# NONPARAMETRIC COMPARISON OF REGRESSION FUNCTIONS

Ramidha Srihera and Winfried Stute

*National Institute of Development Administration, Bangkok, Thailand and University of Giessen, Germany*

**Abstract**

In this work we provide a new methodology for comparing regression functions $m_1$ and $m_2$ from two samples. Since apart from smoothness no other (parametric) assumptions are required, our approach is based on a comparison of nonparametric estimators $\hat{m}_1$ and $\hat{m}_2$ of $m_1$ and $m_2$, respectively. The test statistics $\hat{T}$ incorporate weighted differences of $\hat{m}_1$ and $\hat{m}_2$ computed at selected points. Since the design variables may come from different distributions a crucial question is where to compare the two estimators. As our main results we obtain the limit distribution of $\hat{T}$ (properly standardized) under the null hypothesis $H_0 : m_1 = m_2$ and under local and global alternatives. We are also able to choose the weight function so as to maximize power. Furthermore, the tests are asymptotically distribution-free under $H_0$ and shift and scale-invariant. Several of such $\hat{T}$'s may then be combined to get Maximin tests when the dimension of the local alternative is finite. In a simulation study we found out that our tests achieve the nominal level and have excellent power already for small to moderate sample sizes. As to proofs we heavily make use of new results from empirical process theory, U-statistics and nonparametric curve estimation.

*AMS 2000 subject classifications.* 62G10, 62G08.

*Key words and phrases.* Nonparametric regression, test of equality, Maximin test.

## 1 Introduction

In many applied fields, e.g., health science, engineering, agriculture or medicine, it has always been of interest to choose between two complementary courses of action. For example, medical researchers often face the problem, that they have to make a decision whether a new treatment is better than an existing one. Such a comparison should encounter auxiliary information such as the age of the patient or the size of a tumor at surgery. This kind of problem is related with studying the relationship between an independent prognostic factor (dose, input) $X$ and an associated dependent response (output) $Y$. In the real world the relationship between $X$ and $Y$ is not completely deterministic but subject to noise. To be more specific, we rather have

$$Y = m(X) + \varepsilon,$$

where $m$ is the regression function of $Y$ on $X$ and $\varepsilon$ is an error variable orthogonal to $X$, i.e., $\mathbb{E}(\varepsilon|X) = 0$. If $X = x$, then $m(x)$ is the optimal predictor of $Y$.

In the analysis of two populations, one may be interested to compare the two associated regression curves. For example, if $Y$ denotes the disease free survival time after surgery, then $m_1(x)$ and $m_2(x)$ may denote the expected value of $Y$ under treatment and control, respectively, given the covariate at surgery equals $X = x$. If $m_1(x) = m_2(x)$ for all $x$, there will be no systematic difference between the two groups while $m_1 \geq m_2$ but $m_1 \neq m_2$ indicates an improvement under treatment.

Unfortunately, the two functions $m_1$ and $m_2$ are unknown and need to be estimated from two samples of data. A proper test for

$$H_0 : m_1 = m_2 \text{ versus } H_1 : m_1 \neq m_2$$

or some more specified alternatives may then be based on two estimators $\hat{m}_1$ and $\hat{m}_2$, say.

In an unconditional framework, testing for differences in two means has a long history. Under the assumption, that the two samples come from a normal population, this resulted in the famous $t$-test. If this assumption cannot be justified, the distribution of the test statistic admits an approximation through a standard normal distribution. In the context of the Linear Model the $F$-test provides a way to check the equality of two regression functions in a particular parametric framework.

To motivate our approach, some further notation is necessary. Let $(X_1, Y_1)$ and $(X_2, Y_2)$ be two measurements on two populations. Assuming $\mathbb{E}|Y_1| < \infty$ and $\mathbb{E}|Y_2| < \infty$, then the conditional expectations $\mathbb{E}[Y_1|X_1]$ and $\mathbb{E}[Y_2|X_2]$ exist and allow for factorizations

$$\mathbb{E}[Y_1|X_1] = m_1(X_1), \quad \mathbb{E}[Y_2|X_2] = m_2(X_2)$$

through the regression functions $m_1$ and $m_2$. Let $(X_{11}, Y_{11}), \ldots, (X_{1n_1}, Y_{1n_1})$ and $(X_{21}, Y_{21}), \ldots, (X_{2n_2}, Y_{2n_2})$ be two samples of independent replicates of $(X_1, Y_1)$ and $(X_2, Y_2)$, respectively. A general class of nonparametric estimators for $m_1$ and $m_2$ was proposed by Stone (1977). They are of the form

$$\hat{m}_1(x) = \sum_{i=1}^{n_1} Y_{1i} W_{1i}(x) \qquad \hat{m}_2(x) = \sum_{i=1}^{n_2} Y_{2i} W_{2i}(x),$$

where $W_{1i}$ and $W_{2i}$ are proper weights depending on the input data of each sample, preferably satisfying $\sum_{i=1}^{n_1} W_{1i}(x) = 1 = \sum_{i=1}^{n_2} W_{2i}(x)$. Note that these conditions imply that the resulting $\hat{m}$'s are scale and shift-equivariant. This means, that if each $Y_i$ is replaced by $Y_i^* = aY_i + b$ for some constants $a$ and $b$, then the resulting estimator $\hat{m}^*$ satisfies $\hat{m}^*(x) = a\hat{m}(x) + b$. Recall that the sample means are also of this type but with weights $W_{1i}(x) = n_1^{-1}$ and $W_{2i}(x) = n_2^{-1}$ not depending on $x$. Since we want to estimate a function rather than an unknown parameter, our $W_{1i}$ and $W_{2i}$ will depend on $x$. Informally speaking $W_i(x)$ attaches more mass to those $X_i$'s which are closer to $x$ and less weight to the remote $X_i$'s. Two of the most popular estimators of a regression function are

- the Nadaraya-Watson estimator (NW)
- the Nearest-Neighbor estimator (NN).

For the NW-estimator we have for the first sample, e.g.,

$$W_{1i}(x) = \frac{K\left(\frac{X_{1i}-x}{h}\right)}{\sum_{j=1}^{n_1} K\left(\frac{X_{1j}-x}{h}\right)},$$

where $h > 0$ is an appropriate smoothing parameter (window-width) and $K$ is a symmetric kernel function. For details, see Nadaraya (1964). Similarly, for the second sample. For the (symmetrized) NN-estimator, one has to replace $X_{1i}$ by $\hat{F}(X_{1i})$, where with $n = n_1$

$$\hat{F}(x) = \hat{F}_n(x) = n^{-1} \sum_{j=1}^{n} 1_{\{X_{1j} \le x\}}$$

is the empirical distribution function (d.f.) of the sample $X_{11}, \ldots, X_{1n_1}$. In other words, $\hat{F}_n(X_{1i})$ is the normalized rank of $X_{1i}$ within the first data set, and $W_{1i}(x)$ becomes

$$W_{1i}(x) = \frac{K\left(\frac{\hat{F}_n(X_{1i}) - \hat{F}_n(x)}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{\hat{F}_n(X_{1j}) - \hat{F}_n(x)}{h}\right)}.$$

Similarly, for the second sample (with $n = n_2$), we obtain $W_{2i}$, in which $\hat{F}_n$ is replaced by $\hat{G} = \hat{G}_n$, the empirical d.f. of the $X_{2j}$, $1 \le j \le n$.

Though, at first sight, the NN-weights seem to be more complicated than the NW-weights, the resulting NN-estimator has several advantages over the NW-estimator. One disadvantage of the NW-estimator comes from the fact that the denominator of $W_i$ may become zero or at least close to zero. This results in an estimator of $m$ which does not admit finite moments.

In contrast, we will show that the NN-weights have many attractive properties which make them especially suited for the problems discussed in this paper. A pointwise analysis of this estimator may be found in Stute (1984).

The next important question to be discussed is where both $\hat{m}_1$ and $\hat{m}_2$ should be compared. Typically, if $X_1 \sim F$ and $X_2 \sim G$ and $F$ and $G$ have disjoint supports, the testing problem is more difficult since the information about the two samples is located in separate regions. If, on the other hand, $F$ is close to $G$ we may expect both $X$-samples to be mixed up so that comparing $\hat{m}_1$ and $\hat{m}_2$ only there makes sense. As a conclusion one may say that the points where $\hat{m}_1$ and $\hat{m}_2$ are to be compared should be chosen in an adaptive way. In this work we propose averaging each $X_{1i}$ with $X_{2j}$. By this we obtain data-dependent points which are located between the two $X$-samples and therefore constitute a reasonable area on which a test should be based. Particularly, when both $X$-samples are mixed up, then the area where $\hat{m}_1$ and $\hat{m}_2$ are compared coincides more or less with the supports of $F$ and $G$.

The class of test statistics to be studied first will be linear in the sense that we sum up all differences

$$\hat{m}_1\left(\frac{X_{1i}+X_{2j}}{2}\right) - \hat{m}_2\left(\frac{X_{1i}+X_{2j}}{2}\right), \quad 1 \le i \le n_1, 1 \le j \le n_2.$$

As it will turn out it is also important to properly weight each of the above differences, say by $W\left(\frac{X_{1i}+X_{2j}}{2}\right)$. The choice of the weight function $W$ is delicate. We shall show how to choose $W$ in order to maximize power when the direction of the alternative is specified.

Summarizing so far, in this work we first propose and analyze two-sample score test statistics of the form

$$\hat{T} = \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} W\left(\frac{X_{1i}+X_{2j}}{2}\right)\left[\hat{m}_1\left(\frac{X_{1i}+X_{2j}}{2}\right) - \hat{m}_2\left(\frac{X_{1i}+X_{2j}}{2}\right)\right]. \qquad (1.1)$$

Note that since our $\hat{m}_1$ and $\hat{m}_2$ are scale and shift-equivariant, $\hat{T}$ is shift-invariant but scale-equivariant: $\hat{T}^* = a\hat{T}$. After that we show how to combine several of these $\hat{T}$'s to create tests, which are Maximin among tests for $H_0$ versus local alternatives with finite codimension. Moreover, under $H_0$, these tests will turn out to be asymptotically distribution-free and shift- and scale-invariant.

To review the literature, Härdle and Marron (1990) analyzed semiparametric models by comparing nonparametric regression functions under the assumption of fixed equal designs. The main objective of Hall and Hart (1990) was the discussion of using a bootstrap procedure for two-sided tests for $H_0 : m_1 = m_2$ under the assumption, that there are no ties among the design points which are assumed to be identical in the two groups. King, Hart and Wehrly (1991) also presented a test based on the difference between two curve estimators from kernel smoothers when the design points are fixed and equal. Next, Delgado (1993) discussed a test for the equality of nonparametric regression functions which has characteristics analogous to the Kolmogorov-Smirnov statistic. His test did not require smoothing and is easy to implement. At the same time he required the design points to be fixed in advance and being equal for the two samples. Young and Bowman (1995) investigated tests for equality and parallelism across groups with the covariate effect being estimated via Gasser-Müller smoothing. Kulasekera (1995) presented three tests with common fixed design points or equal sample sizes not being necessary. The first two tests are based on Quasi-Residuals techniques while the last test is based on estimators of the variances of the error distributions. Furthermore, Kulasekera and Wang (1997) examined selection of smoothing parameters which effect the power in the three nonparametric tests of Kulasekera (1995).

In Hall, Huber and Speckman (1997) one-sided test statistics for two functional means using an interpolation-based approach were proposed. Their test can be used when means of treatment effects are continuous functions of the covariates and the design sequences are random samples from different densities in two samples. Munk and Dette (1998) investigated a consistent test for the comparison of two regression functions with fixed but unequal design points. In their test they evaluate the difference between two curves based on a weighted $L^2$ distance. Next, Dette and Neumeyer (2001) provided three test statistics, the first of which was based on a linear combination of estimators for the integrated variance function. The second used one-way analysis of variance in the framework of non-parametric curve estimation, while the third applied Munk and Dette's (1998) approach for testing the equality of $k$ regression functions from independent samples. Again, the design points were non-random but nonnecessarily equal. Finally, Neumeyer and Dette (2003) introduced a test for the comparison of two regression curves which is based on a difference of two marked empirical processes based on residuals obtained under the assumption of equal regression curves. Their test can detect alternatives converging to the null model at the rate $(n_1 + n_2)^{-1/2}$, where $n_1$ and $n_2$ denote the two sample sizes.

Summarizing, most of the papers cited so far only deal with fixed design. Many times even equal sample sizes and equal design points were required. Notable exceptions are Scheike (2000) and Neumeyer and Dette (2003) who seem to be the first to study tests for equality of regression curves under random design. Scheike (2000) modified the integrated regression approach proposed by Stute (1997) and replaced the empirical integrals by the Lebesgue integral. By this he compared the two regression estimators on areas which do not depend on the data and therefore may not contain relevant information. Neumeyer and Dette (2003) compare two Nadaraya-Watson estimators. The problem with these estimators is the fact that their denominators may be very unstable. For this reason, Neumeyer and Dette (2003) only compared the numerators which are estimators of $m_1 f$ and $m_2 g$, respectively. Here $f$ and $g$ are the densities of the input variables in each of the two samples. The approximation by the limit distribution is not satisfactory so that a bootstrap is proposed. Also their test is not shift- and scale-invariant.

For other related work on the subject, we also refer to Koul and Schick (1997, 2003), Lavergne (2001), Gørgens (2002) and Pardo-Fernandéz et al. (2006). In most of this work, a detailed analysis of the power of the tests is missing.

Also the discussion of the role played by the design distributions $F$ and $G$ is often misleading. One can often find an argument that $F$ and $G$ can be assumed "without loss of generality" to be supported by the unit interval. In other areas of statistics like Robust Statistics, such a remark would probably raise some "objection", for good reason. For example, suppose that $F$ and $G$ are known and continuous. Then one may transform each $X_{1i}$ to $U_i = F(X_{1i})$, which is uniform on the unit interval. As a consequence the regression of $Y_{1i}$ given $U_i$ becomes $m_1 \circ F^{-1}$, which is defined on the open unit interval and is typically unbounded there. So a transformation of the input data also has some consequences for the regression to the effect that the new regression does not satisfy the regularity assumptions traditionally appearing in the literature on smoothing. As a consequence the distributions $F$ and $G$ do play an intrinsic role for designing tests for equality in regression.

In this paper we restrict ourselves to the null hypothesis $H_0 : m_1 = m_2$. We only mention that our approach can be extended to the null model when $m_1$ and $m_2$ are supposed to differ by a function $u(x, \theta)$. In such a situation we need to replace $\hat{m}_1 - \hat{m}_2$ by $\hat{m}_1 - \hat{m}_2 - u(\cdot, \hat{\theta})$, where $\hat{\theta}$ is a consistent estimator of $\theta$. Details are omitted.

Our final comment is on the decomposition $Y = m(X) + \varepsilon$. For random design, this decomposition just involves orthogonality of $X$ and $\varepsilon$. No independence between $X$ and $\varepsilon$ is to be imposed, nor do we require that $\varepsilon = \sigma(X)\eta$, where $\eta$ is independent of $X$. Actually, our paper also covers the case of discrete $Y$'s, and it is known that, e.g., in dichotomous or Poisson regression such assumptions do not hold.

Summarizing, in this paper we provide a discussion and analysis of tests which take into account

- the design distributions $F$ and $G$
- a detailed study of local power
- distribution-freeness under $H_0$
- heteroscedasticity of the noise variables
- possible discrete $Y$'s
- good finite sample approximations
- shift and scale-invariance
- construction of Maximin tests

# 2 Main Results

In this chapter we will present the main results of our work. Theorem 2.1 contributes a martingale representation of $\hat{T}$, i.e., a representation of $\hat{T}$ as a martingale, a negligible term and a deterministic term which vanishes under $H_0$ but is in charge of the power under $H_1$. It is interesting to note that the martingale part does not have independent identically distributed (i.i.d.) but dependent summands. In Theorem 2.2 we apply the martingale CLT to derive the asymptotic normality of $\sqrt{N}\hat{T}$, where $N$ is a proper standardizing factor depending on the individual sample sizes $n_1$ and $n_2$. After that we answer the question how to choose the weight function $W$ in order to maximize local power of the test, when the two $m$'s differ by a multiple of a fixed function $s$. An interesting extension investigates alternative models, in which the difference is a function spanned by finitely many directions $s_1, \ldots, s_d$.

In the limit the test statistic has a $\chi^2_d$-distribution under $H_0$ so that critical values are readily available. In particular, the test is asymptotically distribution-free under $H_0$. Under $H_1$, the test has a noncentral $\chi^2_d$-distribution. As it will turn out our test is Maximin. We also show that it is consistent when the alternative is fixed. Finally we will see that the test is shift- and scale invariant. In Chapter 3 we report on various simulation results. Proofs will be deferred to Chapter 4.

We already briefly mentioned that the power of our test will heavily depend on $W$. Another issue is the choice of the smoothing kernel $K$ and the bandwidth $h > 0$. As to $K$, we require assumptions which are standard in the literature:

(K) (i) $K(x) = K(-x)$ for $x \in \mathbb{R}$ and $K$ is nonnegative and nondecreasing on the negative real line.

(ii) $\int K(x)dx = 1$.

(iii) $K$ has compact support and is twice continuously differentiable.

Condition (ii) is made only for convenience. In the case $\int K(x)dx = c \neq 0$ we may replace $K$ with $K/c$ without changing the NN-weights and hence our estimators $\hat{m}_1$ and $\hat{m}_2$. Also the assumption (iii) could be weakened and replaced by $K(x) \to 0$ sufficiently fast as $x \to \pm\infty$. It is needed to exploit the local structure of the data. The symmetry condition as always is to control the bias in estimating $m_1$ and $m_2$. The monotonicity is helpful to bound the difference between the Lebesgue-integral of $K$ and approximating Riemann sums.

As to the sample sizes $n_1$ and $n_2$, as always in two-sample problems, we have to guarantee that the information contained in the two samples is approximately proportional. This property may be expressed through

(N) $\dfrac{n_1}{n_1+n_2} \to \lambda$ and $\dfrac{n_2}{n_1+n_2} \to 1-\lambda$ where $0 < \lambda < 1$.

Condition (N) implies some balance between the two samples. In terms of $n_1$ and $n_2$ the standardizing factor for $\hat{T}$ will be $\sqrt{N}$ where

$$N = \frac{n_1 n_2}{n_1 + n_2}.$$

Our next assumption will concern the bandwidth $h$. Actually $h = h_{n_1,n_2}$ with $h \to 0$ as $n_1, n_2 \to \infty$. A proper choice of $h$ is always a delicate question. A larger $h$ would incorporate neighbors at a larger distance and destroy the local flavor of $\hat{m}_1$ and $\hat{m}_2$. On the other hand, a small $h$ would give rise only to few neighbors resulting in $\hat{m}_1$ and $\hat{m}_2$ with a small bias but a larger variance. As a consequence $h$ should converge to zero at a proper rate only. As it will turn out in our situation

(h) $h \to 0$ as $n^{-\beta}$, where $\frac{1}{4} < \beta < \frac{1}{3}$ and $n$ has the order of $n_1$ and $n_2$.

This choice of $h$ guarantees

(i) $nh^3 \to \infty$ as $n \to \infty$

(ii) $nh^4 \to 0$ as $n \to \infty$

Since under $H_0$ the limit distribution of $\sqrt{N}\hat{T}$ is known, a data driven choice of $h$ could be taken from bootstrap samples in such a way, that the bootstrap distribution of $\sqrt{N}\hat{T}$ is the closest to its limit. Due to lack of space this will be, however, not further pursued in this paper.

Also the smoothness conditions to follow are standard:

(S) $f, g, m$ and $W$ are twice continuously differentiable.

Finally, we have to guarantee that the second moments of our (approximating) terms exist:

(M) For $\rho_1^2$ and $\rho_2^2$ from (2.2) and (2.3) below we have $\rho_1^2 < \infty, \rho_2^2 < \infty$.

To formulate our first result, recall that the null hypothesis always is

$$H_0 : m_2 = m_1.$$

The (local) alternative considered in Theorem 2.1 below will be

$$m_2 = m_1 + \frac{cs}{\sqrt{N}},$$

where the function $s$ is specified and determines the direction of the deviation between $m_1$ and $m_2$. The choice of $c = 0$ again leads to $m_1 = m_2$. Also recall that $X_{11} \sim F$ and $X_{21} \sim G$ are the unknown distributions of the design variables with densities $f$ and $g$. Furthermore, let $H$ be the d.f. of the $(X_{1i} + X_{2j})/2$ and $h$ its Lebesgue density.

**Theorem 2.1** *Under (K), (N), (h), (S) and (M), assume*

$$m_2 = m_1 + \frac{cs}{\sqrt{N}}. \tag{2.1}$$

*Then we have the following expansion of $\hat{T}$:*

$$\begin{aligned}\sqrt{N}\hat{T} = \sqrt{1-\lambda}\ n_1^{1/2}\sum_{i=1}^{n_1}(Y_{1i} - m_1(X_{1i})) \int W(x)W_{1i}(x)H(dx) \\ - \sqrt{\lambda}\, n_2^{1/2}\sum_{i=1}^{n_2}(Y_{2i} - m_2(X_{2i})) \int W(x)W_{2i}(x)H(dx) \\ -c\int W(x)s(x)H(dx) + o_{\mathbb{P}}(1).\end{aligned}$$

Note that in each sum the summands form a martingale difference array. Also both sums are independent. An application of the CLT for martingale difference arrays yields the following result.

**Theorem 2.2** *Under the assumptions of Theorem 2.1 we have*

$$\sqrt{N}\hat{T} \xrightarrow{\mathcal{L}} \mathcal{N}(\mu, \sigma^2) \text{ as } N \to \infty,$$

*where*

$$\mu = -c \int W(x)s(x)H(dx)$$

*and*

$$\sigma^2 = (1-\lambda)\rho_1^2 + \lambda\rho_2^2.$$

*Here $\mathcal{L}$ denotes convergence in distribution. Furthermore,*

$$\rho_1^2 = \int \sigma_1^2(x)W^2(x)\frac{h(x)}{f(x)}H(dx) \tag{2.2}$$

$$\rho_2^2 = \int \sigma_2^2(x)W^2(x)\frac{h(x)}{g(x)}H(dx), \tag{2.3}$$

*where $\sigma_1^2(x)$ and $\sigma_2^2(x)$ are the conditional variances of $Y_{11}$ given $X_{11} = x$ and $Y_{21}$ given $X_{21} = x$, respectively. In the homoscedastic case $\sigma_1^2$ and $\sigma_2^2$ are constants.*

Note that for $c = 0$, i.e., under $H_0$, we have $\mu = 0$. Let $\hat{\sigma}$ be a consistent estimator of $\sigma$. Then, under $H_0$,

$$\frac{\sqrt{N}\hat{T}}{\hat{\sigma}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

It is easy to see that $\sigma$ is shift-invariant and scale-equivariant. Typically, $\hat{\sigma}$ also has the same properties. Conclude that $\hat{T}/\hat{\sigma}$ is scale and shift-invariant, as is the test to be discussed now. Let $0 < \alpha < 1$ be a given significance level and denote with $q_{1-\frac{\alpha}{2}}$ the $1-\frac{\alpha}{2}$ quantile of $\mathcal{N}(0,1)$. Then, by Theorem 2.2,

$$\mathbb{P}\left(\left|\frac{\sqrt{N}\hat{T}}{\hat{\sigma}}\right| \geq q_{1-\frac{\alpha}{2}}\right) \to \alpha.$$

Therefore, we reject $H_0$ if and only if $\left|\frac{\sqrt{N}\hat{T}}{\hat{\sigma}}\right| \geq q_{1-\frac{\alpha}{2}}$.

Next we discuss the local power of the test in connection with the choice of $W$. We want to test (2.1) with $c = 0$ versus $c \neq 0$. Under $H_1$,

$$\frac{\sqrt{N}\hat{T}}{\hat{\sigma}} \xrightarrow{\mathcal{L}} \mathcal{N}(\frac{\mu}{\sigma}, 1).$$

Hence the asymptotic power of $|\hat{T}|$ equals

$$\lim_{N\to\infty} \mathbb{P}_{H_1}\left(\left|\frac{\sqrt{N}\hat{T}}{\hat{\sigma}}\right| \geq q_{1-\frac{\alpha}{2}}\right) = \mathbb{P}(|\xi_0 + \frac{\mu}{\sigma}| \geq q_{1-\frac{\alpha}{2}}),$$

where $\xi_0 \sim \mathcal{N}(0,1)$. But

$$\mathbb{P}(|\xi_0 + \frac{\mu}{\sigma}| \geq q_{1-\frac{\alpha}{2}}) = 1 - \left[\Phi(\frac{\mu}{\sigma} + q_{1-\frac{\alpha}{2}}) - \Phi(\frac{\mu}{\sigma} - q_{1-\frac{\alpha}{2}})\right]. \tag{2.4}$$

This is a monotone increasing function of $|\frac{\mu}{\sigma}|$. Therefore it remains to find the $W$ which maximizes $\mu^2/\sigma^2$. Write (with $c = 1$)

$$\frac{\mu^2}{\sigma^2} = \frac{[\int W(x)s(x)H(dx)]^2}{\int a^2(x)W^2(x)H(dx)}. \tag{2.5}$$

with

$$a^2(x) = (1-\lambda)\sigma_1^2(x)\frac{h(x)}{f(x)} + \lambda\sigma_2^2(x)\frac{h(x)}{g(x)}.$$

It is easy to see that (2.5) is maximized for

$$W_0 = \frac{s}{a^2}. \tag{2.6}$$

In fact, for this choice of $W$, (2.5) becomes

$$\frac{\mu^2}{\sigma^2} = \int \frac{s^2}{a^2} dH. \tag{2.7}$$

The asymptotic (local) power (2.4) is determined through $\mu^2/\sigma^2$ as in (2.7). Whereas the function $s$ is given, the function $a^2$ incorporates terms (like $\sigma_1^2, \sigma_2^2, f, g$) which depend on the data and therefore cannot be controlled by the statistician.

Now, $\frac{\mu^2}{\sigma^2}$ and hence power becomes large when $a^2$ is small. This means, that the error of the second kind gets small with $a^2$. On the other hand, $a^2$ becomes small when $\sigma_1^2$ and $\sigma_2^2$ are small. This only expresses the fact that the risk for making an error of the second kind is smaller when $m_1(X_1)$ and $m_2(X_2)$ can be better reconstructed from $Y_1$ and $Y_2$ than in the other case, i.e., when $\sigma_1^2$ and $\sigma_2^2$ are big.

Typically, if the supports of $f$ and $g$ do not have much in common, the testing problem is more difficult since most of the information about the two samples is located in separate regions. By averaging $X_{1i}$ and $X_{2j}$ we obtain data-dependent points which are located between the two $X$-samples and therefore provide a reasonable area at which the two $m$'s may be compared. If the two supports, however, more or less coincide, the situation is less dramatic, since the $X_{1i}$ and $X_{2j}$ fall into the same area so that also the $(X_{1i} + X_{2j})/2$ are located here. In terms of power and hence of error of the second kind the first situation is therefore more difficult. We can easily see that when we look at $a^2$ more closely. Actually, in the first situation both $h/f$ and $h/g$ are typically large on the support of $H$ so that $\mu^2/\sigma^2$ is small and the error of the second kind becomes large. Alternatively, if $F$ and $G$ do not differ much, then the support of $(X_{1i} + X_{2j})/2$ coincides more or less with the support of $F$ and $G$. The functions $h, f$ and $g$ are of a similar order so that $a^2$ is moderately small, as is the error of the second kind.

Next we show how to estimate $\sigma^2$, say by $\hat{\sigma}^2$. Recall

$$\sigma^2 = (1-\lambda)\rho_1^2 + \lambda\rho_2^2,$$

(2.2) and (2.3). Replacing $H$ with $H_n$, the empirical d.f. of the $\frac{X_{1k}+X_{2l}}{2}$, Theorem 2.1 and the proof of Theorem 2.2 suggest the following estimator of $\sigma^2$:

$$\begin{aligned}\hat{\sigma}^2 \quad &:= \quad \frac{n_1 n_2}{n_1+n_2}\sum_{i=1}^{n_1}(Y_{1i}-\hat{m}_1(X_{1i}))^2\left[\int W(x)W_{1i}(x)H_n(dx)\right]^2 \\ &+ \quad \frac{n_1 n_2}{n_1+n_2}\sum_{j=1}^{n_2}(Y_{2j}-\hat{m}_2(X_{2j}))^2\left[\int W(x)W_{2j}(x)H_n(dx)\right]^2.\end{aligned}$$

Note that

$$\begin{aligned}\int W(x)W_{1i}(x)H_n(dx) \quad &= \quad \frac{1}{n_1 n_2}\sum_{k=1}^{n_1}\sum_{l=1}^{n_2} W\left(\frac{X_{1k}+X_{2l}}{2}\right)W_{1i}\left(\frac{X_{1k}+X_{2l}}{2}\right) \\ \int W(x)W_{2j}(x)H_n(dx) \quad &= \quad \frac{1}{n_1 n_2}\sum_{k=1}^{n_1}\sum_{l=1}^{n_2} W\left(\frac{X_{1k}+X_{2l}}{2}\right)W_{2j}\left(\frac{X_{1k}+X_{2l}}{2}\right).\end{aligned}$$

It is easy to see that $\hat{\sigma}$ is shift-invariant but scale-equivariant.

Next we discuss a more general alternative than (2.1), namely

$$m_2 = m_1 + N^{-1/2}\sum_{j=1}^{d}\gamma_j s_j. \tag{2.8}$$

For example, if $s_j(x) = x^{j-1}, 1 \le j \le d$, then (2.8) is tantamount to saying, that $m_1$ and $m_2$ differ by a polynomial of degree $d-1$. In particular, for $d=1$, $m_1$ and $m_2$ differ only by a constant and are therefore parallel. Other choices for $s_j$ are trigonometric polynomials of different frequencies or basic splines. In general, $s_1,\ldots,s_d$ are finitely many functions which may be appropriately chosen once the testing problem (i.e., the alternative model) is specified. The null model corresponds to $\gamma_1 = \ldots = \gamma_d = 0$.

In the following we shall derive Maximin tests for $H_0 : m_1 = m_2$ versus $\|\gamma\| \ge a > 0$, where $\gamma^t = (\gamma_1,\ldots,\gamma_d)$ is the vector of coefficients and $\|\cdot\|$ is a proper norm. Note that the model under $H_1$ is semiparametric since $m_1, m_2$ are not specified and the space spanned by $s_1,\ldots,s_d$ is parametric.

In view of what we found for $d=1$, i.e., model (2.1), we consider the vector of score-statistics (1.1) for $W_1,\ldots,W_d$ with

$$W_j = \frac{s_j}{a^2},$$

say

$$\hat{T} = (\hat{T}^1,\ldots,\hat{T}^d)^t.$$

Then Theorem 2.1 implies, under (2.8), with $s = \sum_{j=1}^d \gamma_j s_j$ and $c=1$, for each $1 \le j \le d$:

$$\begin{aligned}\sqrt{N}\hat{T}^j = \sqrt{1-\lambda}\; n_1^{1/2}&\sum_{i=1}^{n_1}(Y_{1i}-m_1(X_{1i}))\int W_j(x)W_{1i}(x)H(dx) \\ -\sqrt{\lambda}\, n_2^{1/2}&\sum_{i=1}^{n_2}(Y_{2i}-m_2(X_{2i}))\int W_j(x)W_{2i}(x)H(dx)\end{aligned}$$

$$- \int W_j(x) s(x) H(dx) + o_{\mathbb{P}}(1).$$

Therefore,

$$\begin{aligned}
\sqrt{N} \begin{pmatrix} \hat{T}^1 \\ \vdots \\ \hat{T}^d \end{pmatrix} &= \sqrt{1-\lambda} n_1^{1/2} \sum_{i=1}^{n_1} (Y_{1i} - m_1(X_{1i})) \begin{pmatrix} \int W_1(x) W_{1i}(x) H(dx) \\ \vdots \\ \int W_d(x) W_{1i}(x) H(dx) \end{pmatrix} \\
&- \sqrt{\lambda} n_2^{1/2} \sum_{i=1}^{n_2} (Y_{2i} - m_2(X_{2i})) \begin{pmatrix} \int W_1(x) W_{2i}(x) H(dx) \\ \vdots \\ \int W_d(x) W_{2i}(x) H(dx) \end{pmatrix} \\
&- \begin{pmatrix} \int W_1(x) s(x) H(dx) \\ \vdots \\ \int W_d(x) s(x) H(dx) \end{pmatrix} + o_{\mathbb{P}}(1).
\end{aligned}$$

From the multivariate version of Theorem 2.2 the first two sums converge in distribution to $\sqrt{1-\lambda}\ \mathcal{N}_d(\mathbf{0}, \Sigma_1)$ and $\sqrt{\lambda}\, \mathcal{N}_d(\mathbf{0}, \Sigma_2)$, respectively, where $\Sigma_1 = (\rho_{ij}^1)$ and $\Sigma_2 = (\rho_{ij}^2)$, with

$$\begin{aligned}
\rho_{ij}^1 &= \int \sigma_1^2(x) \frac{h^2(x)}{f^2(x)} W_i(x) W_j(x) F(dx) \\
&= \int \sigma_1^2(x) \frac{h(x)}{f(x)} W_i(x) W_j(x) H(dx) \\
\rho_{ij}^2 &= \int \sigma_2^2(x) \frac{h^2(x)}{g^2(x)} W_i(x) W_j(x) G(dx) \\
&= \int \sigma_2^2(x) \frac{h(x)}{g(x)} W_i(x) W_j(x) H(dx).
\end{aligned}$$

By the independence of the two samples we get

$$\sqrt{N}\hat{T} = \sqrt{N} \begin{pmatrix} \hat{T}^1 \\ \vdots \\ \hat{T}^d \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_d(\mathbf{0}, \Sigma) - \begin{pmatrix} \int W_1 s dH \\ \vdots \\ \int W_d s dH \end{pmatrix}$$

with $\Sigma = (\rho_{ij})$ and

$$\rho_{ij} = (1-\lambda)\rho_{ij}^1 + \lambda \rho_{ij}^2 = \int a^2 W_i W_j dH.$$

Write

$$\begin{pmatrix} \int W_1 s dH \\ \vdots \\ \int W_d s dH \end{pmatrix} = \begin{pmatrix} \int W_1 s_1 dH & \ldots & \int W_1 s_d dH \\ \vdots & & \vdots \\ \int W_d s_1 dH & \ldots & \int W_d s_d dH \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_d \end{pmatrix}$$

and notice that for $W_j = \frac{s_j}{a^2}$, $1 \le j \le d$, we have

$$\int W_i s_j dH = \int a^2 W_i W_j dH = \rho_{ij}.$$

In summary, we have

$$\sqrt{N}\hat{T} \xrightarrow{\mathcal{L}} \mathcal{N}_d(\mathbf{0}, \Sigma) - \Sigma \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_d \end{pmatrix} \tag{2.9}$$

Under the null model:

$$\sqrt{N}\hat{T} \xrightarrow{\mathcal{L}} \mathcal{N}_d(\mathbf{0}, \Sigma). \tag{2.10}$$

These results allow us to apply some existing Maximin Theory. See Strasser (1985) for details. Namely, for a given significance level $0 < \alpha < 1$, let $c_{1-\alpha}$ be the $1-\alpha$ quantile of the $\chi^2_d$-distribution. Put

$$t = 1_{\{N\hat{T}^t\Sigma_n^{-1}\hat{T} \geq c_{1-\alpha}\}}. \tag{2.11}$$

**Theorem 2.3** *For a given significance level $0 < \alpha < 1$, the test $t$ from (2.11) is a Maximin test for $H_0 : m_1 = m_2$ versus (2.8) with $H_1 : \gamma^t\Sigma\gamma \geq a$. The asymptotic Maximin power is given by $\mathbb{P}(\chi^2_d(a) \geq c_{1-\alpha})$, where now $a$ is the noncentrality parameter.*

*Proof.* That $t$ is an asymptotic level $\alpha$-test follows from (2.10):

$$\begin{aligned} \mathbb{P}_{H_0}(t = 1) &= \mathbb{P}_{H_0}(N\hat{T}^t\Sigma_n^{-1}\hat{T} \geq c_{1-\alpha}) \\ &\to \mathbb{P}(\xi^t\Sigma^{-1}\xi \geq c_{1-\alpha}), \end{aligned}$$

where $\xi \sim \mathcal{N}_d(\mathbf{0}, \Sigma)$. Write $\xi = A\xi_0$, where $\xi_0 \sim \mathcal{N}_d(\mathbf{0}, I_d)$ and $A$ satisfies $\Sigma = AA^t$. Conclude that

$$\mathbb{P}(\xi^t\Sigma^{-1}\xi \geq c_{1-\alpha}) = \mathbb{P}(\xi_0^t\xi_0 \geq c_{1-\alpha}) = \alpha.$$

Under the local alternative (2.8), we obtain from (2.9) with $\gamma^t = (\gamma_1, \ldots, \gamma_d)$:

$$\begin{aligned} \mathbb{P}_{H_1}(t = 1) &\to \mathbb{P}((\xi - \Sigma\gamma)^t\Sigma^{-1}(\xi - \Sigma\gamma) \geq c_{1-\alpha}) \\ &= \mathbb{P}((\xi_0 - A^t\gamma)^t(\xi_0 - A^t\gamma) \geq c_{1-\alpha}). \end{aligned}$$

Note that $(\xi_0 - A^t\gamma)^t(\xi_0 - A^t\gamma)$ has a $\chi^2_d(a)$-distribution with noncentrality parameter

$$a \equiv \|A^t\gamma\|^2 = \gamma^tAA^t\gamma = \gamma^t\Sigma\gamma.$$

□

The test $t$ is asymptotically distribution-free under $H_0$. In our simulation studies we considered optimal and suboptimal $W$'s. Suboptimal $W$'s need to be considered when the $s_j$'s are not specified. In such a situation we propose for the $W_j$'s a collection of polynomials and trigonometric polynomials. Also some basic splines may be added.

Our final result deals with the case of fixed alternatives. It shows that our test is consistent.

**Theorem 2.4** *Under $H_1 : m_1 \neq m_2$ fixed we have, when $\int W(m_1 - m_2)dH \neq 0$, that $\mathbb{P}(t = 1) \to 1$*

All proofs are postponed to Chapter 4.

## 3 Simulation Study

In this chapter we empirically investigate how our tests perform in finite samples. As in the previous chapters let $m_1$ be the unknown regression function for the first sample. For $m_2$ we assume

$$m_2 = m_1 + N^{-1/2} cs.$$

Here the function $s$ determines the direction in which $m_2$ deviates from $m_1$, while the scalar $c$ is in charge of the amount of deviation. Clearly, $c = 0$ is equivalent to the validity of $H_0$.

We already indicated before that the power of our tests will be influenced by the design distributions $F$ and $G$. It is to be expected that if $F = G$ it may be easier to detect differences between $m_1$ and $m_2$ than in a situation when $F \neq G$. At the same time, since

$$Y_1 = m_1(X_1) + \varepsilon_1 \text{ and } Y_2 = m_2(X_2) + \varepsilon_2,$$

the noise variables $\varepsilon_1$ and $\varepsilon_2$ will also have an impact on the power. For example, the situation may deteriorate when the variances $\sigma_1^2 = \text{Var}(\varepsilon_1)$ and $\sigma_2^2 = \text{Var}(\varepsilon_2)$ increase so that information on $m$ contained in the $Y$'s may be heavily blurred. Needless to say, the power of the test will depend on $c$. When $c = 0$, we expect that the empirical level of the test, i.e., the percentage of times we reject $H_0$ though it is true, is close to the nominal level. Another important feature is the choice of the weight function. It will be interesting to see how the power decreases if rather than optimal weights, we take suboptimal $W$'s. This question is important because we may be interested in the test also with respect to deviations other than into direction $s$.

Simulations were implemented in S-PLUS, Version 6.0 Release 1, of the Data Analysis Products Division of MathSoft Inc., Seattle/Washington, USA, and performed on Sun SPARC stations under Sun OS 5.9.

In each of the simulations the errors $\varepsilon_1$ and $\varepsilon_2$ were independent of $X_1$ and $X_2$, respectively, with $\varepsilon_1 \sim \mathcal{N}(0, 1)$ and $\varepsilon_2 \sim \mathcal{N}(0, 1.5^2)$. The number of replications of each Monte Carlo experiment was $M = 500$. The nominal level always equals $\alpha = 0.05$. For K we took the standard Gaussian kernel, while for $\beta$ in $(h)$ we set $\beta = \frac{7}{24}$, but also other $h$'s were considered to demonstrate the influence of the bandwidth. For $m_1$ we considered the two cases

$$m_1(x) = 1 + 2x \qquad \text{(affine case)} \tag{3.1}$$

and

$$m_1(x) = 1 + 2x + \frac{1}{2}x^2 \qquad \text{(quadratic case).} \tag{3.2}$$

For $s$, i.e., for the alternative models we studied three different examples: $s_1(x) = 9$ (constant shift), $s_2(x) = 9 + 2x$ (affine shift), $s_3(x) = 9 + 2x - \frac{1}{2}x^2$ (quadratic shift).

In the tables to follow we report on the empirical level of the tests under $H_0$ and their power under $H_1$, for various choices of $c$. The reported results are part of a much larger study which because of lack of space cannot be discussed in detail.

In Table 1 below $X_1$ and $X_2$ will have the same standard normal distribution: $F = \mathcal{N}(0,1) = G$. The function $m_1$ equals (3.1) while $h = 0.2$. Under $H_1$, we always put $c = 1$ so that

$$m_2(x) = m_1(x) + \frac{1}{\sqrt{N}} s_i(x) \text{ for } 1 \le i \le 3.$$

Under $H_0$, $c = 0$. Empirical levels and power are always boldface.

Table 1: Percentages of Rejection

| | | $s_1$ | $s_2$ | $s_3$ |
|---|---|---|---|---|
| $n_1 = 25$ | $n_2 = 30$ | $c = 0$ : **0.05** | $c = 0$ : **0.05** | $c = 0$ : **0.05** |
| | | $c = 1$ : **1** | $c = 1$ : **1** | $c = 1$ : **1** |
| $n_1 = 50$ | $n_2 = 60$ | $c = 0$ : **0.06** | $c = 0$ : **0.06** | $c = 0$ : **0.06** |
| | | $c = 1$ : **1** | $c = 1$ : **1** | $c = 1$ : **1** |
| $n_1 = 100$ | $n_2 = 120$ | $c = 0$ : **0.07** | $c = 0$ : **0.07** | $c = 0$ : **0.07** |
| | | $c = 1$ : **1** | $c = 1$ : **1** | $c = 1$ : **1** |

We see that the power is always one! The attained level equals $\alpha$ in the first row but is slightly larger when $n_1$ and $n_2$ increase. This is due to the fact that we kept $h = 0.2$ fixed but changed the $n$'s. Later we shall also adjust $h$.

We mentioned several times that in the random design case the distributions $F$ and $G$ may have an impact on the power of the test. Therefore, in the following, we study situations when $F$ and $G$ vary in several aspects. In the first case, $F$ and $G$ will be again normal with equal means but unequal variances:

$$F = \mathcal{N}(0,1) \qquad \text{and} \qquad G = \mathcal{N}(0,4).$$

The function $m_1$ equals (3.1) while $s_1, s_2$ and $s_3$ are the same as before. Moreover, we have varied $h$ to learn more about the influence of the bandwidth.

Table 2: Percentages of Rejection ($n_1 = 25, n_2 = 30$)

| | $c = 0$ | $c = 1$ |
|---|---|---|
| $h = 0.05$ | $s_1$ : **0.24** | $s_1$ : **1** |
| | $s_2$ : **0.15** | $s_2$ : **1** |
| | $s_3$ : **0.14** | $s_3$ : **1** |
| $h = 0.1$ | $s_1$ : **0.09** | $s_1$ : **1** |
| | $s_2$ : **0.09** | $s_2$ : **1** |
| | $s_3$ : **0.08** | $s_3$ : **1** |
| $h = 0.2$ | $s_1$ : **0.07** | $s_1$ : **1** |
| | $s_2$ : **0.06** | $s_2$ : **1** |
| | $s_3$ : **0.06** | $s_3$ : **1** |

Again the power is one. We also see that the bandwidth $h = 0.05$ leads to a level which is too large compared with $\alpha$. This is also covered by our theoretical results which require $h$ to be of the order $N^{-\beta}, \frac{1}{4} < \beta < \frac{1}{3}$. Note that in our case $N = 13.6$.

We also simulated a model where $F$ and $G$ differ in mean but are equal in variance: $F = \mathcal{N}(0,1)$ and $G = \mathcal{N}(1,1)$. The results were more or less the same. In each case the power was one, and the nominal level was attained for $h = 0.2$. Finally, we studied the case when $F$ and $G$ were as above but $m_1$ was quadratic. Again, the power was one, but for small sample size the attained level was between 0.10 and 0.15. This relatively bad performance under $H_0$ is due to the fact that for these $F$ and $G$ and quadratic $m$'s, the curvature of $m$ is responsible for a less satisfactory fit of $m_1$ on the right side of the center of $G$, namely 1. Hence the quality of the normal approximation suffers from some bias for very small sample size.

The fact that in the preceding examples the power was always one though the alternative model had local character indicates that $c = 1$ was still too big to obtain less than 100% rejections. In the following we therefore study the power for decreasing $c$'s. The parameters were as in Table 1: $F = \mathcal{N}(0,1) = G$ with $\sigma_1 = 1$, $\sigma_2 = 1.5$, $M = 500$, $\alpha = 0.05$, while sample sizes were $n_1 = 50$ and $n_2 = 60$. The function $m_1$ was again $m_1(x) = 1 + 2x$, while $h$ was set $h = 0.10$.

Table 3: Percentages of Rejection

| | $s_1$ | $s_2$ | $s_3$ |
|---|---|---|---|
| $c = 0.5$ | **0.94** | **0.950** | **0.936** |
| $c = 0.1$ | **0.156** | **0.132** | **0.137** |
| $c = 0.05$ | **0.094** | **0.090** | **0.086** |

Similar results were obtained when $m_1$ was quadratic. Even for $c = 0.5$ the power is excellent. Since in our case $N = 27.3$ the case $c = 0.1$ belongs to the alternative $m_2(x) = m_1(x) + 0.02s(x)$, which is very close to $m_1$ on the support of $F$ and $G$, so that the low power is not surprising.

It is interesting to compare loss in power when rather than $W_0$ we choose $W \equiv 1$ not depending on $s$ and the function $a$. According to Table 3, when $s = s_1$ and $c = 0.5$ we get power 0.94 while with the suboptimal $W$ we obtain power 0.924 under $h = 0.1$ and power 0.908 under $h = 0.05$. Hence, in this situation the loss is moderate so that one may say that our tests are robust in neighborhoods of the assumed model.

Table 4: Percentages of Rejection

| $s = s_1$ | $W \equiv 1$ | | | | Power |
|---|---|---|---|---|---|
| $c = 0.50$ | $h = 0.10$ | $n_1 = 25$ | $n_2 = 30$ | $N = 13.6$ | **0.9** |
| | | $n_1 = 50$ | $n_2 = 60$ | $N = 27.3$ | **0.924** |
| | | $n_1 = 100$ | $n_2 = 120$ | $N = 54.5$ | **0.926** |
| | $h = 0.05$ | $n_1 = 25$ | $n_2 = 30$ | $N = 13.6$ | **0.858** |
| | | $n_1 = 50$ | $n_2 = 60$ | $N = 27.3$ | **0.908** |
| | | $n_1 = 100$ | $n_2 = 120$ | $N = 54.5$ | **0.914** |

We end our simulation studies with a comparison of the results obtained by Neumeyer and Dette (2003). In their simulation study they considered, in our notation, only the case when $F$ and $G$ were the same and equal the uniform distribution on $[0, 1]$. Moreover, $\alpha = 0.05, n_1 = 25, n_2 = 50, \sigma_1^2 = \frac{1}{2}$ and $\sigma_2^2 = \frac{1}{4}$. Then the following nine situations were considered:

(i) $m_1 = m_2 = 1$

(ii) $m_1(x) = m_2(x) = \exp x$

(iii) $m_1(x) = m_2(x) = \sin(2\pi x)$

(iv) $m_1(x) = 1, \quad m_2(x) = m_1(x) + x$

(v) $m_1(x) = \exp x, \quad m_2(x) = m_1(x) + x$

(vi) $m_1(x) = \sin(2\pi x), \quad m_2(x) = m_1(x) + x$

(vii) $m_1(x) = 1, \quad m_2(x) = m_1(x) + \sin(2\pi x)$

(viii) $m_1(x) = \exp x, \quad m_2(x) = m_1(x) + \sin(2\pi x)$

(ix) $m_1(x) = \sin(2\pi x), \quad m_2(x) = 2\sin(2\pi x)$

In the following tables we compare the attained levels and the power of our score tests (ST) with those of the Neumeyer-Dette (ND) tests. The results for ND were taken from Neumeyer and Dette (2003), Table 2. For ST we took $h = 0.2$ as in Table 1.

Table 5: Percentages of Rejection

| Model | ND | ST |
|---|---|---|
| (i) | **0.05** | **0.06** (**0.07**) |
| (ii) | **0.06** | **0.07** (**0.06**) |
| (iii) | **0.06** | **0.07** (**0.07**) |
| (iv) | **0.78** | **0.87** |
| (v) | **0.80** | **0.90** |
| (vi) | **0.70** | **0.85** |
| (vii) | **0.18** | **0.99** |
| (viii) | **0.20** | **0.99** |
| (ix) | **0.13** | **0.99** |

The first three situations deal with the null model. We applied our test twice, with the $W$'s associated with $s(x) = x$ and $s(x) = \sin(2\pi x)$. The attained levels are almost identical. The power of the ST-test is excellent and clearly outperforms the ND-test. As Scheike's (2000) test also the ND-test is unable to detect differences in $m_1$ and $m_2$ when they cross each other.

To get a visual impression we plot optimal $W$'s for several selected situations. In each case the variances of the noise variables were as before, namely $\sigma_1^2 = 1$ and $\sigma_2^2 = 1.5^2$. For our $n_1$ and $n_2$, $\lambda$ always equals $\frac{5}{11}$. From (2.6) and the definition of the function $a$ we see that $W_0$ depends on $s$ and the design distributions $F$ and $G$.

In the following, when we talk about down- and upweighting through the function $W$, one should have in mind that the test statistic in (2.11) is invariant w.r.t scaling of $W$, i.e., the test statistic based on $W$ is the same as that for $bW$, where $b \neq 0$ is a non-vanishing constant. Hence the shape of $W$ only provides a comparison of weights in different areas relative to each other. For example, a large value of $W$ in the tails just means that $W$ upweights the influence of the extreme $(X_{1i} + X_{2j})/2$ relative to the central ones.

Figure 1 depicts the optimal $W_0$ when $F = \mathcal{N}(0,1) = G$ and $s \equiv 9$ is a constant. Hence we are testing for $H_0$ versus parallel but different $m$'s. It turns out that $W_0$ downweights $\hat{m}_1 - \hat{m}_2$ at pairs $(X_{1i} + X_{2j})/2$ which are between -1.5 and 1.5 but upweights those which are in the left or right tails. $W_0$ is almost symmetric. The slight asymmetry is caused by the fact that $\lambda \neq \frac{1}{2}$ and $\sigma_1 \neq \sigma_2$. The increase of $W_0$ in the tails comes from the fact that though $f$ and $g$ are small there, the function $h$ being the Gaussian density pertaining to $\mathcal{N}(0, \frac{1}{2})$, is more concentrated at zero and therefore has even less mass in the tails. Informally speaking, since we may expect less $(X_{1i} + X_{2j})/2$ in the tails, it is up to the function $W$ to generate a microscoping effect there for checking possible deviations. Similar but appropriately modified comments also apply to the other cases to be discussed now.

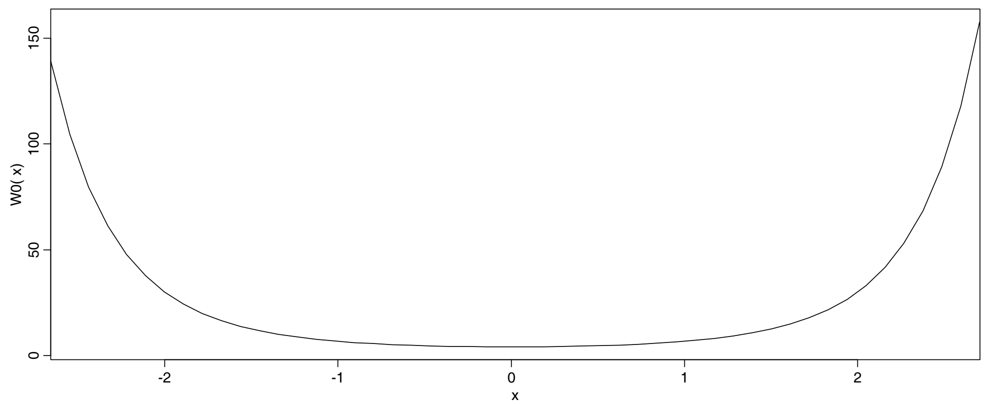

Figure 1: $W_0$ for $F = \mathcal{N}(0,1) = G$ and $s = 9$

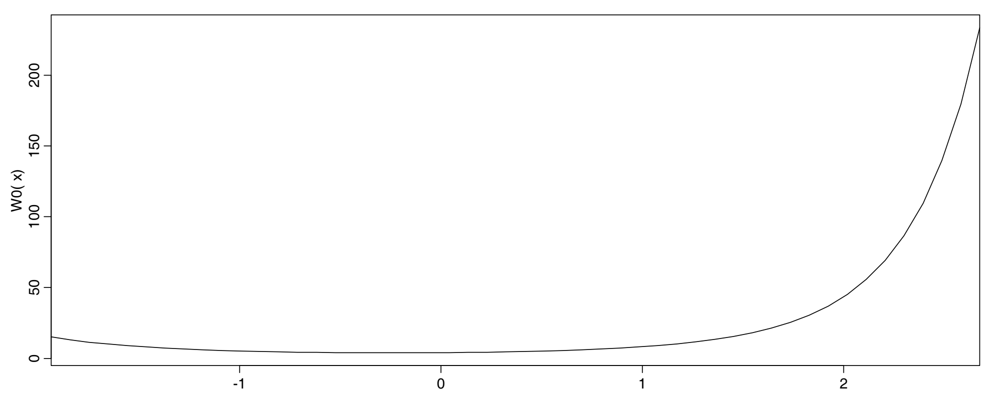

Figure 2: $W_0$ for $F = \mathcal{N}(0,1) = G$ and $s(x) = 9 + 2x$

In Figure 2 $W_0$ is heavily asymmetric. The test statistic upweights differences of $\hat{m}_1$ and $\hat{m}_2$ evaluated at large points. Under $H_1$, large differences may be expected particularly there so that $W_0$ takes special care of this area. Figure 3 belongs to $s \equiv 9$. $F$ and $G$ have equal means but differ in their variances. The resulting $W_0$ has two modes near -2 and +2. Slight asymmetries are again caused by $\lambda \neq \frac{1}{2}$.

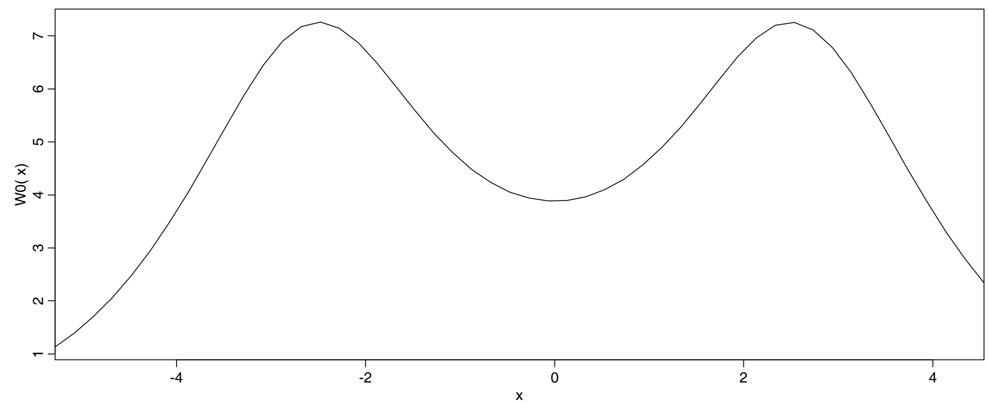


Figure 3: $W_0$ for $F = \mathcal{N}(0,1), G = \mathcal{N}(0,4)$ and $s(x) \equiv 9$

# 4 Proofs

In this section we first derive the martingale representation of $\hat{T}$ as formulated in Theorem 2.1 For this, write

$$\hat{T} = \int\int W\left(\frac{x_1+x_2}{2}\right)\left[\hat{m}_1\left(\frac{x_1+x_2}{2}\right) - \hat{m}_2\left(\frac{x_1+x_2}{2}\right)\right]\hat{F}(dx_1)\hat{G}(dx_2). \tag{4.1}$$

Introducing

$$x = \frac{x_1+x_2}{2}$$

we may expand $\hat{T}$ into

$$\begin{aligned}
\hat{T} &= \int\int W(x)[\hat{m}_1(x) - \hat{m}_2(x)][\hat{F}(dx_1) - F(dx_1)][\hat{G}(dx_2) - G(dx_2)] && (4.2)\\
&+ \int\int W(x)[\hat{m}_1(x) - m_1(x)]F(dx_1)[\hat{G}(dx_2) - G(dx_2)] && (4.3)\\
&- \int\int W(x)[\hat{m}_2(x) - m_2(x)]F(dx_1)[\hat{G}(dx_2) - G(dx_2)] && (4.4)\\
&+ \int\int W(x)[\hat{m}_1(x) - m_1(x)][\hat{F}(dx_1) - F(dx_1)]G(dx_2) && (4.5)\\
&- \int\int W(x)[\hat{m}_2(x) - m_2(x)][\hat{F}(dx_1) - F(dx_1)]G(dx_2) && (4.6)\\
&+ \int\int W(x)[\hat{m}_1(x) - m_1(x)]F(dx_1)G(dx_2) && (4.7)\\
&- \int\int W(x)[\hat{m}_2(x) - m_2(x)]F(dx_1)G(dx_2) && (4.8)\\
&+ \int\int W(x)[m_1(x) - m_2(x)][\hat{F}(dx_1) - F(dx_1)]G(dx_2) && (4.9)
\end{aligned}$$

$$+ \quad \int\int W(x)[m_1(x) - m_2(x)]F(dx_1)[\hat{G}(dx_2) - G(dx_2)] \tag{4.10}$$

$$+ \quad \int\int W(x)[m_1(x) - m_2(x)]F(dx_1)G(dx_2). \tag{4.11}$$

In our first lemma we derive some useful bounds for our NN-weights.

**Lemma 4.1** *Let $X_1, \ldots, X_n$ be a sample of independent random variables from a continuous d.f. $F$ with empirical d.f. $\hat{F}$. Assume (K). Then we have, for all $x \in \mathbb{R}$ and $h > 0$,*

$$\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\hat{F}(X_i) - \hat{F}(x)}{h}\right) \leq 1 + \frac{K(0)}{nh} \tag{4.12}$$

$$\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\hat{F}(X_i) - \hat{F}(x)}{h}\right) \geq \int_{-\hat{F}(x)/h}^{0} K(z)dz + \int_{1/nh}^{\frac{1-\hat{F}(x)}{h}+\frac{1}{nh}} K(z)dz. \tag{4.13}$$

**Proof.** By continuity of $F$, the sample contains, with probability one, no ties. Hence

$$\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\hat{F}(X_i) - \hat{F}(x)}{h}\right) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\frac{i}{n} - \hat{F}(x)}{h}\right).$$

Since $K$ is nondecreasing on $(-\infty, 0]$ the sum over $1 \leq i \leq n\hat{F}(x) - 1$ is bounded from above by $\frac{1}{h}\int_{-\infty}^{\hat{F}(x)} K\left(\frac{y-\hat{F}(x)}{h}\right)dy$, while the sum over $n\hat{F}(x) + 1 \leq i \leq n$ is bounded by $\frac{1}{h}\int_{\hat{F}(x)}^{\infty} K\left(\frac{y-\hat{F}(x)}{h}\right)dy$. Hence (4.12) follows immediately from $\int K(z)dz = 1$. With a similar argument we obtain (4.13). □

Note that the upper bound in (4.12) does not depend on $x$ and $\hat{F}$. In particular, the upper bound tends to one uniformly in $x$ as $n \to \infty$, in view of $nh \to \infty$. The lower bound, however, does depend on $\hat{F}(x)$. To obtain a pointwise limit, fix $x$ such $0 < F(x) < 1$. Since $\hat{F}(x) \to F(x)$ by the Strong Law of Large Numbers (SLLN) the right hand side of (4.13) tends to 1. Together with (4.12) we therefore get

**Corollary 4.2** *For each $x$ with $0 < F(x) < 1$ with probability one*

$$\lim_{n\to\infty} \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\hat{F}(X_i) - \hat{F}(x)}{h}\right) = 1.$$

As to a uniform lower bound, we have to consider two cases separately:

- If $\hat{F}(x) \geq 1/2$, then the right hand side of (4.13) exceeds

$$\int_{-\hat{F}(x)/h}^{0} K(z)dz \geq \int_{-1/2h}^{0} K(z)dz.$$

- If $\hat{F}(x) < 1/2$, then the right hand side of (4.13) exceeds

$$\int_{1/nh}^{\frac{1-\hat{F}(x)}{h}+\frac{1}{nh}} K(z)dz \geq \int_{1/nh}^{1/2h+1/nh} K(z)dz.$$

Now, since $h \to 0$, $K$ has compact support with integral 1 and is symmetric at zero,

$$\int_{-1/2h}^{0} K(z)dz = 1/2 \text{ for all small enough } h > 0.$$

On the other hand

$$\int_{1/nh}^{1/2h+1/nh} K(z)dz = \int_{1/nh}^{\infty} K(z)dz \geq \frac{1}{2} - \varepsilon,$$

for all small enough $h > 0$, where $\varepsilon > 0$ is an arbitrary number.

**Corollary 4.3** *For all small enough $h > 0$, we have uniformly in $x$ and for all samples*

$$\frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{\hat{F}(X_i) - \hat{F}(x)}{h}\right) \geq \frac{1}{2} - \varepsilon.$$

For some purposes the uniform lower bound in Corollary 4.3 is not sufficient. Rather we need an analogue of (4.12), namely

$$\frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{\hat{F}(X_i) - \hat{F}(x)}{h}\right) \geq 1 - \frac{K(0)}{nh}, \tag{4.14}$$

which is valid at least for most of the $x$'s. For this, note that for

$$h \leq \hat{F}(x) \leq 1 - h \tag{4.15}$$

the right-hand side of (4.13) becomes

$$\int_{-\infty}^{0} K(z)dz + \int_{1/nh}^{\infty} K(z)dz = 1 - \int_{0}^{1/nh} K(z)dz \geq 1 - \frac{K(0)}{nh},$$

as desired.

From now on we assume without further mentioning that $K$ is supported by $[-1, 1]$. In the following lemma we are going to bound the first integral (4.2) in the expansion of $\hat{T}$. This bound will enable us to show that (4.2) is asymptotically negligible. Recall $x = \frac{x_1+x_2}{2}$.

**Lemma 4.4** *Under the assumptions of Theorem 2.1 we have*

$$\mathbb{E}\left[\int\int W(x)[\hat{m}_1(x)-\hat{m}_2(x)][\hat{F}(dx_1)-F(dx_1)][\hat{G}(dx_2)-G(dx_2)]\right]^2 = O\left(\frac{1}{n_1n_2h^2}\right). \tag{4.16}$$

**Proof.** Introduce $\varphi_1 = W\hat{m}_1$ and $\varphi_2 = W\hat{m}_2$. To show (4.16) it suffices to prove

$$\mathbb{E}\left[\int\int \varphi(x)[\hat{F}(dx_1)-F(dx_1)][\hat{G}(dx_2)-G(dx_2)]\right]^2 = O\left(\frac{1}{n_1n_2h^2}\right)$$

for $\varphi = \varphi_1, \varphi_2$. Now,

$$\int\int \varphi(x)[\hat{F}(dx_1)-F(dx_1)][\hat{G}(dx_2)-G(dx_2)] = \int\int \varphi^*(x)\hat{F}(dx_1)\hat{G}(dx_2), \tag{4.17}$$

where

$$\begin{aligned}\varphi^*\left(\frac{x_1+x_2}{2}\right) &= \varphi\left(\frac{x_1+x_2}{2}\right) - \int\varphi\left(\frac{x_1+v}{2}\right)G(dv) - \int\varphi\left(\frac{u+x_2}{2}\right)F(du)\\ &\quad + \int\int\varphi\left(\frac{u+v}{2}\right)F(du)G(dv).\end{aligned}$$

But

$$\begin{aligned}&\mathbb{E}\left[\int\int \varphi^*(x)\hat{F}(dx_1)\hat{G}(dx_2)\right]^2\\ &= \frac{1}{n_1^2n_2^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_1}\sum_{l=1}^{n_2}\mathbb{E}\left[\varphi^*\left(\frac{X_{1i}+X_{2j}}{2}\right)\varphi^*\left(\frac{X_{1k}+X_{2l}}{2}\right)\right]\\ &= \frac{(n_1-1)(n_2-1)}{n_1n_2}\mathbb{E}\left[\varphi^*\left(\frac{X_{11}+X_{21}}{2}\right)\varphi^*\left(\frac{X_{12}+X_{22}}{2}\right)\right] \qquad (4.18)\\ &+ \frac{(n_1-1)}{n_1n_2}\mathbb{E}\left[\varphi^*\left(\frac{X_{11}+X_{21}}{2}\right)\varphi^*\left(\frac{X_{12}+X_{21}}{2}\right)\right] \qquad (4.19)\\ &+ \frac{(n_2-1)}{n_1n_2}\mathbb{E}\left[\varphi^*\left(\frac{X_{11}+X_{21}}{2}\right)\varphi^*\left(\frac{X_{11}+X_{22}}{2}\right)\right] \qquad (4.20)\\ &+ \frac{1}{n_1n_2}\mathbb{E}\left[\varphi^*\left(\frac{X_{11}+X_{21}}{2}\right)\right]^2. \qquad (4.21)\end{aligned}$$

We now show that (4.18) and (4.20) vanish for $\varphi_1$. We only deal with (4.18) since the other case is similar. Now, use conditional expectations together with the independence of $(X_{11}, Y_{11}), \ldots, (X_{1n_1}, Y_{1n_1})$, $X_{21}$ and $X_{22}$ to get

$$\begin{aligned}&\mathbb{E}\left[\varphi^*\left(\frac{X_{11}+X_{21}}{2}\right)\varphi^*\left(\frac{X_{12}+X_{22}}{2}\right)\right]\\ &= \mathbb{E}\left\{\mathbb{E}\left[\varphi^*\left(\frac{X_{11}+X_{21}}{2}\right)\varphi^*\left(\frac{X_{12}+X_{22}}{2}\right)\middle| X_{11}, Y_{11}, \ldots, X_{1n_1}, Y_{1n_1}\right]\right\}\\ &= \mathbb{E}\left\{\mathbb{E}\left[\varphi^*\left(\frac{X_{11}+X_{21}}{2}\right)\middle|\ldots\right]\mathbb{E}\left[\varphi^*\left(\frac{X_{12}+X_{22}}{2}\right)\middle|\ldots\right]\right\}.\end{aligned}$$

The first inner conditional expectation equals, however,

$$\begin{aligned}
&\mathbb{E}\left[\varphi^*\left(\frac{X_{11}+X_{21}}{2}\right)\middle|\ldots\right]\\
=&\ \int\varphi\left(\frac{X_{11}+v}{2}\right)G(dv)-\int\varphi\left(\frac{X_{11}+v}{2}\right)G(dv)\\
-&\ \int\int\varphi\left(\frac{u+v}{2}\right)F(du)G(dv)+\int\int\varphi\left(\frac{u+v}{2}\right)F(du)G(dv)=0.
\end{aligned}$$

This proves that (4.18) vanishes. To bound (4.19) and (4.21), we first consider the expectations for $\bar{\bar{\varphi}}^*$ rather than $\varphi^*$, where $\bar{\bar{\varphi}}^*$ equals $\varphi^*$, but with $(X_{1i}, Y_{1i}), 1 \le i \le 2$, deleted from the first sample. Since $\bar{\bar{\varphi}}^*$ is independent of $(X_{ji}, Y_{ji})$, $1 \le i, j \le 2$, we obtain similarly to before:

$$\mathbb{E}\left[\bar{\bar{\varphi}}^*\left(\frac{X_{11}+X_{21}}{2}\right)\bar{\bar{\varphi}}^*\left(\frac{X_{12}+X_{21}}{2}\right)\right]=0.$$

We now bound (4.21) for $\bar{\bar{\varphi}}^*$. For this, write $X = \frac{1}{2}(X_{11}+X_{21})$ for short and note that, after conditioning, by the Cauchy-Schwarz and the triangle inequality for second moments,

$$\begin{aligned}
\mathbb{E}[\bar{\bar{\varphi}}^*(X)]^2 &\le 16\mathbb{E}[\bar{\bar{\varphi}}(X)]^2\\
&= 16\int\int\mathbb{E}\left[\bar{\bar{\varphi}}^2\left(\frac{u+v}{2}\right)\right]F(du)G(dv). \qquad (4.22)
\end{aligned}$$

To bound the inner expectation note that with $n = n_1 - 2$

$$\begin{aligned}
\mathbb{E}\left[\bar{\bar{\varphi}}^2\left(\frac{u+v}{2}\right)\right] &= W^2\left(\frac{u+v}{2}\right)\mathbb{E}\left[\sum_{i=1}^{n}Y_{1i}W_{1i}\left(\frac{u+v}{2}\right)\right]^2\\
&\le W^2\left(\frac{u+v}{2}\right)n^2\mathbb{E}\left[Y_{11}W_{11}\left(\frac{u+v}{2}\right)\right]^2.
\end{aligned}$$

Since

$$W_{11}\left(\frac{u+v}{2}\right)=\frac{K\left(\frac{\hat{F}(X_{11})-\hat{F}\left(\frac{u+v}{2}\right)}{h}\right)}{\sum_{i=1}^{n}K\left(\frac{\hat{F}(X_{1i})-\hat{F}\left(\frac{u+v}{2}\right)}{h}\right)}$$

we obtain from Corollary 4.3 and the boundedness of $K$ that

$$\begin{aligned}
\mathbb{E}[\bar{\bar{\varphi}}^*(X)]^2 &\le h^{-2}C\int\int W^2\left(\frac{u+v}{2}\right)F(du)G(dv)\\
&= O(h^{-2}), \qquad (4.23)
\end{aligned}$$

where $C$ is a constant which may depend on $K$.

Now we bound (4.19) and (4.21) in absolute values from above for the original $\varphi^*$. As to (4.19), ignoring the $X$'s for a moment, we have

$$\begin{aligned}
\mathbb{E}[\varphi^*\varphi^*] &= \mathbb{E}[(\varphi^*-\bar{\bar{\varphi}}^*)(\varphi^*-\bar{\bar{\varphi}}^*)]\\
&+ \mathbb{E}[(\varphi^*-\bar{\bar{\varphi}}^*)\bar{\bar{\varphi}}^*]+\mathbb{E}[\bar{\bar{\varphi}}^*(\varphi^*-\bar{\bar{\varphi}}^*)]+\mathbb{E}\bar{\bar{\varphi}}^*\bar{\bar{\varphi}}^*,
\end{aligned}$$

where the last expectation is already known to vanish. As to the others, we may apply the Cauchy-Schwarz inequality to get

$$|\mathbb{E}[\varphi^*\varphi^*]| \leq \mathbb{E}(\varphi^* - \bar{\bar{\varphi}}^*)^2 + 2\sqrt{\mathbb{E}(\varphi^* - \bar{\bar{\varphi}}^*)^2}O(h^{-1}), \tag{4.24}$$

by (4.23). For (4.21), use the inequality

$$(a+b)^2 \leq 2(a^2+b^2)$$

to obtain, by (4.23)

$$\begin{aligned} \mathbb{E}[\varphi^*]^2 &\leq 2\left\{\mathbb{E}(\varphi^* - \bar{\bar{\varphi}}^*)^2 + \mathbb{E}(\bar{\bar{\varphi}}^*)^2\right\} \\ &= 2\left\{\mathbb{E}(\varphi^* - \bar{\bar{\varphi}}^*)^2 + O(h^{-2})\right\}. \end{aligned} \tag{4.25}$$

Summarizing we see that it suffices to find a proper upper bound for $\mathbb{E}(\varphi^* - \bar{\bar{\varphi}}^*)^2$. As we shall see we have

$$\mathbb{E}(\varphi^* - \bar{\bar{\varphi}}^*)^2 = O((nh)^{-2}). \tag{4.26}$$

This in turn yields the bound $O(n^{-1}h^{-2})$ for (4.24) and $O(h^{-2})$ for (4.25) and completes the proof of the lemma.

Now, to get (4.26), note that $\varphi^* - \bar{\bar{\varphi}}^* = (\varphi - \bar{\bar{\varphi}})^*$. Apply the first inequality in (4.22) to obtain

$$\mathbb{E}(\varphi^* - \bar{\bar{\varphi}}^*)^2 \leq 16\mathbb{E}(\varphi - \bar{\bar{\varphi}})^2.$$

To bound the last expectation we again only deal with the first sample and put $n = n_1$. In this case

$$\begin{aligned} \varphi(x) - \bar{\bar{\varphi}}(x) &= W(x)\left[\sum_{i=1}^{n} W_{1i}(x)Y_{1i} - \sum_{i=3}^{n} W_{n-2,1,i}(x)Y_{1i}\right] \\ &= W(x)[W_{11}(x)Y_{11} + W_{12}(x)Y_{12}] \qquad (4.27)\\ &+ W(x)\left[\sum_{i=3}^{n} Y_{11}(W_{1i}(x) - W_{n-2,1,i}(x))\right]. \qquad (4.28) \end{aligned}$$

Here $W_{n-2,1,i}$ denotes $W_{1i}$ with the first two data deleted from the sample. Since by assumption $W(X)Y$ has a finite second moment and $K$ is bounded from above, the lower bound in Corollary 4.3 yields that the second moment of (4.27) is of the order $O((nh)^{-2})$. To bound the second moment of (4.28), write

$$W_{1i} = \frac{a_i}{b} \text{ and } W_{n-2,1,i} = \frac{c_i}{d}$$

so that

$$W_{1i} - W_{n-2,1,i} = \frac{a_i - c_i}{b} + \frac{c_i}{bd}(d-b).$$

But

$$\begin{aligned} a_i - c_i &= K\left(\frac{\hat{F}_n(X_{1i}) - \hat{F}_n(x)}{h}\right) - K\left(\frac{\hat{F}_{n-2}(X_{1i}) - \hat{F}_{n-2}(x)}{h}\right) \\ &= K'(\Delta_i)\frac{\hat{F}_n(X_{1i}) - \hat{F}_n(x) - \hat{F}_{n-2}(X_{1i}) + \hat{F}_{n-2}(x)}{h}, \end{aligned}$$

where $\Delta_i$ is an appropriate value between the two ratios, and the last ratio is uniformly bounded in absolute values by $4/nh$. We may now again apply the lower bound in Corollary 4.3 to obtain, for some finite constant $C$,

$$\mathbb{E}\left[W(X)\sum_{i=3}^{n} Y_{1i}\frac{a_i - c_i}{b}\right]^2 \le Cn^{-4}h^{-4}\mathbb{E}\left[|W(X)|\sum_{i=3}^{n}|Y_{1i}||K'(\Delta_i)|\right]^2. \tag{4.29}$$

From Lemma 4.5 to follow, with $p = 2$, the last expectation is $O(n^2h^2)$, so that

$$(4.29) = O(n^{-2}h^{-2}),$$

as desired. Finally, we study

$$\begin{aligned} \mathbb{E}\left[W(X)\sum_{i=3}^{n} Y_{1i}\frac{c_i(d-b)}{bd}\right]^2 &= \mathbb{E}\left[W(X)\frac{d-b}{b}\sum_{i=3}^{n} Y_{1i}W_{n-2,1,i}(X)\right]^2 \\ \le\ \mathbb{E}&\left[|W(X)|\frac{|d-b|}{b}\sum_{i=3}^{n}|Y_{1i}|W_{n-2,1,i}(X)\right]^2. \end{aligned} \tag{4.30}$$

To bound $b$ in the denominator, we apply the lower bound from Corollary 4.3. For $d - b$ we have

$$b - d = \sum_{i=1}^{n} a_i - \sum_{i=3}^{n} c_i = a_1 + a_2 + \sum_{i=3}^{n}(a_i - c_i).$$

The boundedness of $a_1$ and $a_2$ follows from the boundedness of $K$. For the remaining sum we have as above

$$\left|\sum_{i=3}^{n}(a_i - c_i)\right| = \sum_{i=3}^{n}|K'(\Delta_i)| \cdot O(1/nh).$$

Apply Lemma 4.5 with $p = 1$ and $Y_{1i} \equiv 1$ to show that the last expression is bounded. Finally, from Stone (1977)

$$\sum_{i=3}^{n}|Y_{1i}|W_{n-2,1,i}(X) \to \mathbb{E}[|Y_1||X] \text{ in } L^2.$$

Conclude that (4.30) is of the order $O((nh)^{-2})$. This completes the proof of the lemma. □

**Lemma 4.5** *Under the assumptions of Theorem 2.1 we have for* $p = 1, 2$*:*

$$\mathbb{E}\left[|W(X)|\sum_{i=3}^{n}|Y_{1i}||K'(\Delta_i)|\right]^p = O((nh)^p).$$

**Proof.** We only deal with $p = 2$. Omitting the sample index 1, the expectation becomes

$$(n-2)\mathbb{E}\left[W^2(X)Y_3^2(K'(\Delta_3))^2\right] \tag{4.31}$$

$$+\quad (n-2)(n-3)\mathbb{E}\left[W^2(X)|Y_3||Y_4||K'(\Delta_3)||K'(\Delta_4)|\right] \tag{4.32}$$

To bound (4.31) recall that $K$ and therefore also $K'$ is w.l.o.g. supported by $[-1,1]$. Since $|K'| \le c < \infty$, we therefore have

$$|K'(\Delta_3)| \le c 1_{\{|\Delta_3| \le 1\}}.$$

Furthermore, when $4/nh < 2$, we have

$$1_{\{|\Delta_3| \le 1\}} \le 1_{\left\{\left|\frac{\hat{F}_n(X_3) - \hat{F}_n(X)}{h}\right| \le 1\right\}} + 1_{\left\{\left|\frac{\hat{F}_{n-2}(X_3) - \hat{F}_{n-2}(X)}{h}\right| \le 1\right\}}. \tag{4.33}$$

Conclude that the expectation in (4.31) is less than or equal to

$$c^2 \mathbb{E}\left[W^2(X) Y_3^2 1_{\{|\frac{\hat{F}_n(X_3) - \hat{F}_n(X)}{h}| \le 1\}}\right] + c^2 \mathbb{E}\left[W^2(X) Y_3^2 1_{\{|\frac{\hat{F}_{n-2}(X_3) - \hat{F}_{n-2}(X)}{h}| \le 1\}}\right].$$

We only deal with the first expectation, the other being the same for sample size $n-2$. By the DKW-inequality for empirical processes, see Dvoretzky, Kiefer and Wolfowitz (1956), we obtain for some constant $C$

$$\mathbb{P}\left(n^{1/2} \sup_x |\hat{F}_n(x) - F(x)| \ge d\right) \le C \exp[-2d^2] \tag{4.34}$$

for all $d > 0$. Similarly for sample size $n-2$. Put $d = L\sqrt{\ln n}$ for some positive constant $L$ to be chosen later. Then (4.34) implies

$$\mathbb{P}\left(n^{1/2} \sup_x |\hat{F}_n(x) - F(x)| \ge d\right) \le C n^{-2L^2}.$$

On the set $\{n^{1/2} \sup_x |\hat{F}_n(x) - F(x)| < d\}$, the inequality

$$|\hat{F}_n(X_3) - \hat{F}_n(X)| \le h \text{ implies the inequality } |F(X_3) - F(X)| \le h + \frac{2d}{\sqrt{n}}.$$

Since by assumption $\frac{nh^2}{\ln n} \to \infty$, we have, at least for all large $n$, that $h + \frac{2d}{\sqrt{n}} \le 2h$. We therefore obtain

$$\begin{aligned}
&\mathbb{E}\left[W^2(X) Y_3^2 1_{\left\{\left|\frac{\hat{F}_n(X_3) - \hat{F}_n(X)}{h}\right| \le 1\right\}}\right] = \mathbb{E}\left[W^2(X) Y_3^2 1_{\{\ldots \le 1, n^{1/2} \sup_x |\hat{F}_n(x) - F(x)| \ge d\}}\right] \\
&+ \mathbb{E}\left[W^2(X) Y_3^2 1_{\{\ldots \le 1, n^{1/2} \sup_x |\hat{F}_n(x) - F(x)| < d\}}\right] \\
&\le \mathbb{E}\left[W^2(X) Y_3^2 1_{\{n^{1/2} \sup_x |\hat{F}_n(x) - F(x)| \ge d\}}\right] \qquad (4.35) \\
&+ \mathbb{E}\left[W^2(X) Y_3^2 1_{\{|F(X_3) - F(X)| \le 2h\}}\right]. \qquad (4.36)
\end{aligned}$$

To bound (4.36), recall that $X$ only depends on the first variables from the two samples and is therefore independent of $(X_3, Y_3)$. Putting $\sigma^2(x) = \mathbb{E}[Y_3^2 | X_3 = x]$, we may therefore condition on $X$ to get that (4.36) equals

$$\begin{aligned}
&\mathbb{E}\left[W^2(X) \mathbb{E}\left[Y_3^2 1_{\{|F(X_3) - F(X)| \le 2h\}} | X\right]\right] \\
&= \mathbb{E}\left[W^2(X) \int_0^1 \sigma^2(F^{-1}(u)) 1_{\{|u - F(X)| \le 2h\}} du\right].
\end{aligned}$$

From differentiation theory this expectation, however, is asymptotically proportional to $4h\mathbb{E}[W^2(X)\sigma^2(X)] = O(h)$.

Next we bound (4.35). By the DKW-inequality, (4.35) is less than or equal to, for $d = L\sqrt{\ln n}$ and any $L > 0$:

$$\sqrt{\mathbb{E}[W^4(X)Y_3^4]}\sqrt{\mathbb{P}(n^{1/2}\sup_x |\hat{F}_n(x) - F(x)| \geq d)} = O(n^{-L^2}).$$

For $L = 1$, the last term is $O(h)$. Altogether we have shown that (4.31) is of the order $O(nh)$ and therefore also $O(n^2h^2)$. We now come to (4.32). Similar to before, we get

$$\mathbb{E}\left[W^2(X)|Y_3||Y_4||K'(\Delta_3)||K'(\Delta_4)|\right] \leq c^2\mathbb{E}\left[W^2(X)|Y_3||Y_4|1_{\{|\Delta_3|\leq 1,|\Delta_4|\leq 1\}}\right].$$

From (4.33) we see that the last expectation may be bounded from above by

$$\begin{aligned} &\mathbb{E}\left[W^2(X)|Y_3||Y_4|1_{\{|\hat{F}_n(X_3)-\hat{F}_n(X)|\leq h,|\hat{F}_n(X_4)-\hat{F}_n(X)|\leq h\}}\right] \\ +\ &\mathbb{E}\left[W^2(X)|Y_3||Y_4|1_{\{|\hat{F}_{n-2}(X_3)-\hat{F}_{n-2}(X)|\leq h,|\hat{F}_{n-2}(X_4)-\hat{F}_{n-2}(X)|\leq h\}}\right] \\ +\ &\mathbb{E}\left[W^2(X)|Y_3||Y_4|1_{\{|\hat{F}_n(X_3)-\hat{F}_n(X)|\leq h,|\hat{F}_{n-2}(X_4)-\hat{F}_{n-2}(X)|\leq h\}}\right] \\ +\ &\mathbb{E}\left[W^2(X)|Y_3||Y_4|1_{\{|\hat{F}_{n-2}(X_3)-\hat{F}_{n-2}(X)|\leq h,|\hat{F}_n(X_4)-\hat{F}_n(X)|\leq h\}}\right]. \end{aligned}$$

We only deal with the first expectation, the others being similar. Using an argument based on the DKW-inequality, the first expectation is less than or equal to

$$\begin{aligned} &\mathbb{E}\left[W^2(X)|Y_3||Y_4|1_{\{|F(X_3)-F(X)|\leq 2h,|F(X_4)-F(X)|\leq 2h\}}\right] \qquad (4.37) \\ +\ &\mathbb{E}\left[W^2(X)|Y_3||Y_4|1_{\{n^{1/2}\sup_x |\hat{F}_n(x)-F(x)|\geq d\}}\right]. \end{aligned}$$

The last expectation is, as before, of the order $O(n^{-L^2})$, which for $L = 1$ is $O(h^2)$. As to (4.37), condition on $X$ and then use the independence of $(X_3, Y_3)$ and $(X_4, Y_4)$ together with a differentiation argument to show that (4.37) is of the order $O(h^2)$.

This concludes the proof of Lemma 4.5. □

Note that with Lemma 4.5 we also completed the proof of Lemma 4.4. We now bound each of the integrals (4.3)-(4.6). Since the analysis of (4.4) and (4.5) will be similar, it suffices to consider (4.5), say.

**Lemma 4.6** *Under the assumptions of Theorem 2.1, we have, with $n = n_1$,*

$$\mathbb{E}\left[\int\int W(x)[\hat{m}_1(x) - m_1(x)][\hat{F}(dx_1) - F(dx_1)]G(dx_2)\right]^2 = o(n^{-1}).$$

**Proof.** Again, we shall omit the index 1 for the sample. By the Cauchy-Schwarz inequality the expectation is less than or equal to

$$\mathbb{E}\left[\int\left[\int W(x)[\hat{m}_1(x)-m_1(x)][\hat{F}(dx_1)-F(dx_1)]\right]^2 G(dx_2)\right]$$

$$= \int \mathbb{E}\left[\int W(x)[\hat{m}_1(x)-m_1(x)][\hat{F}(dx_1)-F(dx_1)]\right]^2 G(dx_2). \qquad (4.38)$$

Set, for each $x_2 \in \mathbb{R}$,

$$\hat{\varphi}_1(x_1) = W(x)[\hat{m}_1(x)-m_1(x)],$$

where as before $x = \frac{x_1+x_2}{2}$. Then the inner expectation in (4.38) becomes

$$\mathbb{E}\left[n^{-1}\sum_{i=1}^{n}\hat{\varphi}_1(X_i) - \int \hat{\varphi}_1(x_1)F(dx_1)\right]^2$$

$$= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left\{(\hat{\varphi}_1(X_i)-\int\hat{\varphi}_1 dF)(\hat{\varphi}_1(X_j)-\int\hat{\varphi}_1 dF)\right\}$$

$$= n^{-1}\mathbb{E}\left\{(\hat{\varphi}_1(X_1)-\int\hat{\varphi}_1 dF)^2\right\} \qquad (4.39)$$

$$+ \frac{n-1}{n}\mathbb{E}\left\{(\hat{\varphi}_1(X_1)-\int\hat{\varphi}_1 dF)(\hat{\varphi}_1(X_2)-\int\hat{\varphi}_1 dF)\right\}. \qquad (4.40)$$

To bound (4.39) from above, we first consider the expectation for $\bar{\varphi}_1$, where $\bar{\varphi}_1$ equals $\hat{\varphi}_1$ but with $(X_1, Y_1)$ deleted from the first sample. But

$$n^{-1}\mathbb{E}\left\{(\bar{\varphi}_1(X_1)-\int\bar{\varphi}_1 dF)^2\right\} = n^{-1}\mathbb{E}\left\{\mathbb{E}[(\ldots)^2|X_i,Y_i,2\le i\le n]\right\}$$

$$\le n^{-1}\mathbb{E}\left\{\mathbb{E}[\bar{\varphi}_1^2(X_1)|X_i,Y_i,2\le i\le n]\right\} = n^{-1}\mathbb{E}\left\{\int\bar{\varphi}_1^2(x_1)F(dx_1)\right\}. \qquad (4.41)$$

We now discuss (4.40), first for the case when the first two data have been deleted from the sample. Denote with $\bar{\bar{\varphi}}_1$ the corresponding $\hat{\varphi}_1$. But then

$$\mathbb{E}\left\{(\bar{\bar{\varphi}}_1(X_1)-\int\bar{\bar{\varphi}}_1 dF)(\bar{\bar{\varphi}}_1(X_2)-\int\bar{\bar{\varphi}}_1 dF)\right\} = \mathbb{E}\left[\mathbb{E}\left\{(\ldots)(\ldots)|X_i,Y_i,3\le i\le n\right\}\right].$$

By independence of $X_1$ and $X_2$, the inner conditional expectation factorizes and therefore vanishes. For the original $\hat{\varphi}_1$, we write

$$\hat{\varphi}_1 - \int\hat{\varphi}_1 dF = \bar{\varphi}_1 - \int\bar{\varphi}_1 dF + \hat{\varphi}_1 - \bar{\varphi}_1 + \int\bar{\varphi}_1 dF - \int\hat{\varphi}_1 dF.$$

Using the inequality $(a+b)^2 \le 2(a^2+b^2)$ we therefore get

$$n^{-1}\mathbb{E}\left\{(\hat{\varphi}_1-\int\hat{\varphi}_1 dF)^2\right\} \le 2n^{-1}\mathbb{E}\left\{(\bar{\varphi}_1-\int\bar{\varphi}_1 dF)^2\right\} + 2n^{-1}\mathbb{E}\left\{(\hat{\varphi}_1-\bar{\varphi}_1+\int\bar{\varphi}_1 dF-\int\hat{\varphi}_1 dF)^2\right\}$$

$$\le 2n^{-1}\mathbb{E}\left\{(\bar{\varphi}_1-\int\bar{\varphi}_1 dF)^2\right\} + 4n^{-1}\mathbb{E}\{(\hat{\varphi}_1-\bar{\varphi}_1)^2\} + 4n^{-1}\int\mathbb{E}\left[(\bar{\varphi}_1-\hat{\varphi}_1)^2\right]dF, \qquad (4.42)$$

where for the last inequality we used the Cauchy-Schwarz inequality and Fubini's Theorem. Up to the factor 2, the first term is bounded from above by (4.41). It follows from Stone (1977) that

$$\mathbb{E}\int\int \bar{\varphi}_1^2(x)F(dx_1)G(dx_2) \to 0 \text{ as } n \to \infty. \tag{4.43}$$

By the same technique leading to (4.26) we obtain for the other two expectations in (4.42):

$$\int \mathbb{E}\left[(\hat{\varphi}_1 - \bar{\varphi}_1)^2\right] G(dx_2) = O\left(\frac{1}{n^2h^2}\right) = o(1)$$

and

$$\int\int \mathbb{E}\left[(\bar{\varphi}_1 - \hat{\varphi}_1)^2\right] dFdG = O\left(\frac{1}{n^2h^2}\right) = o(1).$$

Together with (4.43) we therefore obtain that

$$n^{-1}\int \mathbb{E}\left[\hat{\varphi}_1(X_1) - \int \hat{\varphi}_1 dF\right]^2 dG = o(n^{-1}).$$

We now analyze (4.40) for the original $\hat{\varphi}_1$. Proceeding as before (4.24), we have

$$\begin{aligned}
&\mathbb{E}\left[(\hat{\varphi}_1 - \int \hat{\varphi}_1 dF)(\hat{\varphi}_1 - \int \hat{\varphi}_1 dF)\right] \\
&= \mathbb{E}\left[(\hat{\varphi}_1 - \int \hat{\varphi}_1 dF - \bar{\bar{\varphi}}_1 + \int \bar{\bar{\varphi}}_1 dF)(\hat{\varphi}_1 - \int \hat{\varphi}_1 dF - \bar{\bar{\varphi}}_1 + \int \bar{\bar{\varphi}}_1 dF)\right] \\
&+ \mathbb{E}\left[(\hat{\varphi}_1 - \int \hat{\varphi}_1 dF - \bar{\bar{\varphi}}_1 + \int \bar{\bar{\varphi}}_1 dF)(\bar{\bar{\varphi}}_1 - \int \bar{\bar{\varphi}}_1 dF)\right] \\
&+ \mathbb{E}\left[(\bar{\bar{\varphi}}_1 - \int \bar{\bar{\varphi}}_1 dF)(\hat{\varphi}_1 - \int \hat{\varphi}_1 dF - \bar{\bar{\varphi}}_1 + \int \bar{\bar{\varphi}}_1 dF)\right].
\end{aligned}$$

Each of the three expectations may be bounded from above by using Cauchy-Schwarz. For the first term we obtain the order $(nh)^{-2}$. For the other two we get the upper bound

$$O[(nh)^{-1}]\sqrt{\mathbb{E}(\bar{\bar{\varphi}}_1 - \int \bar{\bar{\varphi}}_1 dF)^2} \le O[(nh)^{-1}]\left[\int \mathbb{E}\bar{\bar{\varphi}}_1^2(x_1)F(dx_1)\right]^{1/2}.$$

To prove the lemma it suffices to show that

$$\int\left[\int \mathbb{E}\bar{\bar{\varphi}}_1^2(x_1)F(dx_1)\right]^{1/2} G(dx_2) = o(h).$$

By Cauchy-Schwarz this will follow from

$$\sqrt{\mathbb{E}\int\int \bar{\bar{\varphi}}_1^2(x_1)F(dx_1)G(dx_2)} = o(h)$$

or, in other words,

$$\mathbb{E}\left[\int\int \bar{\bar{\varphi}}_1^2(x_1)F(dx_1)G(dx_2)\right] = o(h^2). \tag{4.44}$$

This, however, will follow from the next lemma. □

**Lemma 4.7** *Under the assumptions of Theorem 2.1, we have (4.44).*

**Proof.** We show (4.44) for sample size $n$ rather than $n-2$, i.e., for $\hat{\varphi}_1$. Obviously, with $x = \frac{x_1+x_2}{2}$, we have because of $\sum W_{1i} = 1$:

$$\begin{aligned}\hat{\varphi}_1(x) &= W(x)[\hat{m}_1(x) - m_1(x)] = W(x)\sum_{i=1}^{n} W_{1i}(x)[Y_i - m_1(X_i)] \\ &+ W(x)\sum_{i=1}^{n} W_{1i}(x)[m_1(X_i) - m_1(x)].\end{aligned}$$

Conclude that (4.44) equals

$$\int W^2(x)\mathbb{E}\left[\sum_{i=1}^{n} W_{1i}(x)\left[Y_i - m_1(X_i)\right]\right]^2 H(dx) \tag{4.45}$$

$$+\int W^2(x)\mathbb{E}\left[\sum_{i=1}^{n} W_{1i}(x)\left[m_1(X_i) - m_1(x)\right]\right]^2 H(dx), \tag{4.46}$$

upon noticing that the two terms in the brackets are uncorrelated. In (4.45), the summands are also uncorrelated so that (4.45) becomes

$$\int W^2(x)\mathbb{E}\left[\sum_{i=1}^{n} W_{1i}^2(x)\sigma^2(X_i)\right] H(dx), \tag{4.47}$$

where $\sigma^2(x) = \mathrm{Var}(Y_1|X_1 = x)$. Now, recall Corollary 4.3. Together with the boundedness of $K$ we therefore have that (4.47) is of the order

$$O\left(\frac{1}{nh}\right)\int W^2(x)\mathbb{E}\left[\sum_{i=1}^{n} W_{1i}(x)\sigma^2(X_i)\right] H(dx).$$

By Stone (1977), the integral converges to $\int W^2(x)\sigma^2(x)H(dx)$. Hence (4.45) is of the order $O(1/nh)$. By assumption (h), this term is, however, of the order $o(h^2)$.

As to (4.46), the integral becomes

$$\int W^2(x)\mathbb{E}\left[\sum_{i=1}^{n} W_{1i}^2(x)[m_1(X_i) - m_1(x)]^2\right] H(dx) \tag{4.48}$$

$$\begin{aligned}+\int W^2(x)\mathbb{E}\Bigg[\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} &W_{1i}(x)W_{1j}(x)[m_1(X_i) - m_1(x)] \\ &\times [m_1(X_j) - m_1(x)]\Bigg] H(dx).\end{aligned} \tag{4.49}$$

The expression in (4.48) is bounded from above, again by Corollary 4.3, by

$$O\left(\frac{1}{nh^2}\right)\int W^2(x)\mathbb{E}\left\{K^2\left(\frac{\hat F(X_1)-\hat F(x)}{h}\right)[m_1(X_1)-m_1(x)]^2\right\}H(dx).$$

Now use the boundedness of $K$, the property $K \le C1_{[-1,1]}$ and the DKW-inequality to get that the last integral is, similar to the proof of Lemma 4.5, less than or equal to

$$\begin{aligned} & C^2\int W^2(x)\mathbb{E}\left\{1_{\{|\hat F(X_1)-\hat F(x)|\le h\}}[m_1(X_1)-m_1(x)]^2\right\}H(dx) \\ \le\ & C^2\int W^2(x)\mathbb{E}\left\{1_{\{|F(X_1)-F(x)|\le 2h\}}[m_1(X_1)-m_1(x)]^2\right\}H(dx) \\ +\ & O(n^{-L^2}),\ \text{for any } L\ge 1. \end{aligned}$$

The integral equals

$$\int W^2(x)\int_0^1 1_{\{|u-F(x)|\le 2h\}}[m_1(F^{-1}(u))-m_1(x)]^2 du H(dx) = O(h^3).$$

Hence (4.48) is of the order $O(hn^{-1}) = o(h^2)$. It remains to bound (4.49). The integral equals

$$n(n-1)\mathbb{E}\left[\int W^2(x)W_{11}(x)W_{12}(x)[m_1(X_1)-m_1(x)][m_1(X_2)-m_1(x)]H(dx)\right].$$

The $H$-integral equals

$$\begin{aligned} \int_0^1 & W^2(F^{-1}(u))\frac{h(F^{-1}(u))}{f(F^{-1}(u))}W_{11}(F^{-1}(u))W_{12}(F^{-1}(u))[m_1(F^{-1}(U_1))-m_1(F^{-1}(u))] \\ & \times[m_1(F^{-1}(U_2))-m_1(F^{-1}(u))]du. \end{aligned}$$

Note also that, e.g., $W_{11}(F^{-1}(u))$ equals $\frac{K\left(\frac{\bar F_n(U_1)-\bar F_n(u)}{h}\right)}{\sum_{j=1}^n K\left(\frac{\bar F_n(U_j)-\bar F_n(u)}{h}\right)}$. Here $\bar F_n$ is the empirical d.f. of the uniform sample $U_i = F(X_i), 1\le i\le n$. Furthermore,

$$\begin{aligned} m_1(F^{-1}(U_1))-m_1(F^{-1}(u)) &= (U_1-u)(m_1\circ F^{-1})'(u) \\ &+ \frac{1}{2}(U_1-u)^2(m_1\circ F^{-1})''(\Delta_1) \end{aligned}$$

for some $\Delta_1$ between $U_1$ and $u$. Similarly, for $U_2$. Multiplication leads to four integrals of which the first equals

$$n(n-1)\int_0^1 W^2(F^{-1}(u))\frac{h(F^{-1}(u))}{f(F^{-1}(u))}W_{11}(F^{-1}(u))W_{12}(F^{-1}(u))(U_1-u)(U_2-u)[(m_1\circ F^{-1})'(u)]^2du. \tag{4.50}$$

We first consider the integral where in the numerator of $W_{11}$ and $W_{12}$ we have $K\left(\frac{U_1-u}{h}\right)$ and $K\left(\frac{U_2-u}{h}\right)$, respectively. Since the denominators of these $W$'s behave like $nh$ for each $0 < u < 1$, according to Corollary 4.2, it suffices to show that

$$\mathbb{E}\left[\int_0^1 W^2(F^{-1}(u))\frac{h(F^{-1}(u))}{f(F^{-1}(u))}[(m_1\circ F^{-1})'(u)]^2 \right.$$
$$\left. \times K\left(\frac{U_1-u}{h}\right)\left(\frac{U_1-u}{h}\right)K\left(\frac{U_2-u}{h}\right)\left(\frac{U_2-u}{h}\right)du\right] = o(h^2). \qquad (4.51)$$

Setting

$$A(u) = W^2(F^{-1}(u))\frac{h(F^{-1}(u))}{f(F^{-1}(u))}[(m_1\circ F^{-1})'(u)]^2,$$

the expectation in (4.51) equals, by independence of $U_1$ and $U_2$,

$$\int_0^1 A(u)\mathbb{E}^2\left[K\left(\frac{U_1-u}{h}\right)\frac{U_1-u}{h}\right]du$$
$$= \int_0^1 A(u)\left[\int_0^1 K\left(\frac{v-u}{h}\right)\frac{v-u}{h}dv\right]^2 du = \int_0^1 A(u)\left[h\int_{-\frac{u}{h}}^{\frac{1-u}{h}} K(w)wdw\right]^2 du.$$

Since, for $0 < u < 1$, the inner integral tends to

$$\int_{-\infty}^{\infty} K(w)wdw = 0,$$

this shows (4.51).

Next we study

$$h^{-2}\int_0^1 A(u)(U_1-u)(U_2-u)\left[K\left(\frac{\bar{F}_n(U_1)-\bar{F}_n(u)}{h}\right)-K\left(\frac{U_1-u}{h}\right)\right]K\left(\frac{U_2-u}{h}\right)du.$$

By the Mean Value Theorem this expression equals

$$h^{-2}\int_0^1 A(u)(U_1-u)(U_2-u)K'(\Delta_1)\frac{\bar{F}_n(U_1)-\bar{F}_n(u)-U_1+u}{h}K\left(\frac{U_2-u}{h}\right)du \qquad (4.52)$$

for some $\Delta_1$ between the corresponding ratios. Note that the original integral in (4.50) only extends over all $u$'s satisfying $|\bar{F}_n(U_1)-\bar{F}_n(u)| \le h$. By the DKW-inequality and since $\sqrt{n^{-1}\ln n} = o(h)$ this implies, as in the proof of Lemma 4.5, that with large probability we also have $|U_1-u| \le 2h$. Now write

$$\bar{F}_n(U_1)-\bar{F}_n(u)-U_1+u = n^{-1/2}[\bar{\alpha}_n(U_1)-\bar{\alpha}_n(u)],$$

where $\bar{\alpha}_n$ denotes the uniform empirical process based on $U_1, \ldots, U_n$. With large probability this term is in absolute values less than or equal to $n^{-1/2}\bar{w}_n(2h)$, with $\bar{w}_n$ denoting the oscillation modulus of the uniform empirical process. From Theorem 2.14 in Stute (1982) this modulus is of the order $O_{\mathbb{P}}(\sqrt{h \ln h^{-1}})$. Since $K$ is supported by [-1,1], we also need to integrate only over $|U_2 - u| \le h$. Consequently, on the set of relevant $u$'s, $|U_1 - u|/h$ and $|U_2 - u|/h$ are bounded. Hence the integral in (4.52) is bounded from above by

$$C\frac{\sqrt{\ln h^{-1}}}{\sqrt{nh}} \int_0^1 A(u) K\left(\frac{U_2 - u}{h}\right) du,$$

where $C$ is a proper constant. The expectation of the last integral is, however, of the order $h$, so that finally the expectation of (4.52) is of the order

$$O\left(\sqrt{hn^{-1}\ln h^{-1}}\right) = o(h^2),$$

by (h). The same arguments apply to the error term

$$h^{-2} \int_0^1 A(u)(U_1 - u)(U_2 - u) K\left(\frac{U_1 - u}{h}\right) \left[K\left(\frac{\bar{F}_n(U_2) - \bar{F}_n(u)}{h}\right) - K\left(\frac{U_2 - u}{h}\right)\right] du.$$

It remains to bound

$$h^{-2} \int_0^1 A(u)(U_1 - u)(U_2 - u) \left[K\left(\frac{\bar{F}_n(U_1) - \bar{F}_n(u)}{h}\right) - K\left(\frac{U_1 - u}{h}\right)\right]$$
$$\times \left[K\left(\frac{\bar{F}_n(U_2) - \bar{F}_n(u)}{h}\right) - K\left(\frac{U_2 - u}{h}\right)\right] du.$$

By the same arguments as before each term in $[\ldots]$ is of the order $\sqrt{h^{-1}n^{-1}\ln h^{-1}}$, so that overall the integral is uniformly in $U_1$ and $U_2$ bounded by $O(h^{-1}n^{-1}\ln h^{-1}) = o(h^2)$. Summarizing we see that the expectation of (4.50) is $o(h^2)$.

Next we analyze

$$n(n-1) \int_0^1 W^2(F^{-1}(u)) \frac{h(F^{-1}(u))}{f(F^{-1}(u))} W_{11}(F^{-1}(u)) W_{12}(F^{-1}(u))(U_1 - u)(m_1 \circ F^{-1})'(u)$$
$$\times \frac{1}{2}(U_2 - u)^2 (m_1 \circ F^{-1})''(\Delta_2) du.$$

This term is bounded from above by

$$Ch \int_0^1 W^2(F^{-1}(u)) \frac{h(F^{-1}(u))}{f(F^{-1}(u))} |(m_1 \circ F^{-1})'(u)| K\left(\frac{\bar{F}_n(U_1) - \bar{F}_n(u)}{h}\right)$$
$$\times |(m_1 \circ F^{-1})''(\Delta_2)| K\left(\frac{\bar{F}_n(U_2) - \bar{F}_n(u)}{h}\right) du.$$

As before it suffices to study the integral with $K\left(\frac{U_1-u}{h}\right)$ and $K\left(\frac{U_2-u}{h}\right)$. Then the expectation is of the order $O(h^3)=o(h^2)$. Finally, the expectation of

$$\frac{1}{4}n(n-1)\int_0^1 W^2(F^{-1}(u))\frac{h(F^{-1}(u))}{f(F^{-1}(u))}W_{11}(F^{-1}(u))W_{12}(F^{-1}(u))(U_1-u)^2(U_2-u)^2$$
$$\times(m_1\circ F^{-1})''(\Delta_1)(m_1\circ F^{-1})''(\Delta_2)du$$

is of the order $O(h^4)=o(h^2)$. This completes the proof of Lemma 4.7. □

At the same time we also completed the proof of Lemma 4.6.

Next we bound (4.3) and (4.6). Since they are of similar structure, we restrict ourselves to (4.3).

**Lemma 4.8** *Under the assumptions of Theorem 2.1, we have, with $n=n_1$,*

$$\mathbb{E}\left[\int\int W(x)[\hat{m}_1(x)-m_1(x)]F(dx_1)[\hat{G}(dx_2)-G(dx_2)]\right]^2=o(n^{-1}).$$

**Proof.** Put, for each $x_2$,

$$\varphi(x_2)=\int W(x)[\hat{m}_1(x)-m_1(x)]F(dx_1).$$

This function is random but only depends on the first sample. Hence, by independence of the two samples, we have

$$\mathbb{E}\left[\int\varphi(x_2)[\hat{G}(dx_2)-G(dx_2)]\right]^2=\mathbb{E}\left\{\mathbb{E}\left[\left(\int\varphi(d\hat{G}-dG)\right)^2\mid\text{first sample}\right]\right\}$$
$$=\mathbb{E}\left\{\frac{1}{n_2}\left(\int\varphi^2dG-\left(\int\varphi dG\right)^2\right)\right\}\leq\frac{1}{n_2}\mathbb{E}\left[\int\varphi^2dG\right].$$

Now, since $n_1$ and $n_2$ are of the same order, it suffices to show that $\mathbb{E}\int\varphi^2dG\to 0$. This, however, follows from the Cauchy-Schwarz inequality and Stone (1977). □

Summarizing we see, that (4.2)-(4.6) are negligible.

To take care of (4.7) we need the following lemma.

**Lemma 4.9** *Under the assumptions of Theorem 2.1 we have (with $n=n_1$)*

$$\sqrt{n}\int W(x)[\hat{m}_1(x)-m_1(x)]H(dx)$$
$$=n^{1/2}\sum_{i=1}^{n_1}(Y_i-m_1(X_i))\int W(x)W_{1i}(x)H(dx)+o_{\mathbb{P}}(1).\qquad(4.53)$$

**Proof.** The expression (4.7) equals, because of $\sum W_{1i}(x) = 1$,

$$\int W(x)[\hat{m}_1(x) - m_1(x)]H(dx) = \int W(x)\left[\sum_{i=1}^{n}(Y_i - m_1(x))W_{1i}(x)\right]H(dx)$$

$$= \int W(x)\left[\sum_{i=1}^{n}(Y_i - m_1(X_i))W_{1i}(x)\right]H(dx) \tag{4.54}$$

$$+ \int W(x)\left[\sum_{i=1}^{n}(m_1(X_i) - m_1(x))W_{1i}(x)\right]H(dx). \tag{4.55}$$

The expression in (4.54) equals

$$\sum_{i=1}^{n}(Y_i - m_1(X_i))\int W(x)W_{1i}(x)H(dx),$$

while (4.55) may be written as

$$\int W(x)\left[\sum_{i=1}^{n}(m_1(X_i) - m_1(x))\right]W_{1i}(x)H(dx)$$

$$= \sum_{i=1}^{n}\int W(x)\left[m_1(X_i) - m_1(x)\right]W_{1i}(x)H(dx). \tag{4.56}$$

We first study an expression corresponding to (4.56), but with the weights $W_{1i}(x)$ replaced with

$$\bar{W}_{1i}(x) = \frac{K\left(\frac{F(X_i)-F(x)}{h}\right)}{nh}.$$

This results in

$$\sum_{i=1}^{n}\int W(x)\left[m_1(X_i) - m_1(x)\right]\bar{W}_{1i}(x)H(dx), \tag{4.57}$$

a sum of independent identically distributed random variables each of which has expectation

$$\frac{1}{nh}\int\int W(x)[m_1(y) - m_1(x)]K\left(\frac{F(y)-F(x)}{h}\right)H(dx)F(dy).$$

We now show that

$$n^{1/2}\sum_{i=1}^{n}\Bigg[\int W(x)[m_1(X_i) - m_1(x)]\bar{W}_{1i}(x)H(dx)$$
$$- \frac{1}{nh}\int\int W(x)[m_1(y) - m_1(x)]\cdot K\left(\frac{F(y)-F(x)}{h}\right)H(dx)F(dy)\Bigg]$$
$$\longrightarrow 0 \text{ in probability.}$$

Actually, by Bienaymé, the variance of the sum is less than or equal to

$$\begin{aligned}
& n^2 \int \left[ \int W(x)[m_1(y) - m_1(x)] \frac{K\left(\frac{F(y)-F(x)}{h}\right)}{nh} H(dx) \right]^2 F(dy) \\
= \quad & h^{-2} \int \left[ \int W(x)[m_1(y) - m_1(x)] K\left(\frac{F(y)-F(x)}{h}\right) \frac{h(x)}{f(x)} F(dx) \right]^2 F(dy) \\
= \quad & \int \left[ \int_{\frac{F(y)-1}{h}}^{\frac{F(y)}{h}} W(F^{-1}(F(y) - wh))[m_1 \circ F^{-1}(F(y)) - m_1 \circ F^{-1}(F(y) - wh)] \right. \\
& \qquad \left. \times K(w) \frac{h \circ F^{-1}(\ldots)}{f \circ F^{-1}(\ldots)} dw \right]^2 F(dy).
\end{aligned}$$

The last term, however, tends to zero as $h \to 0$.

Next, we study the expectation of (4.57) after multiplication with $n^{1/2}$, namely

$$\begin{aligned}
& \frac{n^{3/2}}{nh} \int \int W(x)[m_1(y) - m_1(x)] K\left(\frac{F(y)-F(x)}{h}\right) H(dx) F(dy) \\
= \quad & \frac{n^{1/2}}{h} \int W(x) \int_0^1 [m_1(F^{-1}(u)) - m_1(x)] K(\frac{u - F(x)}{h}) du H(dx) \\
= \quad & n^{1/2} \int W(x) \int_{\cdots}^{\cdots} \left[ [m_1(F^{-1})]'(F(x)) wh + \frac{1}{2} [m_1(F^{-1})]''(\Delta_1) w^2 h^2 \right] K(w) dw H(dx).
\end{aligned}$$

Since $\int wK(w)dw = 0$, the last term is of the order $O(n^{1/2}h^2)$. Since $nh^4 \to 0$, we see that the expectation of (4.57) (multiplied with $n^{1/2}$) also tends to zero, so that in summary,

$$n^{1/2} \sum_{i=1}^n \int W(x)[m_1(X_i) - m_1(x)] \bar{W}_{1i}(x) H(dx) \to 0 \text{ in probability.}$$

So, to prove Lemma 4.9, it remains to show that

$$n^{1/2} \sum_{i=1}^n \int W(x)[m_1(X_i) - m_1(x)][W_{1i}(x) - \bar{W}_{1i}(x)] H(dx) \to 0 \text{ in probability.}$$

For this, we write

$$\begin{aligned}
W_{1i}(x) - \bar{W}_{1i}(x) \quad = \quad & \frac{K\left(\frac{\hat{F}(X_i)-\hat{F}(x)}{h}\right) - K\left(\frac{F(X_i)-F(x)}{h}\right)}{\sum_{j=1}^n K\left(\frac{\hat{F}(X_j)-\hat{F}(x)}{h}\right)} \\
+ \quad & \frac{K\left(\frac{F(X_i)-F(x)}{h}\right)\left(nh - \sum_{j=1}^n K\left(\frac{\hat{F}(X_j)-\hat{F}(x)}{h}\right)\right)}{nh \sum_{j=1}^n K\left(\frac{\hat{F}(X_j)-\hat{F}(x)}{h}\right)}
\end{aligned}$$

In the first step we show

$$n^{1/2}\sum_{i=1}^{n}\int W(x)[m_1(X_i)-m_1(x)]\frac{K\left(\frac{\hat F(X_i)-\hat F(x)}{h}\right)-K\left(\frac{F(X_i)-F(x)}{h}\right)}{\sum_{j=1}^{n}K\left(\frac{\hat F(X_j)-\hat F(x)}{h}\right)}H(dx)$$
$$\longrightarrow 0 \text{ in probability.}$$

Since the support of $K$ is contained in $[-1,1]$, the above term in absolute values is bound from above, in view of Corollary 4.3, by

$$\frac{C}{nh^2}\sum_{i=1}^{n}\int_{\substack{\{x:|\hat F(X_i)-\hat F(x)|\le h\}\\ \text{or } \{x:|F(X_i)-F(x)|\le h\}}}|W(x)||m_1(X_i)-m_1(x)||K'(\Delta_i)||\bar\alpha_n(F(X_i))-\bar\alpha_n(F(x))|H(dx)$$

As before, using the DKW-bound and the fact that $n^{-1/2}\ll h$, we have that with large probability each set $\{x:|\hat F(X_i)-\hat F(x)|\le h\}$ is included in $\{x:|F(X_i)-F(x)|\le 2h\}$. On this set, the $\bar\alpha_n$ increment is of the order $O(\sqrt{h\ln h^{-1}})$, with probability one. Therefore, the above sum is with probability one bounded from above by

$$\frac{C\|K'\|\sqrt{h\ln h^{-1}}}{nh^2}\sum_{i=1}^{n}\int_{x:|F(X_i)-F(x)|\le 2h}|W(x)||m_1(X_i)-m_1(x)|H(dx). \tag{4.58}$$

Each of the above integrals is of the order $O(h^2)$. More precisely,

$$\begin{aligned}
&\frac{1}{h^2}\int_{x:|F(X_i)-F(x)|\le 2h}|W(x)||m_1(X_i)-m_1(x)|H(dx)\\
&= \frac{1}{h^2}\int_{U_i-2h}^{U_i+2h}|W(F^{-1}(u))||m_1(F^{-1}(U_i))-m_1(F^{-1}(u))|\frac{h(F^{-1}(u))}{f(F^{-1}(u))}du\\
&\le \frac{2}{h}\int_{U_i-2h}^{U_i+2h}|W(F^{-1}(u))||(m_1\circ F^{-1})'(\Delta_i)|\frac{h(F^{-1}(u))}{f(F^{-1}(u))}du\\
&\sim 8|W(F^{-1}(U_i))||(m_1\circ F^{-1})'(U_i)|\frac{h(F^{-1}(U_i))}{f(F^{-1}(U_i))}.
\end{aligned}$$

Hence (4.58) is bounded from above, up to constants, by

$$\frac{\sqrt{h\ln h^{-1}}}{n}\sum_{i=1}^{n}|W(F^{-1}(U_i))||(m_1\circ F^{-1})'(U_i)|\frac{h(F^{-1}(U_i))}{f(F^{-1}(U_i))}.$$

By the SLLN, the sample mean converges so that the last sum is of the order $O_{\mathbb P}(\sqrt{h\ln h^{-1}})=o_{\mathbb P}(1)$, as desired.

Next, we bound

$$n^{1/2}\sum_{i=1}^{n}\int W(x)[m_1(X_i)-m_1(x)]\frac{K\left(\frac{F(X_i)-F(x)}{h}\right)}{\sum_{j=1}^{n}K\left(\frac{\hat{F}(X_j)-\hat{F}(x)}{h}\right)}\left[1-\frac{1}{nh}\sum_{j=1}^{n}K\left(\frac{\hat{F}(X_j)-\hat{F}(x)}{h}\right)\right]H(dx). \tag{4.59}$$

Apply (4.12) and (4.14) to show that for $h \le \hat{F}(x) \le 1-h$, the term

$$1-\frac{1}{nh}\sum_{j=1}^{n}K\left(\frac{\hat{F}(X_j)-\hat{F}(x)}{h}\right) = 1-\frac{1}{nh}\sum_{j=1}^{n}K\left(\frac{\frac{j}{n}-\hat{F}(x)}{h}\right)$$

is in absolute values of the order $O(1/nh)$. Hence, if in (4.59) we restrict integration to the set $\{x : h \le \hat{F}(x) \le 1-h\}$, we obtain that this part of (4.59) is bounded from above in absolute values by

$$\frac{n^{1/2}}{(nh)^2}\sum_{i=1}^{n}\int |W(x)||m_1(X_i)-m_1(x)|K\left(\frac{F(X_i)-F(x)}{h}\right)H(dx).$$

Similar to before, it may be shown that the sum is of the order $O_{\mathbb{P}}(nh^2)$ so that finally the above term is of the order $O_{\mathbb{P}}(n^{-1/2}) = o_{\mathbb{P}}(1)$. To bound, e.g., the integral (4.59) over the set $\{x : \hat{F}(x) < h\}$, we use (4.12) and Corollary 4.3 to get that

$$\begin{aligned} & n^{1/2}\left|\sum_{i=1}^{n}\int_{\{x:\hat{F}(x)<h\}}[\ldots]H(dx)\right| \\ &= O_{\mathbb{P}}\left(\frac{1}{hn^{1/2}}\right)\sum_{i=1}^{n}\int_{\{x:\hat{F}(x)<h\}}|W(x)||m_1(X_i)-m_1(x)|K\left(\frac{F(X_i)-F(x)}{h}\right)H(dx). \end{aligned} \tag{4.60}$$

By DKW, the set $\{x : \hat{F}(x) < h\}$ is eventually contained in the set $\{x : F(x) < 2h\}$. Then use Taylor's expansion and substitution to show as usual that (4.60) is of the order $O_{\mathbb{P}}(h^2 n^{1/2})$. In view of $h^4 n \to 0$, this, however, is $o_{\mathbb{P}}(1)$. This completes the proof of the lemma. □

With the same arguments we obtain the following lemma.

**Lemma 4.10** *Under the assumptions of Theorem 2.1, we have (with $n = n_2$)*

$$\sqrt{n}\int W(x)[\hat{m}_2(x)-m_2(x)]H(dx) = \sqrt{n}\sum_{i=1}^{n_2}(Y_{2i}-m_2(X_{2i}))\int W(x)W_{2i}(x)H(dx) + o_{\mathbb{P}}(1).$$

We are now in a position to give the

**Proof of Theorem 2.1.** From Lemma 4.4 – Lemma 4.8 we have, under $H_0$,

$$\begin{aligned} \sqrt{\frac{n_1 n_2}{n_1+n_2}}\hat{T} &= \sqrt{\frac{n_1 n_2}{n_1+n_2}}\int\int W(x)[\hat{m}_1(x)-m_1(x)]F(dx_1)G(dx_2) \\ &- \sqrt{\frac{n_1 n_2}{n_1+n_2}}\int\int W(x)[\hat{m}_2(x)-m_2(x)]F(dx_1)G(dx_2) + o_{\mathbb{P}}(1), \end{aligned}$$

while Lemmas 4.9 and 4.10 provide the martingale representations of the two integrals. Finally, apply $(N)$.

Under local alternatives we have

$$m_2 = m_1 + \frac{cs}{\sqrt{N}}.$$

In the expansion (4.2) – (4.11), also the last three terms become relevant now. The terms (4.2) – (4.6) are negligible also in this case, while (4.7) and (4.8) are the same as before. The terms (4.9) and (4.10) are also negligible under the alternative. For example,

$$\begin{aligned} & \sqrt{N} \int \int W(x)[m_1(x) - m_2(x)][\hat{F}(dx_1) - F(dx_1)]G(dx_2) \\ = & \; -c \int \int W(x)s(x)[\hat{F}(dx_1) - F(dx_1)]G(dx_2) \to 0 \end{aligned}$$

by the SLLN. We finally come to (4.11). But

$$\sqrt{N} \int \int W(x)[m_1(x) - m_2(x)]F(dx_1)G(dx_2) = -c \int W(x)s(x)H(dx) = \mu,$$

which is the desired noncentrality parameter. □

**Proof of Theorem 2.2.** According to Theorem 2.1 it remains to study the distributional behavior of

$$n_1^{1/2} \sum_{i=1}^{n_1} (Y_{1i} - m_1(X_{1i})) \int W(x)W_{1i}(x)H(dx)$$

and

$$n_2^{1/2} \sum_{i=1}^{n_2} (Y_{2i} - m_2(X_{2i})) \int W(x)W_{2i}(x)H(dx).$$

By independence of the first and second sample it is sufficient to study each sum separately. For the first, say, put

$$\mathcal{F}_{ni} = \sigma(Y_{1j}, 1 \le j \le i, X_{1j}, 1 \le j \le n), n = n_1.$$

Since $\int W(x)W_{1i}(x)H(dx)$ is measurable w.r.t. $\mathcal{F}_{n,i-1}$ and $Y_{1i} - m_1(X_{1i})$ is conditionally centered, the summands

$$\xi_{ni} = n^{1/2}(Y_{1i} - m_1(X_{1i})) \int W(x)W_{1i}(x)H(dx)$$

form a martingale difference array. Brown's (1971) CLT for martingale difference arrays guarantees distributional convergence to $\mathcal{N}(0, \rho_1^2)$, where in our case

$$\rho_1^2 = \lim_{n\to\infty} n \sum_{i=1}^{n} \sigma_1^2(X_{1i}) \left[ \int W(x)W_{1i}(x)H(dx) \right]^2. \tag{4.61}$$

Each of the integrals is asymptotically equal to

$$
\begin{aligned}
&\int W(x)\frac{K\left(\frac{F(X_{1i})-F(x)}{h}\right)}{nh}\frac{h(x)}{f(x)}F(dx) \\
&= \frac{1}{nh}\int_0^1 W(F^{-1}(u))K\left(\frac{F(X_{1i})-u}{h}\right)\frac{h(F^{-1}(u))}{f(F^{-1}(u))}du \\
&= \frac{1}{n}\int_{(F(X_{1i})-1)/h}^{F(X_{1i})/h} W(F^{-1}(F(X_{1i})-wh))K(w)\frac{h(F^{-1}(F(X_{1i})-wh))}{f(F^{-1}(F(X_{1i})-wh))}dw \\
&\sim \frac{1}{n}\int_{-\infty}^{\infty} W(X_{1i})K(w)\frac{h(X_{1i})}{f(X_{1i})}dw = \frac{1}{n}W(X_{1i})\frac{h(X_{1i})}{f(X_{1i})}.
\end{aligned}
$$

Hence the limit in (4.61) equals

$$
\begin{aligned}
\lim_{n\to\infty} n\sum_{i=1}^{n}\sigma_1^2(X_{1i})n^{-2}W^2(X_{1i})\frac{h^2(X_{1i})}{f^2(X_{1i})} &= \mathbb{E}\left[\sigma_1^2(X_{11})W^2(X_{11})\frac{h^2(X_{11})}{f^2(X_{11})}\right] \\
&= \int \sigma_1^2(x)W^2(x)\frac{h^2(x)}{f^2(x)}F(dx), \text{ as desired}
\end{aligned}
$$

□

**Proof of Theorem 2.4.** Let $m_1 \neq m_2$ be fixed but arbitrary.Then Lemmas 4.4 – 4.8 again yield

$$
\begin{aligned}
\sqrt{N}\hat{T} &= \sqrt{N}\int W(x)[\hat{m}_1(x)-m_1(x)]H(dx) \\
&- \sqrt{N}\int W(x)[\hat{m}_2(x)-m_2(x)]H(dx) \\
&+ \sqrt{N}\int\int W(x)[m_1(x)-m_2(x)][\hat{F}(dx_1)-F(dx_1)]G(dx_2) \\
&+ \sqrt{N}\int\int W(x)[m_1(x)-m_2(x)]F(dx_1)[\hat{G}(dx_2)-G(dx_2)] \\
&+ \sqrt{N}\int W(x)[m_1(x)-m_2(x)]H(dx)+o_{\mathbb{P}}(1).
\end{aligned}
$$

According to Lemmas 4.9 and 4.10 the first two terms converge in distribution. The third integral converges in distribution to

$$
\sqrt{1-\lambda}\int\int W(x)[m_1(x)-m_2(x)]B^\circ(F(dx_1))G(dx_2)
$$

while the fourth goes to

$$
\sqrt{\lambda}\int\int W(x)[m_1(x)-m_2(x)]F(dx_1)B^\circ(G(dx_2)).
$$

Here $B^0$ is a Brownian Bridge. See Billingsley (1968). The last term, however, tends to

$$\begin{array}{ll} +\infty \quad , & \text{if } \int W(x)[m_1(x) - m_2(x)]H(dx) > 0 \\ -\infty \quad , & \text{if } \int W(x)[m_1(x) - m_2(x)]H(dx) < 0. \end{array}$$

Conclude that $\sqrt{N}|\hat{T}| \to \infty$ in probability whenever $\int W(m_1 - m_2)dH \neq 0$. Hence $\mathbb{P}_{H_1}(t = 1) \to 1$. $\square$

**Acknowledgements.** The work of Ramidha Srihera was supported by a Grant of the German Academic Exchange Service (DAAD). Her work was done while she was visiting the Math. Institute of the University of Giessen. The authors are also grateful to Gerrit Eichner for his assistance with the simulations.

## References

Billingsley, P. (1968). Convergence of probability measures. Wiley, New York.

Brown, B. M. (1971). Martingale central limit theorems. *Ann. Math. Statist.* **42** 59-66.

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

Dette, H. and Neumeyer, N. (2001). Nonparametric analysis of covariance. *Ann. Statist.* **29** 1361-1400.

Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642-669.

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

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

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

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

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

Koul, H. L. and Schick, A. (1997). Testing for the equality of two nonparametric regression curves. *J. Statist. Plann. Inf.* **65** 293-314.

KOUL, H. L. and SCHICK, A. (2003). Testing for superiority among two regression curves. *J. Statist. Plann. Inf.* **117** 15-33.

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

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

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

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

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

NEUMEYER, N. and DETTE, H. (2003). Nonparametric comparison of regression curves: an empirical process approach. *Ann. Statist.* **31** 880-920.

PARDO-FERNÁNDEZ, J. C., VAN KEILEGOM, I. and GONZÁLEZ-MANTEIGA, W. (2006). Testing for the equality of $k$ regression curves. *Statistica Sinica* To appear.

SCHEIKE, T. H. (2000). Comparison of non-parametric regression functions through their cumulatives. *Statist. Probab. Lett.* **46** 21-32.

STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595-620.

STRASSER, H. (1985). *Mathematical Theory of Statistics.* De Gruyter, Berlin.

STUTE, W. (1982). The oscillation behavior of empirical processes. *Ann. Probab.* **10** 86-107.

STUTE, W. (1984). Asymptotic normality of nearest neighbor regression function estimates. *Ann. Statist.* **12** 917-926.

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

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

Ramidha Srihera
Department of Biostatistics
Faculty of Public Health
Mahidol University
420/1 Rajavithi Road
Phayathai Bangkok 10400
Thailand
srihera_r@yahoo.com

Winfried Stute
Mathematical Institute
University of Giessen
Arndtstr. 2
D-35392 Giessen
Germany
Winfried.Stute@math.uni-giessen.de