Ann. Inst. Statist. Math.
Vol. 54, No. 1, 60–82 (2002)
©2002 The Institute of Statistical Mathematics

A TEST FOR ADDITIVITY IN NONPARAMETRIC REGRESSION

STEPHAN DERBORT, HOLGER DETTE AND AXEL MUNK

Ruhr-Universität Bochum, Fakultät für Mathematik, Universitätsstr. 150, D-44780 Bochum, Germany

(Received June 22, 1999; revised September 25, 2000)

Abstract. A simple consistent test of additivity in a multiple nonparametric regression model is proposed, where data are observed on a lattice. The new test is based on an estimator of the L^2-distance between the (unknown) nonparametric regression function and its best approximation by an additive nonparametric regression model. The corresponding test-statistic is the difference of a classical ANOVA style statistic in a two-way layout with one observation per cell and a variance estimator in a homoscedastic nonparametric regression model. Under the null hypothesis of additivity asymptotic normality is established with a limiting variance which involves only the variance of the error of measurements. The results are extended to models with an approximate lattice structure, a heteroscedastic error structure and the finite sample behaviour of the proposed procedure is investigated by means of a simulation study.

Key words and phrases: Additive models, dimension reduction, test of additivity.

1. Introduction

Consider the nonparametric regression model

$$Y = Y(t) = g(t) + \varepsilon(t) \tag{1.1}$$

where Y is the response variable, $t = (t_1, \dots, t_d)^T$ is a d-dimensional predictor variable and ε is an unobserved error term with mean 0 and variance $\sigma^2 > 0$. The problem of estimating g has a huge literature associated with it. In recent years nonparametric estimators such as kernel or nearest neighbour methods have become increasingly popular, especially in problems where the predictor is one-dimensional. It is well known that with higher-dimensional predictors these methods become less attractive because the efficiency of linear smoothers decreases with an increasing dimension d.

One possibility to avoid problems caused by the so-called "curse of dimensionality" is to approximate the regression function g by a sum of d univariate functions, one for each dimension of t (see Friedman and Stuetzle (1981)). Buja *et al.* (1989) discussed in their landmark paper the use of the most simple projection pursuit regression, i.e. additive models

$$Y = \sum_{j=1}^{d} g_j(t_j) + \varepsilon, \tag{1.2}$$

where $g_1, \dots, g_d$ are unknown smooth functions (see also Hastie and Tibshirani (1990)) A theoretical justification of this approach was given by Stone (1985) who showed that

in the model (1.2) the regression function $g(t) = \sum_{j=1}^{d} g_j(t_j)$ can be estimated with the same rate of the estimation error as in the univarate case. In other words high dimensional nonparametric regression models remain tractable efficiently when the additive structure (1.2) is justified.
For these reasons the application of the additive model (1.2) should be accompanied by a proper "model check of additivity". The problem of testing additivity has found considerable interest in the recent literature. Hastie and Tibshirani (1990) proposed a Tukey-type test of additivity which arises from considering specific interactions of the form $\gamma \prod_{j=1}^{k} g_{i_j}(t_{i_j})$ $(1 \leq i_1 < \cdots < i_k \leq d)$ and testing if $\gamma = 0$ (see also Tukey (1949)). Barry (1993) constructed a likelihood ratio test for the additivity of a regression function in a slightly different regression model, which is based on the local residuals $Y_{i+1,j+1} + Y_{i,j} - Y_{i+1,j} - Y_{i,j+1}$. Eubank *et al.* (1995) reanalyzed the Tukey-type test and showed that it is inconsistent against certain reasonable departures from additivity. In the same paper these authors proposed a second test of additivity which is consistent against essentially any departure from additivity.

In this paper an alternative consistent test of additivity is proposed which is constructed from an estimator for the L^2-distance between the regression function g and its best approximation by an additive model. We propose a class of test-statistics which are based on differences between a classical ANOVA type statistic in a two-way layout with one observation per cell and variance estimators in a homoscedastic nonparametric regression model. For this class which are quadratic forms in the responses we show asymptotic normality under the hypothesis of additivity. This allows the construction of a very intuitive and simple consistent test of additivity which uses the critical values obtained from the standard normal distribution. In Section 2 we state the basic assumptions and main results for the case of a two dimensional predictor, for which the basic ideas of our approach are most transparent. Section 3 presents a small simulation study in order to investigate and compare the finite sample performance of the test proposed in this paper with the methods introduced by Barry (1993) and Eubank *et al.* (1995). In summary, we found that no test performs uniformly better than the others. The power depends heavily on the particular alternative.

Section 4 discusses the extension of our methods to high-dimensional predictors (i.e. $d \geq 3$). If $d \geq 3$ it becomes apparent that the particular choice of an estimatior of the variance σ^2 is crucial for the performance of the proposed test. Guidelines for the selection of an appropriate variance estimator in order to maximize power and to maintain the nominal level with high accuracy are given. In Section 5 we extend our approach to the case of models with nonconstant intrinsic variability $\text{Var}[Y(t)] = \sigma^2(t)$ which is certainly in many applications a more realistic assumption. In the heteroscedastic model estimators, which differ from those used in the constant variance case are required. Roughly speaking, it turns out that the price for a more general model in the above sense is a drastic loss in efficiency. Finally, technical details and proofs are given in Section 6.

We conclude this section by noting that the test proposed in this paper does not require an explicit estimation of the regression and variance function. Only estimators of certain integrals of the regression function g and the variance σ^2 are required and as a consequence the new test does not depend on a choice of kernel estimators and bandwidth. The occuring statistics can be written as quadratic forms, which, together with the asymptotic normality, makes the performance of the test computationally rather simple.

2. Testing additivity: theory

In order to give a transparent presentation of the basic idea of this paper we consider throughout this section the case of a two-dimensional predictor. We assume that the response variable is observed over a lattice $\{(t_{1i}, t_{2j})\}_{i=1,\ldots,n_1}^{j=1,\ldots,n_2}$ on $[0,1]^2$ as it is often the case with image data, i.e.

$$Y_{ij} = g(t_{1i}, t_{2j}) + \varepsilon_{ij} \qquad i = 1, \ldots, n_1; j = 1, \ldots, n_2, \tag{2.1}$$

where $\varepsilon_{11}, \varepsilon_{12}, \ldots, \varepsilon_{n_1 n_2}$ are independent random variables with mean 0, variance $\mathrm{Var}(\varepsilon_{ij}) = \sigma^2$ $(i = 1, \ldots, n_1; j = 1, \ldots, n_2)$ and existing fourth moments. The regression function g is assumed to be Lipschitz-continuous of order $\gamma > \frac{1}{2}$, i.e.

$$g \in \mathrm{Lip}_\gamma[0,1]^2 \tag{2.2}$$

where

$$\mathrm{Lip}_\gamma[0,1]^2 := \{f : [0,1]^2 \to I\!R \mid \exists L > 0 : |f(x) - f(y)| \leq L\|x - y\|_\infty^\gamma\}.$$

For the design points $\{(t_{1i}, t_{2j})\}_{i=1,\ldots,n_1}^{j=1,\ldots,n_2}$ we assume a Sacks-Ylvisacker condition (see Sacks and Ylvisacker (1970))

$$\int_0^{t_{ij}} h_i(u)du = \frac{j}{n_i} \qquad j = 1, \ldots, n_i; i = 1, 2 \tag{2.3}$$

where $h_1, h_2 : [0,1] \to I\!R$ are positive Lipschitz continuous functions of order $\gamma > \frac{1}{2}$ (i.e. $h_i \in \mathrm{Lip}_\gamma[0,1] (i = 1,2)$) which are called the design densities. Observe that this includes the equidistant grid if $h_i \equiv 1$ $(i = 1, 2)$. Defining $h(t_1, t_2) = h_1(t_1)h_2(t_2)$ as the density of the corresponding product measure we introduce

$$L^2[0,1]^2 := \left\{ f : [0,1]^2 \to I\!R \,\middle|\, \int_{[0,1]^2} f^2(t)h(t)dt < \infty \right\}$$

as the Hilbert space of square integrable functions with respect to the product density h. Similarly, let $L_j^2[0,1]$ denote the corresponding Hilbert spaces of square integrable functions (on the unit interval $[0,1]$) with respect to the marginal density h_j $(j = 1,2)$. As a measure of additivity we use

$$\begin{aligned} M^2 = \inf \Bigg\{ \int_0^1 \int_0^1 & [g(t_1, t_2) - g_1(t_1) - g_2(t_2)]^2 \\ & \cdot h(t_1, t_2) dt_1 dt_2 \mid g_j \in L_j^2[0,1], j = 1, 2 \Bigg\}, \end{aligned} \tag{2.4}$$

which is the minimal distance between the regression function g and all additive regression functions g of the form $g(t_1, t_2) = g_1(t_1) + g_2(t_2)$. The hypothesis of an additive model (1.2) can now be rewritten as

$$H_0 : M^2 = 0. \tag{2.5}$$

The following result gives an explicit representation of M^2 and is proved in Section 6.

LEMMA 2.1. *Let $g \in L^2[0,1]^2$, then*

$$M^2 = \int_{[0,1]^2} g^2(t)h(t)dt - \int_0^1 \left[\int_0^1 g(t_1,t_2)h_1(t_1)dt_1\right]^2 h(t_2)dt_2 - \int_0^1 \left[\int_0^1 g(t_1,t_2)h_2(t_2)dt_2\right]^2 h(t_1)dt_1 + \left[\int_{[0,1]^2} g(t)h(t)dt\right]^2 \tag{2.6}$$

and the minimum is attained for the functions

$$g_1(t_1) = \int_0^1 g(t_1,t_2)h_2(t_2)dt_2 - c_1$$

$$g_2(t_2) = \int_0^1 g(t_1,t_2)h_1(t_1)dt_1 - c_2$$

where c_1, c_2 are arbitrary constants satisfying

$$c_1 + c_2 = \int_{[0,1]^2} g(t)h(t)dt.$$

If $g \in \mathrm{Lip}_\gamma[0,1]^2$ for some $\gamma > 0$ then g_1, $g_2 \in \mathrm{Lip}_\gamma[0,1]^2$.

The test proposed in this paper will be based on an estimator of the minimal distance M^2 in (2.6). This will be done by estimating all integrals occuring in (2.6) separately, i.e.

$$\hat{M}_{11}^2 = \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} Y_{ij}^2, \qquad \hat{M}_{12}^2 = \frac{1}{n_1^2 n_2}\sum_{j=1}^{n_2}\left(\sum_{i=1}^{n_1} Y_{ij}\right)^2$$

$$\hat{M}_{21}^2 = \frac{1}{n_1 n_2^2}\sum_{i=1}^{n_1}\left(\sum_{j=1}^{n_2} Y_{ij}\right)^2, \qquad \hat{M}_{22}^2 = \left(\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} Y_{ij}\right)^2.$$

Now the estimator of M^2 is defined by

$$\tilde{M}^2 = \hat{M}_{11}^2 - \hat{M}_{12}^2 - \hat{M}_{21}^2 + \hat{M}_{22}^2 = \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot})^2 \tag{2.7}$$

where $\bar{Y}_{i\cdot} = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_{ij}$, $\bar{Y}_{\cdot j} = \frac{1}{n_1}\sum_{i=1}^{n_1} Y_{ij}$ and $\bar{Y}_{\cdot\cdot} = \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} Y_{ij}$. Observe that the second expression in (2.7) equals the quadratic form for the error term in a complete two–way ANOVA with one observation per cell. This allows a very simple computation of $\tilde{M}^2$ in any statistical software package such as SAS or SPSS. Note that the statistic $\tilde{M}^2$ also appears in Tukey's (1949) test. The following Lemma shows that $\tilde{M}^2$ is a consistent estimator for the sum of the minimal distance M^2 and the variance of the error distribution.

LEMMA 2.2. *If the assumptions* (2.1)–(2.3) *are satisfied and $n_1 \to \infty$, $n_2 \to \infty$ such that $n_1/n_2 \to \lambda \in (0,\infty)$, then*

$$\mathrm{Var}[\tilde{M}^2] = O((n_1 n_2)^{-1})$$

$$E[\tilde{M}^2] = M^2 + \frac{(n_1-1)(n_2-1)}{n_1 n_2}\sigma^2 + O(n_1^{-\gamma}).$$

Especially, if the null hypothesis of additivity (2.5) *is valid we have*

$$E[\tilde{M}^2] = \left(1 - \frac{1}{n_1}\right)\left(1 - \frac{1}{n_2}\right)\sigma^2.$$

Taking into account Lemma 2.2 it is reasonable to define as an estimator for the minimal distance in (2.4)

$$\begin{aligned}\hat{M}^2_{l,\alpha} &= \tilde{M}^2 - r_{n_1,n_2}\hat{\sigma}^2_{l,\alpha} \\ &= \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot})^2 - r_{n_1,n_2}\hat{\sigma}^2_{l,\alpha}\end{aligned} \tag{2.8}$$

where

$$r_{n_1,n_2} = 1 - \frac{1}{n_1} - \frac{1}{n_2} + \frac{1}{n_1 n_2}$$

and $\hat{\sigma}^2_{l,\alpha}$ is the following variance estimator

$$\hat{\sigma}^2_{l,\alpha} = \frac{1}{8(n_1 - 2l)(n_2 - 2l)}\sum_{i=l+1}^{n_1-l}\sum_{j=l+1}^{n_2-l}\sum_{r=-1}^{1}\sum_{s=-1}^{1}\left(\sum_{k=0}^{l}\alpha_k Y_{i+kr,j+ks}\right)^2. \tag{2.9}$$

Here $\alpha = (\alpha_0, \ldots, \alpha_l)$ is a so called "differerence sequence" satisfying $\sum_{k=0}^{l}\alpha_k = 0$; $\sum_{k=0}^{l}\alpha_k^2 = 1$ (see Hall *et al.* (1990)) and (2.9) is a generalization of the variance estimators of Rice (1984) and Gasser *et al.* (1986) to the case of a multiple predictor. Similar estimators were also considered by Hall *et al.* (1991). The choice of the weights $\alpha_0, \ldots, \alpha_l$ will be important for the finite sample and asymptotic behaviour of the proposed test and will be discussed in Sections 3 and 4 more carefully. We also note that the case $l = 1$ yields the variance estimator used by Eubank *et al.* (1995) in their simulation study, i.e.

$$\hat{\sigma}^2 = \frac{1}{16(n_1 - 2)(n_2 - 2)}\sum_{i=2}^{n_1-1}\sum_{j=2}^{n_2-1}\sum_{r=-1}^{1}\sum_{s=-1}^{1}(Y_{i,j} - Y_{i+r,j+s})^2.$$

In the following two theorems we will establish asymptotic normality of $\hat{M}_{l,\alpha}$ if the hypothesis of additivity is valid. The first result gives the asymptotic variance of $\hat{M}^2_{l,\alpha}$.

THEOREM 2.1. *If the null hypothesis* (2.5) *of additivity is valid*, (2.1)–(2.3) *are satisfied and* $n_1, n_2 \to \infty$; $n_1/n_2 \to \lambda \in (0, \infty)$, *then the estimator defined in* (2.8) *satisfies*

$$\lim_{n_1,n_2\to\infty} n_1 n_2 \operatorname{Var}(\hat{M}^2_{l,\alpha}) = \sigma^4 \sum_{r=1}^{l}\left(\sum_{k=0}^{l-r}\alpha_k\alpha_{k+r}\right)^2 =: \mu^2. \tag{2.10}$$

THEOREM 2.2. *If the null hypothesis* (2.5) *of additivity is valid*, (2.1)–(2.3) *are satisfied and* $n_1, n_2 \to \infty$; $n_1/n_2 \to \lambda \in (0, \infty)$, *then the estimator defined in* (2.8) *satisfies*

$$\sqrt{n_1 n_2}\hat{M}^2_{l,\alpha} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mu^2) \tag{2.11}$$

where the asymptotic variance μ^2 *is defined by* (2.10).

Remark 2.1. A careful inspection of the proofs in Section 6 shows that Theorems 2.1 and 2.2 can be extended to designs with an asymptotically approximate lattice structure in the sense of Sacks and Ylvisacker (1970), i.e.

$$\int_0^{t_{ij}} h_i(u)du = \frac{j}{n_i} + o\left(\frac{1}{n_i}\right) \qquad j = 1, \ldots, n_i; i = 1, 2.$$

Remark 2.2. We would like to conclude this section with a general remark concerning the relationship of our statistic to classical ANOVA and maximum likelihood theory. In a model with replications of measurements Y_{ijk}, $i = 1, \ldots, n_1$, $j = 1, \ldots, n_2$, $k = 1, \ldots, K, K \geq 2$, such that $Y_{ijk} \sim^{i.i.d} Y_{ij}$, where Y_{ij} was defined in (2.1), a F-test for no interaction (additivity) is obtained by means of the F-statistic

$$F_{(n_1-1)(n_2-1),n_1n_2(K-1)} = \frac{\frac{1}{(n_1-1)(n_2-1)} \sum_{i,j}(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})^2}{\sigma_K^2} \tag{2.12}$$

where σ_K^2 denotes the UMVU estimator for the variance σ^2 in the general homoscedastic model, i.e.

$$\sigma_K^2 = \frac{1}{n_1n_2(K-1)} \sum_{i,j,k}(Y_{ijk} - \bar{Y}_{ij\cdot})^2.$$

Here the estimator in the nominator equals the norm of the residuals obtained by projecting orthogonally the averages $Y_{\cdot j\cdot}$ onto the subspace of additive vectors (functions) on the grid $\{t_{1i}, t_{2j}\}_{i=1,\ldots,n_1}^{j=1,\ldots,n_2}$. Note that this projection remains estimable for case $K = 1$, i.e. without replications, which yields (up to a scaling factor of order $o(\frac{1}{n_1n_2})$) the statistic $\tilde{M}^2$ defined in (2.7). Hence, the presented theory entails in particular the weak convergence of the nominator of the F-statistic in (2.12) as $n_1, n_2 \to \infty$ provided the smoothness conditions (2.2) and (2.3) are satisfied. The denominator in (2.12), however, estimates σ^2 by projecting on the subspace generated by the replications which is feasible only, if $K \geq 2$. Hence, if $K = 1$ this approach is no more valid and an estimator of σ^2 is required without the assumption of additivity. This could be done, for example, by one of the estimators presented in (2.9). Note, however, that even under a homoscedastic model with normal errors it is not possible to calculate the distribution of the resulting test statistic explicitly, because it will still depend on the unknown additive components g_1, g_2 albeit the dependency on σ^2 vanishes. Furthermore, it can be seen from this approach that the resulting test will be very sensitive against departures from homoscedasticity, due to the ANOVA type character of the test statistic (cf. Jørgensen (1993) p. 192). In contrast, our test is able to deal with heteroscedasticity provided more refined variance estimators are used (cf. Section 5).

3. A test for additivity and simulation results

3.1 *A test for additivity*

The results of Section 2 lead to a very simple test of additivity in the nonparametric regression (1.1). The hypothesis (2.5) is rejected if

$$\sqrt{n_1n_2}\frac{\hat{M}_{l,\alpha}^2}{\hat{\sigma}_{l,\alpha}^2} > u_{1-\alpha}\beta_{l,\alpha} \tag{3.1}$$

where

$$\beta_{l,\alpha}^2 = \sum_{r=1}^{l} \left(\sum_{k=0}^{l-r} \alpha_k \alpha_{k+r} \right)^2 ,$$

$\hat{M}_{l,\alpha}^2$, $\hat{\sigma}_{l,\alpha}^2$ are defined in (2.8) and (2.9), respectively, and $u_{1-\alpha}$ denotes the $(1-\alpha)$ quantile of the standard normal distribution. However, we found in a numerical study that the quality of the approximation by the asymptotic normal law can be significantly improved by a proper choice of the weights $\alpha_0, \ldots, \alpha_l$ and an additional bias correction motivated from a second order approximation.

3.2 *Choice of the difference sequence and improving the asymptotic accuracy*

In this paragraph we will discuss briefly how to choose the particular difference sequence $(\alpha_0, \ldots, \alpha_l)$ in a concrete application. Note, that the limiting variance in Theorem 2.2 and hence the (asymptotic) power of this test is minimized by an appropriate choice of the difference sequence $(\alpha_0, \ldots, \alpha_l)$. For fixed "order" l the asymptotic optimal weights $(\tilde{\alpha}_0, \ldots, \tilde{\alpha}_l)$ minimizing $\beta_{l,\alpha}^2$ were found by Hall *et al.* (1990) in the context of nonparametric regression with one predictor and are listed in Table 1 of their paper. Roughly speaking these weights minimize the asymptotic MSE of the estimator $\hat{\sigma}_{l,\alpha}^2$ which is dominated by its variance. For this sequence we obtain by a similar argument as in Hall *et al.* (1990)

$$\sum_{r=1}^{l} \left(\sum_{k=0}^{l-r} \tilde{\alpha}_k \tilde{\alpha}_{k+r} \right)^2 = \frac{1}{4l}. \tag{3.2}$$

However, for difference type estimators of the variance σ^2 in univariate nonparametric regression Seifert *et al.* (1993) and Dette *et al.* (1999) pointed out that the bias may be significant for small sample sizes although asymptotically neglegible. We will see in Section 4 that this effect is even more serious in the case of multiple predictors owing to the sparseness of the data in the predictor space. In order to reduce the bias Dette *et al.* (1999) proposed to use the difference sequence

$$\alpha_k^* = (-1)^k \binom{l}{k} \binom{2l}{l}^{-1/2} \qquad (k = 0, \ldots, l) \tag{3.3}$$

in the variance estimator (2.9), which will be used for $l = 2$ in the following simulation study. Observe that under H_0 the weights (3.3) lead to unbiased estimation of σ^2 in the case where g_1 and g_2 are polynomials of degree $\leq l-1$.

For the sequence $(\alpha_k^*)_{k=0,\ldots,l}$ we obtain by a straightforward calculation

$$\begin{aligned} \beta_{l,\alpha^*}^2 &= \sum_{r=1}^{l} \left(\sum_{k=0}^{l-r} \alpha_k \alpha_{k+r} \right)^2 = \sum_{r=1}^{l} \left\{ \sum_{k=0}^{l-r} \binom{l}{k} \binom{l}{k+r} \right\}^2 \binom{2l}{l}^{-2} \\ &= \frac{1}{2} \left\{ \binom{4l}{2l} \binom{2l}{l}^{-2} - 1 \right\} \end{aligned} \tag{3.4}$$

and the first few values are given by $\frac{1}{4}$ $(l = 1)$, $\frac{17}{36}$ $(l = 2)$ and $\frac{131}{200}$ $(l = 3)$. Although the (asymptotic) efficiency of the estimator $\hat{M}_{l,\alpha^*}^2$ decreases as l increases, simulation studies similar to those in Dette *et al.* (1999) show that the bias may affect the finite sample distribution heavily, such that an l larger than 1 becomes often necessary, in particular when f_1, f_2 are heavily oscillating functions. In accordance to this finding we can only recommend Hall *et al.*'s optimal difference sequence for very large sample sizes ($n_1 \geq 100, n_2 \geq 100$) and slowly oscillating regression functions. Moreover, we will demonstrate in Section 4 that in the case of more than two predictors the asympototic optimal weights of Hall *et al.* (1990) cannot be used in our approach, because for differentiable regression functions they only produce a bias of order n_1^{-2} independent of further smoothness of the regression function. In fact, for a high dimensional predictor it is required that the bias of the variance estimator is of order $o(n_1^{-d/2})$ which is achieved the difference sequence (3.3) for sufficiently smooth regression functions, where $d \geq 2$ reflects the dimension of the predictor.

In order to obtain a reasonable small bias and asymptotic variance in the two dimensional case we used the estimator with weight (3.3) and order $l = 2$ in our simulation study. We also replaced the limit value β^2_{l,α^*} by its finite sample value $\tilde{\beta}^2_{l,\alpha^*}$ which can be obtained by a straightforward but tedious calculation as

$$\begin{aligned}(3.5)\quad \tilde{\beta}^2_{2,\alpha^*}(n_1,n_2) = {} & \frac{2(n_1-1)^2(n_1-1)^2}{n_1^2n_2^2} + \frac{n_1n_2}{(n_1-4)(n_2-4)} \\ & \times \left[\frac{89}{36} - \frac{487}{288}\left\{\frac{1}{n_1-4} + \frac{1}{n_2-4}\right\} + \frac{335}{192}\frac{1}{(n_1-4)(n_2-4)}\right] \\ & - \frac{1}{n_1n_2}\left[4(n_1-4)(n_2-4) + 13\{n_1+n_2-8\} + \frac{136}{3}\right. \\ & \left. +16\left\{\frac{1}{n_1-4} + \frac{1}{n_2-4}\right\} + \frac{152}{3}\frac{1}{n_1-4}\frac{1}{n_2-4}\right]\end{aligned}$$

whenever $n_1, n_2 \geq 7$. This modification is necessary, because, although $\tilde{\beta}^2_{2,\alpha^*}(n_1,n_2) \to \beta^2_{2,\alpha^*} = \frac{17}{36}$ as $n_1, n_2 \to \infty$, the approximation of $\tilde{\beta}^2_{2,\alpha^*}(n_1,n_2)$ by its limit is rather poor in finite sample situations (e.g. $\tilde{\beta}^2_{2,\alpha^*}(7,7) \approx 5.87$, $\tilde{\beta}^2_{2,\alpha^*}(10,10) \approx 3.23$). For this reason we reject the hypothesis of additivity whenever

$$(3.6)\qquad \sqrt{n_1n_2}\frac{\hat{M}^2_{2,\alpha^*}}{\hat{\sigma}^2_{2,\alpha*}} > u_{1-\alpha}\tilde{\beta}_{2,\alpha^*}(n_1,n_2)$$

where $\tilde{\beta}^2_{2,\alpha^*}(n_1,n_2)$ is defined by (3.5). We finally note that an alternative way of determining critical values is given by matching higher order (asymptotic) moments of the left hand side of (3.6) with a $b(\chi^2_a - a)$ distribution, or by applying the critical value approximation proposed by Buckley and Eagleson (1988).

3.3 *A simulation study*

In the simulation study reported below we also considered the sample sizes $(n_1, n_2) = (5,5)$ and $(n_1, n_2) = (5,20)$ for which the finite sample variances have to be calculated separately, i.e.

$$(3.7)\qquad \tilde{\beta}_{2,\alpha^*}(5,5) \approx 2.93; \qquad \tilde{\beta}_{2,\alpha^*}(5,20) \approx 1.61.$$

Our first table shows the level of the test for various functions g representing the hypothesis of additivity, i.e.

$$
\begin{aligned}
g_1(t_1,t_2) &= 0 \\
g_2(t_1,t_2) &= t_1 + t_2 \\
g_3(t_1,t_2) &= e^{t_1} + \sin(\pi t_2) \\
g_4(t_1,t_2) &= \sin(\pi t_1) + \sin(\pi t_2) \\
g_5(t_1,t_2) &= e^{t_1} + e^{t_2}.
\end{aligned} \tag{3.8}
$$

In our study we used the variances $\sigma^2 = 0.01, 0.25$ and 1 and normally distributed errors ε_{ij} (in accordance with Barry (1993) and Eubank *et al.* (1995)). For the design points we used a uniform grid $\{(\frac{i}{n_1}, \frac{j}{n_2})\}_{i=1,\ldots,n_1}^{j=1,\ldots,n_2}$ which corresponds to the uniform density in (2.3). As all cases yield similar results we only displayed results for $\sigma^2 = 1$. Table 1 shows the relative proportion of rejections of the test (3.6) for the levels 5% and 2.5% based on 5000 simulation runs. We observe a rather accurate approximation of the theoretical level for all regression functions in (3.8).

For a study of the power of the proposed test we considered a similar setup as in Barry (1993) and Eubank *et al.* (1995). More precisely, for the sake of comparison we use the seven alternatives from these references, namely

$$
\begin{aligned}
g_6(t_1,t_2) &= t_1 t_2 \\
g_7(t_1,t_2) &= \exp(5(t_1+t_2))/(1+\exp(5(t_1+t_2))) - 1 \\
g_8(t_1,t_2) &= 0.5(1+\sin(2\pi(t_1+t_2))) \\
g_9(t_1,t_2) &= 64(t_1t_2)^3(1-t_1t_2)^3 \\
g_{10}(t_1,t_2) &= \text{product of two sawtooths} \\
g_{11}(t_1,t_2) &= I\{t_1 > 0.5, t_2 > 0.5\} \\
g_{12}(t_1,t_2) &= (t_1+t_2)/2 + (1 \text{ outlier})
\end{aligned} \tag{3.9}
$$

(For the description of the sawtooths and the outlier see Barry (1993)). Table 2 contains the simulated power for the different sample sizes and alternatives of the test (3.6) with respect to a 5% level based on 5000 simulation runs. The table also shows the corresponding results of Barry's test and the variance was chosen as $\sigma^2 = 1, 0.25$ and 0.01 (see Barry (1993)). Comparing Table 2 with the results of Barry (1993) and Eubank *et al.* (1995) we see that no test performs uniformly better than any other test. While Barry's and Eubank *et al.*'s tests are superior for the alternative g_6,g_7 and g_{11}, the test proposed in this paper performs better for the alternatives g_8, g_9 g_{10} and g_{12}. In summary no "clear winner" among these tests could be found in our study. Finally, it is

Table 1. Relative proportion of rejections of the test (3.6) for various regression functions representing the hypothesis of additivity. The variance of the error distribution is $\sigma^2 = 1$.

	$(n_1,n_2)=(5,5)$		$(n_1,n_2)=(10,10)$		$(n_1,n_2)=(5,20)$	
α	5%	2.5%	5%	2.5%	5%	2.5%
g_1	0.062	0.041	0.044	0.022	0.042	0.024
g_2	0.062	0.041	0.046	0.022	0.042	0.024
g_3	0.057	0.042	0.045	0.021	0.044	0.021
g_4	0.048	0.033	0.049	0.021	0.044	0.022
g_5	0.060	0.039	0.043	0.021	0.044	0.022

Table 2. Simulated power of the test (3.6) ($\alpha = 5\%$) for different alternatives, variances and sample sizes. The numbers in brackets show the corresponding power of Barry's test and are taken from Barry (1993).

g	σ	$(n_1, n_2) = (5, 5)$	$(n_1, n_2) = (20, 5)$	$(n_1, n_2) = (20, 20)$
	0.1	0.648 (0.984)	0.939 (1.000)	1.000 (1.000)
g_6	0.5	0.072 (0.113)	0.083 (0.362)	0.747 (0.892)
	1.0	0.063 (0.073)	0.068 (0.105)	0.318 (0.340)
	0.1	0.118 (0.656)	0.339 (1.000)	0.874 (1.000)
g_7	0.5	0.079 (0.083)	0.153 (0.182)	0.447 (0.689)
	1.0	0.052 (0.059)	0.059 (0.080)	0.151 (0.210)
	0.1	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
g_8	0.5	0.367 (0.305)	1.000 (0.998)	1.000 (1.000)
	1.0	0.153 (0.099)	0.647 (0.595)	0.977 (0.999)
	0.1	0.720 (0.184)	1.000 (0.666)	1.000 (1.000)
g_9	0.5	0.205 (0.069)	0.455 (0.074)	0.891 (0.115)
	1.0	0.083 (0.057)	0.119 (0.063)	0.219 (0.054)
	0.1	0.334 (0.181)	0.798 (0.720)	0.999 (1.000)
g_{10}	0.5	0.067 (0.054)	0.086 (0.069)	0.127 (0.123)
	1.0	0.056 (0.043)	0.069 (0.058)	0.094 (0.080)
	0.1	0.456 (1.000)	1.000 (1.000)	1.000 (1.000)
g_{11}	0.5	0.267 (0.470)	0.586 (0.988)	0.912 (1.000)
	1.0	0.134 (0.221)	0.392 (0.567)	0.592 (0.995)
	0.1	0.313 (0.167)	0.636 (0.127)	0.532 (0.127)
g_{12}	0.5	0.208 (0.138)	0.181 (0.103)	0.135 (0.108)
	1.0	0.065 (0.099)	0.049 (0.122)	0.056 (0.111)

worthwhile to mention that in contrast to Barry (1993) the procedure proposed in the present paper does not require a normal distribution for the noise of the observations, which is an important advantage with respect to any practical application. Moreover, in contrast to Eubank *et al.* (1995) and Barry (1993) we do not need to simulate critical values because we can utilize the asymptotic normality as in Theorem 2.2 together with the correction in (3.5) and (3.6).

4. A high dimensional predictor

The procedure developed in Section 2 can also be extended to the case of more than two predictors. To this end we define as measure of additivity in the regression model (1.1)

$$M^2 = \inf \left\{ \int_0^1 \cdots \int_0^1 \left[g(t_1, \ldots, t_d) - \sum_{j=1}^d g_j(t_j) \right]^2 \right.$$

$$\left. \cdot h(t_1,\dots,t_d)dt_1\dots dt_d \mid g_j \in L_j^2[0,1] \right\}$$

where $h(t_1,\dots,t_d) = \prod_{j=1}^d h_j(t_j)$, $L_j^2[0,1] = \{f : \int_0^1 f^2(t_j)h_j(t_j)dt_j < \infty\}$ and $h_j, j = 1,\dots,d$ satisfy condition (2.2). Observe that again

$$H_0 : M^2 = 0 \quad \Leftrightarrow \quad \exists g_1,\dots,g_d : g = \sum_{j=1}^d g_j$$

where the equality on the right hand side holds almost surely with respect to the Lebesgue measure on the d-dimensional cube. We obtain by similar arguments as given in the proof of Lemma 2.1 the following result:

LEMMA 4.1. *Let $g \in L^2[0,1]^d$, then*

$$\begin{aligned} M^2 = &\int_{[0,1]^d} g^2(t)h(t)dt \\ &-\sum_{j=1}^d \int_0^1 \left[\int_0^1 \cdots \int_0^1 g(t_1,\dots,t_d) \prod_{\substack{i=1\\ i\neq j}}^d h_i(t_i)dt_1\dots dt_{j-1}dt_{j+1}\dots dt_d\right]^2 h(t_j)dt_j \\ &+(d-1)\left[\int_{[0,1]^d} g(t)h(t)dt\right]^2 \end{aligned}$$

and the minimum is attained for any function

$$g_l(t_l) = \int_0^1 \cdots \int_0^1 g(t_1,\dots,t_d) \prod_{\substack{i=1\\ i\neq l}}^d h_i(t_i)dt_1\dots dt_{l-1}dt_{l+1}\dots dt_d - S_l, \qquad l = 1,\dots,d,$$

such that $\sum_{l=1}^d S_l = (d-1)\int_0^1 \cdots \int_0^1 g(t)h(t)dt$. Moreover, if $g \in \mathrm{Lip}_\gamma[0,1]^d$ we have $g_l \in \mathrm{Lip}_\gamma[0,1]$.

Assume that we have $n = n_1 \cdots n_d$ observations over a lattice $\{(t_{1i_1},\dots,t_{di_d}) \mid i_j = 1,\dots,n_j; j = 1,\dots,d\}$ (where $t_{1i_1},\dots,t_{di_d}$ satisfy a similar condition corresponding to (2.3))

$$Y_{i_1,\dots,i_d} = g(t_{1i_1},\dots,t_{di_d}) + \varepsilon_{i_1,\dots,i_d},$$

then the estimator of M^2 is defined by

$$\tilde{M}^2 = \frac{1}{\prod_{j=1}^d n_j} \sum_{i_1=1}^{n_1} \cdots \sum_{i_d=1}^{n_d} \left\{ Y_{i_1,\dots,i_d} - \sum_{j=1}^d {}_j\bar{Y}_{\cdot\cdots\cdot} + (d-1)\bar{Y}_{\cdot\cdots\cdot} \right\}^2 \tag{4.1}$$

where

$${}_j\bar{Y}_{\cdot\cdots\cdot} = \frac{1}{\prod_{i\neq j} n_i} \sum_{i_1=1}^{n_1} \cdots \sum_{i_{j-1}=1}^{n_{j-1}} \sum_{i_{j+1}=1}^{n_{j+1}} \cdots \sum_{i_d=1}^{n_d} Y_{i_1,\dots,i_{j-1},j,i_{j+1},\dots,i_d}, \tag{4.2}$$

$$\bar{Y}_{\cdot\cdots\cdot} = \frac{1}{\prod_{j=1}^d n_j} \sum_{i_1=1}^{d} \cdots \sum_{i_d=1}^{d} Y_{i_1,\dots,i_d}. \tag{4.3}$$

A straightforward calculation shows that under the hypothesis of additivity (1.2)

$$E[\tilde{M}^2] = \sigma^2 r_{n_1,\dots,n_d} \tag{4.4}$$

where

$$r_{n_1,\dots,n_d} = \left(1 - \sum_{j=1}^{d} \prod_{i\neq j} \frac{1}{n_i} + \frac{d-1}{n}\right).$$

Using similar arguments as in the proof of Theorem 2.2 it can be shown that $\sqrt{n}(\tilde{M}^2 - \sigma^2 r_{n_1,\dots,n_d})$ is asymptotically normal with mean 0.

The construction of the test for additivity now needs a variance estimator which is consistent of order $o(n^{1/2})$. To this end define the d-dimensional analogue of (2.9)

$$\hat{\sigma}^2_{l,\alpha} = \frac{1}{(3^d - 1)\prod_{j=1}^{d}(n_j - 2l)} \cdot \sum_{i_1=l+1}^{n_1-l} \cdots \sum_{i_d=l+1}^{n_d-l} \sum_{r_1=-1}^{1} \cdots \sum_{r_d=-1}^{1} \left\{\sum_{k=0}^{l} \alpha_k Y_{i_1+kr_1,\dots,i_d+kr_d}\right\}^2 \tag{4.5}$$

where $\alpha = (\alpha_0, \dots, \alpha_l)$ is a difference sequence, i.e. $\sum_{k=0}^{l} \alpha_k = 0, \sum_{k=0}^{l} \alpha_k^2 = 1$. The estimator $\hat{M}^2_{l,\alpha}$ of M^2 is defined in the same way as in Section 2. It can easily be shown that for an arbitrary difference sequence the order of the bias is $O(n^{-2/d})$ (provided the regression function is differentiable). It can further be proved that this upper bound is attained for the class of estimators considered by Hall *et al.* (1990), which makes them not applicable for our purposes when $d \geq 4$, because we again utilize the asymptotic normality of $\sqrt{n_1 \cdots n_d}\hat{M}^2_{l,\alpha}$ under H_0. Observe that for $d \geq 4$ it follows that $E[\sqrt{n_1 \cdots n_d}\hat{M}^2_{l,\tilde{\alpha}}]$ does not vanish asymptotically and hence bias (instead of variance) reduction by the choice of the difference sequence α becomes necessary. In the next lemma we show that for the weighting pattern α^* in (3.3) a better approximation can be achieved. The proof of the following result is straightforward, observing the fact that the sequence α^* is usually used for numerical differentiation, see Hall *et al.* (1990).

LEMMA 4.2. *If the regression function g is k times continuously differentiable, then the variance estimator $\hat{\sigma}^2_{l,\alpha^*}$ defined in (4.5) for the difference sequence α^* in (3.3) satisfies*

$$E[\hat{\sigma}^2_{l,\alpha^*}] = \sigma^2 + O(n^{-2(k\wedge l)/d}), \qquad \operatorname{Var}[\hat{\sigma}^2_{l,\alpha^*}] = O\left(\frac{1}{n}\right)$$

whenever $n_j \to \infty$ $(j = 1, \dots, d)$, $n_i/n_j \to \lambda_{ij} \in (0, \infty)$.

The construction of an appropriate test for the hypothesis of additivity is now straightforward. Define

$$\hat{M}^2_l = \tilde{M}^2 - r_{n_1,\dots,n_d}\hat{\sigma}^2_{l,\alpha^*}. \tag{4.6}$$

Then the following result can be proved by similar arguments as given in the proof of Theorem 2.2. The corresponding test is constructed following the lines of thought in Section 3.

THEOREM 4.1. *If the regression function g is k times continuously differentiable, such that $k \wedge l > \frac{d}{4}$ and $n_j \to \infty$ $(j = 1, \ldots, d)$, $n_i/n_j \to \lambda_{ij} \in (0, \infty)$, then under H_0 we have for the statistic defined in* (4.6)

$$\sqrt{n}\hat{M}_l^2 \to \mathcal{N}(0, \mu^2)$$

where

$$\mu^2 = \frac{4\sigma^4}{3^d - 1}\left\{\binom{2l}{l}^{-2}\binom{4l}{2l} - 1\right\}. \tag{4.7}$$

Remark 4.1. (*the high dimensional case*) From Theorem 4.1 we draw in particular (for suitably smooth g) that the minimal l, such that the assumption in Theorem 4.1 holds is given as

$$l^* = \left\lfloor \frac{d}{4} \right\rfloor + 1.$$

This gives for large d an approximation for the asymptotic variance in (4.7) as

$$\mu^2 \approx \sigma^4 \frac{\sqrt{2\pi d}}{3^d - 1}. \tag{4.8}$$

At a first glance this might seem rather curious because (4.8) implies that as the dimension d increases, the sample variance of $\sqrt{n}\hat{M}_l^2$ decreases. Observe, however, that in Theorem 4.1 the sample size $n = \prod_{i=1}^d n_i$ depends on d and increases exponentially fast.

5. Heteroscedastic errors

Up to now, we restricted our analysis to the homoscedastic case. In many cases, however, this assumption may not be satisfied and the following model is more appropriate

$$Y(t) = g(t) + \varepsilon(t), \qquad E[\varepsilon(t)] \equiv 0, \quad \mathrm{Var}[Y(t)] = \sigma^2(t), \tag{5.1}$$

where $\sigma^2(\cdot) \in L^2[0,1]^d$ denotes an unknown function describing the variability of the response Y. Throughout the sequel we restrict our considerations to the case $d = 2$, the case for arbitrary d is briefly mentioned at the end of this section. Let us introduce the estimators

$$\begin{aligned}\tilde{M}^2_{n_1-l,n_2-l} = {} & \frac{1}{(n_1-2l)(n_2-2l)}\sum_{i=l+1}^{n_1-l}\sum_{j=l+1}^{n_2-l} Y_{ij}^2 - \frac{1}{(n_1-2l)^2(n_2-2l)}\sum_{j=l+1}^{n_2-l}\left(\sum_{i=l+1}^{n_1-l} Y_{ij}\right)^2 \\ & - \frac{1}{(n_1-2l)(n_2-2l)^2}\sum_{i=l+1}^{n_1-l}\left(\sum_{j=l+1}^{n_2-l} Y_{ij}\right)^2 \\ & + \left(\frac{1}{(n_1-2l)(n_2-2l)}\sum_{i=l+1}^{n_1-l}\sum_{j=l+1}^{n_2-l} Y_{ij}\right)^2\end{aligned}$$

and

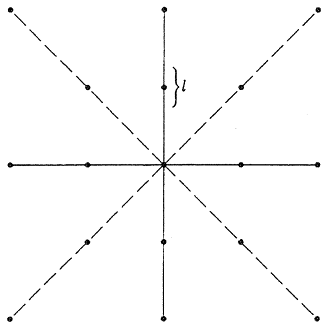

Fig. 1. The local residual estimators $\hat{\sigma}^2_{l,\alpha}$ and $\hat{\tau}^2_{l,\alpha}$. The solid lines give the directions used by $\hat{\tau}^2_{l,\alpha}$ while $\hat{\sigma}^2_{l,\alpha}$ uses all directions displayed in this figure.

$$\hat{\tau}^2_{l,\alpha} = \frac{1}{4} \frac{1}{(n_1 - 2l)(n_2 - 2l)} \sum_{i=l+1}^{n_1-l} \sum_{j=l+1}^{n_2-l} \sum_{\substack{r,s=-1 \\ r=0 \vee s=0}}^{1} \left(\sum_{k=0}^{l} \alpha_k Y_{i+kr,j+ks} \right)^2,$$

where $\alpha = (\alpha_0, \ldots, \alpha_l)$ is an arbitrary difference sequence defined as in Section 2. Observe that $\hat{\tau}^2_{l,\alpha}$ is constructed with local residuals where either $r = 0$ or $s = 0$ in contrast to $\hat{\sigma}^2_{l,\alpha}$ in (2.9) where also residuals on the "diagonals," such as $(r, s) = (1, 1)$ are involved. The difference between $\hat{\tau}^2_{l,\alpha}$ and $\hat{\sigma}^2_{l,\alpha}$ is illustrated in Fig. 1.

A straightforward calculation following the lines of the proof of Theorem 2.1 shows that in the case of $\sigma^2(\cdot) \equiv \sigma^2$ the estimator $\hat{\sigma}^2_{l,\alpha}$ is more efficent than $\hat{\tau}^2_{l,\alpha}$ (c.f. also the next theorem), however, in the heteroscedastic model (5.1) the "full" estimator $\hat{\sigma}^2_{l,\alpha}$ cannot be utilized because of its too large bias. Note, that this does not depend on the particular choice of α as long as residuals with $r \neq 0$ and $s \neq 0$ are encountered.

THEOREM 5.1. *Assume* (2.3), *model* (5.1) *and let* $g, \sigma^2 \in \mathrm{Lip}_\gamma[0,1]^2$, $\gamma > \frac{1}{2}$. *Then we have for any difference scheme* $(\alpha_0, \ldots, \alpha_l)$, $l \geq 1$, *under* H_0 *that*

$$\sqrt{n_1 n_2}(\tilde{M}^2_{n_1-l,n_2-l} - r_{l,n_1,n_2} \hat{\tau}^2_{l,\alpha}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2)$$

as $n_1, n_2 \to \infty$ *where*

$$\tau^2 = 2 \sum_{r=1}^{l} \left(\sum_{k=0}^{l-r} \alpha_k \alpha_{k+r} \right)^2 \int_0^1 \int_0^1 \sigma^4(t_1, t_2) h_1(t_1) h_2(t_2) dt_1 dt_2$$

and $r_{l,n_1,n_2} = \frac{(n_1-2l-1)(n_2-2l-1)}{(n_1-2l)(n_2-2l)}$.

In order to provide a test as in Section 3 it remains to estimate

$$\|\sigma^2\|^2 = \int_0^1 \int_0^1 \sigma^4(t_1, t_2) h_1(t_1) h_2(t_2) dt_1 dt_2 \tag{5.2}$$

in the asymptotic variance of Theorem 5.1.

LEMMA 5.1. *Assume that* (2.3) *holds and that* σ^2, $g \in \mathrm{Lip}_\gamma[0,1]^2$, $\gamma > 0$, *then it follows for any estimator*

$$\|\sigma^2\|^2_{l,\alpha,n} := \frac{1}{8(n_1 - 3l - 1)(n_2 - 3l - 1)} \sum_{i=l+1}^{n_1-2l-1} \sum_{j=l+1}^{n_2-2l-1} \sum_{r=-1}^{1} \sum_{s=-1}^{1}$$

$$\times \left(\sum_{k_1=0}^{l} \alpha_{k_1} Y_{i+k_1 r, j+k_1 s} \right)^2 \left(\sum_{k_2=0}^{l} \alpha_{k_2} Y_{i+k_2 r+l+1, j+k_2 s+l+1} \right)^2$$

such that $\alpha = (\alpha_0, \ldots, \alpha_l)$ *is a difference sequence*

$$\|\sigma^2\|^2_{l,\alpha,n} = \|\sigma^2\|^2 + O_p(n_1^{-\gamma}).$$

The proof of Lemma 5.1 is straightforward taking into account that the functions g, σ^2 are Lipschitz continuous and that the random variables

$$\sum_{k_1=0}^{l} \alpha_{k_1} Y_{i+k_1 r, j+k_1 s} \qquad \text{and} \qquad \sum_{k_2=0}^{l} \alpha_{k_2} Y_{i+k_2 r+l+1, j+k_2 s+l+1}$$

are independent for any $r, s \in \{-1, 0, 1\}$.

Remark 5.1. Comparing the limiting variance in Theorem 5.1 and in Theorem 2.2 shows that the "price" for a nonconstant variance is a loss in efficiency of 50% because we have to use the "reduced" estimator $\hat{\tau}^2_{l,\alpha}$ instead of the "full" estimator $\hat{\sigma}^2_{l,\alpha}$. This does not say that there is no way to estimate M^2 in the heteroscedastic case as efficient as in the homoscedastic case (just use $\hat{\tau}^2_{l,\alpha}$ instead of $\hat{\sigma}^2_{l,\alpha}$), it rather tells us that we have to take into account a loss in the accuracy of the normal approximation.

Remark 5.2. For the general case $d \geq 3$ define

$$\tilde{M}^2_{n_1-l,\ldots,n_d-l} = \frac{1}{\prod_{j=1}^{d}(n_j - 2l)} \sum_{i_1=l+1}^{n_1-l} \cdots \sum_{i_d=l+1}^{n_d-l} \left\{ Y_{i_1,\ldots,i_d} - \sum_{i=1}^{d} {}_j\bar{Y} + (d-1)\bar{Y}_{\cdot} \right\}^2$$

where ${}_j\bar{Y}$ and $\bar{Y}_{\cdot}$ are analoguously defined as in (4.2) and (4.3), i.e.

$${}_j\bar{Y} = \frac{1}{\prod_{i \neq j}(n_i - 2l)} \sum_{i_1=l+1}^{n_1-l} \cdots \sum_{i_{j-1}=l+1}^{n_{j-1}-l} \sum_{i_{j+1}=l+1}^{n_{j+1}-l} \cdots \sum_{i_d=l+1}^{n_d-l} Y_{i_1,\ldots,i_{j-1},j,i_{j+1},\ldots,i_d},$$

$$\bar{Y}_{\cdot} = \frac{1}{\prod_{j=1}^{d}(n_j - 2l)} \sum_{i_1=l+1}^{n_1-l} \cdots \sum_{i_d=l+1}^{n_d-l} Y_{i_1,\ldots,i_d}.$$

Let

$$\hat{\tau}^2_{l,\alpha,d} = (2d)^{-1} \frac{1}{\prod_{j=1}^{d}(n_j - 2l)} \sum_{i_1=l+1}^{n_1-l} \cdots \sum_{i_d=l+1}^{n_d-l} \underbrace{\sum_{r_1=-1}^{1} \cdots \sum_{r_d=-1}^{1}}_{\substack{\text{exactly } d-1 \text{ of the indices} \\ r_1,\ldots,r_d \text{ are } 0}}$$

$$\cdot \left\{ \sum_{k=0}^{l} \alpha_k Y_{i_1+kr_1,\ldots,i_d+kr_d} \right\}^2.$$

Then a careful inspection of the proof of Theorem 5.1 shows that for $g \in \mathrm{Lip}_\gamma[0,1]^d$,

$$E[\hat{\tau}^2_{l,\alpha,d}] = \frac{1}{\prod_{j=1}^d (n_j - 2l)} \Sigma_{n_1-l,\dots,n_d-l} + O(n^{-2\gamma/d})$$

and under H_0

$$E[\tilde{M}^2_{n_1-l,\dots,n_d-l}] = \frac{r_{l,n_1,\dots,n_d}}{\prod_{j=1}^d (n_j - 2l)} \Sigma_{n_1-l,\dots,n_d-l}$$

where

$$r_{l,n_1,\dots,n_d} = 1 - \sum_{j=1}^d \prod_{i \neq j} \frac{1}{n_i - 2l} + (d-1) \prod_{j=1}^d \frac{1}{n_j - 2l} \qquad \text{and}$$

$$\Sigma_{n_1-l,\dots,n_d-l} = \sum_{i_1=l+1}^{n_1-l} \cdots \sum_{i_d=l+1}^{n_d-l} \sigma^2(t_{1i_1}, \dots, t_{di_d})$$

(see the proof of Theorem 5.1). Hence the bias of $\hat{\tau}^2_{l,\alpha,d}$ turns out to be crucial for the validity of a limit law as in Theorem 5.1 and a bias correction by the polynomial weighting scheme α^* becomes necessary in the case $d \geq 4$.

THEOREM 5.2. *Assume* (2.3), *the model* (5.2) *and assume that* H_0 *holds.*
1. *Let* $d = 3$. *Then we have for* σ^2, $g \in \mathrm{Lip}_\gamma[0,1]^3$, $\gamma > \frac{3}{4}$ *that the limit law*

$$\sqrt{n_1 n_2 n_3}(\tilde{M}^2_{n_1-l,n_2-l,n_3-l} - r_{l,n_1,n_2,n_3} \hat{\tau}^2_{l,\alpha,3}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau_3^2)$$

holds for any difference scheme α, *where the limiting variance is given as* $\tau_3^2 = \frac{13}{3}\tau^2$.

2. *Let* $d \geq 4$. *Then we have for* σ^2, $g \in C^k[0,1]^d$ *and for any polynomial scheme* α_l^* *of the form* (3.3), *such that* $l \wedge k > \frac{d}{4}$

$$\sqrt{n_1 \cdots n_d}\{\tilde{M}^2_{n_1-l,\dots,n_d-l} - r_{l,n_1,\dots,n_d} \hat{\tau}^2_{l,\alpha,d}\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau_d^2) \tag{5.3}$$

where

$$\tau_d^2 = \frac{2}{d}\left\{\binom{2l}{l}^{-2}\binom{4l}{2l} - 1\right\} \int_0^1 \cdots \int_0^1 \sigma^4(t_1, \dots, t_d) h_1(t_1) \cdots h_d(t_d) dt_1 \cdots dt_d.$$

Remark 5.3. If we compare the limiting variance τ_d^2 with that in the case of constant variance in Theorem 4.1 we find that the asymptotic relative efficiency is given as $\mathrm{ARE} = \frac{2d}{3^d - 1}$.

6. Proofs

PROOF OF LEMMA 2.1. By Fubini's theorem M^2 can be obtained iteratively by minimizing at first the inner integral

$$\int_0^1 [g(t_1, t_2) - g_1(t_1) - g_2(t_2)]^2 h_1(t_1) dt_1$$

with respect to g_2. This gives

$$g_2^*(t_2) = \int_0^1 g(t_1, t_2)h_1(t_1)dt_1 - \int_0^1 g_1(t_1)h_1(t_1)dt_1. \tag{6.1}$$

In a second step we minimize

$$\int_0^1 [g(t_1, t_2) - g_1(t_1) - g_2^*(t_2)]^2 h_2(t_2)dt_2$$

with respect to g_1. Noting that this optimization is invariant with respect to shifts of $g_1$we obtain

$$g_1^*(t_1) = \int_0^1 g(t_1, t_2)h_2(t_2)dt_2 - \int_{[0,1]^2} g(t)h(t)dt.$$

A straightforward calculation now gives the required representation of M^2. The remaining part of the lemma is obvious and left to the reader. □

PROOF OF LEMMA 2.2. Let $I_k \in \mathbb{R}^{k\times k}$ denote the $k \times k$ unit matrix and $J_k \in \mathbb{R}^{k\times k}$ a matrix with all elements equal 1, define

$$P = I_{n_1} \otimes I_{n_2} - \frac{1}{n_1}J_{n_1} \otimes I_{n_2} - \frac{1}{n_2}I_{n_1} \otimes J_{n_2} + \frac{1}{n_1 n_2}J_{n_1} \otimes J_{n_2} \tag{6.2}$$

and $g_{ij} = g(t_{1i}, t_{2j}); i = 1, \ldots, n_1; j = 1, \ldots, n_2$. With these notations we have $\tilde{M}^2 = \frac{1}{n_1 n_2} Y^T P Y$ where $Y = (Y_{11}, Y_{12}, \ldots, Y_{n_1 n_2})^T \in \mathbb{R}^{n_1 n_2}$ denotes the vector of all responses. Obviously P is a projection of rank $(n_1 - 1)(n_2 - 1)$ and well known results about moments of quadratic forms (see Whittle (1960, 1964)) show

$$\begin{aligned} E[\tilde{M}^2] &= \frac{(n_1 - 1)(n_2 - 1)}{n_1 n_2}\sigma^2 + \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} g_{ij}^2 - \frac{1}{n_1^2 n_2}\sum_{j=1}^{n_2}\left(\sum_{i=1}^{n_1} g_{ij}\right)^2 \\ &\quad - \frac{1}{n_1 n_2^2}\sum_{i=1}^{n_1}\left(\sum_{j=1}^{n_2} g_{ij}\right)^2 + \left(\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} g_{ij}\right)^2, \\ \mathrm{Var}[\tilde{M}^2] &= O\left(\frac{1}{n_1 n_2}\right). \end{aligned} \tag{6.3}$$

In order to show the second assertion of Lemma 2.2 we now prove that the four sums on the right hand side of (6.3) approximate the corresponding integrals on the right hand side of (2.6). For the first term we estimate as follows

$$\left| \int_{[0,1]^2} g^2(t)h(t)dt - \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} g_{ij}^2 \right| \le W_1 + W_2 \tag{6.4}$$

where

$$W_1 = \sum_{i=2}^{n_1}\sum_{j=2}^{n_2}\left\{\int_{t_{1,i-1}}^{t_{1i}}\int_{t_{2,j-1}}^{t_{2j}} |g^2(t_1, t_2)h(t_1, t_2) - g^2(t_{1i}, t_{2j})h(t_{1i}, t_{2j})|dt_1 dt_2\right\}$$

and

$$W_2 = \sum_{i=2}^{n_1}\sum_{j=2}^{n_2} |g^2(t_{1i}, t_{2j})| \left|(t_{1i} - t_{1,i-1})(t_{2j} - t_{2,j-1})h(t_{1i}, t_{2j}) - \frac{1}{n_1 n_2}\right|.$$

From the Lipschitz continuity of g and h we obtain for the first term $W_1 = O(n_1^{-\gamma})$. For the second term in (6.4) we note that g is bounded and that by (2.3) for some $\xi_{ki} \in (t_{k,i-1}, t_{ki})$

$$\left|(t_{ki} - t_{k,i-1})h_k(t_{ki}) - \frac{1}{n_k}\right| = |t_{ki} - t_{k,i-1}||h(t_{ki}) - h(\xi_{ki})| = O\left(\frac{1}{n_k^{1+\gamma}}\right), \qquad k = 1, 2$$

(here the last equality follows from the Lipschitz continuity of h). A straightforward calculation shows $W_2 = O(n_1^{-\gamma})$, which gives

$$\left|\int_{[0,1]^2} g^2(t)h(t)dt - \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} g_{ij}^2\right| = O(n_1^{-\gamma}).$$

The other terms in (6.3) can be treated exactly in the same way which proves the second assertion. For the third part we note that a straightforward calculation yields under the hypothesis of additivity

$$\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} g_{ij}^2 - \frac{1}{n_1}\sum_{j=1}^{n_2}\left(\sum_{i=1}^{n_1} g_{ij}\right)^2 - \frac{1}{n_2}\sum_{i=1}^{n_1}\left(\sum_{j=1}^{n_2} g_{ij}\right)^2 + \frac{1}{n_1 n_2}\left(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} g_{ij}\right)^2 = 0. \qquad (6.5)$$

□

PROOF OF THEOREM 2.1. We first investigate the quadratic form corresponding to the variance estimator in (2.9). To this end let

$$Q_0^{(n_j)} = \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ & & & & 0 \end{pmatrix} \begin{matrix} \\ \leftarrow l+1 \\ \\ \leftarrow n_j - l \\ \\ \end{matrix} \in \mathbb{R}^{n_j \times n_j}$$

$$\begin{matrix} \uparrow & & \uparrow \\ l+1 & & n_j - l \end{matrix}$$

$$Q_r^{(n_j)} = \begin{pmatrix} 0 & & & & & & \\ & 1\ 0 \cdots 0 & 1 & 0 \cdots 0 & 1 & & \\ & & 1\ 0 \cdots 0 & 1 & 0 \cdots 0 & 1 & \\ & & \ddots & \ddots\ \ddots & & \ddots\ \ddots & \\ & & & 1\ 0 \cdots 0 & 1 & 0 \cdots 0 & 1 \\ & & & & & & 0 \end{pmatrix} \begin{matrix} \\ \leftarrow l+1 \\ \\ \\ \leftarrow n_j - l \\ \\ \end{matrix} \in \mathbb{R}^{n_j \times n_j}$$

$$\begin{matrix} \uparrow & \uparrow & \uparrow & \uparrow \\ l+1-r & l+1 & l+1+r & n_j - l + r \end{matrix}$$

$(r = 1, \dots, l)$ (all other elements in this matrix are 0), define $\beta_0 = \sum_{k=0}^{l} \alpha_k^2 = 1$, $\beta_r = 2\sum_{k=0}^{l-r} \alpha_k\alpha_{k+r}$ $(r = 1, \dots, l)$ and

$$Q = \beta_0 Q_0 + \frac{1}{8}\sum_{r=1}^{l} \beta_r Q_r \tag{6.6}$$

where

$$Q_0 = Q_0^{(n_1)} \otimes Q_0^{(n_2)} \tag{6.7}$$

$$Q_r = Q_r^{(n_1)} \otimes Q_r^{(n_2)} - Q_0^{(n_1)} \otimes Q_0^{(n_2)} \qquad (r = 1, \dots, l) \tag{6.8}$$

(here $A \otimes B$ denotes the Kronecker product of A and B). Recall the definition of $\hat{\sigma}_{l,\alpha}^2$ in (2.9) and introduce

$$\tilde{\sigma}_{l,\alpha}^2 = \frac{1}{(n_1 - 2l)(n_2 - 2l)} Y^T Q Y, \tag{6.9}$$

then a careful inspection shows that the number of different elements in the quadratic forms corresponding to $\tilde{\sigma}_{l,\alpha}^2$ and $\hat{\sigma}_{l,\alpha}^2$ is $O(n_1)$ and a straightforward calculation gives

$$n_1 n_2 \operatorname{Var}(\hat{\sigma}_{l,\alpha}^2 - \tilde{\sigma}_{l,\alpha}^2) = o(1). \tag{6.10}$$

Throughout this section define

$$\hat{N}_{l,\alpha}^2 = \frac{1}{(n_1 - 2l)(n_2 - 2l)} Y^T (P - Q) Y \tag{6.11}$$

where P and Q are given by (6.2) and (6.6), respectively. It now follows from (6.10) that it is sufficient to prove the assertion for $\hat{N}_{l,\alpha}^2$, i.e.

$$\lim_{n_1, n_2 \to \infty} n_1 n_2 \operatorname{Var}(\hat{M}_{l,\alpha}^2) = \lim_{n_1, n_2 \to \infty} n_1 n_2 \operatorname{Var}(\hat{N}_{l,\alpha}^2) = \mu^2. \tag{6.12}$$

To this end define $g = (g_{11}, \dots, g_{1n_2}, \dots, g_{n_1 1}, \dots, g_{n_1 n_2})^T$ and $\tilde{Y} = Y - g$, then it follows from equation (6.5) that under the hypothesis of additivity

$$g^T P g = \|Pg\|_2^2 = 0,$$

which shows

$$n_1 n_2 \operatorname{Var}(\hat{N}_{l,\alpha}^2) = \frac{1}{(n_1 - 2l)(n_2 - 2l)} \operatorname{Var}[\tilde{Y}^T (P - Q)\tilde{Y} - 2g^T Q \tilde{Y}] + o(1). \tag{6.13}$$

For the second term in this variance we have

$$\frac{1}{n_1 n_2} \operatorname{Var}(g^T Q \tilde{Y}) = \frac{1}{n_1 n_2} \operatorname{Var}\left[\sum_{i=l+1}^{n_1 - l} \sum_{j=l+1}^{n_2 - l} \sum_{r=-1}^{1} \sum_{s=-1}^{1} \left(\sum_{k=0}^{l} \alpha_k g_{i+rk, j+sk}\right) \times \left(\sum_{k=0}^{l} \alpha_k \tilde{Y}_{i+rk, j+sk}\right)\right] + o(1)$$

where $\tilde{Y}_{ij} = Y_{ij} - g_{ij}$ $(i = 1, \ldots, n_1; j = 1, \ldots, n_2)$. From the Hölder continuity of g, the definition of a difference sequence and from (2.3) we obtain

$$\left| \sum_{k=0}^{l} \alpha_k g_{i+rk,j+sk} \right| = O(n_1^{-\gamma})$$

which shows

$$\frac{1}{n_1 n_2} \operatorname{Var}(g^T Q \tilde{Y}) = O(n_1^{-2\gamma}).$$

Consequently it follows from (6.13) and formula (16) in Whittle (1964)

$$\begin{aligned} \lim_{n_1, n_2 \to \infty} n_1 n_2 \operatorname{Var}(\hat{N}_{l,\alpha}^2) &= \lim_{n_1, n_2 \to \infty} \frac{1}{n_1 n_2} (\tilde{Y}^T (P - Q) \tilde{Y}) \\ &= \frac{\sigma^4}{4} \sum_{r=1}^{l} \beta_r^2 = \sigma^4 \sum_{r=1}^{l} \left(\sum_{k=0}^{l-r} \alpha_k \alpha_{k+r} \right)^2 = \mu^2 \end{aligned}$$

which proves (6.12) and the assertion of Theorem 2.1. □

PROOF OF THEOREM 2.2. Let $\tilde{Y}$ denote the vector of centered random variables $Y_{ij} - g_{ij}$ and define

$$\frac{1}{(n_1 - 2l)(n_2 - 2l)} R$$

as the matrix corresponding to the quadratic form of the variance estimator $\hat{\sigma}_{l,\alpha}^2$ defined in (2.9). With these notations we have (note that $g^T P g = 0$, by (6.5))

$$\begin{aligned} \hat{M}_{l,\alpha}^2 = \tilde{Y}^T & \left(\frac{1}{n_1 n_2} P - \frac{r_{n_1,n_2}}{(n_1 - 2l)(n_2 - 2l)} R \right) \tilde{Y} + 2 \tilde{Y}^T \\ & \cdot \left(\frac{1}{n_1 n_2} P - \frac{r_{n_1,n_2}}{(n_1 - 2l)(n_2 - 2l)} R \right) g \\ & + \frac{r_{n_1,n_2}}{(n_1 - 2l)(n_2 - 2l)} g^T R g + o_p \left(\frac{1}{\sqrt{n_1 n_2}} \right). \end{aligned}$$

The third term in this sum can be shown to be of order $O(n_1^{-2\gamma})$, by the Hölder continuity of the regression function g. The expectation of the second term vanishes while its variance is of order $O(n_1^{-2-2\gamma})$ (which follows by similar arguments as given at the end of the proof of Theorem 2.1). Consequently, we have

$$\hat{M}_{l,\alpha}^2 = \tilde{Y}^T \left(\frac{1}{n_1 n_2} P - \frac{r_{n_1,n_2}}{(n_1 - 2l)(n_2 - 2l)} R \right) \tilde{Y} + o_p \left(\frac{1}{\sqrt{n_1 n_2}} \right)$$

and it is sufficient to prove the assertion for the centered random variables $\tilde{Y}_{ij}$. Throughout this proof define a matrix $\tilde{R}$ as follows. We change the diagonal elements of R such that the resulting matrix

$$\frac{1}{n_1 n_2} P - \frac{r_{n_1,n_2}}{(n_1 - 2l)(n_2 - 2l)} \tilde{R}$$

has vanishing diagonal elements (all other elements of R remain unchanged). Let

$$\hat{\kappa}^2_{l,\alpha} = \frac{1}{(n_1-2l)(n_2-2l)}\tilde{Y}^T\tilde{R}\tilde{Y}$$

denote the corresponding quadratic form and

$$\hat{A}^2_{l,\alpha} = \tilde{M}^2 - r_{n_1,n_2}\hat{\kappa}^2_{l,\alpha} = \tilde{Y}^T\left(\frac{1}{n_1n_2}P - \frac{r_{n_1,n_2}}{(n_1-2l)(n_2-2l)}\tilde{R}\right)\tilde{Y}$$
$$\hat{B}^2_{l,\alpha} = r_{n_1,n_2}(\hat{\kappa}^2_{l,\alpha} - \hat{\sigma}^2_{l,\alpha}),$$

where $\tilde{M}^2$ is defined in (2.7). We have $\hat{M}^2_{l,\alpha} = \hat{A}^2_{l,\alpha}+\hat{B}^2_{l,\alpha}$, $E[\hat{M}^2_{l,\alpha}] = 0$ (because $E[\tilde{Y}] = 0$ implies $E[\tilde{M}^2] = r_{n_1,n_2}\sigma^2$; $E[\hat{\sigma}^2_{l,\alpha}] = \sigma^2$) and $E[\hat{A}^2_{l,\alpha}] = 0$, because, by definition of $\hat{\kappa}^2_{l,\alpha}$, the matrix corresponding to the quadratic form $\hat{A}^2_{l,\alpha}$ has vanishing diagonal elements and the $\tilde{Y}_{ij}$ are centered random variables. This implies $E[\hat{B}^2_{l,\alpha}] = 0$ and it follows from a straightforward calculation that

$$\begin{aligned} \operatorname{Var}(\hat{B}^2_{l,\alpha}) &= o((n_1n_2)^{-1}) \\ \hat{B}^2_{l,\alpha} &= r_{n_1,n_2}(\hat{\kappa}^2_{l,\alpha} - \hat{\sigma}^2_{l,\alpha}) \\ &= o_p((n_1n_2)^{-1/2}) \\ \hat{M}^2_{l,\alpha} - \hat{A}^2_{l,\alpha} &= o_p((n_1n_2)^{-1/2}). \end{aligned} \tag{6.14}$$

The first assertion is needed for the proof of the second one and follows from the fact that among the n_1n_2 diagonal elements of the matrix $R-\tilde{R}$ there are at most $O(n_1)$ elements which do not converge to zero. Consequently it is sufficient to prove the assertion of the theorem for $\hat{A}^2_{l,\alpha}$, i.e.

$$\sqrt{n_1n_2}\hat{A}^2_{l,\alpha} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\mu^2). \tag{6.15}$$

To this end define

$$W = \sqrt{n_1n_2}\left(\frac{1}{n_1n_2}P - \frac{r_{n_1,n_2}}{(n_1-2l)(n_2-2l)}\tilde{R}\right),$$

where P is given by (6.2) and $\tilde{R}$ is defined at the beginning of this proof, and consider the quadratic form

$$\tilde{W}_{n_1,n_2} = \tilde{Y}^T W\tilde{Y} = \sqrt{n_1n_2}\hat{A}^2_{l,\alpha}. \tag{6.16}$$

We will prove asymptotic normality of $\tilde{W}_{n_1,n_2}$ by an application of Theorem 5.2 in de Jong (1987). To this end we show that the assumption (1)–(3) of this theorem are satisfied.

A similar argument as in the proof of Theorem 2.1 shows $\operatorname{Var}(\tilde{W}_{n_1,n_2} - \sqrt{n_1n_2}\hat{M}^2_{l,\alpha}) = o(1)$ and by Theorem 2.1 we have for $n_1, n_2 \to \infty$, $n_1/n_2 \to \lambda \in (0,\infty)$

$$\sigma^2(n_1,n_2) := \operatorname{Var}(\tilde{W}_{n_1,n_2}) \to \mu^2$$

where μ^2 is defined by (2.10). The elements of the matrix $W = (w_{ij})_{i,j=1,\dots,n_1n_2}$ can easily be shown to satisfy for sufficiently large n_1, n_2

$$\begin{aligned} \sum_{j=1}^{n_1n_2} w^2_{ij} &\le \frac{c_1}{(n_1n_2)^3}\{(n_1-1)^2n_2 + (n_2-1)^2n_1 + (n_1-1)(n_2-1)\} + \frac{c_2}{n_1n_2} \\ &= O((n_1n_2)^{-1}) \end{aligned} \tag{6.17}$$

where the constants c_1, c_2 do not depend on n_1, n_2. Here the first term on the right hand side of (6.17) corresponds to the matrix P and the second to the matrix $\tilde{R}$. For this reason condition (1) in de Jong's (1987) Theorem 5.2 is satisfied for any sequence $K(n_1, n_2)$ of real numbers which converges to infinity such that $K(n_1, n_2) = o((n_1 n_2)^{1/4})$. Since $\tilde{Y}_{ij}^2$ are i.i.d. and integrable condition (2) also holds for these sequences. For the remaining condition in de Jong's Theorem 5.2 we need to estimate the eigenvalues of W. To this end we note that it follows by similar arguments as in (6.17) for sufficiently large n_1, n_2

$$\sum_{j=1}^{n_1 n_2} |w_{ij}| \leq \frac{c_3}{\sqrt{n_1 n_2}}. \tag{6.18}$$

By Gerschgorin's theorem we therefore obtain that the eigenvalues μ_i $(i = 1, \ldots, n_1 n_2)$ of the matrix W are bounded by $c_3/\sqrt{n_1 n_2}$ for sufficiently large n_1, n_2 (note that the diagonal elements of W are 0). Consequently assumption (3) of Theorem 5.2 in de Jong (1987) is also satisfied and this theorem shows $[n_1, n_2 \to \infty,\ n_1/n_2 \to \lambda \in (0, \infty)]$

$$\tilde{W}_{n_1,n_2} = \tilde{Y}^T W \tilde{Y} = \sqrt{n_1 n_2} \hat{A}_{l,\alpha}^2 \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mu^2) \tag{6.19}$$

which proves the assertion for the centered random variables $\tilde{Y}_{ij} = Y_{ij} - g_{ij}$. The general statement (6.15) follows by the discussion at the beginning of this proof. □

PROOF OF THEOREM 5.1. A similar calculation as in the proof of Lemma 2.2 gives under H_0 that

$$E[\tilde{M}_{n_1-l,n_2-l}^2] = \frac{r_{l,n_1,n_2}}{(n_1 - 2l)(n_2 - 2l)} \Sigma_{n_1-l,n_2-l}$$

where

$$\Sigma_{n_1-l,n_2-l} = \sum_{i=l+1}^{n_1-l} \sum_{j=l+1}^{n_2-l} \sigma^2(t_{1i}, t_{2j}).$$

Furthermore,

$$\begin{aligned} E[\hat{\tau}_{l,\alpha}^2] &= \frac{1}{4} \frac{1}{(n_1 - 2l)(n_2 - 2l)} \sum_{\substack{r,s=-1 \\ r=0 \vee s=0 \\ (r,s) \neq (0,0)}}^{1} \sum_{i,j=l+1}^{n_1-l,n_2-l} \sum_{k=0}^{l} \alpha_k^2 \sigma_{i+kr,j+ks}^2 \\ &+ \frac{1}{4} \frac{1}{(n_1 - 2l)(n_2 - 2l)} \sum_{\substack{r,s=-1 \\ r=0 \vee s=0 \\ (r,s) \neq (0,0)}}^{1} \sum_{i,j=l+1}^{n_1-l,n_2-l} \left\{ \sum_{k=0}^{l} \alpha_k g_{i+kr,j+ks} \right\}^2 \end{aligned} \tag{6.20}$$

where $g_{ij} = g(t_{1i}, t_{2j})$, $(i = 1, \ldots, n_1; j = 1, \ldots, n_2)$ and the second term vanishes at a rate $o((n_1 n_2)^{-1/2})$. The first summand on the right hand side of (6.20) is rewritten as

$$\begin{aligned} &\frac{1}{4} \frac{1}{(n_1 - 2l)(n_2 - 2l)} \sum_{k=0}^{l} \alpha_k^2 \sum_{i,j=l+1}^{n_1-l,n_2-l} \left\{ \sum_{s=-1,1} \sigma_{i,j+ks}^2 + \sum_{r=-1,1} \sigma_{i+kr,j}^2 \right\} \\ &= \frac{1}{4} \frac{1}{(n_1 - 2l)(n_2 - 2l)} \left\{ 4\Sigma_{n_1-l,n_2-l} \sum_{k=0}^{l} \alpha_k^2 + O(n_1^{1-\gamma}) \right\} \\ &= \frac{1}{(n_1 - 2l)(n_2 - 2l)} \Sigma_{n_1-l,n_2-l} + O(n_1^{-(1+\gamma)}). \end{aligned}$$

Hence, it follows that

$$E[\tilde{M}^2_{n_1-l,n_2-l} - r_{l,n_1,n_2}\hat{\tau}^2_{l,\alpha}] = o((n_1 n_2)^{-1/2}).$$

It is easy to see that

$$\mathrm{Var}[\tilde{M}^2_{n_1-l,n_2-l} - r_{l,n_1,n_2}\hat{\tau}^2_{l,\alpha}) = O((n_1 n_2)^{-1})$$

and the asymptotic normality and the limiting variance are proved along the lines of the proof of Theorem 2.2. □

Acknowledgements

The authors are grateful to I. Spreckelsen, who typed this paper with considerable technical experience, and to T. Wagner for some helpful comments regarding variance estimators. We are also indebted to two unknown referees for some helpful comments which led to a substantial improvement of an earlier version of this paper.

References

Barry, D. (1993). Testing for additivity of a regression function, *Ann. Statist.*, **21**, 235–254.

Buckley, M. J. and Eagleson, G. K. (1988). An approximation to the distribution of quadratic forms in nonnormal random variables, *Austral. J. Statist. A*, **30**, 150–159.

Buja, A., Hastie, T. and Tibshirani, R. (1989). Linear smoothers and additive models, *Ann. Statist.*, **17**, 453–555.

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

Dette, H., Munk, A. and Wagner, T.(1999). A review of variance estimators with extensions to multivariate nonparametric regression models, *Multivariate Design and Sampling* (ed. S. Ghosh), 469–499, Marcel Dekker, New York (to appear).

Eubank, R. L., Hart, J. D., Simpson, D. G. and Stefanski, L. A. (1995). Testing for additivity in nonparametric regression, *Ann. Statist.*, **23**(6), 1896–1920.

Friedman, J. H. and Stuetzle, W. (1981). Projection pursuit regression, *J. Amer. Statist. Assoc.*, **76**, 817–823.

Gasser, T., Sroka, L. and Jennen-Steinmetz, G. (1986). Residual variance and residual pattern in nonlinear regression, *Biometrika*, **73**, 625–633.

Hall, P., Kay, J. W. and Titterington, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression, *Biometrika*, **77**, 521–528.

Hall, P., Kay, J. W. and Titterington, D. M. (1991). On estimation of noise variance in two–dimensional processing, *Adv. in Appl. Prob.*, **23**, 476–495.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, Chapman and Hall, London.

Jørgensen, B. (1993). *The Theory of Linear Models*, Chapman and Hall, London.

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

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

Seifert, B., Gasser, T. and Wolf, A. (1993). Nonparametric estimation of residual variance revisited, *Biometrika*, **80**, 373–383.

Stone, C. J. (1985). Additive regression and other nonparametric models, *Ann. Statist.*, **13**, 689–705.

Tukey, J. (1949). One degree of freedom test for non-additivity, *Biometrics*, **5**, 232–242.

Whittle, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables, *Theory Probab. Appl.*, **5**, 302–305.

Whittle, P. (1964). On the convergence to normality of quadratic forms in independent variables, *Theory. Probab. Appl.*, **9**, 103–108.