Reprinted from

journal of statistical planning and inference

Journal of Statistical Planning and Inference 55 (1996) 77–93

A nonsequential design procedure for parameter estimation and model discrimination in nonlinear regression models

Timothy E. O'Brien[a,*], John O. Rawlings[b]

[a] *Program in Statistics, Washington State University, Pullman, WA 99164-3144, USA*
[b] *Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA*

Received 11 January 1994; Revised 20 August 1995

ELSEVIER

ELSEVIER

Journal of Statistical Planning and Inference 55 (1996) 77–93

journal of statistical planning and inference

A nonsequential design procedure for parameter estimation and model discrimination in nonlinear regression models

Timothy E. O'Brien[a,*], John O. Rawlings[b]

[a] *Program in Statistics, Washington State University, Pullman, WA 99164-3144, USA*
[b] *Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA*

Received 11 January 1994; Revised 20 August 1995

Abstract

This paper presents and illustrates a new nonsequential design procedure for simultaneous parameter estimation and model discrimination for a collection of nonlinear regression models. This design criterion is extended to make it robust to initial parameter choices by using a Bayesian design approach, and is also extended to yield efficient estimation–discrimination designs which take account of curvature.

AMS classification: Primary 62K05; Secondary 53–00

Keywords: Bayesian designs, Bias; Curvature; Model discrimination; Multipurpose experimental design; Optimal design; Quadratic design criteria

1. Introduction

For a given process, several regression model functions can often be given which do a reasonable job of modelling the underlying mechanism. For example, the Gompertz, Logistic, Richards, Morgan-Mercer-Flodin, and Weibull sigmoidal growth curves discussed in Ratkowsky (1983) and Seber and Wild (1989) all provide good fits to the agricultural data sets given in Ch. 4 of Ratkowsky (1983). Another example is discussed in Box and Hill (1967), in which at least ten rival model functions are used to adequately describe the water gas shift reaction:

$$CO + H_2O \rightarrow CO_2 + H_2.$$

*Correspondence address: Department of Statistics, University of Georgia, 204 Statistics Building, Athens, GA 30602-1952, USA. Research partially funded by The Dow Chemical Company (Midland, MI, USA) and the National Science Foundation's International Postdoctoral Fellows Program, and completed while the first author was at the Universität Augsburg (Germany).

SSDI 0378-3758(95)00182-4

Design methodology for a collection of nonlinear regression models has generally focused on either efficiently estimating all of the model parameters of the collection (i.e., parameter estimation) or highlighting which of the model functions is more appropriate for describing the underlying process (i.e., model discrimination). For example, several of the design procedures given in Läuter (1974a, 1976) and Steinberg and Hunter (1984) focus solely on efficient parameter estimation, whereas the procedures given in Box and Hill (1967), Hill (1978), Atkinson (1988) and Pukelsheim and Rosenberger (1993) focus solely on model discrimination.

As estimation-only design procedures tend to do a rather poor job of model discrimination, and vice versa, a combined estimation–discrimination strategy is clearly needed. Sequential design procedures for simultaneous parameter estimation and model discrimination, such as those developed in Hill et al. (1968) and Borth (1975), are often impractical. Läuter (1974a) gives a non-sequential design procedure for estimation and discrimination for two linear models. This paper provides a non-sequential design procedure for simultaneous estimation and discrimination for a collection of nonlinear regression models. This new criterion, which depends on an initial choice of the model parameter vectors, is extended to make it robust to this initial parameter choice using the Bayesian design approach of Läuter (1974b) and Chaloner and Larntz (1989). Extensions to take account of the curvature associated with the model functions are also given.

In Section 2, design theory and estimation-only and discrimination-only design procedures are reviewed. We introduce and illustrate our estimation–discrimination procedure in Section 3, and provide Bayesian and second-order extensions in Section 4.

2. Background

2.1. Design theory

The design problem for the single nonlinear model

$$y_i = \eta(x_i, \underset{\sim}{\theta}) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

typically involves choosing a design with n design points, ξ, to estimate some function of the p-dimensional parameter vector, $\underset{\sim}{\theta}$, with high efficiency. Here ξ can be written as

$$\xi = \begin{Bmatrix} x_1, x_2, \ldots, x_n \\ \omega_1, \omega_2, \ldots, \omega_n \end{Bmatrix}, \tag{2.2}$$

where the design points $x_1, x_2, \ldots, x_n$ are elements of the design space X and are not necessarily distinct, and associated weights $\omega_1, \omega_2, \ldots, \omega_n$ are nonnegative real numbers which sum to one.

In contrast to the expression given in (2.2), ξ can be thought of in terms of its $r(r \leqslant n)$ *distinct* design points, s_1, $s_2, \ldots, s_r$, called its support points, and their associated design weights $\lambda_1, \lambda_2, \ldots, \lambda_r$. Whenever $n\lambda_i$ is integral for each i, ξ is said to be an exact design; otherwise, it is said to be a discrete design (see Rasch, 1990). We use ξ here to denote a discrete design with n design points and ξ^e to denote an exact n-point design. Whether a design is discrete or exact, $n\lambda_i$ can be thought of as the 'number of observations' taken at the experimental level s_i. Throughout this paper optimal designs will be sought in the class of all discrete designs.

When the residuals in (2.1) are uncorrelated normal random variables with zero mean and constant variance σ^2, the (Fisher) information per observation is given by

$$M(\xi,\underset{\sim}{\theta}) = \sum_{i=1}^{n} \omega_i \left(\frac{\partial\eta(x_i)}{\partial\underset{\sim}{\theta}}\right)\left(\frac{\partial\eta(x_i)}{\partial\underset{\sim}{\theta}}\right)^{\mathrm{T}} = V^{\mathrm{T}}\Omega V. \tag{2.3}$$

Here V is the $n\times p$ Jacobian of η with the ith row equal to $(\partial\eta(x_i)/\partial\underset{\sim}{\theta})^{\mathrm{T}}$, and $\Omega = \mathrm{diag}\{\omega_1,\ldots,\omega_n\}$. Also, the variance function (White, 1973) of η is given by

$$d(x,\xi,\underset{\sim}{\theta}) = \left(\frac{\partial\eta(x)}{\partial\underset{\sim}{\theta}}\right)^{\mathrm{T}} M^{-1}(\xi,\underset{\sim}{\theta})\left(\frac{\partial\eta(x)}{\partial\underset{\sim}{\theta}}\right), \tag{2.4}$$

where $\partial\eta(x)/\partial\underset{\sim}{\theta}$ is of dimension $p\times 1$, and a generalized inverse is used whenever M is singular.

For the (single) nonlinear model given in (2.1), optimal designs typically minimize some convex function of M^{-1}. For example, designs which minimize the determinant $|M^{-1}(\xi,\underset{\sim}{\theta}^0)|$ are called the locally D-optimal, and those which minimize the maximum (over $x\in X$) of $d(x,\xi,\underset{\sim}{\theta}^0)$ are called locally G-optimal. The term 'locally' is used here to emphasize the fact that the design is based on an initial parameter choice, $\underset{\sim}{\theta}^0$; see Chernoff (1953) and Box and Lucas (1959) for a discussion of local optimality. Further, the General Equivalence Theorem (Kiefer and Wolfowitz, 1960; White, 1973) establishes that D-optimal designs are equivalently G-optimal over the class of discrete designs. One advantage to using D-optimal designs in regressions settings is that these designs are invariant to a (even nonlinear) reparameterization of the model function. Other optimality criteria are discussed in Pázman (1986), Atkinson and Donev (1992), and Pukelsheim (1993).

Optimal designs can be obtained with computer programs which minimize the particular objective function by using a numerical search procedure such as the OPTMUM algorithm provided in the GAUSS[1] programming language. The GAUSS computer programs used to obtain optimal designs given here (available from the first author) require starting values for the design points and weights, and use a numerical search procedure which alternates between four numerical algorithms to obtain the corresponding optimal design.

2.2. Parameter estimation for a collection of nonlinear models

Often several model functions exist to describe a given process. Suppose that the m model functions $\eta_1(x,\underset{\sim}{\theta}_1),\ldots,\eta_m(x,\underset{\sim}{\theta}_m)$ are available, where $\underset{\sim}{\theta}_i$, the parameter vector

[1]GAUSS is a trademark of Aptech Systems, Inc., Kent, WA.

of the ith model, is of dimension $p_i \times 1$. Our focus here is on efficient estimation of the $p = p_1 + \cdots + p_m$ parameters in the collection $\mathbb{C} = \{\eta_1, \ldots, \eta_m\}$.

Corresponding to the ith model function in $\mathbb{C}$ is the $p_i \times p_i$ information matrix, $M_i(\xi, \underset{\sim}{\theta}_i) = V_i^{\mathrm{T}} \Omega V_i$, where the Jacobian matrix V_i, is of dimension $n \times p_i$. Here the number of support points of ξ must be greater than or equal to the maximum of the p_i's so that each $\underset{\sim}{\theta}_i$ is estimable. Further, the estimation efficiency (Atkinson and Donev, 1992, p. 116) associated with the ith model function is given by

$$E_i(\xi, \underset{\sim}{\theta}_i^0) = \left\{ \frac{|M_i(\xi, \underset{\sim}{\theta}_i^0)|}{|M_i(\xi_i^*, \underset{\sim}{\theta}_i^0)|} \right\}^{1/p_i}, \tag{2.5}$$

where ξ_i^* is the locally D-optimal design for the ith model function. The estimation efficiency is a relative measure of the information contained in ξ regarding the model parameters $\underset{\sim}{\theta}_i$, it compares the estimation information of the arbitrary design ξ with the estimation information of the D-optimal design, and the exponent in (2.5) reflects the number of parameters being considered.

When the nonnegative model weights $\pi_1, \ldots, \pi_m$ (which sum to unity) are chosen to reflect prior probabilities associated with the model functions $\eta_1, \ldots, \eta_m$, designs may be obtained to maximize $\sum \pi_i \log E_i$. Designs which maximize this weighted sum of the log estimation efficiencies, called locally $\mathrm{D_E}$-optimal designs here, equivalent maximize the estimation measure

$$E(\xi, \underset{\sim}{\theta}^0) = \sum_{i=1}^{m} \frac{\pi_i}{p_i} \log |M_i(\xi, \underset{\sim}{\theta}_i^0)|, \tag{2.6}$$

where $\underset{\sim}{\theta}^0 = (\underset{\sim}{\theta}_1^{0\mathrm{T}}, \ldots, \underset{\sim}{\theta}_m^{0\mathrm{T}})^{\mathrm{T}}$ is an initial choice of $\underset{\sim}{\theta} = (\underset{\sim}{\theta}_1^{\mathrm{T}}, \ldots, \underset{\sim}{\theta}_m^{\mathrm{T}})^{\mathrm{T}}$. Eq. (2.6) is a slight generalization of the estimation measure used in Atkinson and Donev (1992, p. 253), where focus is on parameter estimation in polynomial models. In the following example, we obtain a locally $\mathrm{D_E}$-optimal design for two nonlinear models.

Example 1. Two rival growth model functions are the one-parameter simple exponential model (SE1) and the one-parameter inverse linear model (IL1) given respectively by

$$\eta_1(x, \theta_{11}) = 1 - \mathrm{e}^{-\theta_{11} x} \quad \text{and} \quad \eta_2(x, \theta_{21}) = 1 - \frac{1}{1 + \theta_{21} x}. \tag{2.7}$$

For the initial parameter choices $\theta_{11}^0 = 0.1$ and $\theta_{21}^0 = 0.2$, and design space $X = [0, 30]$, the equal-interest (i.e., $\pi_1 = \pi_2 = \frac{1}{2}$) locally $\mathrm{D_E}$-optimal design (over all discrete designs) takes all observations at $x = 7.81$. By way of comparison, the locally D-optimal design for the first model function puts all weight at $x = 10.0$ and the locally D-optimal design for the second model function puts all weight at $x = 5.0$; the estimation efficiencies (see Eq. (2.5) above) of the $\mathrm{D_E}$-optimal design relative to these designs are $E_1 = 94\%$ and $E_2 = 91\%$.

Incidentally, the initial parameter value $\theta_{21}^{0} = 0.2$ was chosen here since this is the least-squares estimate obtained for θ_{21} when we used selected data points from the SE1 model with $\theta_{11} = 0.1$.

We note that the estimation measure given in (2.6) is preferred to the S-optimality estimation measure given in Läuter (1974a, 1976) since the measure given in (2.6) is invariant to a (linear or nonlinear) reparameterization of the model function, whereas the measure of Läuter is not.

2.3. Model discrimination for a collection of nonlinear models

Suppose now that our goal is to obtain a design that highlights which of the model functions in the collection $\mathbb{C}$ is best in describing the given process under study. The 'Bayesian' augmented model function for the collection $\mathbb{C}$ is of the form

$$\eta(x,\underset{\sim}{\theta}) = \pi_1\eta_1(x,\underset{\sim}{\theta}_1) + \cdots + \pi_m\eta_m(x,\underset{\sim}{\theta}_m), \tag{2.8}$$

where the positive weights $\pi_1,\ldots,\pi_m$ (which sum to one) are again chosen to reflect prior probabilities associated with the model function of $\mathbb{C}$. The $n \times p$ Jacobian for this augmented model is

$$\boldsymbol{V} = \frac{\partial\eta}{\partial\underset{\sim}{\theta}} = [\pi_1\boldsymbol{V}_1 \vdots \pi_2\boldsymbol{V}_2 \vdots \cdots \vdots \pi_m\boldsymbol{V}_m], \tag{2.9}$$

which is of dimension $n \times p$ where $n \geqslant p$. Also, the information matrix associated with the augmented model is (2.8) is $\boldsymbol{M}(\xi,\underset{\sim}{\theta}) = \boldsymbol{V}^{\mathrm{T}}\Omega\boldsymbol{V}$. In what follows, we assume that the columns of $\boldsymbol{V}$ are linearly independent. Inherent in this assumption is the requirement that no more than one of the m models in $\mathbb{C}$ be a linear model or have a linear component; see Pukelsheim and Rosenberger (1993) for the case where the models of interest are each linear models and Atkinson and Donev (1992) for the case where at least two of the models have linear components.

The discrimination efficiency (Atkinson and Cox, 1974, p. 323) associated with the ith model function of $\mathbb{C}$ is given by

$$D_i(\xi,\underset{\sim}{\theta}^0) = \left\{\frac{|\boldsymbol{M}(\xi,\underset{\sim}{\theta}^0)|/|\boldsymbol{M}_i(\xi,\underset{\sim}{\theta}_i^0)|}{|\boldsymbol{M}(\xi_i^*,\underset{\sim}{\theta}^0)|/|\boldsymbol{M}_i(\xi_i^*,\underset{\sim}{\theta}_i^0)|}\right\}^{1/s_i}, \tag{2.10}$$

where $s_i = p - p_i$ and ξ_i^* maximizes $|\boldsymbol{M}(\xi,\underset{\sim}{\theta}^0)|/|\boldsymbol{M}_i(\xi,\underset{\sim}{\theta}_i^0)|$. The discrimination efficiency is a relative measure of the information in ξ for detecting departures from the ith model function in the direction of the other model functions; it compares the discrimination information of the arbitrary design ξ with the discrimination information of ξ_i^*, and the exponent reflects the number of parameters in the collection $\mathbb{C}$ which are not in the ith model function.

When the nonnegative model weights $\pi_1,\ldots,\pi_m$ (which sum to unity) are again chosen to reflect prior probabilities associated with the model functions $\eta_1,\ldots,\eta_m$, designs may be obtained to maximize $\sum \pi_i \log D_i$. Designs which maximize this weighted sum of the log discrimination efficiencies, called locally D_D-optimal designs here, equivalently maximize the discrimination measure

$$D(\xi, \underset{\sim}{\theta}^0) = \sum_{i=1}^{m} \frac{\pi_i}{s_i} \log \frac{|M(\xi, \underset{\sim}{\theta}^0)|}{|M_i(\xi, \underset{\sim}{\theta}_i^0)|}, \tag{2.11}$$

again where $\underset{\sim}{\theta}^0 = (\underset{\sim}{\theta}_i^{0\mathrm{T}}, \ldots, \underset{\sim}{\theta}_m^{0\mathrm{T}})^{\mathrm{T}}$ is an initial choice of $\underset{\sim}{\theta} = (\underset{\sim}{\theta}_1^{\mathrm{T}}, \ldots, \underset{\sim}{\theta}_m^{\mathrm{T}})^{\mathrm{T}}$. Locally D_D-optimal designs highlight the adequacy of the m model functions in the collection $\mathbb{C}$. Note that the discrimination measure given in (2.11) represents a slight generalization of the equal-interest discrimination measure presented in Atkinson and Cox (1974), where each of the model weights is chosen to equal $1/m$. The following example illustrates equal-interest model discrimination.

Example 1 (*continued*). For the SE1 and IL1 model functions and the initial parameter choices given in Section 2.2, the equal-interest locally D_D-optimal design (over all discrete designs) for $n = p = 2$ design points associates the weight $\omega_1 = 0.60$ with the point $x_1 = 1.73$, and the weight $\omega_2 = 0.40$ with the point $x_2 = 13.19$. Further, the discrimination efficiencies (Eq. (2.10)) associated with this design are $D_1 = D_2 = 96\%$.

Not surprisingly, the estimation efficiencies associated with the design given in the previous example are only $E_1 = 46\%$ and $E_2 = 60\%$, illustrating that designs which do a good job of discrimination can do a rather poor job of parameter estimation (and vice versa). What is therefore needed is a design criterion which combines the dual goals of efficient parameter estimation and model discrimination.

3. Simultaneous estimation and discrimination for a collection of nonlinear models

Instead of treating parameter estimation and model discrimination for the collection of model functions $\mathbb{C}$ as separate problems, we now present a design procedure which provides for some control of the efficiency of both parameter estimation and model discrimination. The estimation and discrimination measures given in (2.6) and (2.11) can be combined into the single measure

$$\begin{aligned} B(\xi, \underset{\sim}{\theta}) &= \alpha E(\xi, \underset{\sim}{\theta}) + (1-\alpha) D(\xi, \underset{\sim}{\theta}) \\ &= b \log |M(\xi, \underset{\sim}{\theta})| + \sum_{i=1}^{m} c_i \log |M_i(\xi, \underset{\sim}{\theta}_i)|, \end{aligned} \tag{3.1}$$

where $\alpha \in [0, 1]$, $b = [1-\alpha)\sum(\pi_i/s_i)$, and $c_i = \alpha\pi_i/p_i - (1-\alpha)\pi_i/s_i$. Here α controls the degree of emphasis placed on estimation relative to discrimination. Note that the estimation–discrimination measure in (3.1) includes the alternate measure

$$B'(\xi, \underset{\sim}{\theta}) = \frac{\lambda}{E^*} E(\xi, \underset{\sim}{\theta}) + \frac{1-\lambda}{D^*} D(\xi, \underset{\sim}{\theta}) \tag{3.2}$$

as a special case by choosing

$$\alpha = \frac{\lambda}{\lambda + (1-\lambda)R} \quad \text{for } R = \frac{E^*}{D^*} \text{ and } \lambda \in [0, 1]; \tag{3.3}$$

here E^* and D^* are the maximum values of $E(\xi,\underset{\sim}{\theta})$ and $D(\xi,\underset{\sim}{\theta})$, respectively. An obvious advantage of $B(\xi,\underset{\sim}{\theta})$ in (3.1) over $B'(\xi,\underset{\sim}{\theta})$ in (3.2) is that $B(\xi,\underset{\sim}{\theta})$ is easier to use since it does not require knowing E^* and D^*. Designs which maximize $B(\xi,\underset{\sim}{\theta}^0)$ in (3.1), called locally $\mathrm{D_B}$-optimal designs here, are equivalently B-optimal (see Läuter, 1974a), and provide for simultaneous efficient parameter estimation and model discrimination.

Example 1 (*continued*). For the SE1 and IL1 model functions and the initial parameter choices in Section 2.2, the locally $\mathrm{D_B}$-optimal design (over all discrete designs) with the weights $\alpha = \pi_1 = \pi_2 = \frac{1}{2}$ (hereafter called the 'equal-interest' $\mathrm{D_B}$-optimal design) for $n = 2$ design points places the weight $\omega = \frac{1}{2}$ at each of the points $x_1 = 2.28$ and $x_2 = 11.99$. The estimation and discrimination efficiencies associated with this design are $E_1 = 64\%$, $E_2 = 72\%$, $D_1 = 93\%$ and $D_2 = 83\%$. To the extent that these discrimination efficiencies exceed the estimation efficiencies, this design tends to favor discrimination over estimation.

An alternative strategy is to maximize $B(\xi,\underset{\sim}{\theta}^0)$ in (3.1) subject to the constraint that E_1 and E_2 exceed some constant, c. For example, for $c = 70\%$, the constrained $\mathrm{D_B}$-optimal design for $n = 2$ observations associates the weight $\omega_1 = 0.40$ with the point $x_1 = 2.51$ and the weight $\omega_2 = 0.60$ with the point $x_2 = 11.49$. For this latter design, the estimation and discrimination efficiencies are $E_1 = 70\%$, $E_2 = 75\%$, $D_1 = 85\%$ and $D_2 = 68\%$. Of course, another way to obtain a design with higher estimation efficiencies would be to choose α in (3.1) to exceed $\frac{1}{2}$, although the exact relationship between α and the estimation and discrimination efficiencies is not readily apparent.

In all examples studied, 'equal-interest' $\mathrm{D_B}$-optimal designs had higher discrimination efficiencies than estimation efficiencies. This phenomenon is analogous to the experience associated with the sequential estimation–discrimination design criterion presented in Hill et al. (1968), in which initial runs tend to focus on model discrimination and subsequent runs tend to focus on parameter estimation once the model has been identified.

The variance function (Läuter, 1974a) which corresponds to the weighted sum of information matrices in (3.1) is given by

$$d_B(x,\xi,\underset{\sim}{\theta}) = b\,d(x,\xi,\underset{\sim}{\theta}) + \sum_{i=1}^{m} c_i d_i(x,\xi,\underset{\sim}{\theta}_i), \tag{3.4}$$

where $d(x,\xi,\underset{\sim}{\theta})$, $d_1(x,\xi,\underset{\sim}{\theta}_1),\ldots,d_m(x,\xi,\underset{\sim}{\theta}_m)$ are the variance functions associated with η in (2.8), $\eta_1,\ldots,\eta_m$, respectively, and are of the form given in (2.4). Further, the following extension to the General Equivalence Theorem and subsequent corollary are useful in obtaining D_B-optimal designs.

Theorem 3.1. D_B-*optimal designs (over the class of all discrete designs) can be equivalently characterized by the following conditions*:

1. ξ^* *maximizes* $B(\xi, \underline{\theta})$;
2. ξ^* *minimizes the maximum (over* X*) of* $d_B(x, \xi, \underline{\theta})$;
3. *the maximum of* $d_B(x, \xi^*, \underline{\theta})$ *over* X *is equal to one.*

Corollary 3.1. *The variance function* $d_B(x, \xi^*, \underline{\theta})$ *attains its maximum value at the support points of* ξ^*.

The proof of this theorem is similar to those of other extensions to the General Equivalence Theorem given in Whittle (1973) and Läuter (1974a), and is omitted here. Further, Corollary 3.1 can be used to show that a given design is indeed D_B-optimal, as is illustrated in the following example.

Example 2. Suppose that two rival model functions for a given growth process are the SE1 model in (2.7) with initial choice $\theta_{11}^0 = 0.1$ and the quadratic model, $\eta_2 = \beta_0 + \beta_1 x + \beta_2 x^2$, over the range $X = [0, 30]$. The equal-interest D_B-optimal design (over all discrete designs) for $n = p = 4$ design points associates the weights $\omega = 0.20$, 0.34, 0.27, and 0.19 with the points $x = 0$, 6.50, 19.61, and 30, respectively. The corresponding variance function, graphed in Fig. 1, verifies that this design is indeed D_B-optimal since this function reaches its maximum value (of $y = 1$) at the support points of the design. Incidentally, designs of this form can be converted into practical designs by using the algorithm presented in O'Brien and Rawlings (1993).

4. Extensions

Two important criticisms of design procedures based on information matrices of the form given in (2.3) are that these procedures are often valid only when the true vector of parameters is in a neighborhood of the initial choice (cf., Pilz, 1991), and that these procedures take no account of the curvature of the corresponding expectation surface. In this section, we demonstrate how the estimation–discrimination design procedure presented in the previous section can be extended to temper the impact of these potential problems.

4.1. A Bayesian estimation–discrimination procedure

For a single nonlinear model function, Bayesian D-optimal designs have been introduced in Läuter (1974b) and Chaloner and Larntz (1989) so as to relax the sensitivity of locally D-optimal designs to the initial parameter choice. Bayesian

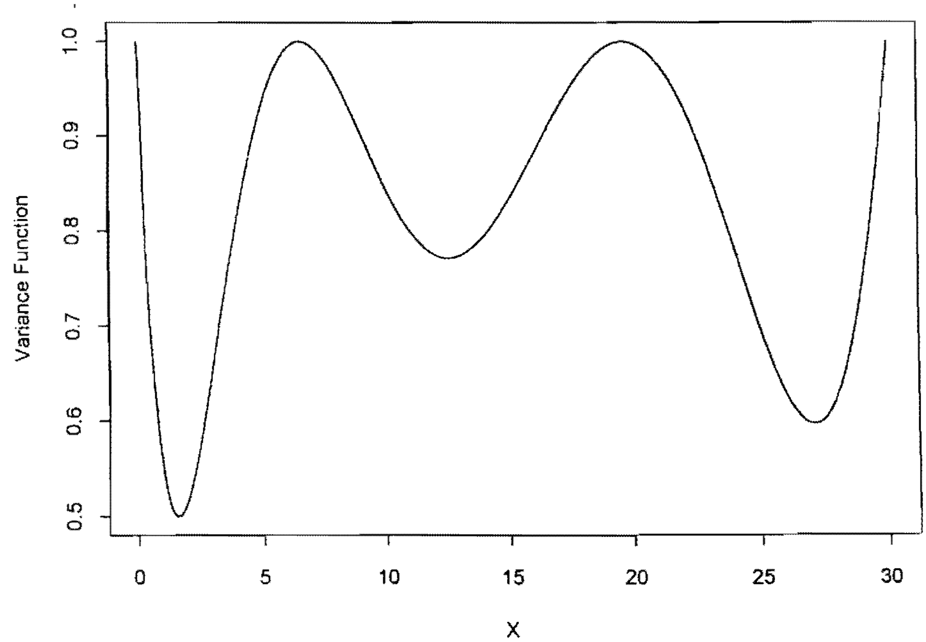

Fig. 1. Variance function for Example 2.

D-optimal designs maximize the expected log information, $\int \log |\boldsymbol{M}(\xi, \underset{\sim}{\theta})| p(\underset{\sim}{\theta}) \,\mathrm{d}\underset{\sim}{\theta}$ where $p(\underset{\sim}{\theta})$ is some prior distribution function hypothesized for $\underset{\sim}{\theta}$. Arumugham (1992) shows that for the Weibull model function Bayesian D-optimal designs are preferred to locally D-optimal designs when some uncertainty exists regarding the true value of $\underset{\sim}{\theta}$. Further, Chaloner and Larntz (1989) report that as the dispersion of $p(\underset{\sim}{\theta})$ increases, so too does the number of support points of the corresponding Bayesian D-optimal design.

In a similar manner, we define Bayesian D_B-optimal designs as those which maximize $\int B(\xi, \underset{\sim}{\theta}) p(\underset{\sim}{\theta}) \,\mathrm{d}\underset{\sim}{\theta}$ for $B(\xi, \underset{\sim}{\theta})$ given in (3.1) and for a given prior distribution function. An extension of the General Equivalence Theorem to Bayesian D_B-optimal designs is straightforward, and is omitted here; and an analogue of Corollary 3.1 may be used to show that a given design is indeed Bayesian D_B-optimal. The following example highlights the difference between Bayesian and non-Bayesian D_B-optimality designs.

Example 1 (*continued*). For the SE1 and IL1 model functions given in (2.7), suppose that θ_{11} and θ_{21} are independently and uniformly distributed on the intervals [0.07, 0.13] and [0.15, 0.25], respectively. For this prior distribution, the equal-interest Bayesian D_B-optimal design (over all discrete designs) for at least 2 design points put the weight $\omega = \frac{1}{2}$ at each of the points $x_1 = 2.22$ and $x_2 = 11.65$. This design represents a slight shift from the (non-Bayesian) locally D_B-optimal design, denoted ξ_1, which associates the weight $\omega = \frac{1}{2}$ with each of the points $x_1 = 2.28$ and $x_2 = 11.99$.

In contrast, if θ_{11} and θ_{21} are independently and uniformly distributed on the intervals [0.05, 0.15] and [0.10, 0.30], the Bayesian D_B-optimal design for at least 3 design points, denoted ξ_2, associates the weights $\omega = 0.42, 0.45$, and 0.13 with the points $x = 2.31$, 11.15, and 30, respectively. This increase in the number of support points for Bayesian D_B-optimal designs with an increase in uncertainty about $\underset{\sim}{\theta}$ is analogous to the experience of Chaloner and Larntz (1989) where Bayesian D-optimal designs are obtained for a single nonlinear model.

To compare the robustness regarding $\underset{\sim}{\theta}$ of this Bayesian D_B-optimal design (ξ_2) with that of the locally D_B-optimal design (ξ_1), we may use the D_B-efficiency

$$\mathrm{EFF} = \left[\frac{|\boldsymbol{M}(\xi_1,\underset{\sim}{\theta})|}{|\boldsymbol{M}(\xi_2,\underset{\sim}{\theta})|}\right]^{1/2} \tag{4.1}$$

evaluated at the extremes of the region

$$\{(\theta_{11},\theta_{21}) : 0.05 \leqslant \theta_{11} \leqslant 0.15 \text{ and } 0.10 \leqslant \theta_{21} \leqslant 0.30\},$$

where $\boldsymbol{M}$ corresponds to the Jacobian given in (2.9). Since the D_B-efficiencies at the points $(\theta_{11},\theta_{21}) = (0.05, 0.10), (0.05, 0.30), (0.15, 0.10)$ and $(0.15, 0.30)$ are $\mathrm{EFF} = 0.90$, 0.99, 0.49 and 1.01, respectively, we conclude that the Bayesian D_B-optimal design may be preferred to the locally D_B-optimal design since the efficiency of the latter design can be very poor as $\underset{\sim}{\theta}$ moves away from $\underset{\sim}{\theta}^{0}$ in certain directions (e.g., towards $\theta_{11} = 0.15$, $\theta_{21} = 0.10$).

4.2. *Quadratic estimation–discrimination design criteria*

For a single nonlinear model function, quadratic design criteria take account of the curvature of the corresponding expectation surface; see Bates and Watts (1980, 1988) and Seber and Wild (1989) for a discussion of curvature and O'Brien (1993) for a discussion of quadratic design criteria. Two such criteria are the D_2-optimality design criterion presented in O'Brien and Rawlings (1993) and the Q-optimality design criterion given in Hamilton and Watts (1985). In this section, the estimation–discrimination criterion given in Section 3 is extended to yield designs which take account of curvature.

4.2.1. D_{2B}-*optimality for a collection of nonlinear models*

For a nonlinear model of the form (2.1), designs which minimize the second-order generalized mean squared error (GMSE) of the least-squares estimate of $\underset{\sim}{\theta}$, called D_2-optimal designs, are introduced and illustrated in O'Brien and Rawlings (1993) and highlighted in Clarke and Haines (1995). D_2-optimal designs minimize the determinant of the second-order MSE

$$\boldsymbol{G} = \boldsymbol{G}(\xi,\underset{\sim}{\theta}) = \boldsymbol{S} + \boldsymbol{b}\boldsymbol{b}', \tag{4.2}$$

where $\boldsymbol{S}$ is the corresponding second-order variance estimate of Clarke (1980) and $\boldsymbol{b}$ is the second-order bias estimate of Box (1971). Since $\boldsymbol{G}$ is a function of $\underset{\sim}{\theta}$, either locally D_2-optimal designs or Bayesian D_2-optimal designs are obtained.

One extension of the estimation–discrimination design criterion presented in Section 3 to take account of curvature is to use D_2-optimality in place of D-optimality in Eq. (3.1). Thus, instead of obtaining designs to maximize $B(\xi,\underset{\sim}{\theta})$ given in (3.1), designs can be chosen to *minimize* the second-order estimation–discrimination measure

$$B_2(\xi,\underset{\sim}{\theta}) = b\log|\boldsymbol{G}(\xi,\underset{\sim}{\theta})| + \sum_{i=1}^{m} c_i \log|\boldsymbol{G}_i(\xi,\underset{\sim}{\theta}_i)|. \tag{4.3}$$

Here $\boldsymbol{G}$ is given in (4.2), and $\boldsymbol{G}_i = \boldsymbol{S}_i + \boldsymbol{b}_i\boldsymbol{b}_i'$ where $\boldsymbol{S}_i$ and $\boldsymbol{b}_i$ are the second-order variance and bias estimates associated with the ith model function.

Whereas estimation–discrimination designs based on D-optimality are called D_B-optimal designs, those based on D_2-optimality are called D_{2B}-optimal here; and since D_{2B}-optimal designs are functions of $\underset{\sim}{\theta}$ and σ, either locally D_{2B}-optimal designs or Bayesian D_{2B}-optimal designs are obtained. The following example illustrates the difference between locally D_B-optimal and locally D_{2B}-optimal designs.

Example 1 (*continued*). For the SE1 and IL1 model functions and initial parameter choices given in Section 2.2, the equal-interest locally D_B-optimal design associates the weight $\omega = \frac{1}{2}$ with each of the points $x_1 = 2.28$ and $x_2 = 11.99$. In contrast, the equal-interest locally D_{2B}-optimal design using $\sigma^0 = 0.3$ places the weight $\omega_1 = 0.69$ at the point $x_1 = 1.72$, and the weight $\omega_2 = 0.31$ at the point $x_2 = 10.87$, representing a nontrivial shift from the locally D_B-optimal design.

Various curvature measures have been introduced to assess the degree of nonlinearity (or curvature) of the corresponding expectation surface. Two important measures of the curvature associated with a particular design are the marginal curvature measure of Clarke (1987) and the 'measure of the importance of the biases' of Box (1971, p. 179). Analogous to the bias measure of Box (1971) is our bias–variance ratio (BVR), given by

$$\mathrm{BVR} = \frac{|\boldsymbol{G}| - |\boldsymbol{S}|}{p|\boldsymbol{S}|} = \frac{1}{p}\boldsymbol{b}'\boldsymbol{S}^{-1}\boldsymbol{b}. \tag{4.4}$$

Our BVR measure is preferred to the bias measure of Box (1971), since, whereas Box's measure compares a second-order bias estimate with a first-order variance estimate, ours compares a second-order bias estimate with a second-order variance estimate.

For this example, the length of the $p \times 1$ marginal curvature vector associated with the D_{2B}-optimal design is 7% less than that of the D_B-optimal design, and the length of the $m \times 1$ BVR vector associated with the D_{2B}-optimal design is 2% less than that of the D_B-optimal design. It follows that the D_{2B}-optimal design results in a slight to moderate reduction in curvature when compared with the D_B-optimal design.

A second quadratic design procedure is presented in Hamilton and Watts (1985), and in the following section this procedure is adapted to achieve parameter estimation and model discrimination for the model functions in the collection $\mathbb{C}$.

4.2.2. Q_B-optimality for a collection of nonlinear models

The $n \times p$ Jacobian associated with the Bayesian augmented model function given in (2.8) is

$$V = \frac{\partial \eta}{\partial \underset{\sim}{\theta}} = [\pi_1 V_1 \vdots \pi_2 V_2 \vdots \cdots \vdots \pi_m V_m] \tag{4.5}$$

and the corresponding $n \times p \times p$ array of second derivatives is

$$W = \operatorname{diag}\{\pi_1 W_1, \pi_2 W_2, \ldots, \pi_m W_m\}, \tag{4.6}$$

where each $W_i = \partial^2 \eta_i / \partial \underset{\sim}{\theta}_i^2$ is of dimension $n \times p_i \times p_i$.

Further, for the single model (2.1), the second-order approximation to the volume of the $100(1-\alpha)\%$ confidence region for $\underset{\sim}{\theta}$ developed in Hamilton and Watts (1985) can be written as

$$v(\xi, \underset{\sim}{\theta}) = d|V'V|^{-1/2}|D|^{-1/2}(1 + k^2 \operatorname{tr}(D^{-1}N)), \tag{4.7}$$

where d and k are constants, and D and N are functions of W; detailed expressions for these terms are given in Hamilton and Watts (1985). Designs which minimize $v(\xi, \underset{\sim}{\theta})$ in (4.7), called Q-optimal designs, are discussed in O'Brien (1992). As Q-optimal designs are functions of $\underset{\sim}{\theta}$ and σ, either locally Q-optimal designs or Bayesian Q-optimal designs are obtained.

A second extension to the estimation–discrimination design criterion presented in Section 3 to take account of curvature is to use Q-optimality in place of D-optimality in Eq. (3.1). Thus, instead of obtaining designs to maximize $B(\xi, \underset{\sim}{\theta})$ given in (3.1), designs can be chosen to *minimize* the second-order estimation–discrimination measure

$$B_Q(\xi, \underset{\sim}{\theta}) = b \log v(\xi, \underset{\sim}{\theta}) + \sum_{i=1}^{m} c_i \log v_i(\xi, \underset{\sim}{\theta}_i), \tag{4.8}$$

where v is the volume given in (4.7) corresponding to the augmented model function (2.8) and v_i is the volume corresponding to the ith model function of the collection $\mathbb{C}$. Designs which minimize $B_Q(\xi, \underset{\sim}{\theta}^0)$ in (4.8), called locally Q_B-optimal here, use the initial parameter choices $\underset{\sim}{\theta}^0$ and σ^0. The following example highlights the difference between locally D_B-optimal and locally Q_B-optimal designs for model functions which possess a moderate amount of curvature.

Example 3. A rival to the two-parameter intermediate product (IP2) model function

$$\eta_1 = \frac{\theta_{11}}{\theta_{11} - \theta_{12}}(e^{-\theta_{12}x} - e^{-\theta_{11}x}) \quad (\text{for } x \geqslant 0;\ \theta_{11}, \theta_{12} > 0) \tag{4.9}$$

is the two-parameter inverse quadratic (IQ2) model function

$$\eta_2 = \frac{\theta_{21}x}{(1 + \theta_{21}x)(1 + \theta_{22}x)} \quad (\text{for } x \geqslant 0;\ \theta_{21}, \theta_{22} > 0). \tag{4.10}$$

Suppose that reasonable initial parameter choices are $(\theta_{11}^0, \theta_{12}^0, \theta_{21}^0, \theta_{22}^0) = (0.7, 0.2, 1.8, 0.2)$ and $\sigma^0 = 0.1$, and that the corresponding design space is the interval [0, 10]. The equal-interest locally D_B-optimal design for $n = p = 4$ design points places the weight $\omega = \frac{1}{4}$ at each of the points $x = 0.241$, 1.104, 3.036, and 7.087. In contrast, the locally Q_B-optimal design associates weight $\omega = \frac{1}{4}$ with each of the points $x = 0.169$, 0.874, 2.710, and 5.896, representing a substantial shift in the design points. Further, since the Q_B-optimal design results in a 4% reduction in the length of the marginal curvature vector and a 9% reduction in the length of the BVR vector when compared with the D_B-optimal design, the Q_B-optimal design represents a moderate curvature reduction.

Incidentally, the initial parameter choices $\theta_{21}^0 = 1.8$ and $\theta_{22}^0 = 0.2$ were chosen here since these were the least-squares estimates obtained for θ_{21} and θ_{22} when we used selected data points from the IP2 model with $\theta_{11} = 0.7$ and $\theta_{12} = 0.2$.

5. Remarks

Table 1 lists the various optimality criteria and objective functions for the design strategies discussed here. In situations where more than one regression function can be used to model a given process, our recommendation is to obtain D_B-optimal designs whenever the curvature associated with the model functions is small and D_{2B}- and Q_B-optimal designs whenever curvature cannot be ignored. Although a practical

Table 1
First- and second-order estimation and discrimination criteria

Optimality criteria	Objective function	Goal
Local D_E-optimality	$E(\xi, \theta^0)$ in (2.6)	1st-order estimation (E) only
Local D_D-optimality	$D(\xi, \theta^0)$ in (2.11)	1st-order discrimination (D) only
Local D_B-optimality	$B(\xi, \theta^0)$ in (3.1)	1st-order E and D
Bayesian D_B-optimality	$\int B(\xi, \theta) p(\theta) d\theta$	Robust 1st order E and D
Local D_{2B}-optimality	$B_2(\xi, \theta^0)$ in (4.2)	2nd-order E and D
Bayesian D_{2B}-optimality	$\int B_2(\xi, \theta) p(\theta) d\theta$	Robust 2nd order E and D
Local Q_B-optimality	$B_Q(\xi, \theta^0)$ in (4.7)	2nd-order E and D
Bayesian Q_B-optimality	$\int B_Q(\xi, \theta) p(\theta) d\theta$	Robust 2nd order E and D

limitation of these second-order design procedures is that they can break down for moderate to large values of σ^* ($= \sigma/\sqrt{k}$ where k is the number of replicates of a given design used), large noise levels can be overcome by increased replication.

Many of the examples where designs for estimation and discrimination are needed involve only two (or possibly three) model functions, each typically having the same number of parameters; see, for example, Atkinson and Cox (1974), Sparrow (1979), and Atkinson and Donev (1992). In these situations, each of the 'c_i/b' terms in (3.1) is zero (or nearly so), and so D_B-optimal designs essentially maximize $|M(\xi,\underset{\sim}{\theta})|$, D_{2B}-optimal designs essentially minimize $|G(\xi,\underset{\sim}{\theta})|$, and Q_B-optimal designs essentially minimize $v(\xi,\underset{\sim}{\theta})$. In all the examples studied, whenever $|c_i/b|$ was less than or equal to $\frac{1}{3}$, locally D_B-optimal designs were practically indistinguishable from those which maximize $|M(\xi,\underset{\sim}{\theta}^0)|$. Interestingly, the converse also seems to hold: for Example 2 given in Section 3, $|c_1/b| = |c_2/b| = \frac{1}{2} > \frac{1}{3}$, and the difference between the locally D_B-optimal design and the design which maximizes $|M(\xi,\underset{\sim}{\theta}^0)|$ is nontrivial. The point here is that in many practical situations, the estimation–discrimination design procedures presented here are very simple to use: simply find the design which maximizes $|M|$ (equivalently $|V|$ in many instances, where V is the augmented Jacobian given in (2.9) ignoring the π's). By way of comparison, the (sequential) estimation–discrimination proposed in Hill et al. (1968) is quite involved and time-consuming.

Another application of the estimation–discrimination procedures given here is to situations where a researcher requires a design to both efficiently estimate the p parameters of a given (single) model function and to provide for a check of lack of fit of the model function (see O'Brien, 1994, 1995). One shortcoming of currently used design procedures is that these procedures typically yield designs with only p design points, thereby providing no opportunity to test for the adequacy of the assumed model. A reasonable alternative strategy is to find a similar second model function, and use one of the estimation–discrimination procedures given above to obtain a design with 'extra' design points. For example, if the IP2 model function given in (4.9) adequately describes the given process, then a constrained Bayesian D_{2B}-optimal design (with four design points) could be obtained using the IQ2 model given in (4.10) as an alternative model and using the constraint that the estimation efficiency of the IP2 function parameters be at least, say 80%. This design would then provide the opportunity for at least a visual assessment of how well the IP2 model fits the actual data.

Subsequent to choosing the example given in Section 4.1 to illustrate our Bayesian D_B-optimality design procedure, we have gathered some relevant empirical evidence. Specifically, our concern was with the assumption of independence of the prior distribution used for θ_{11} (from the SE1 model function) and θ_{21} (from the IL1 model function) in the example. Our subsequent empirical experience was based on a study in which several values of θ_{11} were chosen from the interval [0.05, 0.15], and for each choice of θ_{11} least-squares estimates of θ_{21} were obtained for various designs and using points on the SE1 function as observations. Based on this study, we feel a more appropriate (empirically based) prior distribution for θ_{11} and θ_{21} is one where θ_{11} is assumed to be uniformly distributed on the interval [0.05, 0.15], and conditional on this value equalling θ_{11}^*, θ_{21} has a gamma distribution with parameters which depend on θ_{11}^*. Since in our study we never observed a least-squares estimate for θ_{21} less than the corresponding value of θ_{11}, our empirically based prior distribution function assigns zero probability to the region $\theta_{21} < \theta_{11}$. Also, note that for our empirically based prior distribution function, the skewness in the gamma distributions increases from 0.4 at $\theta_{11} = 0.05$ to 1.7 at $\theta_{11} = 0.15$.

Recall that the Bayesian D_B-optimal design for this example using the independent uniform prior distribution given in Section 4.1 associates the weights $\omega = 0.42, 0.45$ and 0.13 with the points $x = 2.31$, 11.15 and 30. Interestingly, the Bayesian D_B-optimal design using the empirically based prior distribution function described here is quite similar in that it associates the weights $\omega = 0.47, 0.48$ and 0.05 with the points $x = 2.10$, 11.33 and 30. Further, to compare the D_B-efficiency of the locally D_B-optimal design (which associates the weight $\omega = \frac{1}{2}$ with each of the points $x = 2.28$ and $x = 11.99$), denoted ξ_1, relative to the empirically based Bayesian D_B-optimal design, denoted ξ_2, we use the expected D_B-efficiency

$$\mathrm{E(EFF)} = \iint \mathrm{EFF}\, p(\theta_{11}, \theta_{21})\, \mathrm{d}\theta_{11}\, \mathrm{d}\theta_{21}, \tag{5.1}$$

where EFF is the D_B-efficiency measure given in (4.1) and $p(\theta_{11}, \theta_{21})$ is the empirically based prior distribution function. The empirically based Bayesian D_B-optimal design (ξ_2) is preferred to the locally D_B-optimal design (ξ_1) here since, first, the expected D_B-efficiency if 98.8% (so, on average, ξ_2 is more efficient than ξ_1), and, second, although EFF is as high as 110% for one choice of $(\theta_{11}, \theta_{21})$, it is as low as 5% for another choice. This again highlights the lack of robustness of this locally D_B-optimal design to departures from the hypothesized values of θ_{11} and θ_{21} in certain directions.

We conclude by noting the following connection between the model discrimination measure given in (2.11) for the collection $\mathbb{C} = \{\eta_1(x, \underset{\sim}{\theta}_1), \eta_2(x, \underset{\sim}{\theta}_2)\}$ and the divergence (Kullback, 1959) between the hypotheses

$$\begin{aligned} &\mathrm{H}_1\colon\ \ y \sim \mathrm{N}(\eta_1(x, \underset{\sim}{\theta}_1), \sigma_1^2 I_n), \\ &\mathrm{H}_2\colon\ \ y \sim \mathrm{N}(\eta_2(x, \underset{\sim}{\theta}_2), \sigma_2^2 I_n). \end{aligned} \tag{5.2}$$

When we choose $\pi_1 = p_1/p$ in (2.11), exact D_D-optimal designs maximize

$$|V_1' V_1|\,|V_2' V_2|\,\Delta^2 \tag{5.3}$$

where the vector alienation coefficient (Hotelling, 1936) is given by

$$\Delta = \frac{|V_1'(I_n - P(V_2))V_1|}{|V_1' V_1|} = \frac{|V_2'(I_n - P(V_1))V_2|}{|V_2' V_2|}, \tag{5.4}$$

for $\boldsymbol{P}(\boldsymbol{A}) = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$. The vector alienation coefficient, which lies between zero and one, is a measure of the collinearity between $C(V_1)$, the column space of V_1, and $C(V_2)$, the column space of V_2. When $\Delta = 0$, $C(V_1)$ and $C(V_2)$ are perfectly collinear, and when

$\Delta = 1$, these column spaces are perfectly orthogonal; see Mardia et al. (1979) and Fox and Monette (1992). Further, since the 'distance' or divergence between the hypotheses H_1 and H_2 in (5.2) is proportional to

$$D(1,2) = \underset{\sim}{\theta}_1' V_1' V_1 \underset{\sim}{\theta}_1 + \underset{\sim}{\theta}_2' V_2' V_2 \underset{\sim}{\theta}_2 - 2\underset{\sim}{\theta}_1' V_1' V_2 \underset{\sim}{\theta}_2, \tag{5.5}$$

we note that designs which simultaneously make $V_1' V_1$ and $V_2' V_2$ 'large' and make $C(V_1)$ and $C(V_2)$ reasonably orthogonal result in larger values of both the divergence measure given in (5.5) and the discrimination measure given in (2.11).

Acknowledgements

The authors would like to express their appreciation to G. Peter Clarke, Linda M. Haines, Dieter Rasch and Suzanne M. Zweizig for helpful contributions.

References

Arumugham, T. (1992). Curvature and experimental design for the Weibull model. Ph.D. thesis, North Carolina State University, Raleigh, NC.

Atkinson, A.C. (1988). Recent developments in the methods of optimum and related experimental designs. *Internat, Statist. Rev.* **56**, 99–115.

Atkinson, A.C. and D.R. Cox (1974). Planning experiments for discriminating between models (with discussion). *J. Roy. Statist. Soc. Ser. B* **36**, 321–348.

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

Bates, D.M. and D.G. Watts (1980). Relative curvature measures of nonlinearity (with discussion). *J. Roy. Statist. Soc. Ser. B* **42**, 1–25.

Bates, D.M. and D.G. Watts (1988). *Nonlinear Regression Analysis and its Applications.* Wiley, New York.

Borth, D.M. (1975). A total entropy criterion for the dual problem of model discrimination and parameter estimation. *J. Roy. Statist. Soc. Ser. B* **37**, 77–87

Box, M.J. (1971). Bias in nonlinear estimation (with discussion). *J. Roy. Statist. Soc. Ser. B* **33**, 171–201.

Box, G.E.P. and W.J. Hill (1967). Discrimination among mechanistic models. *Technometrics* **9**, 57–71.

Box, G.E.P. and H.L. Lucas (1959). Design of experiments in non-linear situations. *Biometrika* **46**, 77–90.

Chaloner, K. and K. Larntz (1989). Optimal Bayesian design applied to logistic regression experiments. *J. Statist. Plann. Inference* **21**, 191–208.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* **24**, 586–602.

Clarke, G.P.Y. (1980) Moments of the least squares estimators in a non-linear regression model. *J. Roy. Statist. Soc. Ser B* **42**, 227–237.

Clarke, G.P.Y. (1987). Marginal curvatures and their usefulness in the analysis of nonlinear regression models. *J. Amer. Statist. Assoc.* **82**, 844–850.

Clarke, G.P.Y. and L.M. Haines (1995). Optimal design for models incorporating the Richards function. In: G.U.H. Seeber et al., Eds., *Proc. 10th Internat. Workshop on Statistical Modelling*. Springer, New York, 61–66.

Fox, J. and G. Monette (1992). Generalized collinearity diagnostics. *J. Amer Statist. Assoc.* **87**, 178–183.

Hamilton, D.C. and D.G. Watts (1985). A quadratic design criterion for precise estimation in nonlinear regression models *Technometrics* **27**, 241–250.

Hill, P.D.H. (1978). A review of experimental design procedures for regression model discrimination. *Technometrics* **20**, 15–21.

Hill, W.J., W.G. Hunter and D.W. Wichern (1968). A joint design criterion for the dual problem of model discrimination and parameter estimation. *Technometrics* **10**, 145–160.

Hotelling, H. (1936). Relations between two sets of variables. *Biometrika* **23**, 321–377.

Kiefer, J. and J. Wolfowitz (1960). The equivalence of two extremum problems *Canad. J. Math.* **12**, 363–366.

Kullback, S. (1959). *Information Theory and Statistics.* Wiley, New York.

Läuter, E. (1974a). Experimental design in a class of models. *Math. Oper. Statist.* **5**, 379–398.

Läuter, E. (1974b). A method of designing experiments for nonlinear models. *Math. Oper. Statist.* **5**, 625–636.

Läuter, E. (1976). Optimal multipurpose designs for regression models. *Math. Oper. Statist.* **7**, 51–68.

Mardia, K.V., J.T. Kent and J.M. Bibby (1979) *Multivariate Analysis.* Academic Press, London.

O'Brien, T.E. (1992). A note on quadratic designs for nonlinear regression models. *Biometrika* **79**, 847–849.

O'Brien, T.E. (1993). Design strategies for nonlinear regression models. Ph.D. thesis, North Carolina State University, Raleigh, NC.

O'Brien, T.E. (1994). A new robust design strategy for sigmoidal models based on model nesting. In: R. Dutter and W. Grossmann, Eds., *CompStat 1994.* Physica-Verlag, Heidelberg, 97–102.

O'Brien T.E. (1995). Optimal design and lack of fit in nonlinear regression models. In: G.U.H. Seeber et al., Eds., *Proc. 10th Internat. Workshop on Statistical Modelling.* Springer, New York, 201–206.

O'Brien, T.E. and J.O. Rawlings (1993). A practical design methodology for nonlinear regression models. Unpublished manuscript.

Pázman, A. (1986). *Foundations of Optimum Experimental Design.* Reidel, Dordrecht.

Pilz, J. (1991). *Bayesian Estimation and Experimental Design in Linear Regression Models.* Wiley, New York.

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

Pukelsheim, F. and J.L. Rosenberger (1993). Experimental designs for model discrimination. *J. Amer. Statist. Assoc.* **88**, 642–649.

Rasch, D. (1990). Optimum experimental design in nonlinear regression. *Comm. Statist. Theory Methods* **19**, 4789–4806.

Ratkowsky, D.A. (1983). *Nonlinear Regression Modeling: A Unified Practical Approach.* Marcel Dekker, New York.

Seber, G.A.F. and C.J. Wild (1989). *Nonlinear Regression.* Wiley, New York.

Sparrow P.E. (1979). Nitrogen response curves of spring barley. *J. Agric. Sci. Cambridge* **92**, 307–317.

Steinberg, D.M. and W.G. Hunter (1984). Experimental design: review and comment (with discussion). *Technometrics* **26**, 71–130.

White, L.V. (1973). An extension of the general equivalence theorem to nonlinear models. *Biometrika* **60**, 345–348.

Whittle, P. (1973). Some general points in the theory of optimal experimental design. *J. Roy. Statist. Soc. Ser. B* **35**, 123–130.