

A CONSISTENT TEST FOR THE FUNCTIONAL FORM OF A REGRESSION BASED ON A DIFFERENCE OF VARIANCE ESTIMATORS¹

BY HOLGER DETTE

Ruhr-Universität Bochum

In this paper we study the problem of testing the functional form of a given regression model. A consistent test is proposed which is based on the difference of the least squares variance estimator in the assumed regression model and a nonparametric variance estimator. The corresponding test statistic can be shown to be asymptotically normal under the null hypothesis and under fixed alternatives with different rates of convergence corresponding to both cases. This provides a simple asymptotic test, where the asymptotic results can also be used for the calculation of the type II error of the procedure at any particular point of the alternative and for the construction of tests for precise hypotheses. Finally, the finite sample performance of the new test is investigated in a detailed simulation study, which also contains a comparison with the commonly used tests.

1. Introduction. In the present paper we consider the nonparametric regression model

$$y = y(t) = m(t) + \varepsilon,$$

where m is an (unknown) regression function, ε is a random error and t is the predictor. Parametric regression models are attractive among practitioners because they describe in a concise way the relation between the response y and the predictor t and allow extrapolation in many cases. However, misspecification of such a model may lead to serious errors in the subsequent data analysis, and in practice it is always advisable to test the goodness-of-fit of the postulated model. To be precise, consider the problem of testing for a linear regression. Assume that

$$\mathcal{M} = \{g^T(t)\theta \mid \theta \in \Theta\}$$

is a given family of functions, where $\Theta \subset \mathbb{R}^p$ is a proper parameter set and $g = (g_1, \dots, g_p)^T$ is a vector of given linear independent regression functions. The hypothesis of a linear model is

$$(1.1) \quad H_0: m \in \mathcal{M}$$

and significant effort has been devoted to the problem of testing H_0 during the last two decades [see, e.g., Shillington (1979), Neil and Johnson (1985),

Received March 1998; revised February 1999.

¹Supported in part by the Deutsche Forschungsgemeinschaft De 502/9-1.

AMS 1991 subject classifications. Primary 62G10; secondary 62G20, 62J05, 62J02.

Key words and phrases. Variance estimation, model checks, least squares estimator, limit theorems for quadratic forms.

Eubank and Hart (1992), Wooldridge (1992), Yatchew (1992), Azzalini and Bowman (1993), Härdle and Mammen (1993), Brodeau (1993), Zheng (1996), Stute, González Manteiga and Presedo Quindimil (1998), Dette and Munk (1998)]. Many authors compare a parametric and a nonparametric fit of the regression curve which requires square root consistent estimation of the parameters [see, e.g., Härdle and Mammen (1993) or Weirather (1993)]. Eubank and Hart (1992) and Kuchibhatla and Hart (1996) used a method based on order selection criteria while Stute, González Manteiga and Presedo Quindimil (1998) investigated a marked empirical process based on the residuals.

In the present paper we propose a new goodness-of-fit test for a parametric regression, which is based on a comparison of a nonparametric and a parametric estimator of the integrated variance function

$$\int \sigma^2(t)f(t) dt$$

(here f denotes the design density and $\sigma^2(\cdot)$ the variance function, that is, $\text{Var}(y(t)) = \sigma^2(t)$). Our approach to the problem of testing linearity is similar to Yatchew (1992) but in contrast to this work, our procedure does not require homoscedasticity and sample splitting in order to obtain independent variance estimators. More precisely, the test statistic, say T_n , proposed in this paper is simply the difference between the sums of squared residuals based on a parametric and a nonparametric fit. Asymptotic normality of T_n is established under the hypothesis (1.1) and under fixed alternatives $H_1: m \notin \mathcal{M}$, but the rates of convergence are different in both cases. While under the hypothesis of linearity the variance of T_n is of order $(n^2h)^{-1}$ (here h denotes the bandwidth of a kernel estimator of the regression function) it turns out that under the alternative $m \notin \mathcal{M}$ this variance is of order n^{-1} . This provides a simple asymptotic test for the hypothesis (1.1), where the asymptotic results under fixed alternatives can also be used to estimate the type II error of such a procedure at any particular point of the alternative.

The paper will be organized as follows. The test statistic and the main results can be found in Section 2, where the case of a fixed design and a linear model is considered. Section 3 discusses several extensions of our approach. Roughly speaking the results can be transferred to model checks for nonlinear regression models, random design and multivariate predictors. Surprisingly it turns out that there is a difference in the asymptotic variance under the alternative between the random and fixed design assumption. In the same section we also discuss two tests for the hypothesis (1.1) which were considered by Härdle and Mammen (1993) and Zheng (1996) and are most similar in spirit with the method proposed in this paper. The statistic of the first-named authors is based on an integrated L^2 -distance between a parametric and nonparametric fit of the regression curve, while the last-named author considers an appropriate estimator of the integrated distance between the regression function m and its best approximation by elements of the model space \mathcal{M} . In both papers asymptotic normality of the corresponding statistic is established under the null hypothesis (1.1) at a rate of order $(n\sqrt{h})^{-1}$. The present paper

extends these results and establishes additionally asymptotic normality under fixed alternatives at a rate $n^{-1/2}$. Detailed numerical comparisons of the different procedures proposed in the literature are presented in Section 4 while all proofs are deferred to Section 5.

2. Parametric versus nonparametric variance estimation. For the sake of transparency, all results presented in this paper are formulated for a univariate explanatory variable, but the proposed methods can be directly transferred to regression models with multivariate predictors [see Remark 2.7]. Consider the common fixed design regression model

$$y_{j,n} = y(t_{j,n}) = m(t_{j,n}) + \varepsilon_{j,n}, \quad j = 1, \dots, n,$$

where $t_{1,n}, \dots, t_{n,n} \in [0, 1]$ are distinct points and m is an (unknown) mean function. The errors $\varepsilon_{j,n} = \varepsilon(t_{j,n})$ are assumed to form a triangular array of row-wise independent variables with mean zero and variances $\sigma^2(t_{j,n}) = E[\varepsilon_{j,n}^2]$ ($j = 1, \dots, n$). We further assume that the fourth moments of the errors are uniformly bounded, that is,

$$(2.1) \quad E[\varepsilon_{j,n}^4] \leq C < \infty, \quad j = 1, \dots, n; \quad n \in \mathbb{N}.$$

The index n is omitted whenever this dependence will be clear from the context. Under the assumption of a linear model and a homoscedastic error structure, the standard variance estimator is the sum of squared residuals

$$\hat{\sigma}_{\text{LSE}}^2 = \frac{1}{n-p} \sum_{j=1}^n e_j^2$$

where $e_j = y_j - g^T(t_j)\hat{\theta}_n$ is the residual at the point t_j and $\hat{\theta}_n$ is the LSE of θ . The following proposition gives the asymptotic expectation of this estimator in a general regression which is not necessarily linear or homoscedastic. Throughout this paper $\text{Lip}_\gamma[0, 1]$ denotes the class of Lipschitz continuous functions of order $\gamma > 0$.

LEMMA 2.1. *Assume that the design points t_1, \dots, t_n satisfy*

$$(2.2) \quad \frac{i}{n+1} = \int_0^{t_i} f(t) dt, \quad i = 1, \dots, n$$

for a positive design density $f \in \text{Lip}_\gamma[0, 1]$ [see Sacks and Ylvisaker (1970)] and $m, g_1, \dots, g_p, \sigma^2 \in \text{Lip}_\gamma[0, 1]$, then

$$E[\hat{\sigma}_{\text{LSE}}^2] = \int_0^1 \sigma^2(t)f(t) dt + M^2 + O(n^{-\gamma}),$$

where

$$(2.3) \quad M^2 = \min_{u \in \mathcal{A}} \int_0^1 (m(t) - u(t))^2 f(t) dt$$

denotes the minimal L^2 -distance between the unknown regression function and the class \mathcal{A} of parametric models.

Note that the hypothesis of linearity is valid if and only if $M^2 = 0$ and as a consequence H_0 could be rejected for large values of $\hat{\sigma}_{LSE}^2$. However, such a procedure requires explicit knowledge of the integrated variance $\int_0^1 \sigma^2(t)f(t) dt$, which has to be estimated in practice. To this end we use a sum of squared residuals based on a nonparametric fit of the regression function. Following Hall and Marron (1990) we define the weights

$$w_{ij} = \frac{K((t_i - t_j)/h)}{\sum_{l=1}^n K((t_i - t_l)/h)}, \quad i, j = 1, \dots, n,$$

where K is a kernel function and h a bandwidth. Based on the nonparametric residuals

$$\hat{\varepsilon}_i = y_i - \sum_{j=1}^n w_{ij}y_j, \quad i = 1, \dots, n,$$

Hall and Marron (1990) proposed

$$\hat{\sigma}_{HM}^2 = \frac{1}{v} \sum_{j=1}^n \hat{\varepsilon}_j^2$$

as an estimator of the variance in a homoscedastic nonparametric regression, where

$$v = n - 2 \sum_{i=1}^n w_{ii} + \sum_{i,k=1}^n w_{ik}^2$$

is a normalizing constant, motivated by the fact that $E[\hat{\sigma}_{HM}^2] = \sigma^2$ when $m(t) \equiv 0$. It is demonstrated in Dette, Munk and Wagner (1998) that $\hat{\sigma}_{HM}^2$ has a reasonable performance in many regression problems.

We assume throughout this paper that the kernel K has compact support, is of order $r \geq 2$, that is,

$$(2.4) \quad \int_{-\infty}^{\infty} K(u)u^j du = \begin{cases} 1, & \text{if } j = 0, \\ 0, & \text{if } 1 \leq j \leq r - 1, \end{cases}$$

and define

$$\kappa_r = \frac{(-1)^r}{r!} \int_{-\infty}^{\infty} u^r K(u) du.$$

For the asymptotic inference, the bandwidth h is supposed to satisfy

$$(2.5) \quad h = O(n^{-2/(4r+1)}), \quad nh^2 \rightarrow \infty$$

if $n \rightarrow \infty$. Finally, the design density f , the regression function m , the variance function σ^2 and the basis functions g_1, \dots, g_p are assumed to be sufficiently smooth, that is,

$$(2.6) \quad m, f \in C^{(r)}([0, 1]), \quad \sigma^2, g_1, \dots, g_p \in C^{(1)}([0, 1]).$$

For a homoscedastic error structure, it was shown by Hall and Marron (1990) that under suitable modification of the estimator at the boundary we have

$$(2.7) \quad E[\hat{\sigma}_{\text{HM}}^2] = \sigma^2 + C_2 h^{2r} + o(h^{2r}),$$

where

$$(2.8) \quad C_2 = \kappa_r^2 \int_0^1 \{(mf)^{(r)}(u) - mf^{(r)}(u)\}^2 \frac{du}{f(u)}.$$

The following lemma shows that a slightly modified version of (2.7) holds in the heteroscedastic setup. The proof is similar to Hall and Marron (1990) and therefore omitted.

LEMMA 2.2. *If (2.2), (2.4), (2.5) and (2.6) are satisfied and $n \rightarrow \infty$, then*

$$(2.9) \quad E[\hat{\sigma}_{\text{HM}}^2] = \int_0^1 \sigma^2(t)f(t) dt + C_2 h^{2r} + \frac{C_3}{nh} + o(h^{2r}) + O\left(\frac{1}{n}\right),$$

where

$$(2.10) \quad C_3 = \left(2K(0) - \int_{-\infty}^{\infty} K^2(t) dt\right) \left(\int_0^1 \sigma^2(t)f(t) dt - \int_0^1 \sigma^2(t) dt\right).$$

Observing Lemmas 2.1 and 2.2 it is reasonable to base a test for a parametric regression model on the difference of the parametric and nonparametric variance estimator, and therefore we define

$$(2.11) \quad T_n = \hat{\sigma}_{\text{LSE}}^2 - \hat{\sigma}_{\text{HM}}^2,$$

which is an asymptotically unbiased estimator of M^2 . The following two results give the asymptotic distribution of T_n under the hypothesis of linearity ($M^2 = 0$) and the alternative ($M^2 > 0$). The proofs are cumbersome and therefore deferred to Section 5.

THEOREM 2.3. *If (2.1), (2.2), (2.4), (2.5) and (2.6) are satisfied, $M^2 = 0$ and $n \rightarrow \infty$, then*

$$(2.12) \quad n\sqrt{h} \left(T_n + C_2 h^{2r} + \frac{C_3}{nh}\right) \rightarrow_{\mathcal{D}} \mathcal{N}(0, \mu_0^2),$$

where $\mathcal{N}(0, \mu_0^2)$ denotes a centered normal distribution with variance given by

$$(2.13) \quad \mu_0^2 = 2 \int_{-\infty}^{\infty} (2K(u) - K * K(u))^2 du \int_0^1 \sigma^4(u) du$$

and $K_1 * K_2$ denotes the convolution of K_1 with K_2 .

THEOREM 2.4. *If (2.1), (2.2), (2.4), (2.5) and (2.6) are satisfied, $M^2 > 0$ and $n \rightarrow \infty$, then*

$$(2.14) \quad \sqrt{n}(T_n - M^2) \rightarrow_{\mathcal{D}} \mathcal{N}(0, \mu_1^2),$$

where the asymptotic variance is given by

$$(2.15) \quad \mu_1^2 = 4 \int_0^1 \sigma^2(u) \Delta^2(u) f(u) du,$$

$\Delta = m - P_{\{g_1, \dots, g_p\}} m$ and $P_{\{g_1, \dots, g_p\}} m$ is the projection of m onto $\text{span}\{g_1, \dots, g_p\}$ with respect to the inner product $\langle p, q \rangle = \int_0^1 p(u)q(u)f(u) du$.

It is remarkable that the normalizing factor is of different order in Theorems 2.3 and 2.4. Under the null hypothesis $H_0: m \in \mathcal{M}$, the variance of T_n is of order $(n^2h)^{-1}$ while under the alternative $m \notin \mathcal{M}$, this is of order n^{-1} . In principle, Theorem 2.3 can be used to construct an asymptotic level α test for the hypothesis of linearity and Theorem 2.4 establishes the consistency of such a test. It will also provide useful information about the type II error of such a test at any particular point of the alternative (see Section 4). This is particularly important for the application of a goodness-of-fit test, because the acceptance of the null will lead to a subsequent data analysis adapted to the model \mathcal{M} , and it is desirable to control the corresponding error of this procedure. Moreover, the result of Theorem 2.4 allows the construction of tests for the problem of precise hypotheses

$$(2.16) \quad H: M^2 > \Delta, \quad K: M^2 \leq \Delta$$

[see Berger and Delampady (1987)] where M^2 is defined in (2.3) and measures the deviation of the regression function m from the linear model \mathcal{M} . The hypothesis (2.16) is rejected if $\sqrt{n}(T_n - \Delta) \leq \hat{\mu}_1 u_\alpha$ where $\hat{\mu}_1^2$ is an appropriate estimator of μ_1^2 and u_α denotes the α -quantile of the standard normal distribution. Note that in this case rejection of H in (2.16) allows us to assess the validity of the model \mathcal{M} within an L^2 -neighborhood at a controlled error rate. Finally, it is also notable that Theorem 2.4 allows the construction of confidence intervals for the parameter M^2 , which measures the deviation from the linear model \mathcal{M} .

In all cases the choice of the bandwidth h becomes an important and non-trivial problem, which will be illustrated for the problem of testing the classical hypothesis $M^2 = 0$. It follows from Hall and Marron (1990) that in a homoscedastic regression the asymptotic optimal (with respect to the MSE criterion) bandwidth is of order $n^{-2/(4r+1)}$, and a straightforward calculation shows that this result carries over to the heteroscedastic case, where the optimal bandwidth is given by

$$h_{\text{opt}} = \left\{ \frac{\mu_0^2 n^{-2}}{4rC_2^2} \right\}^{1/(4r+1)}.$$

Note that under the null hypothesis (1.1), this choice produces a nonvanishing bias in Theorem 2.3, that is,

$$(2.17) \quad \frac{n\sqrt{h_{\text{opt}}}}{\mu_0} T_n + \frac{1}{2\sqrt{r}} + \frac{C_3}{\sqrt{h_{\text{opt}}\mu_0}} \rightarrow_{\mathcal{D}} \mathcal{N}(0, 1)$$

but under the alternative we have

$$\frac{\sqrt{n}}{\mu_1}(T_n - M^2) \rightarrow_{\mathcal{G}} \mathcal{N}(0, 1).$$

In principle (2.17) provides an asymptotic test for the hypothesis $H_0: m \in \mathcal{M}$ where μ_0^2 is estimated by

$$(2.18) \quad \hat{\mu}_0^2 = 2 \int_{-\infty}^{\infty} (2K(u) - K * K(u))^2 du \sum_{i=1}^{n-1} (t_{i+1} - t_i) [y_i - \hat{\theta}_n^T g(t_i)]^2 \times [y_{i+1} - \hat{\theta}_n^T g(t_{i+1})]^2$$

(for consistency of $\hat{\mu}_0^2$ see Lemma 2.5 below) and h_{opt} is obtained by some (consistent) cross validation procedure. However, for moderate sample sizes the bias in (2.17) has a serious impact on the quality of the normal approximation. Our numerical studies show that in the problem of testing the hypothesis $H_0: m \in \mathcal{M}$, the bias of T_n is more important than the variance and a balance between bias and variance seems only appropriate for very large sample sizes. Based on an extensive simulation study, we recommend the bandwidth

$$(s^2/n)^{2/(2r+1)}$$

for the testing problem (1.1); here r is the order of the corresponding kernel and $s^2 = \int_0^1 \sigma^2(t)f(t) dt$ the integrated variance of the error. This specific order can also be motivated by the requirement that the “bias” $nC_2h^{2r+1/2}$ converges to 0 at a reasonable rate which is $n^{-2r/(2r+1)}$ for the proposed choice. For this choice, the hypothesis of linearity is rejected if

$$(2.19) \quad n\sqrt{h} \left(T_n + \frac{C_3}{nh} \right) > u_{1-\alpha} \hat{\mu}_0; \quad h = o(n^{-2/(4r+1)}),$$

where $u_{1-\alpha}$ is the $1 - \alpha$ quantile of the standard normal distribution and the estimator of the variance $\hat{\mu}_0^2$ is defined by (2.18). Note that $C_3 = 0$ under the assumption of a uniform design or a heteroscedastic error. The following lemma establishes the consistency of the variance estimator $\hat{\mu}_0^2$; the proof can be found in Section 5.

LEMMA 2.5. *Under the assumptions of Theorem 2.3 we have for the estimator in (2.18),*

$$\hat{\mu}_0^2 \rightarrow^P \mu_0^2,$$

where μ_0^2 is defined in (2.13).

Note that in the case of a homoscedastic error structure we can use a modified estimator for the asymptotic variance, that is,

$$(2.20) \quad \tilde{\mu}_0^2 = 2 \int_{-\infty}^{\infty} (2K(u) - K * K(u))^2 du \cdot \hat{\sigma}_{\text{LSE}}^4.$$

The consistency of this estimator under the null hypothesis in a homoscedastic setup is obvious by Lemma 2.1 and

$$(2.21) \quad \text{Var}(\hat{\sigma}_{\text{LSE}}^2) = O(n^{-1}).$$

REMARK 2.6. As pointed out by a referee, similar results to those stated in Theorem 2.3 and 2.4 can be obtained by using different smoothing procedures for the calculation of the residuals $\hat{\varepsilon}_i = y_i - \sum_{j=1}^n w_{ij} y_j$ corresponding to the nonparametric variance estimator. Roughly speaking, the results of Theorems 2.3 and 2.4 remain valid, where the bias term $C_2 h^{2r}$ and the asymptotic variance μ_0^2 in (2.13) depend on the specific smoothing procedure under consideration, while the asymptotic variance μ_1^2 in (2.15) is not changed. We will illustrate this statement by considering two specific examples. The first example is the estimator of Gasser and Müller (1979) which uses the weights

$$w_{ij} = \frac{1}{h} \int_{s_{j-1}}^{s_j} K\left(\frac{t - t_i}{h}\right) dt,$$

where $s_0 = 0, s_n = 1, s_j = (t_j + t_{j+1})/2 (j = 1, \dots, n - 1)$. For this smoothing procedure a careful inspection of the proofs in Section 5 shows that Theorems 2.3 and 2.4 remain valid, where the constant C_2 in (2.10) is replaced by $\kappa_r^2 \int_0^1 (m^{(r)}(t))^2 dt$.

A similar argument applies to our second example, the local polynomial estimator [see, e.g., Fan (1992) or Fan and Gijbels (1996)]. Here the residuals $\hat{\varepsilon}_i$ in the estimator of Hall and Marron (1990) are replaced by

$$\tilde{\varepsilon}_i = y_i - \hat{\beta}_{i0}$$

and $(\hat{\beta}_{i0}, \dots, \hat{\beta}_{ip})$ is the minimizer of

$$\sum_{k=1}^n \left\{ y_k - \sum_{j=0}^p \beta_{ij} (t_k - t_i)^j \right\}^2 K\left(\frac{t_k - t_i}{h}\right)$$

and the case $p = 0$ corresponds to the Nadaraya–Watson estimator. It is pointed out by Wand and Jones (1995) that the estimator $\hat{\beta}_{i0}$ of $m(t_i)$ is asymptotically equivalent to

$$\tilde{m}(t_i) = \frac{1}{nhf(t_i)} \sum_{j=1}^n K_{(p)}\left(\frac{t_j - t_i}{h}\right) y_j,$$

where the kernel $K_{(p)}$ is a higher order kernel [see Gasser, Müller and Mammitzsch (1985)] defined by

$$K_{(p)}(u) = \frac{|M_p(u)|}{|N_p|} K(u),$$

N_p is the $(p+1) \times (p+1)$ matrix having (i, j) entry equal to $\int_{-\infty}^{\infty} u^{i+j-2} K(u) du$ and the matrix $M_p(u)$ is obtained from N_p by replacing the first column by

the vector $(1, u, \dots, u^p)$. In this case (under appropriate smoothness assumptions) Theorem 2.3 and 2.4 remain valid, where the bias in (2.10) has to be replaced by

$$\rho_{p+1}^2 \int_0^1 (m^{(p+1)}(t))^2 f(t) dt$$

or

$$\rho_{p+2}^2 \int_0^1 [(p+2)m^{(p+1)}(t)f'(t)/f(t) + m^{(p+2)}(t)]^2 f(t) dt$$

corresponding to the case of an odd or even order p and

$$\rho_k = \frac{h^k}{k!} \int_{-\infty}^{\infty} u^k K_{(p)}(u) du.$$

Moreover, the kernel $K_{(p)}$ appears in (2.13) instead of K .

REMARK 2.7. Theorem 2.3 and 2.4 can be extended without any difficulties to the case of a multivariate predictor. If a product kernel

$$K(x_1, \dots, x_d) = \prod_{j=1}^d K_j(x_j)$$

is used, the normalizing factor under the null hypothesis is $n\sqrt{h_1 \dots h_d}$ where h_j is the bandwidth used for the j th marginal kernel ($j = 1, \dots, d$). Because of the curse of dimensionality, the choice of h_j is even more critical in this case. Based on a first numerical experience in the two-dimensional case, we recommend for a kernel of order 2 the bandwidth

$$h_j = \left(\frac{1}{n} \int_0^1 \sigma^2(t) f(t) dt \right)^{2/(d+4)}, \quad j = 1, 2, \dots, d.$$

REMARK 2.8. For local alternatives of the form $m_n(t) = \theta_0^T g(t) + \delta_n l(t)$ ($\theta_0 \in \Theta, \delta_n = (n\sqrt{h})^{-1/2}$) a careful inspection of the proof of Theorem 2.3 shows that

$$n\sqrt{h} \left(T_n + C_2 h^{2r} + \frac{C_3}{nh} \right) \rightarrow_{\mathcal{D}} \mathcal{N}(\mu, \mu_0^2),$$

where T_n, μ_0^2 are defined in (2.11), (2.13), respectively, and $\mu = \int_0^1 l^2(t) f(t) dt$.

REMARK 2.9. The idea of comparing a parametric and nonparametric variance estimator has been exploited previously by Yatchew (1992). Yatchew’s test can detect local alternatives at a rate $1/\sqrt{n}$ and is in this asymptotic sense more efficient (see the previous remark). However, this theoretical advantage is compensated by several practical drawbacks, because Yatchew’s approach is based on the knowledge of several parameters and assumptions which makes its application difficult. It requires (uniform) bounds for the regression function and its first and second derivative which basically gives three “smoothing

parameters.” The specification of these bounds is usually difficult in practice. Secondly it uses a sample splitting which usually results in a loss of power for moderate sample sizes. Finally Yatchew’s test is only applicable for a homoscedastic error.

3. Further discussion. In this section we present several extensions of the test based on the difference of variance estimators and discuss its relation to other lack-of-fit tests proposed in the literature.

3.1. *Testing the validity of nonlinear models.* The results presented in Section 2 remain true if \mathcal{M} is a class of nonlinear models,

$$(3.1) \quad \mathcal{M} = \{m(t, \theta) \mid \theta \in \Theta\},$$

where $\Theta \subset \mathbb{R}^p$, and the minimum of

$$(3.2) \quad M_{\Theta}^2 = \min_{\theta \in \Theta} \int_0^1 [m(t) - m(t, \theta)]^2 f(t) dt$$

is unique and attained at an interior point $\theta_0 \in \Theta$. Then under regularity assumptions [see, e.g., Seber and Wild (1989), pages 572–574 or Gallant (1987), Chapter 4], Theorems 2.3 and 2.4 remain valid, where M^2 in (2.14) has to be replaced by M_{Θ}^2 . The proof uses a Taylor expansion and the fact that the sum of squared residuals in the nonlinear model can be approximated by

$$\hat{\sigma}_{\text{LSE}}^2 = \min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n (y_i - m(t_i, \theta))^2 = \frac{1}{n} \eta^T (I - G(G^T G)^{-1} G^T) \eta + O_p\left(\frac{1}{n}\right),$$

where $G^T = ((\partial m / \partial \theta)(t_i, \theta_0))_{i=1}^n \in \mathbb{R}^{p \times n}$ and $\eta = (y_i - m(t_i, \theta_0))_{i=1}^n$. Roughly speaking, this means that the nonlinear model can be treated as the linear model with the p regression functions

$$g_1(t) = \frac{\partial}{\partial \theta_1} m(t, \theta_0), \dots, g_p(t) = \frac{\partial}{\partial \theta_p} m(t, \theta_0),$$

where the regression $m(t)$ has to be replaced by $m(t) - m(t, \theta_0)$. The asymptotic normality now follows along the lines of Section 2 observing that w.l.o.g. the regression functions g_1, \dots, g_p can be assumed as orthonormal with respect to the design density f . This implies for the projection of a function q onto $\text{span}\{g_1, \dots, g_p\}$,

$$P_{\{g_1, \dots, g_p\}} q = \sum_{l=1}^p \langle q, g_l \rangle g_l.$$

Consequently, the quantity M^2 in Theorem 2.4 is given by

$$\begin{aligned} M^2 &= \int_0^1 [m(t) - m(t, \theta_0) - \sum_{l=1}^p \langle m - m_{\theta_0}, g_l \rangle g_l(t)]^2 f(t) dt \\ &= \int_0^1 (m(t) - m(t, \theta_0))^2 f(t) dt, \end{aligned}$$

where $m_{\theta_0} = m(\cdot, \theta_0)$ and the last equality is a consequence of (3.2) which implies

$$0 = \frac{\partial}{\partial \theta_l} \int_0^1 [m(t) - m(t, \theta)]^2 f(t) dt \Big|_{\theta=\theta_0} = -2 \langle m - m_{\theta_0}, g_l \rangle, \quad l = 1, \dots, p.$$

3.2. Random design. The approach of using the difference of two variance estimators as a goodness-of-fit statistic can be transferred to the case of a random design where t_1, \dots, t_n are realizations of i.i.d. random variables U_1, \dots, U_n with positive density f on the interval $[0, 1]$.

THEOREM 3.1. *Assume that (2.1), (2.4), (2.5) and (2.6) are satisfied, $n \rightarrow \infty$, and let T_n be the statistic in (2.11) where t_1, \dots, t_n are replaced by i.i.d. random variables U_1, \dots, U_n with positive design density f on the interval $[0, 1]$.*

(a) *If $M^2 = 0$, then $n\sqrt{h}(T_n + C_2 h^{2r} + C_3/nh) \rightarrow_{\mathcal{D}} \mathcal{N}(0, \mu_0^2)$, where μ_0^2 is defined in (2.13).*

(b) *If $M^2 > 0$, then $\sqrt{n}(T_n - M^2) \rightarrow_{\mathcal{D}} \mathcal{N}(0, \tilde{\mu}_1^2)$, where*

$$(3.3) \quad \tilde{\mu}_1^2 = 4E[\Delta^2(U_1)\sigma^2(U_1)] + \text{Var}[\Delta^2(U_1)] = \mu_1^2 + \Delta_1^2$$

and $\Delta = m - P_{\{g_1, \dots, g_p\}}m$ denotes the difference between m and its projection onto $\text{span}\{g_1, \dots, g_p\}$.

The proof of Theorem 3.1 can be obtained by similar arguments to those given for the fixed design in Section 5 and is therefore omitted. Comparing this result with Theorems 2.3 and 2.4, we observe no difference between the random and fixed design assumption under the null hypothesis $M^2 = 0$. Surprisingly, there appears an additional term in the asymptotic variance under the alternative $M^2 > 0$ which results in a loss of power of the corresponding test caused by the randomness of the predictor. This term can be explained by the well-known formula

$$\text{Var}(Z) = \text{Var}[E[Z | U]] + E[\text{Var}[Z | U]],$$

where the first and second term on the right-hand side correspond to Δ_1^2 and μ_1^2 in (3.3).

3.3. Alternative goodness-of-fit tests. The methods which are most similar in spirit to the procedures presented in this paper were proposed by Härdle and Mammen (1993) and Zheng (1996), who obtained the same rate of convergence under the null hypothesis. They discussed the random design and also considered alternatives converging to the null with a rate $\delta_n = (n\sqrt{h})^{-1/2}$. These results can be directly transferred to the case of a fixed design satisfying assumption (2.2). We now investigate the asymptotic behavior of the procedures proposed by Härdle and Mammen (1993) and Zheng (1996) under fixed alternatives.

Härdle and Mammen (1993) based their criterion on a (weighted) L^2 -distance between a parametric and a nonparametric fit of the regression function, that is,

$$(3.4) \quad T_n^{\text{HM}} = \int_0^1 \{ \hat{m}_n(t) - \mathcal{K}_{h,n}(\hat{\theta}_n^T g(t)) \}^2 \pi(t) dt,$$

where \hat{m}_n denotes the Nadaraya–Watson estimator of the regression curve [see Nadaraya (1964), Watson (1964)] and $\mathcal{K}_{h,n}$ is a smoothing operator defined by

$$(3.5) \quad \mathcal{K}_{h,n}(m(t)) = \frac{\sum_{i=1}^n K((t_i - t)/h)m(t_i)}{\sum_{i=1}^n K((t_i - t)/h)}.$$

Under alternatives converging with a rate $\delta_n = (n\sqrt{h})^{-1/2}$ to the null, they obtained results similar to Theorem 2.3. We now extend their result to the case of fixed alternatives and present an analogue of Theorems 2.4 and 3.1, for the statistic (3.4).

THEOREM 3.2. *Let (2.1), (2.4), (2.5) and (2.6) be satisfied, $M^2 > 0$, $n \rightarrow \infty$, and define $\Delta(t) = m(t) - P_{\{g_1, \dots, g_p\}}m(t)$.*

(a) *In the fixed design case with assumption (2.2) we have for the statistic T_n^{HM} of Härdle and Mammen (1993) defined by (3.4),*

$$(3.6) \quad \sqrt{n}(T_n^{\text{HM}} - b_h) \rightarrow_{\mathcal{D}} \mathcal{N}(0, \mu_2^2),$$

where the bias is given by

$$(3.7) \quad \begin{aligned} b_h &= \frac{1}{nh^2} \int_0^1 \int_0^1 K^2\left(\frac{x-v}{h}\right) f(v) \sigma^2(v) \frac{\pi(x)}{f^2(x)} dv dx \\ &+ \int_0^1 \left\{ \int_0^1 \frac{1}{h} K\left(\frac{y-x}{h}\right) (\Delta f)(y) dy \right\}^2 \frac{\pi(x)}{f^2(x)} dx \end{aligned}$$

and the asymptotic variance is

$$\mu_2^2 = 4 \int_0^1 \sigma^2(y) \left\{ \left(\pi \frac{\Delta}{f} \right)(y) - P_{\{g_1, \dots, g_p\}} \left(\pi \frac{\Delta}{f} \right)(y) \right\}^2 f(y) dy,$$

which reduces for $f \equiv \pi$ to

$$\mu_2^2 = 4 \int_0^1 \sigma^2(y) \Delta^2(y) f(y) dy.$$

(b) *In the random design case the assertion (3.6) is still valid where the asymptotic variance μ_2^2 has to be replaced by*

$$\begin{aligned} \tilde{\mu}_2^2 &= 4E \left[\sigma^2(U_1) \left\{ \left(\pi \frac{\Delta}{f} \right)(U_1) - P_{\{g_1, \dots, g_p\}} \left(\pi \frac{\Delta}{f} \right)(U_1) \right\}^2 \right] \\ &+ 4 \text{Var} \left[\left(\frac{\Delta}{f} \right)^2(U_1) (\pi f)(U_1) \right], \end{aligned}$$

which reduces for $f \equiv \pi$ to

$$\tilde{\mu}_2^2 = 4E[\sigma^2(U_1)\Delta^2(U_1)] + 4 \text{Var}[\Delta^2(U_1)].$$

Note that the bias in Theorem 3.2 can be rewritten as

$$b_h = \frac{1}{nh} \int_{-\infty}^{\infty} K^2(u) du \int_0^1 \frac{\sigma^2(u)}{f(u)} \pi(u) du + \int_0^1 \Delta^2(u) \pi(u) du + O(h^r),$$

where the terms depending on n are asymptotically negligible in (3.6) if $nh^2 \rightarrow \infty$, $nh^{2r} \rightarrow 0$ and the variance function is assumed to be r times continuously differentiable.

Roughly speaking, the test proposed in this paper and the test of Härdle and Mammen (1993) are based on an estimator of the integrated deviation $\Delta^2(t) = \{m(t) - P_{\{g_1, \dots, g_p\}}(m(t))\}^2$. A minor difference is that the last-named authors consider a weight function π which corresponds in our approach to the use of weighted sums of squared residuals in (2.11).

An important difference between both methods is the different order of the bias in the normalized statistic in the case of a uniform design or a homoscedastic error. Under the null hypothesis, there appears an additional term of order $h^{-1/2}$ in the statistic $n\sqrt{h} T_n^{\text{HM}}$ which cannot be diminished by choosing an appropriate order of convergence for the bandwidth (as we did at the end of Section 2). The bias $C_2 h^{2r}$ in Theorem 2.3 does not appear in Theorem 3.2 because of the additional smoothing of the parametric regression function in (3.4). A similar procedure could be used to reduce the bias in Theorems 2.3 and 3.1 as well.

An alternative approach was discussed in Zheng (1996) who considered a test based on an estimate of

$$E[\{y_1 - P_{\{g_1, \dots, g_p\}}m(U_1)\}E(y_1 - P_{\{g_1, \dots, g_p\}}m(U_1) | U_1)f(U_1)],$$

that is,

$$(3.8) \quad V_n = \frac{1}{hn(n-1)} \sum_{i=1}^n \sum_{j \neq i} e_i e_j K\left(\frac{t_i - t_j}{h}\right),$$

where e_i is the residual at the point t_i from the parametric fit. Under alternatives converging to the null with a rate $\delta_n = (n\sqrt{h})^{-1/2}$, Zheng (1996) obtained an analogue of Theorems 2.3 and 3.1 using degenerate U -statistics theory. The following result extends his findings and gives the asymptotic distribution of V_n under fixed alternatives.

THEOREM 3.3. *Assume that (2.1), (2.4), (2.5) and (2.6) are satisfied, $M^2 > 0$, $n \rightarrow \infty$, and define $\Delta(t) = m(t) - P_{\{g_1, \dots, g_p\}}m(t)$.*

(a) *In the fixed design case with assumption (2.2), we have for the statistic of Zheng (1996) defined by (3.8),*

$$(3.9) \quad \sqrt{n} \left\{ V_n - \int_0^1 K(u)(\Delta f)(x)(\Delta f)(x - uh) du dx \right\} \rightarrow_{\mathcal{D}} \mathcal{N}(0, \mu_3^2),$$

where the asymptotic variance is given by

$$(3.10) \quad \mu_3^2 = 4 \int_0^1 \sigma^2(u) \{(\Delta f)(u) - P_{\{g_1, \dots, g_p\}}(\Delta f)(u)\}^2 f(u) du.$$

(b) Under the random design assumption, the assertion (3.9) is still valid where the asymptotic variance in (3.10) has to be replaced by

$$\tilde{\mu}_3^2 = 4E[\sigma^2(U_1)\{(\Delta f)(U_1) - P_{\{g_1, \dots, g_p\}}(\Delta f)(U_1)\}^2] + 4 \text{Var}[\Delta^2(U_1)f(U_1)].$$

Note that for a kernel of order r the bias in (3.9) can be rewritten as

$$(3.11) \quad \int_0^1 \Delta^2(u)f^2(u) du + \kappa_r h^r \int_0^1 (\Delta f)(u)(\Delta f)^{(r)}(u) du + o(h^r),$$

and as a consequence Zheng’s (1996) procedure is comparable with Härdle and Mammen’s (1993) method and with the test (2.19) based on weighted sums of squared residuals, where the weight function is chosen as $\pi = f^2$. The proof of Theorem 3.3 can be found in Section 5. The proof of Theorem 3.2 is similar and therefore omitted.

4. Finite sample performance. In this section we study the finite sample behavior of the test (2.19) introduced in Section 2 and compare it with other procedures proposed in the literature. Unless stated otherwise, we consider a random and a fixed uniform design on the interval $[0, 1]$ (i.e., $f \equiv 1$) and the sample size is chosen as $n = 50, 100, 200$ and 400 . We use the kernel $K(x) = \frac{3}{4}(1 - x^2)I_{[-1, 1]}(x)$ of order 2 and the corresponding bandwidth $h = (s^2/n)^{2/5}$ with $s^2 = \int_0^1 \sigma^2(t)f(t) dt$, which corresponds to the situation considered in (2.19). In practice s^2 can be estimated by a preliminary variance estimator [see, e.g., Rice (1984) or Gasser, Skroka and Jennen-Steinmetz (1986)] and our simulation results show that this estimation has a negligible impact on the distributional behavior of the test statistic (these results are not displayed). The asymptotic variance μ_0^2 in Theorems 2.3 and 3.1 was estimated by $\tilde{\mu}_0^2$ in (2.20) for the case of a homoscedastic error (see Examples 4.1, 4.2 and 4.4) and by $\hat{\mu}_0^2$ in (2.18) for the case of a heteroscedastic error (see the first part of Example 4.3). We also performed simulations with higher order kernels [see Gasser, Müller and Mammitzsch (1985)] and local polynomial estimators [see Remark 2.6]. These results did not yield a substantial difference with respect to power and approximation of the level and for the sake of brevity are not displayed.

EXAMPLE 4.1 (Testing procedures with optimal rates). It is demonstrated in Section 2 that the test (2.19) detects local alternatives converging to the null at a rate $n^{-1/2}h^{-1/4}$, which reduces to $n^{-r/(2r+1)}$ for the proposed bandwidth $h = (s^2/n)^{2/(2r+1)}$. There are several procedures proposed in the literature which are able to detect local alternatives that converge to the null at the optimal rate $1/\sqrt{n}$ and are in this (asymptotic) sense more efficient [see,

e.g., Eubank and Hart (1992), Yatchew (1992), Stute (1997), Stute, Gonzalés Manteiga and Presedo Quindimil (1998)].

Eubank and Hart (1992) used a Fourier-series approach based on an extended model and a test if the coefficients in this extension vanish. Stute (1997) introduced a marked empirical process and analyzed the principal components of the covariance kernel of a nonstandard Gaussian process. Because this procedure is rather difficult to implement, Stute, Gonzalés Manteiga and Presedo Quindimil (1998) proposed a bootstrap version of the test. Yatchew's (1992) test is also based on a difference of two variance estimators but uses a sample splitting which is unrealistic in any application. Additionally, it requires a restriction to a class of regression functions with uniform bounds on the first two derivatives and, as a consequence, a constrained least squares estimation which makes its implementation in a detailed simulation study difficult. For these reasons and for space considerations, we include in our comparison only the procedures proposed by Eubank and Hart (1992) and Stute, Gonzalés Manteiga and Presedo Quindimil (1998). Because the test of the last-named authors requires the bootstrap we take their setup as a reference example (see Tables 1 and 2 of their paper). More precisely, we consider the model

$$(4.1) \quad m(t) = 5t + at^2$$

for various values of a where the case $a = 0$ corresponds to the null hypothesis of a linear regression through the origin. The results are listed in Table 1 for a homoscedastic error and show the relative proportion of rejection calculated by 1000 simulations on the basis of a 5% level.

The asymptotics provided by Theorem 2.3 yields a very accurate approximation of the nominal level, even for $n = 50$. Comparing the power of both tests we observe a slightly higher power for the test of Stute, Gonzáles Manteiga and Presedo Quindimil (1998) in most (but not all) cases. On the other

TABLE 1
Simulated rejection probabilities of the test (2.19) in the model (4.1) for various values of a , σ^2 , n and a fixed and random uniform design on the interval $[0, 1]$

$\sigma^2 a$	$n = 50$		$n = 100$		$n = 200$		$n = 400$		
	Fixed	Random	Fixed	Random	Fixed	Random	Fixed	Random	
0	0.053	0.042	0.054	0.052	0.053	0.053	0.051	0.051	
1	1	0.095	0.071	0.116	0.111	0.167	0.146	0.297	0.283
	2	0.181	0.149	0.343	0.297	0.597	0.536	0.892	0.872
0	0.060	0.049	0.056	0.052	0.055	0.057	0.048	0.054	
2	1	0.089	0.065	0.084	0.100	0.130	0.125	0.184	0.178
	2	0.120	0.108	0.194	0.188	0.335	0.295	0.594	0.589
0	0.059	0.051	0.054	0.050	0.054	0.058	0.058	0.054	
3	1	0.081	0.083	0.092	0.101	0.103	0.110	0.144	0.143
	2	0.089	0.086	0.147	0.146	0.221	0.215	0.427	0.415

TABLE 2

Simulated rejection probabilities of the test of Eubank and Hart (1992) in the model (4.1) for various values of α , σ^2 , n and a fixed uniform design

σ^2	α	$n = 50$	$n = 100$	$n = 200$	$n = 400$
1	0	0.076	0.053	0.056	0.058
	1	0.103	0.124	0.170	0.283
	2	0.166	0.290	0.491	0.804
2	0	0.067	0.070	0.045	0.053
	1	0.087	0.091	0.115	0.182
	2	0.122	0.181	0.290	0.478
3	0	0.072	0.052	0.055	0.058
	1	0.082	0.073	0.083	0.146
	2	0.098	0.127	0.207	0.358

hand, the test discussed in this paper does not require any bootstrap estimation and as a consequence it is easier to apply. Moreover, Theorems 2.4 and 3.1 even provide the asymptotic distribution of the test statistic in (2.19) under the alternative, which allows a calculation of the probability for the type II error at any particular point m , if the test accepts the hypothesis $H_0: m \in \mathcal{M}$. From a practical point of view, this is particularly important, because the acceptance of the null will lead to a subsequent data analysis adapted to the specific model \mathcal{M} , and it is desirable to estimate the corresponding probability of an error in this procedure at any particular point in the alternative. Note that by Theorems 2.3 and 2.4 the power of test (2.19) at a particular $m \notin \mathcal{M}$ depends asymptotically only through the L^2 -distance M^2 on m and we obtain approximately

$$(4.2) \quad p = P(\text{“rejection”}) \approx \Phi\left(\frac{\sqrt{n}}{\mu_1} \left\{ M^2 - \frac{u_{1-\alpha}\mu_0}{n\sqrt{h}} \right\}\right)$$

for the probability of a type II error of the test (2.19). As a numerical example we estimated the probability of rejection in the case of a fixed design, $\alpha = 2$, $n = 100$, $\sigma^2 = 1$ and obtained $p \approx 0.334$, while the corresponding simulated probability is 0.343 (see Table 1). It is also worthwhile to mention that Theorems 2.4 and 3.1 can be used for the construction of confidence intervals for M^2 and for testing precise hypotheses of the form (2.16) (see the discussion in Section 2).

Table 2 shows the simulated power for the test proposed by Eubank and Hart (1992) for the regression function (4.1). Note that this test requires the specification of $(n - 1)$ additional regression functions which we chose as

$$u_{jn}(x) = \cos(\pi j(2x - 1)), \quad j = 1, \dots, n - 1.$$

We observe a less accurate approximation of the nominal level for moderate sample sizes ($n = 50$) and a lower power of the test of Eubank and Hart (1992) compared to that of the test (2.19) and the procedure of Stute, Gonzalés Manteiga and Presedo Quindimil (1998). It is also worthwhile to mention that

these results depend sensitively on the choice of the additional regression functions $\{u_{jn}(x) \mid j = 1, \dots, n - 1\}$ and the corresponding ordering. For an illustration of this phenomenon, see Dette and Munk (1998). Finally, we note that in contrast to our work, Eubank and Hart's (1992) procedure cannot be used in a random design, while the test of Stute, Gonzalés Manteiga and Presedo Quindimil (1998) is not directly applicable under the fixed design assumption.

EXAMPLE 4.2 (Testing procedures based on smoothing methods). Our second example compares our procedure with the methods which use kernel estimation for the calculation of the residuals. As pointed out in Section 3.3, the procedures of Härdle and Mammen (1993) and Zheng (1996) are most similar in spirit to the test proposed in this paper. A further natural competitor is a test introduced by Azzalini and Bowman (1993). Roughly speaking, this is obtained as a pseudolikelihood ratio test by replacing the original observations in the difference of the variance estimators under a parametric and nonparametric fit by the residuals under the parametric fit. Under the assumption of a normally distributed error and a fixed design, these authors demonstrated that the distribution of the corresponding test statistic is given by a linear combination of independent χ^2 -distributions. For practical purposes, Azzalini and Bowman (1993) recommended an approximation by a $a + b\chi_c^2$ distribution by fitting the first three cumulants, which could also be used for nonnormal errors. A principal drawback of this test is that it cannot be applied under the random design assumption.

Because the test of Härdle and Mammen (1993) requires an application of the bootstrap, we choose their setup as a reference example, that is,

$$(4.3) \quad m(t) = 2t - t^2 + c\left(t - \frac{1}{4}\right)\left(t - \frac{1}{2}\right)\left(t - \frac{3}{4}\right)$$

and $\mathcal{M} = \{a_0 + a_1t + a_2t^2 \mid a_0, a_1, a_2 \in \mathbb{R}\}$. The variance is assumed to be constant and given by $\sigma^2 = 0.01$. Some results for selected cases are given in Table 4 for Zheng's (1996) test [where the asymptotic variance is estimated by formula (3.9) of his paper] and in Table 3 for the test (2.19) of the present paper. Finally, the corresponding results for the test of Azzalini and Bowman (1993) are shown in Table 5 in the case of a fixed design. The bandwidth was chosen as $h = (\sigma^2/n)^{2/5}$ in all cases according to the proposed choice in Section 2. The case $n = 100$ corresponds to the situation considered by Härdle and Mammen (1993).

We observe that all tests are comparable with respect to the power behavior. The test proposed in this paper yields a slightly better power than its competitors. Compared with Azzalini and Bowman's (1993) and Zheng's (1996) test, the improvement with respect to power is even more substantial, especially for relatively small sample sizes (e.g., $n = 50$). Note also that the test of Zheng (1996) is conservative in most cases, and all tests are not too sensitive with respect to the choice of the bandwidth. These results are not displayed and confirm recent findings of Azzalini and Bowman (1993) and Zheng (1996). The approximation of the level of Azzalini and Bowman's test (1993) is extremely

TABLE 3

Simulated rejection probabilities of the test (2.19) in the model (4.3) for various values of c and a fixed and random uniform design on the interval $[0, 1]$. The variance is constant and given by $\sigma^2 = 0.01$

c	$n = 50$		$n = 100$		$n = 200$		$n = 400$	
	Fixed	Random	Fixed	Random	Fixed	Random	Fixed	Random
0.0	0.065	0.055	0.062	0.058	0.061	0.057	0.057	0.048
0.5	0.112	0.084	0.134	0.106	0.163	0.135	0.234	0.235
1.0	0.215	0.144	0.328	0.287	0.571	0.478	0.851	0.822
1.5	0.432	0.308	0.673	0.608	0.929	0.884	1.000	0.994
2.0	0.698	0.531	0.922	0.844	1.000	0.993	1.000	1.000

accurate, which can be explained by the application of the χ^2 -approximation. It is worthwhile to mention that the test of Azzalini and Bowman (1993) is not extendable to the case of a random design. Moreover, a χ^2 -approximation will also improve the approximation of the nominal level of the tests of Zheng (1996), Härdle and Mammen (1993) and the test (2.19). This is demonstrated in Example 4.4 for the last-named test.

We finally remark that [in contrast to Härdle and Mammen (1993)], Azzalini and Bowman’s (1993), Zheng’s (1996) test and the test (2.19) proposed in this paper do not require any bootstrap estimates.

EXAMPLE 4.3 (Heteroscedasticity and nonuniformity). In this example we investigate the impact of deviations from homoscedasticity and nonuniformity (of the design) on power and level of the test (2.19). Our first example considers a heteroscedastic situation for the model (4.1) where the variance function is given by $\sigma^2(t) = 3(1 + ct^2)/(3 + c)$. The results are given in Table 6 under the same setup as considered in Example 4.1. Note that the function σ^2 has been normalized such that $\int_0^1 \sigma^2(t)f(t) dt = 1$. Compared to the homoscedastic case ($c = 0$) we observe no significant loss in the accuracy of the approximation of the nominal level and a loss of power with increasing values of c .

TABLE 4

Simulated rejection probabilities of Zheng’s (1996) test in the model (4.3) for various values of c and a fixed and random uniform design on the interval $[0, 1]$. The variance is constant and given by $\sigma^2 = 0.01$

c	$n = 50$		$n = 100$		$n = 200$		$n = 400$	
	Fixed	Random	Fixed	Random	Fixed	Random	Fixed	Random
0.0	0.044	0.043	0.043	0.044	0.053	0.048	0.052	0.054
0.5	0.059	0.067	0.087	0.078	0.130	0.123	0.229	0.229
1.0	0.136	0.311	0.271	0.256	0.543	0.423	0.843	0.843
1.5	0.311	0.277	0.609	0.556	0.913	0.868	0.998	0.999
2.0	0.550	0.487	0.889	0.799	0.998	0.985	1.000	1.000

TABLE 5

Simulated rejection probabilities of the test of Azzalini and Bowman (1993) in the model (4.3) for various values of c and a fixed uniform design on the interval $[0, 1]$. The variance is constant and given by $\sigma^2 = 0.01$

c	$n = 50$	$n = 100$	$n = 200$	$n = 400$
0.0	0.047	0.049	0.053	0.050
0.5	0.080	0.109	0.131	0.262
1.0	0.194	0.297	0.523	0.832
1.5	0.349	0.606	0.886	0.998
2.0	0.604	0.881	0.991	1.000

Our second example investigates the impact of a nonuniform design on power and level of the test (2.19). To this end we consider the model (4.1) with a homoscedastic error. In this case the argument in (4.2) shows that the power of the test is increasing with

$$\mu_1^2 = \mu_1^2(f) = 4\sigma^2 \int_0^1 f(t)[m(t) - P_{\{g_1, \dots, g_p\}}m(t)]^2 dt.$$

If $m(t) = 5t + at^2$, $g_1(t) = t$, this yields for the uniform distribution $\mu_1^2 \approx 0.05\sigma^2 a^2$. We considered two nonuniform densities,

$$f_1(t) = \frac{1}{2} + t, \quad f_2(t) = 2(1 - t),$$

for which the corresponding values are $\mu_1^2 \approx 0.053\sigma^2 a^2$ and $\mu_1^2 \approx 0.027\sigma^2 a^2$. Consequently, we expect similar results for the uniform and the density f_1 . The results of the simulation of this scenario are shown in Table 7 and basically reflect our asymptotic findings. In most cases we do not observe any substantial difference with respect to the approximation of the level and power

TABLE 6

Simulated power of the test (2.19) in the model (4.1) for various scenarios of heteroscedasticity and alternatives. The design is a uniform distribution, the variance function is given by $\sigma^2(t) = 3(1 + ct^2)/(3 + c)$ and normalized by $\int_0^1 \sigma^2(t)f(t) dt = 1$

c	a	$n = 50$		$n = 100$		$n = 200$		$n = 400$	
		Fixed	Random	Fixed	Random	Fixed	Random	Fixed	Random
0	0	0.048	0.050	0.050	0.057	0.054	0.050	0.053	0.049
0	1	0.093	0.071	0.094	0.092	0.168	0.163	0.296	0.285
	2	0.168	0.132	0.328	0.254	0.616	0.537	0.886	0.886
	0	0.047	0.051	0.043	0.047	0.058	0.045	0.055	0.045
1	1	0.079	0.060	0.088	0.085	0.152	0.156	0.301	0.260
	2	0.158	0.142	0.323	0.254	0.578	0.522	0.896	0.875
	0	0.046	0.043	0.045	0.050	0.052	0.048	0.048	0.048
2	1	0.074	0.059	0.097	0.092	0.153	0.151	0.261	0.255
	2	0.149	0.129	0.306	0.243	0.431	0.546	0.882	0.850

TABLE 7

Simulated rejection probabilities of the test (2.19) in the model (4.1) for various values of a , σ^2 , n and a fixed and random design with density $f_1(t) = \frac{1}{2} + t$

$\sigma^2 a$	$n = 50$		$n = 100$		$n = 200$		$n = 400$	
	Fixed	Random	Fixed	Random	Fixed	Random	Fixed	Random
0	0.051	0.047	0.056	0.052	0.054	0.052	0.053	0.052
1 1	0.070	0.067	0.104	0.103	0.165	0.182	0.281	0.285
2	0.147	0.133	0.328	0.275	0.623	0.571	0.911	0.867
0	0.055	0.055	0.054	0.055	0.057	0.047	0.057	0.053
2 1	0.063	0.061	0.078	0.079	0.109	0.121	0.177	0.169
2	0.106	0.088	0.190	0.155	0.334	0.323	0.577	0.597
0	0.056	0.058	0.052	0.059	0.054	0.047	0.050	0.056
3 1	0.061	0.059	0.069	0.065	0.074	0.094	0.139	0.139
2	0.076	0.072	0.129	0.124	0.224	0.225	0.449	0.422

for the density f_1 . A few cases show a (slight) loss of power if the design is not uniform.

Similarly, the use of the density f_2 should yield a more substantial loss with respect to the power of the test (2.19). This is reflected in Table 8 where (compared to the uniform design in Table 1) we observe a larger difference in power for most cases but still a sufficiently accurate approximation of the nominal level. We finally remark that it follows from Elfving’s theorem [Elfving (1952)] that

$$\sup_f \mu_1^2(f) \approx 0.1178c^2\sigma^2,$$

where the sup on the left-hand side is taken with respect to all positive densities on the interval $[0, 1]$. Moreover, the sup is not attained in this class but can be approximated arbitrarily close by approximating the discrete measure with mass $1/\sqrt{2}$ and $1 - 1/\sqrt{2}$ at the points $\sqrt{2} - 1$ and 1, respectively.

EXAMPLE 4.4 (Small sample sizes and χ^2 -approximation). It is demonstrated in the previous examples that the normal approximation based on Theorems 2.3 and 2.4 is sufficiently accurate for a sample size $n \geq 50$. For smaller sample sizes we observe a loss of accuracy in the approximation of the nominal level. In this case we propose an approximation for the distribution of the statistic which is similar to Azzalini and Bowman (1993). More precisely, we used an approximation by an $a\chi_b^2 + c$ distribution for the statistic T_n , where a, b, c are determined by fitting the first three moments. Table 9 shows some results of the simulated level in the model (4.1). We observe a sufficiently accurate approximation in all cases and a substantial improvement compared to the normal approximation. Of course the accuracy of the approximation of the level could be further improved by fitting more moments

TABLE 8

Simulated rejection probabilities of the test (2.19) in the model (4.1) for various values of n , a , σ^2 and a fixed and random design with density $f_2(t) = 2(1 - t)$

$\sigma^2 a$	$n = 50$		$n = 100$		$n = 200$		$n = 400$	
	Fixed	Random	Fixed	Random	Fixed	Random	Fixed	Random
0	0.047	0.055	0.052	0.052	0.052	0.057	0.052	0.054
1 1	0.061	0.060	0.076	0.063	0.110	0.106	0.173	0.158
2	0.115	0.111	0.219	0.165	0.349	0.312	0.604	0.551
0	0.055	0.050	0.047	0.054	0.055	0.044	0.051	0.051
2 1	0.060	0.053	0.070	0.070	0.085	0.087	0.114	0.101
2	0.094	0.067	0.132	0.129	0.199	0.191	0.322	0.267
0	0.052	0.044	0.046	0.043	0.054	0.047	0.050	0.045
3 1	0.059	0.047	0.066	0.061	0.072	0.079	0.093	0.100
2	0.065	0.064	0.109	0.102	0.140	0.123	0.255	0.235

to a distribution with more parameters (e.g., Pearson or Johnson curves) as proposed by Azzalini and Bowman (1993).

5. Proofs. Throughout this section we assume without loss of generality that the regression functions are orthonormal with respect to the design density f , that is,

$$(5.1) \quad \langle g_j, g_i \rangle := \int_0^1 g_j(t)g_i(t)f(t) dt = \delta_{ij}, \quad i, j = 1, \dots, p,$$

where δ_{ij} denotes Kronecker’s symbol.

PROOF OF LEMMA 2.1. From the Lipschitz continuity we obtain by a straightforward calculation,

$$(5.2) \quad \int_0^1 u(t)f(t) dt - \frac{1}{n} \sum_{i=1}^n u(t_i) = O(n^{-\gamma}),$$

TABLE 9

Simulated 5% level of the test (2.19) in the model (4.1) using a $a\chi_b^2 + c$ approximation for various sample sizes. The design is uniform and the error homoscedastic

σ^2	$n = 10$		$n = 20$		$n = 30$	
	Fixed	Random	Fixed	Random	Fixed	Random
1	0.051	0.053	0.048	0.049	0.048	0.045
2	0.051	0.058	0.049	0.050	0.049	0.052
3	0.048	0.060	0.055	0.057	0.051	0.051

whenever $u, f \in \text{Lip}_\gamma[0, 1]$. Therefore it follows from the orthonormality of the regression functions g_1, \dots, g_p that

$$\begin{aligned} E[\hat{\sigma}_{\text{LSE}}^2] &= \frac{1}{n} \sum_{i=1}^n (m^2(t_i) + \sigma^2(t_i)) - \sum_{l=1}^p \left(\frac{1}{n} \sum_{i=1}^n m(t_i) g_l(t_i) \right)^2 + O(n^{-\gamma}) \\ &= \int_0^1 \sigma^2(t) f(t) dt + \int_0^1 m^2(t) f(t) dt - \sum_{l=1}^p \left(\int_0^1 m(t) g_l(t) f(t) dt \right)^2 \\ &\quad + O(n^{-\gamma}) \\ &= \int_0^1 \sigma^2(t) f(t) dt + M^2 + O(n^{-\gamma}), \end{aligned}$$

where M^2 is defined in Section 2 and the last equality is obtained from a basic fact in Fourier analysis. \square

PROOF OF THEOREMS 2.3 AND 2.4. The proof can be divided into three parts. At first we derive an asymptotically equivalent statistic. In the second part of the proof we calculate the corresponding asymptotic variance and finally we prove the asymptotic normality.

(a) Recalling the definition of T_n in (2.11) we obtain

$$\begin{aligned} T_n &= \hat{\sigma}_{\text{LSE}}^2 - \hat{\sigma}_{\text{HM}}^2 \\ &= \frac{1}{n} \sum_{i=1}^n y_i^2 - \sum_{l=1}^p \left(\frac{1}{n} \sum_{i=1}^n g_l(t_i) y_i \right)^2 \\ (5.3) \quad &\quad - \frac{1}{v} \sum_{i=1}^n \left(y_i - \sum_{j=1}^n w_{ij} y_j \right)^2 + O_p\left(\frac{1}{n}\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \tilde{b}_{ij} y_i y_j + O_p\left(\frac{1}{n}\right), \end{aligned}$$

where the second equality follows from the orthonormality of the regression functions [see (5.1)] and the third equality by the definition

$$(5.4) \quad \tilde{b}_{ij} = \left\{ \frac{1}{n} \left(1 - \frac{n}{v} \right) \delta_{ij} + \frac{1}{v} \left(2w_{ij} - \sum_{k=1}^n w_{ki} w_{kj} \right) - \frac{1}{n^2} \sum_{l=1}^p g_l(t_i) g_l(t_j) \right\}$$

($i, j = 1, \dots, n$). Define $\tilde{B} = (\tilde{b}_{ij})_{i,j=1}^n$; then it is easy to see that

$$(5.5) \quad E[y^T \tilde{B} y] = E[T_n] + O\left(\frac{1}{n}\right) = M^2 - C_2 h^{2r} - \frac{C_3}{nh} + o(h^{2r}) + O\left(\frac{1}{n}\right),$$

where C_2 and C_3 are defined in (2.8) and (2.10), respectively, and the second equality is a consequence of Lemmas 2.1 and 2.2. Consequently, it is suffi-

cient to prove the assertions of Theorems 2.3 and 2.4 for the random variable $y^T \tilde{B}y - E[y^T \tilde{B}y]$. For a further simplification we note that

$$\begin{aligned} y^T \tilde{B}y - E[y^T \tilde{B}y] &= \sum_{i \neq j} \tilde{b}_{ij}(y_i y_j - m(t_i)m(t_j)) + \sum_{i=1}^n \tilde{b}_{ii}(y_i^2 - \sigma^2(t_i) - m^2(t_i)) \\ &= T_n^{(1)} + T_n^{(2)} \end{aligned}$$

and observe that $E[T_n^{(2)}] = 0$, $\text{Var}(T_n^{(2)}) = O(1/n^3 h^2)$, which implies $T_n^{(2)} = o_p((n\sqrt{h})^{-1})$. Define $B = (b_{ij})_{i,j=1}^n$ as the matrix corresponding to the quadratic form $T_n^{(1)}$, that is, $b_{ij} = \tilde{b}_{ij}(1 - \delta_{ij})$ ($i, j = 1, \dots, n$); then we obtain by (5.3) and (5.5),

$$\begin{aligned} (5.6) \quad T_n - M^2 + C_2 h^{2r} + \frac{C_3}{nh} \\ = T_n^{(1)} + o_p((n\sqrt{h})^{-1}) + o(h^{2r}) + O\left(\frac{1}{n}\right), \end{aligned}$$

where

$$(5.7) \quad T_n^{(1)} = y^T B y - \tilde{m}^T B \tilde{m} = \tilde{y}^T B \tilde{y} + \tilde{y}^T B \tilde{m} + \tilde{m}^T B \tilde{y},$$

$\tilde{m} = (m(t_1), \dots, m(t_n))^T$ and \tilde{y} is the centered vector of observations, that is, $\tilde{y} = y - \tilde{m}$.

(b) The variances and covariances of the random variables on the right-hand side of (5.7) can now be calculated by straightforward but tedious algebra. More precisely, we obtain from Whittle (1964) for the variance of the first term,

$$\begin{aligned} (5.8) \quad n^2 h \text{Var}(\tilde{y}^T B \tilde{y}) &= 2n^2 h \sum_{i, j=1}^n b_{ij}^2 \sigma^2(t_i) \sigma^2(t_j) \\ &= 2h \sum_{i=1}^n \sum_{j=1}^n \sigma^2(t_i) \sigma^2(t_j) \left\{ 2w_{ij} - \sum_{k=1}^n w_{ki} w_{kj} \right\}^2 + o(1) \\ &= 2 \int_0^1 \sigma^4(v) dv \int_{-\infty}^{\infty} [2K(u) - K * K(u)]^2 du + o(1). \end{aligned}$$

For the variance of the second term it follows that

$$\begin{aligned} R_n &= n \text{Var}(\tilde{y}^T B \tilde{m} + \tilde{m}^T B \tilde{y}) \\ &= \frac{n}{v^2} \sum_{i=1}^n \left(\sum_{j=1}^n \left[2w_{ij} - \sum_{k=1}^n w_{ki} w_{kj} - \frac{v}{n^2} \sum_{l=1}^p g_l(t_i) g_l(t_j) \right] m(t_j) \right. \\ &\quad \left. + \sum_{j=1}^n \left[2w_{ji} - \sum_{k=1}^n w_{ki} w_{kj} - \frac{v}{n^2} \sum_{l=1}^p g_l(t_i) g_l(t_j) \right] m(t_j) \right)^2 \sigma^2(t_i), \end{aligned}$$

which gives by a straightforward but tedious computation,

$$(5.9) \quad R_n = o\left(\frac{1}{nh}\right)$$

if $M^2 = 0$ and

$$(5.10) \quad R_n = 4 \int_0^1 f(t)\sigma^2(t) \left\{ m(t) - \sum_{l=1}^p \langle m, g_l \rangle g_l(t) \right\}^2 dt + o(1)$$

if $M^2 > 0$. Finally, we have

$$\text{Cov}(\tilde{y}^T B \tilde{y}, \tilde{y}^T B \tilde{m} + \tilde{m}^T B \tilde{y}) = 0$$

and are now in the position to prove Theorems 2.3 and 2.4.

(c) PROOF OF THEOREM 2.3. Under the hypothesis of linearity we have $M^2 = 0$ and from (5.6) $T_n + C_2 h^{2r} + C_3/nh = T_n^{(1)} + o_p((n\sqrt{h})^{-1})$. Therefore it is sufficient to show the assertion for $T_n^{(1)}$. To this end we note that

$$(5.11) \quad n\sqrt{h}T_n^{(1)} = n\sqrt{h}\{\tilde{y}^T B \tilde{y} + \tilde{y}^T B \tilde{m} + \tilde{m}^T B \tilde{y}\} = n\sqrt{h}\tilde{y}^T B \tilde{y} + o_p(1),$$

where the first equality in (5.11) follows from (5.7). The second equality is obtained from $E[y^T B \tilde{m} + \tilde{m}^T B \tilde{y}] = 0$ and (5.9) which implies

$$nhR_n = n^2h \text{Var}(\tilde{y}^T B \tilde{m} + \tilde{m}^T B \tilde{y}) = o(1).$$

For the first term on the right-hand side of (5.11) we have from (5.8),

$$(5.12) \quad \begin{aligned} \sigma^2(n) &= \text{Var}(n\sqrt{h}\tilde{y}^T B \tilde{y}) \\ &= 2 \int_0^1 \sigma^4(u) du \int_0^1 (2K(u) - K * K(u))^2 du + o(1) = \mu_0^2 + o(1). \end{aligned}$$

In order to establish the asymptotic normality of $n\sqrt{h}\tilde{y}^T B \tilde{y}$ we now apply Theorem 5.2 in de Jong (1987) to the quadratic form $X^T A X$ where A is the $n \times n$ matrix with elements $a_{ij} = n\sqrt{h}b_{ij}\sigma(t_i)\sigma(t_j)$ and $X_i = \tilde{y}_i/\sigma(t_i)$ ($i, j = 1, \dots, n$). A straightforward calculation shows

$$\max_{i=1}^n \sum_{j=1}^n a_{ij}^2 = n^2h \sum_{j=1}^n \sigma^2(t_i)\sigma^2(t_j)b_{ij}^2 = O\left(\frac{1}{n}\right),$$

which implies assumptions (1) and (2) in de Jong's theorem with $K(n) = \log n$. For the remaining assumption regarding the eigenvalues μ_1, \dots, μ_n of the matrix A , we apply Gerschgorin's theorem and obtain

$$\max_{i=1}^n |\mu_i| \leq \max_{i=1}^n \sum_{j=1}^n |a_{ij}| = n\sqrt{h} \max_{i=1}^n \sum_{j=1}^n \sigma(t_i)\sigma(t_j)|b_{ij}| = O(\sqrt{h}).$$

The assertion of Theorem 2.3 is now an immediate consequence of (5.11), (5.12) and de Jong's Theorem 5.2. \square

PROOF OF THEOREM 2.4. If $M^2 > 0$, it follows from (5.8) and (5.10) that the dominating term in the variance of (5.7) is of order n^{-1} , that is,

$$\begin{aligned} \sigma^2(n) &:= \text{Var}(\sqrt{n}T_n^{(1)}) = n \text{Var}(\tilde{m}^T B \tilde{y} + \tilde{y}^T B \tilde{m}) + o(1) = R_n + o(1) \\ &= 4 \int_0^1 f(t) \sigma^2(t) \left\{ m(t) - \sum_{l=1}^p \langle m, g_l \rangle g_l(t) \right\}^2 dt + o(1) = \mu_1^2 + o(1). \end{aligned}$$

Because B has vanishing diagonal elements, we have $E[\tilde{y}^T B \tilde{y}] = 0$ and as a consequence from (5.8),

$$\sqrt{n}T_n^{(1)} = \sqrt{n}(\tilde{y}^T B \tilde{m} + \tilde{m}^T B \tilde{y}) + o_p(1).$$

The term on the right-hand side is asymptotically normal with variance μ_1^2 defined in (2.15) by the central limit theorem. This proves the assertion of Theorem 2.4. \square

PROOF OF LEMMA 2.6. In order to keep the notation simple we consider the case $p = 1$; the general case follows exactly the same lines. Let $\hat{\theta}_n = (\sum_{j=1}^n g_1^2(t_j))^{-1} \sum_{j=1}^n g_1(t_j) y_j$ denote the least squares estimator for θ , then the random part in the estimator (2.18) can be rewritten as

$$\begin{aligned} \hat{R}^2 &= \sum_{i=1}^{n-1} \frac{(t_{i+1} - t_i)}{s^8} \left[\sum_{j \neq i} g_1(t_j) (\varepsilon_i g_1(t_j) - g_1(t_i) \varepsilon_j) \right]^2 \\ &\quad \times \left[\sum_{j \neq i+1} g_1(t_j) (\varepsilon_{i+1} g_1(t_j) - g_1(t_{i+1}) \varepsilon_j) \right]^2, \end{aligned}$$

where $s^2 = \sum_{j=1}^n g_1^2(t_j) = n + o(n)$, by the orthonormality assumption (5.1). A straightforward calculation yields for the expectation

$$\begin{aligned} E[\hat{R}^2] &= \sum_{i=1}^{n-1} \frac{(t_{i+1} - t_i)}{s^8} \sum_{j, j' \neq i} \sum_{k, k' \neq i+1} (g_1(t_j) g_1(t_{j'}) g_1(t_k) g_1(t_{k'}))^2 \sigma^2(t_i) \sigma^2(t_{i+1}) \\ &\quad + O\left(\frac{1}{n}\right), \\ &= \sum_{i=1}^{n-1} (t_{i+1} - t_i) \sigma^2(t_i) \sigma^2(t_{i+1}) + O\left(\frac{1}{n}\right) \\ &= \int_0^1 \sigma^4(u) du + O\left(\frac{1}{n}\right). \end{aligned}$$

A similar calculation shows that $\text{Var}(\hat{R}^2) = O(1/n)$ and we obtain $\hat{R}^2 \xrightarrow{p} \int_0^1 \sigma^4(u) du$. The assertion of Lemma 2.6 now follows from the definition of $\hat{\mu}_0^2$ in (2.18). \square

PROOFS OF THEOREMS 3.2 AND 3.3. We will only present a proof of Theorem 3.3; the result of Theorem 3.2 follows by similar arguments.

In the fixed design case, define $\varepsilon_i = y_i - m(t_i)$ ($i = 1, \dots, n$), $\Delta(t) = (m - P_{\{g_1, \dots, g_p\}}m)(t) = m(t) - \theta_0^T g(t)$ and let $\hat{\theta}_n$ denote the LSE of θ . Following Zheng (1996) we rewrite V_n as

$$(5.13) \quad V_n = V_{1n} - 2\{V_{2n}^{(1)} - V_{2n}^{(2)}\} + \{V_{3n}^{(1)} - 2V_{3n}^{(2)} + V_{3n}^{(3)}\},$$

where

$$\begin{aligned} V_{1n} &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{h} K\left(\frac{t_i - t_j}{h}\right) \varepsilon_i \varepsilon_j, \\ V_{2n}^{(1)} &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{h} K\left(\frac{t_i - t_j}{h}\right) \varepsilon_i (\hat{\theta}_n^T - \theta_0^T) g(t_j), \\ V_{2n}^{(2)} &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{h} K\left(\frac{t_i - t_j}{h}\right) \varepsilon_i \Delta(t_j), \\ V_{3n}^{(1)} &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{h} K\left(\frac{t_i - t_j}{h}\right) (\hat{\theta}_n - \theta_0)^T g(t_i) (\hat{\theta}_n - \theta_0)^T g(t_j), \\ V_{3n}^{(2)} &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{h} K\left(\frac{t_i - t_j}{h}\right) \Delta(t_i) (\hat{\theta}_n - \theta_0)^T g(t_j), \\ V_{3n}^{(3)} &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{h} K\left(\frac{t_i - t_j}{h}\right) \Delta(t_i) \Delta(t_j). \end{aligned}$$

From Zheng (1996) we have

$$(5.14) \quad \begin{aligned} V_{1n} &= O_p((n\sqrt{h})^{-1}); & V_{2n}^{(1)} &= o_p((n\sqrt{h})^{-1}); \\ V_{3n}^{(1)} &= o_p((n\sqrt{h})^{-1}) \end{aligned}$$

and for the term $V_{3n}^{(3)}$ it follows

$$(5.15) \quad \begin{aligned} V_{3n}^{(3)} &= \int_0^1 \int_0^1 K(u) (\Delta f)(v) (\Delta f)(v - uh) du dv \\ &\quad - \frac{K(0)}{nh} \int_0^1 \Delta^2(u) f(u) du + O\left(\frac{1}{n}\right). \end{aligned}$$

The remaining terms can be treated by the central limit theorem; from the orthonormality of the regression function we have

$$\hat{\theta}_n - \theta = \left(\frac{1}{n} \sum_{i=1}^n g_l(t_i) \varepsilon_i \right)_{l=1}^p + O_p\left(\frac{1}{n}\right),$$

which implies

$$\begin{aligned}
 & \sqrt{n}(V_{3n}^{(2)} - V_{2n}^{(2)}) \\
 (5.16) \quad &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i \left\{ \sum_{l=1}^p \frac{g_l(t_i)}{(n-1)n} \sum_j \sum_{k \neq j} \frac{1}{h} K\left(\frac{t_k - t_j}{h}\right) \Delta(t_k) g_l(t_j) \right. \\
 & \quad \left. - \frac{1}{n-1} \sum_{j \neq i} \frac{1}{h} K\left(\frac{t_i - t_j}{h}\right) \Delta(t_j) \right\} + o_p(1) \\
 & \rightarrow_{\mathcal{D}} \mathcal{N}(0, \mu_3^2),
 \end{aligned}$$

where

$$\begin{aligned}
 \mu_3^2 &= \lim_{\substack{n \rightarrow \infty \\ h \rightarrow 0}} \frac{1}{n} \sum_{i=1}^n \sigma^2(t_i) \left\{ \sum_{l=1}^p \frac{g_l(t_i)}{(n-1)n} \sum_j \sum_{k \neq j} \frac{1}{h} K\left(\frac{t_k - t_j}{h}\right) \Delta(t_k) g_l(t_j) \right. \\
 & \quad \left. - \frac{1}{n-1} \sum_{j \neq i} \frac{1}{h} K\left(\frac{t_i - t_j}{h}\right) \Delta(t_j) \right\}^2 \\
 &= \int_0^1 \sigma^2(u) f(u) \left\{ (\Delta f)(u) - \sum_{l=1}^p \langle \Delta f, g_l \rangle g_l(u) \right\}^2 du \\
 &= \int_0^1 \sigma^2(u) f(u) \left\{ (\Delta f)(u) - P_{\{g_1, \dots, g_p\}}(\Delta f)(u) \right\}^2 du.
 \end{aligned}$$

The assertion for the fixed design case now follows, combining (5.13)–(5.16).

The remaining part for the random design is obtained by the same arguments, observing that in this case the contributing terms in the asymptotic variance are the uncorrelated random variables $V_{2n}^{(2)} - V_{3n}^{(3)}$ and $V_{3n}^{(3)}$. Under the random design assumption the asymptotic variance of $\sqrt{n}(V_{2n}^{(2)} - V_{3n}^{(2)})$ is also given by μ_3^2 defined in (5.16) while the random variable $V_{3n}^{(3)}$ yields

$$\begin{aligned}
 n \text{Var}(V_{3n}^{(3)}) &= \frac{1}{n(n-1)^2} E \left[\sum_{\substack{i \neq j \\ i' \neq j'}} \frac{1}{h^2} K\left(\frac{U_i - U_j}{h}\right) K\left(\frac{U_{i'} - U_{j'}}{h}\right) \right. \\
 & \quad \left. \times \Delta(U_i) \Delta(U_j) \Delta(U_{i'}) \Delta(U_{j'}) \right] \\
 & \quad - \left(E \left[\frac{1}{h} K\left(\frac{U_1 - U_2}{h}\right) \Delta(U_1) \Delta(U_2) \right] \right)^2 \\
 &= \frac{4(n-2)}{n-1} E \left[\frac{1}{h^2} K\left(\frac{U_1 - U_2}{h}\right) K\left(\frac{U_1 - U_3}{h}\right) \Delta^2(U_1) \Delta(U_2) \Delta(U_3) \right] \\
 & \quad - 4 \left(E \left[\frac{1}{h} K\left(\frac{U_1 - U_2}{h}\right) \Delta(U_1) \Delta(U_2) \right] \right)^2 + o(1) \\
 &= 4 \text{Var}[\Delta^2(U_1) f(U_1)] + o(1),
 \end{aligned}$$

where U_1, \dots, U_n are i.i.d. random variables with density f . This proves the assertion in the case of a random design. \square

Acknowledgments. The author is grateful to the referees and Associate Editor for very constructive comments on an earlier version of this paper. The author also thanks N. Neumeyer and A. Munk for several interesting comments and I. Gottschlich who typed numerous versions of this manuscript with considerable technical expertise.

REFERENCES

- AZZALINI, A. and BOWMAN, A. (1993). On the use of nonparametric regression for checking linear relationships. *J. Roy. Statist. Soc. Ser. B* **55** 549–559.
- BERGER, J. O. and DELAMPADE, M. (1987). Testing precise hypotheses. *Stat. Sci.* **2** 317–352.
- BRODEAU, F. (1993). Tests for the choice of approximative models in nonlinear regression when the variance is unknown. *Statistics* **24** 95–106.
- DE JONG, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields* **75** 261–277.
- DETTE, H. and MUNK, A. (1998). Validation of linear regression models. *Ann. Statist.* **26** 778–800.
- DETTE, H., MUNK, A. and WAGNER, T. (1998). A review of variance estimators with applications to multivariate nonparametric regression models. In *Multivariate Analysis, Design of Experiments and Survey Sampling* (S. Ghosh, ed.) 469–498. Dekker, New York.
- ELFVING, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist.* **23** 255–262.
- EUBANK, R. L. and HART, J. D. (1992). Testing goodness of fit in regression via order selection criteria. *Ann. Statist.* **20** 1412–1425.
- FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Soc.* **87** 998–1004.
- FAN, J. and GLJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.
- GALLANT, A. R. (1987). *Nonlinear Statistical Models*. Wiley, New York.
- GASSER, T. and MÜLLER, H.-G. (1979). Kernel estimation of regression functions. *Smoothing Techniques for Curve Estimation. Lecture Notes in Math.* **757** 23–68. Springer, New York.
- GASSER, T., MÜLLER, H.-G. and MAMITZSCH, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B* **47** 238–252.
- GASSER, T., SKROKA, L. and JENNEN-STEINMETZ, G. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73** 626–633.
- HALL, P. and MARRON, J. S. (1990). On variance estimation in nonparametric regression. *Biometrika* **77** 415–419.
- HÄRDLE, W. and MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947.
- KUTCHIBHATLA, M. and HART, J. D. (1996). Smoothing-based lack-of-fit tests: variations on a theme. *Nonparametr. Statist.* **7** 1–22.
- NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **10** 186–190.
- NEIL, J. W. and JOHNSON, D. E. (1985). Testing linear regression function adequacy without replication. *Ann. Statist.* **13** 1482–1489.
- RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215–1230.
- SACKS, J. and YLVIKAKER, D. (1970). Designs for regression problems for correlated errors. *Ann. Math. Statist.* **41** 2057–2074.
- SEBER, G. A. F. and WILD, G. J. (1989). *Nonlinear Regression*. Wiley, New York.
- SHILLINGTON, E. R. (1979). Testing lack of fit in regression without replication. *Canad. J. Statist.* **7** 137–146.
- STUTE, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613–641.
- STUTE, W., GONZÁLES MANTEIGA, W. and PRESEDO QUINDIMIL, M. (1998). Bootstrap approximation in model checks for regression. *J. Amer. Statist. Assoc.* **93** 141–149.

- WAND, M. P. and JONES, M. C. (1995). *Kernel Smoothing*. Chapman and Hall, London.
- WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372.
- WEIRATHER, G. (1993). Testing a linear regression model against nonparametric alternatives. *Metrika* **40** 367–379.
- WHITTLE, P. (1964). On the convergence to normality of quadratic forms in independent variables. *Theory. Probab. Appl.* **9** 103–108.
- WOOLDRIDGE, J. M. (1992). A test for a functional form against nonparametric alternatives. *Econometric Theory* **8** 452–475.
- YATCHEW, A. J. (1992). Nonparametric regression tests based on least squares. *Econometric Theory* **8** 435–451.
- ZHENG, J. X. (1996). A consistent test of a functional form via nonparametric estimation techniques. *J. Econometrics* **75** 263–289.

FAKULTÄT FÜR MATHEMATIK
RUHR-UNIVERSITÄT BOCHUM
44780 BOCHUM
GERMANY
E-MAIL: holger.dette@ruhr-uni-bochum.de