Ann. Inst. Statist. Math.
Vol. 50, No. 2, 253–275 (1998)

A SIMPLE GOODNESS-OF-FIT TEST FOR LINEAR MODELS UNDER A RANDOM DESIGN ASSUMPTION

HOLGER DETTE AND AXEL MUNK

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

(Received February 6, 1997; revised June 9, 1997)

Abstract. Let (X, Y) denote a random vector with decomposition $Y = f(X) + \varepsilon$ where $f(x) = E[Y \mid X = x]$ is the regression of Y on X. In this paper we propose a test for the hypothesis that f is a linear combination of given linearly independent regression functions $g_1, \dots, g_d$. The test is based on an estimator of the minimal L^2-distance between f and the subspace spanned by the regression functions. More precisely, the method is based on the estimation of certain integrals of the regression function and therefore does not require an explicit estimation of the regression. For this reason the test proposed in this paper does not depend on the subjective choice of a smoothing parameter. Differences between the problem of regression diagnostics in the nonrandom and random design case are also discussed.

Key words and phrases: Nonparametric regression check, validation of goodness of fit, L^2-distance, equivalence of regression functions, random design.

1. Introduction

Consider a two dimensional random vector (X, Y) which allows a decomposition

$$Y = f(X) + \sigma(X)\varepsilon \tag{1.1}$$

where f is the regression of Y on X, ε denotes a centered random variable which is independent of X and $\sigma(\,) > 0$ is an unknown variance function $V[Y \mid X = x] = \sigma^2(x)$, i.e. $V[\varepsilon] = 1$. In other words, we consider a nonparametric (heteroscedastic) regression model, where the explanatory variable x is a realisation of a random variable X with unknown distribution. Let $g_1, \dots, g_d$ be given linearly independent regression functions and define $U = \text{span}\{g_1, \dots, g_d\}$. In this paper we propose a simple test for the hypothesis that the regression of Y on X is a linear function of $g_1, \dots, g_d$, i.e.

$$H_0 : f \in U \tag{1.2}$$

versus

$$H_1 : f \notin U. \tag{1.3}$$

Much effort has been devoted to the problem of model diagnostics, because the statistical analysis for linear models is more tractable and allows a direct interpretation of the observed effects in terms of the parameters. Among others we refer to the work of Zwanzig (1980), Brodeau (1993), Azzalini and Bowman (1993), Eubank and Hart (1992), who considered the case of a nonrandom design variable. Kozek (1991), Staniswalis and Severini (1991), Härdle and Mammen (1993), Stute and Manteiga (1996) and Stute (1997) studied the testing problem (1.2) versus (1.3) under the assumption of a random design as above. Most authors propose a test statistic based on the sup- or L^2-distance between a nonparametric and a parametric fit (see e.g. Kozek (1991), Stute and Manteiga (1996) or Härdle and Mammen (1993)) which requires a square-root consistent estimation of the parameter vector. For a deterministic design and homoscedastic errors Dette and Munk (1998) based their test criterion on an estimator of the minimal L^2-distance between f and the subspace U of regression functions with respect to a weighted L^2-norm. The particular case of linear and polynomial regression was treated by means of spline smoothing techniques in Eubank and Spiegelman (1990) and recently by Jayasuriya (1996). In this paper we generalize Dette and Munk's (1998) approach to the case of a random design and heteroscedastic error structure. Surprisingly, this situation turns out to be rather different compared to the fixed design case. This is caused by the need to estimate additionally the design density in the Fourier-expansion of the L^2-distance

$$M^2 := \min_{g \in U} E[(f(X) - g(X))^2]$$

between U and f.

In Section 2 we provide an estimator for M^2 which is shown to be asymptotically normal. This is applied in Section 3 to goodness-of-fit testing of linearity. It is shown, that, compared to the fixed design case, the asymptotic variance is additionally increased by terms which depend essentially on the variability of the design variable X. Only, under the null hypothesis (1.2) both variances coincide. Hence, inference for (1.2) may become rather noninformative as the variance of the explanatory variable X increases. Therefore, the practical merits of this paper consist to some extent in the understanding of the differences between the fixed and random design assumption for the statistical analysis of regression models.

In order to overcome systematic drawbacks of testing the classical hypotheses (1.2) in the context of model checking (this will be made precise in the following) we suggest to use tests for precise hypotheses (Berger and Delampady (1987))

$$H_\pi : M^2 > \pi \qquad \text{versus} \qquad K_\pi : M^2 \leq \pi.$$

Rejection of H_π allows to assess the validity of the model U within an L^2 neighborhood at a controlled error rate α. Here π denotes a specified measure of discrepancy which is assumed to be a tolerable deviation from the model. The reformulation of the null H_0 into H_π can be motivated by the demand of various authors (see

Berkson (1942), or Hauck and Anderson (1996)) to 'interchange' the classical hypotheses H_0 and H_1 in many situations of practical interest. Typically, this will be the case when the wrong decision for H_0, will be a more serious error as a rejection of H_0. This argument applies in general for goodness of fit testing, particularly in model diagnostics (see MacKinnon (1992) for a careful discussion) because here the acceptance of H_0 implies a subsequent data analysis within a parametric framework, i.e. under the assumption of $f \in U$ – without any evidence for H_0. We will see in Section 3 that a proper test for H_π requires the estimation of additional terms occurring from the variability of the design variable X (compared to the case of testing H_0) in the limiting variance of the empirical counterpart of M^2. When testing H_0 these terms do not occur which implies that even a large observed P-value of the pivot statistic for H_0 bears no evidence in favour of H_0. The lack of a naive use of P-values in many statistical applications was already criticised by various authors in a different context than regression checks (see Berger and Sellke (1987) or Schervish (1996) for an overview).

We finally mention that our approach avoids an explicit (nonparametric) estimation of the regression function. Only integrals of the regression and the basis functions $g_1, \ldots, g_d$ have to be estimated and as a consequence the test proposed in this paper does not depend on any specific choices of a smoothing parameter.

2. The empirical L^2-distance: asymptotic theory

2.1 *A weighted L^2-distance*

For the sake of simplicity we shall only consider real valued X's. Most of the presented results have immediate extensions to the case of a multivariate explanatory variable (see the discussion in Remark 2.4). The main results of the paper require the following basic assumptions regarding the distribution of the (real) random variables X and ε.

(2.1) X has a density, say h, with compact support, say $\mathcal{H} \subseteq \mathbb{R}$, which is Hölder continuous of order $\gamma > \frac{1}{2}$ and bounded away from zero on its support.

$$X \text{ and } \varepsilon \text{ are independent.} \tag{2.2}$$

$$E[\varepsilon] = 0, \qquad V[\varepsilon] = 1, \qquad E[\varepsilon^4] < \infty. \tag{2.3}$$

For the variance function $\sigma^2(\cdot)$ and the regression functions we will make the following basic assumptions

$$\sigma, f, g_1, \ldots, g_d \in \text{Höl}_\gamma(\mathcal{H}), \qquad E[\sigma^2(X)] > 0 \tag{2.4}$$

where $\text{Höl}_\gamma(\mathcal{H})$ denotes the set of functions defined on $\mathcal{H}$ which are Hölder continuous of order $\gamma > 1/2$.

As a measure of goodness of fit for testing the hypothesis of linearity we consider the best approximation of the regression function f by the linear model $U = \text{span}\{g_1, \ldots, g_d\}$ with respect to the L^2-norm induced by the distribution of X, i.e.

$$M^2 = \min_{g \in U} E[(f(X) - g(X))^2]. \tag{2.5}$$

With this notation the hypotheses (1.2) and (1.3) can be conveniently rewritten as

$$H_0 : M^2 = 0, \qquad H_1 : M^2 > 0. \tag{2.6}$$

As mentioned in the Introduction we will additionally consider the following 'precise testing' problem

$$H_\pi : M^2 > \pi \qquad \text{versus} \qquad K_\pi : M^2 \leq \pi \tag{2.7}$$

where $\pi > 0$ denotes a fixed preassigned bound for which the regression function is considered as sufficiently close to the (linear) model space U. Observe that the rejection of H_π implies the existence of constants a_i, $i = 1, \ldots, d$, such that

$$E\left[\left(f(X) - \sum_{i=1}^{d} a_i g_i(X)\right)^2\right] \leq \pi$$

at a controlled error rate, say α.

2.2 *Estimation of M^2 and asymptotics*

In order to construct asymptotic tests for H_π and H_0 we utilize that the distance in (2.5) can be expressed as a function of the inner products (see Achieser (1956))

$$\begin{aligned} A_0 &= E[f^2(X)] \\ A_\ell &= E[f(X) g_\ell(X)], \qquad 1 \leq \ell \leq d \\ G_{\ell k} &= E[g_\ell(X) g_k(X)], \qquad 1 \leq \ell \leq k \leq d. \end{aligned} \tag{2.8}$$

More precisely, putting

$$\begin{aligned} A &:= (A_0, \ldots, A_d)^T \in \mathbb{R}^{d+1}, \\ G &:= ((G_{11}, \ldots, G_{dd}), (G_{12}, \ldots, G_{1d}), (G_{23}, \ldots, G_{2d}), \ldots, (G_{d-1,d}))^T \\ &= (G_1^T, \ldots, G_d^T)^T \in \mathbb{R}^{d(d+1)/2}, \end{aligned} \tag{2.9}$$

where $G_i \in \mathbb{R}^{d-i+1}$, $i = 1, \ldots, d$, we have (see Achieser (1956), p. 16)

$$M^2 = \rho(A, G) = \frac{\Gamma(A, G)}{\Gamma(G)}. \tag{2.10}$$

Here $\Gamma(A,G)$ denotes the Gram's determinant

$$\Gamma(A,G) = \begin{vmatrix} A_0 & A_1 & A_2 & \cdots & A_d \\ A_1 & G_{11} & G_{12} & \cdots & G_{1d} \\ A_2 & G_{12} & G_{22} & \cdots & G_{2d} \\ \vdots & \vdots & \vdots & & \vdots \\ A_d & G_{1d} & G_{2d} & \cdots & G_{dd} \end{vmatrix} \tag{2.11}$$

and $\Gamma(G)$ is the determinant obtained from $\Gamma(A,G)$ by deleting the first row and column. Note that ρ is well defined, i.e. $\Gamma(G) > 0$, because $\Gamma^2(G)$ gives the volume with respect to the measure P^X of the parallelpepid spanned by the linearly independent vectors $g_1, \ldots, g_d$. Therefore, a test for the hypotheses (2.6) can either be based on an empirical counterpart of M^2 or of $\Gamma(A,G)$ which will both be carried out in the following (see Theorems 2.1 and 2.2 below).

In the subsequent discussion let $(Y_1, X_1), \ldots, (Y_n, X_n)$ denote independent copies of (Y,X) where realizations of the random variables correspond to the observed values of the regression. Consistent estimates of the above determinants can easily be obtained by estimating the expectations in (2.8). To this end let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics of $X_1, \ldots, X_n$ and define $R_1^{-1}, \ldots, R_n^{-1}$ as the antiranks of $X_1, \ldots, X_n$, i.e. $(X_{(1)}, Y_{R_1^{-1}}), \ldots, (X_{(n)}, Y_{R_n^{-1}})$ is the ordered (according to the x-values) sample of the observations. The estimates of (2.8) are defined by

$$\hat{A}_0^{(n)} = \frac{1}{n-1} \sum_{j=2}^{n} Y_{R_j^{-1}} Y_{R_{j-1}^{-1}} \tag{2.12}$$

$$\hat{A}_\ell^{(n)} = \frac{1}{n} \sum_{j=1}^{n} g_\ell(X_{(j)}) Y_{R_j^{-1}} = \frac{1}{n} \sum_{j=1}^{n} g_\ell(X_j) Y_j, \qquad 1 \le \ell \le d \tag{2.13}$$

$$\hat{G}_{\ell k}^{(n)} = \frac{1}{n} \sum_{j=1}^{n} g_\ell(X_j) g_k(X_j), \qquad 1 \le \ell \le k \le d. \tag{2.14}$$

Finally, we estimate M^2 as

$$\hat{M}_n^2 = \rho(\hat{A}^{(n)}, \hat{G}^{(n)}) \tag{2.15}$$

and $\Gamma(A,G)$ as

$$\hat{\Gamma}_n = \Gamma(\hat{A}^{(n)}, \hat{G}^{(n)}) \tag{2.16}$$

where

$$\hat{A}^{(n)} = (\hat{A}_0^{(n)}, \ldots, \hat{A}_d^{(n)})^T \tag{2.17}$$

$$\begin{aligned} \hat{G}^{(n)} &:= ((\hat{G}_{11}^{(n)}, \ldots, \hat{G}_{dd}^{(n)}), (\hat{G}_{12}^{(n)}, \ldots, \hat{G}_{1d}^{(n)}), \\ &\qquad (\hat{G}_{23}^{(n)}, \ldots, \hat{G}_{2d}^{(n)}), \ldots, (\hat{G}_{d-1,d}^{(n)}))^T \\ &= (\hat{G}_1^{(n)^T}, \ldots, \hat{G}_d^{(n)^T})^T. \end{aligned} \tag{2.18}$$

Observe, that the estimator $\hat{M}_n^2$ is well defined if and only if

$$\Gamma(\hat{G}^{(n)}) > 0, \tag{2.19}$$

which is asymptotically valid because Kolmogorov's SLLN and condition (2.4) imply

$$\Gamma(\hat{G}^{(n)}) \to \Gamma(G) > 0, \qquad P^X \quad \text{a.s.},$$

by the independence of X and ε. The asymptotic theory for $\hat{M}_n^2$ and $\hat{\Gamma}^{(n)}$ is given in the following two theorems. The proofs are complicated and therefore deferred to Section 4. Throughout this paper let $\mathcal{N}(\mu, \sigma^2)$ denote a normally distributed random variable with expectation μ and variance σ^2.

THEOREM 2.1. *If the assumptions* (2.1)–(2.4) *are satisfied and if* $n \to \infty$, *then*

$$\sqrt{n}(\hat{M}_n^2 - M^2) \overset{\mathcal{D}}{\Rightarrow} \mathcal{N}(0, \tau_1^2)$$

where

$$\tau_1^2 = E[\sigma^4(X)] + 4E[\sigma^2(X)\{(f - P_U f)(X)\}^2] + V[\{(f - P_U f)(X)\}^2] \tag{2.20}$$

and $P_U f$ *denotes the projection of* f *onto the subspace* $U = \operatorname{span}\{g_1, \ldots, g_d\}$ *with respect to the inner product* $E[r(X)s(X)]$; $r, s \in L^2(P^X)$.

THEOREM 2.2. *If the assumptions* (2.1)–(2.4) *are satisfied and if* $n \to \infty$, *then*

$$\sqrt{n}(\hat{\Gamma}_n - \Gamma(A, G)) \overset{\mathcal{D}}{\Rightarrow} \mathcal{N}(0, \tau_2^2)$$

where

$$\begin{aligned} \tau_2^2 = {} & \Gamma^2(G)\{E[\sigma^4(X)] + 4E[\sigma^2(X)\{(f - P_U f)(X)\}^2]\} \\ & + V\left[\sum_{\ell=0}^{d} \eta(g_1, \ldots, g_{\ell-1}, g_0, g_{\ell+1}, \ldots, g_d)\{(g_\ell - P_{U_\ell} g_\ell)(X)\} g_\ell(X)\right] \end{aligned} \tag{2.21}$$

where $g_0 = f$ *and* P_{U_ℓ} *denotes the projection onto the subspace* $U_\ell = \operatorname{span}\{g_0, \ldots, g_{\ell-1}, g_{\ell+1}, \ldots, g_d\}$ ($U_0 = U$) *with respect to the inner product* $E[r(X)s(X)]$, $r, s \in L^2(P^X)$, *and*

$$\eta(f_1, \ldots, f_d) = |(E[f_\ell(X) f_k(X)])_{\ell,k=1}^d|. \tag{2.22}$$

Remark 2.3. Note that under the null hypothesis of linearity, $H_0 : f \in U$, we have

$$f - P_U f = 0, \qquad g_\ell - P_{U_\ell} g_\ell = 0 \qquad (\ell = 0, \ldots, d)$$

and the asymptotic variances in Theorems 2.1 and 2.2 reduce to

$$\tau_1^2 = E[\sigma^4(X)], \qquad \tau_2^2 = \Gamma^2(G) E[\sigma^4(X)], \tag{2.23}$$

respectively. It turns out (see Section 4) that a proof of Theorem 2.2 is substantially more difficult than the proof of the corresponding result in Theorem 2.1. A simple intuitive explanation for this phenomenon is the following: the statistic $\hat{M}_n^2$ is invariant with respect to a change of the basis $g_1, \ldots, g_d$ of the model space U_d. However, the statistic $\hat{\Gamma}_n$ considered in Theorem 2.2 does not enjoy this invariance property, which results into a more complicated formula for the asymptotic variance.

Remark 2.4. It is also worthwhile to mention that the estimator $A_0^{(n)}$ in (2.12) can be rewritten as

$$A_0^{(n)} = \frac{1}{n}\sum_{j=1}^{n} Y_j^2 - \frac{1}{2n}\sum_{j=2}^{n}(Y_{R_j^{-1}} - Y_{R_{j-1}^{-1}})^2 + o\left(\frac{1}{n}\right). \tag{2.24}$$

Note that the first term in (2.24) is a consistent estimator of $E[f^2(X) + \sigma^2(X)]$, while the second term consistently estimates $E[\sigma^2(X)]$ (see Rice (1984) or Hall *et al.* (1990)).

This point of view allows us to indicate an immediate generalization of the proposed method to the case of a multivariate explanatory variable, i.e. X has a density with compact support $\mathcal{H} \subseteq \mathbb{R}^d$, which is Hölder continuous of order $\gamma > 1/2$. We use the estimator $\hat{M}_n^2$ (or $\hat{\Gamma}_n$) where $\hat{A}_\ell^{(n)}$ and $\hat{G}_{\ell k}^{(n)}$ are defined by (2.13) and (2.14) exactly as in the univariate case ($1 \le \ell, k \le d$) and $\hat{A}_0^{(n)}$ is given by

$$\hat{A}_0^{(n)} = \frac{1}{n}\sum_{j=1}^{n} Y_j^2 - \hat{\sigma}_n^2 \tag{2.25}$$

where $\hat{\sigma}_n^2$ is a consistent estimator of $E[\sigma^2(X)]$ (see e.g. Breiman and Meisel (1976) or Herrmann *et al.* (1995)). We finally note that the asymptotic variance of $\hat{M}_n^2$ (or $\hat{\Gamma}_n$) depends on the specific choice of this variance estimator.

3. Goodness-of-fit tests and further discussion

As pointed out in Section 2 a test for the hypothesis of linearity $H_0 : f \in U$ versus $H_1 : f \notin U$ can be based on $\hat{M}_n^2$ or $\hat{\Gamma}_n$ and the corresponding asymptotic result in Theorem 2.1 or 2.2, respectively. For the sake of brevity we restrict ourselves to the empirical distance $\hat{M}_n^2$. Similar remarks apply to the goodness-of-fit test based on $\hat{\Gamma}_n$.

3.1 *A simple model check*

A test based on $\hat{M}_n^2$ only requires the additional estimation of the asymptotic variance τ_1^2. From (2.23) we observe that, if $f \in U$, then $\tau_1^2 = E[\sigma^4(X)]$, and the following lemma provides a consistent estimate. The proof is deferred to Section 4.

LEMMA 3.1. *Assume that* (2.1)–(2.4) *are satisfied; if* $n \to \infty$, *then*

$$\hat{\sigma}_n^4 := \frac{1}{4(n-3)}\sum_{j=2}^{n-2}(Y_{R_{j+2}^{-1}} - Y_{R_{j+1}^{-1}})^2(Y_{R_j^{-1}} - Y_{R_{j-1}^{-1}})^2 \xrightarrow{P} E[\sigma^4(X)]. \tag{3.1}$$

Note that the estimator $\hat{\sigma}_n^4$ is a generalization of the variance estimator proposed by Rice (1984)

$$(3.2)\qquad \tilde{\sigma}_n^2 = \frac{1}{2(n-1)} \sum_{j=2}^{n} (Y_{R_j^{-1}} - Y_{R_{j-1}^{-1}})^2$$

in a homoscedastic setup $\sigma^2(x) \equiv \sigma^2$. Observe further, that in this case $E[\sigma^4(X)] = \sigma^4$ could also be estimated by $\tilde{\sigma}_n^4$. Moreover, there are numerous other difference based estimators (see e.g. Gasser *et al.* (1986) or Hall *et al.* (1990)) which generalize (3.2) and can easily be modified for the estimation of $E[\sigma^4(X)]$. Theorem 2.1 and Lemma 3.1 imply that, under the null hypothesis of linearity, $\sqrt{n/\hat{\sigma}_n^4}\hat{M}_n^2$ is asymptotically standard normal. Hence an (asymptotic) level α test for the hypothesis $H_0 : f \in U$ is obtained by rejecting H_0 if

$$(3.3)\qquad \sqrt{n}\frac{\hat{M}_n^2}{\sqrt{\hat{\sigma}_n^4}} > u_{1-\alpha}$$

where $u_{1-\alpha}$ denotes the $(1-\alpha)$ quantile of the standard normal distribution. Similarly, the same test can be obtained from Theorem 2.2. Note these tests have asymptotically the power

$$(3.4)\qquad P_{C,\sigma^2}\left(\sqrt{n}\frac{\hat{M}_n^2}{\sqrt{\hat{\sigma}_n^4}} > u_{1-\alpha}\right) = 1 - \Phi(u_{1-\alpha} - C(E[\sigma^4(X)])^{-1/2}) + o(1)$$

under contiguous alternatives $M_n^2 = Cn^{-1/2}$, $C > 0$.

3.2 *Testing precise hypotheses and confidence bounds for* M^2

In the Introduction we sketched some arguments in favor of testing hypotheses (2.7) instead of (2.6) because the variation of the design density may inflate the power of a test for H_0, leading consequently to a large rate of misspecifications, i.e. deciding for H_0 although M^2 is large. The additional difficulty encountered with testing precise hypotheses certainly consists in the specification of the tolerance bound π for the deviation from the linear model U which the experimenter will tolerate. In general, the hypotheses, which have to be tested—precise or point null—, will depend on the subsequent data analysis, e.g. estimation and interpretation of the parameters within the linear submodel, prediction of f or additional testing of specific subhypotheses of the model U. However, if one has decided for such a bound π, a level α test for $H_\pi : M^2 > \pi$ is given by the rejection region

$$(3.5)\qquad \sqrt{n}\frac{\hat{M}_n^2 - \pi}{\hat{\tau}_1} < u_\alpha,$$

where $\hat{\tau}_1^2$ denotes any consistent estimate of τ_1^2 defined in (2.20). In some concrete practical applications it may be difficult to determine such an exact bound π.

In these cases we can simply evaluate an asymptotic one sided $1-\alpha$ confidence interval for M^2 as

$$CI_{1-\alpha} = [0, \hat{M}_n^2 + n^{-1/2} u_{1-\alpha} \hat{\tau}_1). \tag{3.6}$$

This provides us certainly with more information concerning the evidence of the presence of the linear model U than a pure test decision of a test for the simple hypothesis H_0.

We finally discuss the homoscedastic case where the variance $\sigma^2(t)$ is known to be independent of t. Then the asymptotic variance τ_1^2 simplifies to

$$\tilde{\tau}_1^2(M^2) = \sigma^4 + M^2(4\sigma^2 - M^2) + m_4(f, U),$$

where $m_4(f, U) = E[\{(f - P_U f)(X)\}^4]$. Hence, if $\sigma^2(t) \equiv \sigma^2$, we obtain under contiguous alternatives such that $M_n^2 = \pi - Cn^{-1/2}$, $C > 0$, the asymptotic power of (3.5) as

$$1 - \Phi(u_{1-\alpha} - C\{\sigma^4 + 4M^2(4\sigma^2 - M^2) + m_4(f_\pi, U)\}^{-1/2}), \tag{3.7}$$

where f_π denotes a regression function such that $M^2 = \pi$. Observe further that in this case an alternative confidence region is given by the set

$$\tilde{CI}_{1-\alpha} = \{M^2 : H(M^2) > u_\alpha n^{-1/2}\} \cap \boldsymbol{R}_0^+,$$

where $H(M^2) := (\hat{M}_n^2 - M^2)/\tilde{\tau}_1(M^2)$. A simple calculation shows that H is a decreasing function on $[0, \infty)$ if and only if

$$2\sigma^2 \geq \hat{M}_n^2.$$

Hence we find that only in this case the confidence region is a simply connected interval of the form $[0, c)$.

3.3 *A comparison between random and nonrandom design*

It is of some interest to compare the differences between the random and fixed design assumption in the present situation of model checking. For the sake of simplicity let $\mathcal{H} = [0, 1]$ and consider a nonparametric (heteroscedastic) regression with a fixed design, i.e. we observe the outcomes of

$$Z_j = f(t_j) + \sigma(t_j)\varepsilon_j, \qquad j = 1, \ldots, n$$

independent random variables where $\varepsilon_1, \ldots, \varepsilon_n$ are independently distributed with mean 0 and variance 1. Let $0 \leq t_0 < \cdots < t_n \leq 1$ denote the design points which form an asymptotically regular sequence in the sense of Sacks and Ylvisacker (1970), i.e.

$$\max_{i=1}^{n} \left| \int_{t_{i-1}}^{t_i} h(t)dt - \frac{1}{n} \right| = o(n^{-3/2}) \tag{3.8}$$

where $h \in \text{Höl}_\gamma[0,1]$, $\gamma > 1/2$, is a positive (possibly unknown) density on $[0,1]$ exactly as in the random case. Although, the interpretation of h is rather different for fixed and random designs it is near at hand to think that both situations are mathematically equivalent, simply interpreting the fixed design density as a density which generates 'randomly' design points. Let $\hat{L}_n^2$ be defined exactly as $\hat{M}_n^2$ where $X_{(i)}$ and $Y_{R_i^{-1}}$ are replaced by t_i and Z_i. More precisely, we define

$$\hat{L}_n^2 = \rho(\hat{C}^{(n)}, \hat{D}^{(n)})$$

where $\hat{C}^{(n)} = (\hat{C}_0^{(n)}, \ldots, \hat{C}_d^{(n)})^T$, $\hat{D}^{(n)} = (\hat{D}_{1,1}^{(n)}, \ldots, \hat{D}_{d,d}^{(n)}, \hat{D}_{1,2}^{(n)}, \ldots, \hat{D}_{d-1,d}^{(n)})^T$ and

$$\hat{C}_0^{(n)} = \frac{1}{n-1} \sum_{j=2}^{n} Z_j Z_{j-1} \tag{3.9}$$

$$\hat{C}_\ell^{(n)} = \frac{1}{n} \sum_{j=1}^{n} g_\ell(t_j) Z_j, \qquad \ell = 1, \ldots, d \tag{3.10}$$

$$\hat{D}_{\ell,k}^{(n)} = \frac{1}{n} \sum_{j=1}^{n} g_\ell(t_j) g_k(t_j), \qquad \ell, k = 1, \ldots, d. \tag{3.11}$$

Note that the variables $D_{\ell,k}^{(n)}$ are not random (in contrast to the case of a random design) but are used to approximate the inner products $\int_0^1 g_\ell(t) g_k(t) h(t) dt$ with respect to the (unknown) density h. The following result generalizes recent findings of Dette and Munk (1998), who considered a different estimator which incorporates the assumption of a constant variance $\sigma^2(t) \equiv \sigma^2$ and a known design density in (3.8).

THEOREM 3.2. *If the assumptions* (2.3), (2.4) *and* (3.8) *are satisfied for a positive design density* $h \in \text{Höl}_\gamma([0,1])$ *for some* $\gamma > 1/2$, *then, if* $n \to \infty$,

$$\sqrt{n}(\hat{L}_n^2 - M^2) \overset{\mathcal{D}}{\Rightarrow} \mathcal{N}(0, \lambda^2)$$

where the asymptotic variance is given by

$$\lambda^2 = \int_0^1 \sigma^4(t) h(t) dt + 4 \int_0^1 \sigma^2(t) \{(f - P_U f)(t)\}^2 h(t) dt. \tag{3.12}$$

Note that (3.12) is the 'fixed design' counterpart of (2.20) and that

$$\lambda^2 = \tau_1^2 \Leftrightarrow f \in U,$$

where τ_1^2 is the asymptotic variance of $\sqrt{n}\hat{M}_n^2$. From (3.4) it follows that under contiguous alternatives approaching $H_0 : M^2 = 0$ with rate $O(n^{-1/2})$ there is (asymptotically) no first order difference between model discriminating in the random and non-random design case. However, if it is to be tested that $H_\pi : M^2 > \pi$, for some $\pi > 0$, then, under local alternatives approaching $M^2 = \pi$ with rate $O(n^{-1/2})$, the difference of the limiting variances is given by

$$\tau_1^2 - \lambda^2 = V[\{(f - P_U f)(X)\}^2] > 0. \tag{3.13}$$

Only, when M^2 vanishes these additional estimators do not inflate the variance. We illustrate this effect in the following where the variance σ^2 is assumed to be constant. A similar calculation as in (3.7) for the fixed design case yields for the asymptotic relative efficiency between the test in the fixed and the random design

$$ARE = \left\{1 + \frac{m_4(f_\pi, U) - \pi^2}{\sigma^4 + 4\sigma^2\pi}\right\}^{1/2}.$$

This loss in efficiency can be substantial if the tolerance bound π is small or if the variance of the errors is small. Observe further, that by varying only the design density h within the class of densities satisfying the Sacks-Ylvisacker condition (3.8) the difference in (3.13) and hence the ARE can be made arbitrarily large for a fixed alternative f.

Note finally, that the additional term in the random design variance (3.13) affects the rate of misspecifications of the model U, i.e. the type II error. In fact, we find that the hypothesis $H_0 : f \in U$ will be falsely accepted with increasing probability as $V[\{(f - P_U f)(X)\}^2]$ increases, fixing an alternative $M^2 = C > 0$. Because this quantity will be unknown in general this observation supports certainly the 'precise testing' approach in the random design case. Note, that also the confidence interval in (3.6) involves additional estimation of $V[\{(f - P_U f)(X)\}^2]$ (such as the test for the precise hypothesis H_π) in contrast to the test for H_0 in (3.3). Here, the test statistic does not discriminate between fixed and random design which may result in a large type II error solely caused by the variability of X. Hence testing of (2.7) or the additional consideration of confidence intervals for M^2 is strictly recommended.

3.4 *Testing equality of regression curves*

In the following we apply our approach to the assessment of the equality of two regression curves, say f and g. Tests based on kernel estimators for the hypothesis $\bar{H}_0 : f = g$ have been suggested by Hall and Hart (1990), Härdle and Marron (1990) and King *et al.* (1991). Finally, Delgado (1993) proposed a Kolmogorov Smirnov type statistic which does not depend on kernel estimators.

Assume the unknown regression curves to be Hölder continuous of order $\gamma > 1/2$, where the corresponding measure δ of discrepancy between f and g is the weighted L^2-distance

$$\delta^2 = \delta^2(f, g) = E[(f(X) - g(X))^2]. \tag{3.14}$$

Because δ^2 may be considered as the minimal distance of $f - g$ to the subspace $U_0 = \{0\}$, the assessment of similarity of two regression curves can be regarded as the validation of the linear regression model $f \equiv 0$. Assume that we observe triples of random variables (X_i, Y_i, Z_i) $(i = 1, \ldots, n)$, where

$$Y_i = f(X_i) + \sigma_Y(X_i)\epsilon_i$$
$$Z_i = g(X_i) + \sigma_Z(X_i)\epsilon_i, \qquad 1 \le i \le n$$

and the variance functions satisfy $\sigma_Y^2, \sigma_Z^2 \in \text{Höl}_\gamma(\mathcal{H})$, $\gamma > 1/2$ and define

$$\hat{\delta}_n^2 := \frac{1}{n-1} \sum_{i=2}^{n} (Y_{R_i^{-1}} - Z_{R_i^{-1}})(Y_{R_{i-1}^{-1}} - Z_{R_{i-1}^{-1}})$$

similar as in (2.12).

THEOREM 3.3. *If the assumptions* (2.1)–(2.4) *are satisfied and if* $n \to \infty$, *then*

$$\sqrt{n}(\hat{\delta}_n^2 - \delta^2) \overset{\mathcal{D}}{\Rightarrow} \mathcal{N}(0, \xi^2)$$

where

$$\xi^2 = E[\sigma^4(X)] + 4E[\sigma^2(X)\{(f-g)(X)\}^2] + V[\{(f-g)(X)\}^2] \tag{3.15}$$

and $\sigma(t) = \sigma_Y(t) - \sigma_Z(t)$ $(t \in \mathcal{H})$.

The proof is implicitly contained in the proof of Theorem 2.1 and therefore omitted. We finally remark that confidence limits and hypotheses tests for δ can be constructed in the same way as demonstrated in the previous sections.

4. Proofs

The first two lemmata give the asymptotic distribution of the vector $(\hat{A}^{(n)^T}, \hat{G}^{(n)^T})^T$ introduced in formulas (2.17) and (2.18) in Section 2. Here and in the following we define a complete ordering on the set of indices of the elements of a symmetric matrix by $(l', k') < (l, k)$ when $l' < l$ or when $l' = l$ and $k' < k$ in accordance to the ordering of the elements of the vector $(G_2^T, \ldots, G_d^T)^T$ in (2.9). We start our investigations with a discussion of a statistic which turns out to be asymptotically equivalent to the vector $(\hat{A}^{(n)^T}, \hat{G}^{(n)^T})^T$ (see the proof of Lemma 4.2).

LEMMA 4.1. *Let*

$$\hat{S}_0^{(n)} = \frac{1}{n-1} \sum_{j=2}^{n} (f(X_{(j)}) + \sigma(X_{(j)})\varepsilon_{R_j^{-1}})(f(X_{(j)}) + \sigma(X_{(j)})\varepsilon_{R_{j-1}^{-1}}), \tag{4.1}$$

and define $\tilde{A}^{(n)} = (\hat{S}_0^{(n)}, \hat{A}_1^{(n)}, \ldots, \hat{A}_d^{(n)})^T$, $\hat{G}^{(n)} = (\hat{G}_1^{(n)^T}, \ldots, \hat{G}_d^{(n)^T})^T$, *where* $\hat{A}_\ell^{(n)}$ *and* $\hat{G}_\ell^{(n)}$ *are given by* (2.13) *and* (2.14). *Assume that the basic assumptions* (2.1)–(2.4) *are satisfied, then*

$$E[\tilde{A}^{(n)}] = A + o(n^{-1/2}) \tag{4.2}$$

$$E[\hat{G}^{(n)}] = G \tag{4.3}$$

$$\lim_{n\to\infty} nV\left[\begin{pmatrix}\tilde{A}^{(n)}\\ \hat{G}^{(n)}\end{pmatrix}\right] = \Sigma := \begin{pmatrix}\Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \Sigma_{14}\\ \Sigma_{12}^T & \Sigma_{22} & \Sigma_{23} & \Sigma_{24}\\ \Sigma_{13}^T & \Sigma_{23}^T & \Sigma_{33} & \Sigma_{34}\\ \Sigma_{14}^T & \Sigma_{24}^T & \Sigma_{34}^T & \Sigma_{44}\end{pmatrix} \tag{4.4}$$

where the elements of the matrices $\Sigma_{11} \in \mathbb{R}^1$, $\Sigma_{12}^T, \Sigma_{13}^T \in \mathbb{R}^d$, $\Sigma_{14}^T \in \mathbb{R}^{d(d-1)/2}$, $\Sigma_{22}, \Sigma_{23}, \Sigma_{33} \in \mathbb{R}^{d\times d}$, $\Sigma_{24}^T, \Sigma_{34}^T \in \mathbb{R}^{d\times d(d-1)/2}$ *and* $\Sigma_{44} \in \mathbb{R}^{(d-1)d/2\times(d-1)d/2}$ *are given by*

$$\begin{aligned}
&\Sigma_{11} = E[\sigma^4(X)] + 4E[\sigma^2(X)f^2(X)] + V[f^2(X)]\\
&(\Sigma_{12}^T)_\ell = 2E[\sigma^2(X)f(X)g_\ell(X)] + C[f^2(X), g_\ell(X)f(X)], \qquad \ell = 1,\ldots,d\\
&(\Sigma_{13}^T)_\ell = C[f^2(X), g_\ell^2(X)], \qquad \ell = 1,\ldots,d\\
&(\Sigma_{14}^T)_{(\ell,k)} = C[f^2(X), g_\ell(X)g_k(X)], \qquad 1 \le \ell < k \le d\\
&(\Sigma_{22})_{\ell,k} = E[\sigma^2(X)g_\ell(X)g_k(X)] + C[f(X)g_\ell(X), f(X)g_k(X)], \qquad 1 \le \ell, k \le d\\
&(\Sigma_{23})_{\ell k} = C[f(X)g_\ell(X), g_k^2(X)], \qquad \ell, k = 1,\ldots,d\\
&(\Sigma_{24}^T)_{\ell,(\ell',k')} = C[f(X)g_\ell(X), g_{\ell'}(X)g_{k'}(X)], \qquad \ell = 1,\ldots,d, \quad 1 \le \ell' < k' \le d\\
&(\Sigma_{33})_{\ell k} = C[g_\ell^2(X), g_k^2(X)], \qquad \ell, k = 1,\ldots,d\\
&(\Sigma_{34}^T)_{\ell,(\ell',k')} = C[g_\ell^2(X), g_{\ell'}(X)g_{k'}(X)], \qquad \ell = 1,\ldots,d, \quad 1 \le \ell' < k' \le d\\
&(\Sigma_{44})_{(\ell,k),(\ell',k')} = C[g_\ell(X)g_k(X), g_{\ell'}(X)g_{k'}(X)],\\
&\qquad\qquad\qquad\qquad\qquad\qquad 1 \le \ell < k \le d, \quad 1 \le \ell' < k' \le d,
\end{aligned}$$

PROOF. The relation (4.3) is obvious observing (2.14) and (2.8). Similarly, we have from the definition (2.13)

$$\hat{A}_\ell^{(n)} = \frac{1}{n}\sum_{j=1}^n g_\ell(X_{(j)})Y_{R_j^{-1}} = \frac{1}{n}\sum_{j=1}^n g_\ell(X_j)(f(X_j) + \varepsilon_j)$$

for $\ell = 1, \ldots, d$ which proves (4.2) for the last d components of $\tilde{A}^{(n)}$, by the independence of X and ε. For the first component of $\tilde{A}^{(n)}$ we obtain

$$\begin{aligned}
E[\hat{S}_0^{(n)}] &= \frac{1}{n-1}\sum_{j=2}^n E[(f(X_{(j)}) + \sigma(X_{(j)})\varepsilon_{R_j^{-1}})(f(X_{(j)}) + \sigma(X_{(j)})\varepsilon_{R_{j-1}^{-1}})]\\
&= \frac{1}{n-1}\sum_{j=2}^n E[f^2(X_{(j)})] = A_0 + O\left(\frac{1}{n}\right).
\end{aligned}$$

In order to prove the representation (4.4) we derive exemplarily the asymptotic covariances $nC(\hat{S}_0^{(n)}, \hat{A}_\ell^{(n)})$ and $nC(\hat{A}_\ell^{(n)}, \hat{G}_{ij}^{(n)})$ which appear in the matrices Σ_{12}, Σ_{23} and Σ_{33}. All other cases are treated exactly in the same way and therefore left to the reader. For the first term we note that by (4.2) and (2.13) and the independence of X and ε

$$
\begin{aligned}
&nC[\hat{S}_0^{(n)}, \hat{A}_\ell^{(n)}] \\
&\quad = \frac{1}{n}\sum_{i,j=2}^{n}\{E[(f(X_{(i)}) + \sigma(X_{(i)})\varepsilon_{R_i^{-1}})(f(X_{(i)}) + \sigma(X_{(i)})\varepsilon_{R_{i-1}^{-1}}) \\
&\qquad\qquad \cdot (f(X_{(j)}) + \sigma(X_{(j)})\varepsilon_{R_j^{-1}})g_\ell(X_{(j)})] \\
&\qquad\qquad - E[(f(X_{(i)}) + \sigma(X_{(i)})\varepsilon_{R_i^{-1}})(f(X_{(i)}) + \sigma(X_{(i)})\varepsilon_{R_{i-1}^{-1}})] \\
&\qquad\qquad\quad \cdot E[(f(X_{(j)}) + \sigma(X_{(j)})\varepsilon_{R_j^{-1}})g_\ell(X_{(j)})]\} + o(1) \\
&\quad = \frac{1}{n}\sum_{i,j=2}^{n}\{E[f^2(X_{(i)})g_\ell(X_{(j)})f(X_{(j)})] - E[f^2(X_{(i)})]E[g_\ell(X_{(j)})f(X_{(j)})]\} \\
&\qquad + \frac{1}{n}\sum_{i=2}^{n}\{E[\sigma^2(X_{(i)})f(X_{(i)})g_\ell(X_{(i)}) \\
&\qquad\qquad + \sigma(X_{(i-1)})\sigma(X_{(i)})f(X_{(i)})g_\ell(X_{(i-1)})]\} + o(1) \\
&\quad = \frac{1}{n}\Bigg\{\sum_{i=2}^{n} E[f^2(X_{(i)})g_\ell(X_{(i)})f(X_{(i)})] - E[f^2(X_{(i)})]E[g_\ell(X_{(i)})f(X_{(i)})] \\
&\qquad\quad + E[\sigma^2(X_{(i)})f(X_{(i)})g_\ell(X_{(i)}) + \sigma(X_{(i-1)})\sigma(X_{(i)})f(X_{(i)})g_\ell(X_{(i-1)})]\Bigg\} \\
&\qquad + o(1) \\
&\quad = C[f^2(X), g_\ell(X)f(X)] + 2E[\sigma^2(X)f(X)g_\ell(X)] + o(1) = (\Sigma_{12}^T)_\ell + o(1)
\end{aligned}
$$

where the last equality follows from the Hölder continuity of f, σ^2 and g_ℓ (see (2.4)) while the third equality is a consequence of the fact that

$$
\begin{aligned}
&\frac{1}{n}\sum_{2\le i\ne j\le n} E[f^2(X_{(i)})g_\ell(X_{(j)})f(X_{(j)})] - E[f^2(X_{(i)})]E[g_\ell(X_{(j)})f(X_{(j)})] \\
&\quad = \frac{1}{n}\sum_{1\le i\ne j\le n} E[f^2(X_i)g_\ell(X_j)f(X_j)] - E[f^2(X_i)]E[g_\ell(X_j)f(X_j)] + O\left(\frac{1}{n}\right) \\
&\quad = O\left(\frac{1}{n}\right).
\end{aligned}
$$

The second term is treated by a similar argument, i.e.

$$
\begin{aligned}
&nC(\hat{A}_\ell^{(n)}, \hat{G}_{rs}^{(n)}) \\
&\quad = \frac{1}{n}\sum_{i,j=1}^{n} E[g_\ell(X_{(i)})f(X_{(i)})g_r(X_{(j)})g_s(X_{(j)})]
\end{aligned}
$$

$$- E[g_\ell(X_{(i)})f(X_{(i)})]E[g_r(X_{(j)})g_s(X_{(j)})]$$
$$= \frac{1}{n}\sum_{i=1}^{n} E[g_\ell(X_i)f(X_i)g_r(X_i)g_s(X_i)] - E[g_\ell(X_i)f(X_i)]E[g_r(X_i)g_s(X_i)]$$
$$= C[g_\ell(X)f(X), g_r(X)g_s(X)].$$

Because all other terms are treated in the same way the assertion of the Lemma follows. □

LEMMA 4.2. *Under the assumption of Lemma* 4.1 *we have*

$$\sqrt{n}\left\{\begin{pmatrix}\hat{A}^{(n)}\\ \hat{G}^{(n)}\end{pmatrix} - \begin{pmatrix}A\\ G\end{pmatrix}\right\} \overset{\mathcal{D}}{\Rightarrow} \mathcal{N}(0,\Sigma)$$

where the covariance matrix Σ *is defined in* (4.4).

PROOF. Recall the definition of $\hat{A}_0^{(n)}$ in (2.12) and $\hat{S}_0^{(n)}$ in (4.1). Let $\hat{A}_0^{(n)} = \hat{S}_0^{(n)} + \hat{R}_0^{(n)}$, where

$$\hat{R}_0^{(n)} = \frac{1}{n-1}\sum_{j=2}^{n}\{f(X_{(j)}) + \sigma(X_{(j)})\varepsilon_{R_j^{-1}}\}$$
$$\cdot \{f(X_{(j-1)}) - f(X_{(j)}) + \varepsilon_{R_{j-1}^{-1}}[\sigma(X_{(j-1)}) - \sigma(X_{(j)})]\}.$$

By the Hölder-continuity of f and σ^2 we have

$$E|\hat{R}_0^{(n)}| \le \frac{1}{n-1}\sum_{j=2}^{n}\{E[|f(X_{(j)}) - f(X_{(j-1)})||f(X_{(j)}) + \sigma(X_{(j)})\varepsilon_{R_j^{-1}}|]$$
$$+ E[|f(X_{(j)}) + \sigma(X_{(j)})\varepsilon_{R_j^{-1}}||\varepsilon_{R_{j-1}^{-1}}||\sigma(X_{(j-1)}) - \sigma(X_{(j)})|]\}$$
$$\le C_{f,\sigma}\frac{1}{n}\sum_{j=2}^{n}\{E[(X_{(j)} - X_{(j-1)})^{2\gamma}]E[(f(X_{(j)}) + \varepsilon_{R_j^{-1}})^2]\}^{1/2} = O(n^{-\gamma})$$

which implies $\sqrt{n}\hat{R}_0^{(n)} \xrightarrow{P} 0$. Here the last identity is an immediate consequence from the inequality

$$E[|X_{(j)} - X_{(j-1)}|^{2\gamma}] \le (E[|X_{(j)} - X_{(j-1)}|^2])^\gamma = O(n^{-2\gamma}).$$

This follows directly from formula (3.1.6), p. 28 in David (1970) if h is a uniform density on $[0,1]$. The general case is obtained along the lines in David (1970), p. 65 by means of a Taylor expansion observing that the density h of X is bounded away from 0.

Therefore it is sufficient to show the assertion of Lemma 4.2 with $\hat{A}^{(n)}$ replaced by

$$\tilde{A}^{(n)} = (\hat{S}_0^{(n)}, \hat{A}_1^{(n)}, \dots, \hat{A}_d^{(n)})^T.$$

To this end let

$$\ell = (\ell_0, \ldots, \ell_d, k_{11}, \ldots, k_{dd}, k_{12}, \ldots, k_{d-1,d})^T \in \mathbb{R}^{(d+1)(d+2)/2}$$

and consider the linear combination $\sqrt{n}\ell^T((\tilde{A}^{(n)} - A)^T, (\hat{G}^{(n)} - G)^T)^T$. Because of the independence of X and ε the distribution of this random variable is asymptotically equal to the distribution of

$$\frac{1}{\sqrt{n}} \sum_{i=2}^{n} (W_i - E[W_i]) \tag{4.5}$$

where

$$W_i = \ell_0(f(X_i) + \varepsilon_i)(f(X_i) + \varepsilon_{i-1}) + \sum_{j=1}^{d} \{\ell_j g_j(X_i)(f(X_i) + \varepsilon_i) + k_{jj} g_j^2(X_i)\}$$
$$+ \sum_{r=1}^{d} \sum_{s=r+1}^{d} k_{rs} g_r(X_i) g_s(X_i).$$

The random variables $W_1 - E[W_1], W_2 - E[W_2], \ldots$ form a sequence of 2-dependent random variables with mean 0 and asymptotic variance

$$\lim_{n\to\infty} V\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (W_i - E(W_i))\right] = \ell^T \Sigma \ell,$$

by Lemma 4.1, where Σ is defined by (4.4). Now a well known central limit theorem for m-dependent random variables (see Orey (1958)) and (4.5) show that

$$\sqrt{n}\ell^T \begin{pmatrix} \hat{A}^{(n)} - A \\ \hat{G}^{(n)} - G \end{pmatrix}$$

is asymptotically normal with mean zero and variance $\ell^T \sum \ell$. The assertion of Lemma 4.2 now follows from the Cramér-Wold device. □

PROOF OF THEOREM 2.1. Note that M^2 and $\hat{M}_n^2$ are invariant under a change of basis of the subspace U. Therefore, for the proof of Theorem 2.1, it is sufficient to deal with the case where the regression functions $g_1, \ldots, g_d$ are orthogonal with respect to the distribution of the explanatory variable X, i.e.

$$E[g_\ell(X) g_k(X)] = \delta_{\ell k}, \qquad \ell, k = 1, \ldots, d \tag{4.6}$$

where $\delta_{\ell k}$ denotes the Kronecker symbol. Assume first that $d \geq 2$. We calculate the gradient of the function $\rho(x, y)$ defined in (2.10), where $x = (x_0, \ldots, x_d)$, $y = (y_{11}, \ldots, y_{dd}, y_{12}, \ldots, y_{d-1,d})$. To this end we note that for a symmetric matrix $A = (a_{ij})_{ij=1}^m$ $(m \geq 2)$ we have

$$\begin{cases} \dfrac{\partial}{\partial a_{ii}} |A| = A^{ii}, & 2 \leq i \leq m \\ \dfrac{\partial}{\partial a_{ij}} |A| = 2(-1)^{i+j} A^{ij}, & 1 < i < j < m \end{cases} \tag{4.7}$$

where A^{ij} denotes the determinant obtained from $|A|$ by deleting the i-th row and j-th column. The formulas in (4.7) follow easily by Sylvester's identity (see Gantmacher (1959), p. 31) and an expansion of the resulting determinants. Recalling the definition of $\Gamma(x,y)$ and $\Gamma(y)$ in (2.11) and observing the orthogonality (4.6) (i.e. $G = I_d$) yields

$$\begin{aligned}
&\frac{\partial}{\partial x_0}\rho(x,y)\mid_{x=A,y=G}= 1\\
&\frac{\partial}{\partial x_j}\rho(x,y)\mid_{x=A,y=G}= -2A_j, \qquad j=1,\ldots,d\\
&\frac{\partial}{\partial y_{ij}}\rho(x,y)\mid_{x=A,y=G}\\
&\quad= \frac{\frac{\partial}{\partial y_{ij}}\Gamma(x,y)\mid_{x=A,y=G}\Gamma(G)-\Gamma(A,G)\frac{\partial}{\partial y_{ij}}\Gamma(y)\mid_{y=G}}{\Gamma^2(G)}\\
&\quad= \begin{cases}\Gamma^{i+1,i+1}(A,G)-\Gamma(A,G)=A_i^2, & \text{if } i=j\\ 2(-1)^{i+j}\Gamma^{i+1,j+1}(A,G)=2A_iA_j, & \text{if } 1\le i<j\le d\end{cases}
\end{aligned} \tag{4.8}$$

where $\Gamma^{i+1,j+1}(A,G)$ denotes the determinant obtained from $\Gamma(A,G)$ by deleting the $(i+1)$-th row and $(j+1)$-th column. If

$$\begin{aligned}
\mu^T &= \left(\frac{\partial}{\partial x_0}\rho,\ldots,\frac{\partial}{\partial x_d}\rho,\frac{\partial}{\partial y_{11}}\rho,\ldots,\frac{\partial}{\partial y_{dd}}\rho,\frac{\partial}{\partial y_{12}}\rho,\ldots,\frac{\partial}{\partial y_{d-1,d}}\rho\right)\Bigg|_{x=A,y=G}\\
&= (1,-2A_1,\ldots,-2A_d,A_1^2,\ldots,A_d^2,2A_1A_2,\ldots,2A_{d-1}A_d)
\end{aligned}$$

denotes the gradient of μ evaluated at (A,G), then we obtain from Lemma 4.1

$$\begin{aligned}
\mu^T\Sigma\mu &= E[\sigma^4(X)] + 4E[\sigma^2(X)f^2(X)] + V[f^2(X)]\\
&\quad + 4\sum_{\ell,k=1}^{d}\{E[\sigma^2(X)g_\ell(X)g_k(X)] + C[f(X)g_\ell(X),f(X)g_k(X)]\}\cdot A_\ell A_k\\
&\quad + \sum_{\ell,k=1}^{d} C[g_\ell^2(X),g_k^2(X)]A_\ell^2A_k^2\\
&\quad + 4\sum_{\ell<k,\ell'<k'} C[g_\ell(X)g_k(X),g_{\ell'}(X)g_{k'}(X)]A_\ell A_k A_{\ell'}A_{k'}\\
&\quad - 4\sum_{\ell=1}^{d}\{2E[\sigma^2(X)f(X)g_\ell(X)] + C[f^2(X),g_\ell(X)f(X)]\}A_\ell\\
&\quad + 2\sum_{\ell=1}^{d} C[f^2(X),g_\ell^2(X)]A_\ell^2 + 4\sum_{\ell<k} C[f^2(X),g_\ell(X)g_k(X)]A_\ell A_k\\
&\quad - 4\sum_{\ell,k=1}^{d} C[f(X)g_\ell(X),g_k^2(X)]A_\ell A_k^2
\end{aligned}$$

$$- 8\sum_{\ell=1}^{d}\sum_{\ell'<k'} C[f(X)g_\ell(X), g_{\ell'}(X)g_{k'}(X)]A_\ell A_{\ell'} A_{k'}$$

$$+ 4\sum_{\ell=1}^{d}\sum_{\ell'<k'} C[g_\ell^2(X), g_{\ell'}(X)g_{k'}(X)]A_\ell^2 A_{\ell'} A_{k'}$$

$$= E[\sigma^4(X)] + 4E\left[\sigma^2(X)\left\{f(X) - \sum_{\ell=1}^{d} A_\ell g_\ell(X)\right\}^2\right]$$

$$+ V\left[\left\{f(X) - \sum_{\ell=1}^{d} A_\ell g_\ell(X)\right\}^2\right]$$

$$= E[\sigma^4(X)] + 4E[\sigma^2(X)\{(f - P_U f)(X)\}^2] + V[\{(f - P_U f)(X)\}^2]$$

where the last equality follows from the fact that for an orthonormal system of regression functions the best approximation of f in $L^2(P^X)$ by elements of $U = \operatorname{span}\{g_1, \dots, g_d\}$ is given by

$$P_U f = \sum_{\ell=1}^{d} E[f(X)g_\ell(X)]g_\ell = \sum_{\ell=1}^{d} A_\ell g_\ell.$$

Because $\mu^T \Sigma \mu > 0$, the assertion of Theorem 2.1 now follows by Cramér's Theorem. The case $d = 1$ is proved in the same way observing that in this case the gradient can be calculated directly as $\mu^T = (1 - 2A_1, A_1^2)$. □

PROOF OF THEOREM 2.2. In a preliminary step we assume that the regression functions $g_1, \dots, g_d$ are orthonormal with respect to P^X and note that the gradient of $\Gamma(x, y)$ evaluated at (A, G) is given by

$$\nu = (1, -2A_1, \dots, -2A_d, \gamma_1, \dots, \gamma_d, 2A_1A_2, \dots, 2A_{d-1}A_d)$$

where

$$\gamma_\ell = A_0 - \sum_{j\neq\ell} A_j^2.$$

This gives for the asymptotic variance of $\sqrt{n}\Gamma(\hat{A}^{(n)}, \hat{G}^{(n)})$

$$\tau_2^2 = \nu^T\Sigma\nu = E[\sigma^4(X)] + 4E[\sigma^2(X)\{(f - P_U f)(X)\}^2] \tag{4.9}$$
$$+ V\left[f^2(X) - 2\sum_{\ell=1}^{d} A_\ell g_\ell(X)f(X) + \sum_{\ell=1}^{d}\gamma_\ell g_\ell^2(X) + 2\sum_{\ell<k} A_\ell A_k g_\ell(X)g_k(X)\right].$$

The asymptotic normality of $\sqrt{n}(\Gamma(\hat{A}^{(n)}, \hat{G}^{(n)}) - \Gamma(A, G))$ follows again by a routine argument as described in the proof of Theorem 2.1. The expression for τ_2^2

in (2.21) is obtained by a very tedious calculation which will be indicated in the following. Note first that

$$\begin{aligned}\eta_\ell &:= \eta(g_1, \dots, g_{\ell-1}, g_0, g_{\ell+1}, \dots, g_d)(g_\ell - P_{U_\ell} g_\ell) \\ &= \bar{\eta}(g_\ell, g_1, \dots, g_{\ell-1}, g_0, g_{\ell+1}, \dots, g_d)\end{aligned}$$

where

$$\bar{\eta}(f_0, \dots, f_d) = \begin{vmatrix} f_0 & \cdots & f_d \\ E[f_0(X)f_1(X)] & \cdots & E[f_d(X)f_1(X)] \\ \vdots & & \vdots \\ E[f_0(X)f_d(X)] & \cdots & E[f_d^2(X)] \end{vmatrix},$$

and we have used the well known fact that

$$f_0 - P_{\text{span}\{f_1,\dots,f_d\}} f_0 = \frac{\bar{\eta}(f_0, \dots, f_d)}{\eta(f_1, \dots, f_d)}$$

(see Achieser (1956), pp. 15–16). By (4.6) this gives ($g_0 = f$)

$$\bar{\eta}_\ell = \begin{cases} f - \sum_{\ell=1}^{d} A_\ell g_\ell, & \text{if } \ell = 0 \\ \gamma_\ell g_\ell - A_\ell \left(f - \sum_{j \neq \ell} A_j g_j \right), & \text{if } \ell = 1, \dots, d \end{cases}$$

and (4.9) reduces to

$$\begin{aligned}\tau_2^2 &= E[\sigma^4(X)] + 4E[\sigma^2(X)\{(f - P_U f)(X)\}^2] + V\left[\sum_{l=0}^{d} \bar{\eta}_\ell(X) g_\ell(X)\right] \\ &= E[\sigma^4(X)] + 4E[\sigma^2(X)\{(f - P_U f)(X)\}^2] \\ &\quad + V\left[\sum_{\ell=0}^{d} \eta(g_1, \dots, g_{\ell-1}, g_0, g_{\ell+1}, \dots, g_d)\{(g_\ell - P_{U_\ell} g_\ell)(X)\} g_\ell(X)\right]\end{aligned}$$

which is (2.21) for an orthonormal system of regression functions.

For the general case of not necessarily orthonormal regression functions let

$$(g_1, \dots, g_d)^T = H \cdot (\bar{g}_1, \dots, \bar{g}_d)^T$$

where H is a lower triangular matrix and $\bar{g}_1, \dots, \bar{g}_d$ are orthonormal with respect to the distribution P^X. Let $(\bar{A}, \bar{G})$ and $(\bar{A}^{(n)}, \bar{G}^{(n)})$ be the analogue of (A, G) [defined in (2.9)] and $(\hat{A}^{(n)}, \hat{G}^{(n)})$ [defined in (2.17) and (2.18)] where the regression functions g_ℓ are replaced by their orthonormal counterparts $\bar{g}_\ell$ ($\ell = 1, \dots, d$). It follows from Achieser (1956), p. 25, that $\det H = \sqrt{\Gamma(G)}$ and consequently we obtain

$$\begin{aligned}&\Gamma(A, G) = \Gamma(G)\Gamma(\bar{A}, \bar{G}), \\ &\hat{\Gamma}_n = \Gamma(\hat{A}^{(n)}, \hat{G}^{(n)}) = \Gamma(G)\Gamma(\bar{A}^{(n)}, \bar{G}^{(n)}) =: \Gamma(G)\bar{\Gamma}_n.\end{aligned} \tag{4.10}$$

Therefore the first part of the proof shows that

$$\sqrt{n}(\hat{\Gamma}_n - \Gamma(A,G)) = \sqrt{n}\Gamma(G)(\bar{\Gamma}_n - \Gamma(\bar{A},\bar{G})) \stackrel{\mathcal{D}}{\Rightarrow} \mathcal{N}(0,\bar{\tau}_2^2)$$

where

$$\begin{aligned}(4.11)\quad \bar{\tau}_2^2 &= \Gamma^2(G)\Bigg\{E[\sigma^4(X)] + 4E[\sigma^2(X)\{(f-P_Uf)(X)\}^2] \\ &\quad + V\left[\sum_{\ell=0}^{d}\eta(\bar{g}_1,\ldots,\bar{g}_{\ell-1},g_0,\bar{g}_{\ell+1},\ldots,\bar{g}_d)\right. \\ &\quad \left.\left.\cdot(\bar{g}_\ell - P_{U_\ell}\bar{g}_\ell)(X)\bar{g}_\ell(X)\right]\right\} \\ &= \Gamma^2(G)\{E[\sigma^4(X)] + 4E[\sigma^2(X)\{(f-P_Uf)(X)\}^2]\} \\ &\quad + V\left[\sum_{\ell=0}^{d}\eta(g_1,\ldots,g_{\ell-1},g_0,g_{\ell+1},\ldots,g_d)(g_\ell - P_{U_\ell}g_\ell)(X)g_\ell(X)\right]\end{aligned}$$

which gives (2.21). In (4.11) the last identity is a consequence of the identity ($g_0 = f$)

$$\begin{aligned}&\sum_{\ell=0}^{d}\eta(\bar{g}_1,\ldots,\bar{g}_{\ell-1},g_0,\bar{g}_{\ell+1},\ldots,\bar{g}_d)(\bar{g}_\ell - P_{U_\ell}\bar{g}_\ell)\bar{g}_\ell \\ &\quad = \sum_{\ell=0}^{d}\bar{\eta}(\bar{g}_\ell,\bar{g}_1,\ldots,\bar{g}_{\ell-1},g_0,\bar{g}_{\ell+1},\ldots,\bar{g}_d)\bar{g}_\ell \\ &\quad = \frac{1}{(\det H)^2}\sum_{\ell=0}^{d}g_\ell\bar{\eta}(g_\ell,g_1,\ldots,g_{\ell-1},g_0,g_{\ell+1},\ldots,g_d) \\ &\quad = \frac{1}{\Gamma(C)}\sum_{\ell=0}^{d}\eta(g_1,\ldots,g_{\ell-1},g_0,g_{\ell+1},\ldots,g_d)(g_\ell - P_{U_\ell}g_\ell)g_\ell\end{aligned}$$

($l = 0,\ldots,d$). This completes the proof of Theorem 2.2. □

PROOF OF LEMMA 3.1. We write

$$\begin{aligned}\ddot{\sigma}_n^4 = \frac{1}{4(n-3)}\sum_{i=2}^{n-2}&\{H_{i+2} + \sigma(X_{(i+2)})\varepsilon_{R^{-1}_{i+2}} - \sigma(X_{(i+1)})\varepsilon_{R^{-1}_{i+1}}\}^2 \\ &\cdot\{H_i + \sigma(X_{(i)})\varepsilon_{R^{-1}_i} - \sigma(X_{(i-1)})\varepsilon_{R^{-1}_{i-1}}\}^2\end{aligned}$$

where $H_i = f(X_{(i)}) - f(X_{(i-1)})$. By the Hölder continuity of f and σ and the independence of X and ε it follows that terms of the form

$$\frac{1}{n}\sum_{i=2}^{n-2}H_{i+2}H_i\sigma(X_{(i+1)})\varepsilon_{R^{-1}_{i+1}}\{\sigma(X_{(i)})\varepsilon_{R^{-1}_i} - \sigma(X_{(i-1)})\varepsilon_{R^{-1}_{i-1}}\}$$

converge to zero in probability. Therefore, a straightforward calculation shows

$$\hat{\sigma}_n^4 = \frac{1}{4n}\sum_{i=2}^{n-2}\{\sigma(X_{(i)})\varepsilon_{R_i^{-1}} - \sigma(X_{(i-1)})\varepsilon_{R_{i-1}^{-1}}\}^2 \cdot \{\sigma(X_{(i+2)})\varepsilon_{R_{i+2}^{-1}} - \sigma(X_{(i+1)})\varepsilon_{R_{i+1}^{-1}}\}^2 + o_P(1).$$

Similarly, terms of the form

$$\frac{1}{n}\sum_{i=2}^{n-2}\sigma(X_{(i-1)})\sigma(X_{(i)})\sigma(X_{(i+1)})\sigma(X_{(i+2)})\varepsilon_{R_{i-1}^{-1}}\varepsilon_{R_i^{-1}}\varepsilon_{R_{i+1}^{-1}}\varepsilon_{R_{i+2}^{-1}}$$
$$= \frac{1}{n}\sum_{i=2}^{n-2}\sigma^4(X_{(i)})\varepsilon_{R_{i-1}^{-1}}\varepsilon_{R_i^{-1}}\varepsilon_{R_{i+1}^{-1}}\varepsilon_{R_{i+2}^{-1}} + o_P(1)$$

converge to zero in probability, and we obtain

$$\hat{\sigma}_n^4 = \frac{1}{4n}\sum_{i=2}^{n-2}\{\sigma^2(X_{(i-1)})\varepsilon^2_{R_{i-1}^{-1}} + \sigma^2(X_{(i)})\varepsilon^2_{R_i^{-1}}\} \cdot \{\sigma^2(X_{(i+1)})\varepsilon^2_{R_{i+1}^{-1}} + \sigma^2(X_{(i+2)})\varepsilon^2_{R_{i+2}^{-1}}\} + o_P(1).$$

The convergence of $\hat{\sigma}_n^4$ to $E[\sigma^4(X)]$ now follows from the Hölder continuity of σ and the independence of X and ε. □

PROOF OF THEOREM 3.2. Using the notation $\Delta_j = t_j - t_{j-1}$ it follows from (3.8)

$$\max_{i=1}^{n}\left|\Delta_i h(t_i) - \frac{1}{n}\right| = o(n^{-3/2})$$

which implies for any function $k \in \text{Höl}_\gamma([0,1])$

$$\left|\int_0^1 k(t)h(t)dt - \frac{1}{n}\sum_{j=1}^{n}k(t_j)\right| \tag{4.12}$$
$$\leq \left|\int_0^1 k(t)h(t)dt - \sum_{j=1}^{n}\Delta_j h(t_j)k(t_j)\right| + \left|\sum_{j=1}^{n}k(t_j)\left(\frac{1}{n} - \Delta_j h(t_j)\right)\right|$$
$$\leq \sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}|h(t)k(t) - h(t_j)k(t_j)|dt + o(n^{-1/2})$$
$$\leq C_{h,k}\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}|t - t_j|^\gamma dt + o(n^{-1/2}) = o(n^{-1/2}),$$

and shows that

$$\lim_{n\to\infty}\hat{D}_{k,\ell}^{(n)} = D_{k,\ell} = \int_0^1 g_\ell(t)g_k(t)h(t)dt.$$

For the components of the vector $\hat{C}^{(n)}$ the calculation in (4.12) and the central limit theorem for m-dependent random variables (see Orey (1958)) show

$$\sqrt{n}(\hat{C}^{(n)} - C) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma)$$

with mean vector given by

$$C = \left(\int_0^1 f^2(t)h(t)dt, \int_0^1 f(t)g_1(t)h(t)dt, \ldots, \int_0^1 f(t)g_d(t)h(t)dt \right)^T$$

and covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{11} \end{pmatrix}$$

where $\Sigma_{11} \in \mathbb{R}$, $\Sigma_{12}^T \in \mathbb{R}^d$, $\Sigma_{22} \in \mathbb{R}^{d \times d}$ and

$$\Sigma_{11} = \int_0^1 \sigma^4(t)h(t)dt + 4 \int_0^1 \sigma^2(t)f^2(t)h(t)dt$$

$$(\Sigma_{12})_\ell = 2 \int_0^1 \sigma^2(t)f(t)g_\ell(t)h(t)dt, \qquad \ell = 1, \ldots, d$$

$$(\Sigma_{22})_{\ell,k} = \int_0^1 \sigma^2(t)g_\ell(t)g_k(t)h(t)dt, \qquad k, \ell = 1, \ldots, d.$$

The assertion of the theorem now follows by exactly the same reasoning as given for the proof of Theorem 2.1. □

Acknowledgements

The authors are grateful to R. Sawyerr for typing this manuscript and to an unknown Associate Editor and referee for some very helpful remarks, which yield to a substantial improvement of an earlier version of this paper.

References

Achieser, N. I. (1956). *Theory of Approximation*, Dover, New York.

Azzalini, A. and Bowman, A. (1993). On the use of nonparametric regression for checking linear relationships, *J. Roy. Statist. Soc. Ser. B*, **55**, 549–557.

Berger, J. O. and Delampady, M. (1987). Testing precise hypotheses, *Statist. Sci.*, **2**(3), 317–352.

Berger, J. O. and Sellke, T. (1987). Testing a point null hypothesis: the irreconcilability of P-values and evidence (with comments), *J. Amer. Statist. Assoc.*, **82**, 112–122.

Berkson, J. (1942). Tests of significance considered as evidence, *J. Amer. Statist. Assoc.*, **37**, 325–335.

Breiman, L. and Meisel, W. S. (1976). General estimates of the intrinsic variability of data in nonlinear regression models, *J. Amer. Statist. Assoc.*, **71**, 301–307.

Brodeau, F. (1993). Tests for the choice of approximative models in nonlinear regression when the variance is unknown, *Statistics*, **24**, 95–106.

David, H. A. (1970). *Order Statistics*, Wiley, New York.

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

Dette, H. and Munk, A. (1998). Validation of linear regression models, *Ann. Statist.* (to appear).

Eubank, R. L. and Hart, J. D. (1992). Testing goodness-of-fit in regression via order selection criteria, *Ann. Statist.*, **20**, 1412–1425.

Eubank, R. L. and Spiegelman, C. H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques, *J. Amer. Statist. Assoc.*, **85**, 387–392.

Gantmacher, F. R. (1959). *The Theory of Matrices*, Vol. I, II, Chelsea Publishing, New York.

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

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

Härdle, W. and Marron, J. S. (1990). Semiparametric comparison of regression curves, *Ann. Statist.*, **18**, 83–89.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits, *Ann. Statist.*, **21**, 1926–1947.

Hauck, W. W. and Anderson, S. (1996). Comment on 'Bioequivalence trials, intersection-union tests and equivalence confidence sets by R. L. Berger and J. C. Hsu', *Statist. Sci.*, **11**, 283–319.

Hermann, E., Wand, M. P., Engel, J. and Gasser, T. (1995). A bandwidth selector for bivariate regression, *J. Roy. Statist. Soc. Ser. B*, **57**, 171–180.

Jayasuriya (1996). Testing for polynomial regression using nonparametric regression techniques, *J. Amer. Statist. Assoc.*, **91**, 1626–1631.

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

Kozek, A. S. (1991). A nonparametric fit of a parametric model, *J. Multivariate Anal.*, **37**, 66–75.

MacKinnon, J. G. (1992). Model specification tests and artificial regressions, *Journal of Economic Literature*, **30**, 102–146.

Orey, S. (1958). A central limit theorem for m-dependent random variables, *Duke Math. J.*, **52**, 543–546.

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.

Schervish, M. J. (1996). P-values: what they are and what they are not, *Amer. Statist.*, **50**, 203–206.

Staniswalis, J. and Severini, T. (1991). Diagnostics for assessing regression models, *J. Amer. Statist. Assoc.*, **86**, 684–692.

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

Stute, W. and Manteiga, W. G. (1996). NN goodness-of-fit tests for linear models, *J. Statist. Plann. Inference*, **53**, 75–92.

Zwanzig, S. (1980). The choice of approximative models in nonlinear regression, *Math. Operationsforschung Statist., Ser. Statistics*, **11**, 23–47.