Optimal designs for testing the functional form of a regression via nonparametric estimation techniques

Stefanie Biedermann
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: Stefanie.Biedermann@ruhr-uni-bochum.de

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: Holger.Dette@ruhr-uni-bochum.de
FAX: +49 2 34 32 14 559

July 21, 2000

Abstract

For the problem of checking linearity in a heteroscedastic nonparametric regression model under a fixed design assumption we study maximin designs which maximize the minimum power of a nonparametric test over a broad class of alternatives from the assumed linear regression model. It is demonstrated that the optimal design depends sensitively on the used estimation technique (i.e. weighted or ordinary least squares) and on an inner product used in the definiton of the class of alternatives. Our results extend and put recent findings of Wiens (1991) in a new light, who established the maximin optimality of the uniform design for lack-of-fit tests in homoscedastic multiple linear regression models.

AMS 1991 Subject classifications: Primary 62K05; 62G10; Secondary 62G20; 62J05
Keywords and phrases: goodness-of-fit test, weighted least squares, optimal design, maximin optimality, D_1-optimality

1 Introduction

Consider the common nonparametric regression model

(1.1) $$Y_{i,n} = m(t_{i,n}) + \varepsilon_{i,n} \qquad i = 1, \ldots, n,$$

where the $\varepsilon_{i,n}$ form a triangular array of rowwise independent zero mean random variables with variance $\mathrm{Var}(\varepsilon_{i,n}) = \sigma^2(t_{i,n})$, $m(\cdot), \sigma^2(\cdot)$ are unknown functions and $\{t_{i,n} \mid i = 1, \ldots, n\}$ is a fixed design in the interval $[0, 1]$. Parametric assumptions for the regression model are attractive among practitioners and much effort has been devoted to the problem of checking

the functional form of the regression $m(\cdot)$ [see e.g. Shillington (1979), Neil and Johnson (1985), Azzalini and Bowman (1993), Härdle and Mammen (1993), González Manteiga and Cao (1993), Zheng (1996), Alcalá, Christóbal, González Manteiga (1999), Dette (1999, 2000) among many others].
In the present paper we study the impact of the underlying design on the asymptotic power of certain kernel based methods for checking linearity which were recently proposed by Azzalini and Bowman (1993), González Manteiga and Cao (1993) and Dette (1999). In Section 2 we investigate the asymptotic behaviour of these procedures under fixed alternatives using similar techniques as proposed by Dette (1999). These results are used for the determination of the (asymptotically) dominating term in the power function, which turns out to depend on the variance and regression function and the underlying design. The maximization of the minimum power (with respect to a certain class of alternatives for variance and regression function) is considered in Section 3 as an optimal design problem. On the one hand this extends an optimality criterion introduced by Wiens (1991) (in a slightly different context) to the heteroscedastic regression model. On the other hand it is demonstrated that the uniform distribution [which was identified by Wiens (1991) as the maximin optimal design for his criterion] is not necessarily a good choice for testing linearity. Its particular optimality properties depend sensitively on the method of estimation (i.e. weighted or unweighted least squares) and on the specific choice of an inner product used in the definition of the class of alternatives.

2 Testing the lack-of-fit in nonparametric regression — a new optimality criterion for the choice of a design

Consider the nonparametric regression model (1.1) and a p-dimensional vector g of linearly independent regression functions $g_1, \dots, g_p$. We are interested in the problem of testing the hypothesis of linearity

$$m \in \mathcal{U}_p := \text{ span}\,\{g_1, \dots, g_p\} \tag{2.1}$$

where $\mathcal{U}_p = \text{ span}\,\{g_1, \dots, g_p\}$ denotes the linear space generated by the functions $g_1, \dots, g_p$. Throughout this paper we assume that the errors in (1.1) have uniformly bounded fourth moments, i.e.

$$E[\varepsilon_{j,n}^4] \le C < \infty \quad j = 1, \dots, n; n \in \mathbb{N}, \tag{2.2}$$

that the design points $\{t_{i,n} \mid i = 1, \dots, n\}$ satisfy a Sacks and Ylvisaker (1970) condition, i.e.

$$\frac{i}{n} = \int_0^{t_{i,n}} f(t)dt \quad i = 1, \dots, n; n \in \mathbb{N}, \tag{2.3}$$

and that all functions appearing in (2.1) and (2.3) are sufficiently smooth, i.e.

$$m,\ f,\ \sigma^2,\ g_1, \dots,\ g_p \in C^{(r)}([0,1]), \tag{2.4}$$

where $r \ge 2$. Let w denote a positive weight function on the interval $[0,1]$ and $\hat{\theta}_n = (\hat{\vartheta}_1, \dots, \hat{\vartheta}_p)^T$ be the weighted least squares estimator for the parameter θ (with respect to the function w) in the linear model $y = \theta^T g(t) + \varepsilon$, where $g(t) = (g_1(t), \dots, g_p(t))^T$.
Our first approach of testing the lack-of-fit of the linear regression was introduced by González Manteiga and Cao (1993) and compares a parametric and nonparametric fit of the regression

curve [see also Härdle and Mammen (1993) for a similar method]. To be precise, let K denote a kernel with compact support, say $[-1,1]$, of order $r \geq 2$, i.e.

$$\frac{(-1)^j}{j!}\int_{-1}^{1} u^j K(u)du = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } 1 \leq j \leq r-1 \\ \kappa_r > 0 & \text{if } j = r. \end{cases} \tag{2.5}$$

and let $h = h_n > 0$ denote a bandwidth satisfying

$$h = O(n^{-\frac{1}{2r}}); \qquad nh^2 \to \infty. \tag{2.6}$$

Defining the Nadaraya-Watson weights [see Nadaraya (1964)]

$$w_{ij} = \frac{K\left(\frac{t_{i,n}-t_{j,n}}{h}\right)}{\sum_{\ell=1}^{n} K\left(\frac{t_{i,n}-t_{\ell,n}}{h}\right)} \quad i,j = 1,\ldots,n, \tag{2.7}$$

we obtain nonparametric estimates of the regression function evaluated at the design points by

$$\hat{m}(t_{i,n}) = \sum_{j=1}^{n} w_{ij} Y_{j,n}. \tag{2.8}$$

González Manteiga and Cao (1993) used

$$T_n^{(1)} = \frac{1}{n}\sum_{i=1}^{n}\Big\{\hat{m}_n(t_{i,n}) - \sum_{j=1}^{p}\hat{\vartheta}_j g_j(t_{i,n})\Big\}^2 w(t_{i,n}) \tag{2.9}$$

as a measure of linearity and showed the consistency of the test which rejects linearity for large values of the statistic $T_n^{(1)}$. Similarly, Azzalini and Bowman (1993) proposed the statistic

$$T_n^{(2)} = \frac{\hat{\varepsilon}^T\hat{\varepsilon} - \hat{\varepsilon}^T M \hat{\varepsilon}}{\hat{\varepsilon}^T M \hat{\varepsilon}} \tag{2.10}$$

for testing the linearity of a homoscedastic nonparametric regression model. Here $\hat{\varepsilon}_i = Y_i - \hat{\theta}_n^T g(t_{i,n})$ $(i = 1,\ldots n)$, $\hat{\varepsilon} = (\sqrt{w(t_{1,n})}\hat{\varepsilon}_1,\ldots,\sqrt{w(t_{n,n})}\hat{\varepsilon}_n)^T$ is the vector of (weighted) residuals formed from a weighted least squares fit, $M = (I_n - W)^T(I_n - W)$ and $W = (w_{ij})_{ij=1}^n$ is the matrix defined by the weights (2.7). It was proved by Dette (2000) that rejecting (2.1) for large values of $T_n^{(2)}$ provides a consistent test for linearity even in the heteroscedastic case.
Our final measure of linearity was introduced by Dette (1999) as a difference of a variance estimator in the linear regression model (2.1) and the nonparametric regression model. More precisely, this author considered the statistic

$$T_n^{(3)} = \hat{\sigma}^2_{LSE} - \hat{\sigma}^2_{HM}, \tag{2.11}$$

where $\hat{\sigma}^2_{LSE}$ is the weighted least squares estimator of the variance in the linear regression model and $\hat{\sigma}^2_{HM}$ is a slightly modified weighted version of the nonparametric estimator introduced by Hall and Marron (1990). This estimator is defined by

$$\hat{\sigma}^2_{HM} = \frac{1}{n}\sum_{i=1}^{n}\Big(Y_{i,n} - \sum\nolimits_{j=1}^{n} w_{ij}Y_{j,n}\Big)^2 w(t_{i,n}),$$

where the weights are given by (2.7). It was shown by González Manteiga and Cao (1993) and Dette (1999, 2000) that under the hypothesis of linearity these statistics are asymptotically normally distributed, that is

$$(2.12) \qquad n\sqrt{h}(T_n^{(j)} - \frac{B_j}{nh}) \to \mathcal{N}(0, \lambda_j^2) \qquad j = 1, 2, 3$$

as $n \to \infty$, where the asymptotic bias and variance are listed in Table 1 for the different cases. Similarly, Dette (1999, 2000) established asymptotic normality for the unweighted (i.e. $w \equiv 1$) versions of $T_n^{(2)}$ and $T_n^{(3)}$ under fixed alternatives. A generalization of these arguments to arbitrary weight functions and to the statistic $T_n^{(1)}$ considered by González Manteiga (1993) yields under the alternative $m \notin \mathcal{U}_p$

$$(2.13) \qquad \sqrt{n}(T_n^{(j)} - \frac{B_j}{nh} - M_j^2) \to \mathcal{N}(0, \mu_j^2) \qquad j = 1, 2, 3,$$

where the asymptotic bias M_j^2 and variance are listed in Table 2 for the different cases.
A straightforward calculation shows that the power function of the test which rejects the hypothesis of linearity for large values of $T_n^{(j)}$ $(j = 1, 2, 3)$, is given by

$$(2.14) \qquad P(\text{"rejection"}) = \Phi\Big\{2\sqrt{n\ \psi^w_{m,\sigma^2}(\xi)}\Big\} + o(1),$$

where ξ is the probability measure with density f [i.e. $d\xi(x) = f(x)dx$], Φ denotes the cdf. of the standard normal distribution,

$$(2.15) \qquad \psi^w_{m,\sigma^2}(\xi) = \frac{(M_j^2)^2}{\mu_j^2} = \frac{[\int_0^1 \Delta^2(x)w(x)d\xi(x)]^2}{\int_0^1 \Delta^2(x)w^2(x)\sigma^2(x)d\xi(x)} = \frac{[\int_0^1 (m - P^{w,\xi}_{\mathcal{U}_p} m)^2(x)w(x)d\xi(x)]^2}{\int_0^1 (m - P^{w,\xi}_{\mathcal{U}_p} m)^2(x)w^2(x)\sigma^2(x)d\xi(x)},$$

$\Delta = m - P^{w,\xi}_{\mathcal{U}_p} m$ and $P^{w,\xi}_{\mathcal{U}_p}$ denotes the orthogonal projection onto the linear space $\mathcal{U}_p$ with respect to the inner product induced by the measure $w(x)d\xi(x)$. Note that the asymptotic power in (2.14) also depends on the weight function w. Our preliminary result identifies an optimal weight function which maximizes ψ^w_{m,σ^2} uniformly over the class of all regression functions and over the class of all designs. The proof is a straightforward application of Cauchy's inequality and therefore omitted.

j	B_j	λ_j^2
1	$t^2 \cdot \int_{-1}^1 K^2(z)dz$	$2\rho^2 \cdot \int_{-2}^2 (K * K)^2(z)dz$
2	$\frac{t^2}{s^2}[2K(0) - \int_{-1}^1 K^2(z)dz]$	$2\frac{\rho^2}{s^4} \cdot \int_{-2}^2 [2K - K * K]^2(z)dz$
3	$t^2[2K(0) - \int_{-1}^1 K^2(z)dz]$	$2\rho^2 \cdot \int_{-2}^2 [2K - K * K]^2(z)dz$

Table 1: *Asymptotic bias and variance of the statistics $T_n^{(j)}$ $(j = 1, 2, 3)$ under the hypothesis of linearity (see (2.12)). Here $s^2 = \int_0^1 \sigma^2(x)w(x)f(x)dx$, $t^2 = \int_0^1 \sigma^2(x)w(x)dx$, $\rho^2 = \int_0^1 \sigma^4(x)w^2(x)dx$ and $K_1 * K_2$ denotes the convolution of K_1 and K_2.*

j	M_j^2	μ_j^2
1	$\int_0^1 \Delta^2(x)w(x)f(x)dx$	$4\int_0^1 \sigma^2(x)w^2(x)\Delta^2(x)f(x)dx$
2	$\frac{1}{s^2}\int_0^1 \Delta^2(x)w(x)f(x)dx$	$\frac{4}{s^4}\int_0^1 \sigma^2(x)w^2(x)\Delta^2(x)f(x)dx$
3	$\int_0^1 \Delta^2(x)w(x)f(x)dx$	$4\int_0^1 \sigma^2(x)w^2(x)\Delta^2(x)f(x)dx$

Table 2: *Asymptotic bias and variance of the statistic $T_n^{(j)}$ $(j = 1, 2, 3)$ under the alternative of nonlinearity [i.e. $m \notin \mathcal{U}_p$], (see (2.13)). Here $s^2 = \int_0^1 \sigma^2(x)w(x)f(x)dx, \Delta = m - P_{\mathcal{U}_p}^{w,\xi}m$ and $P_{\mathcal{U}_p}^{w,\xi}$ denotes the orthogonal projection onto $\mathcal{U}_p$ with respect to the inner product induced by the measure $w(x)d\xi(x) = w(x)f(x)dx$.*

Proposition 1: *Let*

$$w^*(x) = \sigma^{-2}(x) \cdot \Big\{ \int_0^1 \sigma^{-2}(t)dt \Big\}^{-1} \tag{2.16}$$

then for any m, σ^2 and any design ξ (such that the criterion (2.15) is well defined)

$$\sup_w \psi_{m,\sigma^2}^w(\xi) = \psi_{m,\sigma^2}^{w^*}(\xi).$$

Note that Proposition 1 gives a guideline for the choice of the weight function in weighted least squares estimation provided some knowledge about the heteroscedastic structure is available. By using weights which are inverse proportional to the variance the power of the tests for detecting departures from any linear model is uniformly maximized. Especially in the homoscedastic case it is strictly recommended to use ordinary least squares estimation. Note that for the optimal choice (2.16) the criterion (2.15) reduces to the well-known D_1-optimality criterion for estimating the parameter a in the linear model

$$\sigma^{-1}(x)\Big\{\sum_{j=1}^p \vartheta_j g_j(x) + am(x)\Big\}$$

[see Fedorov (1972)], that is

$$\psi_{m,\sigma^2}^{w^*}(\xi) = \int_0^1 (m - P_{\mathcal{U}_p}^{w^*,\xi}m)^2(x)\frac{d\xi(x)}{\sigma^2(x)}. \tag{2.17}$$

This follows readily by a standard result of approximation theory in Hilbert spaces [see e.g. Achieser (1956), p. 15]. For the constant weight function (in other words: for the homoscedastic regression model with unweighted least squares estimation) maximin optimal designs with respect to the criterion (2.17) were discussed by Wiens (1991).
In most cases knowledge about the variance function is not available, so the choice of the weight function for the least squares estimation is not obvious at all. Nevertheless, the asymptotic power in (2.14) is an increasing function of $\psi_{m,\sigma^2}^w(\xi)$, and consequently a good design ξ for testing the lack-of-fit of the linear model should somehow maximize the criterion (2.15). Note that $\psi_{m,\sigma^2}^w(\cdot)$ depends on the particular nonlinear alternative $m(\cdot)$ and the unknown variance

function $\sigma^2(\cdot)$ and is in this sense a local optimality criterion [see Chernoff (1953)]. It is also worthwhile to mention that the local optimality criterion $\log \psi^w_{m,\sigma^2}(\xi)$ is a difference of two D_1-optimality criteria [see Pukelsheim (1993), p. 289 and Kiefer and Wolfowitz (1959)] and for this reason not necessarily concave as a function on the space of all design measures. In general the lack of concavity complicates the determination of optimal designs, because standard techniques cannot be applied [see e.g. Silvey (1980)].

3 Minimax designs — why is the uniform distribution optimal?

In this section we investigate a maximin version of the optimality criterion (2.15). Our approach is similar as in Wiens (1991), but illuminates the particular optimality properties of the uniform distribution from a different point of view. To be precise, let v denote a positive weight function (not necessarily equal to w) and define for $\eta, \varepsilon > 0$

$$\begin{aligned}\mathcal{F} = \Big\{(m,\sigma^2) |\ & \textstyle\int_0^1 m^2(x)v(x)dx \geq \eta^2;\ \sup_{x\in[0,1]} |w(x)\sigma^2(x)| \leq \varepsilon^2; \\ & \textstyle\int_0^1 m(x)g_j(x)v(x)dx = 0; \qquad j=1,\ldots,p\Big\}\end{aligned} \tag{3.1}$$

as the set of pairs (m, σ^2) of regression and variance functions such that the weighted sup-norm of σ^2 is bounded, m is orthogonal to the linear space $\mathcal{U}_p$ spanned by the regression functions (with respect to the inner product induced by the measure $v(x)dx$) and has L^2-norm bounded from below by η^2.

Definition 2: *A design ξ^* is called maximin optimal for testing the lack-of-fit of the linear model $\mathcal{U}_p$, if and only if it maximizes*

$$\inf\{\psi^w_{m,\sigma^2}(\xi) \mid (m,\sigma^2) \in \mathcal{F}\}. \tag{3.2}$$

It might be desirable to use only L^2-norms in the definition of the set $\mathcal{F}$ in (3.2). However, the following result indicates that this is in general impossible.

Proposition 3: *Let $p = 1, g_1(x) = x - \frac{1}{2}$, λ denote the Lebesgue measure and for $\varepsilon, \eta > 0$*

$$\tilde{\mathcal{F}} = \Big\{(m,\sigma^2) \mid \int_0^1 m^2(x)dx \geq \eta^2; \int_0^1 \sigma^2(x)dx \leq \varepsilon^2; \int_0^1 m(x)g_1(x)dx = 0\Big\}.$$

then we have for any design ξ with positive and continuous density in a neighbourhood of the point $x = \frac{1}{2}$:

$$\inf\{\psi^\lambda_{m,\sigma^2}(\xi) \mid (m,\sigma^2) \in \tilde{\mathcal{F}}\} = 0.$$

Proof: Let $h = \frac{d\xi}{dx}$ denote the density of the measure ξ, assume without loss of generality $\int_0^1 g_1^2(t)d\xi(t) = 1$, let $M_n = [\frac{1}{2} - \frac{1}{n}, \frac{1}{2} + \frac{1}{n}]$ and define a regression and variance function by

$$\begin{aligned} m_n(x) &= \eta\sqrt{\frac{n}{2}} I_{M_n}(x) \\ \sigma_n^2(x) &= \varepsilon^2 \frac{n}{2} I_{M_n}(x), \end{aligned}$$

where I_{M_n} denotes the indicator function of the set M_n. Obviously, we have $(m_n, \sigma_n^2) \in \tilde{\mathcal{F}}$ for all $n \in I\!N$, and a straightforward calculation shows that

$$P_{\mathcal{U}_1}^{\lambda,\xi} m(x) = \langle g_1, m\rangle g_1(x), \tag{3.3}$$

where $\langle g_1, m\rangle = \int_0^1 g_1(t)m(t)d\xi(t)$ denotes the inner product with respect to the measure $d\xi(t)$. Observing the definition of m_n, we obtain for the numerator in (2.15) [note that $w = \lambda, p = 1, g_1(t) = t - \frac{1}{2}$]

$$\begin{aligned} &\int_0^1 (m_n - P_{\mathcal{U}_1}^{\lambda,\xi} m_n)^2(x) d\xi(x) = \langle m_n, m_n\rangle - \langle m_n, g_1\rangle^2 \qquad (3.4) \\ &= \frac{n}{2}\eta^2 \Big[\int_{M_n} h(t)dt - \Big\{\int_{M_n} g_1(t)h(t)dt\Big\}^2\Big] \\ &= \eta^2 h(\frac{1}{2}) + o(1), \end{aligned}$$

because of the continuity of the function h in a neighbourhood of the point $x = \frac{1}{2}$. A similar reasoning gives for the denominator

$$\begin{aligned} &\int_0^1 (m_n - P_{\mathcal{U}_1}^{\lambda,\xi} m_n)^2(x)\sigma_n^2(x) d\xi(x) \qquad (3.5) \\ &= \langle m_n, m_n\sigma_n^2\rangle - 2\langle m_n, g_1\rangle\langle m_n, g_1\sigma_n^2\rangle + \langle m_n, g_1\rangle^2\langle g_1^2, \sigma_n^2\rangle \\ &= \frac{n}{2}\eta^2\varepsilon^2 h(\frac{1}{2}) + o(1). \end{aligned}$$

Consequently, we obtain from (3.4) and (3.5)

$$\lim_{n\to\infty} \psi_{m_n,\sigma_n^2}^{\lambda}(\xi) = 0,$$

which proves the assertion.

□

Theorem 4: *Let ξ^* denote the measure corresponding to the density*

$$\frac{v(x)}{w(x)}\Big\{\int_0^1 \frac{v(t)}{w(t)}dt\Big\}^{-1},$$

then ξ^* is maximin optimal for testing the lack-of-fit of the linear model $\mathcal{U}_p$.

Proof: The proof essentially follows by similar arguments as given in Wiens (1991). More precisely, we prove in two steps

$$\text{(1)} \qquad \forall\ \xi \quad \exists\ (m, \sigma^2) \in \mathcal{F}: \quad \psi_{m,\sigma^2}^{w}(\xi) \le \frac{\eta^2}{\varepsilon^2}$$

$$\text{(2)} \qquad \inf_{(m,\sigma^2)\in\mathcal{F}} \psi_{m,\sigma^2}^{w}(\xi^*) = \frac{\eta^2}{\varepsilon^2},$$

where ξ^* denotes the design with density proportional to $\frac{v}{w}$. This establishes maximin optimality of ξ^*. For the sake of a transparent notation assume without loss of generality that v and w are appropriately normalized such that

$$\int_0^1 \frac{v(t)}{w(t)} dt = 1.$$

In order to prove the assertion (2), we note that $P_{\mathcal{U}_p}^{w,\xi^*}$ is the projection with respect to the inner product induced by the measure $w(x)d\xi^*(x) = v(x)dx$ and the definition of the set $\mathcal{F}$ yields $P_{\mathcal{U}_p}^{w,\xi^*} m = 0$ whenever $(m, \sigma^2) \in \mathcal{F}$. For such pairs we have

$$\begin{aligned}
\inf_{(m,\sigma^2)\in\mathcal{F}} \psi_{m,\sigma^2}^w(\xi^*) &= \inf_{(m,\sigma^2)\in\mathcal{F}} \frac{[\int_0^1 m^2(x)v(x)dx]^2}{\int_0^1 m^2(x)v(x)w(x)\sigma^2(x)dx} \\
&\geq \inf_{(m,\sigma^2)\in\mathcal{F}} \frac{1}{\varepsilon^2} \int_0^1 m^2(x)v(x)dx = \frac{\eta^2}{\varepsilon^2}.
\end{aligned}$$

On the other hand it follows for the variance function $\tau^2(x) := \frac{\varepsilon^2}{w(x)}$

$$\begin{aligned}
\inf_{(m,\sigma^2)\in\mathcal{F}} \psi_{m,\sigma^2}^w(\xi^*) &\leq \inf_{(m,\sigma^2)\in\mathcal{F}} \psi_{m,\tau^2}^w(\xi^*) \\
&= \frac{1}{\varepsilon^2} \inf_{(m,\sigma^2)\in\mathcal{F}} \int_0^1 m^2(x)v(x)dx = \frac{\eta^2}{\varepsilon^2},
\end{aligned}$$

which establishes assertion (2). In order to prove (1), let μ denote the signed measure $\mu = \xi - \xi^*$. By the Hahn decomposition theorem there exists a measurable set A_1 with $\xi^*(A_1) > 0$ and $\mu(A_1) > 0, \mu(A_1^c) = -\mu(A_1) < 0$. Obviously $\xi^*(A_1^c) = (\xi - \mu)(A_1^c) > \xi(A_1^c) \geq 0$, which implies the existence of sets $A_2, \ldots, A_{p+2}$ satisfying $\xi^*(A_j) > 0$ such that

$$A_1^c = \sum_{j=2}^{p+2} A_j \quad \text{and} \quad \mu(A_j) \leq 0 \quad j = 2, \ldots, p+2.$$

Let $\alpha = (\alpha_2, \ldots, \alpha_{p+2})^T$ denote a nontrivial solution of

$$\sum_{j=2}^{p+2} \alpha_j \int_{A_j} g_k(x)v(x)dx \;=\; 0 \quad (k = 1, \ldots, p),$$

define

$$m_*(x) := \eta \sum_{j=2}^{p+2} \alpha_j I_{A_j}(x) \cdot \Big\{ \sum_{j=2}^{p+2} \alpha_j^2 \rho(A_j) \Big\}^{-1/2},$$

where ρ is the measure corresponding to the density v (with respect to the Lebesgue measure) and consider the variance function

$$\sigma_*^2(x) := \frac{\varepsilon^2}{w(x)}. \tag{3.6}$$

A straightforward calculation shows that $(m_*, \sigma_*^2) \in \mathcal{F}$, and

$$P_{\mathcal{U}_p}^{w,\xi^*} m_* = P_{\mathcal{U}_p}^{v,\lambda} m_* = 0, \tag{3.7}$$

which yields

$$\psi^{w}_{m_*,\sigma_*^2}(\xi^*) = \frac{1}{\varepsilon^2}\int_0^1 m_*^2(x)v(x)dx = \frac{\eta^2}{\varepsilon^2}. \tag{3.8}$$

On the other hand we have from (3.6) - (3.8)

$$\begin{aligned}
\psi^{w}_{m_*,\sigma_*^2}(\xi) - \psi^{w}_{m_*,\sigma_*^2}(\xi^*) &= \frac{1}{\varepsilon^2}\Big\{\int_0^1 (m_* - P^{w,\xi}_{\mathcal{U}_p} m_*)^2(x)w(x)d\xi(x) - \int_0^1 m_*^2(x)w(x)d\xi^*(x)\Big\} \\
&\le \frac{1}{\varepsilon^2}\int_0^1 m_*^2(x)w(x)d\mu(x) = \frac{\eta^2}{\varepsilon^2}\frac{\sum_{j=2}^{p+2}\alpha_j^2\int_{A_j} w(x)d\mu(x)}{\sum_{j=2}^{p+2}\alpha_j^2\,\rho(A_j)} \le 0. \qquad (3.9)
\end{aligned}$$

A combination of (3.8) and (3.9) yields (1) and completes the proof of the theorem.

□

Note that Theorem 4 contains the result of Wiens (1991) as a special case. More precisely, this author considered the unweighted least squares estimation (i.e. $w(x) \equiv 1$) in a homoscedastic regression for which the criterion reduces to

$$\psi^{\lambda}_{m,\sigma^2}(\xi) = \frac{1}{\sigma^2}\int_0^1 (m - \mathcal{P}^{\lambda,\xi}_{\mathcal{U}_p} m)^2(x)d\xi(x).$$

For the set

$$\mathcal{F}^* = \Big\{m \mid \int_0^1 m^2(x)dx \ge \eta^2;\ \int_0^1 m(x)g_j(x)dx = 0; \quad j = 1,\ldots,p\Big\} \tag{3.10}$$

Wiens (1991) proved the maximin optimality of the uniform design. A careful inspection of the proof of Theorem 4 shows that the design with density v is maximin optimal, if the Lebesgue measure dx is replaced by the measure $v(x)dx$ in the definition (3.10) of the inner product in the class of alternatives $\mathcal{F}^*$. In other words the optimality property of a particular design with respect to Wiens (1991) maximin criterion is intimately related to the particular measure used in the definition of the set $\mathcal{F}^*$.
We conclude with the discussion of a further important special case which uses the optimal weight function defined in (2.16). The proof is similar to Theorem 4 and therefore omitted.

Corollary 5: *The design which maximizes*

$$\inf\Big\{\int_0^1 (m - P^{\sigma^{-2},\xi}_{\mathcal{U}_p} m)^2(x)\frac{d\xi(x)}{\sigma^2(x)} \mid (m,\sigma^2) \in \mathcal{F}\Big\}$$

is given by the measure

$$\frac{v(x)dx}{\int_0^1 v(t)dt}.$$

Acknowledgements: The authors would like to thank I. Gottschlich for typing most parts of this paper with considerable technical expertise. The financial support of the Deutsche

Forschungsgemeinschaft (SFB 475, Reduction of complexity in multivariate data structures) is gratefully acknowledged.

References

Achieser, N.J. (1956). Theory of Approximation. Dover, N.Y.

Alcalá, J.T., Christóbal, J.A., González Manteiga, W. (1999). Goodness-of-fit test for linear models based on local polynomials. Statist. & Probab. Letters 42, 39-46.

Azzalini, A., Bowman, A. (1993). On the use of nonparametric regression for checking linear relationships. J. Roy. Statist. Soc., Ser. B., 55, 549-559.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. Ann. Math. Statist. 24, 589-602.

Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. Ann. Statist., 27, 1012-1050.

Dette, H. (2000). On a nonparametric test for linear relationships. Statist. Probab. Lett. 42, 39-46.

Fedorov, V.V. (1972). Theory of optimal experiments. Academic Press, N.Y.

González Manteiga, W., Cao, R. (1993). Testing hypothesis of general linear model using nonparametric regression estimation. TEST 2, 161-189.

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

Hall, P., Marron, J.S. (1990). On variance estimation in nonparametric regression. Biometrika 77, 415-419.

Kiefer, J.C., Wolfowitz, J. (1959). Optimum designs in regression problems. Ann. Math. Statist. 30, 271-294.

Nadaraya, E.A. (1964). On estimating regression. Theory Probab. Appl., 10, 186-190.

Neil, J.W., Johnson, D.E. (1985). Testing linear regression function adequacy without replication. Ann. Statist., 13, 1482-1489.

Pukelsheim, F. (1993). Optimal design of experiment. Wiley, N.Y.

Sacks, J., Ylvisaker, D. (1970). Designs for regression problems for correlated errors. Ann. Math. Statist. 41, 2057-2074.

Shillington, E.R. (1979). Testing lack-of-fit in regression without replication. The Canad. J. of Statist., 7, 137-146.

Silvey, S. (1980). Optimal design. Chapman and Hall, London.

Wiens, D. (1991). Designs for approximately linear regression: two optimality properties of uniform designs. Statist. Probab. Lett. 12, 217-221.

Zheng, J.X. (1996). A consistent test of a functional form via nonparametric estimation techniques. J. of Econometrics 75, 263-289.