The Annals of Statistics
2003, Vol. 31, No. 3, 880–920

NONPARAMETRIC COMPARISON OF REGRESSION CURVES: AN EMPIRICAL PROCESS APPROACH[1]

BY NATALIE NEUMEYER AND HOLGER DETTE

Ruhr-Universität

We propose a new test for the comparison of two regression curves that is based on a difference of two marked empirical processes based on residuals. The large sample behavior 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 $N^{-1/2}$ and that, in contrast to all other available procedures based on marked empirical processes, the new test allows the optimal choice of bandwidths for curve estimation (e.g., $N^{-1/5}$ in the case of twice differentiable regression functions). As a by-product we explain the problems of a related test proposed by Kulasekera [*J. Amer. Statist. Assoc.* **90** (1995) 1085–1093] and Kulasekera and Wang [*J. Amer. Statist. Assoc.* **92** (1997) 500–511] 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 test and to compare it with the so far available procedures. Finally, heteroscedastic data is analyzed in order to demonstrate the benefits of the new test compared to the so far available procedures which require homoscedasticity.

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 with a sample of $N = n_1 + n_2$ observations,

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

where X_{ij} ($j = 1, \dots, n_i$) are independent observations with positive density r_i on the interval [0, 1] ($i = 1, 2$) and ε_{ij} are independent identically distributed random variables with mean 0 and variance 1. In (1.1) f_i and σ_i denote the regression and variance functions in the ith sample ($i = 1, 2$). In this paper we are interested in the problem of testing the equality of the mean functions, that is,

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

Received November 2000; revised July 2002.

[1]Supported in part by Deutsche Forschungsgemeinschaft SFB 415.

AMS 2000 subject classifications. Primary 62G05; secondary 60F15, 60F17.

Key words and phrases. Comparison of regression curves, goodness-of-fit, marked empirical process, *VC*-classes, *U*-processes.

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), Bowman and Young (1996), Hall, Huber and Speckman (1997), Munk and Dette (1998) or Dette and Neumeyer (2001)]. 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) and 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 first considered by Munk and Dette (1998) who considered the fixed design case 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 functions (i.e., differentiability is not required). Recently Dette and Neumeyer (2001) 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 and from a theoretical point of view.

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 Theorems 2.1 and 2.2 in Kulasekera (1995), because this author ignores the variability caused by the nonparametric estimation of the regression function in the application of Donsker's invariance principle. We present a correct version of Kulasekera's result in Section 3.

A different method was proposed in a recent paper by Cabus (2000), who used U-processes for the construction of a test statistic. However, her approach assumes knowledge of the variance function and design density and is therefore difficult to implement in practice. In order to avoid this drawback we investigate a wild bootstrap version of Cabus' test in Section 4. Finally, it is worthwhile to mention an important difference between our method and Kulasekera's (1995) and Cabus' (2000) approaches. The method proposed in this paper allows "optimal" choices of the bandwidth (e.g., $N^{-1/5}$ in the case of twice differentiable functions) and therefore avoids undersmoothing of the regression estimate [see, e.g., assumption $\mathcal{A}_2$ in Kulasekera (1995)]. This theoretical advantage is achieved by the application of a difference of two marked empirical processes, which basically yields a cancellation of lower order bias terms. As a consequence we obtain a better approximation of the nominal level from a theoretical point of view, which is reflected for finite samples in a simulation study presented in Section 4.

The remaining parts of the paper are organized as follows. Section 2 introduces the marked empirical processes and the corresponding test statistics and gives their asymptotic behavior. Some comments regarding the test of Kulasekera (1995) and a clarification of its asymptotic properties are given in Section 3. The finite sample behavior 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 version of the test proposed in this paper and discusses a data example. In this example we demonstrate the potential benefits of our approach by reanalyzing heteroscedastic data which were analyzed previously with testing procedures assuming homoscedasticity. Finally, all proofs are deferred to Section 5.

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 density r_i on the interval $[0, 1]$ such that $r_i(x) \geq c > 0$ for all $x \in [0, 1]$ $(i = 1, 2)$. The regression functions

f_1, f_2 and the densities r_1, r_2 are supposed to be d (≥ 2) times continuously differentiable; that is,

$$r_i, f_i \in C^d([0, 1]), \qquad i = 1, 2. \tag{2.1}$$

Throughout this paper let

$$\widehat{r}_i(x) = \frac{1}{n_i h} \sum_{j=1}^{n_i} K\left(\frac{x - X_{ij}}{h}\right) \tag{2.2}$$

denote the density estimator from the ith sample ($i = 1, 2$) and

$$\widehat{r}(x) = \frac{n_1}{N} \widehat{r}_1(x) + \frac{n_2}{N} \widehat{r}_2(x)$$

the density estimator from the combined sample $X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2}$. For the sake of transparency we assume an equal bandwidth h for all estimators satisfying

$$h \to 0, \qquad N h^{4d} \to 0, \qquad N h^2 \to \infty, \tag{2.3}$$

but we note that all results of this paper can be generalized to the case of different bandwidths satisfying (2.3). In (2.2) the function K is a symmetric kernel with compact support of order $d \geq 2$; that is,

$$\int K(u) u^j \, du \quad \begin{cases} = 1, & j = 0, \\ = 0, & 1 \leq j \leq d - 1, \\ \neq 0, & j = d \end{cases} \tag{2.4}$$

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

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

$(0 = a_0 < a_1 < \cdots < a_{m-1} < a_m = \infty)$ such that for some $\varepsilon \in \{-1, 1\}$ the function εK 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

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

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

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

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

$$
\begin{aligned}
\widehat{f}(x) &= \frac{1}{Nh}\sum_{i=1}^{2}\sum_{j=1}^{n_i} K\left(\frac{x-X_{ij}}{h}\right) Y_{ij}\frac{1}{\widehat{r}(x)} \\
&= \frac{(n_1/N)\widehat{r}_1(x)\widehat{f}_1(x) + (n_2/N)\widehat{r}_2(x)\widehat{f}_2(x)}{\widehat{r}(x)}
\end{aligned}
\tag{2.7}
$$

and consistently estimates

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

where

$$
\widehat{f}_i(x) = \frac{1}{n_i h}\sum_{j=1}^{n_i} K\left(\frac{x-X_{ij}}{h}\right) Y_{ij}\frac{1}{\widehat{r}_i(x)}
$$

is the curve estimator from the ith sample ($i = 1,2$). Note that under the null hypothesis of equal regression curves we have $f_1 = f_2 = f$. For $i = 1,2$ we define residuals

$$
e_{ij} = \frac{n_{3-i}}{N}\big(Y_{ij} - \widehat{f}(X_{ij})\big)\widehat{r}(X_{ij})\widehat{r}_{3-i}(X_{ij}),
\tag{2.8}
$$

$$
f_{ij} = \frac{N}{n_i}\big(Y_{ij} - \widehat{f}(X_{ij})\big)/\widehat{r}_i(X_{ij}),
\tag{2.9}
$$

and consider the marked empirical processes

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

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

where $t \in [0,1]$ and $I\{\cdot\}$ denotes the indicator function. The multiplication of the residuals by the density estimators $\widehat{r}(x)\widehat{r}_{3-i}(x)$ and $1/\widehat{r}_i(x)$ is motivated by a cancellation of the lower order bias terms in the marked empirical processes under the null hypothesis of equal regression curves (see the following Proposition 2.1 and its proof and Theorem 2.2). The form of $\widehat{R}_N^{(2)}$ is attractive because it essentially 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). As pointed out by a referee the residual f_{ij} is not defined in the case $\widehat{r}_i(X_{ij}) = 0$, which may occur with positive probability if the kernel K also attains negative values. In such cases the residuals f_{ij} should be multiplied by the factor $I\{\widehat{r}_i(X_{ij}) \neq 0\}$. Because the densities satisfy $r_i(x) \geq c > 0$, these factors converge uniformly to 1 and the effect of the multiplication is asymptotically negligible. 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 Section 5.

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

(i) *Under the null hypothesis of equal regression curves we have*

$$E[\widehat{R}_N^{(\ell)}(t)] = O\left(\frac{1}{Nh}\right) + O(h^{2d}) = o\left(\frac{1}{\sqrt{N}}\right), \qquad \ell = 1, 2.$$

(ii) *Under the alternative of unequal regression curves we have*

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

$$E[\widehat{R}_N^{(2)}(t)] = \kappa_1\kappa_2 \int_0^t \big(f_1(x) - f_2(x)\big)\, dx + O(h^d).$$

Note that

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

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 ($i = 1, 2$)

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

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 $\widehat{R}_N^{(1)}$ and $\widehat{R}_N^{(2)}$ 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* $\sqrt{N}\widehat{R}_N^{(1)}$ *defined by* (2.10) *converges weakly to a centered Gaussian process* $Z^{(1)}$ *in the space* $D[0, 1]$ *with covariance function*

$$\begin{aligned} &H^{(1)}(s, t) \\ &\quad = \int_0^{s\wedge t} \big(\sigma_1^2(x)\kappa_2 r_2(x) + \sigma_2^2(x)\kappa_1 r_1(x)\big)\kappa_1 r_1(x)\kappa_2 r_2(x) r^2(x)\, dx. \end{aligned} \tag{2.12}$$

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

$$H^{(2)}(s,t) = \int_0^{s\wedge t} \big(\sigma_1^2(x)\kappa_2 r_2(x) + \sigma_2^2(x)\kappa_1 r_1(x)\big)\frac{1}{\kappa_1 r_1(x)\kappa_2 r_2(x)}\,dx. \tag{2.13}$$

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 mathematical complexity. Note that local polynomial estimators have various practical and theoretical advantages such as better boundary behavior 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, e.g., the proof of Lemma 5.3a]. 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 and the data example 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 $\sqrt{N}\widehat{R}_N^{(i)}(\cdot)$ $(i = 1, 2)$ converge weakly to Gaussian processes with respective covariance kernels $H^{(i)}(\cdot,\cdot)$ given in Theorem 2.2 and means

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

respectively. Another approach to obtain tests which achieve nontrivial power for $N^{-1/2}$ distant alternatives is based on comparisons of smoothers with fixed bandwidths. These tests will usually not detect alternatives where the regression functions differ by an oscillating function. On the other hand one can construct examples where nonparametric tests based on comparison of local smoothers perform better than tests based on marked empirical processes [see Dette and Neumeyer (2001)]. Because the main application of the test is to answer the question if it is more appropriate to fit the two regression functions by one fit obtained from the pooled sample or by separately smoothing the two samples, some care is necessary with the interpretation of the corresponding p-value. In practice the statement of the resulting p-value should always be accompanied by a corresponding fit of the data under the null hypothesis and the alternative. Of course this remark applies to all goodness-of-fit tests proposed in the literature for comparing regression curves. The corresponding curve estimators are already used in the construction of the processes $\widehat{R}_N^{(1)}$ and $\widehat{R}_N^{(2)}$ and therefore easily available if a test is based on these processes.

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

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

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

$$\begin{aligned} e_{j\ell}^{(i)} &= \big(Y_{j\ell} - \widehat{f}^{(i)}(X_{j\ell})\big)\widehat{r}^{(i)}(X_{j\ell}) \\ &\quad \times \left(\frac{n_i}{n_i + n_{i+1}}\widehat{r}_i(X_{j\ell})I_{\{j=i+1\}} + \frac{n_{i+1}}{n_i + n_{i+1}}\widehat{r}_{i+1}(X_{j\ell})I_{\{j=i\}}\right) \end{aligned}$$

$(i = 1, \dots, k-1,\ j \in \{i, i+1\},\ \ell \in \{1, \dots, n_j\})$ where $\widehat{f}^{(i)}$ and $\widehat{r}^{(i)}$ denote the Nadaraya–Watson and the density estimators from the combined ith and $(i+1)$st samples. If $N = \sum_{i=1}^k n_i$ denotes the total sample size, $\frac{n_i}{N} = \kappa_i + O(\frac{1}{N})$ $[\kappa_i \in (0,1);\ i = 1, \dots, k]$ and

$$\begin{aligned} \widehat{R}_{Ni}^{(1)} &= \frac{1}{n_i + n_{i+1}} \sum_{\ell=1}^{n_i} e_{i\ell}^{(i)} I\{X_{i\ell} \le t\} \\ &\quad - \frac{1}{n_i + n_{i+1}} \sum_{\ell=1}^{n_{i+1}} e_{i+1,\ell}^{(i)} I\{X_{i+1,\ell} \le t\}, \qquad i = 1, \dots, k-1, \end{aligned}$$

then it follows that $\sqrt{N}\widehat{R}_N^{(1)}(t) := \sqrt{N}(\widehat{R}_{N1}^{(1)}(t), \dots, \widehat{R}_{Nk-1}^{(1)}(t))^T$ converges weakly to a $(k-1)$-dimensional Gaussian process $(Z_1^{(1)}, \dots, Z_{k-1}^{(1)})^T$ with covariance structure

$$\operatorname{Cov}\big(Z_i^{(1)}(t), Z_j^{(1)}(s)\big) = k_{ij}(s \wedge t)$$

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

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

where

$$r^{(i)}(x) = \frac{\kappa_i}{\kappa_i + \kappa_{i+1}} r_i(x) + \frac{\kappa_{i+1}}{\kappa_i + \kappa_{i+1}} r_{i+1}(x), \qquad i = 1, \dots, k.$$

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) and 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 (2001), 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), Kulasekera (1995) and 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 $\widehat{R}_N^{(2)}$ reduces in this case essentially 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 and in the following section that these deficiencies are on the one hand caused by the use of incorrect (asymptotic) critical values and on the other hand by a nonnegligible bias in the calculated residuals for finite samples.

To be precise, consider the model (1.1) in the case of a fixed design $X_{ij} = t_{ij}$ ($j = 1, \dots, n_i$; $i = 1, 2$) satisfying a Sacks and Ylvisaker (1970) condition,

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

let $\widehat{f}_i$ denote the Nadaraya–Watson estimator from the ith sample ($i = 1, 2$) and define residuals by

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

The corresponding partial sums are given by

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

and the following result specifies the asymptotic distribution of these marked empirical processes. For the appropriate asymptotic statement we require a different bandwidth condition,

$$h \to 0, \qquad Nh^{2d} \to 0, \qquad Nh^2 \to \infty. \tag{3.3}$$

THEOREM 3.1. *If the assumptions* (2.1), (2.4), (2.6), (3.1) *and* (3.3) *are satisfied, then under the null hypothesis of equal regression curves the marked empirical process μ_1 defined in* (3.2) *converges weakly to a centered Gaussian process with covariance function*

$$m_{12}(s,t) = \int_0^{R_1^{-1}(s\wedge t)} \bigl(\sigma_1^2(x)\kappa_2 r_2(x) + \sigma_2^2(x)\kappa_1 r_1(x)\bigr)\frac{r_1(x)}{\kappa_2 r_2(x)}\,dx, \tag{3.4}$$

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 μ_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 μ_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 $m_{i,3-i}(s,t)$. For this reason some care is necessary if the test of Kulasekera 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

$$\lambda_N^{(i)}(t) = \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} \bigl(Y_{ij} - \widehat{f}_{3-i}(X_{ij})\bigr) I\{X_{ij} \le t\}, \qquad i = 1, 2, \tag{3.5}$$

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

THEOREM 3.2. *If the assumptions* (2.1), (2.4), (2.6) *and* (3.3) *are satisfied, then under the null hypothesis of equal regression curves the marked empirical process $\lambda_N^{(1)}$ defined by* (3.5) *converges weakly to a centered Gaussian process with covariance function $m_{12}(R_1(s), R_1(t))$ where m_{12} is defined in* (3.4) *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 R_2 is the distribution function of X_{2j}.*

REMARK 3.3. Note that in contrast to the new tests presented in Section 2, Kulasekera's (1995) method requires a regression estimate with bandwidth h satisfying $Nh^{2d} \to 0$ and the "optimal" choices for the bandwidth (e.g., $N^{-1/5}$ in the case of twice differentiable functions) cannot be used for this test [see, e.g., assumption $\mathcal{A}_2$ in Kulasekera (1995)]. A consequence of this undersmoothing of the regression estimate is a substantially less precise approximation of the nominal level for nonconstant regression functions if optimal bandwidths for curve estimation are used in these procedures. This loss of accuracy can also be observed for the corresponding bootstrap procedures (see our simulation study in Section 4).

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

$$U_N(t) = \frac{1}{n_1 n_2 h} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (Y_{1i} - Y_{2j}) K\left(\frac{X_{1i} - X_{2j}}{h}\right) I\{X_{1i} \le t,\ X_{2j} \le t\}. \tag{3.6}$$

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

$$\frac{1}{\kappa_1 \kappa_2} \int_0^{s \wedge t} \big(\sigma_1^2(x) \kappa_2 r_2(x) + \sigma_2^2(x) \kappa_1 r_1(x)\big) r_1(x) r_2(x)\, dx$$

and assumed knowledge of the variance function and design density for the construction of a test of equality of the regression functions. Using similar arguments as given in Section 5, the consistency of a wild bootstrap version of this test can be established. However, this procedure has the same problems with respect to its finite sample performance as mentioned for Kulasekera's approach (see Remark 3.3 and the simulation study in Section 4). It requires undersmoothing which results in a substantial loss in the approximation of the nominal level if the regression functions are not constant and optimal bandwidths for curve estimation are used (see the simulation study in the following section).

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

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

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 Theorems 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). The finite sample properties of the resulting tests are then investigated by means of a simulation study. To be precise, let $\widehat{f}_g(x)$ 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

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

and bootstrap residuals by

$$\varepsilon^*_{ij} := \widehat{\varepsilon}_{ij} V_{ij}, \tag{4.3}$$

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

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

We obtain the bootstrap sample

$$Y^*_{ij} := \widehat{f}_g(X_{ij}) + \varepsilon^*_{ij} \tag{4.5}$$

and the corresponding marked empirical processes

$$\begin{aligned} \widehat{R}_N^{(1)*}(t) = \frac{1}{N} \sum_{\ell=1}^{2} \sum_{j=1}^{n_\ell} (-1)^{\ell-1} \bigl(Y^*_{\ell j} - \widehat{f}^*_h(X_{\ell j})\bigr) \\ \times \widehat{r}_h(X_{\ell j}) \frac{n_{3-\ell}}{N} \widehat{r}_{3-\ell,h}(X_{\ell j}) I\{X_{\ell j} \le t\}, \end{aligned} \tag{4.6}$$

$$\widehat{R}_N^{(2)*}(t) = \frac{1}{N} \sum_{\ell=1}^{2} \sum_{j=1}^{n_\ell} (-1)^{\ell-1} \bigl(Y^*_{\ell j} - \widehat{f}^*_h(X_{\ell j})\bigr) \frac{N}{n_\ell} \frac{1}{\widehat{r}_{\ell,h}(X_{\ell j})} I\{X_{\ell j} \le t\}, \tag{4.7}$$

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 $\widehat{f}^*_h$ and $\widehat{r}_h$) 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 $K_N^{(i)} \ge k^{*(i)}_{N,1-\alpha}$, where $k^{*(i)}_{N,1-\alpha}$

denotes the critical value obtained from the bootstrap distribution, that is,

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

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 $\sqrt{N}\widehat{R}_N^{(i)}$ and $\sqrt{N}\widehat{R}_N^{(i)*}$ in probability conditionally on the sample $\mathcal{Y}_N$.

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

$$g \to 0, \qquad \sqrt{N} g h \to \infty, \qquad N g^{2d+1} = O(1), \tag{4.8}$$
$$\frac{g^{2d+1}}{h^{2d} \log^2 N} \to \infty, \qquad \frac{g^{2d}}{h} \to 0$$

are satisfied, then the bootstrapped marked empirical process $\sqrt{N}\widehat{R}_N^{(i)*}$ *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 the bootstrap version of the tests based on the approaches 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

$$L_N = \max\Big(\sup_{t \in [0,1]} |\lambda_N^{(1)}(t)|, \sup_{t \in [0,1]} |\lambda_N^{(2)}(t)| \Big), \tag{4.9}$$

where the processes $\lambda_N^{(1)}(\cdot)$ and $\lambda_N^{(2)}(\cdot)$ have been defined in (3.5). Similarly, we consider the statistic

$$C_N = \sup_{t \in [0,1]} |U_N(t)|, \tag{4.10}$$

where U_N is the process introduced by Cabus (2000) and defined by (3.6). The wild bootstrap version of these tests is essentially the same as explained in the previous paragraph.

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 considered the following scenario:

$$
\begin{aligned}
&\text{(i)} && f_1(x) = f_2(x) = 1,\\
&\text{(ii)} && f_1(x) = f_2(x) = \exp(x),\\
&\text{(iii)} && f_1(x) = f_2(x) = \sin(2\pi x),\\
&\text{(iv)} && f_1(x) = 1; \quad f_2(x) = 1 + x,\\
&\text{(v)} && f_1(x) = \exp(x); \quad f_2(x) = \exp(x) + x,\\
&\text{(vi)} && f_1(x) = \sin(2\pi x); \quad f_2(x) = \sin(2\pi x) + x,\\
&\text{(vii)} && f_1(x) = 1; \quad f_2(x) = 1 + \sin(2\pi x),\\
&\text{(viii)} && f_1(x) = \exp(x); \quad f_2(x) = \exp(x) + \sin(2\pi x),\\
&\text{(ix)} && f_1(x) = \sin(2\pi x); \quad f_2(x) = 2\sin(2\pi x),
\end{aligned}
\tag{4.11}
$$

where the first three cases correspond 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,

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

All bandwidths for the estimation from the combined and individual samples were chosen data adaptively. We investigated two selection rules. The first method is a plug-in method [see Gasser, Kneip and Köhler (1991)] leading to bandwidths with the optimal rate $h = cN^{-1/5}$. For the consideration of a second (simpler) method, we performed simulations for the simple rule of thumb,

$$h = \left\{ \frac{n_1 \widehat{\sigma}_2^2 + n_2 \widehat{\sigma}_1^2}{(n_1 + n_2)^2} \right\}^{1/5}, \qquad h_i = \left(\frac{\widehat{\sigma}_i^2}{n_i} \right)^{1/5}, \qquad i = 1, 2,$$

where $\widehat{\sigma}_i^2$ denotes the estimator of Rice (1984) for the integrated variance function $\int_0^1 \sigma_i^2(t) r_i(t)\,dt$ for the ith sample ($i = 1, 2$). The bandwidths in the bootstrap steps were chosen slightly larger, that is, $g = h^{5/6}$ (although this is not necessary for the asymptotic theory). The results for both selection procedures are not distinguishable and only the first case will be displayed.

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 Tables 1 and 2 for the statistics $K_N^{(1)}$, $K_N^{(2)}$, respectively, which show the relative proportion of rejections based on 1,000 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

TABLE 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* (4.11). *The errors are homoscedastic and have variances* $\sigma_1^2 = 0.5$, $\sigma_2^2 = 0.25$

	n_2:	25			50			100		
n_1	α:	**2.5%**	**5%**	**10%**	**2.5%**	**5%**	**10%**	**2.5%**	**5%**	**10%**
25	(i)	0.023	0.046	0.109	0.018	0.039	0.098	0.026	0.060	0.112
	(ii)	0.029	0.052	0.106	0.028	0.058	0.112	0.037	0.061	0.121
	(iii)	0.019	0.039	0.100	0.022	0.052	0.115	0.031	0.054	0.097
	(iv)	0.449	0.568	0.686	0.557	0.665	0.781	0.596	0.703	0.898
	(v)	0.484	0.607	0.716	0.567	0.676	0.777	0.584	0.701	0.808
	(vi)	0.346	0.462	0.601	0.425	0.549	0.655	0.512	0.645	0.749
	(vii)	0.214	0.322	0.477	0.260	0.404	0.554	0.294	0.408	0.558
	(viii)	0.197	0.316	0.456	0.260	0.403	0.565	0.277	0.416	0.578
	(ix)	0.108	0.184	0.312	0.230	0.325	0.458	0.296	0.435	0.586
50	(i)	0.023	0.052	0.114	0.021	0.044	0.098	0.031	0.042	0.111
	(ii)	0.027	0.049	0.094	0.029	0.050	0.107	0.021	0.047	0.096
	(iii)	0.027	0.047	0.096	0.025	0.046	0.089	0.029	0.053	0.096
	(iv)	0.622	0.748	0.838	0.799	0.878	0.923	0.865	0.917	0.954
	(v)	0.615	0.741	0.836	0.802	0.861	0.919	0.888	0.939	0.971
	(vi)	0.535	0.661	0.759	0.734	0.717	0.883	0.827	0.881	0.932
	(vii)	0.226	0.357	0.535	0.439	0.579	0.737	0.583	0.728	0.881
	(viii)	0.199	0.329	0.534	0.452	0.379	0.745	0.490	0.719	0.855
	(ix)	0.112	0.204	0.358	0.314	0.465	0.617	0.508	0.667	0.830

$K_N^{(1)}$ and $K_N^{(2)}$ shows that the application of the marked empirical process $\widehat{R}_N^{(2)}$ usually yields an improvement with respect to the power of approximately 5–10% (see Tables 1 and 2) and in most cases also a better approximation of the nominal level. Tables 3 and 4 show a few of the corresponding results for the wild bootstrap tests based on the statistics L_N and C_N. The results of the first three rows demonstrate that these procedures yield a less accurate approximation of the nominal level, except in the case (i), where the regression functions are assumed to be constant. In all other cases the level is underestimated. The reason for these problems (as mentioned in Remark 3.3) is that in general the corresponding partial sum processes for Kulasekera's (1995) and Cabus' (2000) tests have only a stochastic expansion of order h^d under the null hypothesis of equal regression curves (except in the case, where $f_1 = f_2$ is constant). This nonnegligible bias also appears if critical values are obtained by the wild bootstrap. The processes $\widehat{R}_N^{(1)}$ and $\widehat{R}_N^{(2)}$ are based on a difference of two marked empirical processes and this difference operation produces a stochastic expansion of order h^{2d} under the null hypothesis (see Proposition 2.1 and its proof). This theoretical advantage is partially supported by our simulation study. The wild bootstrap tests based on

TABLE 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* (4.11). *The errors are homoscedastic and have variances* $\sigma_1^2 = 0.5$, $\sigma_2^2 = 0.25$

	n_2:	25			50			100		
n_1	α:	**2.5%**	**5%**	**10%**	**2.5%**	**5%**	**10%**	**2.5%**	**5%**	**10%**
25	(i)	0.027	0.057	0.119	0.024	0.052	0.109	0.028	0.055	0.107
	(ii)	0.032	0.061	0.110	0.030	0.058	0.114	0.023	0.053	0.108
	(iii)	0.031	0.052	0.105	0.033	0.059	0.103	0.025	0.054	0.108
	(iv)	0.599	0.724	0.804	0.681	0.783	0.873	0.719	0.806	0.895
	(v)	0.644	0.752	0.838	0.709	0.800	0.881	0.707	0.810	0.885
	(vi)	0.523	0.634	0.745	0.599	0.699	0.794	0.633	0.749	0.837
	(vii)	0.111	0.223	0.421	0.067	0.177	0.359	0.098	0.207	0.375
	(viii)	0.112	0.226	0.409	0.111	0.201	0.372	0.128	0.239	0.404
	(ix)	0.067	0.160	0.322	0.081	0.132	0.261	0.090	0.143	0.279
50	(i)	0.029	0.052	0.093	0.035	0.052	0.117	0.033	0.059	0.113
	(ii)	0.023	0.047	0.112	0.034	0.051	0.107	0.031	0.056	0.109
	(iii)	0.022	0.048	0.091	0.023	0.049	0.097	0.028	0.053	0.097
	(iv)	0.761	0.845	0.915	0.919	0.950	0.982	0.953	0.974	0.990
	(v)	0.726	0.814	0.883	0.911	0.948	0.975	0.958	0.979	0.992
	(vi)	0.723	0.815	0.877	0.877	0.932	0.966	0.946	0.976	0.988
	(vii)	0.176	0.308	0.525	0.391	0.574	0.777	0.490	0.651	0.802
	(viii)	0.137	0.291	0.505	0.438	0.619	0.810	0.457	0.629	0.803
	(ix)	0.117	0.265	0.476	0.317	0.513	0.692	0.321	0.480	0.677

TABLE 3

Rejection probabilities of a wild bootstrap version of the test based on L_N [*see* (4.9)] *for various sample sizes and the regression functions specified in* (4.11). *The errors are homoscedastic and have variances* $\sigma_1^2 = 0.5$, $\sigma_2^2 = 0.25$

	n_2:	25			50			100		
n_1	α:	**2.5%**	**5%**	**10%**	**2.5%**	**5%**	**10%**	**2.5%**	**5%**	**10%**
25	(i)	0.029	0.054	0.097	0.028	0.057	0.114	0.036	0.058	0.110
	(ii)	0.015	0.032	0.076	0.013	0.035	0.083	0.019	0.038	0.085
	(iii)	0.010	0.036	0.080	0.019	0.038	0.076	0.021	0.043	0.082
	(iv)	0.593	0.715	0.793	0.670	0.770	0.863	0.654	0.766	0.862
	(vii)	0.090	0.152	0.302	0.033	0.134	0.269	0.089	0.181	0.327
50	(i)	0.024	0.052	0.101	0.024	0.056	0.117	0.035	0.068	0.107
	(ii)	0.019	0.037	0.084	0.017	0.041	0.087	0.021	0.041	0.081
	(iii)	0.014	0.029	0.081	0.021	0.039	0.079	0.019	0.038	0.086
	(iv)	0.779	0.872	0.923	0.912	0.949	0.976	0.941	0.967	0.982
	(vii)	0.123	0.234	0.399	0.350	0.539	0.692	0.401	0.615	0.731

TABLE 4
Rejection probabilities of a wild bootstrap version of the test based on C_N [see (4.10)] for various sample sizes and the regression functions specified in (4.11). The errors are homoscedastic and have variances $\sigma_1^2 = 0.5$, $\sigma_2^2 = 0.25$

	n_2:	25			50			100		
n_1	α:	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
25	(i)	0.032	0.054	0.106	0.025	0.048	0.113	0.032	0.064	0.120
	(ii)	0.014	0.038	0.081	0.019	0.037	0.079	0.022	0.046	0.091
	(iii)	0.015	0.034	0.082	0.021	0.039	0.085	0.021	0.048	0.093
	(iv)	0.590	0.791	0.804	0.627	0.739	0.863	0.626	0.744	0.821
	(vii)	0.081	0.152	0.321	0.046	0.124	0.289	0.076	0.144	0.268
50	(i)	0.026	0.056	0.107	0.023	0.045	0.103	0.031	0.059	0.101
	(ii)	0.019	0.038	0.084	0.020	0.041	0.091	0.022	0.043	0.091
	(iii)	0.018	0.042	0.088	0.022	0.043	0.086	0.021	0.041	0.089
	(iv)	0.781	0.860	0.919	0.886	0.925	0.963	0.909	0.955	0.976
	(vii)	0.145	0.217	0.424	0.312	0.497	0.712	0.401	0.589	0.721

$K_N^{(1)}$ and $K_N^{(2)}$ yield a sufficiently accurate approximation of the nominal level in all considered cases, while for small sample sizes, L_N and C_N yield only a comparable approximation of the level for constant regression functions. (Even for linear regression functions the error is substantial; these results are not displayed.) For these reasons only the cases (iv) and (vii) are displayed in Tables 3 and 4 as illustration for the performance of these tests under alternatives. Here we observe in the case $1 = f_1(x) = f_2(x) - x$ a similar behavior as for the statistics $K_N^{(1)}$ and $K_N^{(2)}$ while in the situation $1 = f_1(x) = f_2(x) - \sin(2\pi x)$ of an oscillating alternative the tests based on L_N and C_N have less power. In all other cases (f_1, f_2 not constant) the loss of power when applying the wild bootstrap tests based on L_N and C_N is of similar order. This loss of power can be partially explained by the fact that the tests of Kulasekera (1995) and Cabus (2000) underestimate the nominal level. Based on our (limited) numerical experience, the wild bootstrap version of the test based on $K_N^{(2)}$ should be preferred to the procedure based on $K_N^{(1)}$ and to Kulasekera's (1995) and Cabus' (2000) tests because it approximates the nominal level most accurately. We finally mention that (based on a further numerical study) these differences are relatively stable with respect to different selection rules of the bandwidth.

EXAMPLE 4.2. We conclude this section by illustrating the advantage of our procedure with a data example involving heteroscedasticity. To this end we reanalyze data measuring the concentration of sulfate in the rain of North Carolina, which was obtained as a part of the National Atmospheric Deposition Program. This data was investigated by Hall and Hart (1990), who compared the sulfate

TABLE 5
Semiannual variance estimates of the rainfall data discussed by Hall and Hart (1990)

	Variance estimate									
Period	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Coweeta	0.279	0.205	0.385	0.223	0.276	0.629	0.361	0.182	0.407	0.613
Lewiston	0.929	0.584	0.614	0.769	0.551	1.088	0.549	0.551	0.573	0.459

concentration as a function of time in the two towns, Coweeta and Lewiston. The data available for the analysis were weekly observations of the concentration of sulfate in the rain over a five-year period 1979–1983 [see Figure 1 in Hall and Hart (1990)]. A residual analysis gave no indication that the error terms were correlated over time. Because there were several weeks for which data was not available, Hall and Hart (1990) only used the weeks with no missing data (189 weeks). There were 215 weeks of data in Lewiston and 220 weeks of data in Coweeta and consequently these authors do not use 13% of the data [note that Kulasekera (1995) pointed out that Hall and Hart's (1990) test does not work with sufficient accuracy for different design points and therefore there is no way to include the unused data in the analysis with this test]. The test of Hall and Hart (1990) clearly rejects the hypothesis of equal regression curves which is also obvious from Figure 1 in the same reference. The same figure indicates that the two curves could only differ by a simple shift and it is easy to see that a test for this hypothesis can be obtained by applying Hall and Hart's (1990) procedure to the data

$$Y_{ij} - \bar{Y}_{i\cdot}.$$

Hall and Hart (1990) obtained for this procedure a p-value of 0.097 and argued that this provides some evidence that the curves differ by more than a simple shift. We would like to point out here that some care is necessary with this argument because the test of Hall and Hart (1990) requires homoscedasticity, which is rarely available in seasonal data. In order to investigate the question of heteroscedasticity, we estimated the integrated variance functions

$$\int \sigma_i^2(t)\,dt, \qquad i = 1, 2,$$

for the two locations semiannually using Rice's (1984) estimator. The corresponding estimates are shown for the two towns, Coweeta and Lewiston, in Table 5 and clearly indicate a heteroscedastic structure in the data.

For this reason we investigate whether the conclusion that the two curves differ by more than a shift is obtained by the application of a (inappropriate) procedure requiring homoscedastic errors to heteroscedastic data (or even by neglecting 13% of the data). To this end we applied the bootstrap test based on $K_N^{(2)}$ (which works under heteroscedasticity and for unequal design points) to the rainfall data considered by Hall and Hart (1990). The resulting p-value obtained from $B = 1{,}000$ bootstrap replications is 0.405. We obtained nearly the same p-value if we only used the 189 weeks, where data was available at both locations. Consequently we conclude that there is in fact some evidence from the data that the two curves differ only by a simple shift as suggested in Figure 1 of Hall and Hart (1990). The different conclusion obtained by these authors is probably caused by ignoring heteroscedasticity in the data.

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

5.1. *Proof of Proposition* 2.1. The expectation of the residuals in (2.8) is obtained as

$$
\begin{aligned}
&E[e_{ij} I\{X_{ij} \le t\}] \\
&\quad = E\biggl[E\biggl[\bigl(Y_{ij}\widehat{r}(X_{ij}) - \widehat{f}(X_{ij})\widehat{r}(X_{ij})\bigr) \\
&\qquad\qquad \times \frac{n_{3-i}}{N}\widehat{r}_{3-i}(X_{ij}) I\{X_{ij} \le t\} \,\Big|\, X_{11}, \dots, X_{2n_2}\biggr]\biggr] \\
&\quad = \frac{1}{Nh}\sum_{\ell=1}^{2}\sum_{k=1}^{n_\ell} E\biggl[K\biggl(\frac{X_{\ell k} - X_{ij}}{h}\biggr)\bigl(f_i(X_{ij}) - f_\ell(X_{\ell k})\bigr) I\{X_{ij} \le t\} \\
&\qquad\qquad \times \frac{1}{Nh}\sum_{\nu=1}^{n_{3-i}} K\biggl(\frac{X_{ij} - X_{3-i,\nu}}{h}\biggr)\biggr] \\
&\quad = \frac{\kappa_{3-i}(n_i - 1)}{Nh}\int_0^1\int_0^t K\biggl(\frac{x-y}{h}\biggr)\bigl(f_i(x) - f_i(y)\bigr) r_i(x) r_i(y) \\
&\qquad\qquad \times \frac{1}{h}\int_0^1 K\biggl(\frac{x-z}{h}\biggr) r_{3-i}(z)\,dz\,dx\,dy \\
&\qquad + \frac{\kappa_{3-i} n_{3-i}}{Nh}\int_0^1\int_0^t K\biggl(\frac{x-y}{h}\biggr)\bigl(f_i(x) - f_{3-i}(y)\bigr) r_i(x) r_{3-i}(y) \\
&\qquad\qquad \times \frac{1}{h}\int_0^1 K\biggl(\frac{x-z}{h}\biggr) r_{3-i}(z)\,dz\,dx\,dy \\
&\qquad + O\biggl(\frac{1}{Nh}\biggr)
\end{aligned}
$$

and a Taylor expansion and a tedious calculation yield, under the hypothesis of equal regression curves,

$$
\begin{aligned}
E\big[\widehat{R}_N^{(1)}(t)\big]
&= \kappa_1\kappa_2 \int_0^t \int_0^1 \frac{1}{h} K\left(\frac{x-y}{h}\right)[f(x)-f(y)]r(y) \\
&\qquad \times \left[r_1(x)\frac{1}{h}\int_0^1 K\left(\frac{x-z}{h}\right) r_2(z)\,dz \right. \\
&\qquad\qquad \left. - r_2(x)\frac{1}{h}\int_0^1 K\left(\frac{x-z}{h}\right) r_1(z)\,dz\right] dy\,dx + O\left(\frac{1}{Nh}\right) \\
&= O(h^{2d}) + O\left(\frac{1}{Nh}\right).
\end{aligned}
$$

Similarly, we obtain under the alternative,

$$
E[e_{ij} I\{X_{ij} \le t\}] = \kappa_{3-i} \int_0^t \big(f_i(x) - f_{3-i}(x)\big) r_i(x) r_{3-i}^2(x)\,dx + O(h^d)
$$

and the definition of $\widehat{R}_N^{(1)}$ yields

$$
\begin{aligned}
E\big[\widehat{R}_N^{(1)}(t)\big] &= \kappa_1\kappa_2^2 \int_0^t \big(f_1(x) - f_2(x)\big) r_1(x) r_2^2(x)\,dx \\
&\qquad - \kappa_2\kappa_1^2 \int_0^t \big(f_2(x) - f_1(x)\big) r_2(x) r_1^2(x)\,dx + O(h^d),
\end{aligned}
$$

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

5.2. *Proof of Theorem* 2.2. We begin with an auxiliary result, which shows that the $\widehat{r}_{3-i}(X_{ij})$ weights in the residuals e_{ij} can be replaced by $r_{3-i}(X_{ij})$ without changing the asymptotic properties of the test statistic. The proof follows by similar arguments as given in Lemma 5.3 and is left to the reader.

LEMMA 5.0. *Define*

$$
\bar{e}_{ij} = \big(Y_{ij} - \widehat{f}(X_{ij})\big)\widehat{r}(X_{ij})\kappa_{3-i} r_{3-i}(X_{ij})
$$

and

$$
\bar{R}_N(t) = \frac{1}{N}\sum_{j=1}^{n_1} \bar{e}_{1j} I\{X_{1j} \le t\} - \frac{1}{N}\sum_{j=1}^{n_2} \bar{e}_{2j} I\{X_{2j} \le t\}.
$$

If the assumptions of Theorem 2.2 *are satisfied*, *then*

$$
\sup_{t\in[0,1]} \big|\widehat{R}_N^{(1)}(t) - \bar{R}_N(t)\big| = o_p\left(\frac{1}{\sqrt{N}}\right),
$$

where the process $\widehat{R}_N^{(1)}(t)$ *is defined in* (2.10).

Recalling the definition of the residuals $\bar{e}_{ij}$,

$$\begin{aligned}\bar{e}_{ij} &= \big(\sigma_i(X_{ij})\varepsilon_{ij}\widehat{r}(X_{ij}) + f(X_{ij})\widehat{r}(X_{ij}) - \widehat{f}(X_{ij})\widehat{r}(X_{ij})\big)\kappa_{3-i}r_{3-i}(X_{ij}) \\ &= \Bigg(\sigma_i(X_{ij})\varepsilon_{ij}\widehat{r}(X_{ij}) \\ &\qquad + \frac{1}{Nh}\sum_{\ell=1}^{2}\sum_{k=1}^{n_\ell} K\bigg(\frac{X_{ij}-X_{\ell k}}{h}\bigg)\big(f(X_{ij}) - f(X_{\ell k})\big) \\ &\qquad\qquad - \frac{1}{Nh}\sum_{\ell=1}^{2}\sum_{k=1}^{n_\ell} K\bigg(\frac{X_{ij}-X_{\ell k}}{h}\bigg)\sigma_\ell(X_{\ell k})\varepsilon_{\ell k}\Bigg)\kappa_{3-i}r_{3-i}(X_{ij}),\end{aligned} \tag{5.1}$$

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

$$\bar{R}_N(t) = R_N(t) + S_N(t) + W_N(t) + V_N(t), \tag{5.2}$$

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

$$R_N(t) := \sum_{\ell=1}^{2}\frac{\kappa_{3-\ell}}{N}\sum_{j=1}^{n_\ell}(-1)^{3-\ell}\sigma_\ell(X_{\ell j})\varepsilon_{\ell j}r(X_{\ell j})r_{3-\ell}(X_{\ell j})I\{X_{\ell j}\le t\}, \tag{5.3}$$

$$\begin{aligned}S_N(t) := \sum_{\ell,i=1}^{2}(-1)^{\ell}\frac{\kappa_{3-\ell}}{N^2h}\sum_{j=1}^{n_i}\sigma_i(X_{ij})\varepsilon_{ij}\sum_{k=1}^{n_\ell}K\bigg(\frac{X_{ij}-X_{\ell k}}{h}\bigg) \\ \times r_{3-\ell}(X_{\ell k})I\{X_{\ell k}\le t\},\end{aligned} \tag{5.4}$$

$$\begin{aligned}W_N(t) := \sum_{\ell,i=1}^{2}(-1)^{\ell-1}\frac{\kappa_{3-\ell}}{N^2h}\sum_{j=1}^{n_\ell}\sum_{k=1}^{n_i}K\bigg(\frac{X_{\ell j}-X_{ik}}{h}\bigg)\big(f(X_{\ell j}) - f(X_{ik})\big) \\ \times r_{3-\ell}(X_{\ell j})I\{X_{\ell j}\le t\},\end{aligned} \tag{5.5}$$

$$\begin{aligned}V_N(t) := \sum_{i=1}^{2}(-1)^{i-1}\kappa_{3-i}\frac{1}{N}\sum_{j=1}^{n_i}\sigma_i(X_{ij})\varepsilon_{ij}\big(\widehat{r}(X_{ij}) - r(X_{ij})\big) \\ \times r_{3-i}(X_{ij})I\{X_{ij}\le t\}.\end{aligned} \tag{5.6}$$

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* $T_N(t) = \sqrt{N}R_N(t)$ *converges weakly to a centered Gaussian process in the space* $D[0, 1]$ *with covariance function given by* (2.12).

PROOF. With the notation

$$\Delta_{ij}(t) := \sigma_i(X_{ij})\big[(-1)^{i-1}\kappa_{3-i}r_{3-i}(X_{ij})r(X_{ij})I\{X_{ij}\le t\}\big] \tag{5.7}$$

$(i=1,2)$, we decompose the process T_N as follows:

$$T_N(t) = \sqrt{N}R_N(t) = \sum_{i=1}^{2}\frac{1}{\sqrt{N}}\sum_{j=1}^{n_i}\varepsilon_{ij}\Delta_{ij}(t).$$

For the covariance we obtain, by a straightforward calculation,

$$\begin{aligned}\operatorname{Cov}(T_N(t),T_N(s)) &= E\Bigg[\frac{1}{N}\sum_{j=1}^{n_1}\sigma_1^2(X_{1j})\varepsilon_{1j}^2\Delta_{1j}(t)\Delta_{1j}(s)\\ &\qquad + \frac{1}{N}\sum_{j=1}^{n_2}\sigma_2^2(X_{2j})\varepsilon_{2j}^2\Delta_{2j}(t)\Delta_{2j}(s)\Bigg]\\ &= \kappa_1\kappa_2^2\int_0^1\sigma_1^2(y)r^2(y)r_2^2(y)I\{y\le t\}I\{y\le s\}r_1(y)\,dy\\ &\quad + \kappa_2\kappa_1^2\int_0^1\sigma_2^2(y)r^2(y)r_1^2(y)I\{y\le t\}I\{y\le s\}r_2(y)\,dy\\ &\quad + o(1)\\ &= H(s,t)+o(1).\end{aligned}$$

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

$$\begin{aligned}&E\big[\big(T_N(w)-T_N(v)\big)^2\big(T_N(v)-T_N(u)\big)^2\big]\le C(w-u)^2\\ &\qquad\qquad \text{for all } 0\le u\le v\le w\le 1\end{aligned} \tag{5.8}$$

can be established [see Billingsley (1968), page 128, or Shorack and Wellner (1986), pages 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

$$\begin{aligned}&E\Bigg[\Bigg(\sum_{i=1}^{n_1}\alpha_i+\sum_{j=1}^{n_2}\gamma_j\Bigg)^2\Bigg(\sum_{i=1}^{n_1}\beta_i+\sum_{j=1}^{n_2}\delta_j\Bigg)^2\Bigg]\\ &\quad\le n_1E[\alpha_1^2\beta_1^2]+3n_1^2E[\alpha_1^2]E[\beta_1^2]\\ &\qquad + n_2E[\gamma_1^2\delta_1^2]+3n_2^2E[\gamma_1^2]E[\delta_1^2]+n_1n_2E[\alpha_1^2]E[\delta_1^2]\\ &\qquad + n_1n_2E[\gamma_1^2]E[\beta_1^2]+4n_1n_2E[\alpha_1\beta_1]E[\gamma_1\delta_1]\end{aligned} \tag{5.9}$$

holds, which follows by similar arguments as stated in the proof of Theorem 13.1

in Billingsley (1968). We now apply (5.9) for the random variables

$$\begin{aligned}
\alpha_i &= \varepsilon_{1i}\big(\Delta_{1i}(w) - \Delta_{1i}(v)\big), & \beta_i &= \varepsilon_{1i}\big(\Delta_{1i}(v) - \Delta_{1i}(u)\big),\\
\gamma_j &= \varepsilon_{2j}\big(\Delta_{2j}(w) - \Delta_{2j}(v)\big), & \delta_j &= \varepsilon_{2j}\big(\Delta_{2j}(v) - \Delta_{2j}(u)\big).
\end{aligned} \tag{5.10}$$

A straightforward calculation yields

$$\begin{aligned}
E[\alpha_1^2] &= \int_0^1 \sigma_1^2(x)\big(\kappa_2 r_2(x) r(x) I\{v < x \le w\}\big)^2 r_1(x)\,dx\\
&= \int_v^w \sigma_1^2(x)\kappa_2^2 r_2^2(x) r^2(x) r_1(x)\,dx\\
&\le O(1)(w-v)
\end{aligned}$$

and similar arguments show that the terms $E[\beta_1^2]$, $E[\gamma_1^2]$, $E[\delta_1^2]$, $E[\alpha_1\beta_1]$, $E[\gamma_1\delta_1]$, $E[\alpha_1^2\beta_1^2]$ and $E[\gamma_1^2\delta_1^2]$ are of the same order. Now, a combination of these results with (5.10) and (5.9) yields

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

which establishes (5.8) 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 *defined by* (5.4),

$$\sup_{t\in[0,1]} |S_N(t)| = o_p\bigg(\frac{1}{\sqrt{N}}\bigg). \tag{5.11}$$

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),

$$\sup_{t\in[0,1]} |V_N(t)| = o_p\bigg(\frac{1}{\sqrt{N}}\bigg), \tag{5.12}$$

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

In order to prove Lemmas 5.2 and 5.3 we need some basic terminology from recent U-process 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

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

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

$$\mathcal{N}_1(\varepsilon, Q, \mathcal{F}, F) \le A\varepsilon^{-V} \qquad \text{whenever } 0 < QF < \infty.$$

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

$$\mathcal{D} = \{G_f \mid f \in \mathcal{F}\}$$

with

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

forms a polynomial class (or *VC*-class); that is, there exists a polynomial $p(\cdot)$ such that

$$\#\{D \cap F \mid D \in \mathcal{D}\} \le p(\#F)$$

for every fixed finite subset F of S. We finally note that *VC*-classes are Euclidean [see Pollard (1984), Lemma II25] 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 (5.12); the remaining case (5.13) is very similar and left to the reader. Recalling the definition of V_N in (5.6) we obtain the decomposition,

$$V_N(t) = V_N^{(1)}(t) + V_N^{(2)}(t) - V_N^{(3)}(t) - V_N^{(4)}(t) + o_p\left(\frac{1}{\sqrt{N}}\right), \tag{5.14}$$

where

$$\begin{aligned} V_N^{(1)}(t) = \frac{\kappa_2}{N^2 h} \sum_{j=1}^{n_1} \sum_{k=1}^{n_1} \sigma_1(X_{1j})\varepsilon_{1j}\left(K\left(\frac{X_{1j} - X_{1k}}{h}\right) - hr_1(X_{1j})\right) \\ \times r_2(X_{1j}) I\{X_{1j} \le t\}, \end{aligned} \tag{5.15}$$

$$V_N^{(2)}(t) = \frac{\kappa_2}{N^2 h} \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} \sigma_1(X_{1j})\varepsilon_{1j}\left(K\left(\frac{X_{1j} - X_{2k}}{h}\right) - hr_2(X_{1j})\right) \times r_2(X_{1j}) I\{X_{1j} \le t\}, \tag{5.16}$$

$$V_N^{(3)}(t) = \frac{\kappa_1}{N^2 h} \sum_{j=1}^{n_2} \sum_{k=1}^{n_1} \sigma_2(X_{2j})\varepsilon_{2j}\left(K\left(\frac{X_{2j} - X_{1k}}{h}\right) - hr_1(X_{2j})\right) \times r_1(X_{2j}) I\{X_{2j} \le t\}, \tag{5.17}$$

$$V_N^{(4)}(t) = \frac{\kappa_1}{N^2 h} \sum_{j=1}^{n_2} \sum_{k=1}^{n_2} \sigma_2(X_{2j})\varepsilon_{2j}\left(K\left(\frac{X_{2j} - X_{2k}}{h}\right) - hr_2(X_{2j})\right) \times r_1(X_{2j}) I\{X_{2j} \le t\}; \tag{5.18}$$

the remainder in (5.14) is obtained replacing κ_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 (5.14) 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* (5.15) *and* (5.18),

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

PROOF. Both terms are treated exactly in the same way and we only consider $V_N^{(1)}$ which can be written as

$$\begin{aligned} V_N^{(1)}(t) \;=\; & \frac{\kappa_2}{N^2 h} \sum_{j=1}^{n_1} \sum_{k=1,k\neq j}^{n_1} \sigma_1(X_{1j})\varepsilon_{1j}\left(K\left(\frac{X_{1j} - X_{1k}}{h}\right) - hr_1(X_{1j})\right) \\ & \times r_2(X_{1j}) I\{X_{1j} \le t\} \\ & + \frac{\kappa_2}{N^2 h} \sum_{j=1}^{n_1} \sigma_1(X_{1j})\varepsilon_{1j}\big(K(0) - hr_1(X_{1j})\big) r_2(X_{1j}) I\{X_{1j} \le t\} \\ =:\; & I_N(t) + I_N^{(1)}(t), \end{aligned} \tag{5.19}$$

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

$$\sup_{t\in[0,1]} \big|I_N^{(1)}(t)\big| = O_p\biggl(\frac{1}{Nh}\biggr) = o_p\biggl(\frac{1}{\sqrt{N}}\biggr), \tag{5.20}$$

where we have used the assumptions for the bandwidth stated in (2.3). The treatment of the remaining term I_N in (5.19) 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

$$\sqrt{N} I_N - \frac{\kappa_1^{3/2}}{2h} U_{n_1}(\varphi) = o_p(1) \tag{5.21}$$

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

$$U_{n_1}(\varphi) := \frac{\sqrt{n_1}}{n_1(n_1-1)} \sum_{i=1}^{n_1} \sum_{\substack{j=1\\ i\neq j}}^{n_1} \varphi(\xi_i, \xi_j) \tag{5.22}$$

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

$$\begin{aligned} \varphi(\xi_i, \xi_j) &= \kappa_2 \varepsilon_{1j} \biggl(K\biggl(\frac{X_{1i}-X_{1j}}{h}\biggr) - h r_1(X_{1j})\biggr) \\ &\qquad \times \sigma_1(X_{1j}) r_2(X_{1j}) I\{X_{1j} \le t\} \\ &\quad + \kappa_2 \varepsilon_{1i} \biggl(K\biggl(\frac{X_{1i}-X_{1j}}{h}\biggr) - h r_1(X_{1i})\biggr) \\ &\qquad \times \sigma_1(X_{1i}) r_2(X_{1i}) I\{X_{1i} \le t\}. \end{aligned} \tag{5.23}$$

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}; that is,

$$U_{n_1}(\varphi) = U_{n_1}(\tilde\varphi) + \frac{2}{\sqrt{n_1}} \sum_{i=1}^{n_1} \varphi_1(\xi_i), \tag{5.24}$$

where

$$\tilde\varphi(x, y) = \varphi(x, y) - \varphi_1(x) - \varphi_1(y) \tag{5.25}$$

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

$$\mathcal{F} = \{\varphi_{h,t} \mid t \in [0, 1],\ h > 0\}, \tag{5.26}$$

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

$$\begin{aligned} \varphi_{h,t}(x, y) &= \kappa_2 x_2 \biggl(K\biggl(\frac{x_1-y_1}{h}\biggr) - h r_1(x_1)\biggr) \sigma_1(x_1) r_2(x_1) I\{x_1 \le t\} \\ &\quad + \kappa_2 y_2 \biggl(K\biggl(\frac{x_1-y_1}{h}\biggr) - h r_1(y_1)\biggr) \sigma_1(y_1) r_2(y_1) I\{y_1 \le t\}. \end{aligned} \tag{5.27}$$

It can be shown by a tedious calculation and similar arguments as in Nolan and Pollard [(1987), Lemma 16] and Pollard [(1984), Examples II26, II38] that the class $\mathcal{F}$ and the induced class

$$P\mathcal{F} = \{\varphi_1 \mid \varphi_1(x) = E[\varphi(\xi_1, \xi_2)|\xi_2 = x], \varphi \in \mathcal{F}\} \tag{5.28}$$

are Euclidean. Note that the proof of this property requires the special assumption on the kernel K stated in the paragraph following (2.4) [see Pollard (1984), Example II38 and Problem II28, 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 fulfilled. Now the second part in the proof of this theorem shows

$$\sup_{\varphi \in \mathcal{F}} |U_{n_1}(\tilde{\varphi})| = O_p\left(\frac{1}{\sqrt{N}}\right). \tag{5.29}$$

The assertion of the first part in Lemma 5.3a now follows from (5.29), (5.24), (5.21), (5.19) and (5.20) if the estimate

$$\sup_{t \in [0,1]} \left| \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \frac{1}{h} \varphi_{1,t,h_{n_1}}(\xi_i) \right| = o_p(1) \tag{5.30}$$

can be established, where

$$\begin{aligned} \varphi_{1,t,h}(\xi_i) = \varphi_1(\xi_i) = \kappa_2 \varepsilon_{1i} &\left(\int K\left(\frac{x - X_{1i}}{h}\right) r_1(x)\,dx - h r_1(X_{1i}) \right) \\ &\times \sigma_1(X_{1i}) r_2(X_{1i}) I\{X_{1i} \leq t\}. \end{aligned} \tag{5.31}$$

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,

$$\mathcal{F}_{n_1} := \{\varphi_{1,t,h_{n_1}} \mid t \in [0, 1]\}. \tag{5.32}$$

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

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

where c is a constant chosen such that

$$
\begin{aligned}
P(\varphi_{1,t,h_{n_1}}^2) &= \int_0^t \sigma_1^2(z)\biggl(\int K\biggl(\frac{x-z}{h_{n_1}}\biggr) r_1(x)\,dx - h_{n_1} r_1(z)\biggr)^2 \kappa_2^2 r_2^2(z) r_1(z)\,dz \\
&= h_{n_1}^2 \int_0^t \sigma_1^2(z)\biggl(\int K(u)\bigl(r_1(z+h_{n_1}u) - r_1(z)\bigr)\,du\biggr)^2 \kappa_2^2 r_2^2(z) r_1(z)\,dz \\
&\le h_{n_1}^2 h_{n_1}^{2d} c.
\end{aligned}
$$

Let F_1 denote the envelope of the class $P\mathcal{F}$ defined by (5.28) (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

$$
\mathbb{P}\biggl(|P_{n_1}F_1 - PF_1| > \frac{k_1}{2}\biggr) \xrightarrow{N\to\infty} 0,
$$

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 (5.30) 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° denote the symmetrization of P_n [see Pollard (1984), page 15]; then we obtain for $\varepsilon_{n_1} = \varepsilon\delta_{n_1}^2\alpha_{n_1}$ $(\varepsilon > 0)$,

$$
\begin{aligned}
&\mathbb{P}\biggl(\sup_{\varphi\in\mathcal{F}_{n_1}} |P_{n_1}(\varphi)| > 8\varepsilon_{n_1}\biggl(\frac{k_1}{2} + k_2\biggr)\biggr) \\
&\quad\le 4\mathbb{P}\biggl(\sup_{\varphi\in\mathcal{F}_{n_1}} |P_{n_1}^\circ(\varphi)| > 2\varepsilon_{n_1}\biggl(\frac{k_1}{2} + k_2\biggr)\biggr) \\
&\quad\le 4\mathbb{P}\biggl(\sup_{\varphi\in\mathcal{F}_{n_1}} |P_{n_1}^\circ(\varphi)| > 2\varepsilon_{n_1} P_{n_1}F_1\biggr).
\end{aligned} \tag{5.33}
$$

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

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

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

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{\varphi\in\mathcal{F}_{n_1}} |P^\circ_{n_1}(\varphi)| > 2\varepsilon_{n_1} P_{n_1} F_1\Big) \\
&\qquad \le 2A\varepsilon_{n_1}^{-V}\exp\Big(-\frac{1}{8}\frac{k_1^2\varepsilon_{n_1}^2 n_1}{\delta_{n_1}^2}\Big) + \mathbb{P}\Big(\sup_{\varphi\in\mathcal{F}_{n_1}} P_{n_1}(\varphi^2) > \delta_{n_1}^2\Big)
\end{aligned} \tag{5.34}$$

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

$$\begin{aligned}
E\Big|\sup_{\varphi\in\mathcal{F}_{n_1}} P_{n_1}(\varphi^2)\Big| &\le \frac{1}{n_1}E\Big|\sum_{i=1}^{n_1}\varepsilon_{1i}^2\Big(\int K\Big(\frac{x-X_{1i}}{h_{n_1}}\Big)r_1(x)\,dx - h_{n_1}r_1(X_{1i})\Big)^2 \\
&\qquad\qquad \times \sigma_1^2(X_{1i})\kappa_2^2 r_2^2(X_{1i})\Big| \\
&= O\big(h_{n_1}^{2d+2}\big)
\end{aligned}$$

and Markov's inequality yields (using the definition of δ_{n_1})

$$\mathbb{P}\Big(\sup_{\varphi\in\mathcal{F}_{n_1}} P_{n_1}(\varphi^2) > \delta_{n_1}^2\Big) = O\big(h_{n_1}\big). \tag{5.35}$$

A combination of (5.33)–(5.35) finally gives

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

which establishes the remaining estimate (5.30) [note that $\delta_{n_1}^2\alpha_{n_1} = O(h_{n_1}/\sqrt{n_1})$]. □

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* (5.16) *and* (5.17),

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

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

PROOF. The proof essentially follows the arguments given in the proof of Lemma 5.3a and we will restrict ourselves to indicating the main difference, which is a derivation of an analogue of the estimate (5.29). 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 analogous 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 $X_1, \dots, X_{2n} \sim P$ and $Y_1, \dots, Y_{2m} \sim Q$ denote independent samples and $\sigma_1, \dots, \sigma_n$ and $\tau_1, \dots, \tau_m$ denote independent samples (also independent of the X_i and Y_j) such that

$$\mathbb{P}(\sigma_i = 1) = \mathbb{P}(\sigma_i = -1) = 1/2,$$
$$\mathbb{P}(\tau_i = 1) = \mathbb{P}(\tau_i = -1) = 1/2.$$

Introducing the notation

$$\xi_i = I\{\sigma_i = 1\}X_{2i} + I\{\sigma_i = -1\}X_{2i-1},$$
$$\xi_i' = I\{\sigma_i = 1\}X_{2i-1} + I\{\sigma_i = -1\}X_{2i},$$
$$\zeta_j = I\{\tau_j = 1\}Y_{2j} + I\{\tau_j = -1\}Y_{2j-1},$$
$$\zeta_j' = I\{\tau_j = 1\}Y_{2j-1} + I\{\tau_j = -1\}Y_{2j},$$

we obtain again independent samples $\xi_1, \dots, \xi_n, \xi_1', \dots, \xi_n' \sim P$ and $\zeta_1, \dots, \zeta_m$, $\zeta_1', \dots, \zeta_m' \sim Q$.

For a function $\varphi \in \mathcal{F}$ consider the two-sample U-statistic,

$$S_{nm}(\varphi) := \sum_{i=1}^{n} \sum_{j=1}^{m} \varphi(\xi_i, \zeta_j), \tag{5.36}$$

and its standardized version,

$$U_{nm}(\varphi) := \frac{\sqrt{n+m}}{nm} S_{nm}(\varphi). \tag{5.37}$$

Let

$$\varphi_1(x) = E[\varphi(\xi_1, \zeta_1) | \xi_1 = x],$$
$$\varphi_2(y) = E[\varphi(\xi_1, \zeta_1) | \zeta_1 = y],$$

and define the kernel

$$\tilde{\varphi}(x, y) = \varphi(x, y) - \varphi_1(x) - \varphi_2(y); \tag{5.38}$$

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

$$T_{nm}(\varphi) := \sum_{i=1}^{n} \sum_{j=1}^{m} [\varphi(\xi_i, \zeta_j) + \varphi(\xi_i, \zeta_j') + \varphi(\xi_i', \zeta_j) + \varphi(\xi_i', \zeta_j')]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \varphi(X_{2i}, Y_{2j}) + \varphi(X_{2i}, Y_{2j-1}) + \varphi(X_{2i-1}, Y_{2j}) + \varphi(X_{2i-1}, Y_{2j-1}) \tag{5.39}$$

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

$$\sup_{n,m} E[\mathcal{J}(1, T_{nm}, \mathcal{F}, F)^2] < \infty, \tag{5.40}$$

$$\mathcal{J}(1, P \otimes Q, \mathcal{F}, F) < \infty, \tag{5.41}$$

$$\sup_{n} E[\mathcal{J}(1, P_n, Q\mathcal{F}, QF)^2] < \infty, \tag{5.42}$$

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

imply the estimate

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

which gives

$$\sup_{\varphi \in \mathcal{F}} |U_{nm}(\tilde{\varphi})| = O_p\Big(\frac{1}{\sqrt{N}}\Big). \tag{5.44}$$

In the specific situation of $V_N^{(2)}$ or $V_N^{(3)}$ the assumptions (5.40)–(5.43) now follow, 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). $\square$

5.4. *Proof of Lemma* 5.2. Recalling the definition of S_N in (5.4) and observing the equality

$$\sum_{\ell,i=1}^{2} (-1)^\ell \frac{\kappa_{3-\ell}}{hN} \sum_{j=1}^{n_i} \sigma_i(X_{ij})\varepsilon_{ij} \frac{1}{h} \int_0^t K\Big(\frac{X_{ij} - x}{h}\Big) r_{3-\ell}(x)\kappa_\ell r_\ell(x)\, dx = 0,$$

it follows that S_N is a linear combination of four terms of the form

$$\frac{2\kappa_{3-k}}{hn_\ell n_k}\sum_{i=1}^{n_\ell}\varepsilon_{\ell i}\Biggl(\sum_{j=1,i\neq j}^{n_k} K\biggl(\frac{X_{\ell i}-X_{kj}}{h}\biggr)r_{3-k}(X_{kj})I\{X_{kj}\le t\} \\ - \int_0^t K\biggl(\frac{X_{\ell i}-x}{h}\biggr)r_k(x)r_{3-k}(x)\,dx\Biggr)\sigma_\ell(X_{\ell i}),$$

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 (5.40)–(5.44) 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

$$\begin{aligned}\varphi(\xi_i,\xi_j) = \kappa_2\varepsilon_{1i}\biggl(&K\biggl(\frac{X_{1i}-X_{1j}}{h}\biggr)r_2(X_{1j})I\{X_{1j}\le t\} \\ &- \int_0^t K\biggl(\frac{X_{1i}-x}{h}\biggr)r_1(x)r_2(x)\,dx\biggr)\sigma_1(X_{1i}) \\ + \kappa_2\varepsilon_{1j}\biggl(&K\biggl(\frac{X_{1i}-X_{1j}}{h}\biggr)r_2(X_{1i})I\{X_{1i}\le t\} \\ &- \int_0^t K\biggl(\frac{X_{1j}-x}{h}\biggr)r_1(x)r_2(x)\,dx\biggr)\sigma_1(X_{1j}),\end{aligned}$$

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

$$\frac{1}{h}\sup_{\varphi\in\mathcal{F}}\bigl|U_{n_1,n_1}^{(1,1)}(\varphi)\bigr| = \frac{1}{h}O_p\biggl(\frac{1}{N}\biggr) = o_p\biggl(\frac{1}{\sqrt{N}}\biggr).$$

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

5.5. *Proof of Theorems* 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

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

where m_{12} is defined by (3.4).

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 $\widehat{R}_N^{(1)*}$ (the remaining case is treated exactly in the same way) and obtain

$$
\sup_{t\in[0,1]} \bigl|\widehat{R}_N^{(1)*}(t) - \tilde{R}_N^{(1)*}(t)\bigr| = o_p\Bigl(\frac{1}{\sqrt{N}}\Bigr), \tag{5.45}
$$

where

$$
\begin{aligned}
\tilde{R}_N^{(1)*}(t) &= \frac{1}{N} \sum_{\ell=1}^{2} \sum_{j=1}^{n_\ell} (-1)^{\ell-1} \bigl(Y_{\ell j}^* - \widehat{f}_h^*(X_{\ell j})\bigr) \widehat{r}_h(X_{\ell j}) \\
&\qquad \times I\{X_{\ell j} \le t\} \kappa_{3-\ell} r_{3-\ell}(X_{\ell j})
\end{aligned} \tag{5.46}
$$

and $\widehat{R}_N^{(1)*}$ is defined in (4.6). This shows that it is sufficient to prove the corresponding statement for the process $\tilde{R}_N^{(1)*}$, for which we obtain the decomposition

$$\tilde{R}_N^{(1)*}(t) = R_N^*(t) + S_N^*(t) + W_N^*(t) + V_N^*(t), \tag{5.47}$$

where the processes on the right-hand side are defined by

$$\begin{aligned} R_N^*(t) &:= \frac{\kappa_2}{N}\sum_{j=1}^{n_1}\varepsilon_{1j}^* r(X_{1j})r_2(X_{1j})I\{X_{1j}\le t\} \\ &\qquad - \frac{\kappa_1}{N}\sum_{j=1}^{n_2}\varepsilon_{2j}^* r(X_{2j})r_1(X_{2j})I\{X_{2j}\le t\}, \end{aligned} \tag{5.48}$$

$$\begin{aligned} S_N^*(t) &:= \sum_{i=1}^{2}\frac{1}{N}\sum_{j=1}^{n_i}\varepsilon_{ij}^*\Biggl(\frac{1}{Nh}\sum_{\ell=1}^{2}(-1)^{\ell}\kappa_{3-\ell}\sum_{k=1}^{n_\ell}K\biggl(\frac{X_{ij}-X_{\ell k}}{h}\biggr) \\ &\qquad\qquad \times r_{3-\ell}(X_{\ell k})I\{X_{\ell k}\le t\}\Biggr), \end{aligned} \tag{5.49}$$

$$\begin{aligned} W_N^*(t) &:= \sum_{\ell,i=1}^{2}(-1)^{\ell-1}\frac{\kappa_{3-\ell}}{N^2h}\sum_{j=1}^{n_\ell}\sum_{k=1}^{n_i}K\biggl(\frac{X_{\ell j}-X_{ik}}{h}\biggr)\bigl(\widehat{f}_g(X_{\ell j})-\widehat{f}_g(X_{ik})\bigr) \\ &\qquad\qquad \times r_{3-\ell}(X_{\ell j})I\{X_{\ell j}\le t\}, \end{aligned} \tag{5.50}$$

$$\begin{aligned} V_N^*(t) &:= \sum_{i=1}^{2}(-1)^{i-1}\kappa_{3-i}\frac{1}{N}\sum_{j=1}^{n_i}\varepsilon_{ij}^*\bigl(\widehat{r}_h(X_{ij})-r(X_{ij})\bigr) \\ &\qquad\qquad \times r_{3-i}(X_{ij})I\{X_{ij}\le t\}. \end{aligned} \tag{5.51}$$

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

LEMMA 5.4. *If the assumptions of Theorem* 2.2 *and* (4.8) *are satisfied we have for all* $\delta > 0$,

$$\mathbb{P}\Bigl(\sqrt{N}\sup_{t\in[0,1]}|V_N^*(t)| > \delta\Big|\mathcal{Y}_N\Bigr) = o_p(1), \tag{5.52}$$

$$\mathbb{P}\Bigl(\sqrt{N}\sup_{t\in[0,1]}|S_N^*(t)| > \delta\Big|\mathcal{Y}_N\Bigr) = o_p(1), \tag{5.53}$$

$$\mathbb{P}\Bigl(\sqrt{N}\sup_{t\in[0,1]}|W_N^*(t)| > \delta\Big|\mathcal{Y}_N\Bigr) = o_p(1). \tag{5.54}$$

Observing Lemma 5.4 it follows that the processes

$$T_N^* := \sqrt{N} R_N^*$$

and $\sqrt{N}\widehat{R}_N^{(1)*}$ are (conditionally on $\mathcal{Y}_N$) asymptotically equivalent in probability; that is,

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

The following lemma shows that T_N^* 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 Δ_{ij} are defined in (5.7).

LEMMA 5.5. *If the assumptions of Theorem* 2.2 *and* (4.8) *are satisfied we have*

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

The assertion of Theorem 4.1 now follows from (5.57) and (5.55) which demonstrate that it is sufficient to consider the asymptotic behavior 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 Lemmas 5.4 and 5.5.

PROOF OF LEMMA 5.4. For a proof of (5.52) 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; that is,

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

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

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

where

$$V_N^{*(1)}(t) = \frac{1}{N}\sum_{i=1}^{2}(-1)^{i-1}\kappa_{3-i}\sum_{j=1}^{n_i} V_{ij}\varepsilon_{ij}\sigma_i(X_{ij})\big(\widehat{r}_h(X_{ij}) - r(X_{ij})\big) \tag{5.60}$$
$$\times r_{3-i}(X_{ij})I\{X_{ij} \le t\},$$

$$V_N^{*(2)}(t) = \sum_{i=1}^{2}(-1)^{i-1}\frac{\kappa_{3-i}}{N}\sum_{j=1}^{n_i} V_{ij}\big(f(X_{ij}) - \widehat{f}_g(X_{ij})\big)\big(\widehat{r}_h(X_{ij}) - r(X_{ij})\big) \tag{5.61}$$
$$\times r_{3-i}(X_{ij})I\{X_{ij} \le t\}.$$

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

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

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

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

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 (5.53) recall the definition of S_N^* in (5.49) and observe

$$S_N^* = S_N^{*(1)} + S_N^{*(2)},$$

where

$$S_N^{*(1)}(t) := \sum_{i=1}^{2}\frac{1}{N}\sum_{j=1}^{n_i} V_{ij}\varepsilon_{ij}\sigma_i(X_{ij})$$
$$\times \frac{1}{Nh}\sum_{\ell=1}^{2}(-1)^{\ell}\sum_{k=1}^{n_\ell} K\bigg(\frac{X_{ij} - X_{\ell k}}{h}\bigg)I\{X_{\ell k} \le t\}\kappa_{3-\ell}r_{3-\ell}(X_{\ell k}),$$

$$S_N^{*(2)}(t) := \sum_{i=1}^{2} \frac{1}{N} \sum_{j=1}^{n_i} V_{ij}\big(f(X_{ij}) - \widehat{f}_g(X_{ij})\big)$$
$$\times \frac{1}{Nh} \sum_{\ell=1}^{2} (-1)^{\ell} \sum_{k=1}^{n_\ell} K\left(\frac{X_{ij} - X_{\ell k}}{h}\right) I\{X_{\ell k} \le t\} \kappa_{3-\ell} r_{3-\ell}(X_{\ell k}).$$

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

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

The second term is estimated as follows:

$$\sup_{t \in [0,1]} \big|S_N^{*(2)}(t)\big| \le \sum_{i=1}^{2} \frac{1}{N} \sum_{j=1}^{n_i} |V_{ij}|\big|f(X_{ij}) - \widehat{f}_g(X_{ij})\big|\big\{U_{Nij}^{(1)} + U_{Nij}^{(2)}\big\}, \tag{5.64}$$

where

$$U_{Nij}^{(\ell)} = \frac{\kappa_{3-\ell}}{h} \sup_{t \in [0,1]} \left| \frac{1}{N} \sum_{k=1}^{n_\ell} K\left(\frac{X_{ij} - X_{\ell k}}{h}\right) r_{3-\ell}(X_{\ell k}) I\{X_{\ell k} \le t\} \right.$$
$$\left. - \int_0^t K\left(\frac{X_{ij} - z}{h}\right) \kappa_\ell r_\ell(z) r_{3-\ell}(z)\, dz \right|$$

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

$$\sup_{t,x \in [0,1]} \left| \frac{1}{n_1} \sum_{k=1}^{n_1} K\left(\frac{x - X_{1k}}{h}\right) r_2(X_{1k}) I\{X_{1k} \le t\} - \int_0^t K\left(\frac{x - z}{h}\right) r_1(z) r_2(z)\, dz \right|$$
$$= \sup_{\varphi \in \mathcal{F}_{n_1}} |P_{n_1}\varphi - P\varphi|,$$

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

$$\mathcal{F}_{n_1} = \left\{ \varphi_{h_{n_1},t,x} \,\middle|\, \varphi_{h_{n_1},t,x}(y) = K\left(\frac{x - y}{h_{n_1}}\right) r_2(y) I\{y \le t\},\ x, t \in [0, 1] \right\}$$

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

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

the estimate

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

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

$$\sup_{t\in[0,1]} |S_N^{*(2)}(t)| \le o_p\big(\sqrt{g_{n_1}}\big)\sum_{i=1}^{2}\frac{1}{N}\sum_{j=1}^{n_i}|V_{ij}|\big|f(X_{ij})-\widehat{f}_g(X_{ij})\big| = o_p\bigg(\frac{1}{\sqrt{N}}\bigg),$$

where the last estimate follows from Markov's inequality. A combination of this estimate with (5.63) gives $\sqrt{N}\sup_{t\in[0,1]}|S_N^*(t)| = o_p(1)$ and the assertion (5.53) follows again from Markov's inequality. $\square$

PROOF OF LEMMA 5.5. Defining $(i=1,2)$

$$\tilde{\Delta}_{ij}(t) := (-1)^{i-1}\kappa_{3-i} r(X_{ij}) r_{3-i}(X_{ij}) I\{X_{ij}\le t\} \tag{5.65}$$

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

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

[note that $\Delta_{ij}(t) = \tilde{\Delta}_{ij}(t)\sigma_i(X_{ij})$, by the definition of Δ_{ij} in (5.7)]. The first term is estimated as follows:

$$\begin{aligned}
\sup_{t\in[0,1]}|A_N(t)| &\le \sum_{i=1}^{2}\frac{1}{\sqrt{N}}\sum_{j=1}^{n_i}\sup_{t\in[0,1]}\big|\tilde{\Delta}_{ij}(t)\big|\frac{1}{r(X_{ij})}\Big|f(X_{ij})-\widehat{f}_g(X_{ij})\Big|\\
&\qquad\qquad \times\big|r(X_{ij})-\widehat{r}_g(X_{ij})\big|\\
&= O_p\bigg(\frac{1}{\sqrt{Ng}}\bigg) = o_p(1),
\end{aligned}$$

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

$$\tilde{B}_N(t) := \frac{1}{n_1\sqrt{N}} \sum_{j=1}^{n_1} \sum_{k=1}^{n_1} \tilde{\Delta}_{1j}(t) \frac{1}{g} K\left(\frac{X_{1j} - X_{1k}}{g}\right)\big(f(X_{1j}) - f(X_{1k})\big) \times V_{1j} \frac{1}{r(X_{1j})}$$

$$+ \frac{1}{n_1\sqrt{N}} \sum_{j=1}^{n_1} \sum_{k=1}^{n_1} \tilde{\Delta}_{1j}(t) \frac{1}{g} K\left(\frac{X_{1j} - X_{1k}}{g}\right) \varepsilon_{1k}\sigma_1(X_{1k}) \times V_{1j} \frac{1}{r(X_{1j})}, \tag{5.68}$$

which are all treated similarly. We obtain

$$\tilde{B}_N(t) = I_1(t) + I_2(t), \tag{5.69}$$

where

$$I_1(t) := \frac{1}{n_1\sqrt{N}} \sum_{j=1}^{n_1} \sum_{k=1}^{n_1} \kappa_2 r_2(X_{1j}) I\{X_{1j} \le t\} \frac{1}{g} K\left(\frac{X_{1j} - X_{1k}}{g}\right) \times \big(f(X_{1j}) - f(X_{1k})\big) V_{1j},$$

$$I_2(t) := \frac{1}{n_1\sqrt{N}} \sum_{j=1}^{n_1} \sum_{k=1}^{n_1} \kappa_2 r_2(X_{1j}) I\{X_{1j} \le t\} \frac{1}{g} K\left(\frac{X_{1j} - X_{1k}}{g}\right) \varepsilon_{1k}\sigma_1(X_{1k}) V_{1j}.$$

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

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

This implies

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

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

Acknowledgments. This paper contains parts of the first author's Ph.D. thesis, written at the Ruhr-Universität Bochum. The authors are grateful for the technical expertise of Isolde Gottschlich who typed most of this paper 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 and J. Hart for sending us the rainfall data analyzed in Example 4.2. We would also like to thank the Associate Editor and the referees for extremely constructive comments, which resulted in a remarkable improvement of an earlier version of this paper.

REFERENCES

AN, H.-Z. and BING, C. (1991). A Kolmogorov–Smirnov type statistic with application to test for nonlinearity in time series. *Internat. Statist. Rev.* **59** 287–307.

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

BOWMAN, A. and YOUNG, S. (1996). Graphical comparison of nonparametric curves. *Appl. Statist.* **45** 83–98.

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

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

DELGADO, M. A. and GONZÁLEZ-MANTEIGA, W. (2001). Significance testing in nonparametric regression based on the bootstrap. *Ann. Statist.* **29** 1469–1507.

DETTE, H. and NEUMEYER, N. (2001). Nonparametric analysis of covariance. *Ann. Statist.* **29** 1361–1400.

FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Soc.* **87** 998–1004.

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

GASSER, T., KNEIP, A. and KÖHLER, W. (1991). A flexible and fast method for automatic smoothing. *J. Amer. Statist. Assoc.* **86** 643–652.

GASSER, T., MÜLLER, H.-G. and MAMMITZSCH, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B* **47** 238–252.

HALL, P. and HART, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *J. Amer. Statist. Assoc.* **85** 1039–1049.

HALL, P., HUBER, C. and SPECKMAN, P. L. (1997). Covariate-matched one-sided tests for the difference between functional means. *J. Amer. Statist. Assoc.* **92** 1074–1083.

HÄRDLE, W. and MARRON, J. S. (1990). Semiparametric comparison of regression curves. *Ann. Statist.* **18** 63–89.

HJELLVIK, V. and TJØSTHEIM, D. (1995). Nonparametric tests of linearity for time series. *Biometrika* **77** 351–368.

KING, E. C., HART, J. D. and WEHRLY, T. E. (1991). Testing the equality of two regression curves using linear smoothers. *Statist. Probab. Lett.* **12** 239–247.

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

KULASEKERA, K. B. and WANG, J. (1997). Smoothing parameter selection for power optimality in testing of regression curves. *J. Amer. Statist. Assoc.* **92** 500–511.

MUNK, A. and DETTE, H. (1998). Nonparametric comparison of several regression functions: Exact and asymptotic theory. *Ann. Statist.* **26** 2339–2368.

NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.

NOLAN, D. and POLLARD, D. (1987). U-processes: Rates of convergence. *Ann. Statist.* **15** 780–799.

NOLAN, D. and POLLARD, D. (1988). Functional limit theorems for U-processes. *Ann. Probab.* **16** 1291–1298.

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

RICE, J. A. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215–1230.

SACKS, J. and YLVISAKER, D. (1970). Designs for regression problems with correlated errors. III. *Ann. Math. Statist.* **41** 2057–2074.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

STUTE, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613–641.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372.

WU, C. F. Y. (1986). Jacknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14** 1261–1350.

YOUNG, S. G. and BOWMAN, A. W. (1995). Nonparametric analysis of covariance. *Biometrics* **51** 920–931.

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

FAKULTÄT FÜR MATHEMATIK
RUHR-UNIVERSITÄT BOCHUM
44780 BOCHUM
GERMANY
E-MAIL: natalie.neumeyer@ruhr-uni-bochum.de
holger.dette@ruhr-uni-bochum.de