Comparison of Separable Components in Different Samples

Natalie Neumeyer
Ruhr-Universität Bochum
Fakultät für Mathematik
D – 44780 Bochum, Germany
natalie.neumeyer@ruhr-uni-bochum.de

Stefan Sperlich
Universidad Carlos III de Madrid
c / Madrid, 126
E – 28903 Getafe, Madrid, Spain
stefan@est-econ.uc3m.es

July 29, 2003

Abstract

Imagine we have two different samples and are interested in doing semi- or nonparametric regression analysis in each of them, possibly on the same econometric model. In this article we consider the problem of testing whether a specific covariate has different impacts on the regression curve in these two samples. So we compare regression curves of different samples but being interested in specific differences instead of testing for equality of the whole regression function. Our procedure does not only allow for random designs and different sample sizes but also for different variance functions, different sets of regressors with different impact functions, etc. Actually, it is as general as the comparison of particular coefficients in different parametric regression models but now on the level of (nonparametric) functionals. As we use the marginal integration approach this method can be applied to any strong, weak or latent separable model to compare the (lower dimensional) separable components between the different samples. Thus, in the case of separable models our procedure includes the possibility of comparing the whole regression curves avoiding the curse of dimensionality that otherwise would render such a task impractical. In practice, resampling methods are necessary for applying our test to real data. However, it will be shown that for our approach bootstrap fails in practice and theory. Instead, we propose a subsampling procedure with automatic parameter choice. We give complete asymptotic theory, and its excellent performance is demonstrated by an extensive simulation study. [i]

Keywords and Phrases: comparison of regression curves, nonparametric testing, subsampling, bootstrap, marginal effects, separable models

---

[i]This research was supported by the Spanish "Dirección General de Enseñanza Superior" BEC2001-1270 and the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures").

1 Introduction

Let us consider two different samples $\{Y_{1j}, X_{1j}, W_{1j}\}_{j=1}^{n_1}$, $\{Y_{2j}, X_{2j}, W_{2j}\}_{j=1}^{n_2}$ with Y_{1j}, Y_{2j}, X_{1j}, $X_{2j} \in I\!R$, and $W_{1j} \in I\!R^{d_1}$, $W_{2j} \in I\!R^{d_2}$. In practice, typically one has $d_1 = d_2$. For each sample we have in mind a regression model of the form

$$(1) \qquad E[Y_{1j}|X_{1j}, W_{1j}] = F_1(X_{1j}, W_{1j}) = G_1\{f_1(X_{1j}), g_1(W_{1j})\}, \quad j = 1, \ldots, n_1$$

$$(2) \qquad E[Y_{2j}|X_{2j}, W_{2j}] = F_2(X_{2j}, W_{2j}) = G_2\{f_2(X_{2j}), g_2(W_{2j})\}, \quad j = 1, \ldots, n_2,$$

where we can estimate f_1, f_2 nonparametrically, and the other functions g_1, g_2, G_1, G_2 either non-, semi- or parametrically. Notice that for each particular case of specification, there exists a considerable amount of literature how these models can be estimated non- or semiparametrically. The articles that provide estimators with its complete asymptotic theory for the most general models are, to our knowledge, Horowitz (2001), Rodríguez-Póo, Sperlich and Vieu (2003), Mammen and Nielsen (2003), Härdle, Huet, Mammen and Sperlich (2003), or Gozalo and Linton (2001).

In econometrics it is a common problem to know whether f_1 and f_2 are equal or whether there exist significant differences across populations. Typical examples are e.g. the comparison of labor supply offered by women and by men, comparisons between economic behavior in East and West Germany after unification, comparisons between races to find discrimination effects, etc.. In multivariate linear regression this has been a relatively easy task where possibly different variance functions could cause the only difficulty. In contrast, for nonlinear flexible regression models, especially for non- or semiparametric models, this task becomes a rather difficult one.

What concerns the comparison of $F_1(\cdot, \cdot)$, $F_2(\cdot, \cdot)$ much effort has been devoted to the problem of testing the equality of the whole regression curves [see Härdle and Marron (1990), Hall and Hart (1990), King, Hart and Wehrly (1991), Delgado (1993), Young and Bowman (1995), Kulasekera (1995), Hall, Huber and Speckman (1997), Dette and Munk (1998), Cabus (1998), Yatchew (1999), Kulasekera and Wang (2001), Dette and Neumeyer (2001), Lavergne (2001), Gørgens (2002), Neumeyer and Dette (2003)]. However, to our knowledge little (or even nothing) is known about the more sophisticated problem of comparing only parts of the regression function, even though in practice this question is somewhat more interesting.

Moreover, in the case of high dimensional predictors, due to the so–called curse of dimensionality, a direct comparison of F_1 and F_2 is even not reasonable at all. Then, if something is known about the structure of the regression as e.g. additivity or multiplicativity, this can help to surrender the curse of dimensionality, see Stone (1986) for generalized additive models, or Rodríguez-Póo, Sperlich and Vieu (2001, 2003) for weak or latent separable models. When we study the literature on estimation of models like (1), (2), we will realize that most of them make use of the marginal integration principle. Therefore, we will concentrate on this method applying the version of Hengartner and Sperlich (2001). To make reading easier, we first develop the asymptotic theory for models where f_1, respectively f_2, are strongly separable from the rest of the model and later discuss extensions. Further simplifications are $d_1 = d_2 = d$ and the assumption of having independent data. Extensions to $d_1 \neq d_2$ and dependent data structures are deferred to some remarks in the text.

To be more specific, we consider the following two (additive) nonparametric regression models in the general setting of heteroscedastic errors and different sample sizes in the two samples,

$$Y_{1j} \;=\; c_1 + f_1(X_{1j}) + g_1(W_{1j}) + \sigma_1(X_{1j}, W_{1j})\varepsilon_{1j}\,, \quad j = 1, \ldots, n_1\,, \tag{3}$$

$$Y_{2j} \;=\; c_2 + f_2(X_{2j}) + g_2(W_{2j}) + \sigma_2(X_{2j}, W_{2j})\varepsilon_{2j}\,, \quad j = 1, \ldots, n_2\,. \tag{4}$$

The two models are supposed to be independent of each other. The design points (X_{ij}, W_{ij}), $j = 1, \ldots, n_i$ shall have the joint density $p_i(x, w)$, $i = 1, 2$. The marginal densities are denoted by

$$\begin{aligned} r_i(x) &= \int p_i(x, w)\, dw \quad (\text{of } X_{ij}) \\ \text{and}\quad s_i(w) &= \int p_i(x, w)\, dx \quad (\text{of } W_{ij}), \quad i = 1, 2. \end{aligned}$$

We assume that X_{ij} is one dimensional and W_{ij} d–dimensional. The errors ε_{ij}, $j = 1, \ldots, n_i$, $i = 1, 2$, shall be independent of the covariates, independent and identically distributed with mean zero and variance one. For identification of the functions we assume

$$\begin{aligned} E[f_i(X_{ij})] &= \int f_i(x) r_i(x)\, dx = \int f_i(x) p_i(x, w)\, d(x, w) = 0\,, \\ E[g_i(W_{ij})] &= \int g_i(w) s_i(w)\, dw = \int g_i(w) p_i(x, w)\, d(x, w) = 0 \quad (i = 1, 2)\,, \end{aligned}$$

compare Sperlich, Tjøstheim and Yang (2002) for more details.

Our aim is to test the equality of the marginal impact, respectively of the separable function, i.e. the covariates X_{ij} $(i = 1, 2)$,; that is to test the following hypotheses,

$$H_0 : f_1 = f_2 \quad \text{versus} \quad H_1 : f_1 \neq f_2\,. \tag{5}$$

To get a first idea, take a most simple (i.e. parametric) model

$$F_i(X_{ij}, W_{ij}) \;=\; G_{\gamma_i}\{\alpha_i X_{ij}, \beta_i^T W_{ij}\} \quad , \; \gamma_i \in I\!R^q,\; q \in I\!N\,,$$

with unknown α_i, β_i, γ_i, $i = 1, 2$. Then, the synonym of our null hypothesis is simply $H_0 :\; \alpha_1 = \alpha_2$ versus $H_1 :\; \alpha_1 \neq \alpha_2$.

In Section 4 we will discuss the extension to other models of the form (1), (2) that are not necessarily additive or even not separable at all. However, the hypothesis and alternative will not change; what will change is only the question where f_1, f_2 are standing for.

Remark 1.1 Often one would like to test

$$H_0 : f_1 = S \circ f_2 \circ A \ , \quad S, A \ : \ I\!R \longrightarrow I\!R \ ,$$

where S, A are linear, maybe unknown, transformations. If the transformations are known, such an extension is straight forward. In the contrary case, we recommend to apply the method proposed by Härdle and Marron (1990) who considered such a case when f_1, f_2 are the regression functions itselves ($d = 0$).

A canonical test statistic for testing the equality of f_1 and f_2 could be a weighted L^2-distance between the estimated regression functions, that is,

$$T_N = \int \{\hat{f}_1(t) - \hat{f}_2(t)\}^2 v(t)\, dt \tag{6}$$

[see e. g. Härdle and Mammen (1993)] and the null hypothesis of equality of f_1 and f_2 should be rejected for large values of this test statistic. Here $v(\cdot)$ is a continuous and positive weight function. This test statistic T_N was considered by Dette and Neumeyer (2001) for testing the equality of regression functions in the case $f_i = F_i$ ($i = 1, 2$), see also the references listed there. The asymptotic distribution of this test statistic can be hard to estimate, depending on the regression model that is considered. In any case, it is well known that the asymptotics are little helpful in nonparametric testing to find the correct critical value, see Hjellvik, Yao and Tjøstheim (1998). We therefore first consider the possibility of applying wild bootstrap even though this is a discussable approach due to the fact that we do not always want to estimate the whole regression functions F_1, F_2 under the null, if this is possible at all. It will be seen that bootstrap fails to work anyway, not only in practice but also in theory. Fortunately, subsampling offers a viable alternative, see Politis, Romano and Wolf (1999, 2001) and Delgado, Rodríguez-Póo and Wolf (2001). We will introduce this procedure in detail because it never before has been applied in this context nor in one that at least comes close to ours, i.e. neither in a two sample problem, nor in nonparametric curve estimation or testing.

Notice that we have chosen the test statistic (6) only due to practical considerations. We do not state optimality but emphasize feasibility. Other test statistics and extensions will be discussed in Section 3. For the test statistic T_N defined in (6) asymptotic normality under the null hypothesis of equal functions $f_1 = f_2$ and (with a different convergence rate) under fixed alternatives is proved. The proposed tests can detect local alternatives converging to the null hypothesis with convergence rate $1/(N\sqrt{h_N})^{1/2}$ where h_N denotes a sequence of bandwidths tending to zero.

The rest of the paper is organized as follows. In Section 2 we introduce the estimators that we use and the regularity assumptions that are needed for the investigation of the asymptotic results. The main theoretical results are stated in Section 3, i.e. asymptotic distributions of T_N and alternative test statistics. The extension of the procedure to other models is considered in Section 4. In Section 5 it is shown that wild bootstrap can not be used for this test. Instead, we propose a subsampling procedure. We discuss practical questions and investigate the finite sample properties of the developed test in Section 6. The proofs of the theoretical results of Section 3 and 5 are deferred to the appendix.

2 The Estimator and Necessary Assumptions

In this section we first introduce briefly the estimator that is used to get the $\hat{f}_i$, $i = 1, 2$. Afterwards we give and discuss the assumptions we need to establish the asymptotics of the test.

The so called marginal integration estimator has first been introduced by Newey (1994) as an average mean estimator, and by Tjøstheim and Auestad (1994) and Linton and Nielsen (1995) as an estimator for additive and multiplicative models. Kim, Linton and Hengartner (1999), and Hengartner and Sperlich (2001) have modified and extended this estimator to make it computationally easier, faster and in its performance more robust against correlation among the covariates that unfortunately worsens a lot the performance of the classical version of the marginal integration estimator (CMIE). They call this modification internalized marginal integration estimator (IMIE). As mentioned in Section 1, since its introduction this estimation method has been applied to many different estimation problems and turned out to be quite helpful for handling rather general and complex models.

The internalized marginal integration estimators of f_1 and f_2 as we will use them for the test statistic T_N are defined as follows:

$$\hat{f}_i(x) \;=\; \frac{1}{n_i}\sum_{j=1}^{n_i} \gamma_{ij}(x) Y_{ij} - \hat{c}_i \tag{7}$$

for $i = 1, 2$, where we set

$$\gamma_{ij}(x) \;=\; K_h(X_{ij} - x)\frac{\hat{s}_i(W_{ij})}{\hat{p}_i(X_{ij}, W_{ij})}. \tag{8}$$

Here, $\hat{p}_i$ and $\hat{s}_i$ are estimators of the joint and marginal densities of (X_{ij}, W_{ij}), and W_{ij} respectively:

$$\hat{p}_i(x, w) \;=\; \frac{1}{n_i}\sum_{k=1}^{n_i} K_h(X_{ik} - x) L_b^{(d)}(W_{ik} - w) \tag{9}$$

$$\hat{s}_i(w) \;=\; \frac{1}{n_i}\sum_{k=1}^{n_i} L_b^{(d)}(W_{ik} - w) \tag{10}$$

with $K_h(\cdot) = \frac{1}{h}K(\cdot\; h^{-1})$ being a kernel function from $I\!R$ to $I\!R$, and $L_b^{(d)}(\cdot) = \frac{1}{b^d}L^d(\cdot\; b^{-1})$ a kernel from $I\!R^d$ to $I\!R$. Finally, we will use

$$\hat{c}_i \;=\; \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$$

as an estimator of the unconditional expectations $E[Y_{ij}] = c_i$ $(i = 1, 2)$. Since a lot is written about the performance and extensions of this estimator (see Sperlich, 1998) we skip further discussion and refer to the literature mentioned above and in Section 1.

We turn now to the regularity assumptions. They actually will turn out to be rather mild in practice. Throughout the paper we assume the following conditions to be satisfied:

[A1] When $N = n_1 + n_2$ denotes the total sample size we assume the existence of constants $\kappa_i \in (0, 1)$ such that

$$\lim_{n_1, n_2 \to \infty} \frac{n_i}{N} = \kappa_i \ , \quad i = 1, 2 \ .$$

Obviously this assumption is purely technical to establish more easily the convergence rates, but it has no further implication in practice.

[A2] We assume continuity of g_i for $i = 1, 2$.

[A3] The regression functions f_i and the marginal design densities r_i shall be r–times continuously differentiable $(i = 1, 2)$ for some $r \geq 2$.

[A4] Let K be a symmetric kernel with compact support, $\int K^2(u)\, du < \infty$, and of order r (from [A3]), that is,

$$\int K(u) u^j \, du = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } j = 1, \ldots, r-1 \end{cases} \quad .$$

Let $h = h_N$ denote a sequence of bandwidths and set in the following $K_h(x) = K(x/h)/h$.

These last conditions are rather natural and common in kernel smoothing to guarantee some smoothness of the regression functions, see Gasser, Müller and Mammitzsch (1985).

[A5] The marginal densities s_i shall be s–times continuously partially differentiable for $i = 1, 2$ and some $s \geq 2$.

[A6] Let L denote a symmetric kernel of order s (from [A5]) with compact support, and $b = b_N$ a sequence of bandwidths. Then $L^{(d)}$ denotes the product kernel $L^{(d)}(w_1, \ldots, w_d) = L(w_1) \cdots L(w_d)$ with notation $L_b^{(d)}(w) = L^{(d)}(w/b)/b^d$.

[A7] We assume the design densities p_i to be $(r \wedge s)$–times continuously partially differentiable and bounded away from zero for $i = 1, 2$ (r, s from [A3] to [A6]).

These three conditions are related to the use of marginal integration but much less restrictive and more realistic than those necessary for the CMIE. Notice that concerning the covariates not of interest (i.e. W_{ij}) we only ask for smoothness in the densities but not in the functions g_i. For a detailed discussion and how these conditions can even be relaxed further, see Hengartner, Sperlich (2001). Only for the ease of notation in the proof, we set the support of the p_i equal to the cube $[0, 1]^{d+1}$.

[A8] The variance functions σ_i^2, $i = 1, 2$, are continuous and bounded away from zero and infinity. The distributions of the errors have a finite fourth moment, that is $E[\varepsilon_{ij}^4] < \infty$ $\forall\ i, j$.

It is clear that this is assumed to establish consistency and convergence in distribution.

[A9] We assume the following bandwidth conditions to be satisfied:

$$h \to 0,\ b \to 0,\ Nh^{3/2}b^d \to \infty,\ Nb^{2d}h \to \infty$$

$$Nh^{2r+1} = o(1),\ N\sqrt{h}b^{2s} = o(1),\ Nb^s h^{r+\frac{1}{2}} = o(1),\ h^{r+\frac{1}{2}}/b^d = o(1).$$

To yield a consistent test, these conditions are certainly somewhat stronger as those needed for the consistency of the IMIE only . Actually, they might look worse than they are: e.g. in the special case where $r = s$, $d = 1$, and bandwidths h, b chosen to be of the same convergence rate, this gives just the constraints $h \to 0$, $Nh^3 \to \infty$ and $Nh^{2r+\frac{1}{2}} \to 0$.

However, we will give one example how these bandwidth conditions can already be relaxed. For the ease of presentation, implementation and for historical reasons we have given the definition of the IMIE in (7) with the typical kernel density estimates (10) and (9) for the p_i, s_i. This will cause bias terms of order $1/(Nb^d)$ and $1/(Nhb^d)$ in the density estimators $\hat{s}_i$ and $\hat{p}_i$, respectively. If we modify slightly the estimators used in definitions (7) and (8) in the following way:

$$(11)\ \widetilde{p_i}(X_{ij}, W_{ij}) = \frac{1}{n_i}\sum_{\substack{k=1\\k\neq j}}^{n_i} K_h(X_{ik} - X_{ij})L_b^{(d)}(W_{ik} - W_{ij}) = \hat{p}_i(X_{ij}, W_{ij}) - \frac{K(0)L^{(d)}(0)}{n_i h b^d}$$

$$(12)\qquad \widetilde{s_i}(W_{ij}) = \frac{1}{n_i}\sum_{\substack{k=1\\k\neq j}}^{n_i} L_b^{(d)}(W_{ik} - W_{ij}) = \hat{s}_i(W_{ij}) - \frac{L^{(d)}(0)}{n_i b^d},$$

the theoretical results remain valid under the following less restrictive bandwidth conditions,

[A9']

$$h \to 0,\ b \to 0,\ Nh \to \infty$$

$$Nh^{2r+1} = o(1),\ N\sqrt{h}b^{2s} = o(1),\ Nb^s h^{r+\frac{1}{2}} = o(1).$$

Then, in the special case where $r = s$, $d = 1$, and h, b chosen to be of the same convergence rate, this gives the constraints $h \to 0$, $Nh \to \infty$ and $Nh^{2r+\frac{1}{2}} \to 0$.

Finally, recall that only for the ease of notation we restrict our presentation of the results to the case where the dimension of covariates $(d+1)$ in the two models (3) and (4) is the same. But realize that everything can be straightforwardly generalized to the case of different dimensions of covariates d_1+1 in model (3) and d_2+1 in model (4). For instance, in that case it can also be reasonable to use different kernels L_i and bandwidths b_i for the two models $i = 1, 2$ (instead of using simply L, b for both samples the same).

3 Main Results

We now investigate the asymptotic behavior of the test statistic

$$T_N \;=\; \int \{\hat{f}_1(t) - \hat{f}_2(t)\}^2 v(t)\, dt$$

and some alternatives to test the hypotheses

$$H_0:\; f_1 = f_2 \qquad vs. \qquad H_1:\; f_1 \neq f_2.$$

As mentioned above, $v(t)$ is a weight function that has to be chosen by the empirical researcher and will be discussed later.

If we consider models of the form (3), (4), i.e. where f_1, f_2 are additively separable from the remainder terms, we can state the following theorem:

Theorem 3.1 *Suppose the assumptions [A1] to [A9] stated in Section 2 are satisfied.*
(i) Under the null hypothesis $H_0 : f_1 = f_2$ the standardized test statistic

$$N\sqrt{h}(T_N - E_{H_0}[T_N])$$

converges in distribution to a mean zero normal distribution with variance β^2 defined by

$$\beta^2 = 2\int (K * K)^2(u)\, du \int v^2(z)\Big(\int H(z,w)\, dw\Big)^2 dz$$

where $K * K$ denotes the convolution of the kernel K with itself and

$$H(z,w) \;=\; s_1^2(w)\frac{\{c_1 + f_1(z) + g_1(w)\}^2 + \sigma_1^2(z,w)}{\kappa_1 p_1(z,w)} + s_2^2(w)\frac{\{c_2 + f_2(z) + g_2(w)\}^2 + \sigma_2^2(z,w)}{\kappa_2 p_2(z,w)}.$$

The expectation is given by

$$E_{H_0}[T_N] \;=\; \frac{1}{Nh}\int K^2(u)\, du \int v(z)H(z,w)\, d(z,w) + o(\frac{1}{N\sqrt{h}}).$$

(ii) Under the alternative $H_1 : f_1 \neq f_2$ the standardized test statistic

$$\sqrt{N}(T_N - E_{H_1}[T_N])$$

converges in distribution to a mean zero normal distribution with variance γ^2 defined by

$$\gamma^2 = \frac{4}{\kappa_1} Var(Z^{(1)}) + \frac{4}{\kappa_2} Var(Z^{(2)})$$

where for $i = 1, 2$

$$Z^{(i)} \;=\; Y_{i1}\Big[\{f_1(X_{i1}) - f_2(X_{i1})\}v(X_{i1})\frac{s_i(W_{i1})}{p_i(X_{i1}, W_{i1})} - \int \{f_1 - f_2\}(x)v(x)\, dx\Big].$$

The expectation is given by

$$E_{H_1}[T_N] \;=\; \int \{f_1 - f_2\}^2(x)v(x)\, dx + O(h^r) + O(b^s) + O(\frac{1}{Nb^d h}).$$

The proof of the theorem which is rather technical and cumbersome is deferred to the appendix.

The first part of the theorem shows that an asymptotic level–α–test is obtained by rejecting the null hypothesis H_0 of equality of f_1 and f_2 whenever

$$N\sqrt{h}\Big(T_N - \frac{1}{Nh}\int K^2(u)\,du \int v(z)H(z,w)\,d(z,w)\Big) \;>\; \beta\Phi^{-1}(1-\alpha).$$

Here Φ denotes the distribution function of the standard normal distribution and β^2 is defined in Theorem 3.1. The formula of the function $H(z,w)$ stated in the Theorem contains unknown features of the data generating process like the design densities p_i , the regression functions g_i and the variance functions σ_i^2 (always for $i = 1,2$). They have to be estimated what can be rather complicated depending on the model under consideration. Moreover, even in the cases where the estimation of the asymptotic bias and variance would be feasible, Hjellvik, Yao and Tjøstheim (1998) showed very well that these asymptotic expressions are little helpful to find the correct critical value in finite samples. For these reasons the proposed test does not provide a satisfactory approximation of the nominal level when there are only small samples available. In Section 5 we consider resampling procedures to circumvent this obstacle.

Due to the second part of the theorem we can establish the consistency of the proposed testing procedure and moreover obtain the following approximation of the power:

$$P\Big(\text{test rejects } H_0 \;\Big|\; H_1\Big) \;\approx\; \Phi\Big(\frac{\sqrt{N}}{\gamma}\int \{f_1 - f_2\}^2(x)v(x)\,dx\Big).$$

Remark 3.2 A careful inspection of the proof of Theorem 3.1 shows that under local alternatives of the form

$$H_{1,N}: \quad f_1(x) = f_2(x) + \frac{\delta(x)}{(N\sqrt{h})^{1/2}}$$

the under H_0 standardized test statistic $N\sqrt{h}(T_N - E_{H_0}[T_N])$ converges in distribution to a normal distribution with variance β^2 and mean

$$\int \delta^2(x)v(x)\,dx.$$

Here $E_{H_0}[T_N]$ and β^2 are defined in Theorem 3.1.

There are several alternative test statistics that can be considered for testing the null hypothesis (5). The maybe most obvious alternative to T_N is to substitute the integral by the average over the design points, i.e.

$$S_N^{(1)} \;=\; \frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\{\hat{f}_1(X_{ij}) - \hat{f}_2(X_{ij})\}^2 w(X_{ij}) \tag{13}$$

with a weight function $w(\cdot)$ that is directly related to $v(\cdot)$. As can be seen easily, $S_N^{(1)}$ is just the empirical approximation of

$$\int \{\hat{f}_1(x) - \hat{f}_2(x)\}^2 \{\kappa_1 r_1(x) + \kappa_2 r_2(x)\} \; w(x)\, dx$$

which is the original test statistic T_N with weight function

$$v(x) \;=\; \{\kappa_1 r_1(x) + \kappa_2 r_2(x)\} \; w(x) \; . \tag{14}$$

Then, under the null hypothesis $H_0: \; f_1 = f_2$ the statistic $S_N^{(1)}$ has the same asymptotic behavior as stated in Theorem 3.1 for T_N. Under the alternative the asymptotic variance changes to

$$\begin{aligned} \gamma_1^2 \;&=\; \frac{1}{\kappa_1} \mathrm{Var}\Big(2Z^{(1)} + \{f_1(X_{11}) - f_2(X_{11})\}^2 w(X_{11})\kappa_1\Big) \\ &\quad + \frac{1}{\kappa_1} \mathrm{Var}\Big(2Z^{(2)} - \{f_1(X_{21}) - f_2(X_{21})\}^2 w(X_{21})\kappa_2\Big) \end{aligned} \tag{15}$$

where $Z^{(1)}$ and $Z^{(2)}$ are defined in Theorem 3.1 with $v(x) = \{\kappa_1 r_1(x) + \kappa_2 r_2(x)\} w(x)$.

This test statistic, even though quite intuitive, has some essential drawbacks. Nonparametric estimates perform quite poor where data are sparse, neither they should be used for extrapolation. Imagine, you have some areas where the sample density $\hat{r}_1$ is close (or equal) to zero but the sample density $\hat{r}_2$ is large, and / or the other way around. Right those areas you would like to weight down or even trim out in your test statistic. You still could arrange this with a proper choice of $w(\cdot)$ but it would be quite cumbersome to do it that way. Especially in the case of having pretty different design densities in the two samples it is more reasonable to evaluate $\hat{f}_i$ only in points X_{ij} from the ith sample $(i = 1, 2)$.

These considerations lead us to the following test statistic:

$$S_N^{(2)} \;=\; \frac{1}{n_1 n_2} \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} \{\hat{f}_1(X_{1j}) - \hat{f}_2(X_{2k})\}^2 \; \frac{1}{a} \; k\Big(\frac{X_{1j} - X_{2k}}{a}\Big) \; ,$$

where $a = a_N$ denotes a new sequence of bandwidths converging to zero, and k is a kernel function. This test statistic now has the asymptotic expectation

$$\int\!\!\int \{f_1(x) - f_2(y)\}^2 \frac{1}{a} k\Big(\frac{x-y}{a}\Big) r_1(x) r_2(y)\, dx\, dy \;\approx\; \int \{f_1(x) - f_2(x)\}^2 r_1(x) r_2(x)\, dx.$$

The asymptotic distribution of $S_N^{(2)}$ under the null hypothesis H_0 is thus similar to the one stated for T_N in Theorem 3.1 but with weight function $v(x) = r_1(x) r_2(x)$. To get this asymptotic result we need some conditions to be satisfied additional to the assumptions made in Section 2 :

[B1] The densities r_1 and r_2 of the first covariates of each model shall be ρ–times continuously differentiable with $1 \leq \rho \leq r$.

[B2] The kernel $k(\cdot)$ shall be of order ρ.

[B3] The bandwidths have to satisfy the constraints

$$a^\rho = o(\sqrt{h}),\ Nh^{2r+\frac{1}{2}} = O(1),\ Nh^2 \to \infty,\ h^{4r+1} = o(a),\ b^{4s}h = o(a),\ N\sqrt{ah} \not\to 0.$$

Under the alternative H_1 an extra bias term of order $O(a^\rho)$ appears in the expectation and the asymptotic variance of $\sqrt{N}(S_N^{(2)} - E_{H_1}[S_N^{(2)}])$ is of the form

$$\gamma_2^2 = \frac{1}{\kappa_1}\mathrm{Var}\Big(2Z^{(1)} + \{f_1(X_{11}) - f_2(X_{11})\}^2 r_2(X_{11})\Big) + \frac{1}{\kappa_1}\mathrm{Var}\Big(2Z^{(2)} - \{f_1(X_{21}) - f_2(X_{21})\}^2 r_1(X_{21})\Big) . \tag{16}$$

Again, $Z^{(1)}$ and $Z^{(2)}$ are defined in Theorem 3.1 with $v(x) = r_1(x)r_2(x)$.

To summarize we can state the following results:

Corollary 3.3 *Assume that conditions [A1] to [A9] hold. Then*

(i) the test statistic $N\sqrt{h}S_N^{(1)}$ *behaves asymptotically like* $N\sqrt{h}T_N$ *with* $v(\cdot)$ *as in equation (14), but its variance changes under the alternative* H_1 *to* γ_1^2 *defined in equation (15).*

(ii) If further hold the assumptions [B1] to [B3], the test statistic $S_N^{(2)}$ *behaves asymptotically like* $N\sqrt{h}T_N$ *with* $v(t) = r_1(t)r_2(t)$*, but its variance changes under the alternative* H_1 *to* γ_2^2 *defined in equation (16).*

Instead of considering estimators for the L^2–distance of the functions f_1, f_2, testing problems of this kind have also been handled with empirical process approaches, see for example Delgado (1993), Cabus (1998) or Neumeyer and Dette (2003). They applied this for testing the equality of the whole regression functions F_1, F_2 in a nonparametric context. In our context, one could consider, for instance, the following process in $t \in [0, 1]$:

$$S_N^{(3)}(t) = \frac{1}{n_1 n_2}\sum_{j=1}^{n_1}\sum_{k=1}^{n_2}\{\hat{f}_1(X_{1j}) - \hat{f}_2(X_{2k})\}\frac{1}{a}k\Big(\frac{X_{1j} - X_{2k}}{a}\Big)I\{X_{1j} \leq t\}I\{X_{2k} \leq t\} .$$

This obviously is an estimator for the integral

$$\int_0^t \{f_1(x) - f_2(x)\}r_1(x)r_2(x)\,dx.$$

Then, the null hypothesis $H_0 : f_1 = f_2$ should be rejected for large values of a Kolmogorov–Smirnov type statistic

$$\sup_{t\in[0,1]} |S_N^{(3)}(t)|$$

or a Cramer–von–Mises type statistic

$$\int_0^t \{S_N^{(3)}(t)\}^2\, dt.$$

The development of the asymptotic theory for this type of tests is quite different from the one we did for the test statistics mentioned above and therefore would be beyond the scope of this paper. For more details we therefore refer to Neumeyer and Dette (2003).

Finally, coming back to the aforementioned approaches we could also make use of a pooled sample estimator under the null hypothesis H_0. Indeed, the observations from both additive regression models can be used to estimate a common function $f = f_1 = f_2$, i.e. we consider a pooled sample of the form

$$Y_{ij} \;=\; f(X_{ij}) + c_i + g_i(W_{ij}) + \sigma_i(X_{ij}, W_{ij})\varepsilon_{ij}\,, \quad j = 1, \ldots, n_i,\; i = 1, 2.$$

In generalization of definition (7) we then define the estimator $\tilde f$ (to get f) by

$$\tilde f(x) \;=\; \frac{1}{N}\sum_{\ell=1}^{2}\sum_{k=1}^{n_\ell}\frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i} K_h(X_{ij}-x)L_b^{(d)}(W_{ij}-W_{\ell k})\frac{(Y_{ij}-\hat c_i)}{\hat p(X_{ij}, W_{ij})}\,,$$

where we set

$$\hat p(x, w) \;=\; \frac{1}{N}\sum_{i=1}^{2}\sum_{k=1}^{n_i} K_h(X_{ik}-x)L_b^{(d)}(W_{ik}-w)\;.$$

This is an estimator for the convex combination of densities p_1, p_2, in particular

$$p(x, w) \;=\; \kappa_1 p_1(x, w) + \kappa_2 p_2(x, w).$$

It can be shown that $\tilde f(x)$ is an estimator for

$$\begin{aligned}
&\sum_{\ell=1}^{2}\sum_{i=1}^{2}\kappa_\ell\kappa_i \int\!\!\int K_h(y-x)L_b^{(d)}(w-v)\{f(y)+g_i(w)\}\frac{p_i(y,w)}{p(y,w)}s_\ell(v)\, d(y,w)\, dv \\
\approx\; & f(x) + \int \{g_1(w)p_1(x,w) + g_2(w)p_2(x,w)\}\frac{1}{p(x,w)}(s_1(w)+s_2(w))\, dw.
\end{aligned}$$

Now an estimator $\hat f$ for the common function $f = f_1 = f_2$ is obtained by centering, i.e.

$$\hat f(x) \;=\; \tilde f(x) - \frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\tilde f(X_{ij}). \tag{17}$$

A possible test statistic based on the pooled estimate is, for example,

$$S_N^{(4)} = \frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\{\hat{f}(X_{ij}) - \hat{f}_i(X_{ij})\}^2,$$

see for comparison Young and Bowman (1995) or Dette and Neumeyer (2001) who used such a statistic to test the equality of regression functions.

4 Extensions to Other Models

As we mentioned at the very beginning, the marginal integration estimator has been used to estimate many different models apart from the one described in equation (3) and (4). Most of them are highly interesting for applied econometrics. Therefore we think it is worthwhile to give a brief review discussing the application of our proposed test statistic T_N on these different models. The extensions to the alternative test statistics are straight forward and therefore skipped here.

We start with the most general model. We can consider separable models of the form

$$Y_{ij} = G_i\{f_i(X_{ij}), g_i(W_{ij})\} + \sigma_i(X_{ij}, W_{ij})\varepsilon_{ij}\,, \quad j = 1, \ldots, n_i,\ i = 1, 2,$$

where G_i is not necessarily additive. Here, f_1 and f_2 can be interpreted as the marginal impact of the first components X_{1j} and X_{2j} or as (the first) separable component functions. Under the assumption of the existence of functions q_i such that

$$\int G_i\{f_i(x), g_i(w)\}q_i(w)\, dw = f_i(x) \quad \text{for } i = 1, 2\ ,$$

the functions f_1 and f_2 can be identified and estimated by the IMIEs $\hat{f}_1$ and $\hat{f}_2$ defined by

$$\hat{f}_i(x) = \frac{1}{n_i}\sum_{j=1}^{n_i}\gamma_{ij}(x)Y_{ij}\ .$$

Here, the estimators for the densities s_i in the definition of $\gamma_{ij}(x)$ in equation (8) have to be substituted by the functions q_i $(i = 1, 2)$. Actually, this is not surprising as already in Sperlich, Linton and Härdle (1999) and later in Sperlich, Tjøstheim and Yang (2002) or Hengartner and Sperlich (2001) it was shown that all versions of the marginal integration estimate consistently the marginal impact whatever the real underlying model is. Appropriate identification conditions do the rest.

Thus, for testing the hypothesis of equality $H_0 : f_1 = f_2$ we can consider the same testing procedure T_N defined in (6) as we did it in the additive case. A careful inspection of the proof of Theorem 3.1 shows that the asymptotic behavior of T_N stated in that theorem is also valid in this model. The asymptotic expectations and variances actually change only slightly. For instance, the function $H(z,w)$ defined in Theorem 3.1 has now the form

$$H(z,w) \;=\; q_1^2(w)\frac{G_1^2\{f_1(z),g_1(w)\}+\sigma_1^2(z,w)}{\kappa_1 p_1(z,w)} + q_2^2(w)\frac{G_2^2\{f_2(z),g_2(w)\}+\sigma_2^2(z,w)}{\kappa_2 p_2(z,w)} \;.$$

Similarly, we can consider generalized additive models

$$E[Y_i|X_i,W_i] \;=\; G_i\{f_i(X_i)+g_i(W_i)\}\;,\; i=1,2\;,$$

see e.g. Horowitz (2001), Gozalo and Linton (2001), or even generalized additive partial linear models where some of the covariates of W_{ij} enter linearly. This can be important if some dummy variables form part of these covariates, see Härdle, Huet, Mammen and Sperlich (2003).

If the link functions G_i depend further on some parameter, this will not change our asymptotics as long as these parameter can be estimated with the parametric rate and consequently do not affect the slower nonparametric expressions, see e.g. Gozalo and Linton (2001) or Rodríguez-Póo, Sperlich and Vieu (2003).

The particular case of estimating additive partial linear models with IMIE, i.e. modelling some parts of W_i linearly and G_i being the identity, can also be found in Hengartner and Sperlich (2001). From that article it can be seen that the theory worked out for our test in Section 3 works also through for additive partial linear models without any additional problems.

Finally, consider the interaction model (denoting now $U_i = (X_i, W_i)$)

$$E[Y_i|U_i] \;=\; \sum_{\alpha=1}^{d+1}\eta_{i,\alpha}(U_{i\alpha}) + \sum_{\alpha<\beta}\eta_{i,\alpha\beta}(U_{i\alpha},U_{i\beta})\;,\; i=1,2,$$

with $\eta_{i,\alpha}$, $\eta_{i,\alpha\beta}$ being unknown smooth functions, as introduced in Sperlich, Tjøstheim and Yang (2002). Again, all these components can be estimated by marginal integration. If we denote $f_i = \eta_{i,1}$, then again it would be straight forward to give the asymptotics of our test statistic as the asymptotics for the $\hat{f}_i$ do not change compared to the strong separable models (3), (4). However, if the included interactions between X_i and W_i turn out to be significant the interpretation of the null hypothesis becomes harder in empirical research. The existence of those nonparametric interactions can be tested for significance by Härdle, Sperlich and Spokoiny (2001) or Sperlich, Tjøstheim and Yang (2002).

Finally, for the case of dependent data, Camlong-Viot (2000) has extended the main results of IMIE to data that fulfill $\beta-$mixing conditions. Moreover, she demonstrated that also the typical additivity tests with IMIE can be extended to dependent data with $\beta-$mixing. From that results it is quite clear that this is also possible for our test statistic. However, as the proofs for our theoretical results are rather technical, we have given all theorems only for the independent case.

5 Resampling Methods

As we mentioned above, in finite samples it will be necessary to simulate the distribution of the test statistic T_N under the null hypothesis. Let us first consider the wild bootstrap, afterwards the subsampling procedure. Even though there is nothing to prove for the consistency of the subsampling procedure whereas for bootstrap it is, bootstrap always is preferable as long as it is consistent. The reason is simply that it converges faster and has thus a much better power. On the other hand, the subsampling method has one interesting advantage: we are not obliged to estimate the whole regression models, i.e. F_1 and F_2, under the null. Having at hand consistent estimators for f_1, f_2, respectively for T_N , is therefore enough.

5.1 The Wild Bootstrap

Again, we restrict to the following model

$$Y_{ij} \;=\; c_i + f_i(X_{ij}) + g_i(W_{ij}) + \sigma_i(X_{ij}, W_{ij})\varepsilon_{ij} \;, \quad j = 1, \ldots, n_i,\; i = 1, 2 \;.$$

Having estimated the f_i and g_i by IMIE and c_i as described above in Section 2, we define the residuals by

$$\begin{aligned} \hat{\varepsilon}_{ij} \;&=\; Y_{ij} - \hat{c}_i - \hat{f}_i(X_{ij}) - \hat{g}_i(W_{ij}) \\ &=\; \sigma_i(X_{ij}, W_{ij})\varepsilon_{ij} + c_i - \hat{c}_i + f_i(X_{ij}) - \hat{f}_i(X_{ij}) + g_i(W_{ij}) - \hat{g}_i(W_{ij}) \;. \end{aligned}$$

Then, the bootstrap observations are drawn from the following data generating process under the null hypothesis H_0:

$$Y^*_{ij} \;=\; \hat{c}_i + \hat{f}(X_{ij}) + \hat{g}_i(W_{ij}) + V_{ij}\hat{\varepsilon}_{ij} \qquad \forall\; i, j \;.$$

Notice that now, $\hat{f}$ has to be an estimator under the null. We get this by taking the estimator defined in equation (17) from the pooled sample. The V_{ij} denote independent identically distributed random variables with expectations $E[V_{ij}] = 0$ and variances $\mathrm{Var}(V_{ij}) = 1$, and are independently from from the sample

$$\mathcal{Y}_N \;=\; \{(X_{ij}, W_{ij}, Y_{ij}) \mid j = 1, \ldots, n_i, i = 1, 2\}.$$

Let $E^*_N = E[\cdot \mid \mathcal{Y}_N]$ and $\mathrm{Var}^*_N = \mathrm{Var}(\cdot \mid \mathcal{Y}_N)$ denote the conditional expectation and variance, respectively, conditioned on the original sample $\mathcal{Y}_N$. Then we have

$$E^*_N[Y^*_{ij}] \;=\; \hat{c}_i + \hat{f}(X_{ij}) + \hat{g}_i(W_{ij}), \tag{18}$$

$$E^*_N[(Y^*_{ij})^2] \;=\; \Big\{\hat{c}_i + \hat{f}(X_{ij}) + \hat{g}_i(W_{ij})\Big\}^2 + \hat{\varepsilon}^2_{ij}, \tag{19}$$

$$\text{and} \qquad \mathrm{Var}^*_N(Y^*_{ij}) \;=\; \hat{\varepsilon}^2_{ij}.$$

Analogous to (7) we estimate the functions f_1, f_2 from the bootstrap samples, i.e.

$$\hat{f}_i^*(x) = \frac{1}{n_i}\sum_{j=1}^{n_i}\{\gamma_{ij}(x) - 1\}Y_{ij}^*,$$

where the weights

$$\gamma_{ij}(x) = K_h(X_{ij} - x)\frac{\hat{s}_i(W_{ij})}{\hat{p}_i(X_{ij}, W_{ij})}$$

are defined in (8), Section 2. We end up with a bootstrap version of the test statistic under the null hypothesis by defining

$$T_N^* = \int \{\hat{f}_1^*(t) - \hat{f}_2^*(t)\}^2 v(t)\, dt. \tag{20}$$

When calculating the asymptotic distribution, to simplify matters we use the same bandwidths for the bootstrap test statistic as for the original procedure. Additionally to the assumptions made in Section 2 we need the following bandwidth conditions:

[C1] We assume that $Nh^{2r} \to 0$, $Nb^{2s} \to 0$, and $Nh^2b^{2d} \to \infty$.

The following theorem gives the variance of the asymptotic distribution of the bootstrap statistic (20).

Theorem 5.1 *The conditional variance $\beta_N^{*2} = Var_N^*(N\sqrt{h}T_N^*)$ of the bootstrap version $N\sqrt{h}T_N^*$ of the test statistic, conditioned on the sample $\mathcal{Y}_N$, converges in probability to*

$$\begin{aligned}\tilde{\beta}^2 &= \int (K*K)^2(u)\, du \int v^2(z)\Big[\int \sum_{i=1}^{2} s_i^2(w)\frac{\{c_i + f_i(z) + g_i(w)\}^2 + \sigma_i^2(z,w)}{\kappa_i p_i(z,w)}\, dw\Big]^2 dz \\ &\quad - \int (K*K)^2(u)\, du \int v^2(z)\Big[\int \sum_{i=1}^{2} s_i^2(w)\frac{\{c_i + f_i(z) + g_i(w)\}^2}{\kappa_i p_i(z,w)}\, dw\Big]^2 dz \\ &\neq \beta^2,\end{aligned}$$

where β^2 is the asymptotic variance of the original test statistic $N\sqrt{h}T_N$ (under the null hypothesis) defined in Theorem 3.1.

The proof of Theorem 5.1 is deferred to the appendix.

Corollary 5.2 *Conditional on the sample $\mathcal{Y}_N$, the bootstrap test statistic $N\sqrt{h}T_N^*$ does <u>not</u> converge in distribution to the limit distribution of the original test statistic $N\sqrt{h}T_N$ (under the null hypothesis H_0), as stated in Theorem 3.1.*

Further problems arise from the observation, that the bootstrap bias $N\sqrt{h}(T_N^* - E_{H_0}[T_N])$ (where $E_{H_0}[T_N]$ is stated in Theorem 3.1) does <u>not</u> converge to zero in probability. In fact, the difference $E_{H_0}^*[T_N^*] - E_{H_0}[T_N]$ includes a term of the form

$$\int \frac{1}{n_1^2} \sum_{j=1}^{n_1} \sum_{\substack{k=1 \\ k \neq j}}^{n_1} \Big\{ (\gamma_{1j}(x) - 1)(\hat{c}_1 + \hat{f}(X_{1j}) + \hat{g}_1(W_{1j})) - \hat{f}(x) \Big\}$$
$$\Big\{ (\gamma_{1k}(x) - 1)(\hat{c}_1 + \hat{f}(X_{1k}) + \hat{g}_1(W_{1k})) - \hat{f}(x) \Big\} v(x) dx$$

and it holds

$$\int \frac{1}{n_1^2} \sum_{j=1}^{n_1} \sum_{\substack{k=1 \\ k \neq j}}^{n_1} (\gamma_{1j}(x) - 1)(\gamma_{1k}(x) - 1) v(x) dx \; = \; O_p(\frac{1}{N\sqrt{h}}) \; .$$

In practice this means that one has either to find other estimation methods or test statistics that do not suffer from this defect, or one has to apply different resampling methods. In Section 6 this result is illustrated by a simulations study.

Remark 5.3 It is clear from Theorem 5.1 and confirmed by the simulations that the wild bootstrap procedure does not work in particular in the case where the variances are small compared to the functionals f_i , g_i . For that reason it is interesting to consider what happens in the degenerate case where

$$\sigma_i^2 \longrightarrow 0 \quad (i = 1, 2).$$

We then have $\tilde{\beta}^2 = 0$ for the asymptotic variance in Theorem 5.1, but still $\beta^2 > 0$ for the asymptotic variance from Theorem 3.1. That means

$$\mathrm{Var}(N\sqrt{h}T_N^* \mid \mathcal{Y}_N) \xrightarrow{P} 0$$

but $N\sqrt{h}(T_N - E_{H_0}[T_N])$ converges (under the null hypothesis) to a nondegenerate normal distribution.

On this account comparing $N\sqrt{h}T_N$ with the quantiles of the distribution of $N\sqrt{h}T_N^*$ (conditioned on $\mathcal{Y}_N$) leads far to often to a rejection of the null hypothesis.

As indicated above, other estimation or bootstrap procedures might work. However, to find such procedure is not trivial. Let us demonstrate this by a simple example. Drawing the bootstrap observations from the data generating process

$$Y_{ij}^* = \hat{c}_i + \hat{f}(X_{ij}) + \hat{g}_i(W_{ij}) + V_{ij}\hat{\varepsilon}_{ij}$$

as suggested above might not be quite appropriate in our model for the following reason: For the aim of comparing the functions f_1 and f_2, the impact of the covariates not of interest $g_i(W_{ij})$ has the nature of an additional "error" variable. Hence, one could also suggest to estimate the "total error" by

$$e_{ij} = Y_{ij} - \hat{c}_i - \hat{f}_i(X_{ij})$$

and draw bootstrap observations from

$$Y^*_{ij} = \hat{c}_i + \hat{f}(X_{ij}) + V_{ij} e_{ij} \ .$$

But notice that this approach does not attain the goal neither, i.e. $\text{Var}^*_N(N\sqrt{h}T^*_N)$ does not converge to the variance of the asymptotic distribution of $N\sqrt{h}T_N$ under H_0. One has to find different proceedings for applying a bootstrap procedure for this testing problem, if this is possible at all.

5.2 The Subsampling Method

As mentioned before, even if one could find a testing method for our problem for that bootstrap does work, it is often preferable to have a resampling method at hand that does not require the estimation of the whole model (under the null). This is guaranteed when we use subsampling.

We try to concentrate in the introduction of subsampling on the particular setup of our testing problem. For more details we refer to Politis, Romano and Wolf (1999, 2001) and Delgado, Rodríguez-Póo and Wolf (2001). As so far this method has never been introduced for the two-sample context, neither for the nonparametric analysis of functionals, our notation is slightly different from the traditional one.

We have given a test statistic

$$\Upsilon_N \ = \ N\sqrt{h_N} t_N(\mathcal{Y}_N) \ \text{ where } \ t_N(\mathcal{Y}_N) = T_N - E_{H_0}[T_N] \ ,$$

and $t_N(\mathcal{Y}_N)$ converges in probability to zero under the null and to some positive constant under the alternative. Define now

$$Q_N(x,P) \ = \ Prob_P(\Upsilon_N \leq x) \ = \ Prob_P(N\sqrt{h_N} t_N(\mathcal{Y}_N) \leq x) \ ,$$

where P is the data generating process. Note that we used here the index N for bandwidth h to emphasize its rate-dependence on N. Why this is important will be seen later. Let $\mathcal{Y}^1_M, \cdots, \mathcal{Y}^{N_M}_M$ be equal to the N_M subsets of $\mathcal{Y}_N$ of size $M < N$, but always with the same proportions $\hat{\kappa}_1 = n_1/m_1$, $\hat{\kappa}_2 = n_2/m_2$, $m_1 + m_2 = M$, m_i being the subsample size of sample i. The sampling distribution of Υ_N can then be approximated by the subsampling distribution by

$$\widehat{Q}_{N,M}(x) \ = \ \frac{1}{N_M} \sum_{l=1}^{N_M} \mathbb{1}\{M\sqrt{h_M}\; t_N(\mathcal{Y}^l_M) \leq x\} \ .$$

Obviously, the critical value for our test can be obtained as the $1-\alpha$ quantile of this estimated sampling distribution $\widehat{Q}_{N,M}$ by $\hat{q}_{N,M}(1-\alpha) = \inf\{x : \widehat{Q}_{N,M}(x) \geq \alpha\}$, and the nominal level α test rejects the null hypothesis iff $\Upsilon_N > \hat{q}_{N,M}(1-\alpha)$. From Theorem 3.1 and Politis, Romano and Wolf (2001) we get

Corollary 5.4 *Under the assumptions of Theorem 3.1, $M/N \to 0$, and $M \to \infty$ as $N \to \infty$ it holds that*
(i) Under H_0, $Prob_{P_0}\{\Upsilon_N > \hat{q}_{N,M}(1-\alpha)\} \longrightarrow \alpha$ as $N \to \infty$
(ii) Under H_1, $Prob_P\{\Upsilon_N > \hat{q}_{N,M}(1-\alpha)\} \longrightarrow 1$ as $N \to \infty$.

In the work of Politis, Romano and Wolf (2001) only one sample examples are considered. Consequently, they have $N_M = \begin{pmatrix} N \\ M \end{pmatrix}$ admissible subsamples in the i.i.d. case. In the time series case, the ordering of observations does matter and so it is necessary to do block wise subsampling. This yields (in the one sample case) to $N_M = N - M + 1$ subsamples. For the reasons of implementation and computational expense this is what usually is done in practice in the i.i.d. case, too, see e.g. Delgado, Rodríguez-Póo and Wolf (2001). In a two-sample time series problem, it is only reasonable to consider the models over time intervals of the same length, i.e. $n_1 = n_2$ and you would have $N_M = n_1 - m_1 + 1 = n_2 - m_2 + 1$. If you have independent data but you want to apply blockwise subsampling, you have $N_M = (n_1 - m_1 + 1)(n_2 - m_2 + 1)$ different subsamples. In practice, actually, it is sufficient to draw randomly N_M subsamples out of the admissible ones with N_M chosen arbitrarily by the empirical researcher. This can be seen from Politis, Romano and Wolf (1999), Chapters 2 and 3 (stochastic approximations). For the theory certainly, N_M has to go to infinity, in practice it should be chosen "large". This approach is what we have implemented in Section 6.

A data driven choice of subsample size M, respectively m_1 and m_2 can be found in Delgado, Rodríguez-Póo and Wolf (2001) and its mathematical justification in Politis, Romano and Wolf (2001). We therefore give just the idea and algorithm here. The idea is that in practice, the actual rejection level $\hat{\alpha}$ is a function of M. With some pseudo sequences $\mathcal{Y}_N^{*,k}$, $k = 1, \ldots, K$ for $\mathcal{Y}_N$ (in the i.i.d. case take e.g. the naive bootstrap to generate them), calculate $\Upsilon_N^{*,k}$ and compute $\hat{\alpha}(M) = \frac{1}{K}\sum_{k=1}^{K} \mathbb{1}\{\Upsilon_N^{*,k}$ rejects $\}$. You will take that subsample size M that minimizes $|\hat{\alpha}(M) - \alpha|$ for the wanted level α. The algorithm is then

1. Fix a grid of reasonable subsample sizes M and the wanted rejection level α.

2. Generate pseudo sequences $\mathcal{Y}_N^{*,k}$, $k = 1, \ldots, K$, i.i.d. (according to the sample distribution)

3. For each $\mathcal{Y}_N^{*,k}$ construct its test $\Upsilon_N^{*,k}$ and calculate $\hat{\alpha}(M) = \frac{1}{K}\sum_{k=1}^{K} \mathbb{1}\{\Upsilon_N^{*,k}$ rejects $H_0\}$

4. Find susample size M that minimizes $|\hat{\alpha}(M) - \alpha|$

Remark 5.5 This method does not maximize power but guarantees that we hold the first error level without getting too conservative. Notice that Politis, Romano and Wolf (2001) even doubt that a power maximizing method can be found. In any case, this discussion is clearly beyond the scope of this paper.

When we apply now this subsampling procedure to our test, note that we have $t_N = T_N - E_{H_0}[T_N]$, where $E_{H_0}[T_N] = O_p(\frac{1}{Nh_N})$. There are two possibilities to proceed: Either estimate $E_{H_0}[T_N]$, $\hat{t}_N$ by

$$\widehat{E_{H_0}[T_N]} = \frac{1}{Nh_N} \int K^2(u)du \left(\frac{1}{n_1} \sum_{j=1}^{n_1} Y_{1j}^2 \frac{\hat{s}_1^2(W_{1j})v(X_{1j})}{\kappa_1 \hat{p}_1^2(X_{1j}, W_{1j})} + \frac{1}{n_2} \sum_{j=1}^{n_2} Y_{2j}^2 \frac{\hat{s}_2^2(W_{2j})v(X_{2j})}{\kappa_2 \hat{p}_2^2(X_{2j}, W_{2j})} \right)$$

$$\hat{T}_N = T_N - \widehat{E_{H_0}[T_N]} \ . \tag{21}$$

This obviously is consistent with all necessary assumptions for the subsampling hypothesis testing.

Note that $N\sqrt{h_N}\ E_{H_0}[T_N]$ alone does unfortunately not go to zero. On the other hand, a careful study of the results of Politis, Romano and Wolf (1999, 2001) give some hope that also in finite samples this *bias term* $E_{H_0}[T_N]$ is replicated reasonable well by the subsampling procedure and could therefore be neglected. Then we compare directly $N\sqrt{h_N}T_N$ with the K subsample statistics $M\sqrt{h_M}T_M$. It sometimes can be rather cumbersome to estimate the bias term, so in practice it would be nice if the test would work without this "bias correction". Therefore our simulkation study in Section 6 mainly focuses on this approach. However, we come back to this discussion in that section.

Finally, note that Politis, Romano and Wolf (1999, 2001) derived the results we need to apply subsampling on our testing problem also for dependent (e.g. time series) data. They only had to assume stationarity and α-mixing. As the stronger β- mixing was already assumed by Camlong-Viot (2000) to derive its results on the IMIE, subsampling would thus not require additional assumptions here.

6 The Procedure in Practice

In this section we mainly focus on two points: feasibility and small samples performance of the proposed methods including the choice of smoothing parameter and illustration of the break down of the bootstrap. We will concentrate here only on the original test statistic T_N with a particular weight function $v(\cdot)$. On the one hand, the weight function should be chosen along the objectives of the empirical researcher, on the other hand it is well known that in non- and semiparametric testing for small samples trimming is recommended. From a mathematical point of view one even could calculate from the asymptotics stated in Theorem 3.1 a weight function that minimizes variances and maximizes $E_{H_1}[T_N]$. This would maximize power, however, only asymptotically. Further, such an expression consists mainly of unknown expressions and it can be doubted whether an estimated optimal weight $\widehat{v}_{opt}(\cdot)$ is really optimal in practice. Due to the discussion in Section 3, and to the conclusions of Dette and von Lieres und Wilkau (2001) we have decided for

$$v(t) \;=\; \frac{\sqrt{\hat{r}_1(t)\cdot\hat{r}_2(t)}}{\int\sqrt{\hat{r}_1(u)\cdot\hat{r}_2(u)}du}, \tag{22}$$

where $\hat{r}_i(t)$, $i=1,2$, are kernel estimates of the marginal densities of the X_i. In our simulation study, the chosen weight works pretty well. We tried also other weight functions, e.g. as in (22) but without taking the root so that at least under the null, the asymptotics are equivalent to those of $S_N^{(2)}$. We could not find major differences and will therefore only present the results obtained for weight function (22).

Further, we did simulations comparing the results based on estimators $\hat{p}_i$, $\hat{s}_i$ from (9), (10) with those based on $\widetilde{p}_i$, $\widetilde{s}_i$ from (11), (12). The power seems to be slightly better when using $\widetilde{p}_i$, $\widetilde{s}_i$ but this can be just by chance. In any case, the results given in the tables and graphs refer to the use of these estimators.

The following models are considered: Variables (X_i', W_i') are drawn from a bivariate standard normal distribution with correlation $\mathrm{Cov}(X_i', W_i') = -0.1$. Then, to facilitate the bandwidths selection, we set $X_i = \arctan(X_i')2.5\pi^{-1}$, $W_i = \arctan(W_i')2.5\pi^{-1}$ so that for $i=1,2$ we get $(X_i, W_i) \in [-1.1, 1.1]^{(d+1)}$. For $\varepsilon \sim N(0,1)$ and $\sigma = 0.5$, respectively 0.1, we generated

$$Y_{1j} \;=\; 0.5\sin(\pi X_{1j}) + 0.5\sin(\pi W_{1j}) + \sigma\varepsilon_{1j} \tag{23}$$

$$\text{and} \quad Y_{2j} \;=\; 0.5\sin(\{1.0+\delta\}\pi X_{2j}) + 0.5\sin(\pi W_{2j}) + \sigma\varepsilon_{2j}\ . \tag{24}$$

Obviously, with δ increasing we can see the power performance of our test. The marginal density of X_1, functional f_1 and its (IMIE) estimate are shown in Figure 1.

A main problem in the practice of non- and semiparametric testing is the choice of smoothing parameters. Recently there can be found an increasing amount of literature on "adaptive" testing (for kernel methods see e.g. Horowitz and Spokoiny, 2001). But they commonly consider problems of parametric null hypothesis so that they only need to choose a smoothing parameter for the alternative. They choose the smoothing parameter in a way that it maximizes the test statistic value. Computationally this becomes rather expensive when a multivariate regression is considered. Further, to our knowledge only Härdle, Sperlich and Spokoiny (2001) consider nonparametric nuisance directions as we have them given by the functions $g_1(\cdot)$, $g_2(\cdot)$. All in all, the existing procedures to select the smoothing parameters adaptively can not straight forwardly be applied to our context. Moreover, we need several bandwidths: *a)* for the test statistic h_1, b_1 for the estimates

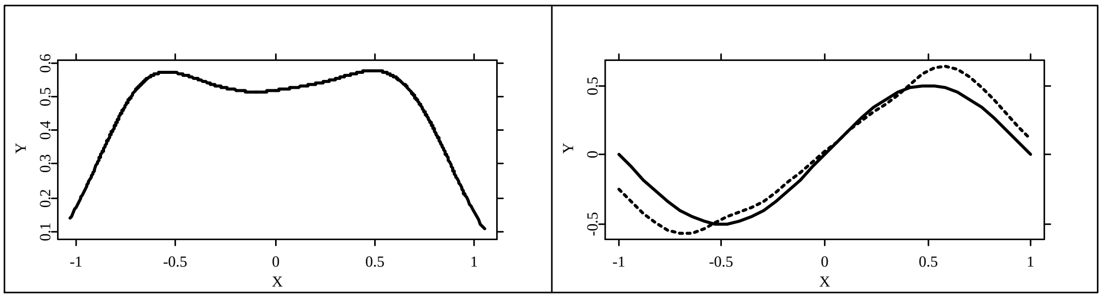

Figure 1: On the left side, marginal density of X_1. On the right side, function $f_1(\cdot)$ of model (23) (solid) with an estimate (dashed) for $n_1 = 200$, using bandwidths $h_1 = 0.4$, $b_1 = 3h_1$.

of the first sample, respectively h_2 and b_2 for the second sample. When doing bootstrap we need: *b)* h_f, b_f for the estimation of $f(\cdot)$ from the pooled data, and some bandwidths to estimate the functions $g_1(\cdot)$, $g_2(\cdot)$. When doing subsampling we need instead of *b)*: *c)* h_{ib}, b_{ib} for the estimation in the subsamples. They should have the sizes corresponding to the subsample size, i.e. $h_{ib} = h_i \cdot (n_i/(m_i))^{1/4}$ (compare [A9'] needed when using $\widetilde{p}_i$, $\widetilde{s}_i$) and so on.

We tried out several methods, e.g. for the bootstrap following the recommendations of Härdle and Marron (1991) for the estimation of the null model to generate the bootstrap samples. Among all data-driven methods the following turned out to work best: calculate cross validation (cv) bandwidths to estimate f_1, f_2, g_1, g_2. In the case of models (3), (4) this can be done by leave-one-out cross validation. The cross validation bandwidth for $\hat{f}_i$, $i = 1, 2$, should be of order $n^{-1/5}$. For testing, when $d = 1$, we can take bandwidths of order $n^{-1/4}$ (condition [A9']), so we set $h_i = h_{i,cv} n_1^{1/5-1/4}$, where $h_{i,cv}$ stands for the optimal cross validation bandwidths. The same way we proceed for all the other bandwidths of *a)*. For the estimators to generate the bootstrap samples we take

$$h_f = \{h_{1,cv} n_1^{1/5} + h_{2,cv} n_2^{1/5}\} / \{2\ (n_1 + n_2)^{1/5}\}$$

and analogue for b_f.

We are certainly aware that not for all models discussed in Section 4 cross validation or similar data driven bandwidths selection rules are available. But to develop them here is clearly beyond the scope of this paper. For the kernels K, L we always take the second order quartic kernel according to conditions [A4], [A6].

6.1 Applying Wild Bootstrap

In this part of the section we resample with the bootstrap procedure presented in Section 5 drawing always 500 bootstrap samples to get the critical value. We restrict here to consider the sample sizes $n_1 = 200$, $n_2 = 250$ and $n_1 = 400$, $n_2 = 500$.

From Theorem 5.1 it can be seen that for small σ^2 the error the bootstrap commits will be strong whereas for large σ^2 one might not even see it in simulations. In Table 1 we have given the error of the first kind after 500 replications. Clearly, for $\sigma = 0.5$ we can not detect any problem with the bootstrap, whereas for $\sigma = 0.1$ it fails to work. When we increase the sample sizes, this impression does hardly change, only that for $\sigma = 0.1$ we reject even much more.

			wanted first error level (α)			
n_1,n_2	σ	p-value	0.01	0.05	0.10	0.15
200,250	0.5	0.481	0.010	0.056	0.118	0.182
	0.1	0.330	0.052	0.168	0.274	0.354
400,500	0.5	0.446	0.012	0.048	0.126	0.20
	0.1	0.301	0.036	0.154	0.294	0.396

Table 1: First error probabilities after 500 simulation runs when applying wild bootstrap for different models.

6.2 Applying Subsampling

The bandwidth selection could be handled as before. However, it is no longer a crucial point here as we do not need to estimate the null hypothesis to generate new samples. In particular, it will not depend on the quality of estimation whether we hold the first error probability level or not. Instead, this now depends on the choice of the subsample size. Therefore we fix the bandwidth over all simulation runs to the following: $h_1 = 0.4$, $b_1 = 1.2$ when $n_1 = 200$, $n_2 = 250$, all the remaining bandwidth the same but adjusting for the particular (sub-)sample size, i.e. $h_2 = 0.4 \cdot (n_1/n_2)^{1/4}$. In order to underpin that this procedure seems to work quite well for any sample size, we have added results for sample sizes $n_1 = 100$, $n_2 = 150$ with $h_1 = 0.45$, $b_1 = 1.35$ and $n_1 = 400$, $n_2 = 500$ with $h_1 = 0.35$, $b_1 = 1.4$. These bandwidths come close to the average value we got for the bandwidths determined with the aid of cross validation and adjusted for the convergence rates required by [A9'].

As discussed in Section 5 a question to clarify is whether it makes some qualitative difference to apply the subsampling with or without the bias correction (21) and if yes of what type. In our simulation comparison we found that both procedures work very well but could not find one to be definitely superior to the other. Therefore we limit the future discussion on the method without bias correction as this, at least in our simulations, performed always somewhat better. To facilitate the simulation we further restricted to continuous block wise subsampling, see Section 5.

In Section 5 it has been introduced how to determine the subsample sizes m_i that guarantee the correct first error level. So we first do a simulation study to find block sizes for which the test will hold the first error level. In a second step we will focus on whether determining the block size as described in Section 5 does work in our context. We investigate this only for $n_1 = 200$, $n_2 = 250$, $N_M = 400$ subsamples, $K = 100$ naive bootstrap samples and 100 simulation runs. This has been done for the case when the null hypothesis is true ($\delta = 0.0$), and if it is completely false ($\delta = 1.0$), see model (24). We have summarized some results of the first step in Table 2 where we have restricted on the case $\sigma = 0.5$. It can be seen clearly, that our procedure suggests, depending on the chosen first error level α, to take block sizes between $m_i \approx n_i/2$ and $m_i \approx n_i/3$. Whether this suggestion is a good one we will see in the next step.

		wanted first error level (α)			
δ	m_i	0.01	0.05	0.10	0.15
0.0	$[n_i/2]$	0.039	0.069	0.101	0.138
	$[n_i/2.5]$	0.024	0.047	0.078	0.113
	$[n_i/3]$	0.018	0.041	0.070	0.104
1.0	$[n_i/2]$	0.043	0.077	0.112	0.143
	$[n_i/2.5]$	0.030	0.056	0.088	0.121
	$[n_i/3]$	0.023	0.046	0.078	0.110

Table 2: For $n_1 = 200$, $n_2 = 250$, the estimated first error levels $\hat{\alpha}(m_i)$ as a function of block size m_i.

We continue our simulation study based on the above recommended block sizes m_i. In Tables 3 and 4 are given the first error levels (i.e. the rejection levels for data generated from the null hypothesis model) when $n_1 = 100$, $n_2 = 125$, for $n_1 = 200$, $n_2 = 250$, and for $n_1 = 400$, $n_2 = 500$, always for models with noise level $\sigma = 0.5$ and for those with $\sigma = 0.1$. Depending on the sample size we draw (from the smallest to the biggest n_i, $i = 1, 2$) $N_M = 300, 400$, respectively 500 subsamples, compare Section 5. The rejection levels given in Table 3 (when $\sigma = 0.5$) and Table 4 (when $\sigma = 0.1$) are always calculated from 500 simulation runs. It can be seen clearly that the block size search as well as the subsampling test seem to work well even for extremely small sample sizes, e.g. for only $m_1 = [100/3] = 33$.

Finally, we have studied the power performance of the subsampling test when δ from model (24) runs from zero to one. In Figure 2 are given the power functions for the different sample sizes when the noise level is $\sigma = 0.5$ and the rejection level $\alpha = 0.05, 0.1$ respectively. In Figure 3 is given the analoge when $\sigma = 0.1$. The shown power functions are calculated on 100 simulation runs for all $\delta \in [0, 1]$. We always used $m_i = n_i/3$, $i = 1, 2$, except for $n = 100$, $\sigma = 0.1$ where we set $m_i = n_i/2$, $i = 1, 2$. The figures illustrate that

m_i	n_1	p-value	wanted first error level (α) 0.01	0.05	0.10	0.15
$[n_i/2]$	100	0.540	0.020	0.042	0.080	0.124
	200	0.518	0.030	0.048	0.078	0.120
	400	0.502	0.036	0.062	0.098	0.156
$[n_i/2.5]$	100	0.528	0.008	0.030	0.070	0.100
	200	0.522	0.016	0.022	0.044	0.090
	400	0.514	0.024	0.044	0.078	0.128
$[n_i/3]$	100	0.525	0.006	0.024	0.052	0.084
	200	0.547	0.006	0.018	0.050	0.076
	400	0.521	0.014	0.042	0.062	0.104

Table 3: First error probabilities when data are generated with $\sigma = 0.5$, using subsampling. Numbers are calculated from 500 simulation runs.

m_i	n_1	p-value	wanted first error level (α) 0.01	0.05	0.10	0.15
$[n_i/2]$	100	0.534	0.032	0.050	0.086	0.126
	200	0.544	0.028	0.056	0.096	0.114
	400	0.548	0.030	0.058	0.084	0.122
$[n_i/2.5]$	100	0.530	0.024	0.044	0.074	0.110
	200	0.547	0.020	0.038	0.066	0.102
	400	0.555	0.020	0.044	0.066	0.108
$[n_i/3]$	100	0.524	0.012	0.030	0.062	0.106
	200	0.557	0.010	0.034	0.064	0.086
	400	0.560	0.016	0.030	0.060	0.084

Table 4: First error probabilities when data are generated with $\sigma = 0.1$, using subsampling. Numbers are calculated from 500 simulation runs.

obviously the proposed procedure works well even for small samples sizes ($n_1 = 100$), strong disturbance ($\sigma = 0.5$), and functions difficult to estimate, i.e. $f_i(t)$ and nuisance terms $g_i(t)$ are sine functions equal to $\sin(\pi t)$ on $t \in [-1.1, 1.1]$. For small disturbance ($\sigma = 0.1$) or more reasonable sample sizes ($n_1 = 400$) the power improves a lot.

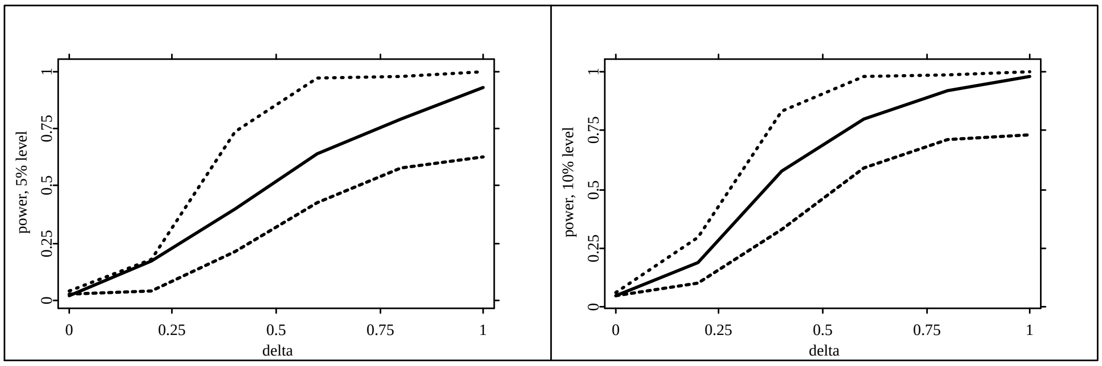

Figure 2: Power performance when data are generated with $\sigma = 0.5$ for rejection level $\alpha = 0.05$ on the left, $\alpha = 0.10$ on the right. Power function for sample sizes $n_1 = 100$, $n_2 = 125$ is dashed, for $n_1 = 200$, $n_2 = 250$ solid, and dotted for $n_1 = 400$, $n_2 = 500$.

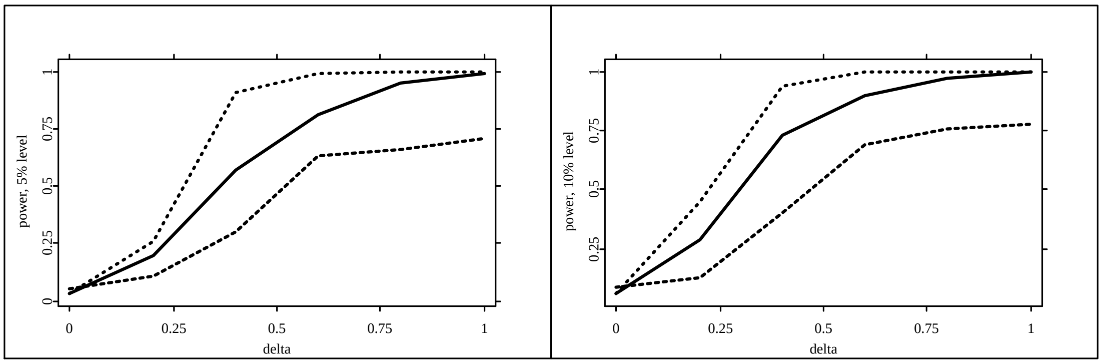

Figure 3: Power performance when data are generated with $\sigma = 0.1$ for rejection level $\alpha = 0.05$ on the left, $\alpha = 0.10$ on the right. Power function for sample sizes $n_1 = 100$, $n_2 = 125$ is dashed, for $n_1 = 200$, $n_2 = 250$ solid, and dotted for $n_1 = 400$, $n_2 = 500$.

7 Appendix

7.1 Proof of Theorem 3.1

The proof is similar to the proof of Theorem 2.1 of Dette and Neumeyer (2001) (where testing the equality of the whole regression functions F_1, F_2 was considered) with certain discrepancies arising from the different model considered in this paper.

For the ease of presentation we restrict the proof to the case $d = 1$. The general case can be treated exactly the same way. We start with a decomposition of the test statistic,

$$T_N = \int (\hat{f}_1(x) - \hat{f}_2(x))^2 v(x)\, dx = T_N^{(1)} + T_N^{(2)} + 2T_N^{(3)} \tag{25}$$

with the following definitions,

$$T_N^{(1)} = \int (\hat{f}_1(x) - f_1(x) - \hat{f}_2(x) + f_2(x))^2 v(x)\, dx \tag{26}$$

$$T_N^{(2)} = \int (f_1(x) - f_2(x))^2 v(x)\, dx \tag{27}$$

$$T_N^{(3)} = \int (f_1(x) - f_2(x))(\hat{f}_1(x) - f_1(x) - \hat{f}_2(x) + f_2(x)) v(x)\, dx. \tag{28}$$

Note that $T_N^{(2)}$ is deterministic, and under $H_0 : f_1 = f_2$ we have $T_N = T_N^{(1)}$.

7.1.1 Expectations

We will make use of the following

Lemma 7.1 *Under the null hypothesis H_0 we have for the expectation of the test statistic*

$$E[T_N] = E[T_N^{(1)}] = \frac{1}{Nh} \int K^2(u)\, du \int H(z,w) v(z)\, d(z,w) + o(\frac{1}{N\sqrt{h}}),$$

where $H(z,w)$ is defined in Theorem 3.1.

Proof. By using the definition (7) of the IMIEs $\hat{f}_1$ and $\hat{f}_2$ we derive a decomposition of $T_N^{(1)}$ as follows.

$$\begin{aligned} T_N^{(1)} &= \int \Big(\frac{1}{n_1} \sum_{j=1}^{n_1} \gamma_{1j}(x) Y_{1j} - \hat{c}_1 - f_1(x) - \frac{1}{n_2} \sum_{j=1}^{n_2} \gamma_{2j}(x) Y_{2j} + \hat{c}_2 + f_2(x) \Big)^2 v(x)\, dx \\ &= T_N^{(1,1)} - 2T_N^{(1,2)} + T_N^{(2,2)} \end{aligned} \tag{29}$$

where $T_N^{(i,\ell)}$ denote the following quadratic forms,

$$T_N^{(i,\ell)} = \frac{1}{n_i n_\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{n_\ell} \int \Big((\gamma_{ij}(x) - 1) Y_{ij} - f_i(x) \Big) \Big((\gamma_{\ell k}(x) - 1) Y_{\ell k} - f_\ell(x) \Big) v(x)\, dx, \tag{30}$$

where $i, \ell \in \{1,2\}$ and $T_N^{(1,2)} = T_N^{(2,1)}$. We start with a straightforward calculation of the expectation $E[T_N^{(1,1)}]$ and use the definition (8) of the functions $\gamma_{ij}(x)$.

$$\begin{aligned} E[T_N^{(1,1)}] &= \frac{n_1(n_1-1)}{n_1^2} \int E\Big[\Big((\gamma_{11}(x) - 1) Y_{11} - f_1(x) \Big) \Big((\gamma_{12}(x) - 1) Y_{12} - f_1(x) \Big) \Big] v(x)\, dx \\ &\quad + \frac{1}{n_1} \int E\Big[\Big((\gamma_{11}(x) - 1) Y_{11} - f_1(x) \Big)^2 \Big] v(x)\, dx \end{aligned}$$

$$
\begin{aligned}
= \frac{(n_1-1)}{n_1} \int E\Big[\Big(\Big\{ & K_h(X_{11}-x)\Big[\frac{1}{n_1}\sum_{\ell=2}^{n_1} L_b(W_{11}-W_{1\ell}) + \frac{L(0)}{n_1 b}\Big]\frac{1}{\hat{p}_1(X_{11},W_{11})} - 1\Big\} \\
& \times \{c_1 + f_1(X_{11}) + g_1(W_{11}) + \sigma_1(X_{11},W_{11})\varepsilon_{11}\} - f_1(x)\Big) \\
& \times \Big(\Big\{K_h(X_{12}-x)\Big[\frac{1}{n_1}\sum_{\substack{k=1\\k\neq 2}}^{n_1} L_b(W_{12}-W_{1k}) + \frac{L(0)}{n_1 b}\Big]\frac{1}{\hat{p}_1(X_{12},W_{12})} - 1\Big\} \\
& \times \{c_1 + f_1(X_{12}) + g_1(W_{12}) + \sigma_1(X_{12},W_{12})\varepsilon_{12}\} - f_1(x)\Big)\Big] v(x)\, dx \\
+ \frac{1}{n_1}\int E\Big[\Big(\Big\{ & K_h(X_{11}-x)\Big[\frac{1}{n_1}\sum_{\ell=2}^{n_1} L_b(W_{11}-W_{1\ell}) + \frac{L(0)}{n_1 b}\Big]\frac{1}{\hat{p}_1(X_{11},W_{11})} - 1\Big\} \\
& \times \{c_1 + f_1(X_{11}) + g_1(W_{11}) + \sigma_1(X_{11},W_{11})\varepsilon_{11}\} - f_1(x)\Big)^2\Big] v(x)\, dx
\end{aligned}
$$

$$
\begin{aligned}
= (1+O(\frac{1}{N}))\int & \Big(\int\int \Big\{K_h(y-x)\Big[L_b(w-v) + \frac{L(0)}{n_1 b}\Big]\frac{1}{p_1(y,w)} - 1\Big\} p_1(y,w) \\
& (c_1 + f_1(y) + g_1(w)) s_1(v)\, d(y,w)\, dv \;-\; f_1(x)\Big)^2 v(x)\, dx \\
& \Big(1 + O(h^r) + O(b^s) + O(\frac{1}{n_1 h b})\Big) \\
+ (\frac{1}{n_1} + O(\frac{1}{N^2}))\int\int\int\int & \Big(\Big\{K_h(y-x)\Big[L_b(w-v) + \frac{L(0)}{n_1 b}\Big]\frac{1}{p_1(y,w)} - 1\Big\} \\
& (c_1 + f_1(y) + g_1(w)) - f_1(x)\Big) p_1(y,w) \\
& \Big(1 + O(h^r) + O(b^s) + O(\frac{1}{n_1 h b})\Big)\Big(\Big\{K_h(z-x)\Big[L_b(u-v) + \frac{L(0)}{n_1 b}\Big]\frac{1}{p_1(z,u)} - 1\Big\} \\
& (c_1 + f_1(z) + g_1(u)) - f_1(x)\Big) p_1(z,u) s_1(v)\, dv\, d(y,w)\, d(z,u)\, v(x)\, dx
\end{aligned}
$$

$$
\begin{aligned}
+ O(\frac{1}{N^2})\int & \Big(\int\int K_h^2(y-x)\Big[L_b(w-v) + \frac{L(0)}{n_1 b}\Big]^2 \frac{1}{p_1(y,w)} \\
& \Big(1 + O(h^r) + O(b^s) + O(\frac{1}{n_1 h b})\Big) \\
& \{(c_1 + f_1(y) + g_1(w))^2 + \sigma_1^2(y,w)\}\, d(y,w)\, s_1(v)\, dv + O(1)\Big)\, v(x)\, dx \\
+ (\frac{1}{n_1} + O(\frac{1}{N^2}))\int & \Big(\int\int\int K_h^2(y-x)\Big[L_b(w-v) + \frac{L(0)}{n_1 b}\Big]\Big[L_b(w-u) + \frac{L(0)}{n_1 b}\Big] \\
& \frac{1}{p_1(y,w)}\Big(1 + O(h^r) + O(b^s) + O(\frac{1}{n_1 h b})\Big)\{(c_1 + f_1(y) + g_1(w))^2 + \sigma_1^2(y,w)\} \\
& s_1(v) s_1(u)\, d(y,w)\, dv\, du + O(1)\Big)\, v(x)\, dx.
\end{aligned}
$$

The replacement of the reciprocal of the density estimate $1/\hat{p}_1$ by the true $1/p_1$ can be explained in detail in the manner of the proof of Theorem 1 of Kim, Linton and Hengartner (1999) [note, that the variance of the estimator $\hat{\gamma}_1^{pi}$ in section 2 of the cited paper in principle matches the bias calculated in this proof]. The remainder $\hat{p}_1(y,w) - p_1(y,w)$ is of order $O(h^r) + O(b^s) + O(1/(n_1 hb))$. Further we obtain for the expectation

$$\begin{aligned}
E[T_N^{(1,1)}] \;=\; & (1+O(\frac{1}{N}))\int\Big(c_1 + \int K_h(y-x)f_1(y)\,dy + \int\int L_b(w-v)g_1(w)s_1(v)\,dv\,dw \\
& \qquad + O(\frac{1}{n_1 b}) - c_1 - f_1(x)\Big)^2 v(x)\,dx\Big(1+O(h^r)+O(b^s)+O(\frac{1}{n_1 hb})\Big) \\
& +O(\frac{1}{N}) + O(\frac{1}{N^2 hb}) \\
& + (\frac{1}{n_1} + O(\frac{1}{N^2}))\int\Big(\int K_h^2(y-x)\Big[\int L_b(w-v)s_1(v)\,dv\Big]^2 \\
& \qquad \times\{(c_1+f_1(y)+g_1(w))^2 + \sigma_1^2(y,w)\}\frac{1}{p_1(y,w)}\,d(y,w) + O(\frac{1}{n_1 h^2 b})\Big)\,v(x)\,dx.
\end{aligned}$$

The first summand now can be simplified by a substitution in the integrals and a Taylor expansion of the functions $f_1(y)$ and $s_1(v)$. Using the order of the kernels [A4], [A6], the assumption $\int g_1(w)s_1(w)\,dw = 0$, and the bandwidth conditions [A9] we obtain:

$$\begin{aligned}
(31)\; E[T_N^{(1,1)}] \;=\; & \int\Big(\int K_h(y-x)f_1(y)\,dy - f_1(x)\Big)^2 v(x)\,dx \\
& + O(h^{2r})\Big(1+O(h^r)+O(b^s)+O(\frac{1}{n_1 hb})\Big) \\
& + \Big(O(b^s)+O(\frac{1}{Nb})\Big)^2 + \Big(O(b^s)+O(\frac{1}{Nb})\Big)O(h^r) + O(\frac{1}{N^2h^2b}) \\
& + \frac{1}{Nh}\int K^2(u)\,du \int \frac{(c_1+f_1(z)+g_1(w))^2+\sigma_1^2(z,w)}{\kappa_1 p_1(y,w)} s_1^2(w)v(z)\,d(z,w).
\end{aligned}$$

An analogous calculation shows for the last term in the decomposition (29):

$$\begin{aligned}
(32)\; E[T_N^{(2,2)}] \;=\; & \int\Big(\int K_h(y-x)f_2(y)\,dy - f_2(x)\Big)^2 v(x)\,dx + o(\frac{1}{N\sqrt{h}}) \\
& + \frac{1}{Nh}\int K^2(u)\,du \int \frac{(c_2+f_2(z)+g_2(w))^2+\sigma_2^2(z,w)}{\kappa_2 p_2(y,w)} s_2^2(w)v(z)\,d(z,w)
\end{aligned}$$

and with similar calculations we obtain:

$$\begin{aligned}
(33) \qquad E[T_N^{(1,2)}] \;=\; & \int\Big(\int K_h(y-x)f_1(y)\,dy - f_1(x)\Big) \\
& \times\Big(\int K_h(y-x)f_2(y)\,dy - f_2(x)\Big)v(x)\,dx + o(\frac{1}{N\sqrt{h}}).
\end{aligned}$$

Under the null hypothesis $H_0 : f_1 = f_2$ we then get the desired results from the above extensions (31), (32) and (33) :

$$
\begin{aligned}
E[T_N^{(1)}] &= E[T_N^{(1,1)}] + E[T_N^{(2,2)}] - 2E[T_N^{(1,2)}] \\
&= \frac{1}{Nh} \int K^2(u)\,du \int \frac{(c_1 + f_1(z) + g_1(w))^2 + \sigma_1^2(z,w)}{\kappa_1 p_1(y,w)} s_1^2(w) v(z)\,d(z,w) \\
&\quad + \frac{1}{Nh} \int K^2(u)\,du \int \frac{(c_2 + f_2(z) + g_2(w))^2 + \sigma_2^2(z,w)}{\kappa_2 p_2(y,w)} s_2^2(w) v(z)\,d(z,w) \\
&\quad + o(\frac{1}{N\sqrt{h}}) \\
&= \frac{1}{Nh} \int K^2(u)\,du \int H(z,w) v(z)\,d(z,w) + o(\frac{1}{N\sqrt{h}}).
\end{aligned}
$$

□

Under the alternative $H_1 : f_1 \neq f_2$ we directly obtain from the above calculations (31), (32) and (33) and the fact that $\int K_h(y-x) f_i(y) dy - f_i(x) = O(h^r)$,

$$
E[T_N^{(1)}] = O(h^{2r}) + O(\frac{1}{Nh}) + o(\frac{1}{N\sqrt{h}}).
$$

which is the first statement of the next lemma.

Lemma 7.2 *Under the fixed alternative $H_1 : f_1 \neq f_2$ we have for the expectation of the statistics $T_N^{(1)}$, $T_N^{(2)}$, $T_N^{(3)}$ defined in (26)–(28),*

$$
\begin{aligned}
E[T_N^{(1)}] &= O(h^{2r}) + O(\frac{1}{Nh}) \\
E[T_N^{(2)}] &= T_N^{(2)} = \int (f_1(x) - f_2(x))^2 v(x)\,dx \\
E[T_N^{(3)}] &= O(h^r) + O(b^s) + O(\frac{1}{Nbh}).
\end{aligned}
$$

Proof. The first statement is shown above, and the second is just the definition of $T_N^{(2)}$. The last result is obtained from the following calculation of the expectation $E[\hat{f}_1(x)]$ and the analogous considerations for $E[\hat{f}_2(x)]$. From the definitions (7) und (8) we have

$$
\begin{aligned}
E[\hat{f}_1(x)] &= \\
&= E\Big[\frac{1}{n_1}\sum_{j=1}^{n_1}(\gamma_{1j}(x)-1)Y_{1j}\Big] \;=\; E[(\gamma_{11}(x)-1)(c_1+f_1(X_{11})+g_1(W_{11}))] \\
&= \frac{n_1-1}{n_1}E\Big[\Big\{K_h(X_{11}-x)L_b(W_{11}-W_{12})\frac{1}{\hat{p}_1(X_{11},W_{11})}-1\Big\}(c_1+f_1(X_{11})+g_1(W_{11}))\Big] \\
&\quad + O(\frac{1}{n_1 b}) \\
&= (1-\frac{1}{n_1})\int\int K_h(y-x)L_b(w-v)(c_1+f_1(y)+g_1(w))s_1(v)\,d(y,w)\,dv \\
&\qquad \Big(1+O(h^r)+O(b^s)+O(\frac{1}{Nbh})\Big) \\
&\quad -(1-\frac{1}{n_1})\int(c_1+f_1(y)+g_1(w))p_1(y,w)\,d(y,w)+O(\frac{1}{n_1 b}) \\
&= \int K_h(y-x)f_1(y)\,dy+\int\Big(\int L_b(w-v)s_1(v)\,dv\Big)g_1(w)\,dw+O(h^r)+O(b^s)+O(\frac{1}{n_1 bh}) \\
&= f_1(x)+O(h^r)+O(b^s)+O(\frac{1}{Nbh})
\end{aligned}
$$

where we have applied Taylor expansions of the functions $f_1(y)$ and $s_1(v)$ and have used assumptions [A4] and [A6] about the order of the kernels and the fact that $\int f_1(y)p_1(y,w)\,d(y,w)=\int g_1(w)p_1(y,w)\,d(y,w)=0$. Substituting this in the definition of $T_N^{(3)}$ gives the third statement of this lemma. □

7.1.2 Variances

Lemma 7.3 *Under the null hypothesis $H_0 : f_1 = f_2$ we have for the variance of the test statistic*

$$\mathrm{Var}(T_N)=\mathrm{Var}(T_N^{(1)}) \;=\; \frac{1}{N^2 h}\beta^2+o(\frac{1}{N^2 h})$$

where β^2 is defined in Theorem 3.1.

Proof. We use the decomposition (29) of $T_N^{(1)}$ and first consider the variance of $T_N^{(1,1)}$ defined in (30). The calculation is straightforward, but rather technical and cumbersome,

$$
\begin{aligned}
&\mathrm{Var}(T_N^{(1,1)}) \;= \\
&= \frac{1}{n_1^4}\sum_{j=1}^{n_1}\sum_{k=1}^{n_1}\sum_{j'=1}^{n_1}\sum_{k'=1}^{n_1}\Big\{E\Big[\int\Big((\gamma_{1j}(x)-1)Y_{1j}-f_1(x)\Big)\Big((\gamma_{1k}(x)-1)Y_{1k}-f_1(x)\Big)v(x)\,dx \\
&\qquad\qquad \times\int\Big((\gamma_{1j'}(y)-1)Y_{1j'}-f_1(y)\Big)\Big((\gamma_{1k'}(y)-1)Y_{1k'}-f_1(y)\Big)v(y)\,dy\Big] \\
&\qquad - E\Big[\int\Big((\gamma_{1j}(x)-1)Y_{1j}-f_1(x)\Big)\Big((\gamma_{1k}(x)-1)Y_{1k}-f_1(x)\Big)v(x)\,dx\Big] \\
&\qquad \times E\Big[\int\Big((\gamma_{1j'}(y)-1)Y_{1j'}-f_1(y)\Big)\Big((\gamma_{1k'}(y)-1)Y_{1k'}-f_1(y)\Big)v(y)\,dy\Big]\Big\}
\end{aligned}
$$

$$
\begin{aligned}
&= \Big(\frac{2}{n_1}+O(\frac{1}{N^2})\Big)\Bigg\{\int\int\Big\{E\Big[\Big((\gamma_{11}(x)-1)Y_{11}-f_1(x)\Big)\Big((\gamma_{11}(y)-1)Y_{11}-f_1(y)\Big)\Big] \\
&\qquad \times E\Big[(\gamma_{11}(x)-1)Y_{11}-f_1(x)\Big]E\Big[(\gamma_{11}(y)-1)Y_{11}-f_1(y)\Big]v(x)v(y)\,dx\,dy \\
&\qquad - \Big(\int\Big(E\Big[(\gamma_{11}(x)-1)Y_{11}-f_1(x)\Big]\Big)^2 v(x)\,dx\Big)^2\Bigg\} \\
&\quad + \Big(\frac{2}{n_1^2}+O(\frac{1}{N^3})\Big)\Bigg\{\int\int\Big(E\Big[\Big((\gamma_{11}(x)-1)Y_{11}-f_1(x)\Big)\Big((\gamma_{11}(y)-1)Y_{11}-f_1(y)\Big)\Big]\Big)^2 \\
&\qquad v(x)v(y)\,dx\,dy \quad - \Big(\int\Big(E\Big[(\gamma_{11}(x)-1)Y_{11}-f_1(x)\Big]\Big)^2 v(x)\,dx\Big)^2\Bigg\} \\
&\quad + \Big(\frac{4}{n_1^2}+O(\frac{1}{N^3})\Big)\Bigg\{\int\int E\Big[\Big((\gamma_{11}(x)-1)Y_{11}-f_1(x)\Big)^2\Big((\gamma_{11}(y)-1)Y_{11}-f_1(y)\Big)\Big] \\
&\qquad \times E\Big[\Big((\gamma_{11}(y)-1)Y_{11}-f_1(y)\Big)\Big]v(x)v(y)\,dx\,dy \\
&\qquad - \Big(\int\Big(E\Big[(\gamma_{11}(x)-1)Y_{11}-f_1(x)\Big]\Big)^2 v(x)\,dx\Big) \\
&\qquad \times\Big(\int E\Big[\Big((\gamma_{11}(x)-1)Y_{11}-f_1(x)\Big)^2\Big]v(x)\,dx\Big)\Bigg\} \\
&\quad + \frac{4}{n_1^3}\Bigg\{\int\int E\Big[\Big((\gamma_{11}(x)-1)Y_{11}-f_1(x)\Big)^2\Big((\gamma_{11}(y)-1)Y_{11}-f_1(y)\Big)^2\Big]v(x)v(y)\,dx\,dy \\
&\qquad - \Big(\int E\Big[\Big((\gamma_{11}(x)-1)Y_{11}-f_1(x)\Big)^2\Big]v(x)\,dx\Big)^2\Bigg\}.
\end{aligned}
$$

We have $E[(\gamma_{11}(x)-1)Y_{11}-f_1(x)] = O(h^r)+O(b^s)+O(1/(Nhb))$ and therefore obtain:

$$
\begin{aligned}
&\mathrm{Var}(T_N^{(1,1)}) \;=\\
&= O(\frac{1}{N})\Big(O(h^r)+O(b^s)+O(\frac{1}{Nbh})\Big)^2\\
&\quad+\Big(\frac{2}{n_1^2}+O(\frac{1}{N^3})\Big)\\
&\quad\times\Bigg\{\int\int\Big(E\Big[\int\Big(\Big\{K_h(z-x)\frac{s_1(w)}{p_1(z,w)}-1\Big\}(c_1+f_1(z)+g_1(w)+\sigma_1(z,w)\varepsilon_{11})-f_1(x)\Big)\\
&\quad\times\Big(\Big\{K_h(z-y)\frac{s_1(w)}{p_1(z,w)}-1\Big\}(c_1+f_1(z)+g_1(w)+\sigma_1(z,w)\varepsilon_{11})-f_1(y)\Big)p_1(z,w)\,d(z,w)\Big]\Big)^2\\
&\quad\times v(x)v(y)\,dx\,dy-\Big(O(h^r)+O(b^s)+O(\frac{1}{Nbh})\Big)^4\Bigg\}\\
&\quad+O(\frac{1}{N^2})\Big\{O(\frac{1}{h})\Big(O(h^r)+O(b^s)+O(\frac{1}{Nb})\Big)-O(\frac{1}{h})\Big(O(h^r)+O(b^s)+O(\frac{1}{Nbh})\Big)^2\Big\}\\
&\quad+O(\frac{1}{N^3})O(\frac{1}{h^2})\\
&= \frac{2}{n_1^2}\int\int\Bigg\{\int K_h(z-x)K_h(z-y)\frac{s_1^2(w)}{p_1(z,w)}\{(c_1+f_1(z)+g_1(w))^2+\sigma_1^2(z,w)\}\,d(z,w)\\
&\qquad-\int K_h(z-y)s_1(w)\{(c_1+f_1(z)+g_1(w))^2+\sigma_1^2(z,w)\}\,d(z,w)\\
&\qquad-\int K_h(z-x)s_1(w)\{(c_1+f_1(z)+g_1(w))^2+\sigma_1^2(z,w)\}\,d(z,w)\\
&\qquad+\int\{(c_1+f_1(z)+g_1(w))^2+\sigma_1^2(z,w)\}p_1(z,w)\,d(z,w)\\
&\qquad-f_1(x)\Big(\int K_h(z-y)s_1(w)(c_1+f_1(z)+g_1(w))\,d(z,w)\\
&\qquad\qquad-\int(c_1+f_1(z)+g_1(w))p_1(z,w)\,d(z,w)\Big)\\
&\qquad-f_1(y)\Big(\int K_h(z-x)s_1(w)(c_1+f_1(z)+g_1(w))\,d(z,w)\\
&\qquad\qquad-\int(c_1+f_1(z)+g_1(w))p_1(z,w)\,d(z,w)\Big)+f_1(x)f_1(y)\Bigg\}^2 v(x)v(y)\,dx\,dy\\
&\quad+o(\frac{1}{N^2h})\\
&= \frac{2}{n_1^2}\int\int\int\int K_h(z-x)K_h(z-y)K_h(z'-x)K_h(z'-y)\frac{s_1^2(w)s_1^2(w')}{p_1(z,w)p_1(z',w')}\\
&\qquad\{(c_1+f_1(z)+g_1(w))^2+\sigma_1^2(z,w)\}\{(c_1+f_1(z')+g_1(w'))^2+\sigma_1^2(z',w')\}\\
&\qquad d(z,w)\,d(z',w')v(x)v(y)\,dx\,dy\\
&\quad+O(\frac{1}{N^2})+o(\frac{1}{N^2h})
\end{aligned}
$$

$$= \frac{2}{n_1^2 h} \int\!\!\int\!\!\int\!\!\int\!\!\int\!\!\int K(s)K(t)K(u)K(u-s+t) \frac{s_1^2(w)s_1^2(w')}{p_1(z,w)p_1(z+h(u-s),w')}$$
$$\{(c_1+f_1(z)+g_1(w))^2+\sigma_1^2(z,w)\}\{(c_1+f_1(z+h(u-s))+g_1(w'))^2$$
$$+\sigma_1^2(z+h(u-s),w')\}v(z-hs)v(z-ht)\,dz\,dw\,dw'\,ds\,dt\,du$$
$$+ o(\frac{1}{N^2h})$$

$$= \frac{2}{N^2h}\int (K*K)^2(u)\,du \int \Big(\int \frac{s_1^2(w)}{\kappa_1 p_1(z,w)}\{(c_1+f_1(z)+g_1(w))^2+\sigma_1^2(z,w)\}\,dw\Big)^2 v^2(z)\,dz$$
$$+ o(\frac{1}{N^2h})$$

where we have used the bandwidth conditions [A9] and similar transformations as in the proofs of the Lemmata 7.1 und 7.2. The analogous calculation for $T_N^{(2,2)}$ defined in (30) yields

$$\mathrm{Var}(T_N^{(2,2)})$$

$$= \frac{2}{N^2h}\int (K*K)^2(u)\,du \int \Big(\int \frac{s_2^2(w)}{\kappa_2 p_2(z,w)}\{(c_2+f_2(z)+g_2(w))^2+\sigma_2^2(z,w)\}\,dw\Big)^2 v^2(z)\,dz$$
$$+ o(\frac{1}{N^2h})$$

and similarly we obtain

$$\mathrm{Var}(T_N^{(1,2)})$$

$$= \frac{1}{n_1n_2}\int\!\!\int E\Big[\Big((\gamma_{11}(x)-1)Y_{11}-f_1(x)\Big)\Big((\gamma_{11}(y)-1)Y_{11}-f_1(y)\Big)\Big]$$
$$E\Big[\Big((\gamma_{21}(x)-1)Y_{21}-f_2(x)\Big)\Big((\gamma_{21}(y)-1)Y_{21}-f_2(y)\Big)\Big]v(x)v(y)\,dx\,dy + o(\frac{1}{N^2h})$$

$$= \frac{1}{N^2h}\int (K*K)^2(u)\,du \int \Big(\int \frac{s_1^2(w)}{\kappa_1 p_1(z,w)}\{(c_1+f_1(z)+g_1(w))^2+\sigma_1^2(z,w)\}\,dw\Big)$$
$$\times\Big(\int \frac{s_2^2(w)}{\kappa_2 p_2(z,w)}\{(c_2+f_2(z)+g_2(w))^2+\sigma_2^2(z,w)\}\,dw\Big)v^2(z)\,dz + o(\frac{1}{N^2h})$$

and the covariances

$$\mathrm{Cov}(T_N^{(1,1)},T_N^{(1,2)}) = o(\frac{1}{N^2h}),\ \mathrm{Cov}(T_N^{(2,2)},T_N^{(1,2)}) = o(\frac{1}{N^2h}),\ \mathrm{Cov}(T_N^{(1,1)},T_N^{(2,2)}) = 0.$$

Now from the above calculations and the decomposition (29) follows the result of Lemma 7.3,

$$
\begin{aligned}
\mathrm{Var}(T_N^{(1)}) &= \mathrm{Var}(T_N^{(1,1)}) + 4\mathrm{Var}(T_N^{(1,2)}) + \mathrm{Var}(T_N^{(2,2)}) + o(\frac{1}{N^2h}) \\
&= \frac{2}{N^2h}\int (K*K)^2(u)\,du \int \Big(\int \frac{s_1^2(w)}{\kappa_1 p_1(z,w)}\{(c_1+f_1(z)+g_1(w))^2+\sigma_1^2(z,w)\}\,dw \\
&\qquad + \int \frac{s_2^2(w)}{\kappa_2 p_2(z,w)}\{(c_2+f_2(z)+g_2(w))^2+\sigma_2^2(z,w)\}\,dw\Big)^2 v^2(z)\,dz \\
&= \frac{1}{N^2h}\beta^2 + o(\frac{1}{N^2h}).
\end{aligned}
$$

□

From the proof of Lemma 7.3 follows the first statement of the next Lemma.

Lemma 7.4 *Under the fixed alternative $H_1 : f_1 \neq f_2$ we have for the variances and covariances of the statistics defined in (26) and (28)*

$$
\begin{aligned}
\mathrm{Var}(T_N^{(1)}) &= O(\frac{1}{N^2h}) \;=\; o(\frac{1}{N}) \\
\mathrm{Cov}(T_N^{(1)}, T_N^{(3)}) &= o(\frac{1}{N}) \\
\mathrm{Var}(T_N^{(3)}) &= \frac{1}{N}\gamma^2 + o(\frac{1}{N})
\end{aligned}
$$

where γ^2 is defined in Theorem 3.1.

Proof. The calculations for $\mathrm{Cov}(T_N^{(1)}, T_N^{(3)})$ and $\mathrm{Var}(T_N^{(3)})$ are similar to the proof of Lemma 7.3 and therefore only stated briefly. On account of definition (28) we decompose $T_N^{(3)}$ into a difference $T_N^{(3,1)} - T_N^{(3,2)}$ where $(i = 1, 2)$

$$
T_N^{(3,i)} = \frac{1}{n_i}\sum_{j=1}^{n_i}\int (f_1(x)-f_2(x))\Big((\gamma_{ij}(x)-1)Y_{ij} - f_i(x)\Big)v(x)\,dx.
$$

For the proof of $\mathrm{Cov}(T_N^{(1)}, T_N^{(3)}) = o(1/N)$ we only show the calculation of $\mathrm{Cov}(T_N^{(1,1)}, T_N^{(3,1)})$ where $T_N^{(1,1)}$ is defined in (29). We obtain

$$
\begin{aligned}
\mathrm{Cov}(T_N^{(1,1)}, T_N^{(3,1)}) &= \\
= \frac{1}{n_1^3}\sum_{j=1}^{n_1}\sum_{k=1}^{n_1}\sum_{\ell=1}^{n_1}\Big\{ & E\Big[\int \Big((\gamma_{1j}(x)-1)Y_{1j} - f_1(x)\Big)\Big((\gamma_{1k}(x)-1)Y_{1k} - f_1(x)\Big)v(x)\,dx \\
& \int (f_1(y)-f_2(y))\Big((\gamma_{1\ell}(y)-1)Y_{1\ell} - f_1(y)\Big)v(y)\,dy\Big] \\
& - E\Big[\int \Big((\gamma_{1j}(x)-1)Y_{1j} - f_1(x)\Big)\Big((\gamma_{1k}(x)-1)Y_{1k} - f_1(x)\Big)v(x)\,dx\Big] \\
& E\Big[\int (f_1(y)-f_2(y))\Big((\gamma_{1\ell}(y)-1)Y_{1\ell} - f_1(y)\Big)v(y)\,dy\Big]\Big\}
\end{aligned}
$$

$$
\begin{aligned}
&= 2\frac{n_1(n_1-1)}{n_1^3}\int\int E\Big[\Big((\gamma_{11}(x)-1)Y_{11}-f_1(x)\Big)\Big((\gamma_{11}(y)-1)Y_{11}-f_1(y)\Big)\Big] \\
&\qquad (f_1(y)-f_2(y))v(y)v(x)E\Big[(\gamma_{11}(x)-1)Y_{11}-f_1(x)\Big]\,dx\,dy \\
&\quad -2\frac{n_1(n_1-1)}{n_1^3}\int\int\Big(E\Big[(\gamma_{11}(x)-1)Y_{11}-f_1(x)\Big]\Big)^2 E\Big[(\gamma_{11}(y)-1)Y_{11}-f_1(y)\Big] \\
&\qquad (f_1(y)-f_2(y))v(y)v(x)\,dx\,dy \quad + O(\frac{h}{n_1^2}) \\
&= o(\frac{1}{N})
\end{aligned}
$$

because $E\Big[(\gamma_{11}(x)-1)Y_{11}-f_1(x)\Big] = O(h^r)+O(b^s)+O(1/(Nbh))$.

For deriving the variance of $T_N^{(3)}$ defined in (28) we write

$$
\begin{aligned}
T_N^{(3)} &= \frac{1}{n_1}\sum_{j=1}^{n_1}\int (f_1(x)-f_2(x))(\gamma_{1j}(x)-1)v(x)\,dx\,Y_{1j} \\
&\quad -\frac{1}{n_2}\sum_{j=1}^{n_2}\int (f_1(x)-f_2(x))(\gamma_{2j}(x)-1)v(x)\,dx\,Y_{2j} - \int (f_1(x)-f_2(x))^2 v(x)\,dx
\end{aligned}
$$

and obtain

$$
\begin{aligned}
\mathrm{Var}(T_N^{(3)}) &= \frac{1}{n_1^2}\sum_{j=1}^{n_1}\sum_{k=1}^{n_1}\int\int (f_1(x)-f_2(x))(f_1(y)-f_2(y)) \\
&\qquad E\Big[(\gamma_{1j}(x)-1)(\gamma_{1k}(y)-1)Y_{1j}Y_{1k}\Big]v(x)v(y)\,dx\,dy \\
&\quad +\frac{1}{n_2^2}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}\int\int (f_1(x)-f_2(x))(f_1(y)-f_2(y)) \\
&\qquad E\Big[(\gamma_{2j}(x)-1)(\gamma_{2k}(y)-1)Y_{2j}Y_{2k}\Big]v(x)v(y)\,dx\,dy \\
&\quad -\Big(\int E\Big[(\gamma_{11}(x)-1)Y_{11}\Big](f_1(x)-f_2(x))v(x)\,dx\Big)^2 \\
&\quad -\Big(\int E\Big[(\gamma_{21}(x)-1)Y_{21}\Big](f_1(x)-f_2(x))v(x)\,dx\Big)^2
\end{aligned}
$$

$$
\begin{aligned}
&(34)\quad = \frac{1}{n_1}\int\int (f_1-f_2)(x)(f_1-f_2)(y)E\Big[(\gamma_{11}(x)-1)(\gamma_{11}(y)-1)Y_{11}^2\Big]v(x)v(y)\,dx\,dy \\
&(35)\qquad -\frac{1}{n_1}\Big(\int (f_1-f_2)(x)E\Big[(\gamma_{11}(x)-1)Y_{11}\Big]v(x)\,dx\Big)^2 \\
&(36)\qquad +\frac{1}{n_2}\int\int (f_1-f_2)(x)(f_1-f_2)(y)E\Big[(\gamma_{21}(x)-1)(\gamma_{21}(y)-1)Y_{21}^2\Big]v(x)v(y)\,dx\,dy \\
&(37)\qquad -\frac{1}{n_2}\Big(\int (f_1-f_2)(x)E\Big[(\gamma_{21}(x)-1)Y_{21}\Big]v(x)\,dx\Big)^2.
\end{aligned}
$$

For the first term on the right hand side (34) we have

$$\int\int (f_1-f_2)(x)(f_1-f_2)(y)E\Big[(\gamma_{11}(x)-1)(\gamma_{11}(y)-1)Y_{11}^2\Big]v(x)v(y)\,dx\,dy$$

$$= \int\int\int (f_1-f_2)(x)(f_1-f_2)(y)v(x)v(y)\Big(K_h(z-x)\frac{s_1(w)}{p_1(z,w)}-1\Big)\Big(K_h(z-y)\frac{s_1(w)}{p_1(z,w)}-1\Big)$$
$$\{(c_1+f_1(z)+g_1(w))^2+\sigma_1^2(z,w)\}p_1(z,w)\,d(z,w)\,dx\,dy+o(1)$$

$$= \int\Big(\int K_h(z-x)(f_1-f_2)(x)v(x)\,dx\Big)^2\frac{s_1^2(w)}{p_1^2(z,w)}\{(c_1+f_1(z)+g_1(w))^2+\sigma_1^2(z,w)\}$$
$$p_1(z,w)\,d(z,w)$$
$$-2\int\Big(\int(f_1-f_2)(x)v(x)\,dx\Big)\Big(\int K_h(z-y)(f_1-f_2)(y)v(y)\,dy\Big)$$
$$\times\frac{s_1(w)}{p_1(z,w)}\{(c_1+f_1(z)+g_1(w))^2+\sigma_1^2(z,w)\}p_1(z,w)\,d(z,w)$$
$$+\int\Big(\int(f_1-f_2)(x)v(x)\,dx\Big)^2\{(c_1+f_1(z)+g_1(w))^2+\sigma_1^2(z,w)\}p_1(z,w)\,d(z,w)+o(1)$$

$$= \int\Big((f_1-f_2)(z)v(z)\frac{s_1(w)}{p_1(z,w)}-\int(f_1-f_2)(x)v(x)\,dx\Big)^2$$
$$\times\{(c_1+f_1(z)+g_1(w))^2+\sigma_1^2(z,w)\}p_1(z,w)\,d(z,w)+o(1)$$

$$= E[(Z^{(1)})^2]+o(1)$$

where we have used $\int K_h(z-x)(f_1-f_2)(x)v(x)\,dx=(f_1-f_2)(z)v(z)+o(1)$ and $Z^{(1)}$ is defined in Theorem 3.1. For the calculation of (35) we observe:

$$\int(f_1-f_2)(x)E\Big[(\gamma_{11}(x)-1)Y_{11}\Big]v(x)\,dx$$
$$= \int(f_1-f_2)(x)v(x)\int\Big(K_h(z-x)\frac{s_1(w)}{p_1(z,w)}-1\Big)(c_1+f_1(z)+g_1(w))p_1(z,w)\,d(z,w)\,dx+o(1)$$
$$= \int(f_1-f_2)(x)v(x)f_1(x)\,dx+o(1).$$

Further we have for the expectation of $Z^{(1)}$:

$$E[Z^{(1)}]$$
$$= \int\Big((f_1-f_2)(z)v(z)\frac{s_1(w)}{p_1(z,w)}-\int(f_1-f_2)(x)v(x)\,dx\Big)\{c_1+f_1(z)+g_1(w)\}p_1(z,w)\,d(z,w)$$
$$= \int(f_1-f_2)(z)v(z)s_1(w)\{c_1+f_1(z)+g_1(w)\}\,d(z,w)-\int(f_1-f_2)(x)v(x)\,dx\,c_1$$
$$= \int(f_1-f_2)(z)v(z)f_1(z)\,dz+o(1).$$

From this and the analogous calculations for (36) and (37) we obtain for the variance

$$\mathrm{Var}(T_N^{(3)}) \;=\; \frac{1}{n_1}\mathrm{Var}(Z^{(1)}) + \frac{1}{n_2}\mathrm{Var}(Z^{(2)}) + o(\frac{1}{N}) \;=\; \frac{1}{N}\gamma^2 + o(\frac{1}{N}).$$

□

7.1.3 Asymptotic normality under the fixed alternative.

From the decomposition (25) of the test statistic and the Lemmata 7.2 and 7.4 we have under $H_1 : f_1 \neq f_2$:

$$\sqrt{N}(T_N - E[T_N]) \;=\; 2\sqrt{N}(T_N^{(3)} - E[T_N^{(3)}]) + o_p(1).$$

Further, analogous considerations as in the proofs of the Lemmata 7.2 and 7.4 show

$$\sqrt{N}(T_N^{(3)} - E[T_N^{(3)}]) \;=\; \sqrt{N}(\tilde{T}_N^{(3)} - E[\tilde{T}_N^{(3)}]) + o_p(1)$$

where $\tilde{T}_N^{(3)}$ is defined by

$$\tilde{T}_N^{(3)} \;=\; \frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(-1)^{i-1}\frac{1}{\kappa_i}\int (f_1 - f_2)(x)v(x)\{(\tilde{\gamma}_{ij}(x) - 1)Y_{ij} - f_i(x)\}\,dx$$

and

$$\tilde{\gamma}_{ij}(x) \;=\; K_h(X_{ij} - x)\frac{s_i(W_{ij})}{p_i(X_{ij}, W_{ij})}, \tag{38}$$

see the definition of $T_N^{(3)}$ in (28), $\hat{f}_i(x)$ in (7) and $\gamma_{ij}(x) = \tilde{\gamma}_{ij}(x) + O(h^r) + O(b^s) + O(1/(Nbh))$ in (8). $\tilde{T}_N^{(3)}$ is a sum of independent random variables and we can apply the central limit theorem for triangular arrays to prove asymptotic normality. Writing

$$\sqrt{N}(\tilde{T}_N^{(3)} - E[\tilde{T}_N^{(3)}]) \;=\; \frac{1}{\sqrt{N}}\sum_{i=1}^{2}\sum_{j=1}^{n_i} Z_{ij}$$

the asymptotic normal distribution follows from Lyapunov's condition:

$$\frac{1}{N^2}\sum_{i=1}^{2}\sum_{j=1}^{n_i} E[|Z_{ij}|^4] \;=\; O(\frac{1}{N}) \;=\; o(1)$$

which can be deduced in a similar manner as the proof of Lemma 7.4. □

7.1.4 Asymptotic normality under the null hypothesis

Under $H_0 : f_1 = f_2$ we have $T_N = T_N^{(1)}$ defined in (26). With the decomposition (29), the definition of the terms $T_N^{(i,\ell)}$ in (30) and an inspection of the proofs of the Lemmata 7.1 and 7.3 we obtain

$$T_N^{(1)} - E[T_N^{(1)}] \;=\; \tilde{T}_N^{(1,1)} - 2\tilde{T}_N^{(1,2)} + \tilde{T}_N^{(2,2)} + o_p(\frac{1}{N\sqrt{h}})$$

where $(i, \ell \in \{1, 2\})$

$$\begin{aligned}
\tilde{T}_N^{(i,\ell)} &= \frac{1}{n_i n_\ell} \sum_{j=1}^{n_i} \sum_{\substack{k=1\\(i,j)\neq(\ell,k)}}^{n_\ell} \Delta_{jk}^{(i,\ell)} \qquad \text{with} \\
\Delta_{jk}^{(i,\ell)} &= \int \Big((\tilde{\gamma}_{ij}(x) - 1) Y_{ij} - \int K_h(z-x) f_i(z)\, dz \Big) \\
&\qquad \Big((\tilde{\gamma}_{\ell k}(x) - 1) Y_{\ell k} - \int K_h(z-x) f_\ell(z)\, dz \Big) v(x)\, dx
\end{aligned}$$

and $\tilde{\gamma}_{ij}$ is defined in (38). Note that $E[\Delta_{jk}^{(i,\ell)}] = 0$ for $(i, j) \neq (\ell, k)$.

We define random variables $Z_1, \ldots, Z_N$ by

$$\begin{aligned}
Z_1 = (X_{11}, W_{11}, \varepsilon_{11}), \ldots, Z_{n_1} = (X_{1n_1}, W_{1n_1}, \varepsilon_{1n_1}),\\
Z_{n_1+1} = (X_{21}, W_{21}, \varepsilon_{21}), \ldots, Z_N = (X_{2n_2}, W_{2n_2}, \varepsilon_{2n_2}).
\end{aligned}$$

Then the standardized test statistic $N\sqrt{h}(T_N - E[T_N])$ is asymptotically equivalent to a quadratic form with vanishing diagonal elements, that is,

$$W(N) \;=\; \sum_{i=1}^{N} \sum_{\substack{j=1\\ i\neq j}}^{N} w_{ij,N}(Z_i, Z_j) \;=\; \sum_{1\le i<j\le N} V_{ij}$$

where $V_{ij} = (w_{ij,N}(Z_i, Z_j) + w_{ji,N}(Z_j, Z_i))/2$ and

$$w_{ij,N} \;=\; \begin{cases}
\dfrac{N\sqrt{h}}{n_1^2} \Delta_{ij}^{(1,1)} & : \; i, j = 1, \ldots, n_1 \\[2ex]
\dfrac{N\sqrt{h}}{n_2^2} \Delta_{i-n_1, j-n_1}^{(2,2)} & : \; i, j = n_1 + 1, \ldots, N \\[2ex]
-\dfrac{N\sqrt{h}}{n_1 n_2} \Delta_{i, j-n_1}^{(1,2)} & : \; i = 1, \ldots, n_1, j = n_1 + 1, \ldots, N \\[2ex]
-\dfrac{N\sqrt{h}}{n_1 n_2} \Delta_{i-n_1, j}^{(1,2)} & : \; i = n_1 + 1, \ldots, N, j = 1, \ldots, n_1.
\end{cases}$$

Because $W(N)$ is clean, that is $E[V_{ij}|Z_i] = 0$, we can apply Theorem 2.1 of de Jong (1987) and have to show

(a) $\lim_{N\to\infty} \frac{1}{\sigma^2(N)} \max_{1\le i\le N} \sum_{j=1}^{N} E[V_{ij}^2] \;=\; 0$

where $\sigma^2(N) = \mathrm{Var}(W(N)) = \mathrm{Var}(N\sqrt{h}(T_N - E[T_N])) + o(1) = O(1)$

(b) $E[(W(N))^4] \;=\; 3(\mathrm{Var}(W(N)))^2 + o(1)$ for $N \to \infty$.

To deduce condition (a) we consider only the case $i, j = 1, \ldots, n_1$ (the other cases are treated in the same way) and obtain

$$\begin{aligned}
E[V_{ij}^2] &= O(\frac{h}{N^2})E[(\Delta_{12}^{(1,1)})^2] \\
&= O(\frac{h}{N^2})\int\int v(x)v(y)\Big\{E\Big[\Big((\tilde\gamma_{11}(x)-1)Y_{11} - \textstyle\int K_h(z-x)f_1(z)\,dz\Big) \\
&\qquad\qquad\qquad\times\Big((\tilde\gamma_{12}(y)-1)Y_{12} - \textstyle\int K_h(z-y)f_1(z)\,dz\Big)\Big]\Big\}^2\,dx\,dy \\
&= O(\frac{h}{N^2})\Big(\int\int\int\int K_h(z-x)K_h(z-y)K_h(z'-x)K_h(z'-y)\ldots\,dx\,dy\,dz\,dz' + O(1)\Big) \\
&= O(\frac{1}{N^2})
\end{aligned}$$

uniformly in i, j. We obtain $1/\sigma^2(N)\max_i \sum_j E[V_{ij}^2] \;=\; O(1/N) \;=\; o(1)$ which proves condition (a).

To verify condition (b) we have

$$\begin{aligned}
E[(W(N))^4] &= 12\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N}{}^{\neq}\; E[V_{ij}^2]E[V_{k\ell}^2] + c_1\sum_{i=1}^{N}\sum_{j=1}^{N}{}^{\neq}\; E[V_{ij}^4] \\
&\quad + c_2\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N}{}^{\neq}\; E[V_{ij}V_{jk}V_{k\ell}V_{\ell i}] + c_3\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}{}^{\neq}\; E[V_{ij}V_{ik}^2V_{jk}] \\
(\mathrm{Var}(W(N)))^2 &= 4\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N}{}^{\neq}\; E[V_{ij}^2]E[V_{k\ell}^2] + c_4\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} E[V_{ij}^2]E[V_{ik}^2] \\
&\quad + c_5\sum_{i=1}^{N}\sum_{j=1}^{N}(E[V_{ij}^2])^2
\end{aligned}$$

where $c_1, \ldots, c_5$ denote some constants and the notation $\sum^{\neq}$ means summation over disjoint indices. Now condition (b) follows from some straightforward calculations establishing:

$$\sum_{i=1}^{N}\sum_{j=1}^{N}{}^{\neq} E[V_{ij}^4] = o(1), \quad \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N}{}^{\neq} E[V_{ij}V_{jk}V_{k\ell}V_{\ell i}] = o(1),$$
$$\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}{}^{\neq} E[V_{ij}V_{ik}^2V_{jk}] = o(1), \quad \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} E[V_{ij}^2]E[V_{ik}^2] = o(1),$$
$$\sum_{i=1}^{N}\sum_{j=1}^{N}(E[V_{ij}^2])^2 = o(1).$$

Asymptotic normality of $W(N)$ is deduced from Theorem 2.1 of de Jong (1987). $\square$

7.2 Proof of Theorem 5.1

We use a decomposition of the bootstrap test statistic T_N^* defined in (20) analogous to the decomposition (29) stated in the proof of Lemma 7.1,

$$T_N^* \;=\; T_N^{(1,1)*} - 2T_N^{(1,2)*} + T_N^{(2,2)*}\;,$$

where

$$T_N^{(i,\ell)*} \;=\; \frac{1}{n_i n_\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{n_\ell}\int\Big((\gamma_{ij}(x)-1)Y_{ij}^* - \hat f(x)\Big)\Big((\gamma_{\ell k}(x)-1)Y_{\ell k}^* - \hat f(x)\Big)v(x)\,dx\;,$$

where $i,\ell \in \{1,2\}$ and $T_N^{(1,2)*} = T_N^{(2,1)*}$.

We obtain the conditional variance of $N\sqrt{h}T_N^{(1,1)*}$ similar to the proof of Lemma 7.3:

$$\mathrm{Var}_N^*(N\sqrt{h}T_N^{(1,1)*}) \;=\; N^2hE\left[\Big(T_N^{(1,1)*} - E[T_N^{(1,1)*}\mid \mathcal{Y}_N]\Big)^2\;\Big|\;\mathcal{Y}_N\right] \;=$$

$$\begin{aligned}
= \frac{N^2h}{n_1^4}\sum_{j=1}^{n_1}\sum_{k=1}^{n_1}\sum_{j'=1}^{n_1}\sum_{k'=1}^{n_1}\Big\{&E_N^*\Big[\int\Big((\gamma_{1j}(x)-1)Y_{1j}^*-\hat f(x)\Big)\Big((\gamma_{1k}(x)-1)Y_{1k}^*-\hat f(x)\Big)v(x)\,dx\\
&\times\int\Big((\gamma_{1j'}(y)-1)Y_{1j'}^*-\hat f(y)\Big)\Big((\gamma_{1k'}(y)-1)Y_{1k'}^*-\hat f(y)\Big)v(y)\,dy\Big]\\
&-E_N^*\Big[\int\Big((\gamma_{1j}(x)-1)Y_{1j}^*-\hat f(x)\Big)\Big((\gamma_{1k}(x)-1)Y_{1k}^*-\hat f(x)\Big)v(x)\,dx\Big]\\
&\times E_N^*\Big[\int\Big((\gamma_{1j'}(y)-1)Y_{1j'}^*-\hat f(y)\Big)\Big((\gamma_{1k'}(y)-1)Y_{1k'}^*-\hat f(y)\Big)v(y)\,dy\Big]\Big\}.
\end{aligned}$$

The addends of the sum vanish if $\{j,k\}\cap\{j',k'\}=\emptyset$. Via a rather technical calculation, that is omitted, we can show that the dominating part of the sum is obtained in the cases

$j = j' \neq k = k'$ or $j = k' \neq j' = k$, compare the proof of Lemma 7.3. Therefore we have

$$
\begin{aligned}
&\mathrm{Var}_N^*(N\sqrt{h}T_N^{(1,1)*}) = \\
&= \frac{2N^2h}{n_1^4}\sum_{j=1}^{n_1}\sum_{\substack{k=1\\ j\neq k}}^{n_1}\int\int E_N^*\Big[(\gamma_{1j}(x)-1)(\gamma_{1j}(y)-1)(Y_{1j}^*)^2 - \hat{f}(y)(\gamma_{1j}(x)-1)Y_{1j}^* \\
&\qquad\qquad - \hat{f}(x)(\gamma_{1j}(y)-1)Y_{1j}^* + \hat{f}(x)\hat{f}(y)\Big] \\
&\qquad\quad E_N^*\Big[(\gamma_{1k}(x)-1)(\gamma_{1k}(y)-1)(Y_{1k}^*)^2 - \hat{f}(y)(\gamma_{1k}(x)-1)Y_{1k}^* \\
&\qquad\qquad - \hat{f}(x)(\gamma_{1k}(y)-1)Y_{1k}^* + \hat{f}(x)\hat{f}(y)\Big]v(x)v(y)\,dx\,dy \\
&\qquad - \Big(\int\Big\{(\gamma_{1j}(x)-1)E_N^*[Y_{1j}^*] - \hat{f}(x)\Big\}\Big\{(\gamma_{1k}(x)-1)E_N^*[Y_{1k}^*] - \hat{f}(x)\Big\}v(x)\,dx\Big)^2 \\
&\quad + o_p(1).
\end{aligned}
$$

Now using the notation $\hat{h}_{1j} = \hat{c}_1 + \hat{f}(X_{1j}) + \hat{g}_1(W_{1j})$ and observing (18) and (19) from Section 5.1 follows

$$
\begin{aligned}
&\mathrm{Var}_N^*(N\sqrt{h}T_N^{(1,1)*}) \;= \\
&= \frac{2N^2h}{n_1^4}\sum_{j=1}^{n_1}\sum_{\substack{k=1\\ j\neq k}}^{n_1}\int\int\Big\{(\gamma_{1j}(x)-1)(\gamma_{1j}(y)-1)(\gamma_{1k}(x)-1)(\gamma_{1k}(y)-1)(\hat{h}_{1j}^2+\hat{\varepsilon}_{1j}^2)(\hat{h}_{1k}^2+\hat{\varepsilon}_{1k}^2) \\
&\quad - \{\hat{f}(y)(\gamma_{1j}(x)-1)+\hat{f}(x)(\gamma_{1j}(y)-1)\}(\gamma_{1k}(x)-1)(\gamma_{1k}(y)-1)\hat{h}_{1j}\hat{\varepsilon}_{1k}^2 \\
&\quad - \{\hat{f}(y)(\gamma_{1k}(x)-1)+\hat{f}(x)(\gamma_{1k}(y)-1)\}(\gamma_{1j}(x)-1)(\gamma_{1j}(y)-1)\hat{h}_{1k}\hat{\varepsilon}_{1j}^2 \\
&\quad + \hat{f}(x)\hat{f}(y)(\gamma_{1k}(x)-1)(\gamma_{1k}(y)-1)\hat{\varepsilon}_{1k}^2 + \hat{f}(x)\hat{f}(y)(\gamma_{1j}(x)-1)(\gamma_{1j}(y)-1)\hat{\varepsilon}_{1j}^2 \\
&\quad - (\gamma_{1j}(x)-1)(\gamma_{1j}(y)-1)(\gamma_{1k}(x)-1)(\gamma_{1k}(y)-1)\hat{h}_{1j}^2\hat{h}_{1k}^2\Big\}v(x)v(y)\,dx\,dy \\
&\quad + o_p(1) \\
&= \frac{2N^2h}{n_1^2}\int\int\Big(\frac{1}{n_1}\sum_{j=1}^{n_1}\gamma_{1j}(x)\gamma_{1j}(y)(\hat{h}_{1j}^2+\hat{\varepsilon}_{1j}^2)\Big)^2 v(x)v(y)\,dx\,dy \\
&\quad - \frac{2N^2h}{n_1^2}\int\int\Big(\frac{1}{n_1}\sum_{j=1}^{n_1}\gamma_{1j}(x)\gamma_{1j}(y)\hat{h}_{1j}^2\Big)^2 v(x)v(y)\,dx\,dy \\
&\quad + o_p(1).
\end{aligned}
$$

Taking into account the definition of the weights γ_{1j} and the convergence of the regression estimators in probability we obtain for the limit of the conditional variance

$$
\begin{aligned}
&\operatorname{Var}_N^*(N\sqrt{h}T_N^{(1,1)*})\\
= \;& \frac{2N^2h}{n_1^2}\int\!\!\int\Big(\int K_h(z-x)K_h(z-y)\frac{s_1^2(w)}{p_1(z,w)}\{(c_1+f_1(z)+g_1(w))^2+\sigma_1^2(z,w)\,d(z,w)\}\Big)^2\\
&\qquad -\Big(\int K_h(z-x)K_h(z-y)\frac{s_1^2(w)}{p_1(z,w)}(c_1+f_1(z)+g_1(w))^2\,d(z,w)\}\Big)^2 v(x)v(y)\,dx\,dy\\
&+o_p(1)\\
= \;& 2\int (K*K)^2(u)\,du\Big\{\int\Big(\int\frac{s_1^2(w)}{\kappa_1 p_1(z,w)}\{(c_1+f_1(z)+g_1(w))^2+\sigma_1^2(z,w)\}\,dw\Big)^2 v^2(z)\,dz\\
&\qquad -\int\Big(\int\frac{s_1^2(w)}{\kappa_1 p_1(z,w)}(c_1+f_1(z)+g_1(w))^2\,dw\Big)^2 v^2(z)\,dz\Big\}\\
&+o_p(1),
\end{aligned}
$$

where the last equality follows analogously to the corresponding calculations of $\operatorname{Var}(T_N^{(1,1)})$ in the proof of Lemma 7.3. The statement of Theorem 5.1 follows with similar calculations for $\operatorname{Var}_N^*(N\sqrt{h}T_N^{(1,2)*})$ and $\operatorname{Var}_N^*(N\sqrt{h}T_N^{(2,2)*})$. □

References

P. Cabus (1998). *Un test de type Kolmogorov–Smirnov dans le cadre de comparaison de fonctions de régression.* C. R. Acad. Sci. Paris Sér. I Math. 327, 11, 939–942. (in French)

C. Camlong-Viot (2000). Modèle additif de régression sous des condicions de mélange. *Ph.D. thesis at University Toulouse III - Paul Sabatier* (in French)

M. A. Delgado (1993). *Testing the equality of nonparametric regression curves.* Statistics and Probability Letters 17, 199–204.

M. A. Delgado, J. M. Rodriguez-Poó and M. Wolf (2001). *Subsampling Cube Root Asymptotics with an Application to Manski's MSE.* Economics Letters 73, 241–250.

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

H. Dette and N. Neumeyer (2001). *Nonparametric analysis of covariance.* The Annals of Statistics 29, 1361–1400.

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

T. Gasser, H.–G. Müller and V. Mammitzsch (1985). *Kernels for Nonparametric Curve Estimation.* Journal of the Royal Statistical Society, Series B 47, 238–252.

P.L. Gozalo and O. B. Linton (2001). *A Nonparametric Test of Additivity in Generalized Nonparametric Regression with Estimated Parameters.* Journal of Econometrics 104, 1–48.

T. Gørgens (2002). *Nonparametric comparison of regression curves by local linear fitting.* Statistics and Probability Letters 60, 81–89.

W. Härdle, S. Huet, E. Mammen and S. Sperlich (2003). *Bootstrap Inference in Semiparametric Generalized Additive Models.* forthcoming in Econometric Theory

W. Härdle and E. Mammen (1993). *Comparing nonparametric versus parametric regression fits.* The Annals of Statistics 21, 1926–1947.

W. Härdle and J. S. Marron (1990). *Semiparametric comparison of regression curves.* The Annals of Statistics 18, 63–89.

W. Härdle and J. S. Marron (1991). *Bootstrap Simultaneous Error Bars for Nonparametric Regression.* The Annals of Statistics 19, 778–796.

W. Härdle, S. Sperlich and V. Spokoiny (2001) *Structural tests in additive regression.* Journal of the American Statistical Association 96, 1333–1347.

P. Hall and J. W. Hart (1990). *Bootstrap test for difference between means in nonparametric regression.* Journal of the American Statistical Association 85, 1039–1049.

P. Hall, C. Huber and P. L. Speckman (1997). *Covariate–matched one–sided tests for the difference between functional means.* Journal of the American Statistical Association 92, 1074–1083.

N. W. Hengartner and S. Sperlich (2001). *Rate Optimal Estimation with the Integration Method in the Presence of Many Covariates.* Working Paper 01-69, Carlos III de Madrid

V. Hjellvik, Q. Yao and D. Tjøstheim (1998). *Linearity testing using local polynomial approximation.* Journal of Statistical Planning and Inference 68, 295-321.

J. Horowitz (2001). *Nonparametric Estimation of a Generalized Additive Model with an Unknown Link Function.* Econometrica 69, 499–513.

J. Horowitz and V. Spokoiny (2001). *An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model against a Nonparametric Alternative.* Econometrica 69, 599–631.

W. Kim, O. B. Linton and N. W. Hengartner (1999). *A Computationally Efficient Oracle Estimator for Additive Nonparametric Regression with Bootstrap Confidence Intervals.* Journal of Computational and Graphical Statistics 8, 278–297.

E. King, J. D. Hart and D. Wehrly (1991). *Testing the equality of two regression curves using linear smoothers.* Statistics and Probability Letters 12, 239–247.

K. B. Kulasekera (1995). *Comparison of regression curves using quasi residuals.* Journal of the American Statistical Association 90, 1085–1093.

K. B. Kulasekera and J. Wang (2001). *A test of equality of regression curves using Gâteaux Scores.* Australian & New Zealand Journal of Statistics 43(1), 89–99.

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

O. B. Linton and J. P. Nielsen (1995). *A kernel method of estimating structured nonparametric regression based on marginal integration.* Biometrika 82, 93–101.

E. Mammen and J. P. Nielsen (2003) *Generalised Structured Models.* Biometrika 90, in press.

N. Neumeyer and H. Dette (2003). *Nonparametric comparison of regression functions – an empirical process approach.* The Annals of Statistics 31, 880–920.

W. K. Newey (1994). *Kernel estimation of partial means.* Econometric Theory 10, 233–253.

D. N. Politis, J. P. Romano and M. Wolf (1999). Subsampling *Springer-Verlag, New York.*

D. N. Politis, J. P. Romano and M. Wolf (2001). *On the Asymptotic Theory of Subsampling.* Statistica Sinica 11, 1105–1124.

J. M. Rodríguez-Póo, S. Sperlich and P. Vieu (2001). *Normalité asymptotique destimateurs de maximum de vraisemblance pour modèles non-paramétriques de régression multidimensionelle.* C. R. Acad. Sci. Paris, 333, I, 61–64. (in French)

J. M. Rodríguez-Póo, S. Sperlich and P. Vieu (2003). *Semiparametric Estimation of Weak and Strong Separable Models.* to appear in Econometric Theory

S. Sperlich (1998). Additive modelling and Testing Model Specification. *Shaker Verlag Aachen.*

S. Sperlich, O. Linton and W. Härdle (1999). *Integration and Backfitting Methods in Additive Models - Finite Sample Properties and Comparison.* Test 8, 419–458.

S. Sperlich, D. Tjøstheim and L. Yang (2002). *Nonparametric Estimation and Testing of Interaction in Additive Models.* Econometric Theory 18, 197–251.

C. J. Stone (1986). *The dimensionality reduction principle for generalized additive models.* Annals of Statistics 14, 90–606.

D. Tjøstheim and B. H. Auestad (1994). *Nonparametric identification of nonlinear time series: projections.* Journal of the American Statistical Association 89, 1398–1409.

A. Yatchew (1999). *An elementary differencing test of equality of regression functions.* Economics Letters 62, 271–278.

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