The Annals of Statistics
1998, Vol. 26, No. 2, 778–800

VALIDATION OF LINEAR REGRESSION MODELS

BY HOLGER DETTE AND AXEL MUNK[1]

Ruhr-Universität Bochum

A new test is proposed in order to verify that a regression function, say g, has a prescribed (linear) parametric form. This procedure is based on the large sample behavior of an empirical L^2-distance between g and the subspace U spanned by the regression functions to be verified. The asymptotic distribution of the test statistic is shown to be normal with parameters depending only on the variance of the observations and the L^2-distance between the regression function g and the model space U. Based on this result, a test is proposed for the hypothesis that "g is not in a preassigned L^2-neighborhood of U," which allows the "verification" of the model U at a controlled type I error rate. The suggested procedure is very easy to apply because of its asymptotic normal law and the simple form of the test statistic. In particular, it does not require nonparametric estimators of the regression function and hence, the test does not depend on the subjective choice of smoothing parameters.

1. Introduction. Consider the homoscedastic regression model

$$Y_{i,n} = Y(t_{i,n}) = g(t_{i,n}) + \varepsilon(t_{i,n}), \qquad 1 \le i \le n, \tag{1.1}$$

where the design $\{t_{1,n}, \ldots, t_{n,n}\} \subset [0, 1] \subset \mathbb{R}$ is assumed to be fixed, $t_{1,n} = 0$, $t_{n,n} = 1$ and $t_{i,n} < t_{j,n}$ $(1 \le i < j \le n)$. The index n is omitted whenever this dependence is clear from the context. For $t \in [0, 1]$ the random variable $Y(t)$ has finite expectation $EY(t) = g(t)$. The errors $\varepsilon_{i,n} = \varepsilon(t_{i,n})$ form a triangular array of rowwise independent random variables with expectation 0 and variance $\sigma^2 < \infty$. Further, let $U_d = \text{span}\{f_1, \ldots, f_d\}$, where $f_1, \ldots, f_d$ denote d linearly independent regression functions which are to be shown to "explain" g; that is, we are concerned with the assessment of $g \in U_d$ at a controlled type I error rate α.

Significant effort has been devoted to this problem during the last two decades [for some early references the reader may consult Shillington (1979) or Neil and Johnson (1985)] because, in applied regression analysis, parametric models are usually preferred to a purely nonparametric approach. In particular, a parametric model $g = \sum_{l=1}^{d} a_l f_l$ admits a direct interpretation of the observed effects in terms of the parameters $a_1, \ldots, a_d$. In addition, the available information of the observations is increased by applying more efficient estimating or testing procedures—provided the assumed model is valid. Therefore, many authors point out that the applied working statistician should

Received January 1996; revised August 1997.

[1]Supported in part by a postdoctoral fellowship, Deutsche Forschungsgemeinschaft, Technische Universität Dresden

AMS 1991 *subject classifications*. Primary 62G05; secondary 62G10, 62G30, 62G07.

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

always check a preassigned model U_d by means of testing the hypothesis

$$H_0\colon g \in U_d \quad \text{versus} \quad K_0\colon g \notin U_d, \tag{1.2}$$

at a controlled error rate α before analyzing the data. In the meantime, a vast amount of literature is devoted to this problem; the works mainly differ in the choice of the underlying model and the technique used to treat the problem (1.2). For an overview, see Davison and Tsai (1992), Härdle and Mammen (1993) or Stute (1997). Various authors assumed a random design. In particular, we refer to Stute (1997) for an approach based on marked empirical processes and to Dieboldt (1995) who obtained exact rates by means of strong approximation techniques. Others focused their interest on more restricted hypotheses [see Cox, Koh, Wahba and Yandell (1988) for the case of polynomial regression or Eubank and Spiegelmann (1990), Barry and Hartigan (1990) or Yanagimoto and Yanagimoto (1987) for the linear case]. Müller (1992) suggested a lack of fit test in the heteroscedastic case which is based on a kernel estimate for f at a given set of design points. Test statistics based on kernel estimation methods in combination with bootstrap techniques were also proposed by Härdle and Mammen (1993). Staniswalis and Severini (1991) based their test criterion on a comparison of the MLE of the regression parameter under the null model and a nonparametric MLE under sufficiently smooth alternatives. This requires kernel estimation of the nonparametric likelihood. Spline smoothing methods and a Fourier analytic approach were suggested by Eubank and Spiegelmann (1990) and Eubank and Hart (1992).

Various authors argue that, even if the null in (1.2) is accepted with a rather large observed p-value, there need not be any empirical evidence for the presence of U_d [see Berger and Delampady (1987), Staudte and Sheather (1990)]. For a careful discussion in the context of model diagnostics, we refer to MacKinnon (1992). These authors point out that it may sometimes be preferable to reformulate the hypothesis in (1.2) into a test problem, which allows us to show g to be close to U_d at a controlled error rate. As a measure of discrepancy M from the model U_d, we suggest the minimal distance between g and U_d with respect to an L^2-norm M, where $M^2 = M^2(g) = \min_{v \in U_d} ||v - g||_2$. In order to overcome the uncertainty which emerges from accepting H_0 in (1.2), we reformulate these hypotheses into

$$H_\Delta\colon M > \Delta \quad \text{versus} \quad K_\Delta\colon M \le \Delta \tag{1.3}$$

and call this a "precise" testing problem. Note that we use the concept of testing precise hypotheses in a slightly different way than Berger and Delampady (1987), because we interchange the alternative and null hypothesis. In formula (1.3), $\Delta > 0$ is a sufficiently small constant such that, whenever $M(g) \le \Delta$, the experimenter agrees to analyze the data by the linear model U_d. Observe that K_Δ is equivalent to the existence of regression parameters $a_1^*, \ldots, a_d^*$ such that $g \approx \sum a_l^* f_l$ up to an error Δ. If $\Delta = 0$, g is explained completely by the linear model U_d. The final decision which of the hypotheses (1.2) or (1.3) is to be tested will certainly depend on the subsequent data analysis of the statistician, such as prediction, estimation of the regression parameters or designing

future experiments. Note that the testing problem (1.3) is conceptually related to bioequivalence testing, which is to show the therapeutical equivalence of different formulations of an active ingredient of a drug. For a detailed discussion of the choice of the null hypotheses and other methodological aspects in the context of bioequivalence testing, we refer to Metzler (1974), Mandallaz and Mau (1981), Schuirmann (1987) and Chow and Liu (1992).

In Section 2 we propose an estimator $\hat{M}_n^2$ of the distance M^2 between g and the model U_d. In Section 3 we prove an asymptotic normal law for $\sqrt{n}(\hat{M}_n^2 - M^2)$ with parameters depending only on M^2 and σ^2. In Section 4 this result is applied for the construction of asymptotic confidence intervals and tests for the hypothesis (1.2) as well as H_Δ. In particular, our approach does not require the estimation of the regression function and hence is independent of a subjective choice of smoothing parameters. It may be especially appealing for the applied working statistician that the proposed test is very simple to perform because only estimation of M^2 and the variance σ^2 is required. Moreover, the goodness-of-fit test proposed in this paper allows a simple calculation of the required sample size in order to control its type II error. In Section 5 we show numerically, for various examples, that the nominal size is maintained with high accuracy, even for small sample sizes. Finally, in a simulation study, the power is compared to the test of Eubank and Hart (1992) which is the most similar in spirit to the work of this paper. Roughly speaking, this study demonstrates that, for different alternatives, each test may improve on the other and vice versa. In summary, the proposed test is uniformly powerful in all nonparametric directions with the same L^2 norm, whereas Eubank and Hart's test is a powerful tool to protect against specific alternatives but does not provide uniform guardance against misspecification.

2. Definition and estimation of M^2. Throughout this paper, let h denote a density function on the interval $[0, 1]$ which is positive and Hölder continuous of order $\gamma > 1/2$; that is, $h \in \text{Höl}_\gamma[0, 1]$, $\gamma > 1/2$. We will call h the design density and assume that the design $\{t_{1,n}, \dots t_{n,n}\}$ satisfies

$$\max_{i=2}^{n} \left| \int_{t_{i-1,n}}^{t_{i,n}} h(t)\, dt - \frac{1}{n} \right| = o\left(\frac{1}{n}\right). \tag{A1}$$

Note that this assumption generalizes designs which are a regular sequence in the sense of Sacks and Ylvisaker (1970), that is,

$$\int_0^{t_{i,n}} h(t)\, dt = \frac{i-1}{n-1}, \qquad i = 1, \dots, n. \tag{2.1}$$

The most important case is obtained for the constant density $h \equiv 1$ which gives a nearly uniform design up to $o(1/n)$. According to the design density, we define L^2 as the set of all functions v on the interval $[0, 1]$ with finite weighted L^2-norm

$$\|v\|_2^2 = \int_0^1 v^2(x) h(x)\, dx$$

and assume that the (unknown) regression function g in (1.1) and the linearly independent functions $f_1, \ldots, f_d$ of the linear model U_d satisfy $g, f_1, \ldots, f_d \in L^2$. We also admit the case $d = 0$, which is interpreted as the subspace of dimension 0, that is, $U_0 = \{0\}$. As a measure of discrepancy M^2 between the regression function g and the subspace $U_d = \text{span}\{f_1, \ldots, f_d\} \subset L^2$, we use the minimal distance

$$M^2 = M^2(g) = \min_{v \in U_d} \|g - v\|_2 \,, \tag{2.2}$$

where the dependence of g is omitted whenever it is clear from the context. If the inner product in L^2 with respect to the design density h is denoted by

$$\langle u, v \rangle := \int_0^1 u(x)v(x)h(x)\,dx, \qquad u, v \in L^2$$

we obtain that the minimal distance in (2.2) can be represented as

$$M^2 = \frac{\Gamma(g, f_1, \ldots, f_d)}{\Gamma(f_1, \ldots, f_d)}, \tag{2.3}$$

where $\Gamma(v_1, \ldots, v_k)$ denotes the Gramian determinant $|(\langle v_i, v_j \rangle)|_{i,\, j=1,\ldots,\, k}$ for $v_1, \ldots, v_k \in L^2$ For a derivation of the identity (2.3), we refer to Achieser (1956) page 16. The expression in (2.3) simplifies to

$$M^2 = \|g\|_2^2 - \sum_{l=1}^{d} |\langle g, f_l \rangle|^2 \tag{2.4}$$

if the system $\{f_i\}_{i=1,\ldots,d}$ of regression functions is orthonormal with respect to the measure with density h.

The multilinearity of M^2 suggests estimating M^2 by replacing in $\Gamma(g, f_1, \ldots, f_d)$ all elements (containing the unknown regression function g) by an empirical counterpart. To this end define for functions $u, v \in L^2$ the vectors $u_n = (u(t_{1,n}), \ldots, u(t_{n,n}))^t$, $v_n = (v(t_{1,n}), \ldots, v(t_{n,n}))^t$ and a bilinear form on $\mathbb{R}^n$ by

$$\langle u_n, v_n \rangle_n := \sum_{i=1}^{n} \Delta_{i,n} w_{i,n} h(t_{i,n}) u(t_{i,n}) v(t_{i,n}), \tag{2.5}$$

which can be considered as a numerical approximation of the inner product $\langle u, v \rangle$ in L^2. Here the $w_i = w_{i,n}$ are (possibly negative) weights and $\Delta_i = \Delta_{i,n} = t_{i,n} - t_{i-1,n}$ $(i = 1, \ldots, n)$ denote the differences of successive locations of the measurements (note that $t_{1,n} = 0$ and $t_{n,n} = 1$). If $Y = (Y_1, \ldots, Y_n)^t$ denotes the vector of observations, then we estimate $\langle g, f_l \rangle$ and $\langle g, g \rangle$ by

$$\langle Y, f_{l,n} \rangle_n = \sum_{i=1}^{n} \Delta_{i,n} w_{i,n} h(t_{i,n}) f_l(t_{i,n}) Y_i, \qquad l = 1, \ldots, d \tag{2.6}$$

and

$$\langle Y, Y \rangle_n - \sigma^2 = \sum_{i=1}^{n} \Delta_{i,n} w_{i,n} h(t_{i,n}) Y_i^2 - \sigma^2, \tag{2.7}$$

respectively, where $f_{l,n} = (f_l(t_{1,n}), \ldots, f_l(t_{n,n}))^t$ and, for the moment, we assume σ^2 to be known. In (2.5), (2.6) and (2.7) it is important that we have the choice of the weights $(w_{i,n})_{i=1,\ldots,n}$ while the $(\Delta_{i,n})_{i=1,\ldots,n}$ are determined by the experiment. This allows a flexible choice of the numerical integration rule for M^2 (see the subsequent discussion). In order to integrate the design density h exactly, we assume further that

$$\sum_{i=1}^{n} \Delta_{i,n} w_{i,n} h(t_{i,n}) = 1 \quad \text{for all } n \geq 1. \tag{2.8}$$

Finally, we obtain a class of estimators for M^2 as

$$M_n^2 := \frac{\Gamma_n(Y, f_{1,n}, \ldots, f_{d,n})}{\Gamma(f_1, \ldots, f_d)} - \sigma^2, \tag{2.9}$$

where $\Gamma_n(\cdot)$ is defined as the determinant which is obtained by replacing the inner products $\langle g, g\rangle$, $\langle g, f_l\rangle$ $(l = 1, \ldots, d)$ in the Gramian determinant $\Gamma(g, f_1, \ldots, f_d)$ by the inner products of the vectors Y, $f_{1,n}, \ldots, f_{d,n} \in \mathbb{R}^n$ according to (2.6) and (2.7), respectively. In order to guarantee that

$$E[M_n^2 - M^2] = o(n^{-1/2}) \quad \text{as } n \to \infty, \tag{2.10}$$

we assume throughout this paper that there exist an integer $K \in \mathbb{N}$ and for every $n \in \mathbb{N}$ a finite set $E_n \subset \{1, \ldots, n\}$, such that $\operatorname{card}(E_n) \leq K$ and

$$\max_{i \in \{1,\ldots,n\} \setminus E_n} |w_{i,n} - 1| = o(n^{-1/2}), \tag{A2}$$

$$g, f_1, \ldots, f_d \in \text{Höl}_\gamma[0, 1] \quad \text{for some } \gamma > 1/2. \tag{A3}$$

If (A1)–(A3) are satisfied, a straightforward but tedious calculation shows that

$$E[M_n^2] = \frac{\Gamma_n(g_n, f_{1,n}, \ldots, f_{d,n})}{\Gamma(f_1, \ldots, f_d)} - \frac{d}{n}\sigma^2 + o\left(\frac{1}{n}\right), \tag{2.11}$$

where $g_n = (g(t_{1,n}), \ldots g(t_{n,n}))^t$ denotes the vector of values of the function g at the design points. The first term on the right-hand side in (2.11) is obtained by replacing the inner products in (2.3), which contain the regression function g, by the corresponding numerical approximations via (2.5). Observe that M^2 (and hence any testing problem concerning M^2) is not affected by changing the basis of U_d. Therefore, it is reasonable to demand that M_n^2 shares this property. The following proposition gives a sufficient condition for this invariance property.

PROPOSITION 2.1. *Assume that there exist weights w_l, $l = 1, \ldots, n$, such that*

$$\langle f_i, f_j\rangle = \textstyle\sum_{l=1}^n w_l \Delta_l h(t_l) f_i(t_l) f_j(t_l) = \langle f_{i,n}, f_{j,n}\rangle_n, \quad 1 \leq i \leq j \leq d. \tag{2.12}$$

Then the following statements hold.

(i) *Let* $u_1, \ldots, u_d$ *denote an arbitrary basis of* U_d *and* $u_{l,n} = (u_l(t_1), \ldots, u_l(t_n))^t$ $(l = 1, \ldots, d)$; *then*

$$\langle u_i, u_j \rangle = \langle u_{i,n}, u_{j,n} \rangle_n, \qquad 1 \le i \le j \le d. \tag{2.13}$$

(ii) *The empirical distance* M_n^2 *in* (2.9) *is invariant with respect to change of the basis of* U_d.

PROOF. Choose weights such that (2.12) holds, recall the definition of $f_{i,n} = (f_i(t_1), \ldots, f_i(t_n))^t$, $i = 1, \ldots, d$ and define a $d \times n$ matrix $\mathbf{f}_n = (f_{1,n}, \ldots, f_{d,n})^t$. Let $u_1, \ldots, u_d$ be a basis of U_d and $\mathbf{A}$ the nonsingular matrix, which represents the change of the basis vectors, that is, $\mathbf{u} = (u_1, \ldots, u_d)^t = \mathbf{A}\mathbf{f}$ where $\mathbf{f} = (f_1, \ldots, f_d)^t$. Define further a $(d+1) \times (d+1)$ matrix

$$\tilde{\mathbf{A}} = 1 \oplus \mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{A} \end{pmatrix},$$

$\tilde{\mathbf{f}}_n = (Y, \mathbf{f}_n^t)^t$, $\mathbf{u}_n = \mathbf{A}\mathbf{f}_n$, $\tilde{\mathbf{u}}_n = \tilde{\mathbf{A}}\tilde{\mathbf{f}}_n$ and denote the $n \times n$ diagonal matrix with diagonal elements $h(t_i)\Delta_i w_i$ $(i = 1, \ldots, n)$ as $\mathbf{W}$. Then it follows

$$\begin{aligned} \big(\langle u_i, u_j \rangle\big)_{i,j=1,\ldots,d} &= \int \mathbf{u}(t)\mathbf{u}^t(t)h(t)\,dt = \mathbf{A}\int \mathbf{f}(t)\mathbf{f}^t(t)h(t)\,dt\mathbf{A}^t \\ &= \mathbf{A}\mathbf{f}_n\mathbf{W}\mathbf{f}_n^t\mathbf{A}^t = \mathbf{u}_n\mathbf{W}\mathbf{u}_n^t = \big(\langle u_{i,n}, u_{j,n} \rangle_n\big)_{i,j=1,\ldots,d}, \end{aligned}$$

which proves assertion (i). Note, that $\Gamma(f_1, \ldots, f_d) = \Gamma_n(\mathbf{f}_n^t) \neq 0$ because of the linear independence of $f_1, \ldots, f_d$. Then we have from (2.13),

$$\begin{aligned} M_n^2 + \sigma^2 &= \frac{\Gamma_n(Y, f_{1,n}, \ldots, f_{d,n})}{\Gamma(f_1, \ldots, f_d)} = \frac{\Gamma_n(Y, f_{1,n}, \ldots, f_{d,n})}{\Gamma_n(f_{1,n}, \ldots, f_{d,n})} \\ &= \frac{|\tilde{\mathbf{f}}_n\mathbf{W}\tilde{\mathbf{f}}_n^t|}{|\mathbf{f}_n\mathbf{W}\mathbf{f}_n^t|} = \frac{|\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{u}}_n\mathbf{W}\tilde{\mathbf{u}}_n^t(\tilde{\mathbf{A}}^{-1})^t|}{|\mathbf{A}^{-1}\mathbf{u}_n\mathbf{W}\mathbf{u}_n^t(\mathbf{A}^{-1})^t|} = \frac{|\tilde{\mathbf{u}}_n\mathbf{W}\tilde{\mathbf{u}}_n^t|}{|\mathbf{u}_n\mathbf{W}\mathbf{u}_n^t|} = \frac{\Gamma_n(Y, u_{1,n}, \ldots, u_{d,n})}{\Gamma(u_1, \ldots, u_d)}, \end{aligned}$$

where the last equality follows from (i). This implies that for given weights $w_1, \ldots, w_n$ satisfying (2.12), the estimator M_n^2 remains invariant under change of basis of U_d. □

Note that the condition for the weights in (2.12) means that the numerical quadrature formula integrates products of functions of the subspace U_d exactly. It is strictly recommended to determine the weights with respect to (2.12) and (2.8) because the accuracy of the normal approximation of M_n^2 derived in the following section is increased significantly (see the simulation results in Section 5). In order to guarantee that (2.12) and (2.8) hold, we have

to solve

$$\mathbf{b} = \begin{pmatrix} 1 \\ \langle f_1, f_1 \rangle \\ \langle f_1, f_2 \rangle \\ \vdots \\ \langle f_d, f_d \rangle \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \cdots & 1 \\ f_1(t_1)f_1(t_1) & \cdots & f_1(t_n)f_1(t_n) \\ f_1(t_1)f_2(t_1) & \cdots & f_1(t_n)f_2(t_n) \\ \vdots & \cdots & \vdots \\ f_d(t_1)f_d(t_1) & \cdots & f_d(t_n)f_d(t_n) \end{pmatrix} \begin{pmatrix} w_1\Delta_1 h(t_1) \\ w_2\Delta_2 h(t_2) \\ \vdots \\ w_n\Delta_n h(t_n) \end{pmatrix} = \mathbf{F}\mathbf{w}, \tag{2.14}$$

which is a system of $d(d+1)/2+1$ equations for n unknowns. Note that the vector $\mathbf{w}$ in (2.14) is defined as $\mathbf{w}= (w_1\Delta_1 h(t_1), \ldots, w_n\Delta_n h(t_n))^t$ and determines the integration weights $w_1, \ldots, w_n$ [because $\Delta_j h(t_j)$ is known]. In general, a solution of (2.14) exists if and only if $\mathbf{b} \in \text{Range}(\mathbf{F})$, which will be true in most cases of practical interest. For example, if $n \geq d(d+1)/2+1$ and $\tilde{U}_d := \text{span}\{1, f_i f_j\}_{1\leq i\leq j\leq d}$ is a Tchebycheff space [Zielke (1979)] we have $rg(\mathbf{F}) = d(d+1)/2+1$ and hence a solution of (2.14) exists. Note that this solution will not be unique in general. A possible choice for the weights w_i is to use $n - d(d+1)/2 - 1$ equal weights and to determine the remaining $d(d+1)/2+1$ weights such that (2.12) and (2.8) are valid. Observe further, that in some cases $\dim \tilde{U}_d := k+1 < d(d+1)/2+1$. For example, in a d-dimensional polynomial regression model $f_i(t) = t^{i-1}$, $i = 1, \ldots, d$ we have $\tilde{U}_d = \text{span}\{1, x, \ldots, x^{2d-2}\}$. This implies $rg(\mathbf{F}) = 2d-1$ because $\mathbf{F}$ contains a minor of dimension $2d-1$ which is the Vandermonde determinant. Therefore we can find weights which satisfy (2.14) provided that $n \geq 2d-1$. In those cases where (2.14) is not solvable, it is still reasonable to choose weights such that as many equations as possible in (2.14) are satisfied. However, the following result shows that if n increases the weights can always be determined in accordance to (2.14) and (A2).

THEOREM 2.1. *For any linear regression space U_d with basis functions $f_1, \ldots, f_d$ satisfying* (A3), *there exists an $N \in \mathbb{N}$, such that for all $n \geq N$ one can find weights $(w_{1,n}, \ldots, w_{n,n})$ satisfying* (2.14) *and condition* (A2).

PROOF. Note that $1 \leq \dim(\tilde{U}_d) = k+1 \leq d(d+1)/2+1$. Let $u_i = f_i^1 f_i^2$, $i = 1, \ldots, k+1$ be a basis of $\tilde{U}_d$ where $f_i^1, f_i^2 \in \{1, f_1, \ldots, f_d\}$. Then (2.14) is equivalent to

$$(\langle f_1^1, f_1^2 \rangle, \ldots, \langle f_{k+1}^1, f_{k+1}^2 \rangle)^t = \tilde{\mathbf{F}}\mathbf{w}, \tag{2.15}$$

where

$$\tilde{\mathbf{F}} = \left(f_j^1(t_i) f_j^2(t_i)\right)_{\substack{i=1,\dots,n \\ j=1,\dots,k+1}}.$$

Choose $t_1, \dots, t_{k+1}$, such that

$$|u_j(t_i)|_{\substack{i=1,\dots,k+1 \\ j=1,\dots,k+1}} \neq 0. \tag{2.16}$$

Now there exists an open neighborhood $U = U(t_1, \dots, t_{k+1})$ such that (2.16) holds in U because the basis functions u_i are continous by (A3). From (A1) it follows that there exists N, such that for all $n \geq N$ $rg(\tilde{\mathbf{F}}) = k+1$, that is, $\operatorname{Im} \tilde{\mathbf{F}} = \mathbb{R}^{k+1}$. Hence there exists a solution of (2.15) and consequently an $n-(k+1)$-dimensional subspace of solutions of (2.14). Condition (A2) can therefore easily be satisfied by choosing a solution with $n-k-1$ components equal to one and solving (2.14) for the remaining components. □

It is also worthwhile to mention that the scaling property

$$\frac{\Gamma(\alpha g, \alpha f_1, \dots, \alpha f_d)}{\Gamma(\alpha f_1, \dots, \alpha f_d)} = \alpha^2 \frac{\Gamma(g, f_1, \dots, f_d)}{\Gamma(f_1, \dots, f_d)}$$

carries over to M_n^2, provided condition (2.12) is valid. In the particular case when $w_i \geq 0$, $i = 1, \dots, n$ we may interpret $\{M_n^2 + \sigma^2\}^{1/2}$ as the minimal distance between the vector of observations Y and the regression space $U_{d,n} = \operatorname{span}\{f_{1,n}, \dots, f_{d,n}\}$ with respect to the weighted inner product $\langle \cdot, \cdot \rangle_n$, with corresponding norm

$$\|v_n\|_n^2 = |\langle v_n, v_n \rangle_n|^2 = \sum_{i=1}^n \Delta_i w_i h(t_i) v^2(t_i),$$

which is distorted by the variance σ^2. In other words, if (2.12) is valid, we have

$$M_n^2 + \sigma^2 = \sum_{i=1}^n \Delta_i w_i h(t_i) \left(Y_i - \sum_{l=1}^d \hat{a}_l f_l(t_i) \right)^2, \tag{2.17}$$

where $\hat{a} = (\hat{a_1}, \dots, \hat{a_d})^t$ denotes the weighted LSE of the parameter vector $(a_1, \dots, a_d)^t$ for which the minimum in (2.17) is attained. For an unweighted LSE approach see Zwanzig (1980) and Bordeau (1993).

We stress that the recommendations given above concerning the choice of the weights in $\hat{M}_n^2$ can be interpreted also as a guidance for the choice of weights of the weighted least squares statistics in (2.17). This certainly reinforces any weighted least squares approach. The estimates $\hat{a}_l$ now have an immediate interpretation as the Fourier coefficients of the best approximation by a certain subspace with respect to the norm $\|\cdot\|_n$. In the case of (2.12) this is just the regression space U_d. This may serve as a rough explanation why the accuracy of the subsequent asymptotics is extremely high when the model is approximately valid (cf. Section 5).

All that remains is the estimation of the unknown variance σ^2 in (2.9). Various procedures for this purpose have been presented in the literature by

Breiman and Meisel (1976), Whaba (1978), Silvermann (1985), Rice (1984), Gasser, Sroka and Jennen-Steinmetz (1986) and Hall and Marron (1990). An extensive simulation study [see Dette, Munk and Wagner (1998)] motivates the use of the following estimator [see Rice (1984) or Gasser, Sroka and Jennen-Steinmetz (1986)]. We define pseudoresiduals

$$\tilde{\varepsilon}_i = \frac{t_{i+1}-t_i}{t_{i+1}-t_{i-1}} Y(t_{i-1}) + \frac{t_i - t_{i-1}}{t_{i+1}-t_{i-1}} Y(t_{i+1}) - Y(t_i), \qquad i = 2, \ldots, n-1$$

and an estimate of σ^2 as

$$S_\varepsilon^2 = \frac{1}{n-2} \sum_{i=2}^{n-1} c_i^2 \tilde{\varepsilon}_i^{\,2}, \tag{2.18}$$

where $c_i^2 = (a_i^2 + b_i^2 + 1)^{-1}$, $a_i = (t_{i+1} - t_i)/(t_{i+1} - t_{i-1})$ and $b_i = (t_i - t_{i-1})/(t_{i+1} - t_{i-1})$. Define C as the $(n-2) \times (n-2)$ diagonal matrix with elements $C_{ii} = c_{i+1}$ and A as the $(n-2) \times n$ tridiagonal matrix with elements $A_{ii} = a_{i+1}$, $A_{i,\,i+1} = -1$, $A_{i,\,i+2} = b_{i+1}$ then S_ε^2 can be conveniently written as a quadratic form

$$S_\varepsilon^2 = \frac{1}{n-2} Y^t D Y =: Y^t D_n Y, \tag{2.19}$$

where $D = A^t C^2 A$. In the particular case of an equidistant design $t_i = (i-1)/(n-1)$, $i = 1, \ldots, n$ [which corresponds to the constant design density in (2.1)], the estimator in (2.19) reduces

$$\frac{1}{6(n-2)} \sum_{i=2}^{n-1} (Y_{i+1} + Y_{i-1} - 2Y_i)^2 \;=\; Y^t B_n Y,$$

where

$$\begin{aligned} B_n &:= \frac{1}{(n-2)} B \\ &= \frac{1}{6(n-2)} \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & & & \\ -2 & 5 & -4 & 1 & 0 & \cdots & & \\ 1 & -4 & 6 & -4 & 1 & 0 & \cdots & \\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & 0 & 1 & -4 & 6 & -4 & 1 & 0 \\ & \cdots & 0 & 1 & -4 & 6 & -4 & 1 \\ & & \cdots & 0 & 1 & -4 & 5 & -2 \\ & & & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}. \end{aligned} \tag{2.20}$$

Observing (2.9) and (2.18), we now define our estimator of the L^2-distance M^2 by its empirical counterpart

$$\hat{M}_n^2 = \frac{\Gamma_n(Y, f_{1,n}, \ldots, f_{d,n})}{\Gamma(f_1, \ldots, f_d)} - \left(1 - \frac{d}{n}\right) S_\varepsilon^2 \tag{2.21}$$

where the term d/nS_ε^2 corresponds to the second-order approximation of the expectation of M_n^2 in (2.11). Recall that $\Gamma_n(\cdot)$ is defined as the determinant which is obtained by replacing the inner products $\langle g, g\rangle$, $\langle g, f_l\rangle$ $(l = 1, \ldots, d)$ in the Gramian determinant $\Gamma(g, f_1, \ldots, f_d)$ by the inner products of the vectors Y, $f_{1,n}, \ldots, f_{d,n} \in \mathbb{R}^n$ according to (2.6) and (2.7), respectively.

REMARK 2.1. Note that Gasser, Sroka and Jennen-Steinmetz (1986) require a Hölder constant $\gamma > 1/4$ for the regression function g in order to establish the asymptotic normality of the variance estimator S_ε^2. The stronger condition $\gamma > 1/2$ will become necessary in the following section where (2.10) is derived.

REMARK 2.2. Various authors [see Stute (1997) or Müller (1992), among others] based their test criterion for problem (1.2) on the asymptotic normality of the LSE $\hat{\theta}$ under H_0 [Jennrich (1969)]. In contrast, our approach relates directly to the L^2-distance between the parametric submodel U_d and the unknown regression function g. In Section 3 we demonstrate that this allows the determination of the limit distribution of $\sqrt{n}(\hat{M}_n^2 - M^2)$ for arbitrary $g \in \text{Höl}_\gamma[0, 1], \gamma > 1/2$. This will be applied to obtain tests for the precise hypotheses H_Δ in (1.3) and confidence intervals for the discrepancy M between the unknown regression function g and the subspace U_d spanned by the functions $f_1, \ldots, f_d$.

3. The asymptotic distribution of $\hat{M}_n^2$. For the asymptotic theory we may assume without loss of generality that $\langle f_i, f_j\rangle = \delta_{ij}$, $i, j = 1, \ldots, d$; that is, $\{f_i\}_{i=1,\ldots,d}$ is an orthonormal system, because the particular choice of the basis of U_d remains M^2 invariant. Hence, by (2.4) and (2.21), we have

$$\begin{aligned} \hat{M}_n^2 &= \sum_{i=1}^n \Delta_i w_i h(t_i) Y_i^2 - \left(1 - \frac{d}{n}\right) S_\varepsilon^2 - \sum_{l=1}^d \left| \sum_{i=1}^n \Delta_i w_i h(t_i) f_l(t_i) Y_i \right|^2 \\ &=: Y^t T Y - \left(1 - \frac{d}{n}\right) S_\varepsilon^2, \end{aligned} \tag{3.1}$$

where T denotes the $n \times n$ matrix

$$T = \mathbf{W} - \sum_{l=1}^d (\mathbf{W} f_{l,n})(\mathbf{W} f_{l,n})^t \,, \tag{3.2}$$

$f_{i,n} = (f_i(t_1), \ldots, f_i(t_n))^t$ and $\mathbf{W}$ is defined as quadratic diagonal matrix with diagonal elements $\Delta_i w_i h(t_i)$ $(i = 1, \ldots, n)$. In order to guarantee weak

convergence of $n^{1/2}(\hat{M}_n^2 - M^2)$ to a non-degenerate limit, we finally require the following condition regarding the moments of the error distribution in (1.1)

$$\sup_{t\in[0,1]} E[|\varepsilon(t)|^4] < \infty. \tag{A4}$$

THEOREM 3.1. *Consider the homoscedastic regression model* (1.1), *the linear model* $U_d = \operatorname{span}\{f_1, \ldots, f_d\}$ *and assume that* (A1)–(A4) *are satisfied, then*

$$\mathcal{L}(n^{1/2}(\hat{M}_n^2 - M^2)) \to \mathcal{N}(0, 4\sigma^2 M^2 + 17/9\sigma^4) \quad \text{as } n \to \infty, \tag{3.3}$$

where $\mathcal{N}(\mu, \tau^2)$ *denotes a normal distribution with mean* μ *and variance* τ^2.

PROOF. Let $g_n = (g(t_1), \ldots, g(t_n))^t$, $\varepsilon = Y - g_n$ denote the vector of residuals, I_n the $n \times n$ identity matrix and define

$$R = \frac{1}{n} I_n - \frac{1}{n^2} \sum_{l=1}^{d} f_{l,n} f_{l,n}^t \,. \tag{3.4}$$

With this notation we obtain from (2.19) and (3.1),

$$\begin{aligned}
\hat{M}_n^2 - M^2 &= Y^t T Y - Y^t D_n Y - M^2 + o_p(n^{-1/2}) \\
&= Y^t T Y - Y^t B_n Y - M^2 + o_p(n^{-1/2}) \\
&= \varepsilon^t T \varepsilon + 2\varepsilon^t T g_n + g_n^t T g_n - M^2 - \varepsilon^t B_n \varepsilon - 2\varepsilon^t B_n g_n \\
&\quad - g_n^t B_n g_n + o_p(n^{-1/2}) \\
&= \varepsilon^t R \varepsilon + 2\varepsilon^t R g_n - \varepsilon^t B_n \varepsilon - 2\varepsilon^t B_n g_n + o_p(n^{-1/2}) \\
&= \frac{1}{n}\varepsilon^t (I_n - B)\varepsilon + 2\varepsilon^t \left(R - \frac{1}{n} B \right) g_n + o_p(n^{-1/2}) \\
&=: U_n + o_p(n^{-1/2}),
\end{aligned} \tag{3.5}$$

where we used Lemma A.1(iv) in the Appendix for the second equality, parts (iii), (i), (ii) for the equality between the second and the third line and Lemma A.1(v) for the last equality. Note that the proof of Lemma A.1 requires Hölder continuity of the functions $g, f_1, \ldots, f_d$ of order $\gamma > 1/2$. Observing (2.20) and (3.4), this gives

$$\begin{aligned}
U_n &= \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 - \frac{1}{6n}\sum_{i=2}^{n-1} (\varepsilon_{i+1} + \varepsilon_{i-1} - 2\varepsilon_i)^2 - \frac{2}{n}\sum_{i=1}^{n} \varepsilon_i \tau_{i,n} + \frac{2}{n}\sum_{i=1}^{n} \varepsilon_i g(t_i) \\
&\quad - \frac{1}{3n}\sum_{i=2}^{n-1} (\varepsilon_{i+1} + \varepsilon_{i-1} - 2\varepsilon_i)(g(t_{i+1}) + g(t_{i-1}) - 2g(t_i)) \\
&= \frac{1}{n}\sum_{i=1}^{n-2} L_i + o_p(n^{-1/2}),
\end{aligned} \tag{3.6}$$

where the random variables L_i are defined by

$$L_i = 4/3\varepsilon_i\varepsilon_{i+1} - 1/3\varepsilon_i\varepsilon_{i+2} - 2(\tau_{i,n} - g(t_i))\varepsilon_i, \qquad i = 1, \ldots, n-2, \tag{3.7}$$

and the constants $\tau_{i,n}$ are given by

$$\tau_{i,n} = \frac{1}{n}\sum_{l=1}^{d}\sum_{j=1}^{n} f_l(t_i)f_l(t_j)g(t_j), \qquad i = 1, \ldots, n. \tag{3.8}$$

Observe that the L_i defined in (3.7) are uncorrelated random variables each with expectation 0 and variance

$$V[L_i] = \tfrac{17}{9}\sigma^4 + 4(\tau_{i,n} - g(t_i))^2\sigma^2. \tag{3.9}$$

From (A1), (A3) and (3.8) it follows [see also (A.2) in the Appendix]

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\tau_{i,n}g(t_i) = \lim_{n\to\infty}\sum_{l=1}^{d}\left(\frac{1}{n}\sum_{i=1}^{n}f_l(t_i)g(t_i)\right)^2 = \sum_{l=1}^{d}|\langle f_l, g\rangle|^2$$

and similarly

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\tau_{i,n}^2 = \sum_{l=1}^{d}\sum_{m=1}^{d}\langle f_l, f_m\rangle\langle f_l, g\rangle\langle f_m, g\rangle = \sum_{l=1}^{d}|\langle f_l, g\rangle|^2,$$

where we used the orthogonality of the functions $f_1, \ldots, f_d$ in the last line. Hence,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\tau_{i,n} - g(t_i))^2 = \langle g, g\rangle - \sum_{l=1}^{d}|\langle g, f_l\rangle|^2 = M^2$$

and we obtain

$$\lim_{n\to\infty} V\left(n^{-1/2}\sum_{i=2}^{n}L_i\right) = 4\sigma^2M^2 + 17/9\sigma^4.$$

Convergence of the distribution of $n^{-1/2}\sum_{i=1}^{n}L_i$ to the normal law $\mathcal{N}(0, 4\sigma^2M^2 + 17/9\sigma^4)$ now follows from Orey (1958), Theorem 1, because the L_i form a centered rowwise, 2-dependent array. The assertion of Theorem 3.1 is finally obtained from (3.5) and (3.6). □

4. Applications.

4.1. *Validation of linear models.* The most appealing property of the suggested test statistic $\hat{M}_n^2$ certainly consists in the following aspects. The test statistic can easily be computed and the calculation of critical regions requires only tables of the standard normal distribution. This provides a very simple test for the classical hypotheses (1.2). That is, H_0: $g \in U_d$ is rejected if

$$\left(\frac{9n}{17}\right)^{1/2}\frac{\hat{M}_n^2}{S_\varepsilon^2} > u_{1-\alpha}, \tag{4.1}$$

where u_β denotes the β-quantile of the standard normal distribution. Moreover, Theorem 3.1 gives the asymptotic limit distribution of $\sqrt{n}(\hat{M}_n^2 - M^2)$

for arbitrary regression functions $g \in L^2$. On the one hand, this result can be used for constructing asymptotic confidence intervals for the L^2-distance M between the regression function g and the subspace U_d. For example, an (asymptotic) $(1-\alpha)$ upper confidence bound for M is given by M_n^+, where

$$(M_n^+)^2 = \hat{M}_n^2 + \frac{2u_{1-\alpha}^2}{n} S_\varepsilon^2 + \frac{u_{1-\alpha} S_\varepsilon}{\sqrt{n}} \sqrt{4\hat{M}_n^2 + S_\varepsilon^2 \left(\frac{4u_{1-\alpha}^2}{n} + \frac{17}{9} \right)}. \tag{4.2}$$

On the other hand, Theorem 3.1 provides a method for the statistical assessment of the model U_d instead of a simple model check. The precise hypothesis H_Δ in (1.3) is rejected whenever

$$n^{1/2} \frac{\hat{M}_n^2 - \Delta^2}{\{4S_\varepsilon^2 \Delta^2 + 17/9 S_\varepsilon^4\}^{1/2}} \leq u_\alpha. \tag{4.3}$$

Note that it is not even necessary to determine the weights w_i according to (2.12). The asymptotic result in Theorem 3.1 remains still valid as long as (A1)–(A4) hold. However, the use of the "exact" integrating weights is highly recommended. Simulation results show a substantial improvement of the accuracy in the approximation of the normal distribution if the weights satisfying (2.12) are chosen to be as equal as possible, especially for small sample sizes (see Section 5).

4.2. *Similarity of regression functions.* The assessment of the equality of two regression curves, say f and g is similar to the validation of goodness of fit of a parametric model U_d. Tests for the hypothesis $\overline{H}_0$: $f = g$ have been suggested by Hall and Hart (1990), Härdle and Marron (1990) and King, Hart and Wehrly (1991). The power properties of the proposed tests depend heavily on the particular choice of smoothing parameters. If automatic smoothing procedures are applied, the computational effort increases significantly. For these reasons Delgado (1993) proposed a Kolmogorov–Smirnov type statistic which does not depend on kernel estimators. Assume the unknown regression curves to be $f, g \in L^2$ where the corresponding measure δ of discrepancy between f and g is the weighted L^2-distance

$$\delta^2 = \delta^2(f, g) = \int_0^1 (f(t) - g(t))^2 h(t)\, dt. \tag{4.4}$$

Because δ^2 may be regarded as the minimal distance of $f - g$ to the subspace $U_0 = \{0\}$, the assessment of the similarity of two regression curves is tantamount to the validation of the linear regression model U_0. Following the work of Hall and Hart (1990), we consider two arrays of random variables

$$\begin{aligned} X_{i,n} &= X(t_{i,n}) = f(t_{i,n}) + \varepsilon_{i,n} \\ Y_{i,n} &= Y(t_{i,n}) = g(t_{i,n}) + \eta_{i,n}, \qquad 1 \leq i \leq n, \end{aligned}$$

where the design $\{t_{i,n}\}_{i=1}^n$ is as in (1.1) satisfying assumption (A1) with a positive design density $h \in \text{Höl}_\gamma[0,1]$, $\gamma > 1/2$. For $t \in [0,1]$, the random variables $X(t)$ and $Y(t)$ have finite expectation $EX(t) = f(t)$ and $EY(t) =$

$g(t)$, respectively. The errors $(\varepsilon_i, \eta_i)^t$ form a two-dimensional triangular array of rowwise independent random variables with expectation 0 and unknown variances, respectively. Observe that we do not only allow the variances but also the entire distributions of the errors ε and η to be different. Define new random variables $Z_i = X_i - Y_i$, $i = 1, \ldots, n$ and estimate $\delta^2 = \delta^2(f, g)$ by the quadratic form

$$\hat{\delta}_n^2 = Z^t \mathbf{W} Z - Z^t D_n Z, \tag{4.5}$$

where D_n is defined by (2.19). In order to apply Theorem 3.1, denote $\sigma_z^2 = V[Z_i]$ and observe that $\delta^2(f, g) = M^2(f - g)$ where the distance M^2 is calculated with respect to the subspace $U_0 = \{0\}$. Hence, under the assumptions (A1)–(A4) (transferred to the present situation), it follows that

$$\mathcal{L}(n^{1/2}(\hat{\delta}_n^2 - \delta^2)) \to \mathcal{N}(0, 4\sigma_z^2\delta^2 + 17/9\sigma_z^4) \quad \text{as } n \to \infty. \tag{4.6}$$

Thus as an immediate consequence of this result we obtain a test for the classical hypothesis $\overline{H}_0$: $f = g$ and for the precise hypothesis

$$\overline{H}_{\Delta_0}\colon \delta^2 > \Delta_0 \quad \text{versus} \quad \overline{K}_{\Delta_0}\colon \delta^2 \le \Delta_0. \tag{4.7}$$

Hence, our approach allows one to show that f and g are similar with respect to δ^2 at a controlled type I error rate. Confidence intervals for the discrepancy δ^2 between f and g can be obtained in the same way as in (4.2). Observe that the relation (4.6) and the resulting applications do not require the independence of the errors ε_i and η_i.

5. Numerical results.

5.1. *Accuracy of the asymptotic law.* In order to get more insight into the finite sample properties of the asymptotic approximation in Theorem 3.1, we performed an extensive simulation study of the statistic

$$H_n = n^{1/2} \frac{\hat{M}_n^2 - M^2}{\left(\frac{17}{9} S_\varepsilon^4 + 4 S_\varepsilon^2 M^2\right)^{1/2}} \tag{5.1}$$

for various sample sizes n. We considered three different models

$$\begin{aligned} g_1(x) &= a_1 + a_2 x + \exp(\kappa x) && \text{model I,} \\ g_2(x) &= a_1 + a_2 x + (\kappa x)^2 \sin(\kappa x) && \text{model II,} \\ g_3(x) &= a_1 + a_2 x + \kappa x^2 && \text{model III,} \end{aligned}$$

where κ is a fixed parameter which determines the L^2-distance to the linear regression model $U_2 = \operatorname{span}\{1, x\}$. As a design we used

$$t_{i,n} = \frac{i-1}{n-1}, \qquad i = 1, \ldots, n, \tag{5.2}$$

which corresponds to the design density $h \equiv 1$ in (A1) and (2.1). The weights for the numerical quadrature were chosen in accordance with (2.12) and (A2) to be

$$w_{i,n} = \begin{cases} \dfrac{(n-1)^2}{n(n-2)}, & \text{if } i = 2, \dots n-1, \\ \dfrac{n-1}{2n}, & \text{if } i = 1, n, \end{cases} \tag{5.3}$$

which allows an exact integration of the functions $1, x, x^2$.

In Table 1 we display for $n = 20$ the accuracy of the approximation of the α-quantiles of the standard normal distribution. The numbers of outcomes which were smaller than the corresponding quantile of the normal distribution

TABLE 1
Simulated deviation from the normal distribution for the sample size $n = 20$

	M^2		0				0.5			
	Quantile		**0.05**	**0.1**	**0.9**	**0.95**	**0.05**	**0.10**	**0.9**	**0.95**
f_1	σ^2	0.09	0.042	0.085	0.915	0.962	0.050	0.091	0.879	0.936
		0.25	0.042	0.085	0.915	0.962	0.056	0.102	0.891	0.945
		1	0.042	0.085	0.915	0.962	0.056	0.107	0.906	0.956
		4	0.042	0.085	0.915	0.962	0.047	0.093	0.915	0.959
f_2	σ^2	0.09	0.042	0.085	0.915	0.962	0.031	0.070	0.884	0.938
		0.25	0.042	0.085	0.915	0.962	0.042	0.083	0.901	0.948
		1	0.042	0.085	0.915	0.962	0.048	0.091	0.917	0.956
		4	0.042	0.085	0.915	0.962	0.041	0.081	0.919	0.963
f_3	σ^2	0.09	0.042	0.085	0.915	0.962	0.058	0.099	0.881	0.939
		0.25	0.042	0.085	0.915	0.962	0.065	0.112	0.896	0.946
		1	0.042	0.085	0.915	0.962	0.060	0.113	0.910	0.955
		4	0.042	0.085	0.915	0.962	0.050	0.093	0.916	0.959
	M^2		1				2			
	Quantile		**0.05**	**0.10**	**0.9**	**0.95**	**0.05**	**0.10**	**0.9**	**0.95**
f_1	σ^2	0.09	0.044	0.078	0.869	0.927	0.033	0.067	0.853	0.918
		0.25	0.051	0.096	0.883	0.939	0.045	0.082	0.871	0.930
		1	0.058	0.107	0.900	0.951	0.055	0.097	0.891	0.945
		4	0.050	0.098	0.912	0.958	0.055	0.104	0.905	0.955
f_2	σ^2	0.09	0.028	0.059	0.867	0.930	0.023	0.052	0.857	0.922
		0.25	0.040	0.080	0.892	0.942	0.037	0.078	0.880	0.935
		1	0.052	0.092	0.907	0.952	0.051	0.098	0.902	0.946
		4	0.045	0.089	0.920	0.960	0.054	0.097	0.917	0.955
f_3	σ^2	0.09	0.053	0.091	0.873	0.933	0.043	0.082	0.863	0.926
		0.25	0.060	0.104	0.887	0.941	0.056	0.095	0.877	0.936
		1	0.063	0.116	0.902	0.951	0.065	0.112	0.896	0.946
		4	0.054	0.102	0.913	0.959	0.060	0.113	0.910	0.955

were counted and divided by the total number of simulations. For the sample size and α-quantiles we chose $n = 20, 50, 100$ (only $n = 20$ is displayed in Table 1) and $\alpha = 0.05, 0.1, 0.9, 0.95$, respectively [note that the 5% and 10% quantiles are needed in (4.3) while the 90% and 95% quantiles are used in (4.1)]. For each study we have performed 5000 replications with an SAS-IML random generator. The error, ε, was assumed to be normally distributed with variance σ^2 where $\sigma^2 = 0.09, 0.25, 1, 4$. For the distance M, we chose the values $M^2 = 0, 0.5, 1, 2$. From this simulation study, the following conclusions can be drawn. First of all, we found that the quality of the approximation depends only slightly on this distance. Observe further, that, in general, the estimated probabilities increase whenever the variances increase. When $M = 0$ (or equivalently $\kappa = 0$) we find that the variance σ^2 does not affect the distribution of the test statistic H_n which is in accordance with the observation that H_n in (5.1) is invariant with respect to the group of scale transformations acting on $\mathbb{R}^n$, whenever $M = 0$. From Table 1, we draw further that, even for relatively small sample sizes, the approximation of the level is surprisingly accurate, independent of M^2 and the unknown variance σ^2.

For example, when $n = 20$ we find from Table 1 that the maximal observed deviation between the estimated and nominal probabilities is always less than 60% of the nominal probability q. Further simulations (not displayed for the sake of brevity) show that increasing sample size increases the accuracy of the approximation rapidly. For example, when $n = 50$, the maximal deviation between nominal and actual probabilites was always found to be less than 15% of q and when $n = 100$, less than 5%. We found similar results for different distributional assumptions, even in the case of a nonsymmetric error distribution. For example, if $\varepsilon \sim \sigma^2(\chi_1^2 - 1)/\sqrt{2}$, $M^2 = 0.5$, $\sigma^2 = 0.25$, $n = 50$, we obtained in model III 0.044, 0.097, 0.896 and 0.944 as approximations for the probabilities 0.05, 0.10, 0.90 and 0.95. Other results are omitted for the sake of brevity and are available from the authors on request.

5.2. *A power study.* As pointed out by the referees and the Associate Editor, a comparison of our approach with other procedures for a model check is of some interest. The work most similar in spirit is the work of Eubank and Hart (1992). In order to understand the features of this test completely, a short description of Eubank and Hart's procedure is given here. In addition to the regression functions $f_1, \ldots, f_d$, define $n - d$ functions $u_{jn}(\cdot)$, $j = 1, \ldots n - d$, such that $f_1, \ldots f_d$, $u_{1n}, \ldots, u_{n-d,n}$ are linearly independent. Consider the sample Fourier coefficients

$$\hat{\alpha}_{jn} = \frac{1}{n} \sum_{r=1}^{n} u_{jn}(t_r) Y_r, \qquad j = 1, \ldots, n - d$$

and determine $\hat{k} \in \{0, \ldots, n - d\}$ such that

$$\sum_{r=1}^{n} \left\{ Y_r - \sum_{j=1}^{d} b_{jn} f_j(t_r) - \sum_{j=1}^{k} \hat{\alpha}_{jn} u_{jn}(t_r) \right\}^2 + c_\alpha S_\varepsilon^2 (d + k)$$

is a minimum. Here $b_n = (b_{1n}, \ldots, b_{dn})^t$ is the least squares estimator in the model U_d and c_α denotes a critical value which can be obtained from Table 1 in Eubank and Hart (1992). If the functions u_{jn} are orthonormal with respect to the uniform measure at the design points, this optimization reduces to the minimization of

$$r(k) = \sum_{j=1}^{k} \hat{\alpha}_{jn}^2 - \frac{S_\varepsilon^2 c_\alpha k}{n}, \qquad k = 0, \ldots, n-d,$$

$[r(0) = 0]$ and the hypothesis H_0: $f \in U_d$ is rejected if $\hat{k} \geq 1$. Eubank and Hart (1992) showed that under mild assumptions this test can detect local alternatives converging to the null at the rate $n^{-1/2}$.

First of all, we remark, that in a concrete application Eubank & Hart's test depends sensitively on the choice of the functions $u_{1n}, \ldots, u_{n-d,n}$. Moreover, even if these functions are fixed, the test still depends on the ordering of the functions. This is illustrated in a small example, where we also compare the Fourier analytic approach with the test presented in this paper. Following the work of Eubank and Hart (1992) (Example 1), we consider the problem of testing for no effect of the regression variable, that is, H_0: $g(\cdot) \equiv \beta$, versus K_0: $g(\cdot) \not\equiv \beta$ which corresponds to $d = 1$, $f_1(\cdot) \equiv 1$. In the sequel we consider the alternatives $g(x) = 1 + c\cos(10\pi t)$ for various values of c. The design points are chosen equidistant $t_r = (r - 1/2)/n$, $(r = 1, \ldots, n)$ and the set of additional regression functions is $\{\sqrt{2}\cos(\pi jt); j = 1, \ldots, n-1\}$. In the following Monte Carlo study, we compare the effect on size and power of the test of Eubank and Hart (1992) for three different orderings of these functions:

$$u_{jn}(t) = \sqrt{2}\cos(\pi jt), \qquad j = 1, \ldots, n-1, \tag{5.4}$$

$$u_{jn}(t) = \sqrt{2}\cos(\pi(n-j)t), \qquad j = 1, \ldots, n-1, \tag{5.5}$$

$$u_{jn}(t) = \begin{cases} \sqrt{2}\cos(10\pi t), & \text{if } j = 1, \\ \sqrt{2}\cos(\pi(j-1)t), & \text{if } j = 2, \ldots, 9, \\ \sqrt{2}\cos(\pi jt), & \text{if } j = 11, \ldots, n-1. \end{cases} \tag{5.6}$$

For a sample size of $n = 100$ and standard normal distributed errors, we calculated the power of the test proposed in this paper by (4.1) and the three tests of Eubank and Hart which emerge from the different orderings of the functions in (5.4), (5.5) and (5.6). Then 5000 simulations were performed and the number of rejections of H_0: $g(\cdot) \equiv \beta$ counted. The results are listed in Table 2. We find that the power of the test proposed in this paper is comparable with the power of Eubank and Hart's test if the natural ordering (5.4) is chosen. If the inverse ordering (5.5) is used the last named test has a nearly constant power function $\rho \equiv \alpha$ and cannot be recommended. On the other hand, the ordering (5.6) produces an extremely powerful version of Eubank and Hart's test in order to detect the specific alternative $g(x) = 1 + c\cos(10\pi x)$.

TABLE 2
Power of the test (4.1) *and the test of Eubank and Hart* (1992) *for different orderings of the function* $u_{jn}(t) = \sqrt{n}\cos(\pi jt)^*$

			Test of Eubank and Hart		
α	**c**	**Test (4.1)**	**(5.4)**	**(5.5)**	**(5.6)**
5%	0.00	0.0440	0.0564	0.0424	0.0620
	0.25	0.0754	0.0586	0.0394	0.4114
	0.50	0.1912	0.0994	0.0418	0.9268
	0.75	0.4868	0.3994	0.0374	0.9988
	1.00	0.8164	0.8612	0.0432	1.0000
10%	0.00	0.0966	0.1124	0.0834	0.1150
	0.25	0.1400	0.1256	0.0876	0.5170
	0.50	0.3074	0.2416	0.0876	0.9592
	0.75	0.6270	0.6744	0.0866	0.9996
	1.00	0.8964	0.9688	0.0924	1.000

*The alternative considered here is $g(t) = 1 + c\cos(10\pi t)$ and the sample size is $n = 100$.

This is rather obvious because we have first chosen the regression function which is most similar to the alternative $\cos(10\pi x)$.

We observed similar effects in other set-ups. Summarizing these results, we remark that in many cases the test (4.1) proposed in this paper has power behavior similar to Eubank and Hart's test. However, the last named test is extremely sensitive with respect to the choice and ordering of the additional functions $u_{jn}(\cdot)$ $(j = 1, \ldots, n - d)$. This property yields to a substantial improvement of the power, if the experimenter wants to guard against specific alternatives. Note that the set of these alternatives depends on the sample size. However, Eubank and Hart's procedure loses some of its appeal in many practical applications from the dependence of the test decision on the specific ordering of the set of these specific regression functions. The simulations show that a misspecification of these alternatives leads to a serious loss in power. Therefore, if no information about specific alternatives is available, we strongly recommend the use of the omnibus test (4.1).

6. Remarks and conclusions.

REMARK 6.1. If the random errors $\varepsilon_{i,n}$ in (1.1) are i.i.d. and $E(\varepsilon_{i,n}^3) = 0$, then, under conditions (A1)–(A4), it can be proved that

$$\lim_{n\to\infty} V[n^{1/2}Y^tTY] = (m_4 - 1)\sigma^4 + 4\sigma^2M^2,$$

where $m_4 = E[\varepsilon_{i,n}^4]\sigma^{-4}$. Surprisingly, this shows that the estimation of σ^2 by S_ε^2 reduces the (asymptotic) variability of the statistic

$$n^{1/2}(\hat{M}_n^2 - M^2) = n^{1/2}(Y^tTY - (1 - d/n)S_\varepsilon^2 - M^2)$$

compared with that of

$$n^{1/2}(M_n^2 - M^2) = n^{1/2}(Y^t TY - \sigma^2 - M^2)$$

whenever $m_4 \geq 26/9$. In the case of a normally distributed error, we observe a maximum reduction by a relative amount of approximately 6%, which is attained for $M = 0$.

REMARK 6.2. For testing H_0 versus K_0, one would reject the null hypothesis whenever (4.1) is valid. Then, asymptotically, the power function π_φ of the test φ with critical region (4.1) is

$$\begin{aligned}\pi_\varphi(n, \kappa) = 1 - \Phi(u_{1-\alpha}(1 + 36/17\kappa)^{-1/2} \\ - n^{1/2}(17/9\kappa^{-2} + 4\kappa^{-1})^{-1/2}) + o(1),\end{aligned} \tag{6.1}$$

where $\kappa = M^2/\sigma^2$. This approximation can be used for computing the required sample size in order to control the type II error of the test (4.1) of the classical hypothesis H_0: $g \in U_d$. Such calculations require only information about σ^2 and $M^2 = M^2(g)$ but not about the particular alternative $g \in L^2$. Similar computations can be performed for the test (4.3) of the "precise" hypothesis H_Δ.

REMARK 6.3. It is worthwhile to mention that the definition of the test statistic in (2.21) requires the explicit knowledge of the design density h in (2.1). If such knowledge is not available, the density has to be estimated from the data. To this end the estimates of the inner products in (2.6) and (2.7) have to be defined differently, that is,

$$\langle Y, f_{l,n}\rangle_n := \frac{1}{n}\sum_{i=1}^n w_{i,n} f_l(t_{i,n}) Y_i, \tag{6.2}$$

$$\langle Y, Y\rangle_n := \frac{1}{n}\sum_{i=1}^n w_{i,n} Y_i^2 \ . \tag{6.3}$$

Similary the inner products $\langle f_l, f_k\rangle = \int_0^1 f_l(t) f_k(t) h(t)\, dt$ in $\Gamma(f_1, \ldots, f_d)$ in (2.21) have to be estimated by

$$\langle f_{k,n}, f_{l,n}\rangle_n := \frac{1}{n}\sum_{i=1}^n w_{i,n} f_k(t_{i,n}) f_l(t_{i,n}). \tag{6.4}$$

It can be shown that all results of this paper remain true if the estimates (2.6) and (2.7) are replaced by (6.2) and (6.3) and the inner products of the the regression functions are replaced by their empirical counterparts in (6.4). To this end we require a $o(n^{-3/2})$ instead of $o(n^{-1/n})$ in the asymptotic Sacks-Ylvisaker condition (A1). Applying similar arguments, our approach can also be extended to the case of random predictors. For a random design, an analogue of Theorem 3.1 is valid with a different limiting variance which reflects the randomness of the predictor variable. For details, the reader is referred to the work of Dette and Munk (1997).

APPENDIX

LEMMA A.1. *Consider the homoscedastic regression model* (1.1) *and the matrices* R, T, B_n, D_n *defined by* (3.4), (3.2), (2.20) *and* (2.19) *in Sections* 2 *and* 3. *Let* $g_n = (g(t_1), \dots, g(t_n))^t$. *Under assumptions* (A1)–(A4) *we have the following*:

(i) $g_n^t T g_n - M^2 = o(n^{-1/2})$;
(ii) $g_n^t B_n g_n = o(n^{-1/2})$;
(iii) $\varepsilon^t(R-T)\varepsilon = o_p(n^{-1/2})$, $\varepsilon^t(R-T)g_n = o_p(n^{-1/2})$;
(iv) $Y^t(B_n - D_n)Y = o_p(n^{-1/2})$;
(v) $\varepsilon^t(R - n^{-1}I_n)\varepsilon = o_p(n^{-1/2})$.

PROOF. For (i), (ii), consider an arbitrary $v \in \text{Höl}_\gamma[0,1]$, where $\gamma > 1/2$; then assumption (A2) yields

$$\begin{aligned}
&\left|\sum_{i=1}^n \Delta_i w_i h(t_i) v(t_i) - \int_0^1 v(x)h(x)\,dx\right| \\
&\qquad \le \sum_{i=2}^n \int_{t_{i-1}}^{t_i} |v(t_i)h(t_i) - v(x)h(x)|\,dx + o(n^{-1/2}) \qquad \text{(A.1)} \\
&\qquad \le nC \max_{i=2}^n \int_{t_{i-1}}^{t_i} |x - t_i|^\gamma\,dx + o(n^{-1/2}) = o(n^{-1/2}),
\end{aligned}$$

where we have used $h, v \in \text{Höl}_\gamma[0,1]$, $\gamma > 1/2$, in the second inequality. Applying this result to g^2, $gf_1, \dots gf_d$ proves part (i). Part (ii) of Lemma A.1 follows similarly.

(iii) By assumption (A2) we have

$$\begin{aligned}
V[n^{1/2}\varepsilon^t(n^{-1}I_n - \mathbf{W})\varepsilon] &= n\sum_{i=1}^n (\Delta_i w_i h(t_i) - n^{-1})^2 V[\varepsilon_i^2] \\
&\le C_h\, n^2 \left(\max_{i=1,\dots,n} |\Delta_i h(t_i) - n^{-1}|\right)^2 + o(1)\,.
\end{aligned}$$

A straightforward calculation shows

$$\left|\Delta_i h(t_i) - \frac{1}{n}\right| \le \left|\frac{h(t_i)}{h(\xi_i)}\right|\left|\int_{t_{i-1}}^{t_i} h(x)\,dx - \frac{1}{n}\right| + \frac{1}{n}\left|\frac{h(t_i)}{h(\xi_i)} - 1\right|,$$

where $\xi_i \in (t_{i-1}, t_i)$ $(i = 1, \dots, n)$. Assumptions (A1) and (A3) now imply (note that the design density is positive on $[0,1]$)

$$\max_{i=2}^n \left|\Delta_i h(t_i) - \frac{1}{n}\right| = o\left(\frac{1}{n}\right), \qquad \text{(A.2)}$$

which implies $V(\varepsilon^t(n^{-1}I_n - \mathbf{W})\varepsilon) = o(n^{-1})$. Observing (A.1) we have

$$E[\varepsilon^t(n^{-1}I_n - \mathbf{W})\varepsilon] = \sigma^2\left(\sum_{i=1}^n \Delta_i w_i h(t_i) - 1\right) = o(n^{-1/2})$$

and it follows

$$\varepsilon^t(n^{-1}I_n - \mathbf{W})\varepsilon = o_p(n^{-1/2}).$$

In the same way we obtain for the other terms

$$\varepsilon^t(\mathbf{W}f_{l,n}f_{l,n}^t\mathbf{W}^t - n^{-2}f_{l,n}f_{l,n}^t)\varepsilon = o_p(n^{-1/2}), \qquad l = 1, \ldots, d,$$

which proves the first part of Lemma A.1(iii). The second part is shown by similar arguments.

(iv) From Whittle (1964) we draw that

$$V[Y^t(B_n - D_n)Y] = \sigma^4(n-2)^{-2}\{2\,\mathrm{tr}[(B-D)^2] + (m_4 - 3)\sum_{i=1}^{n}(B-D)_{ii}^2\} + o\left(\frac{1}{n}\right),$$

where $m_4 = E[\varepsilon_i^4]\sigma^{-4}$. A straightforward calculation shows

$$\begin{aligned}\mathrm{tr}[(B-D)^2] = \sum_{i=3}^{n-3}&(c_{i-1}^2 a_{i-1}b_{i-1} - 1/6)^2 \\ &+ (b_{i-1}c_{i-1}^2 + a_i c_i^2 - 2/3)^2 + (c_{i-1}^2 b_{i-1}^2 + c_i^2 + a_{i+1}^2 c_{i+1}^2 - 1)^2 \\ &+ (b_i c_i^2 + a_{i+1}c_{i+1}^2 - 2/3)^2 + (c_{i+1}^2 a_{i+1}b_{i+1} - 1/6)^2 + O(1)\end{aligned}$$

and

$$\sum_{i=1}^{n}(B-D)_{ii}^2 = \sum_{i=3}^{n-3}(c_{i-1}^2 b_{i-1}^2 + c_i^2 + a_{i+1}^2 c_{i+1}^2 - 1)^2 + O(1).$$

Condition (A1) implies

$$\max_{i=1,\ldots,n-1}|\Delta_i/\Delta_{i+1} - 1| = o(1),$$

which shows that

$$\max_{i=2,\ldots,n}|b_i - 1/2| = o(1), \qquad \max_{i=2,\ldots,n}|a_i - 1/2| = o(1)$$

and

$$\max_{i=2,\ldots,n}|c_i^2 - 2/3| = o(1).$$

Combining these arguments we obtain $V(Y^t(B_n - D_n)Y) = o(n^{-1})$. By similar arguments it can be shown that

$$E[Y^t(B_n - D_n)Y] \;=\; g_n^t(B_n - D_n)g_n \;+\; \sigma^2\,\mathrm{tr}(B_n - D_n) \;=\; o(n^{-1/2}),$$

which proves part (iv).

(v) Let $S = R - n^{-1}I_n = \sum_{l=1}^{d} f_{l,n} f_{l,n}^t$ and $\kappa_4 = m_4 - 3m_2^2$ where $m_k = E[\varepsilon_j^k]/\sigma^k$, $k = 1, \ldots, 4$ Then we draw from Whittle (1964) for the variance of $\varepsilon^t S \varepsilon$,

$$V[\varepsilon^t S \varepsilon] = \kappa_4 \sigma^4 \sum_{j=1}^{n} s_{jj}^2 + 2\sigma^4 \sum_{i,j=1}^{n} s_{ij}^2$$

$$= \kappa_4 \sigma^4 \sum_{i=1}^{n} \left(\sum_{l=1}^{d} f_l^2(t_i) \right)^2 + 2\sigma^4 \sum_{i,j=1}^{n} \left(\sum_{l,k=1}^{d} f_l(t_i) f_k(t_j) \right)^2,$$

which leads to $V[\varepsilon^t n^{-2} S \varepsilon] = o(n^{-1})$. The corresponding statement for the expectation is obvious and assertion (v) of Lemma A.1 follows. □

Acknowledgments. The authors acknowledge the hospitality of the Institute of Mathematical Stochastics, Göttingen University and of the Wharton School of Finance, University of Pennsylvania, where parts of this paper were written. The authors are also grateful to Professor W. Stute and Dr. S. Zwanzig for help with the references and M. Sahm and T. Wagner for helpful discussions. Various comments and suggestions of the Editor, the Associate Editor and two referees, which led to a clearer presentation of the results, are acknowledged.

REFERENCES

ACHIESER, N. J. (1956). *Theory of Approximation*. Dover, New York.

BARRY, D. and HARTIGAN, J. A. (1990). An omnibus test for departure from constant mean. *Ann. Statist.* **18** 1340–1357.

BERGER, J. O. and DELAMPADY, M. (1987). Testing precise hypotheses. *Statist. Sci.* **2** 317–352.

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

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

CHOW, S. C. and LIU, J. P. (1992). *Design and Analysis of Bioavailability and Bioequivalence Studies*. Dekker, New York.

COX, D., KOH, E., WAHBA, G. and YANDELL, B. S. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.* **16** 113–119.

DAVISON, A. C. and TSAI, C. L. (1992). Regression model diagnostics. *Internat. Statist. Rev.* **60** 337–353.

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

DETTE, H. and MUNK, A. (1997). A simple goodness-of-fit test for linear models under a random design assumption. *Ann. Inst. Statist. Math.* To appear.

DETTE, H., MUNK, A. and WAGNER, T. (1998). Estimating the variance in nonparametric regression by quadratic forms—what is a reasonable choice? *J. Roy. Statist. Soc. Ser. B.* To appear.

DIEBOLDT, J. (1995). A nonparametric test for the regression function: asymptotic theory. *J. Statist. Plann. Inference* **44** 1–17.

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 SPIEGELMANN, C. H. (1990). Testing the goodness of fit of a linear model via regression techniques. *J. Amer. Statist. Assoc.* **85** 387–392.

GASSER, T., SROKA, 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. and MARRON, J. S. (1990). On variance estimation in nonparametric regression. *Biometrika* **77** 415–419.

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.

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.

MACKINNON, J. G. (1992). Model specification tests and artificial regressions. *J. Economic Lit.* **30** 102–146.

MANDALLAZ, D. and MAU, J. (1981). Comparison of different methods for decision-making in bioequivalence assessment. *Biometrics* **37** 213–222.

METZLER, C. M. (1974). Bioavailiability: a problem in equivalence. *Biometrics* **30** 309–317.

MÜLLER, H. G. (1992). Goodness of fit diagnostics for regression models. *Scand. J. Statist.* **19** 157–172.

NEIL, J. W. and JOHNSON, D. E. (1985). Testing linear regression function adequacy without replication. *Ann. Statist.* **13** 1482–1489.

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

SCHUIRMANN, D. L. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J. Pharmacocinetics and Biopharmaceutics* **15** 657–680.

SHILLINGTON, E. R. (1979). Testing lack of fit in regression without replication. *Canad. J. Statist.* **7** 137–146.

SILVERMAN, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J. Roy. Statist. Soc. Ser. B* **47** 1–52.

STAUDTE, R. G. and SHEATHER, S. J. (1990). *Robust Estimation and Testing*. Wiley, New York.

STANISWALIS, J. G. and SEVERINI, T. A. (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.

WHABA, G. (1978). Improper priors, spline smoothing, and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser. B* **40** 364–372.

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

YANAGIMOTO, T. and YANAGIMOTO, M. (1987). The use of marginal likelihood for a diagnostic test for the goodness of fit of the simple linear regression model. *Technometrics* **29** 95–107.

ZIELKE, R. (1979). *Discontinous Cebysev-Systems. Lecture Notes in Math.* **707**. Springer, Berlin.

ZWANZIG, S. (1980). The choice of approximate models in nonlinear regression. *Statistics* **11** 23–47.

FAKULTÄT UND INSTITUT FÜR MATHEMATIK
RUHR-UNIVERSITÄT BOCHUM
UNIVERSITÄTSSTR. 150, GEBÄUDE NA III
44780 BOCHUM
GERMANY
E-MAIL: holger.dette@rz.ruhr-uni-bochum.de

FAKULTÄT UND INSTITUT FÜR MATHEMATIK
RUHR-UNIVERSITÄT BOCHUM
UNIVERSITÄTSSTR. 150, GEBÄUDE NA III
44780 BOCHUM
GERMANY
E-MAIL: axel.munk@rz.ruhr-uni-bochum.de