Designing Experiments for Adaptively Fitted Models

Norbert Benda
Rainer Schwabe

ABSTRACT: Stepwise regression is a commonly used tool if the order of the linear model to be fitted is not known exactly. In this note strategies are proposed for an optimal choice of both the design and the decision rule which selects the adequate model.

KEYWORDS: Pre–test Estimator; G–optimal Design; Minimax Design Criterion; Stepwise Regression; Efficiency Constraint

1 Introduction

The outcomes of statistical experiments are usually described by the mean response $E(Y(x))$ which expresses the dependence on one or several explanatory variables x. To obtain an appropriate fit – and to ease the interpretations – usually a low degree linear model is fitted arising from a suitable approximation. For example, polynomial approximations are obtained by Taylor expansions. In practice, most commonly a linear or quadratic fit is selected.

Low order approximations are desirable due to their simplicity. However, they may fail to fit all characteristics of the response. Therefore, one is tempted to use a higher order approximation which includes the lower order functions as particular cases. The payoff for a better fit to the observed data is a higher variability in the estimated response due to observational errors which may cover the general structures by noise. To avoid the effects of over–fitting one can choose a lower order model as long as no *significant* deviations are detected and switch to a higher order model otherwise. This kind of pre–test estimation has been treated in linear models e. g. by Judge and Bock (1978), Giles (1991), Boscher (1991), Droge (1993) and Benda (1996), and it is implemented in many statistical software packages as *stepwise regression* procedures.

In the present paper we consider the situation that a higher order linear model truely describes the response while for reasons of simplicity and ac-

curacy it would be desirable to fit a lower order linear model if appropriate. Under these assumptions Judge and Bock (1978) gave a detailed prescription and derived explicit expressions for the mean squared error matrix.
The performance of the pre–test estimator depends on the decision rule which selects the fitted model, on the design according to which the explanatory variables are chosen and on the magnitude of the deviation from the lower order model. In this note the performance is measured by the maximal mean squared error in prediction. A minimax approach is used to get rid of the influence of the actual shape of the response. Two strategies are proposed for obtaining suitable pairs of designs and decision rules. First, following the common idea in stepwise regression the decison rule is fixed and the design is chosen to minimize the overall maximal mean squared error. This may lead to unsatisfactory results. Another strategy introduced by Benda (1996) is based on the idea of achieving a prespecified efficiency level in case the lower order model is true. Under this constraint the design and the decision rule have to be chosen simultaneously.

2 The Pre-test Estimator

We assume that the observations $Y_1, \ldots, Y_N$ are properly described by a linear model

$$Y_n = \sum_{i=1}^{p} f_i(x_n)\beta_i + Z_n = \boldsymbol{f}(x_n)^\top \boldsymbol{\beta} + Z_n \tag{1}$$

where $x_1, \ldots, x_N \in \mathcal{X}$ are the settings of the explanatory variables, $\boldsymbol{f} = (f_1, \ldots, f_p)^\top$ is a set of known regression function, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is the vector of unknown parameters and the Z_n denote uncorrelated homoscedastic random errors. Then the vector of observations $\boldsymbol{Y} = (Y_1, \ldots, Y_N)^\top$ can be written as

$$\boldsymbol{Y} = \boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{Z} \tag{2}$$

where $\boldsymbol{F} = (f_i(x_n))_{n=1,\ldots,N}^{i=1,\ldots,p}$ is the design matrix associated with the design $\boldsymbol{X} = (x_1, \ldots, x_N) \in \mathcal{X}^N$ and $\boldsymbol{Z} = (Z_1, \ldots, Z_N)^\top$ is the error vector, $E(\boldsymbol{Z}) = 0$, $Cov(\boldsymbol{Z}) = \sigma^2 \boldsymbol{I}d_N$.
Although (1) holds true we may be willing to assume that a proper submodel $E(Y_n) = \boldsymbol{f}_0(x_n)^\top \boldsymbol{\beta}_0$ suffices to describe the mean response $E(Y_n) = \boldsymbol{f}(x_n)^\top \boldsymbol{\beta}$, where the regression function $\boldsymbol{f}_0$ of the submodel lies in the function space spanned by $\boldsymbol{f}$, i.e. $\boldsymbol{f}_0 = \boldsymbol{C}_0 \boldsymbol{f}$ for some $(p-q)\times p$ selection matrix $\boldsymbol{C}_0$. Alternatively this can be described by an additional linear constraint $\boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{0}$ on the parameter vector, where $\boldsymbol{L}$ in a $q \times p$ matrix of rank q.
For example if the observations Y are described by a polynomial model $E(Y) = \sum_{i=1}^{p} x^{i-1}\beta_i$ of degree $p-1$ then we may hope that also a polynomial $\sum_{i=1}^{p-q} x^{i-1}\beta_i$ of lower degree $p-q-1$ may be appropriate. In this case the constraint is given by $\boldsymbol{L} = (\boldsymbol{0}, \boldsymbol{I}d_q)$.

In the present situation it is reasonable, first to test whether the hypothesis $\boldsymbol{L\beta} = \mathbf{0}$ is true and then to estimate the response $E(Y)$ by the least squares estimators in model (1) if $\boldsymbol{L\beta} = \mathbf{0}$ is rejected and by the constrained least squares estimator satisfying $\boldsymbol{L}\widehat{\boldsymbol{\beta}} = \mathbf{0}$ if $\boldsymbol{L\beta} = \mathbf{0}$ is accepted.
In the polynomial regression example we fit a polynomial of degree $p-q-1$ if the test confirms that this lower degree polynomial is adequate, and we fit a polynomial of degree $p-1$ otherwise.
For notational convenience let $\boldsymbol{F}$ be of full column rank $p \leq N$. Denote by $\widehat{\boldsymbol{\beta}}^* = (\boldsymbol{F}^\top \boldsymbol{F})^{-1}\boldsymbol{F}^\top \boldsymbol{Y}$ the least squares estimator in the unconstrained model (1) and by $\widehat{\boldsymbol{\beta}}_\infty = (\boldsymbol{F}_0^\top \boldsymbol{F}_0)^{-1}\boldsymbol{F}_0^\top \boldsymbol{Y}$ the least squares estimator in the constrained model, satisfying $\boldsymbol{L}\widehat{\boldsymbol{\beta}}_\infty = \mathbf{0}$, where $\boldsymbol{F}_0 = (f_{0i}(x_n))_{n=1,\ldots,N}^{i=1,\ldots,p_0}$ is the design matrix in the constrained model. Then the pre-test estimator will be defined by

$$\widehat{\boldsymbol{\beta}}_c = \begin{cases} \widehat{\boldsymbol{\beta}}^* & \text{if} \quad U > c \\ \widehat{\boldsymbol{\beta}}_\infty & \text{if} \quad U \leq c \end{cases} \tag{3}$$

where

$$U = \frac{N-p}{q} \, \frac{\boldsymbol{Y}^\top \boldsymbol{L}^\top (\boldsymbol{L}(\boldsymbol{F}^\top \boldsymbol{F})^{-1}\boldsymbol{L}^\top)^{-1}\boldsymbol{LY}}{\boldsymbol{Y}^\top (\boldsymbol{I}d_n - \boldsymbol{F}(\boldsymbol{F}^\top \boldsymbol{F})^{-1}\boldsymbol{F}^\top)\boldsymbol{Y}} \tag{4}$$

is the statistic of the F-test for testing the hypothesis $\boldsymbol{L\beta} = \mathbf{0}$, and c is a prespecified critical value.
This critical value c which governs the pre-test estimator $\widehat{\boldsymbol{\beta}}_c$ will be determined either by a fixed significance level α, i.e. c is the $(1-\alpha)$ quantile of the appropriate F-distribution, which is common use in the implemented statistics software; or c may be chosen in order to guarantee a prespecified precision in the constrained model (Benda, 1996). Alternatively, c may also be chosen to minimize the regret (Droge, 1993).
All these approaches require normality of the random errors $\boldsymbol{Z}$ which we will assume throughout the paper.

3 Asymptotic Risk

For the mean squared error matrix $E((\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta})^\top)$ explicit formulae have been given by Judge and Bock (1978) under the assumption of normal errors. Starting from their results Benda (1996) derived an asymptotic representation for large sample sizes N which is well adapted to the theory of generalized designs interpreted as discrete design measures ξ on the desgin region $\mathcal{X}$.
We will confine to the situation of a one-dimensional constraint $\boldsymbol{l}^\top \boldsymbol{\beta} = 0$,

i.e. $q = 1$. In that case the standardized asymptotic risk matrix is given by

$$\sigma^{-2}\boldsymbol{R}(c,\xi;\boldsymbol{\beta}) = \boldsymbol{I}(\xi)^{-1} + \left[g_c\left(\frac{(\boldsymbol{l}^\top\boldsymbol{\beta})^2}{2\sigma^2\boldsymbol{l}^\top\boldsymbol{I}(\xi)^{-1}\boldsymbol{l}} \right) - 1 \right] \frac{\boldsymbol{I}(\xi)^{-1}\boldsymbol{l}\boldsymbol{l}^\top\boldsymbol{I}(\xi^{-1})}{\boldsymbol{l}^\top\boldsymbol{I}(\xi)^{-1}\boldsymbol{l}} \tag{5}$$

where the information matrix $\boldsymbol{I}(\xi) = \int \boldsymbol{f}(x)\boldsymbol{f}(x)^\top\xi(dx)$ of the generalized design ξ is assumed to be regular and g_c is defined by

$$g_c(\eta) = 1 - F_{\chi_3^2,\eta}(c) + \left(2F_{\chi_3^2,\eta}(c) - F_{\chi_5^2,\eta}(c)\right) 2\eta \tag{6}$$

with $F_{\chi_\nu^2,\eta}$ being the distribution function of the χ^2-distribution with ν degrees of freedom and noncentrality parameter η (see Benda, 1996, for details, in particular, on the properties of the function $g_c(\eta)$).
Here we will consider the case that $\boldsymbol{l} = (0,\ldots,0,1)^\top = \boldsymbol{e}_p$ is the pth unit vector, i.e. we check whether the highest degree parameter β_p equals 0 or not. Hence, we test the model (1) against the proper submodel

$$Y = \sum_{i=1}^{p-1} f_i(x)\beta_i + Z\,. \tag{7}$$

Denote by $\boldsymbol{f}_0 = (f_1,\ldots,f_{p-1})^\top$ and $\boldsymbol{I}_0(\xi) = \int \boldsymbol{f}_0\boldsymbol{f}_0^\top d\xi$ the regression function and the information matrix in the constrained model (7), respectively. Then the standardized asymptotic risk can be rewritten as

$$\sigma^{-2}\boldsymbol{R}(c,\xi;\boldsymbol{\beta}) = \begin{pmatrix} \boldsymbol{I}_0(\xi)^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & 0 \end{pmatrix} + g_c(\eta)\frac{\boldsymbol{I}(\xi)^{-1}\boldsymbol{e}_p\boldsymbol{e}_p^\top\boldsymbol{I}(\xi)^{-1}}{\boldsymbol{e}_p^\top\boldsymbol{I}(\xi)\boldsymbol{e}_p} \tag{8}$$

$$= \begin{pmatrix} \boldsymbol{I}_0(\xi)^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & 0 \end{pmatrix} + g_c(\eta)\boldsymbol{e}_p^\top\boldsymbol{I}(\xi)^{-1}\boldsymbol{e}_p \times$$

$$\times \begin{pmatrix} \boldsymbol{I}_0(\xi)^{-1}\boldsymbol{I}_{01}(\xi)\boldsymbol{I}_{01}(\xi)^\top\boldsymbol{I}_0(\xi)^{-1} & -\boldsymbol{I}_0(\xi)^{-1}\boldsymbol{I}_{01}(\xi) \\ -\boldsymbol{I}_{01}(\xi)^\top\boldsymbol{I}_0(\xi)^{-1} & 1 \end{pmatrix}$$

where $\boldsymbol{I}_{01}(\xi) = \int \boldsymbol{f}_0 f_p d\xi$ and $\eta = \beta_p^2/(2\sigma^2\boldsymbol{e}_p^\top\boldsymbol{I}(\xi)^{-1}\boldsymbol{e}_p)$ is the noncentrality parameter.

4 The G-criterion

Reasonable optimality criteria are aiming at minimizing the mean squared error $E((\boldsymbol{f}(x)^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}))^2)$ of the prediction $\boldsymbol{f}(x)^\top\widehat{\boldsymbol{\beta}}$ compared to the true mean response $\boldsymbol{f}(x)^\top\boldsymbol{\beta}$. In this note we will lay emphasis on the G-criterion of minimizing the maximal variance

$$\sup_{x\in\mathcal{X}} E((\boldsymbol{f}(x)^\top(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}))^2) = \sup_{x\in\mathcal{X}} \boldsymbol{f}(x)^\top E((\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta})^\top)\boldsymbol{f}(x) \tag{9}$$

For this minimax criterion on the design region $\mathcal{X}$ criterion the standardized asymptotic risk becomes

$$\begin{aligned}&\sup_{x\in\mathcal{X}} \sigma^{-2}\boldsymbol{f}(x)^\top \boldsymbol{R}(c,\xi;\boldsymbol{\beta})\boldsymbol{f}(x)\\ &= \sup_{x\in\mathcal{X}}\left[\boldsymbol{f}_0(x)^\top \boldsymbol{I}_0(\xi)^{-1}\boldsymbol{f}_0(x) + g_c(\eta)\frac{(\boldsymbol{f}(x)^\top \boldsymbol{I}(\xi)^{-1}\boldsymbol{e}_p)^2}{\boldsymbol{e}_p^\top \boldsymbol{I}(\xi)^{-1}\boldsymbol{e}_p}\right] \qquad (10)\end{aligned}$$

The noncentrality parameter $\eta = (\boldsymbol{l}^\top\boldsymbol{\beta})^2/(2\sigma^2\boldsymbol{l}^\top\boldsymbol{I}(\xi)^{-1}\boldsymbol{l})$ is a function of the unknown parameters $\boldsymbol{\beta}$ and σ^2 and, hence, the G-criterion given by (10) is local. For example, if the constrained model is true we obtain for the standardized asymptotic risk

$$\sup_{x\in\mathcal{X}} \sigma^{-2}\boldsymbol{f}(x)^\top \boldsymbol{R}(c,\xi;\boldsymbol{\beta})\boldsymbol{f}(x) = \sup_{x\in\mathcal{X}} d_0(\xi,c;x) \qquad (11)$$

uniformly in $\boldsymbol{L\beta}=\boldsymbol{0}$, where

$$d_0(\xi,c;x) = \boldsymbol{f}_0(x)^\top \boldsymbol{I}_0(\xi)^{-1}\boldsymbol{f}_0(x) + \left(1 - F_{\chi^2_{3,0}}(c)\right)\frac{(\boldsymbol{f}(x)^\top \boldsymbol{I}(\xi)^{-1}\boldsymbol{e}_p)^2}{\boldsymbol{e}_p\boldsymbol{I}(\xi)^{-1}\boldsymbol{e}_p} \qquad (12)$$

Following an idea in Benda (1996) we propose a minimax approach which aims at minimizing

$$\sup_{\boldsymbol{\beta}\in I\!R^p}\sup_{x\in\mathcal{X}} \sigma^{-2}\boldsymbol{f}(x)^\top \boldsymbol{R}(c,\xi;\boldsymbol{\beta})\boldsymbol{f}(x) = \sup_{x\in\mathcal{X}} d(\xi,c;x) \qquad (13)$$

to avoid the parameter dependence, where

$$\begin{aligned} d(\xi,c;x) &= \boldsymbol{f}_0(x)^\top \boldsymbol{I}_0(\xi)^{-1}\boldsymbol{f}_0(x) + m(c)\frac{(\boldsymbol{f}(x)^\top \boldsymbol{I}(\xi)^{-1}\boldsymbol{e}_p)^2}{\boldsymbol{e}_p\boldsymbol{I}(\xi)^{-1}\boldsymbol{e}_p} && (14)\\ &= \boldsymbol{f}_0(x)^\top \boldsymbol{I}_0(\xi)^{-1}\boldsymbol{f}_0(x) && (15)\\ &\quad + m(c)\boldsymbol{e}_p^\top\boldsymbol{I}(\xi)^{-1}\boldsymbol{e}_p\left(\boldsymbol{f}_0(x)^\top\boldsymbol{I}_0(\xi)^{-1}\boldsymbol{I}_{01}(\xi) - f_p(x)\right)^2 \end{aligned}$$

is the maximum (in $\boldsymbol{\beta}$) mean squared error function and

$$m(c) = \sup_{\eta\geq 0} g_c(\eta) \qquad (16)$$

is finite and bounded from below by 1 according to Lemma 2 in Benda (1996).

Note that, like all criteria based on the mean squared error of the prediction, both the local and global G-criteria (10) and (13) are invariant with respect to a reparametrization of the model induced by a regular linear transformation $\boldsymbol{f}\to\boldsymbol{Qf}$ of the regression functions, subject to the condition that the constraint is not affected by the transformation, i.e. $\boldsymbol{Ql}=\boldsymbol{l}$. More generally, the minimax G-criterion (13) remains equivariant if the constraint is linearly transformed, i.e. $\boldsymbol{Ql}=\gamma\boldsymbol{l}$ for some non-zero constant γ. (For a treatment of equivariance under constraints we refer e.g. to section 3 in Schwabe, 1996.)

5 Polynomial Regression

In a polynomial regression model $Y(x) = \sum_{i=1}^{p} \beta_i x^{i-1} + Z$ of degree $p-1$ on the standardized design region $\mathcal{X} = [-1, 1]$ the maximum mean squared error function $d(\xi, c; x)$ is a polynomial of degree $2(p-1)$ in the variable x. Denote by $m_i = \int x^i \xi(dx)$, $i = 0, \ldots, 2p-2$, the moments of the design ξ interpreted as a design measure, such that the (i,j)th entry of the information matrix $\boldsymbol{I}(\xi)$ becomes m_{i+j-2}.
Now, for every x and every critical value c fixed, we observe by representation (15) that $d(\xi, c; x)$ is minimized within the class of designs with given lower moments m_i, $i = 1, \ldots, 2p-3$, if the highest moment m_{2p-2} is as large as possible. By the theory of canonical moments (see Karlin and Studden, 1966) the highest order moment m_{2p-2} attains its maximum for prespecified lower moments m_i, $i < 2p-2$, if ξ is concentrated on p different settings for the design variable x including both endpoints -1 and 1 of the interval $\mathcal{X}$.
Hence, every design ξ is uniformly (in x) dominated by a p-point design, and we can confine our search for an optimal design to that class of designs which are supported by $\{-1, x_2, \ldots, x_{p-1}, 1\}$ for some $x_2, \ldots, x_{p-1} \in \mathcal{X}$.

Example 1. For constant versus linear regression, $p = 2$, the unique optimal design assigns equal weights $\frac{1}{2}$ to both endpoints -1 and 1, irrespectively which critical value c is used.

The global G-criterion (13) is invariant with respect to sign change ($x \to -x$). It is, thus, reasonabale that optimal designs should share this symmetry property. Again, every symmetric design is dominated by a symmetric p–point design supported by $\{-1, -x_{m-1}, \ldots, -x_1, x_1, \ldots, x_{m-1}, 1\}$ if $p = 2m$ is even and by $\{-1, -x_{m-1}, \ldots, -x_1, 0, x_1, \ldots, x_{m-1}, 1\}$ if $p = 2m+1$ is odd, respectively, for some $x_1, \ldots, x_{m-1} \in (0, 1)$.

Example 2. For linear versus quadratic regression, $p = 3$, the best symmetric design ξ_w is concentrated on $\{-1, 0, 1\}$ and assigns equal weights $\frac{1}{2}w$ to each endpoint -1 and 1 and the remaining weight $1 - w$ to the midpoint 0 of the interval. The predicted variance $d(\xi_w, c; x)$ is a symmetric 4th order polynomial in x and, hence, attains its maximum at either $x = 0$ or $x = 1$ depending on whether $w \le w_1$ or $w \ge w_1$, where $w_1 = \frac{m(c)+1}{2m(c)+1}$. The variance $d(\xi_w, c; 0)$ is decreasing while $d(\xi_w, c; 1)$ is increasing in w. Thus, for a prespecified significance level α, i.e. a prespecified critical value c satisfying $F_{\chi_1^2, 0}(c) = 1 - \alpha$, the optimal weight $w^* = w^*(c)$ is given by $w^* = w_1 = \frac{m(c)+1}{2m(c)+1}$. The corresponding maximal variance $\sup_{x \in \mathcal{X}} d(\xi_{w^*(c)}, c; x)$ equals $m(c) + 2$.
For various critical values c corresponding to commonly used significance levels α the optimal weights w^* are listed together with $\sup_{x \in \mathcal{X}} d(\xi_{w^*}, c; x)$, the value of the global G-criterion (13), and $\sup_{x \in \mathcal{X}} d_0(\xi_{w^*}, c; x)$, the standardized maximal mean squared error (11) if the constrained model is

α	c	$w^*(c)$	$\sup_{x\in\mathcal{X}} d(\xi_{w^*}, c; x)$	$\sup_{x\in\mathcal{X}} d_0(\xi_{w^*}, c; x)$
0	∞	1	∞	2.00
0.005	7.88	0.53	10.31	2.94
0.01	6.63	0.53	9.00	2.95
0.025	5.02	0.54	7.35	2.99
0.05	3.84	0.55	6.18	3.03
0.10	2.71	0.57	5.09	3.09
0.25	1.32	0.61	3.85	3.12
0.50	0.45	0.65	3.20	3.05
0.75	0.10	0.66	3.02	3.01
1	0	0.67	3.00	3.00

TABLE 1. linear versus quadratic regression: prespecified significance levels

true, $\beta_p = 0$, in Table 1. It is striking that the entries for the standardized maximal mean squared error $\sup_{x\in\mathcal{X}} d_0(\xi_{w^*}, c; x) = d_0(\xi_{w^*}, c; 1) = \frac{1}{w}(2 - (1-w)F_{\chi^2_{3,0}}(c))$ corresponding to the true constrained model are very close or even exceed the threshold $\sup_{x\in\mathcal{X}} d(\xi^*, 0; x) = 3$ which is the standardized maximal mean squared error obtained if the G–optimal design ξ^* and the corresponding least squares estimator $\widehat{\boldsymbol{\beta}}^*$ in the unconstrained model are used. Thus, the performance of this strategy of pre–test estimation is unsatisfactory.

Note also that the optimal weight w^* is decreasing in c from the G–optimal weight $w^*(0) = \frac{2}{3}$ in the unconstrained model to the weight $\frac{1}{2}$ which is optimal for β_2 and provides the best *model discrimination*. In particular, the weight $w^*(c)$ does not tend to the optimal weight $w^*(\infty) = 1$ in the constrained model, as the critical value c approaches infinity.

6 Optimal Critical Values

As can be seen from Table 1 the standardized maximal mean squared error of prediction $\sup_{x\in\mathcal{X}} d(\xi^*(c), c; x)$ of the best design $\xi^* = \xi_{w^*(c)}$ substantially depends on the choice of the critical value c. Therefore the performance of the estimator/design pair can be improved by also choosing the critical value c simultaneously with ξ.

If there are no further restrictions it is straightforward to use $c^* = 0$ as $m(c)$ attains its unique minimum at $c = 0$. Thus, the least squares estimator $\widehat{\boldsymbol{\beta}}^*$ for the unconstrained model minimizes the maximal mean squared error of prediction and the design problem is completely solved.

Alternatively, we impose an efficiency constraint $\sup_{x\in\mathcal{X}} d_0(\xi, c; x) = r_0$ on the precision which guarantees a prespecified performance in case the constrained model is true.

To be more specific, let us recall some general results in the theory of optimal designs (for further readings on this topic we refer to Atkinson and Donev, 1992, and Pukelsheim, 1993). If ξ^* and ξ_0^* are G–optimal in the full model or in the constrained model, respectively, and if the associated least squares estimators $\widehat{\boldsymbol{\beta}}^*$ resp. $\widehat{\boldsymbol{\beta}}_\infty$ are applied, then the standardized maximal mean squared errors equal the corresponding numbers of parameters in the models, i.e. $\sup_{x\in\mathcal{X}} d(\xi^*, 0; x) = \sup_{x\in\mathcal{X}} d_0(\xi^*, 0; x) = p$ and $\sup_{x\in\mathcal{X}} d_0(\xi_0^*, \infty; x) = p - 1$.

We will lay more emphasis on the performance of the estimators $\widehat{\boldsymbol{\beta}}_c$ if the constrained model is valid. To preserve a better performance we specify a bound r_0 on the standardized asymptotic mean squared error d_0, satisfying $p-1 < r_0 < p$, which has to be attained by the competing estimator/design pairs (c, ξ). Within this class where $\sup_{x\in\mathcal{X}} d_0(\xi, c; x) = r_0$ we are looking for the estimator/design pair (ξ^*, c^*) which minimizes the unconstrained standardized maximal mean squared error $\sup_{x\in\mathcal{X}} d(\xi, c; x)$. This pair will be called G–optimal under the efficiency constraint.

Example 3. For constant versus linear regression, $p = 2$, ξ^* assigns equal weights $\frac{1}{2}$ to both endpoints -1 and 1, and c^* is the $(2 - r_0)$–quantile of the χ^2–distribution with 3 degrees of freedom, $F_{\chi^2_{3,0}}(c^*) = 2 - r_0$.

Example 4. For linear versus quadratic regression, $p = 3$, the situation is more complicated. Again, we can confine to the symmetric 3–point designs ξ_w which assign weights $\frac{1}{2}w$ to 1 and -1 and weight $1 - w$ to 0. Because of $d_0(\xi, c; 0) = 1 + \frac{w}{1-w}(1 - F_{\chi^2_{3,0}}(c))$ and $d_0(\xi, c; 1) = 1 + \frac{1}{w} + \frac{1-w}{w}(1 - F_{\chi^2_{3,0}}(c))$ the standardized asymptotic mean squared error $d_0(\xi, c; x)$ attains its maximum at $x = 0$ if $F_{\chi^2_{3,0}}(c) \leq k(w) = \frac{3w-2}{2w-1}$ or at $x = 1$ if $F_{\chi^2_{3,0}}(c) \geq k(w)$. Hence, to meet the efficiency constraint the critical value $c = c(w)$ has to be chosen in dependence on the weight w according to

$$F_{\chi^2_{3,0}}(c) = \begin{cases} \frac{1-(1-w)r_0}{w} & \text{if} \quad w \geq w_0 \\ \frac{2-r_0 w}{1-w} & \text{if} \quad w \leq w_0 \end{cases} \tag{17}$$

where $w_0 = \frac{r_0-1}{2r_0-3}$. Denote by $r(w) = \sup_{x\in\mathcal{X}} d(\xi_w, c(w); x)$ the maximal standardized mean squared error in the unconstrained model which has to be minimized.

For $w \geq w_0$ we observe that $m(c(w))$ is increasing in the weight w and, hence, $r(w) = 1 + \frac{w}{1-w}m(c(w))$ also increases in w. For the alternate case, $w \leq w_0$, $c(w)$ and, hence, $m(c)$ is decreasing in w. Thus, as long as $w \leq \frac{m(c(w))+1}{2m(c(w))+1} < w_0$, the risk $r(w) = 1 + \frac{1}{w} + \frac{1-w}{w}m(c(w))$ decreases in w.

Consequently, we obtain for the optimal weight w^* that

$$\frac{m(c(w))+1}{2m(c(w))+1} \leq w^* \leq w_0 = \frac{r_0 - 1}{2r_0 - 3} \tag{18}$$

and

$$r(w^*) = 1 + \frac{w^*}{1-w^*}m(c(w^*)) \tag{19}$$

r_0	w^*	c^*	$r(w^*)$	α
2.0	1	∞	∞	0
2.1	0.92	6.25	73.61	0.01
2.2	0.86	4.64	30.79	0.03
2.3	0.81	3.66	18.35	0.06
2.4	0.78	2.95	12.60	0.09
2.5	0.75	2.37	9.32	0.12
2.6	0.73	1.87	7.19	0.17
2.7	0.71	1.42	5.69	0.23
2.8	0.69	1.01	4.57	0.32
2.9	0.68	0.58	3.70	0.44
3.0	0.67	0	3.00	1

TABLE 2. linear versus quadratic regression: efficiency constraints

for the corresponding risk. Now, from (17) we get the relation

$$w = \frac{2 - F_{\chi^2_{3,0}}(c(w))}{r_0 - F_{\chi^2_{3,0}}(c(w))} \tag{20}$$

for $w \leq w_0$ and, hence,

$$r(w) = 1 + \frac{1}{r_0 - 2}(2 - F_{\chi^2_{3,0}}(c(w)))m(c(w)). \tag{21}$$

Numerical calculations indicate that $(2 - F_{\chi^2_{3,0}}(c))m(c)$ is increasing in c. thus, $r(w)$ is decreasing in w which yields an optimal weight $w^* = w_0 = \frac{r_0-1}{2r_0-3}$ with corresponding standardized maximal mean squared error $r(w^*) = 1 + \frac{r_0-1}{r_0-2}m(c(w^*))$.
For various values of the efficiency constraint $r_0 = \sup_{x\in\mathcal{X}} d_0(\xi_{w^*}, c^*; x)$ these weights w^* and their associated critical values $c^* = c(w^*)$ are listed in Table 2 together with $r(w^*) = \sup_{x\in\mathcal{X}} d(\xi_{w^*}, c^*; x)$, the value of the global G-criterion (13), and the *significance levels* α corresponding to c^*. Note that w^* decreases in r_0 from the G–optimal weight 1 in the constrained model for $r_0 = 2$ to the G–optimal weight $\frac{2}{3}$ in the unconstrained model for $r_0 = 3$. Moreover, the entries in Table 2 give evidence that higher efficiencies in the constrained model, i. e. smaller values of r_0, result in larger global maximal mean squared errors in the unconstrained model. Note also that the constrained risk $d_0(\xi_{w^*}, c^*; x)$ attains its maximum at $x = 1$ while the unconstrained risk $d(\xi_{w^*}, c^*; x)$ attains its maximum at $x = 0$.

As both, the local and global G-criteria (10) and (13), are equivariant with respect to linear reparametrizations of the model the optimality results obtained in Examples 1 to 4 carry over to arbitray intervals $[a, b]$ for the design region $\mathcal{X}$, if the optimal designs ξ^* are appropriately transformed, e. g. ξ_w assigns equal weights $\frac{1}{2}w$ to both endpoints a and b and the remaining weight $1 - w$ to the midpoint $\frac{a+b}{2}$.

7 Additional Remarks

Remark 1. For a given design ξ the pre-test estimator $\widehat{\boldsymbol{\beta}}_c$ outperforms the unconstrained least squares estimator $\widehat{\boldsymbol{\beta}}^*$ if the parameter $\boldsymbol{\beta}$ lies in a vicinity of the constraint $\boldsymbol{l}^\top\boldsymbol{\beta} = 0$. However, it is doubtful whether this is still true if for fixed c the best design $\xi^*(c)$ is chosen (see Example 2). Hence, the admissibility of the estimator-design pair $(c, \xi^*(c))$has to be checked in each particular case.

Remark 2. If the design ξ^* is both G–optimal in the constrained model and optimal for testing the constraint, then ξ^* is G–optimal for the pre–test estimators uniformly in all choices of the critical value c (see Example 1).

Acknowledgments: The present paper was finished while the second author was visiting the University of Technology at Darmstadt. He wishes to express his sincere thanks for the hospitality he met during this visit. His work was partly supported by the research grant Ku719/2–2 of the *Deutsche Forschungsgemeinschaft* at the Free University, Berlin.

References

Atkinson, A.C. and Donev, A.N. (1992). *Optimum Experimental Designs.* Clarendon Press, Oxford

Benda, N. (1996). Pre–test estimation and design in the linear model. *Journal of Statistical Planning and Inference*, **52**, 225–240

Boscher, H. (1991). Contamination in linear regression models and its influence on estimators. *Statistica Neerlandica*, **45**, 9–19

Droge, B. (1993). On finite–sample properties of adaptive least squares regression estimates. *Statistics*, **24**, 181–203.

Giles, J.A. (1991). Pre–testing in a mis–specified regression model. *Communications in Statistics — Theory and Methods*, **A 20**, 3221–3238

Judge, G.G. and Bock, M.E. (1978). *The Statistical Implications of Pre–test and Stein–rule Estimators in Econometrics.* North–Holland, Amsterdam

Karlin, S. and Studden, W.J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics.* Wiley, New York.

Pukelsheim, F. (1993). *Optimal Design of Experiments.* Wiley, New York.

Schwabe, R. (1996). *Optimum Designs for Multi–Factor Models. Lecture Notes in Statistics*, **113**. Springer, New York.

Norbert Benda
Institut für Medizinische Biometrie
Universität Tübingen
Westbahnhofstrasse 55
D–72 070 Tübingen
Germany

Rainer Schwabe
Technische Hochschule Darmstadt
Fachbereich Mathematik
Schlossgartenstrasse 7
D–64 289 Darmstadt
Germany