quad Nh^{2r} \to 0, \quad h^r \log N \to 0, \quad Nh^2 \to \infty$$

and $K$ is a symmetric kernel with compact support of order $r \geq 2$, i.e.

(2.4) $$\frac{(-1)^j}{j!} \int K(u)u^j\, du = \begin{cases} 1 & : \ j = 0 \\ 0 & : \ 1 \leq j \leq r - 1 \\ k_r \neq 0 & : \ j = r \end{cases}$$

[see Gasser, Müller and Mamitzsch (1985)]. We assume that there exists a decomposition of the nonnegative axis of the form

$$[0,\infty) = \bigcup_{j=1}^{m} [a_{j-1}, a_j)$$

($0 = a_0 < a_1 < \ldots < a_{m-1} < a_m = \infty$) such that for some $\varepsilon \in \{-1, 1\}$ the function $\varepsilon K$ is increasing on the interval $[a_{2j}, a_{2j+1})$ and decreasing on the interval $[a_{2j+1}, a_{2j+2})$. A straightforward argument shows that

$$\hat{r}(x) \xrightarrow{P} r(x) := \kappa_1 r_1(x) + \kappa_2 r_2(x) \tag{2.5}$$

as $N \to \infty$ ,provided that sizes of the individual samples satisfy

$$\frac{n_i}{N} = \kappa_i + O(\frac{1}{N}), \quad i = 1, 2, \tag{2.6}$$

where $\kappa_i \in (0,1), i = 1, 2$. The Nadaraya-Watson estimator of the regression function [see Nadaraya (1964) or Watson (1964)] from the combined sample is defined by

$$\hat{f}(x) \; = \; \frac{1}{Nh} \sum_{i=1}^{2} \sum_{j=1}^{n_i} K\Big(\frac{x - X_{ij}}{h}\Big) Y_{ij} \, \frac{1}{\hat{r}(x)} \tag{2.7}$$

and consistently estimates

$$f(x) := \frac{\kappa_1 r_1(x) f_1(x) + \kappa_2 r_2(x) f_2(x)}{r(x)}.$$

Note that under the null hypothesis of equal regression curves we have $f_1 = f_2 = f$. For $i = 1, 2$ we define residuals

$$e_{ij} \; = \; (Y_{ij} - \hat{f}(X_{ij}))\hat{r}(X_{ij}) \tag{2.8}$$

$$f_{ij} \; = \; Y_{ij} - \hat{f}(X_{ij}) \tag{2.9}$$

and consider the marked empirical processes

$$\hat{R}_N^{(1)}(t) \; = \; \frac{1}{N} \sum_{j=1}^{n_1} e_{1j} I\{X_{1j} \leq t\} - \frac{1}{N} \sum_{j=1}^{n_2} e_{2j} I\{X_{2j} \leq t\} \tag{2.10}$$

$$\hat{R}_N^{(2)}(t) \; = \; \frac{1}{N} \sum_{j=1}^{n_1} f_{1j} I\{X_{1j} \leq t\} - \frac{1}{N} \sum_{j=1}^{n_2} f_{2j} I\{X_{2j} \leq t\} \tag{2.11}$$

where $t \in [0,1]$ and $I\{\cdot\}$ denotes the indicator function. The multiplication of the residuals (2.9) with the density estimator $\hat{r}(x)$ yields the residuals (2.8) and as a consequence a simpler asymptotic analysis of the process $\hat{R}_N^{(1)}$ [see the following Proposition 2.1 and Theorem 2.2]. On the other hand the form of $\hat{R}_N^{(2)}$ is attractive because it reduces for equal design points (i.e. $n_1 = n_2$, $X_{1j} = X_{2j}$, $j = 1, \dots, n_1$) to the process considered by Delgado (1993). The following proposition indicates that the marked empirical processes defined in (2.10) and (2.11) are useful for testing the hypothesis (1.2) of equal regression curves. The proof is given in the appendix.

**Proposition 2.1.** *Assume that (2.1), (2.3), (2.4) and (2.6) are satisfied, then*

$$E\left[\hat{R}_N^{(1)}(t)\right] = 2\kappa_1\kappa_2 \int_0^t (f_1(x) - f_2(x)) r_1(x) r_2(x)\, dx + O(h^r)$$

$$E\left[\hat{R}_N^{(2)}(t)\right] = 2\kappa_1\kappa_2 \int_0^t (f_1(x) - f_2(x)) \frac{r_1(x) r_2(x)}{r(x)}\, dx + O(h^r).$$

Note that

$$\int_0^t (f_1(x) - f_2(x)) r_1(x) r_2(x)\, dx = 0 \quad \forall\, t \in [0,1]$$

if and only if the hypothesis (1.2) is valid. Consequently, a test for the hypothesis of equal regression curves could be based on real valued functionals of the processes (2.10) and (2.11) such as $(i = 1, 2)$

$$\int_0^1 \hat{R}_N^{(i)^2}(t)dt, \qquad \sup_{t\in[0,1]} |\hat{R}_N^{(i)}(t)|.$$

The asymptotic distribution of these statistics can be obtained by the continuous mapping theorem [see e.g. Pollard (1984)] and the following result which establishes weak convergence of the processes $\hat{R}_N^{(1)}$ and $\hat{R}_N^{(2)}$ in the Skorokhod space $D[0,1]$.

**Theorem 2.2.** *Assume that (2.1), (2.3), (2.4) and (2.6) are satisfied, then under the null hypothesis of equal regression curves the marked empirical process $\sqrt{N}\hat{R}_N^{(1)}$ defined by (2.10) converges weakly to a centered Gaussian process $Z^{(1)}$ in the space $D[0,1]$ with covariance function*

$$H^{(1)}(s,t) = 4 \int_0^{s\wedge t} (\sigma_1^2(x)\kappa_2 r_2(x) + \sigma_2^2(x)\kappa_1 r_1(x))\kappa_1 r_1(x)\kappa_2 r_2(x)\, dx. \tag{2.12}$$

*Similarly, the process $\sqrt{N}\hat{R}_N^{(2)}$ defined by (2.11) converges weakly to a centered Gaussian process $Z^{(2)}$ in the space $D[0,1]$ with covariance function*

$$H^{(2)}(s,t) = 4 \int_0^{s\wedge t} (\sigma_1^2(x)\kappa_2 r_2(x) + \sigma_2^2(x)\kappa_1 r_1(x)) \frac{\kappa_1 r_1(x)\kappa_2 r_2(x)}{r^2(x)}\, dx. \tag{2.13}$$

**Remark 2.3.** It is worthwhile to mention that the statement of Theorem 2.2 does not depend on the specific smoothing procedure used in the construction of the processes. For example, a local polynomial estimator [see Fan (1992) or Fan and Gijbels (1996)] can be treated similarly but with a substantial increase of the mathematical complexity. Note that local polynomial estimators have various practical and theoretical advantages such as a better boundary behaviour and they require weaker differentiability assumptions on the design densities. We used the Nadaraya-Watson estimator because for this type of estimator the proof of the VC-property for certain classes of functions is much simpler compared to local polynomial estimators [see, for example, the proof of Lemma 5.2a]. Nevertheless Theorem 2.2 remains valid for local linear (or even higher order) polynomial estimators and we used local linear smoothers in the simulation study presented in Section 4.

**Remark 2.4.** The tests obtained from the continuous mapping theorem and Theorem 2.2 are consistent against local alternatives converging to the null hypothesis at a rate $1/\sqrt{N}$. This follows by a careful inspection of the proof of Theorem 2.2, which shows that for local alternatives of the form $f_1(\cdot)-f_2(\cdot)=\Delta(\cdot)/\sqrt{N}$ the marked empirical processes $\sqrt{N}\hat{R}_N^{(i)}(\cdot)$ ($i=1,2$) converge weakly to Gaussian processes with respective covariance kernels $H^{(i)}(\cdot,\cdot)$ given in Theorem 2.2 and mean

$$\begin{aligned}\mu^{(1)}(t) &= 2\kappa_1\kappa_2\int_0^t \Delta(x)r_1(x)r_2(x)dx \\ \mu^{(2)}(t) &= 2\kappa_1\kappa_2\int_0^t \Delta(x)\frac{r_1(x)r_2(x)}{r(x)}dx,\end{aligned}$$

respectively. These results can be used for an asymptotic comparison of tests based on $\hat{R}_N^{(1)}$ and $\hat{R}_N^{(2)}$, which in general depends on the particular alternative under consideration. For example, if $\Delta(\cdot)\equiv 1$ and $\sigma_1^2(\cdot)=\sigma_2^2(\cdot)\equiv\sigma^2$ we have

$$\frac{[\mu^{(1)}(t)]^2}{H^{(1)}(t,t)} = \frac{\kappa_1\kappa_2[\int_0^t r_1(x)r_2(x)dx]^2}{\sigma^2\int_0^t r_1(x)r_2(x)r(x)dx} \le \frac{\kappa_1\kappa_2}{\sigma^2}\int_0^t\frac{r_1(x)r_2(x)}{r(x)}dx = \frac{[\mu^{(2)}(t)]^2}{H^{(2)}(t,t)},$$

where the inequality follows by the Cauchy-Schwarz inequality. This indicates a better performance of statistics based on the process $\hat{R}_N^{(2)}$ for smooth one-sided local alternatives.

**Remark 2.5.** The results can easily be extended to the comparison of $k$ regression curves in the model

$$Y_{ij}=f_i(X_{ij})+\sigma_i(X_{ij})\varepsilon_{ij},\quad j=1,\dots,n_i;\ i=1,\dots,k.$$

For a generalization of the statistic $\hat{R}_N^{(1)}$ consider the residuals

$$e_{j\ell}^{(i)}=(Y_{j\ell}-\hat{f}^{(i)}(X_{j\ell}))\hat{r}^{(i)}(X_{j\ell}),\quad i=1,\dots,k-1$$

($j\in\{i,i+1\},\ell\in\{1,\dots,n_j\}$) where $\hat{f}^{(i)}$ and $\hat{r}^{(i)}$ denote the Nadaraya-Watson and the density estimator from the combined $i$th and $(i+1)$th sample. If $N=\sum_{i=1}^k n_i$ denotes the total sample size,

$$\frac{n_i}{N}=\kappa_i+O(\frac{1}{N})$$

$(\kappa_i \in (0,1);\ i = 1, \ldots, k)$ and

$$\hat{R}^{(1)}_{Ni} = \frac{1}{N}\sum_{\ell=1}^{n_i} e^{(i)}_{i\ell} I\{X_{i\ell} \le t\} - \frac{1}{N}\sum_{\ell=1}^{n_{i+1}} e^{(i)}_{i+1,\ell} I\{X_{i+1,\ell} \le t\} \quad (i = 1, \ldots, k-1),$$

then it follows that $\hat{R}^{(1)}_N(t) := (\hat{R}^{(1)}_{N1}(t), \ldots, \hat{R}^{(1)}_{Nk-1}(t))^T$ converges weakly to a $(k-1)$-dimensional Gaussian process $(Z^{(1)}_1, \ldots, Z^{(1)}_{k-1})^T$ with covariance structure

$$\mathrm{Cov}(Z^{(1)}_i(t), Z^{(1)}_j(s)) = k_{ij}(s \wedge t)$$

where $k_{ij} = k_{ji}$, $(j \le i)$ and

$$k_{ij}(u) = \begin{cases} 4\int_0^u (\sigma_i^2(x)\kappa_{i+1}r_{i+1}(x) + \sigma_{i+1}^2(x)\kappa_i r_i(x))\kappa_i\kappa_{i+1}r_i(x)r_{i+1}(x)dx & \text{if } j = i \\ \\ -4\int_0^u \sigma_j^2(x)\kappa_{j-1}\kappa_j\kappa_{j+1}r_{j-1}(x)r_j(x)r_{j+1}(x)dx & \text{if } j = i+1 \\ \\ 0 & \text{if } j > i+1\,. \end{cases}$$

# 3 Some remarks on related tests

As pointed out in the introduction the application of empirical processes has already been proposed by several authors. Among many others we refer to An and Bing (1991), Stute (1997), who considered the problem of testing for a parametric form of the regression and to the recent work of Delgado and González-Manteiga (1998), who used this approach in the construction of a test for selecting variables in a nonparametric regression. In the context of comparing regression curves empirical processes were already applied by Delgado (1993) and Kulasekera (1995), Kulasekera and Wang (1997) and recently in an unpublished report by Cabus (2000). Delgado considered equal design points (i.e. $n_1 = n_2; X_{1i} = X_{2i}$ and a homoscedastic error distribution) and the process $R^{(2)}_N$ reduces in this case to the process introduced by Delgado (1993). Kulasekera (1995) and Kulasekera and Wang (1997) discussed the case of not necessarily equal design points and homoscedastic (but potentially different) errors in both samples. In this case these authors proposed a test also based on a marked empirical process and investigated its finite sample performance by means of a simulation study. In the same papers Kulasekera (1995) and Kulasekera and Wang (1997) mention some difficulties with respect to the practical performance of their procedure. They observed levels substantially different from the nominal levels in their study and explained these observations by the sensitive dependency on the bandwidth. We will demonstrate in this section that these deficiencies are partially caused by the use of incorrect (asymptotic) critial values.
To be precise consider the model (1.1) in the case of a fixed design $X_{ij} = t_{ij}$ $(j = 1, \ldots, n_i;\ i = 1, 2)$ satisfying a Sacks and Ylvisacker (1970) condition

$$\int_0^{t_{ij}} r_i(t)dt = \frac{j}{n_i}; \quad j = 1, \ldots, n_i, \quad i = 1, 2\,, \tag{3.1}$$

let $\hat{f}_i$ denote the Nadaraya-Watson estimator from the ith sample ($i = 1, 2$) using bandwidth $h_i$ ($i = 1, 2$) and define residuals by

$$\begin{aligned}
\tilde{e}_{1i} &= Y_{1i} - \hat{f}_2(t_{1i}), \quad i = 1, \ldots, n_1 \\
\tilde{e}_{2j} &= Y_{2j} - \hat{f}_1(t_{2j}), \quad j = 1, \ldots, n_2.
\end{aligned}$$

The corresponding partial sums are given by

$$\mu_i(t) = \sum_{j=1}^{\lfloor n_i t \rfloor} \frac{\tilde{e}_{ij}}{\sqrt{n_i}}, \quad 0 < t < 1; \quad i = 1, 2, \tag{3.2}$$

and the following result specifies the asymptotic distribution of these marked empirical processes.

**Theorem 3.1.** *If the assumptions (2.1), (2.3), (2.4), (2.6) and (3.1) are satisfied, then under the null hypothesis of equal regression curves the marked empirical process $\mu_1$ defined in (3.2) converges weakly to a centered Gaussian process with covariance function*

$$m_{12}(s, t) = \int_0^{R_1^{-1}(s \wedge t)} (\sigma_1^2(x)\kappa_2 r_2(x) + \sigma_2^2(x)\kappa_1 r_1(x)) \frac{r_1(x)}{\kappa_2 r_2(x)} dx \tag{3.3}$$

*where $R_1(t) = \int_0^t r_1(x)\,dx$ denotes the cumulative distribution function corresponding to the design density $r_1$.*
*Similarly, the process $\mu_2$ converges weakly to a centered Gaussian process with covariance function $m_{21}(s, t)$.*

Note that Kulasekera (1995) considered a homoscedastic error and claimed in his proof of Theorem 2.1 [Kulasekera (1995)] weak convergence of $\mu_i$ to a centered Gaussian process with covariance function $\tilde{m}_i(s, t) = \sigma_i^2 \cdot (s \wedge t)$, which is usually different from $m_{i,3-i}(s, t)$ [an exception is the case of the uniform design and equal homoscedastic variances in both groups]. For these reasons some care is necessary if the test of Kualsekera is applied. We finally remark that Kulasekera (1995) and Kulasekera and Wang (1997) discussed several related tests and similar comments apply to these procedures.
In the case of a random design the processes (3.2) have to be modified because in this case the observations are not necessarily ordered. A minor modification given by

$$\lambda_N^{(i)}(t) = \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} (Y_{ij} - \hat{f}_{3-i}(X_{ij})) I\{X_{ij} \le t\}, \quad i = 1, 2, \tag{3.4}$$

could be considered, which yields a slightly simpler covariance structure of the Gaussian process.

**Theorem 3.2.** *If the assumptions (2.1), (2.3), (2.4) and (2.6) are satisfied, then under the null hypothesis of equal regression curves the marked empirical process $\lambda_N^{(1)}$ defined by (3.4) converges weakly to a centered Gaussian process with covariance function $m_{12}(R_1(s), R_1(t))$ where $m_{12}$ is* *defined in (3.3) and $R_1$ denotes the distribution function of $X_{1j}$. Similarly, the process $\lambda_N^{(2)}$ converges weakly to a centered Gaussian process with covariance function $m_{21}(R_2(s), R_2(t))$, where $m_{21}(s,t) = m_{12}(t,s)$ and $R_2$ is the distribution function of $X_{2j}$.*

A rather different method to the problem of comparing regression curves was recently proposed by Cabus (2000), who considered the $U$-process

$$U_N(t) = \frac{1}{n_1 n_2 h} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (Y_{1i} - Y_{2j}) K\Big(\frac{X_{1i} - X_{2j}}{2}\Big) I\{X_{1i} \le t, X_{2j} \le t\}. \tag{3.5}$$

Note that this approach is similar to a method introduced by Zheng (1996) in the context of testing for the functional form of a regression. Cabus (2000) proved weak convergence of the process $\sqrt{N} U_N$ to a centered Gaussian process with covariance function $\frac{1}{4(\kappa_1\kappa_2)^2} H^{(1)}(s,t)$ defined in (2.12). It also follows from Cabus (2000) that the asymptotic behaviour with respect to local alternatives is exactly the same as for the process $\hat{R}_N^{(1)}$ [see Remark 2.4].

# 4 Wild bootstrap and finite sample properties

Throughout this section we will study the finite sample properties of a test based on the Kolmogorov Smirnov distance

$$K_N^{(i)} := \sup_{t \in [0,1]} |\hat{R}_N^{(i)}(t)|, \quad i = 1, 2, \tag{4.1}$$

which rejects the hypothesis of equal regression curves for large values of $K_N^{(i)}$. In principle critical values can be obtained from Theorem 2.2 and the continuous mapping theorem. However, it is well known [see e.g. Hjellvik and Tjøstheim (1995), Hall and Hart (1990)] that in similar problems of specification testing the rate of convergence of the distribution of the test statistic is usually rather slow. Additionally the asymptotic distributions of the Gaussian processes obtained in Theorem 2.2, 3.1 and 3.2 usually depend on certain features of the data generating process and cannot be directly implemented in practice. For this reason we propose in this section the application of a resampling procedure based on the wild bootstrap [see e.g. Wu (1986)] and prove its consistency [see Theorem 4.1 below]. The finite sample properties of the resulting tests are then investigated by means of a simulation study. To be precise let $\hat{f}_g(x)$ denote the Nadaraya-Watson estimator of the regression function from the total sample defined in (2.7) using the bandwidth $g > 0$, where this dependency has now been made explicit in our notation. Define nonparametric residuals by

$$\hat{\varepsilon}_{ij} := Y_{ij} - \widehat{f}_g(X_{ij}) \quad (j = 1, \dots, n_i,\ i = 1, 2) \tag{4.2}$$

and bootstrap residuals by

$$\varepsilon_{ij}^* := \hat{\varepsilon}_{ij} V_{ij} \tag{4.3}$$

where $V_{11}, V_{12}, \dots, V_{1n_1}, V_{21}, \dots, V_{2n_2}$ are bounded i.i.d. zero mean random variables which are independent from the total sample

$$\mathcal{Y}_N := \Big\{ X_{ij}, Y_{ij} \mid i = 1, 2,\ j = 1, \dots, n_i \Big\}. \tag{4.4}$$

We obtain the bootstrap sample

$$Y_{ij}^* := \widehat{f}_g(X_{ij}) + \varepsilon_{ij}^* \tag{4.5}$$

and the corresponding marked empirical processes

$$\hat{R}_N^{(1)*}(t) = \frac{1}{N}\sum_{\ell=1}^{2}\sum_{j=1}^{n_\ell}(Y_{\ell j}^* - \hat{f}_h^*(X_{\ell j}))\hat{r}_h(X_{\ell j})I\{X_{\ell j} \le t\}$$

$$\hat{R}_N^{(2)*}(t) = \frac{1}{N}\sum_{\ell=1}^{2}\sum_{j=1}^{n_\ell}(Y_{\ell j}^* - \hat{f}_h^*(X_{\ell j}))I\{X_{\ell j} \le t\}$$

where throughout this section the index $*$ means that the process has been calculated from the bootstrap sample (4.5). Note that we use the bandwidth $h$ for the calculation of the test statistic (which is indicated by the extra index in $\hat{f}_h^*$ and $\hat{r}_h$) and a bandwidth $g$ for the calculation of the residuals. Let $K_N^{(i)*}$ $(i = 1, 2)$ denote the statistic in (4.1) obtained from the bootstrap sample, then the hypothesis of equal regression curves is rejected if $K_N^{(i)} \ge k_{N,1-\alpha}^{*(i)}$, where $k_{N,1-\alpha}^{*(i)}$ denotes the critical value obtained from the bootstrap distribution i.e.

$$\mathbb{P}(K_N^{(i)*} \ge k_{N,1-\alpha}^* \mid \mathcal{Y}_N) = \alpha, \qquad i = 1, 2.$$

The consistency of this procedure follows from the continuous mapping theorem and the following result, which establishes asymptotic equivalence (in the sense of weak convergence) of the processes $\sqrt{N}\hat{R}_N^{(i)}$ and $\sqrt{N}\hat{R}_N^{(i)*}$ in probability conditionally on the sample $\mathcal{Y}_N$.

**Theorem 4.1.** *If the assumptions of Theorem 2.2 and the bandwidth conditions*

$$g \to 0, \quad \sqrt{N}gh \to \infty, \quad Ng^{2r} \to 0, \quad g^r \log N \to 0, \quad h^r = O(\sqrt{g}) \tag{4.6}$$

*are satisfied, then the bootstrapped marked empirical process $R_N^{(i)*}$ converges under the null hypothesis of equal regression curves weakly to the centered Gaussian process $Z^{(i)} (i = 1, 2)$ of Theorem 2.2 in probability conditionally on the sample $\mathcal{Y}_N$.*

For the sake of comparison we will also discuss tests based on the approach proposed by Kulasekera (1995) and Cabus (2000). More precisely, we use the generalization of Kulasekera's approach to the random design case and reject the hypothesis of equal regression curves for large values of the statistic

$$L_N = \max\{\sup_{t\in[0,1]} |\lambda_N^{(1)}(t)|, \sup_{t\in[0,1]} |\lambda_N^{(2)}(t)|\} \tag{4.7}$$

where the processes $\lambda_N^{(1)}(\cdot)$ and $\lambda_N^{(2)}(\cdot)$ have been defined in (3.4). Similarly, we consider the statistic

$$C_N = \sup_{t\in[0,1]} |U_N(t)| \tag{4.8}$$

where $U_N$ is the process introduced by Cabus (2000) and defined by (3.5). The wild bootstrap version of these tests is essentially the same as explained in the previous paragraph and an analogue of Theorem 4.1 can be established following the steps of its proof in the appendix.
In our investigation of the finite sample performance of these procedures we considered a uniform density for the explanatory variables $X_{1i}$ and $X_{2j}$ (i.e. $r_1 \equiv r_2 \equiv 1$), homoscedastic errors in both samples given by $\sigma_1^2(t) = 0.5, \sigma_2^2(t) = 0.25$ and the sample sizes $(n_1, n_2) = (25, 25), (25, 50), (25, 100), (50, 25), (50, 50), (50, 100)$. For the regression functions we used the following scenario

$$
\begin{aligned}
&\text{(i)} & f_1(x) &= f_2(x) = 0 \\
&\text{(ii)} & f_1(x) &= 0; f_2(x) = x \\
&\text{(iii)} & f_1(x) &= 0; f_2(x) = 1 \\
&\text{(iv)} & f_1(x) &= 0; f_2(x) = \sin(2\pi x) \\
&\text{(v)} & f_1(x) &= 0; f_2(x) = \sqrt{x} \\
&\text{(vi)} & f_1(x) &= 0; f_2(x) = 2x^2
\end{aligned} \tag{4.9}
$$

where the first case corresponds to the null hypothesis of equal regression curves. For the estimation of the regression functions from the total and individual samples we used a local linear estimator [see Fan and Gijbels (1996)] with the Epanechnikov kernel

$$K(x) = \frac{3}{4}(1 - x^2) I_{[-1,1]}(x),$$

which yields an equivalent kernel of order $r = 4$ [see Wand and Jones (1995), p. 125]. For the bandwidths we used

$$h = \left\{ \frac{n_1 \sigma_2^2 + n_2 \sigma_1^2}{(n_1 + n_2)^2} \right\}^{1/5}, g = h^{5/4}$$

for the estimation from the combined samples and

$$h_i = \left( \frac{\sigma_i^2}{n_i} \right)^{1/5}, \quad i = 1, 2$$

in the Nadaraya-Watson estimators of $\hat{f}_1$ and $\hat{f}_2$ from the individual samples. The random variables $V_{ij}$ used in the generation of the bootstrap sample are i.i.d. random variables with masses $(\sqrt{5}+1)/2\sqrt{5}$ and $(\sqrt{5}-1)/2\sqrt{5}$ at the points $(1-\sqrt{5})/2$ and $(1+\sqrt{5})/2$ (note that this distribution satisfies $E[V_{ij}] = 0, E[V_{ij}^2] = E[V_{ij}^3] = 1$). The corresponding results are listed in Table 4.1, 4.2, 4.3 and 4.4 for the statistics $K_N^{(1)}, K_N^{(2)}, L_N$ and $C_N$, respectively, which show the relative proportion of rejections based on 1000 simulation runs, where the number of bootstrap replications was chosen as $B = 200$. We observe a sufficiently accurate approximation of the nominal level in nearly all cases. A comparison of the tests based on $K_N^{(1)}$ and $K_N^{(2)}$ shows that the application of the marked empirical process $\hat{R}_N^{(2)}$ usually yields an improvement with respect to the power of approximately $5-10\%$ [see Table 4.1 and 4.2]. A further comparison with the statistic $L_N$ [essentially proposed by Kulasekera (1995)] shows that this procedure is comparable with the test based on the marked empirical process $\hat{R}_N^{(1)}$, except in the case of the oscillating alternative $f_2(x) - f_1(x) = \sin(2\pi x)$, which is nearly not detected by $L_N$ [see Table 4.1 and 4.3]. However, the natural competitor for $L_N$ is the statistic $K_N^{(2)}$, because in the construction of the marked empirical processes $\lambda_N^{(i)}$ in (3.4) we did not multiply the residuals with the density estimator of the denominator of the Nadaraya and Watson estimate. A comparison of the tests based on $L_N$ and $K_N^{(2)}$ shows a substantial better performance (with respect to power) of the test based on the statistic $K_N^{(2)}$ [see Table 4.2 and 4.3]. Similarly, a comparison with Cabus's approach shows that the test based on $K_N^{(2)}$ is more powerful than the test based on $C_N$ in all considered cases, especially under the oscillating alternative (iv) [see Table 4.2 and 4.4]. Based on these observations and additional simulation results (which are not displayed for the sake of brevity) we recommend to use functionals of the marked empirical process $\hat{R}_N^{(2)}$ in the problem of testing the equality of regression curves.

| $n_1$ | $n_2$ | 25 | | | 50 | | | 100 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | 2.5% | 5% | 10% | 2.5% | 5% | 10% | 2.5% | 5% | 10% |
| | (i) | 0.030 | 0.056 | 0.112 | 0.024 | 0.048 | 0.099 | 0.018 | 0.047 | 0.093 |
| | (ii) | 0.479 | 0.593 | 0.720 | 0.593 | 0.718 | 0.809 | 0.630 | 0.746 | 0.824 |
| 25 | (iii) | 0.991 | 0.998 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | (iv) | 0.171 | 0.283 | 0.455 | 0.381 | 0.544 | 0.719 | 0.630 | 0.772 | 0.916 |
| | (v) | 0.793 | 0.878 | 0.933 | 0.895 | 0.937 | 0.968 | 0.907 | 0.959 | 0.984 |
| | (vi) | 0.603 | 0.713 | 0.792 | 0.683 | 0.776 | 0.837 | 0.629 | 0.717 | 0.820 |
| | (i) | 0.022 | 0.049 | 0.114 | 0.021 | 0.048 | 0.098 | 0.025 | 0.051 | 0.108 |
| | (ii) | 0.657 | 0.766 | 0.840 | 0.828 | 0.886 | 0.931 | 0.921 | 0.949 | 0.973 |
| | (iii) | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 50 | (iv) | 0.180 | 0.295 | 0.458 | 0.441 | 0.598 | 0.785 | 0.755 | 0.868 | 0.959 |
| | (v) | 0.920 | 0.960 | 0.983 | 0.987 | 0.991 | 0.997 | 0.999 | 1.000 | 1.000 |
| | (iv) | 0.765 | 0.843 | 0.899 | 0.919 | 0.956 | 0.971 | 0.958 | 0.999 | 1.000 |

**Table 4.1** *Rejection probabilities of a wild bootstrap version of the test based on $K_N^{(1)}$ [see (4.1)] for various sample sizes and the regression functions specified in (4.9). The errors are homoscedastic and have variances $\sigma_1^2 = 0\ .5, \sigma_2^2 = 0\ .25$.*

| $n_1$ | $n_2$ | 25 | | | 50 | | | 100 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | 2.5% | 5% | 10% | 2.5% | 5% | 10% | 2.5% | 5% | 10% |
| | (i) | 0.024 | 0.055 | 0.120 | 0.029 | 0.053 | 0.111 | 0.021 | 0.056 | 0.110 |
| | (ii) | 0.404 | 0.730 | 0.827 | 0.705 | 0.808 | 0.888 | 0.732 | 0.826 | 0.899 |
| 25 | (iii) | 0.998 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | (iv) | 0.163 | 0.290 | 0.579 | 0.381 | 0.537 | 0.741 | 0.617 | 0.779 | 0.908 |
| | (v) | 0.883 | 0.942 | 0.972 | 0.952 | 0.977 | 0.989 | 0.960 | 0.983 | 0.990 |
| | (vi) | 0.783 | 0.866 | 0.927 | 0.861 | 0.923 | 0.955 | 0.807 | 0.887 | 0.935 |
| | (i) | 0.031 | 0.056 | 0.112 | 0.028 | 0.045 | 0.093 | 0.027 | 0.055 | 0.105 |
| | (ii) | 0.803 | 0.872 | 0.922 | 0.912 | 0.952 | 0.979 | 0.967 | 0.987 | 0.995 |
| | (iii) | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 50 | (iv) | 0.182 | 0.317 | 0.510 | 0.501 | 0.653 | 0.836 | 0.786 | 0.893 | 0.964 |
| | (v) | 0.968 | 0.986 | 0.995 | 0.997 | 0.999 | 1.000 | 0.999 | 1.000 | 1.000 |
| | (iv) | 0.905 | 0.949 | 0.973 | 0.987 | 0.997 | 0.998 | 0.997 | 0.999 | 1.000 |

**Table 4.2** *Rejection probabilities of a wild bootstrap version of the test based on $K_N^{(2)}$ [see (4.1)] for various sample sizes and the regression functions specified in (4.9). The errors are homoscedastic and have variances $\sigma_1^2 = 0\ .5, \sigma_2^2 = 0\ .25$.*

| $n_1$ | $n_2$ | 25 | | | 50 | | | 100 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | 2.5% | 5% | 10% | 2.5% | 5% | 10% | 2.5% | 5% | 10% |
| | (i) | 0.029 | 0.053 | 0.101 | 0.033 | 0.066 | 0.122 | 0.025 | 0.049 | 0.113 |
| | (ii) | 0.594 | 0.704 | 0.810 | 0.649 | 0.759 | 0.857 | 0.667 | 0.777 | 0.834 |
| 25 | (iii) | 0.995 | 0.998 | 1.000 | 0.995 | 0.998 | 0.998 | 1.000 | 1.000 | 1.000 |
| | (iv) | 0.037 | 0.057 | 0.120 | 0.022 | 0.034 | 0.065 | 0.009 | 0.017 | 0.042 |
| | (v) | 0.865 | 0.934 | 0.966 | 0.913 | 0.948 | 0.976 | 0.891 | 0.939 | 0.976 |
| | (vi) | 0.799 | 0.858 | 0.913 | 0.826 | 0.888 | 0.930 | 0.790 | 0.869 | 0.912 |
| | (i) | 0.027 | 0.054 | 0.110 | 0.028 | 0.045 | 0.080 | 0.031 | 0.052 | 0.106 |
| | (ii) | 0.805 | 0.883 | 0.935 | 0.909 | 0.953 | 0.983 | 0.948 | 0.969 | 0.988 |
| | (iii) | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 50 | (iv) | 0.025 | 0.046 | 0.102 | 0.023 | 0.046 | 0.079 | 0.012 | 0.038 | 0.032 |
| | (v) | 0.950 | 0.980 | 0.993 | 0.998 | 0.999 | 0.999 | 0.999 | 1.000 | 1.000 |
| | (iv) | 0.935 | 0.968 | 0.986 | 0.987 | 0.995 | 0.999 | 0.986 | 0.995 | 0.988 |

**Table 4.3** *Rejection probabilities of a wild bootstrap version of the test based on $L_N$ [see (4.7)] for various sample sizes and the regression functions specified in (4.9). The errors are homoscedastic and have variances $\sigma_1^2 = 0\ .5, \sigma_2^2 = 0\ .25$.*

| $n_1$ | $n_2$ | 25 | | | 50 | | | 100 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | 2.5% | 5% | 10% | 2.5% | 5% | 10% | 2.5% | 5% | 10% |
| | (i) | 0.026 | 0.048 | 0.102 | 0.037 | 0.060 | 0.121 | 0.017 | 0.053 | 0.118 |
| | (ii) | 0.582 | 0.702 | 0.808 | 0.636 | 0.757 | 0.844 | 0.653 | 0.763 | 0.834 |
| 25 | (iii) | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | (iv) | 0.062 | 0.090 | 0.137 | 0.048 | 0.057 | 0.103 | 0.030 | 0.045 | 0.062 |
| | (v) | 0.867 | 0.829 | 0.868 | 0.921 | 0.960 | 0.979 | 0.910 | 0.956 | 0.982 |
| | (vi) | 0.359 | 0.736 | 0.826 | 0.689 | 0.775 | 0.748 | 0.655 | 0.754 | 0.830 |
| | (i) | 0.022 | 0.055 | 0.106 | 0.020 | 0.041 | 0.080 | 0.029 | 0.051 | 0.104 |
| | (ii) | 0.792 | 0.876 | 0.915 | 0.889 | 0.935 | 0.965 | 0.928 | 0.951 | 0.974 |
| | (iii) | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 50 | (iv) | 0.046 | 0.079 | 0.149 | 0.048 | 0.076 | 0.134 | 0.022 | 0.038 | 0.068 |
| | (v) | 0.968 | 0.988 | 0.995 | 0.992 | 0.998 | 0.999 | 1.000 | 1.000 | 1.000 |
| | (iv) | 0.854 | 0.906 | 0.955 | 0.929 | 0.957 | 0.980 | 0.941 | 0.961 | 0.983 |

**Table 4.4** *Rejection probabilities of a wild bootstrap version of the test based on $C_N$ [see (4.8)] for various sample sizes and the regression functions specified in (4.9). The errors are homoscedastic and have variances $\sigma_1^2 = 0\ .5, \sigma_2^2 = 0\ .25$.*

# 5 Proofs

For the sake of brevity we restrict ourselves to a consideration of the process $\hat{R}_N^{(1)}$ defined in (2.10). The proofs for the process $\hat{R}_N^{(2)}$ are similar and therefore omitted.

## 5.1 Proof of Lemma 2.1

The expectation of the residuals in (2.8) is obtained as

$$\begin{aligned} E[e_{ij}] &= E\Big[E[Y_{ij}\hat{r}(X_{ij}) - \hat{f}(X_{ij})\hat{r}(X_{ij})|X_{11},\dots,X_{2n_2}]\Big] \\ &= \frac{1}{Nh}\sum_{\ell=1}^{2}\sum_{k=1}^{n_\ell} E\Big[K\Big(\frac{X_{\ell k}-X_{ij}}{h}\Big)\left(f_i(X_{ij}) - f_\ell(X_{\ell k})\right) I\{X_{ij}\le t\}\Big] \\ &= \frac{n_i-1}{Nh}\int_0^1\int_0^t K\Big(\frac{x-y}{h}\Big)(f_i(x)-f_i(y))r_i(x)r_i(y)\,dx\,dy \\ &\quad + \frac{n_{3-i}}{Nh}\int_0^1\int_0^t K\Big(\frac{x-y}{h}\Big)(f_i(x)-f_{3-i}(y))r_i(x)r_{3-i}(y)\,dx\,dy \end{aligned}$$

and a Taylor expansion and a standard argument yield

$$E[e_{ij}] = \kappa_{3-i}\int_0^t (f_i(x)-f_{3-i}(x))r_i(x)r_{3-i}(x)dx.$$

Observing the definition of $\hat{R}_N^{(1)}$ we obtain

$$\begin{aligned} E\left[\hat{R}_N^{(1)}(t)\right] &= \kappa_1\kappa_2 \int_0^t (f_1(x) - f_2(x)) r_1(x) r_2(x)\, dx \\ &\quad - \kappa_2\kappa_1 \int_0^t (f_2(x) - f_1(x)) r_2(x) r_1(x)\, dx + O(h^r), \end{aligned}$$

which establishes the assertion of the Lemma for the process $\hat{R}_N^{(1)}$.

□

## 5.2 Proof of Theorem 2.2

Recalling the definition of the residuals in (2.8)

$$\begin{aligned} e_{ij} &= \sigma_i(X_{ij})\varepsilon_{ij}\hat{r}(X_{ij}) + f(X_{ij})\hat{r}(X_{ij}) - \hat{f}(X_{ij})\hat{r}(X_{ij}) \\ &= \sigma_i(X_{ij})\varepsilon_{ij}\hat{r}(X_{ij}) + \frac{1}{Nh}\sum_{\ell=1}^{2}\sum_{k=1}^{n_\ell} K\Big(\frac{X_{ij}-X_{\ell k}}{h}\Big)(f(X_{ij}) - f(X_{\ell k})) \\ &\quad - \frac{1}{Nh}\sum_{\ell=1}^{2}\sum_{k=1}^{n_\ell} K\Big(\frac{X_{ij}-X_{\ell k}}{h}\Big)\sigma_\ell(X_{\ell k})\varepsilon_{\ell k}, \end{aligned} \tag{5.1}$$

and observing $f_1 = f_2$ under $H_0$ we obtain by a straightforward calculation the decomposition

$$\hat{R}_N^{(1)}(t) = R_N(t) + S_N(t) + W_N(t) + V_N(t) \tag{5.2}$$

where the processes $R_N, S_N, W_N$ and $V_N$ are defined by

$$R_N(t) := \frac{1}{N}\sum_{j=1}^{n_1}\sigma_1(X_{1j})\varepsilon_{1j} r(X_{1j}) I\{X_{1j} \le t\} - \frac{1}{N}\sum_{j=1}^{n_2}\sigma_2(X_{2j})\varepsilon_{2j} r(X_{2j}) I\{X_{2j} \le t\} \tag{5.3}$$

$$S_N(t) := \sum_{i=1}^{2}\frac{1}{N^2 h}\sum_{j=1}^{n_i}\sigma_i(X_{ij})\varepsilon_{ij}\ \Big\{\ \sum_{\ell=1}^{2}(-1)^\ell \sum_{k=1}^{n_\ell} K\Big(\frac{X_{ij}-X_{\ell k}}{h}\Big) I\{X_{\ell k} \le t\}\ \Big\} \tag{5.4}$$

$$W_N(t) := \sum_{\ell=1}^{2}(-1)^{\ell-1}\sum_{i=1}^{2}\frac{1}{N^2 h}\sum_{j=1}^{n_\ell}\sum_{k=1}^{n_i} K\Big(\frac{X_{\ell j}-X_{ik}}{h}\Big)(f(X_{\ell j}) - f(X_{ik})) I\{X_{\ell j} \le t\} \tag{5.5}$$

$$V_N(t) := \sum_{i=1}^{2}(-1)^{i-1}\frac{1}{N}\sum_{j=1}^{n_i}\sigma_i(X_{ij})\varepsilon_{ij}(\hat{r}(X_{ij}) - r(X_{ij})) I\{X_{ij} \le t\}. \tag{5.6}$$

The assertion of Theorem 2.2 now follows from the next Lemma and the following two auxiliary results, which will be proved below.

**Lemma 5.1.** *If the assumptions of Theorem 2.2 are satisfied, the process*

$$T_N(t) = \sqrt{N}(R_N + S'_N)$$

*converges weakly to a centered Gaussian process in the space $D[0,1]$ with covariance function given by (2.12), where $R_N$ is given by (5.3) and the process $S_N'$ is defined by*

$$(5.7) \qquad S_N'(t) := \sum_{i=1}^{2} \frac{1}{N} \sum_{j=1}^{n_i} \sigma_i(X_{ij})\varepsilon_{ij} \Big( \frac{1}{h} \int_0^t K\Big(\frac{X_{ij}-x}{h}\Big)(-\kappa_1 r_1(x) + \kappa_2 r_2(x))\, dx \Big).$$

**Proof.** With the notation

$$(5.8) \quad \Delta_{ij}(t) := \sigma_i(X_{ij})\Big[(-1)^{i-1} r(X_{ij}) I\{X_{ij} \le t\} + \frac{1}{h}\int_0^t K\Big(\frac{X_{ij}-x}{h}\Big)(-\kappa_1 r_1(x) + \kappa_2 r_2(x))\, dx\Big]$$

$(i = 1\,,2)$ we decompose the process $\sqrt{N}(R_N + S_N')$ as follows

$$T_N(t) = \sqrt{N}(R_N(t) + S_N'(t)) = \sum_{i=1}^{2} \frac{1}{\sqrt{N}} \sum_{j=1}^{n_i} \varepsilon_{ij}\Delta_{ij}(t).$$

For the covariance we obtain by a straightforward but cumbersome calculation

$$\begin{aligned}
\mathrm{Cov}(T_N(t), T_N(s)) &= E\Big[\frac{1}{N}\sum_{j=1}^{n_1} \sigma_1^2(X_{1j})\varepsilon_{1j}^2 \Delta_{1j}(t)\Delta_{1j}(s) + \frac{1}{N}\sum_{j=1}^{n_2} \sigma_2^2(X_{2j})\varepsilon_{2j}^2 \Delta_{2j}(t)\Delta_{2j}(s)\Big] \\
&= \kappa_1 \int_0^1 \sigma_1^2(y)\Big[r(y) I\{y \le t\} + \frac{1}{h}\int_0^t K\Big(\frac{y-x}{h}\Big)(-\kappa_1 r_1(x) + \kappa_2 r_2(x))\, dx\Big] \\
&\qquad \Big[r(y) I\{y \le s\} + \frac{1}{h}\int_0^s K\Big(\frac{y-x}{h}\Big)(-\kappa_1 r_1(x) + \kappa_2 r_2(x))\, dx\Big] r_1(y)\, dy \\
&\quad + \kappa_2 \int_0^1 \sigma_2^2(y)\Big[-r(y) I\{y \le t\} + \frac{1}{h}\int_0^t K\Big(\frac{y-x}{h}\Big)(-\kappa_1 r_1(x) + \kappa_2 r_2(x))\, dx\Big] \\
&\qquad \Big[-r(y) I\{y \le s\} + \frac{1}{h}\int_0^s K\Big(\frac{y-x}{h}\Big)(-\kappa_1 r_1(x) + \kappa_2 r_2(x))\, dx\Big] r_2(y)\, dy \\
&= A_1 + A_2
\end{aligned}$$

where the last equation defines the terms $A_1$ and $A_2$. The first term gives for $s \le t$

$$\begin{aligned}
A_1 &= \int_0^s \sigma_1^2(y) r^2(y)\kappa_1 r_1(y)\, dy \\
&\quad + \int_0^t \sigma_1^2(y) r(y) \frac{1}{h}\int_0^s K\Big(\frac{y-x}{h}\Big)(-\kappa_1 r_1(x) + \kappa_2 r_2(x))\, dx\, \kappa_1 r_1(y)\, dy \\
&\quad + \int_0^s \sigma_1^2(y) r(y) \frac{1}{h}\int_0^t K\Big(\frac{y-x}{h}\Big)(-\kappa_1 r_1(x) + \kappa_2 r_2(x))\, dx\, \kappa_1 r_1(y)\, dy \\
&\quad + \int_0^1 \sigma_1^2(y) \frac{1}{h^2}\int_0^t \int_0^s K\Big(\frac{y-x}{h}\Big) K\Big(\frac{y-z}{h}\Big)(-\kappa_1 r_1(x) + \kappa_2 r_2(x)) \\
&\qquad (-\kappa_1 r_1(z) + \kappa_2 r_2(z))\, dx\, dz\, \kappa_1 r_1(y)\, dy
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^s \sigma_1^2(y)(\kappa_1 r_1(y)+\kappa_2 r_2(y))^2 \kappa_1 r_1(y)\,dy \\
&\quad + 2\int_0^s \sigma_1^2(y)(\kappa_1 r_1(y)+\kappa_2 r_2(y))(-\kappa_1 r_1(y)+\kappa_2 r_2(y))\kappa_1 r_1(y)\,dy \\
&\quad + \int_0^s \sigma_1^2(y)(-\kappa_1 r_1(y)+\kappa_2 r_2(y))^2 \kappa_1 r_1(y)\,dy + o(1) \\
&= 4\int_0^s \sigma_1^2(y)\kappa_1 r_1(y)\kappa_2^2 r_2^2(y)\,dy + o(1).
\end{aligned}
$$

Now a similar calculation for the second term yields the claimed covariance structure, i.e.

$$
\begin{aligned}
\mathrm{Cov}(T_N(t),T_N(s)) &= 4\int_0^s \sigma_1^2(y)\kappa_1 r_1(y)\kappa_2^2 r_2^2(y)\,dy + 4\int_0^s \sigma_2^2(y)\kappa_2 r_2(y)\kappa_1^2 r_1^2(y)\,dy + o(1) \\
&= H(s,t)+o(1).
\end{aligned}
$$

The central limit theorem for triangular arrays proves convergence of the finite dimensional distributions of $T_N$. Weak convergence now follows if

$$
E\Big[(T_N(w)-T_N(v))^2(T_N(v)-T_N(u))^2\Big] \le C(w-u)^2 \quad \text{for all} \quad 0\le u\le v\le w\le 1 \tag{5.9}
$$

can be established [see Billingsley (1968); p. 128; or Shorack and Wellner (1986); p. 45-51]. To this end we note that for two independent samples of i.i.d. bivariate centered random vectors $(\alpha_i,\beta_i)_{i=1,\dots,n_1}$ and $(\gamma_i,\delta_i)_{i=1,\dots,n_2}$ the inequality

$$
\begin{aligned}
&E\Big[\Big(\sum_{i=1}^{n_1}\alpha_i+\sum_{j=1}^{n_2}\gamma_j\Big)^2\Big(\sum_{i=1}^{n_1}\beta_i+\sum_{j=1}^{n_2}\delta_j\Big)^2\Big] \le n_1E[\alpha_1^2\beta_1^2]+3\,n_1^2E[\alpha_1^2]E[\beta_1^2] \\
&+n_2E[\gamma_1^2\delta_1^2]+3\,n_2^2E[\gamma_1^2]E[\delta_1^2]+n_1n_2E[\alpha_1^2]E[\delta_1^2] \\
&+n_1n_2E[\gamma_1^2]E[\beta_1^2]+4\,n_1n_2E[\alpha_1\beta_1]E[\gamma_1\delta_1]
\end{aligned} \tag{5.10}
$$

holds which follows by similar arguments as stated in the proof of Theorem 13.1 in Billingsley (1968). We now apply (5.10) for the random variables

$$
\begin{aligned}
\alpha_i &= \varepsilon_{1i}(\Delta_{1i}(w)-\Delta_{1i}(v))\ , \qquad \beta_i = \varepsilon_{1i}(\Delta_{1i}(v)-\Delta_{1i}(u))\ , \\
\gamma_j &= \varepsilon_{2j}(\Delta_{2j}(w)-\Delta_{2j}(v))\ , \qquad \delta_j = \varepsilon_{2j}(\Delta_{2j}(v)-\Delta_{2j}(u))\ .
\end{aligned} \tag{5.11}
$$

A straightforward but cumbersome calculation yields

$$
\begin{aligned}
E[\alpha_1^2] &= \int_0^1 \sigma_1^2(x)\Big(r(x)I\{v\le x\le w\}+\frac{1}{h}\int_v^w K(\frac{x-z}{h})(-\kappa_1 r_1(z)+\kappa_2 r_2(z))\,dz\Big)^2 r_1(x)\,dx \\
&= \int_v^w \sigma_1^2(x)r^2(x)r_1(x)\,dx \\
&\quad + 2\int_v^w \sigma_1^2(x)r(x)\frac{1}{h}\int_v^w K(\frac{x-z}{h})(-\kappa_1 r_1(z)+\kappa_2 r_2(z))\,dz r_1(x)\,dx
\end{aligned}
$$

$$
\begin{aligned}
&+\int_0^1 \sigma_1^2(x)\Big(\frac{1}{h}\int_v^w K(\frac{x-z}{h})(-\kappa_1 r_1(z)+\kappa_2 r_2(z))\,dz\Big)^2 r_1(x)\,dx \\
&\le\ O(1)\ (w-v)+O(1)\int_v^w \Big(\frac{1}{h}\int_0^1 \sigma_1^2(x)K(\frac{x-z}{h})(-\kappa_1 r_1(z)+\kappa_2 r_2(z))r_1(x)\,dx\Big)\,dz \\
&=\ O(1)\ (w-u)
\end{aligned}
$$

and similar arguments show that the terms $E[\beta_1^2], E[\gamma_1^2], E[\delta_1^2], E[\alpha_1\beta_1]$ and $E[\gamma_1\delta_1]$ are of the same order. Similarly we have

$$
\begin{aligned}
E[\alpha_1^2\beta_1^2] &= E[\varepsilon_{11}^4]\int_0^1 \sigma_1^4(x)\Big(r(x)I\{v\le x\le w\}+\frac{1}{h}\int_v^w K\Big(\frac{z-x}{h}\Big)(-\kappa_1 r_1(z)+\kappa_2 r_2(z))\,dz\Big)^2 \\
&\qquad \Big(r(x)I\{u\le x\le v\}+\frac{1}{h}\int_u^v K\Big(\frac{z-x}{h}\Big)(-\kappa_1 r_1(z)+\kappa_2 r_2(z))\,dz\Big)^2 r_1(x)\,dx \\
&= O(1)\ \frac{(w-u)^2}{h^2}, \\
E[\gamma_1^2\delta_1^2] &= O(\frac{1}{h^2})(w-u)^2.
\end{aligned}
$$

Now, a combination of these results with (5.11) and (5.10) yields

$$
\begin{aligned}
E\Big[(T_N(w)-T_N(v))^2(T_N(v)-T_N(u))^2\Big] &= \frac{1}{N^2}E\Big[\Big(\sum_{i=1}^{n_1}\alpha_i+\sum_{j=1}^{n_2}\gamma_j\Big)^2\Big(\sum_{i=1}^{n_1}\beta_i+\sum_{j=1}^{n_2}\delta_j\Big)^2\Big] \\
&= \Big(O(\frac{1}{Nh^2})+O(1)\Big)(w-u)^2\ =\ O(1)(w-u)^2,
\end{aligned}
$$

which establishes (5.9) and completes the proof of Lemma 5.1.

□

**Lemma 5.2.** *If the assumptions of Theorem 2.2 are satisfied we have for the processes $S_N$ and $S_N'$ defined by (5.4) and (5.7)*

$$
\sup_{t\in[0,1]}|S_N(t)-S_N'(t)| = o_p(\frac{1}{\sqrt{N}}). \tag{5.12}
$$

**Lemma 5.3.** *If the assumptions of Theorem 2.2 are satisfied we have for the processes $V_N$ and $W_N$ defined by (5.6) and (5.5)*

$$
\sup_{t\in[0,1]}|V_N(t)|\ =\ o_p(\frac{1}{\sqrt{N}}) \tag{5.13}
$$

$$
\sup_{t\in[0,1]}|W_N(t)|\ =\ o_p(\frac{1}{\sqrt{N}}). \tag{5.14}
$$

In order to prove Lemma 5.2 and 5.3 we need some basic terminology from recent $U$-processes theory. For more details we refer to Nolan and Pollard (1987, 1988) or Pollard (1984). Let $\mathcal{F}$ denote a class of real valued (measurable) functions defined on a set $S$ with envelope $F$. The *covering number* $\mathcal{N}_p(\varepsilon, Q, \mathcal{F}, F)$ of $\mathcal{F}$ (with respect to the probability measure $Q$) is defined as the smallest cardinality for a subclass $\mathcal{F}^*$ of $\mathcal{F}$ such that

$$\min_{f^* \in \mathcal{F}^*} Q|f - f^*|^p \ \leq \ \varepsilon^p Q(F^p) \quad \text{for all} \quad f \in \mathcal{F}$$

and

$$\mathcal{J}(t, Q, \mathcal{F}, F) = \int_0^t \log \mathcal{N}_2(x, Q, \mathcal{F}, F)\, dx$$

is called the *covering integral.* The class $\mathcal{F}$ is called *euclidean*, if there exist constants $A$ and $V$ such that

$$\mathcal{N}_1(\varepsilon, Q, \mathcal{F}, F) \ \leq \ A\varepsilon^{-V}.$$

The class $\mathcal{F}$ is called *VC-class* if its class of graphs

$$\mathcal{D} \ = \ \{G_f \,|\, f \in \mathcal{F}\}$$

with

$$G_f := \{(s,t) \,|\, 0 \leq t \leq f(s) \ \text{ or } \ f(s) \leq t \leq 0\}$$

forms a polynomial class (or VC class); i.e. there exists a polynomial $p(\cdot)$ such that

$$\#\{D \cap F | D \in \mathcal{D}\} \ \leq \ p(\#F)$$

for every fixed finite subset $F$ of $S$. We finally note that VC classes are euclidean [see Pollard (1984), Lemma II 25] and that sums of euclidean classes are euclidean [see Nolan and Pollard (1987), Corollary 17].

## 5.3 Proof of Lemma 5.3

We will restrict ourselves to the process $V_N$ considered in (5.13), the remaining case (5.14) is very similar and left to the reader. Recalling the definition of $V_N$ in (5.6) we obtain the decomposition

$$V_N(t) \ = \ V_N^{(1)}(t) + V_N^{(2)}(t) + V_N^{(3)}(t) + V_N^{(4)}(t) + o_p(\frac{1}{\sqrt{N}}), \tag{5.15}$$

where

$$V_N^{(1)}(t) \ = \ \frac{1}{N^2 h} \sum_{j=1}^{n_1} \sum_{k=1}^{n_1} \sigma_1(X_{1j})\varepsilon_{1j} \left( K\Big(\frac{X_{1j} - X_{1k}}{h}\Big) - h r_1(X_{1j}) \right) I\{X_{1j} \leq t\} \tag{5.16}$$

$$V_N^{(2)}(t) \ = \ \frac{1}{N^2 h} \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} \sigma_1(X_{1j})\varepsilon_{1j} \left( K\Big(\frac{X_{1j} - X_{2k}}{h}\Big) - h r_2(X_{1j}) \right) I\{X_{1j} \leq t\} \tag{5.17}$$

$$(5.18) \qquad V_N^{(3)}(t) = \frac{1}{N^2h}\sum_{j=1}^{n_2}\sum_{k=1}^{n_1}\sigma_2(X_{2j})\varepsilon_{2j}\left(K\Big(\frac{X_{2j}-X_{1k}}{h}\Big)-hr_1(X_{2j})\right)I\{X_{2j}\le t\}$$

$$(5.19) \qquad V_N^{(4)}(t) = \frac{1}{N^2h}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}\sigma_2(X_{2j})\varepsilon_{2j}\left(K\Big(\frac{X_{2j}-X_{2k}}{h}\Big)-hr_2(X_{2j})\right)I\{X_{2j}\le t\},$$

the remainder in (5.15) is obtained replacing $\kappa_i$ by $n_i/N$ and vanishes uniformly with respect to $t \in [0,1]$. The assertion of Lemma 5.3 now follows by showing that all terms in (5.15) are of order $o_p(\frac{1}{\sqrt{N}})$ uniformly with respect to $t \in [0,1]$.

**Lemma 5.3a.** *If the assumptions of Theorem 2.2 are satisfied we have for the statistics $V_N^{(1)}$ and $V_N^{(4)}$ defined by (5.16) and (5.19)*

$$\sup_{t\in[0,1]}|V_N^{(1)}(t)| = o_p(\frac{1}{\sqrt{N}})$$
$$\sup_{t\in[0,1]}|V_N^{(4)}(t)| = o_p(\frac{1}{\sqrt{N}}).$$

**Proof** (of Lemma 5.3a). Both terms are treated exactly in the same way and we only consider $V_N^{(1)}$ which can be written as

$$V_N^{(1)}(t) = \frac{1}{N^2h}\sum_{j=1}^{n_1}\sum_{\substack{k=1\\k\neq j}}^{n_1}\sigma_1(X_{1j})\varepsilon_{1j}\left(K\Big(\frac{X_{1j}-X_{1k}}{h}\Big)-hr_1(X_{1j})\right)I\{X_{1j}\le t\}$$
$$+\frac{1}{N^2h}\sum_{j=1}^{n_1}\sigma_1(X_{1j})\varepsilon_{1j}\left(K(0)-hr_1(X_{1j})\right)I\{X_{1j}\le t\}$$
$$(5.20) \qquad =: I_N(t)+I_N^{(1)}(t)$$

where the last line defines the processes $I_N$ and $I_N^{(1)}$, respectively. For the lastnamed term we obtain by a straightforward calculation

$$(5.21) \qquad \sup_{t\in[0,1]}|I_N^{(1)}(t)| = O_p(\frac{1}{Nh}) = o_p(\frac{1}{\sqrt{N}})$$

where we have used the assumptions for the bandwidth stated in (2.3). The treatment of the remaining term $I_N$ in (5.20) is more complicated and requires some basic results from the treatment of $U$-processes [see e.g. Nolan and Pollard (1987)]. To be precise observe that

$$(5.22) \qquad \sqrt{N}I_N - \frac{\kappa_1^{3/2}}{2h}U_{n_1}(\varphi) = o_p(1)$$

uniformly with respect to $t \in [0,1]$, where $U_{n_1}$ is a $U$-process defined by

$$U_{n_1}(\varphi) := \frac{\sqrt{n_1}}{n_1(n_1-1)} \sum_{i=1}^{n_1} \sum_{\substack{j=1 \\ i \neq j}}^{n_1} \varphi(\xi_i, \xi_j) \tag{5.23}$$

with $\xi_i = (X_{1i}, \varepsilon_{1i})$ and symmetric kernel

$$\begin{aligned} \varphi(\xi_i, \xi_j) \; &= \; \varepsilon_{1j} \left( K\Big(\frac{X_{1i} - X_{1j}}{h}\Big) - hr_1(X_{1j}) \right) \sigma_1(X_{1j}) I\{X_{1j} \leq t\} \\ &\quad + \varepsilon_{1i} \left( K\Big(\frac{X_{1i} - X_{1j}}{h}\Big) - hr_1(X_{1i}) \right) \sigma_1(X_{1i}) I\{X_{1i} \leq t\}. \end{aligned} \tag{5.24}$$

Following Nolan and Pollard (1988) we introduce the notation $\varphi_1(x) = E[\varphi(\xi_1, \xi_2)|\xi_2 = x]$ and obtain a Hoeffding decomposition for the process $U_{n_1}$, i.e.

$$U_{n_1}(\varphi) = U_{n_1}(\tilde{\varphi}) + \frac{2}{\sqrt{n_1}} \sum_{i=1}^{n_1} \varphi_1(\xi_i) \tag{5.25}$$

where

$$\tilde{\varphi}(x,y) = \varphi(x,y) - \varphi_1(x) - \varphi_1(y) \tag{5.26}$$

(note that $E[\varphi(\xi_1, \xi_2)] = 0$). Finally, consider a class of functions

$$\mathcal{F} \; = \; \Big\{ \varphi_{h,t} \mid t \in [0,1], h > 0 \Big\}, \tag{5.27}$$

where $\varphi_{h,t} : [0,1] \times \mathbb{R} \times [0,1] \times \mathbb{R} \to \mathbb{R}$ is defined by

$$\begin{aligned} \varphi_{h,t}(x,y) \; &= \; x_2 \left( K\Big(\frac{x_1 - y_1}{h}\Big) - hr_1(x_1) \right) \sigma_1(x_1) I\{x_1 \leq t\} \\ &\quad + y_2 \left( K\Big(\frac{x_1 - y_1}{h}\Big) - hr_1(y_1) \right) \sigma_1(y_1) I\{y_1 \leq t\}. \end{aligned} \tag{5.28}$$

It can be shown by a tedious calculation and similar arguments as in Noland and Pollard (1987), Lemma 16, and Pollard (1984), Examples II 26, II 38 that the class $\mathcal{F}$ and the induced class

$$P\mathcal{F} \; = \; \{\varphi_1 \mid \varphi_1(x) = E[\varphi(\xi_1, \xi_2)|\xi_2 = x], \varphi \in \mathcal{F}\} \tag{5.29}$$

are euclidean. Note that the proof of this property requires the special assumption on the kernel $K$ stated in the paragraph following equation (2.4) [see Pollard (1984), Example II 38 and problem II 28, who considered the case of a decreasing kernel function on $[0, \infty)$, which is a special case of the situation considered here]. It therefore follows that for $\gamma > 0$ the covering integral satisfies

$$\begin{aligned} \mathcal{J}(\gamma, Q \otimes Q, \mathcal{F}, F) \; &\leq \; a_1 \gamma - b_1(\gamma \log \gamma - \gamma) \\ \mathcal{J}(\gamma, Q, P\mathcal{F}, PF) \; &\leq \; a_2 \gamma - b_2(\gamma \log \gamma - \gamma) \end{aligned}$$

(for given constants $a_1, b_1, a_2, b_2$) and consequently the assumptions of Theorem 5 in Nolan and Pollard (1988) are fullfilled. Now the second part in the proof of this theorem shows

$$\sup_{\varphi\in\mathcal{F}} |U_{n_1}(\tilde{\varphi})| = O_p(\frac{1}{\sqrt{N}}). \tag{5.30}$$

The assertion of the first part in Lemma 5.3a now follows from (5.30), (5.25), (5.22), (5.20) and (5.21) if the estimate

$$\sup_{t\in[0,1]} |\frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \frac{1}{h} \varphi_{1,t,h_{n_1}}(\xi_i)| = o_p(1) \tag{5.31}$$

can be established, where

$$\varphi_{1,t,h}(\xi) \;=\; \varphi_1(\xi_i) = \varepsilon_{1i}\Big(\int K(\frac{x-X_{1i}}{h}) r_1(x)\,dx - h r_1(X_{1i})\Big)\sigma_1(X_{1i}) I\{X_{1i} \le t\}\ . \tag{5.32}$$

To this end we make the dependence of the bandwidth from the sample size explicit by writing $h = h_{n_1}$ and introduce the notation

$$\mathcal{F}_{n_1} \;:=\; \Big\{\varphi_{1,t,h_{n_1}}\ \Big| t\in[0,1]\Big\}\ . \tag{5.33}$$

We use similar arguments as given in the proof of Theorem 37 in Pollard (1984, p. 34). To be precise define

$$\alpha_{n_1} = \frac{1}{\sqrt{n_1 h_{n_1}^{2r}}}\ ,\ \delta_{n_1} = \sqrt{c h_{n_1}^{2r+1}}$$

where $c$ is a constant chosen such that

$$\begin{aligned} P(\varphi^2_{1,t,h_{n_1}}) \;&=\; \int_0^t \sigma_1^2(z)\Big(\int K\Big(\frac{x-z}{h_{n_1}}\Big) r_1(x)\,dx - h_{n_1} r_1(z)\Big)^2 r_1(z)\,dz \\ &=\; h_{n_1}^2 \int_0^t \sigma_1^2(z)\Big(\int K(u)(r_1(z+h_{n_1}u) - r_1(z))\,du\Big)^2 r_1(z)\,dz \;\le\; h_{n_1}^2 h_{n_1}^{2r}\cdot c. \end{aligned}$$

Let $F_1$ denote the envelope of the class $P\mathcal{F}$ defined by (5.29) (note that $\mathcal{F}_{n_1} \subset P\mathcal{F}$ for all $n_1 \in \mathbb{N}$) and assume without loss of generality $0 < k_1 < PF_1 < k_2$. By the strong law of large numbers we have

$$I\!P(|P_{n_1}F_1 - PF_1| > \frac{k_1}{2}) \quad \stackrel{N\to\infty}{\longrightarrow} \ 0$$

where $P_{n_1}$ is the distribution with equal masses at the points $\xi_1, \dots, \xi_{n_1}$. Therefore it is sufficient to prove the assertion (5.31) on the set $\{|P_{n_1}F_1 - PF_1| \le \frac{k_1}{2}\}$ for which $\frac{k_1}{2} < P_{n_1}F_1 < \frac{k_1}{2} + k_2$. The following calculations are restricted to this set without mentioning this explicitly. Let $P_n^\circ$ denote the symmetrization of $P_n$ [see Pollard (1984), p. 15], then we obtain for $\varepsilon_{n_1} = \varepsilon\delta_{n_1}^2\alpha_{n_1}$ $(\varepsilon > 0)$

$$\begin{aligned} I\!P\Big(\sup_{\varphi\in\mathcal{F}_{n_1}} |P_{n_1}(\varphi)| > 8\varepsilon_{n_1}(\frac{k_1}{2}+k_2)\Big) \;&\le\; 4I\!P\Big(\sup_{\varphi\in\mathcal{F}_{n_1}} |P^\circ_{n_1}(\varphi)| > 2\varepsilon_{n_1}(\frac{k_1}{2}+k_2)\Big) \\ &\le\; 4I\!P\Big(\sup_{\varphi\in\mathcal{F}_{n_1}} |P^\circ_{n_1}(\varphi)| > 2\varepsilon_{n_1} P_{n_1}F_1\Big). \end{aligned} \tag{5.34}$$

Conditioning on $\xi = (\xi_1, \dots, \xi_{n_1})$ it therefore follows

$$\mathbb{P}\Big( \sup_{\varphi \in \mathcal{F}_{n_1}} |P^{\circ}_{n_1}(\varphi)| > 2\varepsilon_{n_1} P_{n_1} F_1 \Big| \xi \Big) \le \min \Big\{ 2\mathcal{N}_1(\varepsilon_{n_1}, P_{n_1}, \mathcal{F}_{n_1}, F_1) \exp\Big( -\frac{1}{2} \frac{n_1 \varepsilon^2_{n_1} (P_{n_1} F_1)^2}{\max_j P_{n_1} g_j^2} \Big), 1 \Big\},$$

where the maximum runs over all $m = \mathcal{N}_1(\varepsilon_{n_1}, P_{n_1}, \mathcal{F}_{n_1}, F_1)$ functions of the approximating class $\{g_1, \dots, g_m\}$. Integrating, observing that $P_{n_1} F_1 > \frac{k_1}{2}$ and that $P\mathcal{F}$ is euclidean yields

$$\begin{aligned} \mathbb{P}\Big( \sup_{\varphi \in \mathcal{F}_{n_1}} |P^{\circ}_{n_1}(\varphi)| > 2\varepsilon_{n_1} P_{n_1} F_1 \Big) &\le 2A\varepsilon_{n_1}^{-V} \exp\Big( -\frac{1}{8} \frac{k_1^2\, \varepsilon^2_{n_1}}{64\, \delta^2_{n_1}} \Big) \\ &\quad + \mathbb{P}\Big( \sup_{\varphi \in \mathcal{F}_{n_1}} P_{n_1}(\varphi^2) > 64\delta^2_{n_1} \Big) \end{aligned} \tag{5.35}$$

with positive constants $A$ and $V$. The first term can be treated similarly as in Pollard (1984, p. 34) and converges to 0. The treatment of the second term is different because $\varphi \in \mathcal{F}_{n_1}$ does not necessarily implies $|\varphi| \le 1$. We obtain for the expectation

$$\begin{aligned} E\Big| \sup_{\varphi \in \mathcal{F}_{n_1}} P_{n_1}(\varphi^2) \Big| &\le \frac{1}{n_1} E\Big| \sum_{i=1}^{n_1} \varepsilon^2_{1i} \Big( \int K\big(\frac{x - X_{1i}}{h_{n_1}}\big) r_1(x)\, dx - h_{n_1} r_1(X_{1i}) \Big)^2 \sigma_1^2(X_{1i}) \Big| \\ &= O(h_{n_1}^{2r+2}) \end{aligned}$$

and Markov's inequality yields (using the definition of $\delta_{n_1}$)

$$\mathbb{P}\Big( \sup_{\varphi \in \mathcal{F}_{n_1}} P_{n_1}(\varphi^2) > 64\delta_{n_1} \Big) = O(h_{n_1}) \tag{5.36}$$

A combination of (5.34), (5.35) and (5.36) finally gives

$$\mathbb{P}\Big( \frac{1}{\delta^2_{n_1} \alpha_{n_1}} \sup_{\varphi \in \mathcal{F}_{n_1}} |P_{n_1}(\varphi)| > \varepsilon \Big) \to 0 \quad \text{if } n_1 \to \infty$$

which establishes the remaining estimate (5.31) [note that $\delta^2_{n_1} \alpha_{n_1} = O(h_{n_1}/\sqrt{n_1})$].

**Lemma 5.3b** *If the assumptions of Theorem 2.2 are satisfied we have for the statistics $V_N^{(2)}$ and $V_N^{(3)}$ defined by (5.17) and (5.18)*

$$\begin{aligned} \sup_{t \in [0,1]} |V_N^{(2)}(t)| &= o_p\Big(\frac{1}{\sqrt{N}}\Big) \\ \sup_{t \in [0,1]} |V_N^{(3)}(t)| &= o_p\Big(\frac{1}{\sqrt{N}}\Big). \end{aligned}$$

**Proof.** The proof essentially follows the arguments given in the proof of Lemma 5.3a and we will restrict ourselves indicating the main difference, which is a derivation of an analogue of the estimate (5.30). Because $V_N^{(2)}$ and $V_N^{(3)}$ are $U$-processes formed from two samples the results derived in the proof of Theorem 5 of Nolan and Pollard (1988) are not directly applicable. For this reason we indicate the derivation of an analoguous result for two sample $U$-processes. The application of this result to the two sample $U$-processes obtained from $V_N^{(2)}$ and $V_N^{(3)}$ completes the proof of Lemma 5.3b and follows by exactly the same arguments as given in the proof of Lemma 5.3a.

To be precise let $P, Q$ denote distributions on the spaces $\mathcal{X}$ and $\mathcal{Y}$ and consider a class of real valued measurable functions $\mathcal{F}$ defined on $\mathcal{X} \times \mathcal{Y}$ such that $(P \otimes Q)(\varphi) = 0$ for all $\varphi \in \mathcal{F}$. Assume that there exists an envelope $F$ of $\mathcal{F}$ such that $(P \otimes Q)(F) < \infty$. Let $X_1, \ldots, X_{2n} \sim P$ and $Y_1, \ldots, Y_{2m} \sim Q$ denote independent samples and $\sigma_1, \ldots, \sigma_n$ and $\tau_1, \ldots, \tau_m$ denote independent samples (also independent from the $X_i$ and $Y_j$) such that

$$\begin{aligned} I\!P(\sigma_i = 1) &= I\!P(\sigma_i = -1) = 1\,/2, \\ I\!P(\tau_i = 1) &= I\!P(\tau_i = -1) = 1/2. \end{aligned}$$

Introducing the notation

$$\begin{aligned} \xi_i &= I\{\sigma_i = 1\}X_{2i} + I\{\sigma_i = -1\}X_{2i-1} \\ \xi_i' &= I\{\sigma_i = 1\}X_{2i-1} + I\{\sigma_i = -1\}X_{2i} \\ \zeta_j &= I\{\tau_j = 1\}Y_{2j} + I\{\tau_j = -1\}Y_{2j-1} \\ \zeta_j' &= I\{\tau_j = 1\}Y_{2j-1} + I\{\tau_j = -1\}Y_{2j} \end{aligned}$$

we obtain again independent samples $\xi_1, \ldots, \xi_n, \xi_1', \ldots, \xi_n' \sim P$ and $\zeta_1, \ldots, \zeta_m, \zeta_1', \ldots, \zeta_m' \sim Q$. For a function $\varphi \in \mathcal{F}$ consider the two sample $U$-statistic

$$S_{nm}(\varphi) := \sum_{i=1}^{n} \sum_{j=1}^{m} \varphi(\xi_i, \zeta_j), \tag{5.37}$$

and its standardized version

$$U_{nm}(\varphi) := \frac{\sqrt{n+m}}{nm} S_{nm}(\varphi). \tag{5.38}$$

Let

$$\begin{aligned} \varphi_1(x) &= E[\varphi(\xi_1, \zeta_1) | \xi_1 = x] \\ \varphi_2(y) &= E[\varphi(\xi_1, \zeta_1) | \zeta_1 = y] \end{aligned}$$

and define the kernel

$$\tilde{\varphi}(x, y) = \varphi(x, y) - \varphi_1(x) - \varphi_2(y) \tag{5.39}$$

then it follows that the statistic $U_{nm}(\tilde{\varphi})$ is degenerate [note that $E[\varphi(\xi_i, \zeta_j)] = 0$ by the definiton of $\mathcal{F}$]. Defining

$$\begin{aligned} T_{nm}(\varphi) &:= \sum_{i=1}^{n} \sum_{j=1}^{m} \Big[\varphi(\xi_i, \zeta_j) + \varphi(\xi_i, \zeta_j') + \varphi(\xi_i', \zeta_j) + \varphi(\xi_i', \zeta_j')\Big] \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} \varphi(X_{2i}, Y_{2j}) + \varphi(X_{2i}, Y_{2j-1}) + \varphi(X_{2i-1}, Y_{2j}) + \varphi(X_{2i-1}, Y_{2j-1}) \end{aligned} \tag{5.40}$$

and $P_n$ and $Q_m$ as the empirical distributions based on $\xi_1, \ldots, \xi_n$ and $\zeta_1, \ldots, \zeta_m$, respectively, it can be shown by similar arguments as in Nolan and Pollard (1988) that the conditions

$$\sup_{n,m} E[\mathcal{J}(1, T_{nm}, \mathcal{F}, F)^2] \; < \; \infty \tag{5.41}$$

$$\mathcal{J}(1, P \otimes Q, \mathcal{F}, F) \; < \; \infty \tag{5.42}$$

$$\sup_{n} E[\mathcal{J}(1, P_n, P\mathcal{F}, PF)^2] \; < \; \infty \tag{5.43}$$

$$\sup_{m} E[\mathcal{J}(1, Q_m, Q\,\mathcal{F}, QF)^2] \; < \; \infty \tag{5.44}$$

imply the estimate

$$E\Big[\sup_{\varphi \in \mathcal{F}} |U_{nm}(\tilde{\varphi})|\Big] = O(\frac{1}{\sqrt{N}})$$

which gives

$$\sup_{\varphi \in \mathcal{F}} |U_{nm}(\tilde{\varphi})| = O_p(\frac{1}{\sqrt{N}}). \tag{5.45}$$

In the specific situation of $V_N^{(2)}$ or $V_N^{(3)}$ the assumptions (5.41) - (5.44) now follows, because the classes $\mathcal{F}, P\mathcal{F}$ and $Q\mathcal{F}$ are euclidean [see the first part in the proof of Lemma 5.3a].

□

## 5.4 Proof of Lemma 5.2

Recalling the definition of $S_N$ and $S_N'$ in (5.4) and (5.7), respectively, it follows that the difference $S_N - S_N'$ is a linear combination of four terms of the form

$$\frac{2}{hn_\ell n_k} \sum_{i=1}^{n_\ell} \varepsilon_{\ell i} \Big( \sum_{\substack{j=1 \\ i \neq j}}^{n_k} K\Big(\frac{X_{\ell i} - X_{kj}}{h}\Big) I\{X_{kj} \leq t\} - \int_0^t K\Big(\frac{X_{\ell i} - x}{h}\Big) r_k(x)\, dx \Big) \sigma_\ell(X_{\ell i})$$

which can either be represented as a degenerate one-sample $U$-process [$\ell = k = 1$ ,and $\ell = k = 2$] or a degenerate two-sample $U$-process [$\ell = 1$ , $k=$ 2 and $\ell = 2$ , $k=$ 1] .It now follows either by the arguments in the proof of Theorem 5 in Nolan and Pollard (1988) or by its generalization in (5.41) - (5.44) and (5.45) that the corresponding terms vanish at a rate $O_p(\frac{1}{Nh})$ if the underlying class of indexing functions is euclidean. For example, in the case $\ell = k = 1$ the symmetric kernel is given by

$$\begin{aligned} \varphi(\xi_i, \xi_j) \; &= \; \varepsilon_{1i} \Big( K\Big(\frac{X_{1i} - X_{1j}}{h}\Big) I\{X_{1j} \leq t\} - \int_0^t K\Big(\frac{X_{1i} - x}{h}\Big) r_1(x)\, dx \Big) \sigma_1(X_{1i}) \\ &\quad + \varepsilon_{1j} \Big( K\Big(\frac{X_{1i} - X_{1j}}{h}\Big) I\{X_{1i} \leq t\} - \int_0^t K\Big(\frac{X_{1j} - x}{h}\Big) r_1(x)\, dx \Big) \sigma_1(X_{1j}), \end{aligned}$$

where $\xi_i = (\, X_{1i}, \varepsilon_{1i})$ and the degenerate one sample $U$-process is given by

$$U_{n_1,n_1}^{(1,1)}(\varphi) = \frac{1}{n_1^2} \sum_{i \neq j} \varphi(\xi_i, \xi_j).$$

Note that $\varphi_1(x) = E[\varphi(\xi_1, \xi_2)|\xi_2 = x] = 0$ which implies $\tilde{\varphi} = \varphi$ and $P\mathcal{F} = \{0\}$, which is obviously euclidean. A cumbersome calculation shows that $\mathcal{F}$ is also euclidean and the arguments in the proof of Theorem 5 in Nolan and Pollard (1988) yield

$$\frac{1}{h} \sup_{\varphi \in \mathcal{F}} |U^{(1,1)}_{n_1,n_1}(\varphi)| = \frac{1}{h} O_p(\frac{1}{N}) = o_p(\frac{1}{\sqrt{N}}).$$

The other three cases are treated exactly in the same way establishing the assertion of Lemma 5.2.

□

## 5.5 Proof of Theorem 3.2 and 3.3

The proof follows essentially the steps given for the proof of Theorem 2.2 and therefore we restrict ourselves to the calculation of the asymptotic covariance structure of the process defined by (3.2). A straightforward calculation yields

$$\begin{aligned}
\mathrm{Cov}(\mu_1(t), \mu_1(s)) &= \frac{1}{n_1 n_2^2 h^2} \sum_{i=1}^{\lfloor n_1 s \rfloor \wedge \lfloor n_1 t \rfloor} \sum_{j,\ell=1}^{n_2} K\left(\frac{t_{1i} - t_{2j}}{h}\right) K\left(\frac{t_{1i} - t_{2\ell}}{h}\right) \frac{\sigma_1^2(t_{1i})}{r_2^2(t_{1i})} \\
&\quad + \frac{1}{n_1 n_2^2 h^2} \sum_{i=1}^{\lfloor n_1 s \rfloor} \sum_{k=1}^{\lfloor n_1 t \rfloor} \sum_{j=1}^{n_2} K\left(\frac{t_{1i} - t_{2j}}{h}\right) K\left(\frac{t_{1k} - t_{2j}}{h}\right) \frac{\sigma_2^2(t_{2j})}{r_2(t_{1i}) r_2(t_{1k})} + o(1) \\
&= \frac{1}{h^2} \int_0^{R_1^{-1}(s \wedge t)} \int_0^1 \int_0^1 K\left(\frac{x-y}{h}\right) K\left(\frac{x-z}{h}\right) \frac{\sigma_1^2(x)}{r_2^2(x)} r_1(x) r_2(y) r_2(z) dx dy dz \\
&\quad + \frac{n_1}{n_2} \frac{1}{h^2} \int_0^{R_1^{-1}(t)} \int_0^{R_1^{-1}(s)} \int_0^1 K\left(\frac{x-y}{h}\right) K\left(\frac{z-y}{h}\right) \frac{\sigma_2^2(y) r_1(x) r_2(y) r_1(z)}{r_2(x) r_1(z)} dy dx dz \\
&\quad + o(1) \\
&= m_{12}(s,t) + o(1)
\end{aligned}$$

where $m_{12}$ is defined by (3.3).

□

## 5.6 Proof of Theorem 4.1

The proof essentially follows the proof of Theorem 2.2 and we will only sketch the main arguments. For the sake of simplicity we restrict ourselves to the process $\hat{R}_N^{(1)*}$ (the remaining case is treated exactly in the same way) and start with the decomposition

$$\hat{R}_N^{(1)*}(t) = R_N^*(t) + S_N^*(t) + W_N^*(t) + V_N^*(t) \tag{5.46}$$

where the processes on the right are defined by

$$R_N^*(t) := \frac{1}{N}\sum_{j=1}^{n_1} \varepsilon_{1j}^* r(X_{1j}) I\{X_{1j} \le t\} - \frac{1}{N}\sum_{j=1}^{n_2} \varepsilon_{2j}^* r(X_{2j}) I\{X_{2j} \le t\} \tag{5.47}$$

$$S_N^*(t) := \sum_{i=1}^{2} \frac{1}{N}\sum_{j=1}^{n_i} \varepsilon_{ij}^* \left( \frac{1}{Nh}\sum_{\ell=1}^{2}(-1)^\ell \sum_{k=1}^{n_\ell} K\Big(\frac{X_{ij}-X_{\ell k}}{h}\Big) I\{X_{\ell k} \le t\}\right) \tag{5.48}$$

$$W_N^*(t) := \sum_{\ell=1}^{2}(-1)^{\ell-1}\sum_{i=1}^{2}\frac{1}{N^2 h}\sum_{j=1}^{n_\ell}\sum_{k=1}^{n_i} K\Big(\frac{X_{\ell j}-X_{ik}}{h}\Big)(\widehat{f}_g(X_{\ell j}) - \widehat{f}_g(X_{ik})) I\{X_{\ell j} \le t\} \tag{5.49}$$

$$V_N^*(t) := \sum_{i=1}^{2}(-1)^{i-1}\frac{1}{N}\sum_{j=1}^{n_i} \varepsilon_{ij}^*(\hat{r}_h(X_{ij}) - r(X_{ij})) I\{X_{ij} \le t\}. \tag{5.50}$$

We will prove at the end of this section the following result.

**Lemma 5.4.** *If the assumptions of Theorem 2.2 and (4.6) are satisfied we have for all $\delta > 0$*

$$I\!P\Big(\sqrt{N}\sup_{t\in[0,1]} |V_N^*(t)| > \delta \,\Big|\, \mathcal{Y}_N\Big) = o_p(1) \tag{5.51}$$

$$I\!P\Big(\sqrt{N}\sup_{t\in[0,1]} |S_N^*(t) - S_N'^*(t)| > \delta \,\Big|\, \mathcal{Y}_N\Big) = o_p(1) \tag{5.52}$$

$$I\!P\Big(\sqrt{N}\sup_{t\in[0,1]} |W_N^*(t)| > \delta \,\Big|\, \mathcal{Y}_N\Big) = o_p(1). \tag{5.53}$$

*where the process $S_N'^*$ is defined by*

$$S_N'^*(t) := \frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_1} \varepsilon_{ij}^* \left(\frac{1}{h}\int_0^t K\Big(\frac{X_{ij}-x}{h}\Big)(-\kappa_1 r_1(x) + \kappa_2 r_2(x))\, dx\right). \tag{5.54}$$

Observing Lemma 5.4 it follows that the processes

$$T_N^{(1)*} := \sqrt{N}(R_N^* + S_N'^*)$$

and $\sqrt{N}\hat{R}_N^{(1)*}$ are (conditionally on $\mathcal{Y}_N$) asymptotically equivalent in probability, i.e.

$$I\!P\Big(\sup_{t\in[0,1]} |\sqrt{N}\hat{R}_N^{(1)*}(t) - T_N^{(1)*}(t)| > \delta \Big| \mathcal{Y}_N\Big) = o_p(1). \tag{5.55}$$

The following lemma shows that $T_N^{(1)*}$ in (5.55) can be replaced by

$$T_N'(\cdot) := \sum_{i=1}^{2}\frac{1}{\sqrt{N}}\sum_{j=1}^{n_i} \Delta_{ij}(\cdot) V_{ij}\varepsilon_{ij} \tag{5.56}$$

where the quantities $\Delta_{ij}$ are defined in (5.8).

**Lemma 5.5.** *If the assumptions of Theorem 2.2 and (4.6) are satisfied we have*

$$\mathbb{P}\Big( \sup_{t\in[0,1]} |T_N^{(1)*}(t) - T_N'(t)| > \delta \,\Big|\mathcal{Y}_N \Big) \;=\; o_p(1). \tag{5.57}$$

The assertion of Theorem 4.1 now follows from (5.57) and (5.55) which demonstrate that it is sufficient to consider the asymptotic behaviour of the process $T_N'(\cdot)$ defined in (5.56). But this process can be treated with the conditional multiplier theorem in Section 2.9 of van der Vaart and Wellner (1996), which establishes that conditionally on $\mathcal{Y}_N$ the process $T_N'$ converges to the same Gaussian process $Z^{(1)}$ in probability as the process $T_N$ discussed in the proof of Theorem 2.2. The proof of Theorem 4.1 is now concluded giving some more details for the proof of the auxiliary results in Lemma 5.4 and 5.5.

**Proof of Lemma 5.4.** For a proof of (5.51) we show

$$Z_N \,:=\, \sqrt{N} \sup_{t\in[0,1]} |V_N^*(t)| = o_p(1), \tag{5.58}$$

the assertion is then obvious from Markov's inequality, i.e.

$$\mathbb{P}\Big( \mathbb{P}\Big(Z_N > \delta \,\Big|\mathcal{Y}_N\Big) > \varepsilon \Big) \;\le\; \frac{1}{\varepsilon}\, E\Big[\mathbb{P}\Big(Z_N > \delta \,\Big|\mathcal{Y}_N\Big)\Big] \;=\; \frac{1}{\varepsilon}\, \mathbb{P}\Big(Z_N > \delta\Big) \;=\; o(1).$$

To this end we note that $\varepsilon_{ij}^* = V_{ij}\hat{\varepsilon}_{ij} = V_{ij}\varepsilon_{ij}\sigma_i(X_{ij}) + V_{ij}(f(X_{ij}) - \widehat{f}_g(X_{ij}))$ and obtain the decomposition

$$V_N^* \;=\; V_N^{*(1)} + V_N^{*(2)} \tag{5.59}$$

where

$$V_N^{*(1)}(t) \;=\; \frac{1}{N}\sum_{i=1}^{2}(-1)^{i-1}\sum_{j=1}^{n_i} V_{ij}\varepsilon_{ij}\sigma_i(X_{ij})(\hat{r}_h(X_{ij}) - r(X_{ij}))I\{X_{ij} \le t\} \tag{5.60}$$

$$V_N^{*(2)}(t) \;=\; \frac{1}{N}\sum_{i=1}^{2}(-1)^{i-1}\sum_{j=1}^{n_i} V_{ij}(f(X_{ij}) - \widehat{f}_g(X_{ij}))(\hat{r}_h(X_{ij}) - r(X_{ij}))I\{X_{ij} \le t\}. \tag{5.61}$$

The term in (5.60) can be treated by the same arguments given in the proof of Lemma 5.3 for the term $V_N(\cdot)$ (note that the only difference is the additional factor $V_{ij}$) which gives

$$\sqrt{N} \sup_{t\in[0,1]} |V_N^{*(1)}(t)| \;=\; o_p(1). \tag{5.62}$$

For the second term we use Cauchy's inequality and obtain

$$\begin{aligned} E\Big[\sup_{t\in[0,1]} |V_N^{*(2)}(t)|\Big] \;&\le\; \sum_{i=1}^{2}\frac{1}{N}\sum_{j=1}^{n_i} E|V_{ij}| \cdot \Big(E\Big[(f(X_{ij}) - \widehat{f}_g(X_{ij}))^2\Big] \cdot E\Big[(\hat{r}_h(X_{1j}) - r(X_{1j}))^2\Big]\Big)^{1/2} \\ &=\; O\Big(\frac{1}{N\sqrt{gh}}\Big) \;=\; o(\frac{1}{\sqrt{N}}), \end{aligned}$$

which yields in combination with (5.62) the assertion (5.58) and completes the proof of the first part of Lemma 5.4.

For a proof of the estimate (5.52) recall the definition of $S'^{*}_N$ in (5.54) and observe

$$S^*_N - S'^*_N = S^{*(1)}_N + S^{*(2)}_N$$

where

$$\begin{aligned} S^{*(1)}_N(t) &:= \sum_{i=1}^{2} \frac{1}{N} \sum_{j=1}^{n_i} V_{ij}\varepsilon_{ij}\sigma_i(X_{ij}) \Big[ \frac{-1}{Nh} \sum_{k=1}^{n_1} K\Big(\frac{X_{ij}-X_{1k}}{h}\Big) I\{X_{1k} \le t\} \\ &\quad + \frac{1}{Nh} \sum_{k=1}^{n_2} K\Big(\frac{X_{ij}-X_{2k}}{h}\Big) I\{X_{2k} \le t\} - \frac{1}{h}\int_0^t K\Big(\frac{X_{ij}-x}{h}\Big)(-\kappa_1 r_1(x) + \kappa_2 r_2(x))\,dx \Big] \\ S^{*(2)}_N(t) &:= \sum_{i=1}^{2} \frac{1}{N} \sum_{j=1}^{n_i} V_{ij}(f(X_{ij}) - \widehat{f}_g(X_{ij})) \Big[ \frac{-1}{Nh} \sum_{k=1}^{n_1} K\Big(\frac{X_{ij}-X_{1k}}{h}\Big) I\{X_{1k} \le t\} \\ &\quad + \frac{1}{Nh} \sum_{k=1}^{n_2} K\Big(\frac{X_{ij}-X_{2k}}{h}\Big) I\{X_{2k} \le t\} - \frac{1}{h}\int_0^t K\Big(\frac{X_{ij}-x}{h}\Big)(-\kappa_1 r_1(x) + \kappa_2 r_2(x))\,dx \Big]. \end{aligned}$$

The first term can be treated as in the proof of Lemma 5.2, which yields

$$\sqrt{N} \sup_{t\in[0,1]} |S^{*(1)}_N(t)| \;=\; o_p(1). \tag{5.63}$$

The second term is estimated as follows

$$\sup_{t\in[0,1]} |S^{*(2)}_N(t)| \;\le\; \sum_{i=1}^{2} \frac{1}{N} \sum_{j=1}^{n_i} |V_{ij}| \cdot | f(X_{ij}) - \widehat{f}_g(X_{ij})| \left\{ U^{(1)}_{Nij} + U^{(2)}_{Nij} \right\} \tag{5.64}$$

where

$$U^{(\ell)}_{Nij} \;=\; \frac{1}{h} \sup_{t\in[0,1]} \Big| \frac{1}{N} \sum_{k=1}^{n_\ell} K\Big(\frac{X_{ij}-X_{\ell k}}{h}\Big) I\{X_{\ell k} \le t\} - \int_0^t K\Big(\frac{X_{ij}-z}{h}\Big)\kappa_\ell r_\ell(z)\,dz \Big| \;, \qquad \ell = 1,2 \;.$$

The terms $U^{(\ell)}_{Nij}$ $(i,\ell = 1,2)$ can be treated by Theorem 37 in Pollard (1984). More precisely, for the first term we note

$$\sup_{t,x\in[0,1]} \Big| \frac{1}{n_1} \sum_{k=1}^{n_1} K\Big(\frac{x-X_{1k}}{h}\Big) I\{X_{1k} \le t\} - \int_0^t K\Big(\frac{x-z}{h}\Big) r_1(z)\,dz \Big| \;=\; \sup_{\varphi\in\mathcal{F}_{n_1}} |P_{n_1}\varphi - P\varphi|$$

where $P_{n_1}$ denotes the empirical distribution of the first sample $X_{11},\dots,X_{1n_1}$ and

$$\mathcal{F}_{n_1} = \left\{ \varphi_{h_{n_1},t,x} \mid \varphi_{h_{n_1},t,x}(y) = K\Big(\frac{x-y}{h_{n_1}}\Big) I\{y \le t\}; x,t \in [0,1] \right\}$$

(note that we made the dependency of the bandwidth on the sample size explicit, i.e. $h = h_{n_1}$). Now $\mathcal{F}_{n_1}$ is a subset of a VC-class and the arguments used in the Theorem 37 of Pollard (1984) yield for the sequences

$$\alpha_{n_1} = \sqrt{g} \ , \qquad \delta_{n_1}^2 = c \cdot h_{n_1},$$

the estimate

$$U_{nij}^{(1)} \leq \frac{1}{h_{n_1}} \sup_{\varphi \in \mathcal{F}_{n_1}} |P_{n_1}\varphi - P\varphi| = \frac{1}{h_{n_1}} o_p(\delta_{n_1}^2 \alpha_{n_1}) \ = \ o_p(\sqrt{g}).$$

By a similar argument for the terms $U_{Nij}^{(2)}$ (5.64) simplifies to

$$\sup_{t \in [0,1]} |S_N^{*(2)}(t)| \ \leq \ o_p(\sqrt{g}) \cdot \sum_{i=1}^{2} \frac{1}{N} \sum_{j=1}^{n_i} |V_{ij}| \cdot | f(X_{ij}) - \widehat{f}_g(X_{ij})| \ = \ o_p(\frac{1}{\sqrt{N}})$$

where the last estimate follows from Markov's inequality. A combination of this estimate with (5.63) gives

$$\sqrt{N} \sup_{t \in [0,1]} |S_N^*(t) - S_N'^*(t)| \ = \ o_p(1)$$

and the assertion (5.52) follows again from Markov's inequality.

$\square$

**Proof of Lemma 5.5.** Defining $(i = 1, 2)$

$$(5.65) \quad \tilde{\Delta}_{ij}(t) \ := \ ( -1)^{i-1} r(X_{ij}) I\{X_{ij} \leq t\} + \frac{1}{h} \int_0^t K\Big(\frac{X_{ij} - x}{h}\Big)(-\kappa_1 r_1(x) + \kappa_2 r_2(x))\, dx$$

and recalling the definition of $T_N'$ in (5.56) we obtain

$$\begin{aligned} T_N^{(1)*}(t) - T_N'(t) \ &= \ \sum_{i=1}^{2} \frac{1}{\sqrt{N}} \sum_{j=1}^{n_i} \tilde{\Delta}_{ij}(t) V_{ij}(f(X_{ij}) - \widehat{f}_g(X_{ij})) \\ &= \ \sum_{i=1}^{2} \frac{1}{\sqrt{N}} \sum_{j=1}^{n_i} \tilde{\Delta}_{ij}(t) V_{ij}(f(X_{ij}) - \widehat{f}_g(X_{ij})) \frac{1}{r(X_{ij})}(r(X_{ij}) - \hat{r}_g(X_{ij})) \\ &\quad + \sum_{i=1}^{2} \frac{1}{\sqrt{N}} \sum_{j=1}^{n_i} \tilde{\Delta}_{ij}(t) V_{ij}(f(X_{ij}) - \widehat{f}_g(X_{ij})) \frac{\hat{r}_g(X_{ij})}{r(X_{ij})} \\ &= \ A_N(t) \ + \ B_N(t) \end{aligned} \tag{5.66}$$

[note that $\Delta_{ij}(t) = \tilde{\Delta}_{ij}(t)\sigma_i(X_{ij})$, by the definition of $\Delta_{ij}$ in (5.8)]. The first term is estimated as follows

$$\begin{aligned} \sup_{t \in [0,1]} |A_N(t)| \ \leq \quad & \sum_{i=1}^{2} \frac{1}{\sqrt{N}} \sum_{j=1}^{n_i} \sup_{t \in [0,1]} |\tilde{\Delta}_{ij}(t)| \frac{1}{r(X_{ij})} |f(X_{ij}) - \widehat{f}_g(X_{ij})| \cdot | r(X_{ij}) - \hat{r}_g(X_{ij})| \\ = \ & O_p\Big(\frac{1}{\sqrt{N} g}\Big) \ = \ o_p(1) \end{aligned}$$

where we used Cauchy's inequality and the fact that $\tilde{\Delta}_{ij}(\cdot)$ is uniformly bounded. Now Markov's inequality yields conditionally on the sample $\mathcal{Y}_N$

$$\sup_{t\in[0,1]} |A_N(t)| = o_p(1). \tag{5.67}$$

The second term $B_N(t)$ in (5.66) consists of expressions of the form

$$\begin{aligned} \tilde{B}_N(t) &:= \frac{1}{n_1\sqrt{N}}\sum_{j=1}^{n_1}\sum_{k=1}^{n_1}\tilde{\Delta}_{1j}(t)\frac{1}{g}K\Big(\frac{X_{1j}-X_{1k}}{g}\Big)(f(X_{1j})-f(X_{1k}))V_{1j}\frac{1}{r(X_{1j})} \\ &+ \frac{1}{n_1\sqrt{N}}\sum_{j=1}^{n_1}\sum_{k=1}^{n_1}\tilde{\Delta}_{1j}(t)\frac{1}{g}K\Big(\frac{X_{1j}-X_{1k}}{g}\Big)\varepsilon_{1k}\sigma_1(X_{1k})V_{1j}\frac{1}{r(X_{1j})} \end{aligned} \tag{5.68}$$

which are all treated similary. We obtain

$$\tilde{B}_N(t) = \sum_{\ell=1}^{4} I_\ell(t) \tag{5.69}$$

where

$$\begin{aligned} I_1(t) &:= \frac{1}{n_1\sqrt{N}}\sum_{j=1}^{n_1}\sum_{k=1}^{n_1} I\{X_{1j}\le t\}\frac{1}{g}K\Big(\frac{X_{1j}-X_{1k}}{g}\Big)(f(X_{1j})-f(X_{1k}))V_{1j} \\ I_2(t) &:= \frac{1}{n_1\sqrt{N}}\sum_{j=1}^{n_1}\sum_{k=1}^{n_1} I\{X_{1j}\le t\}\frac{1}{g}K\Big(\frac{X_{1j}-X_{1k}}{g}\Big)\varepsilon_{1k}\sigma_1(X_{1k})V_{1j} \\ I_3(t) &:= \frac{1}{n_1\sqrt{N}}\sum_{j=1}^{n_1}\sum_{k=1}^{n_1}\frac{1}{h}\int_0^t K\Big(\frac{X_{1j}-x}{h}\Big)(-\kappa_1 r_1(x)+\kappa_2 r_2(x))\,dx \\ &\qquad \frac{1}{g}K\Big(\frac{X_{1j}-X_{1k}}{g}\Big)(f(X_{1j})-f(X_{1k}))V_{1j}\frac{1}{r(X_{1j})} \\ I_4(t) &:= \frac{1}{n_1\sqrt{N}}\sum_{j=1}^{n_1}\sum_{k=1}^{n_1}\frac{1}{h}\int_0^t K\Big(\frac{X_{1j}-x}{h}\Big)(-\kappa_1 r_1(x)+\kappa_2 r_2(x))\,dx \\ &\qquad \frac{1}{g}K\Big(\frac{X_{1j}-X_{1k}}{g}\Big)\varepsilon_{1k}\sigma_1(X_{1k})V_{1j}\frac{1}{r(X_{1j})}. \end{aligned}$$

The processes $I_1(\cdot)$ and $I_2(\cdot)$ are treated as in the proof of Lemma 5.3a writing $I_\ell(t)$ as one sample $U$-process $\frac{1}{g}U_N(\varphi)$ indexed by an euclidean class of functions which gives

$$\sup_{t\in[0,1]} |I_\ell(t)| = o_p(1), \quad \ell = 1\ ,2. \tag{5.70}$$

Similarly we have $I_4(t) = \frac{1}{gh}U_N(\varphi)$, where $\varphi$ is the symmetric kernel defined by

$$\begin{aligned} \varphi(\xi_j,\xi_k) &:= \frac{1}{2}\int_0^t K\Big(\frac{X_{1j}-x}{h}\Big)\frac{(-\kappa_1 r_1(x)+\kappa_2 r_2(x))}{r(X_{1j})}\,dx\cdot K\Big(\frac{X_{1j}-X_{1k}}{g}\Big)\varepsilon_{1k}\sigma_1(X_{1k})V_{1j} \\ &+\frac{1}{2}\int_0^t K\Big(\frac{X_{1k}-x}{h}\Big)\frac{(-\kappa_1 r_1(x)+\kappa_2 r_2(x))}{r(X_{1k})}\,dx\cdot K\Big(\frac{X_{1j}-X_{1k}}{g}\Big)\varepsilon_{1j}\sigma_1(X_{1j})V_{1k} \end{aligned} \tag{5.71}$$

and $\xi_j = (X_j, \varepsilon_{1j}, V_{1j})$. A straightforward but tedious calculation shows that the class $\mathcal{F}$ of functions defined by (5.71) is euclidean which gives

$$\sup_{t\in[0,1]} |I_4(t)| = o_p(1).$$

By a similar argument for the process $I_3(\cdot)$ and (5.70) we obtain from (5.69) $\sup_{t\in[0,1]} |\tilde{B}_N(t)| = o_p(1)$. The remaining terms in $B_N(t)$ are treated exactly in the same way, and it follows

$$\sup_{t\in[0,1]} |B_N(t)| = o_p(1)$$

and the assertion of Lemma 5.5 follows from (5.66), (5.67) and Markov's inequality.

□

**Acknowledgements.** The authors are grateful to Isolde Gottschlich who typed most parts of this paper with considerable technical expertise and to P. Cabus for sending us her work before publication. The authors would also like to thank M. Delgado and W. Stute for some help regarding the literature on empirical processes and wild bootstrap. The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, Reduction fo complexity in multivariate data structures) is gratefully acknowledged.

# References

H.S. An, C. Bing (1991). A Kolmogorov-Smirnov type statistic with application to test for non-linearity in time series. Internat. Statist. Rev. 59, 287-307.

P. Billingsley (1968). Convergence of Probability Measures. Wiley, New York.

P. Cabus (2000). Testing for the comparison of non-parametric regression curves. Preprint 99-29, IRMAR, University of Rennes, France.

M.A. Delgado (1993). Testing the equality of nonparametric regression curves. Statist. Prob. Letters 17, 199-204.

M.A. Delgado, W. González Manteiga (1998). Significance testing in nonparametric regression based on the bootstrap. Working paper 98-68 (31), Department of Statistics, Universidad Carlos III de Madrid.

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

H. Dette, N. Neumeyer (1999). Nonparametric analysis of covariance. Technical Report 262, Department of Mathematics, Ruhr-Universität Bochum, Germany (http://www.ruhr-uni-bochum.de/mathematik3/preprint.htm).

J. Fan (1992). Design-adaptive nonparametric regression. J. Amer. Stat. Soc. 87, 998-1004.

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

T. Gasser, H.-G. Müller, V. Mamitzsch (1985). Kernels for nonparametric curve estimation. J. Roy. Statist. Soc., Ser. B, 47, 238-252.

W. Härdle, J.S. Marron (1990). Semiparametric comparison of regression curves . Ann. Statist. 18, 63-89.

P. Hall, C. Huber, P.L. Speckman (1997). Covariate-matched one-sided tests for the difference between functional means. J. Amer. Statist. Assoc. 92, 1074-1083.

P. Hall, J.D. Hart (1990). Bootstrap test for difference between means in nonparametric regression. J. Amer. Statist. Assoc. 85, 1039-1049.

V. Hjellvik, D. Tjøstheim (1995). Nonparametric tests of linearity for time series. Biometrika 77, 351-368.

E.C. King, J.D. Hart, T.E. Wehrly (1991). Testing the equality of regression curves using linear smoothers. Statist. Probab. Lett. 12, 239-247.

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

K.B. Kulasekera, J. Wang (1997). Smoothing parameter selection for power optimality in testing regression curves. J. Amer. Statist. Assoc. 92, 500-511.

E.A. Nadaraya (1964). On estimating regression. Theory Probab. Appl., 10, 186-190.

D. Nolan, D. Pollard (1987). $U$-processes: rates of convergence. Ann. Statist. 15, 780-799.

D. Nolan, D. Pollard (1988). Functional limit theorems for $U$-process. Ann. Prob. 16, 1291-1298.

D. Pollard (1984). Convergence of Stochastic Processes. Springer, New York.

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

G.R. Shorack, J.A. Wellner (1986). Empirical Processes with Applications to Statistics. Wiley, New York.

W. Stute (1997). Nonparametric model checks for regression. Ann. Statist. 25, 613-641.

A.W. van der Vaart, J.A. Wellner (1996). Weak Convergence and Empirical processes. Springer, New York.

M.P. Wand, M.C. Jones (1995). Kernel Smoothing. Chapman and Hall, London.

G.S. Watson (1964). Smooth regression analysis. Sankhya, Ser. A, 26, 359-372.

C.F.Y. Wu (1986). Jacknife, bootstrap and other resampling methods in regression analysis. Ann. Statist. 14, 1261-1350.

S.G. Young, A.W. Bowman (1995). Non-parametric analysis of covariance. Biometrics 51, 920-931.

J.X. Zheng (1996). A consistent test of functional form via nonparametric estimation techniques. J. of Econometrics 75, 263-289.