# Some Methodological Aspects of Validation of Models in Nonparametric Regression

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: holger.dette@ruhr-uni-bochum.de
FAX: +49 234 3214-559

Axel Munk
Universität Paderborn
Fakultät für Mathematik
33095 Paderborn
Germany
email: munk@math.upb.de
FAX: +49 5251 60-3737

February 11, 2002

**Abstract**

In this paper we describe some general methods for constructing goodness of fit tests in nonparametric regression models. Our main concern is the development of statisticial methodology for the assessment (validation) of specific parametric models $\mathcal{M}$ as they arise in various fields of applications.

The fundamental idea which underlies all these methods is the investigation of certain goodness of fit statistics (which may depend on the particular problem and may driven by different criteria) under the assumption that a specified model (which has to be validated) holds true as well as under a broad range of scenaria, where this assumption is violated. This is motivated by the fact that outcomes of tests for the classical hypothesis: "The model $\mathcal{M}$ holds true" (and their associated $p$ values) bear various methodological flaws. Hence, our suggestion is, always to accompany such a test by an analysis of the type II error, which is in goodness of fit problems often the more serious one.

We will give a careful description of the methodological aspects, the required asymptotic theory, and illustrate the main principles in the problem of testing model assumptions such as a specific parametric form or homoscedasticity in nonparametric regression models.

*AMS Subject Classification: Primary 62G05, Secondary 62G10, 62G30, 62G07*

*Keywords: Model check, validation of goodness of fit, $\mathcal{L}^2$-distance, neighborhood hypotheses, p-values, nonparametric regression, test of homoscedasticity.*

# 1 Introduction

In this chapter some methodological concepts of goodness of fit tests will be discussed in detail. We will introduce the concept of neighborhood hypotheses, a method which provides evidence in favour of a model, instead of the abscence of evidence against a certain model (as it is provided by classical goodness of fit tests). Our approach requires knowledge of the (asymptotic) distribution of test statistics at all points in the (possibly infinite dimensional) model space and therefore differs from the usual model checks proposed in the literature. In the subsequent chapters this methodolgy is applied and worked out in detail for various settings in the nonparametric regression model.

## 1.1 What are the right hypotheses?

Without any doubt, one of the main problems of (asymptotic) statistical inference consists in the assessment (at a controlled type I error rate $\alpha$) of a significant difference between several populations. The null hypothesis $H$ of equality is rejected only when the data strongly indicate a departure from equality. This is justified in many applications, namely when the error of falsely rejecting equality is considered as the more serious error (type I) in contrast to accepting $H$ although being false (type II). During the past, it has been recognized in various fields, that there are, however, problems of inference where the type II error (with respect to the classical null hypothesis) is considered as more serious. This has been recognized nowadays in various fields where inference statistics is applied, including medical and pharmaceutical statistics (Altmann and Bland (1995), Chow and Liu (1992), Munk and Czado (1998)), chemistry (Roy, 1997) or in environmental statistics (McBride, 1998).
In the following we would like to show that this holds true particularly for all kinds of goodness of fit problems, where it has to be checked whether a specific statistical model $\mathcal{M}$ describes the data structure appropriately. Although this has been noticed by various authors (at least implicitely) at several places during the past, this has never been systematically taken into account in the development of goodness of fit tests.
Curiously, Fisher's celebrated $\chi^2$-goodness of fit test for testing equality of the cell frequencies from a multinomial sample has already initiated an extensive discussion regarding the interpretation of a test result as a measure of evidence in favour or against the null hypothesis; i.e. the validity of the model of equal cell frequencies. Berkson (1938, 1942, 1943) gave several arguments in a series of papers against Fisher's test (and the use of its associated $P$-value) which may be summarized as follows:

1. Any consistent test will detect any arbitrary small departure from the null model $\mathcal{M}$ with increasing sample size because in the physical world the hypothesis $H : \mathcal{M}$ is never exactly valid (p. 526, 1938).

2. In principle any data set indicating evidence against $\mathcal{M}$ could occur although $\mathcal{M}$ is correct, hence the "believer" in the alternative should determine the decision (p. 531, 1938) "My view is that there is never any valid reason for rejection of the null hypothesis except on the willingness to embrace an alternative one".

3. "What characteristic does the test $\chi^2$-variable represent?" More general, which quantity does the underlying test statistic summarize (p. 536, 1938)?

4. Combination of experiments may lead to a substantial different conclusion than the single experiments (p. 536, 1938).

5. Deviation from the model assumption in another direction as one expects may lead to the rejection of the null model. In order to illustrate this, Berkson discusses an example where the variability of the covariable seems to be responsible for the departure from a straight line regression (p. 329, 1942).

6. Only because the sample size is too small the null model is not rejected (p. 332, 1942).

Summarizing his objections from a practitioners view point (p.242, 1943) Berkson wrote:

> "Time and time again I came upon actual situations in which (1) it was patent to all competent people who viewed the data, that the null hypothesis, on the evidence, *was not* true but a small $P$ did not result, (2) a small $P$ did result, but it was obvious to all who examined the data, that on the evidence the null hypothesis was tested by several tests, each put forward with equally good authority, the $P$'s were considerably *different*."

In 1954 Hodges and Lehmann have rediscovered Berkson's arguments and suggested what we will denote as "(precise) neighborhood hypotheses" in the following. These hypotheses require, in accordance with argument 3. above, always a specification of a measure of discrepancy between the model $\mathcal{M}$ and the "true" unknown model $\mathcal{M}_0$. More specifically, Hodges and Lehmann (1954) considered the quantity

$$\Delta(p, p') := \sum_{i=1}^{r} \omega_i (p'_i - p_i)^2 \tag{1.1}$$

(here the $\omega_i$ are fixed weights) where $p = (p_1, \dots, p_r)$ and $p' = (p'_1, \dots, p'_r)$ denote the hypothetical and the true vector of cell frequencies, respectively. Consequently, they suggested the hypothesis

$$\inf_{p' \in S} \Delta(p, p') \leq c$$

where $S$ denotes the "model surface" to be investigated. This leads to a $\chi^2$-test statistic which is *different* from the usual $\chi^2$-statistic, because the LS estimator has to be found under the constraint

$$\Delta(p, p') = c,$$

and hence depends on $c$. Operationally, this results in a much more complicated test which might be one reason why this "neighborhood"-methodology had no significant impact on practical

statistics over a long period (at least to our knowledge). This is highlighted by other testing problems which require even more complicated procedures compared to the classical null as demonstrated in Hodges and Lehmann (1954) by the construction of an unbiased test for the problem of testing interval hypotheses for the mean of a normal random variables with unknown variance (cf. Brown, Hwang and Munk (1997) or Munk (2001a) for the converse hypotheses).
The problems mentioned above encountered with classical nulls were pointed out again by several authors during the last years in various areas where goodness of fit problems occur. For example McKinnon (1992) discussed again objection 1 of Berkson for the particular case of checking specific regression models with independent observations as well as under specific dependency structures, as they occur in time series analysis. Recently Goutis and Robert (1998) suggested neighborhood hypotheses with respect to the Kullback-Leibler distance for model selection in generalized linear models. Nevertheless, goodness of fit problems were treated throughout the very most part of the literature as testing the null hypothesis, that the null model is true. In this manuscript we follow the idea of neighborhoods around the classical nulls and their associated tests.

## 1.2 Neigborhood hypotheses in regression.

As mentioned before our approach relies on the specification of a proper distance from the true model $\mathcal{M}_0$ to the model space $\mathcal{M}$, in accordance with Hodges and Lehmann's approach. In what follows, we deal only with $\mathcal{L}^2$-distances, which has mainly mathematical reasons, motivated by the inner product structure of $\mathcal{L}^2$. For example, in problems concerning the signal $m$ (which corresponds to $\mathcal{M}_0$) in a nonparametric regression setting we consider throughout the following the minimal distance (or a variation of it)

$$M^2(\mathcal{M}_0, \mathcal{M}) \;=\; M^2(m) := \min_{\theta\in\Theta} \|m - f_\theta\|^2$$

as a measure of discrepancy between $m$ and the parametric model $\mathcal{M} = \{f_\theta\}_{\theta\in\Theta}$. The hypotheses to be tested depend now sensitively on the sort of data analysis to be performed. For example, when one would like to consider a priori $\mathcal{M}_0$ as very likely

$$H_\pi : M(m) \le \pi \quad \text{versus} \quad K_\pi : M(m) > \pi \tag{1.2}$$

seems to be a proper formulation of the problem. Here $\pi$ denotes the bound in which the experimenter would denote $\mathcal{M}$ as "scientifically not relevant" different from $\mathcal{M}_0$. Observe, that this could even include $\pi = 0$, and hence the classical null $m \in \mathcal{M}$. Often, we are, however, concerned with the converse problem $K_\pi$ versus $H_\pi$, i.e. we would like to "validate" the model $\mathcal{M}$ at a controlled error rate $\alpha$ within the bound $\pi$. For example, if prediction is the main concern of the subsequent data analysis a misspecification could have drastic consequences. In contrast, if we commit a type I error of the classical null hypothesis, $\mathcal{M}$ holds true, this will only result in a less efficient data analysis, in general (e.g. when a paramtric $\sqrt{n}$-consistent estimator based on $\mathcal{M}$ will be replaced by a nonparemtric curve estimator).
One of the most difficult tasks with neighborhood hypotheses in any practical application is certainly the specification of the bound $\pi$, of the distance $M$ and the specification of the level

$\alpha$ (why should one use in model testing 0.05?) Therefore, it might be appealing to use the classical hypotheses because here these difficulties do not occur. For example, the bound $\pi$ is canonically given as $\pi = 0$. We mention, however, that this does not mean that classical null hypotheses (and their tests) do avoid these or similar problems, they just cover them (cf. again Berkson's argument). One way out of these difficulties could be to consider model selection more as an explorative data analysis (EDA) sort of issue rather than a testing problem, which also to our believe is most appropriate in many cases.
However, in many practical examples goodness of fit tests are performed and have their own merits. Hence, our claim is, that in these cases any goodnes of fit test should be accompanied by a proper analysis of level and power of its corresponding neighbourhood hypotheses. As we will see this includes as a byproduct also the computation of confidence intervals for the discrepancy $M$, which should always be performed additionally for a valid assessment of a model specification. If such a confidence interval has to be computed for a measure of discrepancy, such as $M$, the asymptotic law under *all* alternatives, $M > 0$, is required and it is not sufficient to derive a limit law under the classical null. Hence, computationally confidence intervals, and tests for neighborhood hypotheses are not different.
In the remaining part of the paper we illustrate a general approach to deal with the questions raised in the last sections. Note that the answer to any of the problems indicated here requires the (asymptotic) distribution of a test statistic at any fixed point in the space $\mathcal{M}$. As pointed out in Section 1.1 this usually results in a much more complicated analysis, which may be one reason why these concepts have not found too much attention in practical statistics. However, the methods proposed in this paper usually yield test statistics which are (asymptotically) normal in all situations and consequently the analysis presented in the last sections is directly applicable.

## 1.3 Two goodness of fit testing problems in regression models

Throughout this paper we consider the nonparametric regression model

$$Y = Y(t) = m(t) + \sigma(t)\varepsilon, \tag{1.3}$$

where $m$ is an (unknown) regression function, $\varepsilon$ is a random error, $\sigma^2$ the (unknown) variance function and $t$ is the predictor. For the sake of brevity we restrict ourselves to a one-dimensional fixed predictor satisfying $t \in [0, 1]$. Some extensions are briefly mentioned in Section 2.7. Among the many possible hypotheses about the regression model we illustrate our approach in the problem of testing for a parametric form and in the problem of testing for homoscedasticity. Parametric regression models are attractive among practioners 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. For these reasons much effort has been devoted to the problem of testing for a parametric form of the regression function. Early work dates back to the papers of Neyman (1937) and von Neumann (1941, 1942) while more recent work can be found in the papers of Bunke and Schmidt (1980), Bunke (1981), Neil and Johnson (1985), Cox, Whaba, Yandell

(1988), Eubank and Spiegelmann (1990), Bierens (1990), Firth, Glosup and Hinkeley (1991), Staniswalis and Severini (1991), Kozek (1991), Wooldridge (1992), Yatchew (1992), Müller (1992), McKinnon (1992), Hart and Wehrly (1992), Eubank and Hart (1992), Azzalini and Bowman (1993), Brodeau (1993), Härdle and Mammen (1993), Weirather (1993), Gonzalez-Manteiga, Cao-Abad (1993), Eubank, Hart , La Riccia (1993), Gozalo (1993), Chen (1994), Djojosugito (1994, 1995), Dieboldt (1995), Gonzalez-Manteiga and Vilar-Fernandez (1995), Fan (1996), Kuchibhatta and Hart (1996), Jayashuriya (1996), Stute and González-Manteiga (1996). . For a nice but not complete review of these results we refer to the monograph of Hart (1997). A rather comprehensive summary is also given by Munk (2001b). More recent references are Stute (1997), Zheng (1997), Dette and Munk (1998a,b), Stute, Gonzalez-Manteiga and Presedo Quindimil (1998), Alcalá, Christóbal, Gonzalez-Manteiga (1999), Dette (1999, 2001), Dette, Munk and Wagner (2000), Biedermann and Dette (2000), Fan and Huang (2001), Fan and Zhang (2001), Horrowitz and Spokoiny (2001). The various publications mainly differ with respect to the techniques, which are applied to deal with the problem of testing for a parametric form of the regression (such as the use of empirical processes, order selection criteria or nonparametric curve estimators). However, most of the results of these papers are only applicable for the problem of testing the classical null hypothesis. $H_0 : m \in \mathcal{M}$ and therefore do not give answers to the questions raised in the introduction. In the following chapter we will review some concepts introduced by Dette and Munk (1998a) which allow to deal with the issues raised in Section 1.2 - 1.3. We present two interpretations of new classes of test statistics which provide the opportunity to generalize our method in various directions. In the following chapter these results are also applied for constructing tests of homoscedasticity. An application of our approach in astrophysics can be found in Bissantz & Munk (2001).
To be concise, we consider for the moment the problem of testing for a linear regression model. Assume that

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

is a given family of functions, where $\Theta \subset \mathbb{R}^d$ is a proper parameter set and $g = (g_1, \ldots, g_d)^T$ are given linear independent regression functions. The classical hypothesis of a linear model is

$$H_0 : m \in \mathcal{M} \ . \tag{1.4}$$

The simplest case appears with $d = 1$ and a constant function $g_1(t) \equiv 1$. Here the question is whether the predictior $t$ has an impact on the response $Y$ or not. Similary, the popular linear regression model

$$m(t) = a + bt, \quad t \in \mathbb{R}$$

can be obtained by chosing $d = 2$ $g_1(t) = 1$ and $g_2(t) = t$. Here we are interested in the problem of checking if the mean of the response is a linear function of the predictor.

Similary to the problem of testing for a parametric form, the problem of detecting heteroscedasticity in the model (1.3) is widely recognized because, if the requirement of homoscedasticity is not met, the loss of efficiency in using procedures for homoscedastic models may be substantial. Checks on the assumption of homoscedasticity are usually based on visual examination of residual plots after a parametric or nonparametric fit. For a precise formulation of the testing problem recall that

$$\sigma^2(t) = \mathrm{Var}(Y(t))$$

is the variance of the response $Y = Y(t)$ at the point $t$, then the classical hypothesis of homoscedasticity can be expressed as

$$H_0 : \sigma^2(t) = \sigma^2 \quad \text{for all } t.$$

The remaining part of this paper is organized as follows. Chapter 2 deals with the testing problem (1.4) , while Section 3 is devoted to the problem of checking homoscedasticity. In these examples we illustrate a general method for the construction of tests and give two interpretations of the corresponding test statistics, which might be appealing from a theoretical and a practical point of view. Our first interpretation is based on classical Hilbert space theory and we estimate a minimal $\mathcal{L}^2$-distance between the true model and the model space induced by the hypothesis. Our second interpretation shows that this generalizes the classical concept of analysis of variance to the nonparametric setting, because the test statistic can be interpreted as the difference between a variance estimator under the null hypothesis and a variance estimator under the full model. We study the asymptotic properties of the corresponding test statistic under the null hypothesis and the alternative (the latter study is important because of the discussion of the previous paragraphs) and derive asymptotic normality in all cases.

# 2 Testing for a parametric form of the regression

Assume that we observe responses of the form

$$Y_{j,n} = Y(t_{j,n}) = m(t_{j,n}) + \sigma(t_{j,n})\varepsilon_{j,n} \quad j = 1, \ldots, n$$

where $t_{1,n}, \ldots, t_{n,n} \in [0, 1]$ 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 rowwise independent random variables with mean zero and variance equal to one, that is $\sigma^2(t_{j,n}) = \mathrm{Var}[Y_{j,n}]$ $(j = 1, \ldots, n)$. We call the function $m$ and $\sigma^2$ regression and variance function, respectively, and assume further that the fourth moments of the errors are uniformly bounded, i.e.

$$E[\varepsilon_{j,n}^4] \le C < \infty, \quad j = 1, \ldots, n;\ n \in \mathbb{N}. \tag{2.1}$$

The index $n$ is omitted whenever this dependence will be clear from the context. Assume that the design points $t_1, \ldots, t_n$ satisfy

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

for a positive design density $f$, which is Lipschitz continuous of order $\gamma > 0$ (see Sacks and Ylvisaker (1970)), that is

$$|f(x) - f(y)| \ \le\ c|x - y|^\gamma$$

for some positive constant $c$.

## 2.1 Estimating a minimal $\mathcal{L}^2$-distance.

Recall that the hypothesis of a linear model (1.4) is completely specified by the given regression functions $g_1, \ldots, g_d$ and a natural measure of goodness of fit is the minimal $\mathcal{L}^2$-distance

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

between the unknown regression function and the class $\mathcal{M}$ of parametric models. Note that this distance is defined by the inner product

$$\langle p, q \rangle := \int_0^1 p(t)q(t)f(t)dt \tag{2.4}$$

and $\|p\| = \langle p, p \rangle^{1/2}$ denotes the corresponding norm. A standard argument from Hilbert space theory [see Achieser (1956), p.16 ] now shows that the minimum distance can be expressed as a ratio of two Gramian determinants

$$M^2 \;=\; \frac{\Gamma(m, g_1, \ldots, g_d)}{\Gamma(g_1, \ldots, g_d)} \tag{2.5}$$

where

$$\Gamma(p_1, \ldots, p_d) = \det(\langle p_i, p_j \rangle)_{i,j=1}^d \;.$$

**Example 2.1.** In the case $d = 1, g_1(t) \equiv 1$ discussed in Section 1.3 the situation is more transparent. Here we are interested in the problem of testing the hypothesis of a constant regression function, i.e. $H_0 : m(t) \equiv c$ for some constant $c \in \mathbb{R}$ and we have

$$\Gamma(g_1) = \langle 1, 1 \rangle = \int_0^1 f(t)dt = 1 \;.$$

Consequently

$$\begin{aligned} M^2 \;&=\; \Gamma(m, g_1) = \begin{vmatrix} \langle m, m \rangle & \langle m, g_1 \rangle \\ \langle m, g_1 \rangle & \langle g_1, g_1 \rangle \end{vmatrix} \\ &=\; \int_0^1 m^2(t)f(t)dt - (\int_0^1 m(t)f(t)dt)^2 \;=\; \mathrm{Var}(m(U)), \end{aligned}$$

where $U$ denotes a random variable with density $f$. Note that $M^2$ vanishes if and only if $m$ is constant, that is the hypothesis of constant mean is valid. Observing the approximations

$$\begin{aligned} E\Big[\frac{1}{n}\sum_{i=1}^n Y_i\Big] \;&=\; \frac{1}{n}\sum_{i=1}^n m(t_i) \approx \int_0^1 m(t)f(t)dt \\ E\Big[\frac{1}{n-1}\sum_{i=2}^n Y_i Y_{i-1}\Big] \;&=\; \frac{1}{n-1}\sum_{i=1}^n m(t_i)m(t_{i-1}) \approx \frac{1}{n}\sum_{i=1}^n m^2(t_i) \;\approx\; \int_0^1 m^2(t)f(t)dt \end{aligned}$$

it is reasonable to estimate $M^2$ by

$$\hat{M}_n^2 = \frac{1}{n-1}\sum_{i=2}^{n} Y_i Y_{i-1} - \Big(\frac{1}{n}\sum_{i=1}^{n} Y_i\Big)^2 \tag{2.6}$$

and to reject the hypothesis of a constant trend for large values of the statistic $\hat{M}_n^2$.

Note that in general we proceed exactly in the same way. The design density $f$ is not known in general, and consequently all terms in the Gramian determinants in (2.5) have to be estimated from the data. To this end the inner products $\langle m, m\rangle$, $\langle g_i, m\rangle$ and $\langle g_i, g_j\rangle$ in these determinants are replaced by their empirical counterparts

$$\begin{aligned} \hat{A}_0 &= \frac{1}{n-1}\sum_{j=2}^{n} Y_j Y_{j-1} \\ \hat{A}_i &= \frac{1}{n}\sum_{j=1}^{n} g_i(t_j) Y_j \quad (i = 1, \ldots, d) \\ \hat{B}_{p,q} &= \frac{1}{n}\sum_{j=1}^{n} g_p(t_j) g_q(t_j) \quad (p, q = 1, \ldots, d)\ , \end{aligned} \tag{2.7}$$

where the dependency on the sample size $n$ is omitted for the sake of simplicity. Recall that the quantities $\hat{B}_{p,q}$ are not random, but are required for the estimation of the inner products $\langle g_p, g_q\rangle$, because the design density $f$ is not assumed to be known. The resulting estimator of the minimal $\mathcal{L}^2$-distance is finally given by

$$\hat{M}_n^2 = \frac{\begin{vmatrix} \hat{A}_0 & \hat{A}_1 & \cdots & \hat{A}_d \\ \hat{A}_1 & \hat{B}_{1,1} & \ldots & \hat{B}_{1,d} \\ \vdots & \vdots & & \vdots \\ \hat{A}_d & \hat{B}_{d,1} & \cdots & \hat{B}_{d,d} \end{vmatrix}}{\begin{vmatrix} \hat{B}_{1,1} & \cdots & \hat{B}_{1,d} \\ \vdots & & \vdots \\ \hat{B}_{d,1} & \cdots & \hat{B}_{d,d} \end{vmatrix}}\ . \tag{2.8}$$

The following result shows that $\hat{M}_n^2$ consistently estimates the minimal $\mathcal{L}^2$ distance $M^2$ defined in (2.3) [or equivalently in (2.5)] and specifies its asymptotic distribution. For a proof of the following Theorem 2.2 we refer to Dette, Munk (1998a) or Dette, Munk and Wagner (2000).

**Theorem 2.2.** *If the regression functions $m, g_1, \ldots, g_d$ and the variance function $\sigma^2$ are Lipschitz continuous of order $\gamma > 1/2$, then as $n \to \infty$*

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

*where the asymptotic variance is given by*

$$\lambda^2 = \int_0^1 \sigma^4(t) f(t) dt + 4 \int_0^1 \sigma^2(t) \{(m - P_{\mathcal{U}_d} m)(t)\}^2 f(t) dt \tag{2.9}$$

*and $P_{\mathcal{U}_d}$ denotes the orthogonal projection onto the space*

$$\mathcal{U}_d \ = \ \text{spann}\{g_1, \ldots, g_d\} \tag{2.10}$$

*with resepct to the inner product $\langle \cdot, \cdot \rangle$.*

**Example 2.3.** In the situation of Example 2.1 [$d = 1, g_1(t) \equiv 1$] the main idea of the proof can easily be illustrated. Note that in this case the statistic $\hat{M}_n^2$ defined in (2.6) [or more generally in (2.8)] can be rewritten as

$$\hat{M}_n^2 = f\Big(\frac{1}{n-1} \sum_{i=2}^n Y_i Y_{i-1}, \frac{1}{n} \sum_{i=1}^n Y_i\Big),$$

where the function $f : \mathbb{R}^2 \to \mathbb{R}$ is given by $f(x, y) = x - y^2$. Similary, we have from Example 2.1

$$M^2 = f\Big(\int_0^1 m^2(t) f(t) dt, \int_0^1 m(t) f(t) dt\Big)$$

and by the central limit theorem for $m$-dependent random variables [see Brockwell and Davis (1991)]

$$\sqrt{n} \begin{pmatrix} \frac{1}{n-1} \sum_{i=2}^n Y_i Y_{i-1} & - & \int_0^1 m^2(t) f(t) dt \\ \frac{1}{n} \sum_{i=1}^n Y_i & - & \int_0^1 m(t) f(t) dt \end{pmatrix} \stackrel{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \Sigma) \ ,$$

where the asymptotic covariance matrix is given by

$$\Sigma = \begin{pmatrix} \int_0^1 \sigma^4(t) f(t) dt + 4 \int_0^1 \sigma^2(t) m^2(t) f(t) dt & \int_0^1 \sigma^2(t) m(t) f(t) dt \\ \int_0^1 \sigma^2(t) m(t) f(t) dt & \int_0^1 \sigma^2(t) f(t) dt \end{pmatrix} .$$

Now the $\delta$-method shows asymptotic normality of

$$\sqrt{n}(\hat{M}_n^2 - M^2) \ = \ \sqrt{n} \Big\{ f\Big(\frac{1}{n-1} \sum_{i=1}^n Y_i Y_{i-1}, \frac{1}{n} \sum_{i=1}^n Y_i\Big) \ - \ f\Big(\int_0^1 m^2(t) f(t) dt, \int_0^1 m(t) f(t) dt\Big) \Big\}$$

with asymptotic variance

$$\lambda_1^2 \ = \ \int_0^1 \sigma^4(t) f(t) dt + 4 \int_0^1 \sigma^2(t) \Big(m - P_{\mathcal{U}_1} m\Big)^2 (t) f(t) dt.$$

It is worthwhile to mention that the proof of Theorem 2.2 in the general case follows exactly the same arguments, where the non trivial part consists in the calculation of the asymptotic variance [see Dette, Munk and Wagner (2000) for more details].

## 2.2 A re-interpretation of the test statistic, generalizations and the principle of ANOVA

In this section we present a different interpretation of the statistic $\hat{M}_n^2$, which yields interesting generalizations of the approach presented in Section 2.1.

- Recall the definition of the term $\hat{A}_0$ in (2.7). A simple calculation shows that this random variable can be rewritten as

  $$\hat{A}_0 = \hat{T} - \hat{\sigma}_1^2,$$

  where

  $$\hat{T} = \frac{1}{2(n-1)} \sum_{i=2}^{n} (Y_i^2 + Y_{i-1}^2),$$

  and

  $$\hat{\sigma}_1^2 = \frac{1}{2(n-1)} \sum_{i=2}^{n} (Y_i - Y_{i-1})^2 \tag{2.11}$$

  denotes a variance estimator in the nonparametric regression model under homoscedasticity, which was introduced by Rice (1984). Under heteroscedasticity $\hat{\sigma}_1^2$ consistently estimates the integrated variance function

  $$\int_0^1 \sigma^2(t) f(t) dt$$

  while $\hat{T}$ is a consistent estimator of the sum

  $$\int_0^1 m^2(t) f(t) dt + \int_0^1 \sigma^2(t) f(t) dt$$

  and the difference is used to estimate the square $\mathcal{L}^2$-norm $\langle m, m \rangle$.

- Recall the definition of the test statistic $\hat{M}_n^2$ in (2.6). A simple calculation shows that $\hat{M}_n^2$ is asymptotically equivalent to the statistic

  $$\tilde{M}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y}_n)^2 - \frac{1}{2(n-1)} \sum_{i=2}^{n} (Y_i - Y_{i-1})^2,$$

  where $\bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$ denotes the sample mean. Note that the first term

  $$\hat{\sigma}_{LSE}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y}_n)^2$$

is the least squares estimator of the variance under the null hypothesis of no effect $H_0 : m(t) \equiv c$ (for some $c \in \mathbb{R}$) in a homoscedastic regression model (1.3). On the other hand it was pointed out by Rice (1984) that the estimator defined in (2.11) is a variance estimator in the general nonparametric (homoscedastic) regression model. In the case of heteroscedasticity both estimators essentially estimate the integrated variance function. This can be easily seen heuristically by looking at the expectations

$$\begin{aligned} E[\hat{\sigma}_1^2] &= \frac{1}{2(n-1)} \sum_{i=2}^{n} E(\sigma(t_i)\varepsilon_i - \sigma(t_{i-1})\varepsilon_{i-1})^2 = \frac{1}{2(n-1)} \sum_{i=2}^{n} \Big\{ \sigma^2(t_i) + \sigma^2(t_{i-1}) \Big\} \\ &\approx \int_0^1 \sigma^2(t) f(t) dt \\ E_{H_0}[\hat{\sigma}_{LSE}^2] &= \frac{1}{n} \sum_{i=1}^{n} E(\sigma^2(t_i)\varepsilon_i^2) - E\Big(\frac{1}{n}\sum_{j=1}^{n} \sigma^2(t_j)\varepsilon_j\Big)^2 \approx \frac{1}{n}\sum_{i=1}^{n} \sigma^2(t_i) \\ &\approx \int_0^1 \sigma^2(t) f(t) dt. \end{aligned}$$

Consequently, the statistic $\hat{M}_n^2$ can approximately be interpreted as a statistic of the form

$$\tilde{M}_n^2 = \hat{\sigma}_{H_0}^2 - \hat{\sigma}_1^2, \tag{2.12}$$

where $\hat{\sigma}_{H_0}^2$ denotes a variance estimator under the null hypothesis [here we use the least squares estimator] and $\hat{\sigma}_1^2$ is a variance estimator in the general nonparametric regression model [here we used the estimator of Rice (1984)]. Note that (2.12) is a nonparametric analogon of the statistic in the numerator of the $F$-test in the linear model and that von Neumann (1941, 1942) discussed a ratio of $\hat{\sigma}_{LSE}^2$ and $\hat{\sigma}_1^2$ for the construction of a test for a trend. The following result is a consequence of standard Hilbert space theory [see Dette and Munk (1998b)] and shows that under the null hypothesis of a linear model (i.e. $M^2 = 0$) the estimator of the empirical minimal $\mathcal{L}^2$-distance $\hat{M}_n^2$ is always approximately equal to a difference of variance estimators.

**Lemma 2.4.** *Let*

$$\hat{\sigma}_{LSE}^2 = \min_{\theta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} (Y_i - \theta^T g(t_i))^2 \tag{2.13}$$

*denote the standard least squares estimator in the linear regression model $y = \theta^T g(t)$ and $\hat{\sigma}_1^2$ the estimator of Rice (1984) defined in (2.11), then*

$$\hat{M}_n^2 = \hat{\sigma}_{LSE}^2 - \hat{\sigma}_1^2 + O_p\Big(\frac{1}{n}\Big) \;=\; M^2 + O_p\Big(\frac{1}{n}\Big).$$

From these observations it is reasonable to use different variance estimators in the statistic $\hat{A}_0$ or in the statistic (2.12). An obvious generalization of (2.11) are the difference based type estimators introduced by Hall, Kay and Titterington (1990), which are defined by

$$\hat{\sigma}^2_{d,r} = \frac{1}{n-r} \sum_{i=r+1}^{n} \Big( \sum_{j=0}^{r} d_i Y_{i-j} \Big)^2 . \tag{2.14}$$

Here the sequence $(d_i)_{i=0}^r$ is called difference sequence and satisfies

$$\sum_{i=0}^{r} d_i = 0; \quad \sum_{i=0}^{r} d_i^2 = 1. \tag{2.15}$$

Note that in the case $r = 1$ we obtain from (2.15) the estimator of Rice (1984) defined in (2.11). Another popular choice is the variance estimator of Gasser, Sroka and Jennen-Steinmetz (1986)

$$\hat{\sigma}^2_G = \frac{1}{6(n-2)} \sum_{i=3}^{n} (Y_{i-2} - 2Y_{i-1} + Y_i)^2 \tag{2.16}$$

and is obtained by the choice

$$d_0 = d_2 = \frac{1}{\sqrt{6}}; \quad d_1 = -\frac{2}{\sqrt{6}}.$$

For a careful discussion of difference type variance estimators we refer to Dette, Munk and Wagner (1998). It can be shown that the asymptotic normality in Theorem 2.2 remains valid, if the estimator $\hat{\sigma}^2_1$ in the definition of $\hat{A}_0$ or in (2.12) is replaced by a general difference based type estimator. In this case the asymptotic distribution of the statistic

$$\sqrt{n}(\hat{M}^2_n - M^2) \;=\; \sqrt{n}(\hat{\sigma}^2_{LSE} - \hat{\sigma}^2_{d,r} + o_p(n^{-1/2})) - M^2 \tag{2.17}$$

is still asymptotically normal with mean 0 and asymptotic variance

$$\lambda^2_D = 4\Big(\sum_{k=1}^{r} (\sum_{j=0}^{r-k} d_j d_{j+k})^2\Big) \int_0^1 \sigma^4(t) f(t) dt + 4 \int_0^1 \sigma^2(t)(m - P_{\mathcal{U}_d} m)^2(t) f(t) dt \tag{2.18}$$

This variance can be minimized with respect to the choice of the weights $d_i$ [see Hall, Kay and Titterington (1990)]. The optimal weights are displayed in Table 2.1 and give

$$\sum_{j=0}^{r-k} d_j d_{j+k} = -\frac{1}{2r} \quad k = 1, \ldots, r,$$

which yields for the smallest variance obtainable by the choice of a difference type estimator of order $r$

$$\lambda^2_r = \frac{1}{r} \int_0^1 \sigma^4(t) f(t) dt + 4 \int_0^1 \sigma^2(t)(m - P_{\mathcal{U}_d} m)^2(t) f(t) dt$$

[the case $r = 1$ is given in Theorem 2.2]. In practice, a reasonable approximation by the normal distribution is obtained by the use of the estimators in (2.11) and (2.16). For an improvement of power it is recommended to choose variance minimizing weights. However, this increases usually also the bias and can only be recommended for large samples or in combination with the application of a jacknife correction or a bootstrap test [see Munk (2001b) for an empirical study].

| $r$ | $d_0$ | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ |
|---|---|---|---|---|---|---|
| 1 | 0.7071 | -0.7071 | | | | |
| 2 | 0.8090 | -0.5 | -0.3090 | | | |
| 3 | 0.1942 | 0.2809 | 0.3832 | -0.8582 | | |
| 4 | 0.2798 | -0.0142 | 0.6909 | -0.4858 | -0.4617 | |
| 5 | 0.9064 | -0.2600 | -0.2167 | -0.1774 | -0.1420 | -0.1103 |

**Table 2.1:** Optimal weights for difference type variance estimators of order $r$

## 2.3 Applications

In this section we illustrate how Theorem 2.2 can be used to solve the problems indicated in Section 1. 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 normal distribution. This provides a very simple test for the classical hypotheses (1.4). Moreover, Theorem 2.2 (or its generalization) gives the asymptotic limit distribution of the statistic $\sqrt{n}(\hat{M}_n^2 - M^2)$ defined in (2.17) for arbitrary regression functions $m$, which can be used to solve the statistical problems mentioned in the introduction.

- **Testing of the classical null hypothesis.** The classical null hypothesis of the linear model $H_0 : m \in \mathcal{M}$ is rejected if

$$n^{1/2} \frac{\hat{M}_n^2}{\hat{s}_n^2} > \delta_r u_{1-\alpha} \ , \tag{2.19}$$

where the statistic $\hat{M}_n^2$ is defined in (2.8), $u_\beta$ denotes the $\beta$-quantile of the standard normal distribution,

$$\delta_r = 2\Big(\sum_{k=1}^{r}(\sum_{j=0}^{r-k} d_j d_{j+k})^2\Big)^{1/2} \quad (r \geq 1). \tag{2.20}$$

and $\hat{s}_n^4$ is an appropriate estimator of the asymptotic variance $\int_0^1 \sigma^4(t) f(t) dt$ under the null hypothesis of a linear model. Note that $\hat{\lambda}_{0,n} = \delta_r^2 \hat{s}_n^4$ defines an estimator of the asmptotic variance in (2.18) under the assumption of a linear model. For the estimator of Rice (1984) defined in (2.11) it follows $\delta_1 = 1$, while for the estimator of Gasser, Sroka, Jennen-Steinmetz (1986) given in (2.16) we have $\delta_2^2 = 17/9$. The choice of the variance estimator depends on the structure of the error distribution. Under homoscedasticity we have $\int_0^1 \sigma^4(t) f(t) dt = \sigma^4$ and any square of the difference based type estimators defined in (2.14) could be used, i.e.

$$\hat{s}_n^2 = \hat{\sigma}_{d,r}^2. \tag{2.21}$$

If heteroscedasticity has to be taken into account a simple estimator is given by

$$\hat{s}_n^4 = \frac{1}{4(n-3)} \sum_{i=2}^{n-2} (Y_i - Y_{i-1})^2 (Y_{i+2} - Y_{i+1})^2 . \tag{2.22}$$

- **Approximation of the power.** Theorem 2.2 establishes also asymptotic normality under (fixed) alternatives, which allows an estimation of the power of the proposed test, i.e.

$$P(\text{“rejection”}) \sim 1 - \Phi((\hat{\lambda}_{0,n} u_{1-\alpha} - \sqrt{n} \hat{M}_n^2)/\hat{\lambda}_{1,n})$$

where $\hat{\lambda}_{1,n}^2$ is an appropriate estimator of the asymptotic variance.

- **Confidence intervals for the measure of linearity.** Theorem 2.2 can be used for constructing asymptotic confidence intervals for the $\mathcal{L}^2$-distance $M^2$ between the regression function $m$ and the subspace $\mathcal{U}_d$ induced by the linear model. For example, an (asymptotic) $(1-\alpha)$ upper confidence bound for $M$ is given by $M_n^+$, where

$$(M_n^+)^2 = \hat{M}_n^2 + \frac{u_{1-\alpha}}{\sqrt{n}} \hat{\lambda}_{1,n} \tag{2.23}$$

and $\hat{\lambda}_{1,n}^2$ is an appropriate estimator of the asymptotic variance defined in (2.18). In the presence of homoscedasticity the problem of estimating this variance in the general model can be simplified by using the upper confidence bound

$$(M_n^+)^2 = \hat{M}_n^2 + \frac{2u_{1-\alpha}^2}{n} \hat{s}_n^2 + \frac{u_{1-\alpha} \hat{s}_n}{\sqrt{n}} \sqrt{4\hat{M}_n^2 + \hat{s}_n^2 \Big( \frac{4u_{1-\alpha}^2}{n} + \delta_r^2 \Big)} , \tag{2.24}$$

where $\hat{s}_n^2$ is defined in (2.21).

- **Testing of neighborhood hypotheses.** Theorem 2.2 provides a method for the statistical assessment of the linear model instead of a simple “model check”. The “neighborhood” hypothesis

$$H_\pi : M > \pi \quad \text{versus} \quad K_\pi : M \leq \pi \tag{2.25}$$

is rejected whenever

$$n^{1/2} \frac{\hat{M}_n^2 - \pi^2}{\hat{\lambda}_{1,n}} \leq u_\alpha , \tag{2.26}$$

where $\hat{\lambda}_{1,n}^2$ is an appropriate estimator of the asymptotic variance defined in (2.18). Note that under the assumption of homoscedasticity the denomimnator on the left hand side of the inequality (2.26) can be replaced by

$$\hat{\lambda}_{1,n} = \{4\hat{s}_n^2\pi^2 + \delta_r^2\hat{s}_n^4\}^{1/2} ,$$

where $\hat{s}_n^2$ is defined in (2.21).

## 2.4 Simulations for goodness of fit tests based on difference type variance estimators.

In this section we will investigate the finite sample properties of the asymptotic normal law. To this end we assumed homoscedasticity and performed an extensive simulation study of the distribution of the statistic

$$H_n = n^{1/2}\frac{\hat{M}_n^2 - M^2}{\left(\frac{17}{9}\hat{s}_n^4 + 4\hat{s}_n^2 M^2\right)^{1/2}} \tag{2.27}$$

for various sample sizes $n$, where $\hat{s}_n^2$ denotes the estimator of Gasser, Sroka and Jennen-Steinmetz (1986) defined in (2.16), for which $\delta_2^2 = 17/36$ (see the discussion of the previous paragraph). We considered three different linear models

$$\begin{aligned} m_1(x) &= a_1 + a_2x + \exp(\kappa x) && \text{model I} \\ m_2(x) &= a_1 + a_2x + (\kappa x)^2\sin(\kappa x) && \text{model II} \\ m_3(x) &= a_1 + a_2x + \kappa x^2 && \text{model III} \end{aligned}$$

where $\kappa$ is a parameter which determines the $\mathcal{L}^2$-distance to the linear regression model $\mathcal{U}_2 = span\ \{1, x\}$. As a design we used

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

which corresponds to the design density $f \equiv 1$. Table 2.2 displays for various sample sizes the accuracy of the approximation of the $\alpha$-quantiles of the standard normal distribution. The numbers of outcomes which were smaller than the corresponding quantile of the normal distribution were counted and divided by the total number of simulations. For the sample size and $\alpha$-quantiles we chosed $n = 20, 50, 100$ and $\alpha = 0.05,\ 0.1,\ 0.9,\ 0.95$, respectively (note that the 5% and 10% quantile are needed in (2.26) while the 90% and 95% quantile are used in (2.19)). For each study we have performed 5000 replications with a SAS-IML random generator. The error, $\varepsilon$, was assumed to be normally distributed with variance $\sigma^2$ where $\sigma^2 = 0.09, 0.25, 1, 4$. For the distance $M$, we chose the values $M^2 = 0,\ 0.5,\ 1$. It turns out 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 see that the variance $\sigma^2$ does not affect the distribution of the test statistic $H_n$ which

is in accordance with the observation that $H_n$ in (2.27) is invariant with respect to the group of scale transformations acting on $\mathbb{R}^n$, whenever $M = 0$.
We observe that, even for relatively small sample sizes, the approximation of the level is surprisingly accurate, independent of $M^2$ and the unknown variance $\sigma^2$. 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 we refer to Dette and Munk (1998a) for more details.

| | $M^2$ | | 0 | | | | .5 | | | | 1 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Quantile | | .05 | .1 | .9 | .95 | .05 | .10 | .9 | .95 | .05 | .10 | .9 | .95 |
| $m_1$ | $\sigma^2$ | .09 | .042 | .085 | .915 | .962 | .050 | .091 | .879 | .936 | .044 | .078 | .869 | .927 |
| | | .25 | .042 | .085 | .915 | .962 | .056 | .102 | .891 | .945 | .051 | .096 | .883 | .939 |
| | | 1 | .042 | .085 | .915 | .962 | .056 | .107 | .906 | .956 | .058 | .107 | .900 | .951 |
| | | 4 | .042 | .085 | .915 | .962 | .047 | .093 | .915 | .959 | .050 | .098 | .912 | .958 |
| $m_2$ | $\sigma^2$ | .09 | .042 | .085 | .915 | .962 | .031 | .070 | .884 | .938 | .028 | .059 | .867 | .930 |
| | | .25 | .042 | .085 | .915 | .962 | .042 | .083 | .901 | .948 | .040 | .080 | .892 | .942 |
| | | 1 | .042 | .085 | .915 | .962 | .048 | .091 | .917 | .956 | .052 | .092 | .907 | .952 |
| | | 4 | .042 | .085 | .915 | .962 | .041 | .081 | .919 | .963 | .045 | .089 | .920 | .960 |
| $m_3$ | $\sigma^2$ | .09 | .042 | .085 | .915 | .962 | .058 | .099 | .881 | .939 | .053 | .091 | .873 | .933 |
| | | .25 | .042 | .085 | .915 | .962 | .065 | .112 | .896 | .946 | .060 | .104 | .887 | .941 |
| | | 1 | .042 | .085 | .915 | .962 | .060 | .113 | .910 | .955 | .063 | .116 | .902 | .951 |
| | | 4 | .042 | .085 | .915 | .962 | .050 | .093 | .916 | .959 | .054 | .102 | .913 | .959 |

**Table 2.2.** *Simulated deviation from the normal distribution for the statistic $H_n$ with sample size $n = 20$*

If higher order difference estimators are used, in general power can be improved essentially. However, in this case, as well for nonlinear models and random design [see Section 2.7 for the corresponding asymptotic result], jacknife corrections are recommended. An extensive simulation study and guidelines for the order of the difference estimator can be found in Munk (2001b).

## 2.5 Goodness of fit tests using kernel based methods

From the arguments surrounding Lemma 2.4 it seems reasonable to replace in the statistic $\hat{\sigma}_{LSE}^2 - \hat{\sigma}_1^2$ the nonparametric estimator $\hat{\sigma}_1^2$ by some refinement. The application of generalized versions of difference based type estimators was illustrated in the previous Section 2.2 and in

the following section we will now discuss variance estimators based on nonparametric curve estimators which will yield to asymptotically more efficient procedures.
More precisely, we use a sum of squared residuals based on a nonparametric fit of the regression function as estimator of the integrated variance function. Following Hall and Marron (1990) we define weights of the form

$$w_{ij} = \frac{K\left(\frac{t_i - t_j}{h}\right)}{\sum_{\ell=1}^{n} K\left(\frac{t_i - t_\ell}{h}\right)}, \quad (i, j = 1, \ldots, n) ,$$

where $K$ is a kernel function with compact support satisfying

$$\int_{-\infty}^{\infty} K(u) u^j du = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } 1 \le j \le r-1 \end{cases}, \tag{2.28}$$

for some $r \geq 2$ and $h$ is called bandwidth. Note that for each $i$ we have $\sum_{j=1}^{n} w_{ij} = 1$ and usually the weight $w_{ij}$ is decreasing with the difference $|i - j|$. The weighted average $\hat{m}_i = \sum_{j=1}^{n} w_{ij} Y_j$ is used as an estimator of the unknown regression function at the point $t_i$ and based on the residuals

$$\hat{\varepsilon}_i = Y_i - \hat{m}_i = Y_i - \sum_{j=1}^{n} w_{ij} Y_j \quad (i = 1, \ldots, n)$$

from this fit. Hall and Marron (1990) proposed

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

as an estimator of the variance in a homoscedastic nonparametric regression. Here the quantity

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

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

$$h = O(n^{-\frac{2}{4r+1}}), \;\; nh^2 \to \infty \tag{2.29}$$

if $n \to \infty$, and the design density $f$, the regression function $m$, the variance function $\sigma^2$ and the basis functions $g_1, \ldots, g_p$ are assumed to be sufficiently smooth, i.e.

$$m, f \in C^{(r)}([0,1]), \;\; \sigma^2, g_1, \ldots, g_p \in C^{(1)}([0,1]) , \tag{2.30}$$

where $C^{(r)}([0,1])$ denotes the space of $r$ times continuously differentiable functions. 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 $E[\hat{\sigma}^2_{HM}] = \sigma^2 + C_2 h^{2r} + o(h^{2r})$, where

$$C_2 = \kappa_r^2 \int_0^1 \{(mf)^{(r)}(u) - mf^{(r)}(u)\}^2 \frac{du}{f(u)}, \tag{2.31}$$

and the constant $\kappa_r$ is defined by

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

In the heteroscedastic case a slight modification of this expansion is necessary and it was shown in Dette (1999) that

$$E[\hat{\sigma}^2_{HM}] = \int_0^1 \sigma^2(t) f(t) dt + C_2 h^{2r} + \frac{C_3}{nh} \; + \; o(h^{2r}) + O\Big(\frac{1}{n}\Big), \tag{2.32}$$

where the constant $C_2$ is defined by (2.31) and

$$C_3 = \Big(2K(0) - \int_{-\infty}^{\infty} K^2(t)dt\Big)\Big(\int_0^1 \sigma^2(t) f(t) dt - \int_0^1 \sigma^2(t) dt\Big). \tag{2.33}$$

Observing the discussion given in Section 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

$$T_n = \hat{\sigma}^2_{LSE} - \hat{\sigma}^2_{HM}. \tag{2.34}$$

A tedious calculation [see Dette (1999)] shows that

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

where $M^2$ is defined in (2.3). Consequently the statistic $T_n$ is an asymptotically unbiased estimator of the minimal $\mathcal{L}^2$-distance $M^2$ between the linear regression model and the unknown regression function. The following result gives the asymptotic distribution of the random variable $T_n$ under the hypothesis of linearity ($M^2 = 0$) and the alternative ($M^2 > 0$). The proofs are cumbersome and can be found in Dette (1999).

**Theorem 2.5.** *Assume that (2.1), (2.2), (2.28), (2.29) and (2.30) are satisfied and $n \to \infty$.*

**a)** *If $M^2 = 0$ , then*

$$n\sqrt{h}(T_n + C_2 h^{2r} + \frac{C_3}{nh}) \stackrel{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \mu_0^2) \; , \tag{2.35}$$

*where the asymptotic variance is given by*

$$\mu_0^2 = 2 \int_0^1 (2K(u) - K * K(u))^2 du \int_0^1 \sigma^4(u) du \tag{2.36}$$

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

**b)** *If $M^2 > 0$ , then*

$$\sqrt{n}(T_n - M^2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mu_1^2) \ , \tag{2.37}$$

*where the asymptotic variance is given by*

$$\mu_1^2 = 4 \int_0^1 \sigma^2(u)[m(u) - P_{\mathcal{U}_d} m(u)]^2 f(u) du, \tag{2.38}$$

*and $P_{\mathcal{U}_d}$ denotes the orthogonal projection onto the space $\mathcal{U}_d$ defined in (2.10) with resepct to the inner product $\langle \cdot, \cdot \rangle$.*

It is remarkable that the normalizing factor is of different order in both parts of Theorem 2.5. Under the null hypohtesis $H_0 : m \in \mathcal{M}$ the variance of $T_n$ is of order $(n^2h)^{-1}$ while under the alternative $m \notin \mathcal{M}$ this is of order $n^{-1}$. Moreover, we have $C_3 = 0$ under the assumption of a uniform design or a homoscedastic error. In all applications of Theorem 2.5 (as indicated in Section 2.3) the choice of the bandwidth $h$ becomes an important and non-trivial problem. 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 it can be shown that this result carries over to the heteroscedastic case.
However, our numerical studies show that in the problem of validating linear models the bias of the statistic $T_n$ 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

$$(\hat{\sigma}_n^2/n)^{2/(2r+1)},$$

where $r$ is the order of the corresponding kernel and $\hat{\sigma}_n^2$ is an estimator of the integrated variance $\int_0^1 \sigma^2(t) f(t) dt$, for example the variance estimator of Rice (1984) defined in (2.11) or the estimator of Gasser, Sroka and Jennen-Steinmetz (1986) given in (2.16). 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 classical hypothesis of linearity

$$H_0 : m \in \mathcal{M}$$

is rejected if

$$n\sqrt{h}\Big(T_n + \frac{C_3}{nh}\Big) > u_{1-\alpha}\hat{\mu}_0; \quad h = o(n^{-2/(4r+1)}), \tag{2.39}$$

where $u_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution and $\hat{\mu}_0^2$ an appropriate estimator of the variance. A simple way of estimating this variance under the null hypothesis and the general assumption of heteroscedasticity is given by

$$\hat{\mu}_0^2 = 2 \int_{-\infty}^{\infty} (2K(u) - K * K(u))^2 du \sum_{i=1}^{n-1} (t_{i+1} - t_i)[Y_i - \hat{\theta}_n^T g(t_i)]^2 [Y_{i+1} - \hat{\theta}_n^T g(t_{i+1})]^2, \tag{2.40}$$

where $\hat{\theta}_n$ denotes the least squares estimator of $\theta$. In the case of a homoscedastic error structure we can use a modified estimator for the asymptotic variance, that is

$$\tilde{\mu}_0^2 = 2\int_{-\infty}^{\infty} (2K(u) - K * K(u))^2 du \cdot \hat{\sigma}^4_{LSE}. \tag{2.41}$$

Similary, all other implications stated in Section 2.3 remain valid for the statistic $T_n$. For example Theorem 2.5 provides useful information about the type II error of the above test at any particular point of the alternative by the approximation

$$p = P(\text{“rejection”}) \approx \Phi\Big(\frac{\sqrt{n}}{\mu_1}\Big\{M^2 - \frac{u_{1-\alpha}\mu_0}{n\sqrt{h}}\Big\}\Big), \tag{2.42}$$

and an asymptotic level $\alpha$ test for the problem of precise hypotheses (2.25) is obtained by rejecting the null hypothesis $H : M > \pi$ , whenever

$$\sqrt{n}\frac{T_n - \pi^2}{\hat{\mu}_1} \ \le\ u_\alpha,$$

where $\hat{\mu}_1^2$ is an appropriate estimator of the asymptotic variance $\mu_1^2$ and $u_\alpha$ denotes the $\alpha$-quantile of the standard normal distribution. Finally, it is also notable that Theorem 2.5 allows the construction of confidence intervals for the parameter $M^2$, which measures the deviation from the linear model $\mathcal{M}$.

**Remark 2.6.** It is worthwhile to mention that the results of Theorem 2.5 do not depend on the special smoothing procedures used for the calculation of the residuals $\hat{\varepsilon}_i = Y_i - \sum_{j=1}^n w_{ij} Y_j$ in the nonparametric variance estimator. Roughly speaking the results of Theorem 2.5 remain valid, where the bias term $C_2 h^{2r}$ and the asymptotic variance $\mu_0^2$ in (2.36) depend on the specific smoothing procedure under consideration, while the asymptotic variance $\mu_1^2$ in (2.38) is not changed. For example, if the kernel estimator of Gasser and Müller (1979) is used for the calculation of the residuals, the weights $w_{ij}$ are defined by

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

where $s_0 = 0, s_n = 1, s_j = (t_j + t_{j+1})/2 \quad (j = 1, \dots, n-1)$ and only the constant $C_2$ has to be changed in Theorem 2.5, that is

$$C_2 = \kappa_r^2 \int_0^1 (m^{(r)}(t))^2 dt.$$

A similar argument applies to the local polynomial estimator [see e.g. Fan (1992) or Fan and Gijbels (1996)], where 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 $(\hat{\beta}_{i0}, \ldots, \hat{\beta}_{ip})$ is the minimizer of

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

[see Dette (1999) for more details]. Note that the local polynomial estimator of order $p = 0$ corresponds to the Nadaraya-Watson estimator discussed in Theorem 2.5.

**Remark 2.7.** Several related tests using kernel based methods have been proposed in the literature [see Azzalini and Bowman (1993), Härdle and Mammen (1993) and Zheng (1997)], for which similar results can be obtained. For later purposes we mention a statistic introduced by Zheng (1997)

$$Z_n = \frac{1}{hn(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \hat{\varepsilon}_i \hat{\varepsilon}_j K(\frac{t_i - t_j}{h}). \tag{2.43}$$

Zheng (1997) showed under the null hypothesis of linearity that

$$n\sqrt{h} V_n \stackrel{\mathcal{D}}{\Rightarrow} \mathcal{N}(0, \mu_0^2),$$

where the asympotic variance is given by

$$\mu_0^2 = \int_0^1 \sigma^4(t) f^2(t) dt \cdot \int_{-1}^1 K^2(x) dx.$$

Under fixed alternatives Dette (1999) proved

$$\sqrt{n} \Big\{ Z_n - \int_0^1 K(u)(\Delta f)(x)(\Delta f)(x - uh) du dx \Big\} \stackrel{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \mu_3^2), \tag{2.44}$$

where $\Delta(t) = m(t) - P_{\mathcal{U}_d} m(t)$, and the asymptotic variance is given by

$$\mu_3^2 = 4 \int_0^1 \sigma^2(u) \Big\{ (\Delta f)(u) - P_{\mathcal{U}_d}(\Delta f)(u) \Big\}^2 f(u) du. \tag{2.45}$$

## 2.6 Simulations for goodness of fit tests based on kernel type variance estimators.

In this example we study the finite sample behaviour of the test (2.39). We consider a random and a fixed uniform design [for the corresponding asymptotic theory see Section 2.7] on the interval $[0, 1]$ (i.e. $f \equiv 1$) and the sample size is chosen as $n = 50, 100, 200$ and 400. Note that for the uniform design we have for the constant $C_3$ in (2.39) $C_3 = 0$. We use the kernel

$$K(t) = \frac{3}{4}(1 - t^2) I\{|t| \leq 1\} \tag{2.46}$$

of order 2 and the corresponding bandwidth $h = (\hat{\sigma}_1^2/n)^{2/5}$, where $\hat{\sigma}_1^2$ denotes the estimator of Rice (1984) defined in (2.11). Our simulation results show that the choice of the estimator has a negligable impact on the distributional behaviour of the test statistic (these results are not displayed). The asymptotic variance $\mu_0^2$ in (2.39) was estimated by the statistic $\tilde{\mu}_0^2$ defined in (2.41) for the case of a homoscedastic error and by the statistic $\hat{\mu}_0^2$ in (2.40) for the case of a heteroscedastic error. 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 nominal level and are not displayed for the sake of brevity.
We consider the model

$$m(t) = 5t + at^2 \tag{2.47}$$

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 2.3 for a homoscedastic error and show the relative proportion of rejection calculated by 1000 simulations on the basis of a 5% level. The normal approximation provided by Theorem 2.5 yields a very accurate approximation of the nominal level, even for $n = 50$. Moreover, the second part of Theorem 2.5 provides the asymptotic distribution of the test statistic in (2.39) 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 $H_0 : m \in \mathcal{M}$. As a numerical example we estimate the probability of rejection in the case of a fixed design, $a = 2, n = 100, \sigma^2 = 1$ and obtain $p \approx 0.334$, while the corresponding simulated probability is 0.343 (see Table 2.3).

| | | $n = 50$ | | $n = 100$ | | $n = 200$ | | $n = 400$ | |
|---|---|---|---|---|---|---|---|---|---|
| $\sigma^2$ | $a$ | 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 2.3.** *Simulated rejection probabilities of the test (2.39) in the model (2.47) for various values of* $a, \sigma^2, n$ *and a fixed and random uniform design on the interval* $[0, 1]$.

Our second example investigates the impact of deviations from homoscedasticity on power and level of the test (2.39). We consider a heteroscedastic situation for the model (2.47) where the variance function is given by $\sigma^2(t) = 3(1 + ct^2)/(3 + c)$. The results are given in Table 2.4 under the same setup as considered in the previous paragraph. Note that the function $\sigma^2$ has been normalized such that $\int_0^1 \sigma^2(t) f(t) dt = 1$. Compared to the homoscedastic case ($a = 0$) we observe no significant loss in the accuracy of the approximation of the nominal level and a loss of power with increasing values of $c$.

| | | $n = 50$ | | $n = 100$ | | $n = 200$ | | $n = 400$ | |
|---|---|---|---|---|---|---|---|---|---|
| c | $a$ | fixed | random | fixed | random | fixed | random | fixed | random |
| | 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 2.4.** *Simulated power of the test (2.39) in the model (2.47) for various scenarios of hetereoscedasticity 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* $\int_0^1 \sigma^2(t) f(t) dt = 1$.

## 2.7 High dimensional predictors, nonlinear models, random designs

The interpretation of the test statistic as an ANOVA statistic allows generalizations in several directions. In this section we mention two possible extensions namely the consideration of multivariate predictors and the problem of testing the goodness of fit in nonlinear models. We also discuss the impact of the random design assumption on the corresponding test procedures.

- Theorem 2.5 can be extended without any difficulties to the case of a multivariate predictor of dimension $k$. If a product kernel

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

is used the normalizing factor under the null hypothesis is $n\sqrt{h_1 \dots h_k}$ where $h_j$ is the bandwidth used for the $j$th marginal kernel ($j = 1, \dots, k$). 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

$$h_j = \Big( \frac{1}{n} \int_0^1 \sigma^2(t) f(t) \Big)^{2/(k+4)} \qquad j = 1, 2, \dots, k.$$

A generalization of Theorem 2.2 to a multivariate predictor seems to be more difficult, because it relies heavily on the application of difference based type variance estimators. If data is available at design points which form approximately a grid results of Hall, Kay and Titterington (1991) can be used for the construction of an analogue of the test statistic $\hat{M}_n^2$ considered in Section 2.2.

- The results presented so far remain true if $\mathcal{M}$ is a class of nonlinear models

$$\mathcal{M} = \{m(t,\theta) | \theta \in \Theta\} \tag{2.48}$$

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

$$M_\Theta^2 = \min_{\theta\in\Theta} \int_0^1 [m(t) - m(t,\theta)]^2 f(t)dt \tag{2.49}$$

is attained at a unique interior point $\theta_0 \in \Theta$. Then under regularity assumptions [see e.g. Seber and Wild (1989), pp. 572-574 or Gallant (1987), Chapter 4], Theorem 2.2 and 2.5 remain valid, where $M^2$ in (2.3) has to be replaced by $M_\Theta^2$. A heuristic argument for this statement uses a Taylor expansion and the fact that the sum of squared residuals in the nonlinear model can be approximated by

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

where $G^T = \left(\frac{\partial m}{\partial \theta}(t_i,\theta_0)\right)_{i=1}^n \in \mathbb{R}^{d\times n}$ 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 $d$ regression functions

$$g_1(t) = \frac{\partial}{\partial\theta_1} m(t,\theta_0), \ldots, g_d(t) = \frac{\partial}{\partial\theta_d} m(t,\theta_0),$$

where the regression $m(t)$ has to be replaced by $m(t) - m(t,\theta_0)$. If the regression functions $g_1,\ldots,g_d$ are assumed as orthonormal with respect to the design density $f$, the orthogonal projection of a function $q$ onto $\mathcal{U}_d = \text{span}\,\{g_1,\ldots,g_d\}$ is given by

$$P_{\mathcal{U}_d} q = \sum_{\ell=1}^d \langle q, g_\ell\rangle g_\ell.$$

Consequently the quantity $M^2$ in Theorem 2.5 is given by

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

where $m_{\theta_0} = m(\cdot,\theta_0)$ and the last equality is a consequence of the representation (2.49), which implies

$$0 = \frac{\partial}{\partial\theta_\ell}\int_0^1 [m(t) - m(t,\theta)]^2 f(t)dt|_{\theta=\theta_0} = -2\langle m - m_{\theta_0}, g_\ell\rangle \ , \qquad \ell = 1,\ldots,d.$$

- 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 $t_1, \ldots, t_n$ are realizations of i.i.d. random variables $U_1, \ldots, U_n$ with positive density $f$ on the interval $[0,1]$. For a generalization of the statistic $\hat{M}_n^2$ to this situation let $U_{(1)}, \ldots, U_{(n)}$ denote the order statistics of $U_1, \ldots, U_n$ and define $R_1^{-1}, \ldots, R_n^{-1}$ as the antiranks of $U_1, \ldots, U_n$, i.e. $(U_{(1)}, Y_{R_1^{-1}}), \ldots, (U_{(n)}, Y_{R_n^{-1}})$ is the ordered sample (according to the $U$-values) of the observations. The analogs of the estimates in (2.7) are then defined by

$$\begin{aligned}
\hat{A}_0 &= \frac{1}{n-1} \sum_{j=2}^{n} Y_{R_j^{-1}} Y_{R_{j-1}^{-1}} \\
\hat{A}_i &= \frac{1}{n} \sum_{j=1}^{n} g_i(U_{(j)}) Y_{R_j^{-1}} = \frac{1}{n} \sum_{j=1}^{n} g_i(U_j) Y_j \\
\hat{B}_{p,q} &= \frac{1}{n} \sum_{j=1}^{n} g_p(U_j) g_q(U_j) \ .
\end{aligned}$$

If these estimates are used in the construction of the test statistic $\hat{M}_n^2$ it can be shown [see Dette and Munk (1998b)] that for $n \to \infty$

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

Here the asymptotic variance is given by

$$\begin{aligned}
\tau_1^2 &= E[\sigma^4(U_1)] + 4E[\sigma^2(U_1)\{(m - P_{\mathcal{U}_d} m)(U_1)\}^2] + \mathrm{Var}[\{(m - P_{\mathcal{U}_d} m)(U_1)\}^2] \\
&= \lambda^2 + \mathrm{Var}[\{(m - P_{\mathcal{U}_d} m)(U_1)\}^2] \ ,
\end{aligned}$$

where $\lambda^2$ is defined in (2.9) and $P_{\mathcal{U}_d}$ denotes the orthogonal projection onto the space $\mathcal{U}_d$ defined in (2.10) with resepct to the inner product $\langle \cdot, \cdot \rangle$.

The generalization of the statistic $T_n$ defined in Section 2.5 to the random design case is substantially easier, because it does not depend on the order of the predictor. In this case we have under the assumption of a linear model (i.e. $M^2 = 0$) for $n \to \infty$

$$\sqrt{h}(T_n + C_2 h^{2r} + \frac{C_3}{nh}) \stackrel{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \mu_0^2)$$

in the case $M^2$=0 and

$$\sqrt{n}(T_n - M^2) \stackrel{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \tilde{\mu}_1^2).$$

Here the asymptotic variance $\mu_0^2$ is defined in (2.36) and

$$\begin{aligned}
\tilde{\mu}_1^2 &= 4E[\sigma^2(U_1)\{(m - P_{\mathcal{U}_d} m)(U_1)\}^2] + \mathrm{Var}[\{(m - P_{\mathcal{U}_d} m)(U_1)\}^2] \\
&= \mu_1^2 + \mathrm{Var}[\{(m - P_{\mathcal{U}_d} m)(U_1)\}^2] \ ,
\end{aligned}$$

where $\mu_1^2$ is defined in (2.38).

Comparing these results with Theorem 2.2 and 2.5 we observe no difference between the random and fixed design assumption under the classical null hypothesis $M^2 = 0$. Surprisingly, there appears an additional term in the asymptotic variance under the alternative $M^2 > 0$ which affects the rate of misspecifications of the model $\mathcal{M}$, i.e. the probability of the type II error. In the case of testing classical hypotheses the test statistic does not discriminate between fixed and random design which may result in a large type II error solely caused by the variability of the predictor $U$. In fact, it can be shown that the hypothesis $H_0 : m \in \mathcal{M}$ will be falsely accepted with increasing probability as $\mathrm{Var}[\{(m - P_{\mathcal{U}_d} m)(U_1)\}^2]$ increases, fixing an alternative $M^2 = C > 0$. Because this quantity will be unknown in general this observation supports certainly the 'precise testing' approach in the random design case. Note, that also the length of the confidence intervals discussed in Section 2.3 involves additional estimation of $\mathrm{Var}[\{(m - P_{\mathcal{U}_d} m)(U_1)\}^2]$. Hence testing of (2.25) or the additional consideration of confidence intervals for the measure of linarity $M^2$ is strictly recommended.

# 3 Testing for homoscedasticity in regression

In many applications of regression models the usual assumptions of homoscedastic disturbances cannot be guaranteed a priori. Therefore, the importance of being able to detect heteroscedasticity is widely recognized because, if the assumption of homoscedasticity is not met, efficient inference for the regression function requires that the heteroscedasticity is taken into account. This may result in transformations of the data, weighted least squares (or modified likelihood) procedures or the choice of a variable bandwidth in nonparametric kernel smoothing [see Müller and Stadtmüller (1987)]. In many cases the loss of efficiency in using procedures for homoscedastic models under heteroscedastic errors may be substantial.

In contrast to the problem of testing for a parametric form of the regression much less attention has been paid to the problem of testing hypotheses regarding the variance structure in a nonparametric regression model. Most of the graphical procedures and formal tests are based on residuals after fitting a model with completely parametrically specified regression and variance function [see for example Harrison and McCabe (1979), Breusch and Pagan (1979), Koenker and Bassett (1981), Cook and Weisberg (1983), Carroll and Ruppert (1988), Sec. 3.4, and Diblasi and Bowman (1997)]. A diagnostic test under a smoothness assumption on the regression function and the assumption of a normal distributed error distribution has been proposed by Eubank and Thomas (1993). Recently Müller and Zhao (1995) considered a semiparametric variance function model where the relation between the mean regression and the variance function is assumed to follow a generalized linear model and the mean response is modelled nonparametrically. They demonstrated that $\sqrt{n}$-consistent inference for the parameters is possible and applied the results to obtain a test for heteroscedasticity under minimal assumptions on the shape of the regression function.

Although the problem of testing the hypothesis of homoscedasticity

$$H_0 : \sigma^2(\cdot) \equiv \sigma^2 \quad \text{versus} \quad H_1 : \sigma^2(\cdot) \not\equiv \sigma^2 \tag{3.1}$$

(for some constant $\sigma^2 > 0$) has been discussed by the above authors under parametric or semiparametric assumptions, this problem has only been solved recently by Dette and Munk (1998c), Zheng (2000) and Dette (2001) in a completely nonparametric regression setup under minimal assumptions on the error structure and the regression and variance function. It is the purpose of this section to present a simple concept for the construction of test statistics for the hypothesis (3.1), which relates this problem to the problem of testing for a parametric form in a transformed regression model. This concept is then successfully applied for the construction of two tests for homoscedasticity.

## 3.1 A a general principle to construct tests for homoscedasticity

Recall the definition of the estimator of the integrated variance function given in (2.11) [see Rice (1984)], which considers a sum of squared local residuals

$$\frac{1}{n-1}\sum_{i=2}^{n} R_{i,n}^2,$$

where

$$R_{i,n} = \frac{1}{\sqrt{2}}(Y_i - Y_{i-1}). \tag{3.2}$$

Under smoothness assumptions on the regression and variance function we have

$$E[R_i^2] = \frac{1}{2}[m(t_i) - m(t_{i-1})]^2 + \frac{1}{2}E[\sigma(t_i)\varepsilon_i - \sigma(t_{i-1})\varepsilon_{i-1}]^2 \approx \sigma^2(t_i). \tag{3.3}$$

Therefore, the problem of testing the hypothesis of homoscedasticity (3.1) is (approximately) equivalent to the problem of testing the hypothesis that the regression of the random variable $R_{i,n}^2$ on the predictor $t_i$ is constant. This problem was extensively discussed in Section 2, and it seems natural to apply one of the procedures developed in the previous sections to the pseudo residuals $R_{i,n}^2$. An obvious difficulty in this construction of the test statistic is that all procedures cited above are derived under the assumptions of independent innovations, but the calculation of pseudo residuals produces obviously $m$-dependent random variables. In the following section we will indicate that nevertheless this basic idea can be successfully applied for the construction of a test for heteroscedasticity with some additional technical difficulties in the asymptotic analysis.

## 3.2 Tests for homoscedasticity using difference based type variance estimators

Recall the situation of Example 2.1, where the test statistic for a constant regression is defined in (2.6). Following the discussion of the last paragraph we propose as a measure of homoscedaticicity the statistic

$$\hat{V}_n^2 = \frac{1}{(n-3)}\sum_{j=2}^{n-2} R_{j,n}^2 R_{j+2,n}^2 - \Big(\frac{1}{(n-1)}\sum_{j=2}^{n} R_{j,n}^2\Big)^2 , \tag{3.4}$$

where the pseudo residuals are defined in (3.2). Note that there is a small modification in (3.4), where we replaced $R^2_{j+1,n}$ by $R^2_{j+2,n}$ in (2.6). This replacement substantially simplifies the asymptotic analysis of the statistic $\hat{V}_n^2$ [see Dette and Munk (1998c)]. If the regression and variance function and the functions

$$(3.5) \qquad m_3(t) = E[(\varepsilon(t))^3] \quad , \quad m_4(t) = E[(\varepsilon(t))^4]$$

are Lipschitz continuous of order $\gamma > 0$ it follows that $\hat{V}_n^2$ is an asymptotically unbiased estimate for a measure of homoscedasticity

$$M_\sigma^2 := \min_{a \in \mathbb{R}^+} \int_0^1 (\sigma^2(t) - a)^2 f(t) dt = \int_0^1 \sigma^4(t) f(t) dt - \Big( \int_0^1 \sigma^2(t) f(t) dt \Big)^2 .$$

These results suggest a simple test for heteroscedasticity in the nonparametric regression model. The null hypothesis (3.1) is rejected for large values of the statistic $\hat{V}_n^2$, where $\hat{V}_n^2$ is the empirical $\mathcal{L}^2$-distance defined in (3.4). Our main result of this section shows that $\hat{V}_n^2$ (appropriately standardized) is asymptotically normally distributed, even under the alternative of heteroscedasticity. A detailed proof is rather involved and can be found in Dette and Munk (1998c). For the sake of brevity we will in the following only display the results for pseudo residuals of order 1 [see Section 2.2], however, we mention that analoguous results can be proved for higher order [see Dette (2001) and Munk, Wagner, Scholz and Bissantz (2001)]. The comparison of procedures with residuals of different order is similar to the case of testing parametric assumptions concerning the regression functions. Again, asymptotic most efficient difference schemes can be determined [see Munk, Wagner, Scholz and Bissantz (2001) for more details].

**Theorem 3.1.** *If the functions $m$, $\sigma^2$, $m_3$ and $m_4$ are Lipschitz continuous of order $\gamma > \frac{1}{2}$, then, as $n \to \infty$,*

$$(3.6) \qquad 4\sqrt{n}(\hat{V}_n^2 - M_\sigma^2) \xrightarrow{\mathcal{D}} N(0, v_\sigma^2),$$

*where the asymptotic variance is given by*

$$(3.7) \qquad \begin{aligned} v_\sigma^2 &= \int_0^1 \Big\{ (6m_4^2(t) + 4m_4(t) + 6)\sigma^8(t) + 64 m_4(t)\sigma^4(t)(\sigma^2(t) - \bar{\sigma}^2)^2 \\ &- 8m_3^2(t)\sigma^6(t)(\sigma^2(t) - 2\bar{\sigma}^2) \Big\} f(t)\, dt \end{aligned}$$

*and $\bar{\sigma}^2 = \int_0^1 \sigma^2(t) f(t) dt$.*

Note that under the hypothesis of homoscedasticity in (3.1) this result simplifies to

$$(3.8) \qquad 4\sqrt{n}\hat{V}_n^2 \xrightarrow{\mathcal{D}} N\Big(0, \sigma^8 \int_0^1 \{4m_4(t) + 6m_4^2(t) + 6 + 8m_3^2(t)\} f(t)\, dt \Big),$$

and in the case of normal responses we have from (3.7) that

$$v_\sigma^2 = 24\Big\{11\int_0^1 \sigma^8(t) f(t)\,dt - 16\int_0^1 \sigma^6(t)\,f(t)\,dt\,\bar\sigma^2 + 8\int_0^1 \sigma^4(t) f(t)\,dt\,\bar\sigma^4\Big\},$$

which reduces under the null hypothesis of homoscedasticity to $v_\sigma^2 = 72\sigma^8$. A test for the classical hypothesis of homoscedasticity can now easily be obtained by rejecting the null hypothesis in (3.1) whenever $4\sqrt{n}\hat M_n^2/\hat v_\sigma \; > \; u_{1-\alpha}$, where $u_{1-\alpha}$ denotes the $1-\alpha$ quantile of the standard normal distribution and $\hat v_\sigma$ is an estimator of the limiting variance $v_\sigma^2$ in (3.8). In principle this can be done in a straightforward but tedious manner. To be precise define

$$\hat A_{1,n} = \frac{4}{(n-3)}\sum_{j=2}^{n-2} R_{j,n}^4 R_{j+2,n}^4, \quad \hat A_{2,n} = \frac{2}{(n-5)}\sum_{j=2}^{n-4} R_{j,n}^4 R_{j+2,n}^2 R_{j+4,n}^2, \tag{3.9}$$

$$\hat A_{3,n} = \frac{2}{(n-3)}\sum_{j=2}^{n-2} R_{j,n}^4 R_{j+2,n}^2, \quad \hat A_{4,n} = \frac{2}{(n-1)}\sum_{j=2}^{n} R_{j,n}^4, \tag{3.10}$$

$$\hat A_{5,n} = \frac{2}{9(n-5)}\sum_{j=3}^{n-3} (R_{j,n} - R_{j-1,n})^3 (R_{j+3,n} - R_{j+2,n})^3, \tag{3.11}$$

$$\hat A_{6,n} = \frac{2}{9(n-7)}\sum_{j=3}^{n-5} (R_{j,n} - R_{j-1,n})^3 (R_{j+3} - R_{j+2,n})^3 R_{j+5,n}^2, \tag{3.12}$$

$$\hat S_{2k,n} = \frac{1}{(n-2k+1)}\sum_{j=2}^{n-2k+2} R_{j,n}^2 \dots R_{j+2k-2,n}^2, \quad k = 1,2,3,4, \tag{3.13}$$

then it follows that $\hat A_{1,n}, \dots, \hat A_{6,n}$ and $\hat S_{2k,n}$ are consistent estimators of

$$\int_0^1 (m_4(t)+3)^2\sigma^8(t)\,f(t)\,dt, \quad \int_0^1 (m_4(t)+3)\sigma^8(t)\,f(t)\,dt, \quad \int_0^1 (m_4(t)+3)\sigma^6(t)\,f(t)\,dt,$$
$$\int_0^1 (m_4(t)+3)\sigma^4(t)\,f(t)\,dt, \quad \int_0^1 m_3^2(t)\sigma^6(t)\,f(t)\,dt, \quad \int_0^1 m_3^2(t)\sigma^8(t)\,f(t)\,dt,$$
$$\int_0^1 \sigma^{2k}(t)\,f(t)\,dt \quad (k = 1,2,3,4),$$

respectively. The estimator of $v_\sigma^2$ is then obtained in a straightforward manner as an appropriate linear combination, i.e.

$$\begin{aligned} \hat v_n^2 \;&=\; 6\hat A_{1,n} + 32\hat A_{2,n} - 128\hat A_{3,n}\hat S_{2,n} + 64\hat A_{4,n}\hat S_{2,n}^2 + 16\hat A_{5,n}\hat S_{2,n} - 8\hat A_{6,n} \\ &\quad - 144\hat S_{8,n} + 384\hat S_{6,n}\hat S_{2,n} - 192\hat S_{4,n}\hat S_{2,n}^2. \end{aligned} \tag{3.14}$$

Under the hypothesis of homoscedasticity we obtain from (3.8)

$$\tilde v_n^2 = 6\hat A_{1,n} - 32\hat A_{2,n} + 48\hat S_{2,n}^8 + 8\hat A_{6,n} \tag{3.15}$$

as a consistent estimator of $v_\sigma^2$. A consistent test for the classical null hypothesis of homoscedasticity is now easily obtained from Theorem 3.1 by rejecting the null hypothesis of homoscedasticity, whenever

$$\frac{4\sqrt{n}\,\hat{T}_n^2}{\tilde{v}_n} > u_{1-\alpha} , \tag{3.16}$$

where $\tilde{v}_n^2$ is the estimator of the asymptotic variance under the assumption of homoscedasticity defined in (3.15), $u_{1-\alpha}$ denotes the upper $(1-\alpha)$ quantile of the standard normal distribution, and

$$\hat{T}_n^2 = \hat{V}_n^2 + \frac{1}{n-1}\{\hat{A}_{4,n} - \hat{S}_{2,n}^2\}. \tag{3.17}$$

The remainder term in (3.17) corresponds to the second order approximation of $E[\hat{V}_n^2]$. This correction is recommended because a numerical study (not displayed) showed that a better approximation of the finite sample distribution is achieved with $\hat{T}_n^2$ in (3.17) instead of $\hat{V}_n^2$. The performance of this test is illustrated by a detailed simulation study in Dette and Munk (1998c).

**Remark 3.2.** We mention that all statistical issues raised in Section 1 can be answered by similar arguments a given in Section 2.3. For example, the second part of Theorem 3.1 provides the possibility to estimate the type II error of the test in (3.16) as

$$P(\text{"}H_0 \text{ is accepted"}) \sim \Phi\left(\frac{u_{1-\alpha}\tilde{v}_n - 4\sqrt{n}\hat{T}_n^2}{\hat{v}_n}\right) , \tag{3.18}$$

where $\Phi$ denotes the standard normal c.d.f. and $\tilde{v}_n, \hat{v}_n$ are consistent estimators of the asymptotic variance under the null hypothesis and alternative, respectively [see (3.8) and (3.7)] . Observe, in addition, that the asymptotic power function of the test in (3.16) is invariant with respect to rescaling the variance function $\sigma^2(t)$ by a constant $c > 0$. Hence the power is not affected by the magnitude of the variance rather by its deviation from a constant.
Again Theorem 3.1 can be used for the construction of tests of precise hypotheses and confidence intervals for the measure of heteroscedasticity $M_\sigma^2$. For example, a $(1-\alpha)$-confidence interval for $M_\sigma^2$ is given by

$$[0, \hat{T}_n^2 + \frac{u_{1-\alpha}\hat{v}_n}{4\sqrt{n}}),$$

where $\hat{v}_n^2$ is a consistent estimator of the asymptotic variance under the alternative of heteroscedasticity [see (3.7)].

## 3.3 A test of homoscedasticity using kernel based methods

Following the discsussion in Section 3.1 we use a modified version of a kernel based test of constant regression for the construction of a test of homoscedasticity. For technical reasons we follow the approach of Zheng (1997) illustrated in Remark 2.7 and define

$$W_n = \frac{1}{(n-1)(n-2)h} \sum_{|i-j|\geq 2} K(\frac{t_i - t_j}{h})(R_{i,n}^2 - \overline{R_n^2})(R_{j,n}^2 - \overline{R_n^2}) \tag{3.19}$$

where $K$ denotes a kernel [see Section 2.5 for more details], $h$ is a bandwidth and

$$\overline{R_n^2} = \frac{1}{n-1} \sum_{i=2}^{n} R_{i,n}^2 \tag{3.20}$$

is the mean of the pseudo residuals defined in (3.2). A generalization of the statistic $W_n$ which is based on pseudo residuals of larger order turns out to be asymptotically more efficient and can be found in Dette (2001). The probabilistic properties of $W_n$ will be derived in the following Theorem.

**Theorem 3.3.** *Assume that $E[|\varepsilon(t)|^8) \leq c < \infty$ for all $t \in [0,1]$, that the regression and variance function are Lipschitz continuous of order $\gamma \geq 1/2$ and that $n \to \infty$, $h \to 0$, $nh^2 \to \infty$.*

*(i) Under the hypothesis (3.1) of homoscedasticity we have for the statistic $W_n$ defined in (3.19)*

$$n\sqrt{h}W_n \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \lambda_0^2), \tag{3.21}$$

*where the asymptotic variance is given by*

$$\lambda_0^2 = 2\sigma^8 \int K^2(x)dx \int_0^1 m_4^2(x) f^2(x)dx \ . \tag{3.22}$$

*(ii) Under the alternative of heteroscedasticity we have*

$$\sqrt{n}\Big\{W_n - \frac{1}{h}\int\int K(\frac{x-y}{h})\Delta(x)\Delta(y)f(x)f(y)dxdy\Big\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \lambda_1^2), \tag{3.23}$$

*where the asymptotic variance is given by*

$$\lambda_1^2 = 4\int_0^1 m_4(x)\sigma^4(x)\{(\Delta f)(x) - \overline{\Delta f}\}^2 f(x)dx, \tag{3.24}$$

*the function $\Delta$ denotes the deviation of $\sigma^2(\cdot)$ from its mean, i.e.*

$$\Delta(x) = \sigma^2(x) - \int_0^1 \sigma^2(t)f(t)dt \ , \tag{3.25}$$

*and $\overline{\Delta f}$ the corresponding mean of $\Delta f = \Delta \cdot f$ with respect to the design density, i.e.*

$$\overline{\Delta f} = \int_{-1}^{1} \Delta(x) f^2(x)dx. \tag{3.26}$$

Note that the bias in (3.23) has the expansion

$$E[W_n] = M^2(h) + o(1),$$

where

$$M^2(h) = \frac{1}{h}\int\int K(\frac{x-y}{h})(\Delta f)(x)(\Delta f)(y)dxdy = \int_0^1 \Delta^2(x)f^2(x)dx + o(1) \tag{3.27}$$

and $\Delta$ is defined in (3.25). The right hand side is of order $o(1)$ if and only if the hypothesis of homoscedasticity is satisfied, and we obtain a consistent test by rejecting the hypothesis (3.1) of homoscedasticity if

$$n\sqrt{h}W_n > u_{1-\alpha}\hat{\lambda}_{0,n}, \tag{3.28}$$

where $u_{1-\alpha}$ denotes the $(1-\alpha)$ quantile of the standard normal distribution and $\hat{\lambda}^2_{0,n}$ is any consistent estimator for the asymptotic variance (3.22). A simple estimator is obtained by similar arguments as in the previous section

$$\hat{\lambda}^2_{0,n} = 2(\hat{A}_{1,n} - 6\hat{A}_{2,n} + 9(\overline{R^2_n})^4)\int K^2(x)dx \tag{3.29}$$

where $\overline{R^2_n}$ is defined in (3.20) and $\hat{A}_{1,n}$, $\hat{A}_{2n}$ are given in (3.9). A simulation study illustrating the finite sample properties of a bootstrap version of this test can be found in Dette (2001).

# 4 Summary, conclsuions and further remarks

We have tried to argue, that the problem of assesing the goodness of fit, if treated as a testing problem, is conceptually different from most of the testing problems established in statistical praxis and theory. The paradigm to choose the null hypothesis in such a way that only strong empirical evidence should force us to switch to an alternative can be reparaphrased in goodness of fit problems in two completely different ways.

1. Choose the null hypothesis as the model to be investigated, because in this case rejection will imply more complicated modelling or giving up a well accepted theory. In general, a wrong rejection will lead to a *loss in efficency* in the subsequent statistical analysis.

2. Choose the model to be investigated as the alternative because controlling the error of acceptance of a model, albeit wrong, is the more serious error. This is, e.g. due to subsequent inconsistent data analysis, i.e. a *loss of consistency.*

We believe, that both formulations have their relative merits, depending of the sort of data analysis to be performed. In this paper, we summarized the required limit laws (based on a specific and subjective selection of test statistics) in two particular cases: testing parametric assumptions about the signal and the variance in regression models with continuous response and independent observations.

We are aware of the fact, that *testing* model assumptions is basically a questionable proceeding and it is not difficult to attack both approaches, 1. and 2., by means of more fundamental arguments, e.g. from a Bayesian perspective. Nevertheless, often goodness of fit tests are applied and will be applied in the future. Hence, we offer to add a bit more understanding of the particular test selected for a problem by considering its associated neighborhood hypotheses and confidence intervals for the underlying measure of discrepancy between the true model and hypothetical model space. This forces us, to consider the type of distance implicitely chosen by each test and its own view on the data. In all examples discussed (we have also investigated various other test statistics, which show a similar behaviour), this is displayed in the asymptotic mean and variance under violation of the null model. Hence, our proposal is to estimate these quantities in order to inspect whether decision in favour of a model is based on empirical grounds or simply because the chosen test has not enough power to detect the underlying deviation from the null model, due to small sample size or large variabilty of data, say. This can simply be done, because we found always asymptotic normality of all statisics under consideration. This is in contrast to the asymptotic distribution of many other goodnes of fit statistics for the hypothesis "the model $\mathcal{M}$ holds true" where often complicated convolutions of $\chi^2$-laws are required for the analysis.
Finally, we mention once again that the construction of asymptotic confidence intervals for the underlying measures of discrepancy is intrinsically related to this approach and requires the same limit laws.

**Acknowledgements.** Many of our results benefitted essentially from joint work, discussions and comments of many colleagues. We are particularly indepted to A. Scholz, S. Biedermann, C. Czado, G. Freitag, W. Härdle, J. Hart, I. Ledwina, E. Mammen, H.G. Müller, N. Neumeyer, F. Ruymgaart, S. Sperlich, I. Spreckelsen, U. Stadtmüller, T. Wagner and S. Zwanzig for their cooperation on this topic during the last years. I. Gottschlich typed various parts of this manuscript with great patience and expertise.
Finally, we would like to thank the Deutsche Forschungsgemeinschaft for their financial support of C. von Lieres and M. Vogt [De 502/9-1,2: Validierung von Hypothesen], who worked on this project with tremendous expertise and enthusiasm. The work of the first author was additionally supported by the DFG, SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen.

## References

Achieser, N.J. (1956). *Theory of Approximation.* Dover Publications, Inc., N.Y.
Alcalá, J.T. , Christóbal, J.A. and González-Manteiga, W. (1999). Goodness-of-fit test for linear models based on local polynomials. *Statistics & Probability Letters* **42**, 39-46.
Altman, D.G., Bland, J.M. (1995). Absence of evidence is not evidence of absence, *Brit. Medic. Journ.* **311**, 485.
Azzalini, A., Bowman, A.W. (1993). On the use of nonparametric regression for checking linear

relationships. *Jour. Roy. Statist. Soc. Ser. B* **55**, 549557.
Berger, J.O., Delampady, M. (1987). Testing precise hypotheses. *Statistical Science* **2**, 317-52.
Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test *Jour. Americ. Statist. Assoc.* **33**, 526-535.
Berkson, J. (1942). Tests of significance considered as evidence. *Jour. Amer. Statist. Assoc.* **37**, 325-335.
Berkson, J. (1943). Experience with tests of significance: A reply to Professor R.A. Fisher. *Jour. Americ. Statist. Assoc.* **38**, 242-246.
Biedermann, S., Dette, H. (2000). Testing linearity of regression models with dependent errors. *Test* **9** 417 - 438
Bierens, H.J. (1990). A consistent conditional moment test of functional form. *Econometrica* **58**, 1443-1458.
Bissantz, N., Munk, A. (2001). Comparison of parametric models with the same or a different number of parameters in noisy inhomogenoeus inverse problems - with applications to recovering the luminosity density from the milky way.
Breusch, T.S. and Pagan, A.R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrika* **47**, 1287-1294.
Brockwell, P.J. , Davis, R.A. (1991). *Time Series: Theory and Methods.* Springer, N.Y.
Brodeau, F. (1993). Tests for the choice of approximative models in nonlinear regression when the variance is unknown. *Statistics* **24**, 95-106.
Brown, L.D. Hwang, J.T.G., Munk, A. (1997). An unbiased test for the bioequivalence problem. *Ann. Statist.* **25**, 2345-2367.
Bunke, H. (1981) A note on parameter estimation in inadequate nonlinear regression models. *Math. Operationsforsch. Statist., Ser. Statistics* **12**, 7-11.
Bunke, H., Schmidt, W.H. (1980). Asymptotic results on nonlinear approximation of regression functions and weighted least squares. *Math. Operationsforsch. Statist. Ser. Statistics* **11**, 3-22.
Carroll, R.J. and Ruppert D. (1988). *Transformation and Weighting in Regression.* Chapman and Hall, New York.
Chen, J.C. (1994). Testing for no effect in nonparametric regression via spline smoothing techniques. *Ann. Inst. Stat. Math.* **46**, 251-65.
Chow, S.C. & Liu, J.P. (1992). *Design and Analysis of Bioequivalence Studies.* STATISTICS: textbooks and monographs. Marcel Dekker, Inc.
Cook, R.D. and Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika* **70**, 1 - 10.
Cox, D., Koh, E., Wahba, G. Yandell, B.S. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.* **18**, 113-19.
Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *Ann. Statist.* **27**, 1012 - 1040.
Dette, H. (2001). A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. *Jour. Statist. Plann. Inf.*, to appear.
Dette, H., Munk, A. (1998a). Validation of linear regression models. *Ann. Statist.* **26**, 778-800.
Dette, H., Munk, A. (1998b) A simple goodness of fit test for linear models under a random design assumption. *Ann. Inst. Statist. Math.* **50**, 253-275.

Dette, H., Munk, A. (1998c). Testing heteroscedasticity in nonparametric regression. *Jour. Roy. Statist. Soc. Ser. B.* **60**, 693-708.

Dette, H., Munk, A., Wagner, T. (1998). Estimating the variance in nonparametric regression - what is a reasonable choice? *Journ. Roy. Statist. Soc. Ser. B.* **60**, 751-764.

Dette, H., Munk, A., Wagner, T. (2000). Testing model assumptions in multivariate linear regression models. *Jour. of Nonpar. Statist.* **12**, 309 –342.

Diblasi, A. and Bowman, A. (1997). Testing for constant variance in a linear model. *Statistics & Probability Letters* **33**, 95-103.

Dieboldt, J. (1995). A nonparametric test for the regression function: Asymptotic theory. *Jour. Statist. Plann. Inf.* **44**, 1-17.

Djojosugito, R.A. (1994). On the use of cubic spline smoothing for testing parametric linear regression models. *Computational Statistics* **9**, 213-231.

Djojosugito, R.A. (1995). A cubic smoothing spline based lack of fit test for nonlinear regression models. *Commun. Statist. - Theory Meth.* **24**, 2183-2197.

Eubank, R.L., Spiegelmann, C.H. (1990). Testing the goodness of a linear model via regression techniques. *Jour. Amer. Statist. Assoc.* **85**, 387-392.

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

Eubank, R.L., Hart, J.D., LaRiccia, V.N. (1993). Testing goodness of fit via nonparametric function estimation techniques. *Comm. Statist. Theory and Meth.* **22**, 3327-3354.

Eubank, J., Thomas, W. (1993). Detecting heteroscedasticity in nonparametric regression. *Jour. Roy. Statist. Soc. Ser. B* **55**, 145-155.

Fan, J. (1992). Design-adaptive nonparametric regression. *Jour. Amer. Stat. Soc.* **87**, 998-1004.

Fan, J. (1996). Test of siginificance based on wavelet thresholding and Neyman's truncation. *Jour. Amer. Statist. Assoc.* **91**, 674 - 688.

Fan , J., Gijbels, I. (1996). Local polynomial modelling and its applications. Chapman and Hall, London.

Fan, J. Huang, L.-S. (2001). Goodness of fit tests for parametric regression models. *Jour. Amer. Statist. Assoc.* **96**, 640 - 652.

Fan, J. Zhang, C. Zhang J. (2001). Generalized likelihodd ratio statistics and Wilks phenomenon. *Ann. Statist.* **29**, 153 - 193.

Firth, D., Glosup, J., Hinkley, D.V. (1991). Model checking with nonparametric curves. *Biometrika* **78**, 245-252.

Gallant, a.R. (1987). *Nonlinear Statistical Models.* Wiley, N.Y.

Gasser,T., Skroka,L., Jennen-Steinmetz, G. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*, **73**, 626-633.

Gasser,T., Müller, H.G., Mamitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc.*, Ser. B, **47**, 238-252.

Gasser, T., Müller, H.G. (1979). Kernel estimation of regression functions. In: Smoothing techniques for curve Estimation. Lecture Notes in Mathematics 757. Springer-Verlag., New York.

González Manteiga, W., Cao-Abad, R. (1993). Testing hypothesis of general linear model using

nonparametric regression estimation. TEST 2, 161-189.
González Manteiga,W., Vilar Fernández, J.M. (1995). Testing linear regression models using non-parametric regression estimators when the errors are correlated. *Comput. Statist. & Data Analysis* **20**, 521-541.
Gozalo, P.L. (1993). A consistent model specification test for nonparametric estimation of regression function models. *Econometric Theory* **9**, 451-477.
Goutis, C., Robert, Ch.P. (1998). Model choice in generalised linear models: A Bayesian approach via Kullback-Leibler projections. *Biometrika* **85**, 29-37.
Härdle, W., Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1926-1947.
Hall, P., Kay, J.W., Titterington, D.M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77**, 521 - 528.
Hall, P. Kay, J.W., Titterington, D.M. (1991). On estimation of noise variance in two dimensional signal processing. *Adv. Appl. Prob.* **23**, 476-495.
Hall, P., Marron, J.S. (1990). On variance estimation in nonparametric regression. *Biometrika* **77**, 415-19.
Hart, J.D. (1997). *Nonparametric smoothing and lack of fit tests.* Springer Series in Statistics. Springer.
Hart, J.D., Wehrly, T.E. (1992). Kernel regression when the boundary region is large, with an application to testing the adequacy of polynomial models. *Jour. Americ. Statist. Assoc.* **87**, 1018-1024.
Harrison, M.J., McCabe, B.P.M. (1979). A test for heteroscedasticity based on least squares residuals. *J. Am. Statist. Assoc.* **74**, 494-500.
Hodges, J.L., Lehmann, E.L. (1954). Testing the approximative validity of statistical hypotheses. *Jour. Roy. Statist. Soc. Ser.B* **16**, 261268.
Horrowitz, J., Spokoiny, V. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. Preprint, Dept. of Statistics, Iowa State University.
Koenker, R. and Bassett, G. (1981). Robust tests for heteroscedasticity based on regression quantiles. *Econometrika* **50**, 43-61.
Jayashuriya (1996). Testing for polynomial regression using nonparametric regression techniques. *Jour. Amer. Statistic. Assoc.* **91**, 1626-1631.
Kozek, A.S. (1991). A nonparametric fit of a parametric model. *J. Multivariate Analys.* **37**, 6675.
Kuchibhatla, M., Hart, J.D. (1996). Smoothing-based lack-of-fit tests: Variations on a theme. *Nonparametric Statistics* **7**, 1-22.
McBride, G. (1998). Equivalence tests can enhance environmental science and management. *Austral. and New Zealand Journ. Statist.* **41**, 19-29.
McKinnon, J.G. (1992). Model specification tests and artificial regressions. *Jour. of Economic Lit.* **30**, 102-46.
Müller, H.G. (1992). Goodness of fit diagnostics for regression models. *Scand. J. Statist.* **19**, 157-72.
Müller, H.G., Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis.

*Ann. Statist.* **15**, 610 - 635.
Müller, H.G., Zhao, P.L. (1995). On a semi-parametric variance function model and a test for heteroscedasticity. *Ann. Statist.* **23**, 946-967.
Munk, A., Czado, C. (1998). Nonparametric validation of similar distributions and assessment of goodness of fit *Journ. Roy. Statist. Soc. Ser. B*, **60**, 223-241.
Munk, A. (2001a). On a problem in pharmaceutical statistics and the iteration of a peculiar nonlinear operator in the upper complex halfplane. *Nonlinear Analysis* **47**, 1513-23.
Munk, A. (2001b). Testing the goodness of fit with random Toeplitz forms. *Scand. J. Statist.*, to appear.
Munk, A., Wagner. T., Scholz, A., Bissantz, N. (2001). A class of simple tests for heteroscedasticity in regression models based on difference estimators. Technical report, Paderborn University.
Neil, J.W., Johnson, D.E. (1985). Testing linear regression function adequacy without replication. *Ann. Statist.* **13**, 1482-1489.
Neyman, J. (1937). Smooth tests for goodness of fit. *Skandinavisk Aktuarietidskrift* **20**, 149-99.
Rice, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12**, 1215-1230.
Roy, T. (1997). Calibrated nonparametric confidence sets. *Journ. Mathemat. Chemistry* **21**, 103-109.
Sacks, J. and Ylvisacker, D. (1970). Designs for regression problems with correlated errors III. *Ann. Math. Statist.* **41**, 2057-2074.
Seber,G.A.F., Wild, G.J. (1989). *Nonlinear Regression.* Wiley, N.Y.
Staniswalis, J.G., Severini, T.A.(1991). Diagnostics for assessing regression models. *Jour. Amer. Statist. Assoc.* **86**, 684-92.
Stute W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25**, 613-41.
Stute, W., González-Manteiga, W. (1996). NN goodness of fit tests for linear models. *Jour. Statist. Plann. Inf.* **53**, 7592.
Stute,W., González-Manteiga, W., Presedo Quindimil, M. (1998). Bootstrap approximation in model checks for regression. *Journ. Amer. Statist. Assoc.* **93**, 141-149.
von Neumann, J. (1941). Distribution of the ratio of the mean squared succesive difference to the variance. *Ann. Math. Statist.* **12**, 367-395.
von Neumann, J. (1942). A further remark concerning the distribution of the ratio of the mean squared succesive difference to the variance. *Ann. Math. Statist.* **13**, 86-88.
Weirather, G. (1993). Testing a linear regression model against nonparametric alternatives. *Metrika* **40**, 367-379.
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. (1997). A consistent test of a functional form via nonparametric estimation techniques. *Journ. of Econometrics* **75**, 263-289.
Zheng, J.X. (2000). A consistent nonparametric test of heteroscedasticity. Preprint, Department of Economics, University of Texas at Austin.