Experimental Designs for Model Discrimination

Author(s): Friedrich Pukelsheim and James L. Rosenberger

Source: *Journal of the American Statistical Association*, Vol. 88, No. 422 (Jun., 1993), pp. 642-649
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: http://www.jstor.org/stable/2290346
Accessed: 06-09-2017 07:12 UTC

Experimental Designs for Model Discrimination

FRIEDRICH PUKELSHEIM and JAMES L. ROSENBERGER*

We present designs that perform well for several objectives simultaneously. Three different approaches are discussed: to augment a given design in an optimal way, to evaluate a mixture of the various criteria, and to optimize one objective subject to achieving a prescribed efficiency level for the others. Our sample designs are for the situation of discriminating between a second- and third-degree polynomial fit, under the D-criterion and geometric mixtures of D-criteria.

KEY WORDS: Augmentation designs; D-optimality; Designs with guaranteed efficiencies; General equivalence theorem; Mixture designs.

1. INTRODUCTION

Many results on optimal experimental designs are derived under the assumption that the statistical model is known at the design stage. But more often than not, this is not the case. Indeed, the experimenter's goal often is to implement a design that is efficient for two or more models that might fit the experiment, to discriminate between them and then select the best one. For an early exposition of the issue, see the seminal paper by Stigler (1971) or the discussion paper by Atkinson and Cox (1974). Practical settings where the problem arises may be found, for example, in Hunter and Reiner (1965) and in Cook and Nachtsheim (1982). The task of weighing and averaging different criteria is also congenial with the Bayesian approach, as discussed by Läuter (1974, 1976).

In this article we review several solutions to the problem that have been proposed in the literature. The examples we list all may be derived from a single equivalence theorem in the spirit of Kiefer and Wolfowitz (1960), thus pulling together seemingly divergent approaches. All rely on maximizing some sort of information; that is, minimizing a function of the variance–covariance matrix of the least-squares estimator. Thus these solutions are complementary to the approach taken by Box and Draper (1959), whose designs optimized a mixture of variance and squared bias.

As a specific example with multiple objectives, we consider the discrimination between a second-degree and a third-degree polynomial model. Suppose that the experimental runs are determined by a single variable $x \in [-1, 1]$. The experimenter hopes that a second-degree polynomial model, designated (A), adequately describes the expected observations Y_x:

$$E(Y_x) = \theta_0 + \theta_1 x + \theta_2 x^2. \tag{A}$$

Yet it is desirable to guard against a third-degree model, designated (B):

$$E(Y_x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3. \tag{B}$$

This calls for a test, in a third-degree model, of whether or not the parameter θ_3 of the cubic term vanishes. If there is significant evidence that θ_3 is not 0, then the third-degree Model (B) is entertained, with parameter vector $\boldsymbol{\theta}_{(B)} = (\theta_0, \theta_1, \theta_2, \theta_3)'$. Otherwise, the second-degree Model (A) is used, with parameter vector $\boldsymbol{\theta}_{(A)} = (\theta_0, \theta_1, \theta_2)'$.

This setting gives rise to the following design problem. Find experimental designs that efficiently serve all three objectives simultaneously: (1) to discriminate between Models (A) and (B) and, depending on the decision, (2) to make inferences in Model (B) or (3) in Model (A).

We review three approaches to the problem that have been proposed in the literature. For the resulting designs we quote the moment matrix $\mathbf{M}$ as the basic quantity that enters into the objective criteria. We also quote $\mathbf{M}^{-1}$, the standardized dispersion matrix of the least squares estimate, because it more easily permits one to study the practical implications of exactly what is given up, in terms of variance, by preferring one design over another. As a numerical measure of goodness, we compare the designs through their efficiencies for each of the objective 1–3, as listed in Table 1.

The efficiencies are defined as follows. Denoting by ξ^* the D-optimal design for the k-dimensional parameter vector $\boldsymbol{\theta}$, any other design ξ has D-efficiency

$$D\text{-eff}(\xi) = \frac{[\det \mathbf{M}(\xi)]^{1/k}}{[\det \mathbf{M}(\xi^*)]^{1/k}}.$$

Here $\mathbf{M}(\xi)$ denotes the $k \times k$ moment matrix of the design ξ. The scaling with the kth root makes the criterion homogeneous of degree 1. We assume throughout that the usual determinant criterion is appropriate to evaluate the individual objectives. Although this notion of efficiency is common to the design literature, a simple transition to what is given up in terms of sample size—similar to the interpretation of Pitman efficiency—is not available.

In Section 2 we first list the D-optimal designs 2.1, 2.2, and 2.3 separately for each of the goals 1, 2, and 3.

In Section 3 we discuss optimal augmentation of a given design (Chaloner 1984; Covey-Crump and Silvey 1970; Welch 1982; Wynn 1977, 1982). The uniform equispaced design 3.1, which assigns the same number of observations to the five equispaced points -1, $-1/2$, 0, $1/2$, 1, has some intuitive appeal, apart from any optimality criterion. The augmentation design 3.2 takes half of the observations from

* Friedrich Pukelsheim is Professor, Institut für Mathematik, Universität Augsburg, D-8900 Augsburg, Germany. James L. Rosenberger is Professor and Head, Department of Statistics, The Pennsylvania State University, University Park, PA 16802. This article was written while the first author was visiting PSU. The authors thank Paul D. Johnson for helping with the computations.

Journal of the American Statistical Association
June 1993, Vol. 88, No. 422, Theory and Methods

Table 1. Efficient Designs for Second- and Third-Degree Model Discrimination

Section	Design	Efficiencies		
		θ_3	$\theta_{(B)}$	$\theta_{(A)}$
2.1	optimal for θ_3, on $[-1, 1]$ (optimal value .06250)	1	.93	.75
2.2	D-optimal for $\theta_{(B)}$, on $[-1, 1]$ (D-optimal value .26750)	.85	1	.87
2.3	D-optimal for $\theta_{(A)}$, on $[-1, 1]$ (D-optimal value .52913)	0	0	1
3.1	uniform, on the five points ± 1, $\pm 1/2$, 0	.72	.94	.84
3.2	half D-optimally augmented for $\theta_{(A)}$, on $[-1, 1]$	.42	.89	.94
4.1	mixture D-optimal for $\theta_{(A)}$ and $\theta_{(B)}$, on $[-1, 1]$	.66	.98	.91
4.2	mixture D-optimal for $\theta_{(A)}$ and $\theta_{(B)}$, on ± 1, $\pm 1/2$, 0	.64	.96	.90
4.3	mixture D-optimal for $\theta_{(A)}$ and θ_3, on ± 1, $\pm 1/2$, 0	1.00	.94	.75
5.1	D-optimal for $\theta_{(A)}$, 50% efficient for θ_3, on $[-1, 1]$	.5	.93	.94
5.2	D-optimal for $\theta_{(A)}$, 50% efficient for θ_3, on ± 1, $\pm 1/2$, 0	.5	.92	.93

NOTE: Efficiencies for the individual component θ_3 are evaluated in the third-degree model, as are the D-efficiencies for the full parameter vector $\theta_{(B)} = (\theta_0, \theta_1, \theta_2, \theta_3)'$. The D-efficiencies for $\theta_{(A)} = (\theta_0, \theta_1, \theta_2)'$ are calculated in the second-degree model, with the exception of 4.3, where both efficiencies are computed in a third-degree model.

the given design 3.1 and adjoins the other half in a way that is D-optimal for $\boldsymbol{\theta}_{(A)}$.

Section 4 presents designs that maximize the mixture of two criteria (Bunke and Bunke 1986; Lau and Studden 1985; Läuter 1974, 1976; Lim and Studden 1988). Design 4.1 mixes the D-criterion for $\boldsymbol{\theta}_{(A)}$ and the D-criterion for $\boldsymbol{\theta}_{(B)}$ on the experimental domain $[-1, 1]$ (Dette 1990). Design 4.2 evaluates the same objective function but is restricted to the five equispaced points ± 1, $\pm 1/2$, 0. Design 4.3 mixes the criterion for the individual parameter θ_3 and the D-criterion for $\boldsymbol{\theta}_{(A)}$, again restricted to the five points ± 1, $\pm 1/2$, 0.

In Section 5 we impose the constraint that our solution must be at least 50% efficient for the individual component θ_3 (Lau 1988; Lee 1987, 1988; Stigler 1971; Studden 1982). We present two designs, both of which are constrained D-optimal for $\boldsymbol{\theta}_{(A)}$. For design 5.1 (Studden 1982) the experimental domain is $[-1, 1]$. Design 5.2 is again restricted to the equispaced points ± 1, $\pm 1/2$, 0.

Of course, the designs presented here are by no means exhaustive. Many other designs satisfy the same purpose, and other approaches place more emphasis on nonlinear modeling and sequential designs (Atkinson and Fedorov 1975a, b). Section 6 provides some guidance for making a choice. Nevertheless, the final selection will reflect the peculiarities of the experimental situation under study, of the experimenter, or of the statistician.

The derivation of these designs, seemingly as diverse as the literature is scattered, may be unified using an appropriate generalization of the celebrated equivalence theorem of Kiefer and Wolfowitz (1960). The pertinent arguments are sketched in the Appendix, following Pukelsheim (1980, 1993).

2. INDIVIDUAL OPTIMALITY

Our designs ξ are given in the form $\xi(x_i) = w_i$, on no more than five support points $x_i \in [-1, 1]$ and with positive weights w_i summing to 1. Such a design ξ directs the experimenter to draw a fraction w_i of all observations under experimental conditions x_i.

Assuming that the observations from model (B) are uncorrelated and homoscedastic, the performance of a design ξ depends on its 4×4 third-degree moment matrix $\mathbf{M}_B(\xi)$, with entries $m_{p,q} = \mu_{p+q-2} = \sum_i w_i x_i^{p+q-2}$ for $p, q = 1, 2, 3, 4$. In model (A) we evaluate the 3×3 top left subblock of $\mathbf{M}_B(\xi)$, which is the second-degree moment matrix $\mathbf{M}_A(\xi)$. Each of our designs is symmetric around 0, and so all the odd moments vanish.

As is usual in the approximate design theory (Kiefer 1974), every probability distribution ξ on $[-1, 1]$ with a finite support is called a design and competes for optimality. A design is optimal for θ_3 when it minimizes the bottom right entry of $\mathbf{M}_B(\xi)^{-1}$ among all designs ξ. A design is D-optimal for $\theta_{(B)}$ when it maximizes the determinant of $\mathbf{M}_{(B)}(\xi)$ and is D-optimal for $\boldsymbol{\theta}_{(A)}$ when it maximizes the determinant of $\mathbf{M}_A(\xi)$. In addition to the design we display the moment matrix and its inverse, $\mathbf{M}_B(\xi)$ and $\mathbf{M}_B(\xi)^{-1}$, as a means to see how the lower order moments of ξ determine its performance.

Because more than one optimality criterion is of interest, we must standardize them so as to enable a meaningful comparison. This is achieved by turning to

(1) $[(\mathbf{M}_B(\xi)^{-1})_{44}]^{-1}$ for θ_3,
(2) $[\det \mathbf{M}_B(\xi)]^{1/4}$ for $\boldsymbol{\theta}_{(B)}$, and
(3) $[\det \mathbf{M}_A(\xi)]^{1/3}$ for $\boldsymbol{\theta}_{(A)}$.

As a function of the moment matrices, the criteria then are all positively homogeneous and concave and take value one for the identity matrix. Because they are information functions, as defined in Pukelsheim (1980), we call the associated optimal values the *optimal information.*

2.1 Optimal Design for θ_3

In a third-degree model, the optimal design ξ for the individual component θ_3 maximizes the criterion $[(\mathbf{M}_B(\xi)^{-1})_{44}]^{-1}$ and is supported by the Chebyshev points ± 1, $\pm 1/2$ (Kiefer and Wolfowitz 1959). The weights, the third-degree moment matrix, and its inverse are

$$\xi(\pm 1) = 1/6, \qquad \xi(\pm 1/2) = 1/3,$$

$$\mathbf{M}_B(\xi) = \begin{pmatrix} 1 & \cdot & .5 & \cdot \\ \cdot & .5 & \cdot & .38 \\ .5 & \cdot & .38 & \cdot \\ \cdot & .38 & \cdot & .34 \end{pmatrix},$$

$$\mathbf{M}_B(\xi)^{-1} = \begin{pmatrix} 3 & \cdot & -4 & \cdot \\ \cdot & 11 & \cdot & -12 \\ -4 & \cdot & 8 & \cdot \\ \cdot & -12 & \cdot & 16 \end{pmatrix}.$$

(Here and elsewhere, dots indicate 0s.) The optimal information for θ_3 is $1/16 = .0625$.

2.2 D-Optimal Design for $\boldsymbol{\theta}_{(\mathrm{B})}$

In a third-degree model, the D-optimal design for the full vector $\boldsymbol{\theta}_{(\mathrm{B})}$, maximizing $[\det \mathbf{M}_{\mathrm{B}}(\xi)]^{1/4}$, is (Kiefer 1959)

$$\xi(\pm 1) = 1/4, \qquad \xi(\pm 1/\sqrt{5}) = 1/4,$$

$$\mathbf{M}_{\mathrm{B}}(\xi) = \begin{pmatrix} 1 & \cdot & .6 & \cdot \\ \cdot & .6 & \cdot & .52 \\ .6 & \cdot & .52 & \cdot \\ \cdot & .52 & \cdot & .50 \end{pmatrix},$$

$$\mathbf{M}_{\mathrm{B}}(\xi)^{-1} = \begin{pmatrix} 3.25 & \cdot & -3.75 & \cdot \\ \cdot & 15.75 & \cdot & -16.25 \\ -3.75 & \cdot & 6.25 & \cdot \\ \cdot & -16.25 & \cdot & 18.75 \end{pmatrix}.$$

The D-optimal information for $\boldsymbol{\theta}_{(\mathrm{B})}$ is $2/(5^{5/4}) = .26750$.

2.3 D-Optimal Design for $\boldsymbol{\theta}_{(\mathrm{A})}$

In a second-degree model, the D-optimal design ξ for the vector $\boldsymbol{\theta}_{(\mathrm{A})}$, maximizing $[\det \mathbf{M}_{\mathrm{A}}(\xi)]^{1/3}$, is $\xi(\pm 1) = 1/3$, $\xi(0) = 1/3$. The D-optimal information for $\boldsymbol{\theta}_{(\mathrm{A})}$ is $4^{1/3}/3 = .52913$. Under this design, in a third-degree model neither the vector $\boldsymbol{\theta}_{(\mathrm{B})}$ nor the component θ_3 are estimable.

3. OPTIMAL AUGMENTATION DESIGNS

3.1 Uniform Equispaced Design

A design with some appeal of symmetry and balance is the uniform design ξ_0 on five equispaced points,

$$\xi_0(\pm 1) = \xi_0(\pm 1/2) = \xi_0(0) = 1/5,$$

$$\mathbf{M}_{\mathrm{B}}(\xi_0) = \begin{pmatrix} 1 & \cdot & .5 & \cdot \\ \cdot & .5 & \cdot & .43 \\ .5 & \cdot & .43 & \cdot \\ \cdot & .43 & \cdot & .41 \end{pmatrix},$$

$$\mathbf{M}_{\mathrm{B}}(\xi_0)^{-1} = \begin{pmatrix} 2.43 & \cdot & -2.86 & \cdot \\ \cdot & 18.06 & \cdot & -18.89 \\ -2.86 & \cdot & 5.71 & \cdot \\ \cdot & -18.89 & \cdot & 22.22 \end{pmatrix}.$$

It has efficiencies of 72%, 94%, and 84% for θ_3, $\boldsymbol{\theta}_{(\mathrm{B})}$, and $\boldsymbol{\theta}_{(\mathrm{A})}$.

Of course, there is no direct merit in the constant spacing. The high efficiencies are explained by the fact that the points ± 1, $\pm 1/2$, 0 are the second- and third-degree Chebyshev points (Kiefer and Wolfowitz 1959; Studden 1968). They already appear as support points for the optimal designs 2.1 and 2.3 and are close to the support points of design 2.2.

3.2 D-Optimal Augmentation for $\boldsymbol{\theta}_{(\mathrm{A})}$

For the second-degree model, the previous design ξ_0 has a D-efficiency of 84% for $\boldsymbol{\theta}_{(\mathrm{A})}$. As an alternative, only half of the observations are drawn according to the old design ξ_0. Subject to this "protected" design portion, the other half is filled in a D-optimal way for $\boldsymbol{\theta}_{(\mathrm{A})}$. That is, the criterion is to find a design ξ_1 that maximizes

$$\left\{\det\left[\frac{1}{2}\mathbf{M}_{\mathrm{A}}(\xi_0) + \frac{1}{2}\mathbf{M}_{\mathrm{A}}(\xi_1)\right]\right\}^{1/3}.$$

The resulting combined design, $\xi = (\xi_0 + \xi_1)/2$, is

$$\xi(\pm 1) = .2987, \qquad \xi(\pm 1/2) = .1, \qquad \xi(0) = .2026$$

$$\mathbf{M}_{\mathrm{B}}(\xi) = \begin{pmatrix} 1 & \cdot & .65 & \cdot \\ \cdot & .65 & \cdot & .61 \\ .65 & \cdot & .61 & \cdot \\ \cdot & .61 & \cdot & .60 \end{pmatrix},$$

$$\mathbf{M}_{\mathrm{B}}(\xi)^{-1} = \begin{pmatrix} 3.20 & \cdot & -3.39 & \cdot \\ \cdot & 35.75 & \cdot & -36.31 \\ -3.39 & \cdot & 5.24 & \cdot \\ \cdot & -36.31 & \cdot & 38.54 \end{pmatrix}.$$

The combined design ξ has efficiencies of 42%, 89%, and 94% for θ_3, $\boldsymbol{\theta}_{(\mathrm{B})}$, and $\boldsymbol{\theta}_{(\mathrm{A})}$. Thus the D-efficiency for $\boldsymbol{\theta}_{(\mathrm{A})}$ has increased by 10% at the cost of the other two efficiencies—in particular, that for θ_3.

We sketch a proof of optimality along the outline presented in the Appendix. To see that ξ is the D-optimal augmentation for $\boldsymbol{\theta}_{(\mathrm{A})}$ of the old design ξ_0, we represent it as $\xi = (\xi_0 + \xi_1)/2$, with $\xi_1(\pm 1) = w$, $\xi_1(0) = 1 - 2w$. That is, the new part ξ_1 is supported by the three second-degree Chebyshev points ± 1, 0. The equivalence theorem for this situation (Welch 1982; Wynn 1977) rests on the evaluation of the polynomial $P(x) = (1, x, x^2)(\mathbf{M}_{\mathrm{A}}(\xi_0/2 + \xi_1/2))^{-1}(1, x, x^2)'$, which gives the standardized variances of the estimated response surface. Optimality of ξ requires $P(\pm 1) = P(0)$, necessitating $w = (1 + \sqrt{71/5})/12 = .3974$. With this value, we find

$$P(x) = 3.20 - 5.24x^2 + 5.24x^4.$$

Now $P(x) \leq 3.20 = P(0) = P(\pm 1)$, for all $x \in [-1, 1]$, establishes the optimality of ξ.

4. OPTIMAL MIXTURE DESIGNS

The objective in this section is to optimize the geometric mean of the design criteria. Of course, the criteria also may be averaged by the arithmetic mean or the harmonic mean (Cook and Nachtsheim 1982).

4.1 D-Optimal Mixture Designs for $\boldsymbol{\theta}_{(\mathrm{A})}$ and $\boldsymbol{\theta}_{(\mathrm{B})}$, on $[-1, 1]$

The geometric mean of the D-optimality criteria for $\boldsymbol{\theta}_{(\mathrm{A})}$ and $\boldsymbol{\theta}_{(\mathrm{B})}$ is

$$\{[\det \mathbf{M}_{\mathrm{A}}(\xi)]^{1/3}[\det \mathbf{M}_{\mathrm{B}}(\xi)]^{1/4}\}^{1/2}.$$

The optimal design with respect to this objective function is (Dette 1990)

$$\xi(\pm 1) = 17/60, \qquad \xi(\pm\sqrt{17/117}) = 13/60,$$

$$\mathbf{M}_{\mathrm{B}}(\xi) = \begin{pmatrix} 1 & \cdot & .63 & \cdot \\ \cdot & .63 & \cdot & .58 \\ .63 & \cdot & .58 & \cdot \\ \cdot & .58 & \cdot & .57 \end{pmatrix},$$

$$\mathbf{M}_{\mathrm{B}}(\xi)^{-1} = \begin{pmatrix} 3.21 & \cdot & -3.51 & \cdot \\ \cdot & 21.79 & \cdot & -22.10 \\ -3.51 & \cdot & 5.57 & \cdot \\ \cdot & -22.10 & \cdot & 24.16 \end{pmatrix}.$$

The efficiencies for θ_3, $\boldsymbol{\theta}_{(B)}$, and $\boldsymbol{\theta}_{(A)}$ are 66%, 98%, and 91%. The value of the optimality criterion is .35553.

Following the general approach of the Appendix, we need to study the polynomial, which we designate as (C):

$$P(x) = \frac{1}{6}(1, x, x^2)\mathbf{M}_A(\xi)^{-1}\begin{pmatrix}1\\x\\x^2\end{pmatrix} + \frac{1}{8}(1, x, x^2, x^3)\mathbf{M}_B(\xi)^{-1}\begin{pmatrix}1\\x\\x^2\\x^3\end{pmatrix} = .94 + .94x^2 - 3.90x^4 + 3.02x^6. \qquad \text{(C)}$$

Because P attains the value 1 at ±1 and at $\pm\sqrt{17/117}$ and has local maxima at $\pm\sqrt{17/117}$, it is bounded by 1 on $[-1, 1]$. This proves that the design ξ maximizes the geometric mean on the experimental domain $[-1, 1]$.

4.2 D-Optimal Mixture Designs for $\theta_{(A)}$ and $\theta_{(B)}$, on ±1, $\pm1/2$, 0

As an alternative, we propose the design that maximizes the same criterion but restrict the support points of the design to the five Chebyshev points ±1, $\pm1/2$, 0. The resulting design is

$$\xi(\pm1) = .279, \qquad \xi(\pm1/2) = .164, \qquad \xi(0) = .114,$$

$$\mathbf{M}_B(\xi) = \begin{pmatrix} 1 & \cdot & .64 & \cdot \\ \cdot & .64 & \cdot & .58 \\ .64 & \cdot & .58 & \cdot \\ \cdot & .58 & \cdot & .56 \end{pmatrix},$$

$$\mathbf{M}_B(\xi)^{-1} = \begin{pmatrix} 3.43 & \cdot & -3.80 & \cdot \\ \cdot & 21.88 & \cdot & -22.48 \\ -3.80 & \cdot & 5.92 & \cdot \\ \cdot & -22.48 & \cdot & 24.87 \end{pmatrix}.$$

The efficiencies for θ_3, $\boldsymbol{\theta}_{(B)}$, and $\boldsymbol{\theta}_{(A)}$ are 64%, 96%, and 90%. The criterion takes the value .34974, which is 98% of the maximum value of the design in 4.1. The efficiencies are excellent, even though the design is inadmissible (Kiefer 1959).

To compute the design and verify its optimality, we proceed as indicated in the Appendix. We conjecture the optimal design to be symmetric, $\xi(\pm1) = w$, $\xi(\pm1/2) = u$, $\xi(0) = 1 - 2w - 2u$, and express the polynomial (C) in terms of u and w. If 0 belongs to the optimal support, then we must have $P(0) = 1$, yielding a relation for u in terms of w. Further, if the optimal support point contains ±1, then we get $P(\pm1) = 1$, leading to an equation that determines w. In summary, we obtain

$$u = \frac{17}{96} - 4w + \sqrt{\left(\frac{17}{96}\right)^2 + \frac{17}{4}w}, \qquad \frac{17}{3\mu_2^2} - \frac{10}{\mu_2} + \frac{1}{2w} = 0.$$

From this, w is computed numerically as .279. The polynomial becomes

$$P(x) = 1 + .78x^2 - 3.89x^4 + 3.11x^6.$$

Now $P(\pm1) = P(\pm1/2) = P(0) = 1$ proves optimality, on the Chebyshev support points ±1, $\pm1/2$, 0.

4.3 D-Optimal Mixture Designs for $\theta_{(A)}$ and θ_3, on ±1, $\pm1/2$, 0

The previous objective function is a mixture of two D-optimality criteria for two different models, the second- and third-degree models. As an alternative approach, one may embed the second-degree model in the third-degree model (Atkinson 1972). Hence $\boldsymbol{\theta}_{(A)}$ no longer is the full parameter vector in the model. Rather, it is considered a subvector of $\boldsymbol{\theta}_{(B)}$.

In the third-degree model, the information matrix for $\boldsymbol{\theta}_{(A)}$ is $\mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21}$, where matrix $\mathbf{M} = \mathbf{M}_B(\xi)$ is partitioned with a top left 3×3 subblock $\mathbf{M}_{11} = \mathbf{M}_A(\xi)$. That is, the information matrix for $\boldsymbol{\theta}_{(A)}$ now is a difference, the information matrix $\mathbf{M}_{11}$ of the second-degree model minus a 'penalty term' $\mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21}$ which reflects the loss of information due to fitting the additional parameter θ_3.

We now maximize the geometric mean of the D-information for $\boldsymbol{\theta}_{(A)}$ and the information for θ_3,

$$\{[\det(\mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21})]^{1/3}[\mathbf{M}_{22} - \mathbf{M}_{21}\mathbf{M}_{11}^{-1}\mathbf{M}_{12}]\}^{1/2},$$

on the Chebyshev points ±1, $\pm1/2$, 0. We obtain

$$\xi(\pm1) = .168, \qquad \xi(\pm1/2) = .332,$$

$$\mathbf{M}_B(\xi) = \begin{pmatrix} 1 & \cdot & .50 & \cdot \\ \cdot & .50 & \cdot & .38 \\ .50 & \cdot & .38 & \cdot \\ \cdot & .38 & \cdot & .35 \end{pmatrix},$$

$$\mathbf{M}_B(\xi)^{-1} = \begin{pmatrix} 3.01 & \cdot & -4.00 & \cdot \\ \cdot & 11.05 & \cdot & -12.04 \\ -4.00 & \cdot & 7.96 & \cdot \\ \cdot & -12.04 & \cdot & 16.00 \end{pmatrix}.$$

The efficiencies for θ_3, $\boldsymbol{\theta}_{(B)}$, and $\boldsymbol{\theta}_{(A)}$ are 100%, 94%, and 75%. As the efficiency for θ_3 indicates, the present design is practically the same as design 2.1, which is optimal for θ_3.

Derivation of the design ξ parallels the steps in 4.2. Again we conjecture the optimal design to be symmetric, $\xi(\pm1) = w$, $\xi(\pm1/2) = u$, $\xi(0) = 1 - 2w - 2u$. Now the equivalence theorem calls for the investigation of the polynomial

$$P(x) = \frac{1}{6}(1, x, x^2, x^3) \times \begin{pmatrix} \frac{\mu_4}{d} & \cdot & -\frac{\mu_2}{d} & \cdot \\ \cdot & \frac{4\mu_6}{D} - \frac{3}{\mu_2} & \cdot & -\frac{4\mu_4}{D} \\ -\frac{\mu_2}{d} & \cdot & \frac{1}{d} & \cdot \\ \cdot & -\frac{4\mu_4}{D} & \cdot & \frac{4\mu_2}{D} - \frac{1}{\mu_6} \end{pmatrix}\begin{pmatrix}1\\x\\x^2\\x^3\end{pmatrix},$$

where $d = \mu_4 - \mu_2^2$ and $D = \mu_2\mu_6 - \mu_4^2$.

Suppose that $x = 0$ is an optimal support point. Then we have $P(0) = 1$, entailing a relation to express u in terms of w. On the other hand, $P(1) = 1$ yields an equation that determines w. Thus we get

$$u = \frac{5}{24} - 4w + \sqrt{\left(\frac{5}{24}\right)^2 + 5w},$$

$$\frac{1}{5\mu_2^2} - \frac{17}{5\mu_2} - \frac{1}{\mu_6} + \frac{2}{w} = 0.$$

The resulting values $w = .19$ and $u = .44$ are not feasible, because the sum $2w + 2u$ exceeds 1. Hence $x = 0$ cannot be an optimal support point.

This leaves us with the relation $u = 1/2 - w$. From $P(1) = 1$ we calculate $w = .168$, and hence the design ξ. Its moments yield the polynomial

$$P(x) = .50 + 5.04x^2 - 14.73x^4 + 10.19x^6.$$

Now $P(0) = .50 < 1 = P(\pm 1/2) = P(\pm 1)$ establishes the optimality of ξ on the five points $\pm 1, \pm 1/2, 0$.

5. *D*-OPTIMAL CONSTRAINT DESIGNS

The original paper of Stigler (1971) proposed maximizing one criterion while securing some efficiency level for another criterion. Our last two designs implement this idea.

5.1 *D*-Optimal Design for $\theta_{(A)}$, Half Efficient for θ_3, on $[-1, 1]$

The *D*-information for $\boldsymbol{\theta}_{(A)}$ in the second-degree model is $[\det \mathbf{M}_A(\xi)]^{1/3}$. The design maximizing this criterion among those designs on the experimental domain $[-1, 1]$ that guarantee 50% efficiency for θ_3 in the third-degree model is (Studden 1982)

$$\xi(\pm 1) = .30095, \qquad \xi(\pm .3236) = .19905,$$

$$\mathbf{M}_B(\xi) = \begin{pmatrix} 1 & \cdot & .64 & \cdot \\ \cdot & .64 & \cdot & .61 \\ .64 & \cdot & .61 & \cdot \\ \cdot & .61 & \cdot & .60 \end{pmatrix},$$

$$\mathbf{M}_B(\xi)^{-1} = \begin{pmatrix} 3.16 & \cdot & -3.35 & \cdot \\ \cdot & 29.95 & \cdot & -30.14 \\ -3.35 & \cdot & 5.21 & \cdot \\ \cdot & -30.14 & \cdot & 32.00 \end{pmatrix}.$$

The efficiencies for θ_3, $\boldsymbol{\theta}_{(B)}$, and $\boldsymbol{\theta}_{(A)}$ are 50%, 93%, and 94%. For all practical purposes, the four support points would be considered equispaced in the interval $[-1, 1]$.

According to the Appendix, the necessary and sufficient condition of the equivalence theorem is in terms of the polynomial $P(x) = (1, x, x^2, x^3)N(1, x, x^2, x^3)'$, where N involves two matrices, one corresponding to the side conditions and the other one to the objective criterion, as well as a Lagrangian multiplier α:

$$N = \alpha \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & 28.40 & \cdot & -30.14 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & -30.14 & \cdot & 32.00 \end{pmatrix} + (1-\alpha) \begin{pmatrix} 1.05 & \cdot & -1.12 & \cdot \\ \cdot & .52 & \cdot & \cdot \\ -1.12 & \cdot & 1.74 & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}.$$

From $P(1) = 1$ we obtain $\alpha = .074$, giving

$$P(x) = .98 + .52x^2 - 2.88x^4 + 2.38x^6.$$

Now $P(1) = P(.3236) = 1$ and the vanishing of the derivative, $P'(.3236) = 0$, imply that on the interval $[-1, 1]$ the polynomial P is bounded by 1. This proves the desired optimality property of ξ.

5.2 *D*-Optimal Design for $\theta_{(A)}$, Half Efficient for θ_3, on ± 1, $\pm 1/2$, 0

As a final example, we take the same criterion as for design 5.1 but again restrict attention to the five Chebyshev points. As a result, we obtain the design

$$\xi(\pm 1) = .292, \qquad \xi(\pm 1/2) = .123, \qquad \xi(0) = .170,$$

$$\mathbf{M}_B(\xi) = \begin{pmatrix} 1 & \cdot & .65 & \cdot \\ \cdot & .65 & \cdot & .60 \\ .65 & \cdot & .60 & \cdot \\ \cdot & .60 & \cdot & .59 \end{pmatrix},$$

$$\mathbf{M}_B(\xi)^{-1} = \begin{pmatrix} 3.28 & \cdot & -3.53 & \cdot \\ \cdot & 29.15 & \cdot & -29.72 \\ -3.53 & \cdot & 5.47 & \cdot \\ \cdot & -29.72 & \cdot & 32.00 \end{pmatrix}.$$

The efficiencies for θ_3, $\boldsymbol{\theta}_{(B)}$, and $\boldsymbol{\theta}_{(A)}$ are 50%, 92%, and 93%. For all practical purposes, these efficiencies are just as good as those of design 5.1. The major difference is that design 5.2 has five support points rather than four.

For the optimality proof, we use the notation of Section 4.3. The efficiency constraint gives $\mu_2/D = 32$, where u is represented in terms of w. The polynomial to be studied is $P(x) = (1, x, x^2, x^3)\mathbf{N}(1, x, x^2, x^3)'$, with

$$\mathbf{N} = \alpha \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \frac{\mu_4^2}{32D^2} & \cdot & -\frac{\mu_4}{D} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & -\frac{\mu_4}{D} & \cdot & 32 \end{pmatrix} + (1-\alpha) \begin{pmatrix} \frac{\mu_4}{3d} & \cdot & -\frac{\mu_2}{3d} & \cdot \\ \cdot & \frac{1}{3\mu_2} & \cdot & \cdot \\ -\frac{\mu_2}{3d} & \cdot & \frac{1}{3d} & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}.$$

From $P(0) = 1$ we obtain a formula for α, whereas w is obtained from $P(1) = 1$. In summary, we get

$$u = \frac{4w}{36w - 1}, \qquad \alpha = \frac{6w^3}{(w - 1/48)(w - 1/36)} - 2.$$

With the resulting weights $w = .292$ and $u = .123$, we calculate $\alpha = .086$ and

$$P(x) = 1.00 + .69x^2 - 3.43x^4 + 2.75x^6.$$

Thus $P(\pm 1) = P(\pm 1/2) = P(0) = 1$ establishes optimality for the design on the Chebyshev points ± 1, $\pm 1/2$, 0.

6. DISCUSSION

In the settings discussed in this article, the design problem relates simultaneously to various models, various parameter systems, or various optimality criteria. There are many ways to combine information arising from these several sources into a single number. Consequently, there are many designs, each of which has good, convincing merits.

The multiplicities to be handled may arise at any one of the following four stages:

1. different models: Depending on the underlying regression functions, a design ξ may give rise to different moment matrices $M_i(\xi)$ of different orders $k_i \times k_i$, for a finite number of models $i = 1, \ldots, m$—Models (A) and (B) in this article.
2. different parameter systems: In model i the full k_i-dimensional parameter vector $\theta_{(i)}$ may be of interest, or only an s_i-dimensional subsystem $\mathbf{K}_i'\theta_{(i)}$, where the $k_i \times s_i$ coefficient matrix $\mathbf{K}_i$ is assumed known.
3. different optimality criteria: Given model i and the parameter system $\mathbf{K}_i'\theta_{(i)}$, different choices are possible for the optimality criterion ϕ_i to evaluate the information matrix for the parameter system of interest, $C_{\mathbf{K}_i}(\mathbf{M}_i)$. Thus as a function of the moment matrix $\mathbf{M}_i$, the objective criterion is a two-fold composition:

$$\psi_i(\mathbf{M}_i) = \phi_i(C_{\mathbf{K}_i}(\mathbf{M}_i)).$$

4. different averaging criteria: The final step is to average the information quantities $\psi_1, \ldots, \psi_m$ that originate with the m models and then merge them into a single number $\Phi(\psi_1, \ldots, \psi_m)$.

The optimality criteria ϕ of greatest interest are the classical D, A, and E criteria, which correspond to maximizing the geometric mean, the harmonic mean, and the minimum of the eigenvalues of the moment matrix. As long as one single model is being investigated, the classical means suffice for all practical purposes.

For the averaging criteria, we may similarly select the geometric mean, the harmonic mean, and the minimum of the information quantities ψ_i. But when the information from the m models is combined, we obtain a grand composition ϕ,

$$\phi(\mathbf{M}_1, \ldots, \mathbf{M}_m) = \Phi(\psi_1(\mathbf{M}_1), \ldots, \psi_m(\mathbf{M}_m)).$$

This terminal composition ϕ is *not* one of the classical means, but does belong to the class of information functions discussed next.

For a unified view of the problem, it thus is imperative to permit a wider class of criteria. The *information functions* of Pukelsheim (1980) serve this purpose well. By definition, they are required to be nonnegative, positively homogeneous, concave, nonconstant, and upper semicontinuous. For example, the vector means of order $p \in [-\infty, 1]$ are information functions on the vectors $\lambda = (\lambda_1, \ldots, \lambda_m)'$ in the nonnegative orthant $\mathbb{R}_+^m$,

$$\begin{aligned}\Phi_p(\lambda_1, \ldots, \lambda_m) &= \left(\frac{1}{m}\sum_{i\le m}\lambda_i^p\right)^{1/p} \quad \text{for} \quad -\infty < p \le 1, \quad p \ne 0\\ &= \left(\prod_{i\le m}\lambda_i\right)^{1/m} \quad \text{for} \quad p = 0\\ &= \min_{i\le m}\{\lambda_i\} \quad \text{for} \quad p = -\infty.\end{aligned}$$

In our examples we have used the geometric mean Φ_0 on the quadrant $\mathbb{R}_+^2$.

Similarly, the matrix means ϕ_p, defined for nonnegative definite $s \times s$ matrices $\mathbf{C}$ through

$$\begin{aligned}\phi_p(\mathbf{C}) &= \left(\frac{1}{s}\sum_{i\le s}\operatorname{trace}\mathbf{C}^p\right)^{1/p} \quad \text{for} \quad -\infty < p \le 1, \quad p \ne 0\\ &= (\det \mathbf{C})^{1/m} \quad \text{for} \quad p = 0\\ &= \text{smallest eigenvalue of } \mathbf{C} \quad \text{for} \quad p = -\infty\end{aligned}$$

are information functions on $\mathrm{NND}(s)$, the cone of nonnegative definite $s \times s$ matrices. Of course, matrix means and vector means are related, in that a matrix mean $\phi_p(\mathbf{C})$ on $\mathrm{NND}(s)$ may be reexpressed as a vector mean Φ_p on $\mathbb{R}_+^s$ applied to the eigenvalues $(\lambda_1(\mathbf{C}), \ldots, \lambda_s(\mathbf{C}))'$ of $\mathbf{C}$.

In general, a composition of such means $\Phi(\lambda_1, \ldots, \lambda_m)$ and $\phi_1(\mathbf{C}_1), \ldots, \phi_m(\mathbf{C}_m)$ fails to produce a classical mean. But any composition of the form

$$\phi = \Phi(\psi_1, \ldots, \psi_m)$$

does enjoy all the properties that constitute an information function, provided only that Φ is an information function on $\mathbb{R}_+^m$ and ψ_i is an information function on $\mathrm{NND}(k_i)$ for all models $i = 1, \ldots, m$. This shows that the concept of information functions is wide enough to embrace the classical criteria and also to permit functional operations, such as forming compositions of information functions.

Hence for the type of problems discussed in this article, a unified view emerges when we use an equivalence theorem of sufficient generality to apply to arbitrary information functions. An appropriate result has been given by Pukelsheim (1980) and is briefly reviewed in the Appendix. Just as in the original equivalence theorem of Kiefer and Wolfowitz (1960), this theorem typically leads to one set of equations that implicitly determine the optimal weights—as in 3.2, 4.2, 4.3, 5.2—while another set determines the optimal support points (see Dette 1990 or Studden 1982).

Four points deserve a final comment. The designs in 4.2, 4.3, and 5.2, on the five equispaced support points ± 1, $\pm 1/2$, 0, are inadmissible. For instance, the results of Kiefer (1959, p. 291), imply that for our Section 5.2 there is a four-point design η with a larger moment matrix, $\mathbf{M}_\mathrm{B}(\eta) \ge \mathbf{M}_\mathrm{B}(\xi)$ and $\mathbf{M}_\mathrm{B}(\eta) \ne \mathbf{M}_\mathrm{B}(\xi)$. Because the objective function and the efficiency side condition in Section 5.2 are given by information functions, and because any information function is monotone under the usual matrix ordering, the design η is at least as good as ξ. Nevertheless, ξ enjoys excellent efficiencies, as do the designs of Sections 4.2 and 4.3. This illustrates one virtue of a general approach: to calculate the efficiencies explicitly and thus provide numerical, indisputable evidence that the designs not only look good, but also do indeed perform well.

Second, all of our examples result in designs symmetric around 0. This demonstrates that symmetry considerations and a reduction by invariance applies to general information functions in the same powerful way that it helps with the classical criteria.

Third, all our results are in the approximate case. The transition to an exact design for sample size n can be carried out in the spirit of Fedorov (1972, chap. 3.1). Among the many available apportionment methods, the method of John Quincy Adams (Balinski and Young 1982, p. 28) is best for the design of experiments.

Fourth, when it comes to combining information (as was done in Section 4 for mixture designs), the most sensitive issue is that of scaling. The information quantities to be averaged must somehow be scaled to be represented in comparable units. The sole exception is the geometric mean of the determinant criterion, which is why we have chosen it for the examples in this article.

The design ordering that originates with the determinant criterion ϕ_0 is invariant under nonsingular affine transformations (Gaffke 1981). A similar order invariance pertains to the geometric mean Φ_0. The reason is that Φ_0 is homogeneous separately in each variable λ_i. This is not true for other information functions on $\mathbb{R}_+^m$. Thus the determinant criterion, and geometric means thereof, lead to the same optimal design irrespective of how the regression functions, and hence the moment matrices, are scaled. This provides a strong argument in their favor.

APPENDIX: EQUIVALENCE THEOREM

The equivalence theorem concentrates on moment matrices rather than on the set Ξ of all designs ξ. We have encountered the following sets $\mathcal{M}$ of moment matrices:

$$3.2 \quad \mathcal{M} = \left\{\frac{1}{2}\mathbf{M}_A(\xi_0) + \frac{1}{2}M_A(\xi_1): \xi \in \Xi\right\} \subseteq \text{NND}(3)$$

$$4.1, 4.2 \quad \mathcal{M} = \{(\mathbf{M}_A(\xi), \mathbf{M}_B(\xi)): \xi \in \Xi\} \subseteq \text{NND}(3) \times \text{NND}(4)$$

$$4.3 \quad \mathcal{M} = \{\mathbf{M}_B(\xi): \xi \in \Xi\} \subseteq \text{NND}(4)$$

$$5.1, 5.2 \quad \mathcal{M} = \{\mathbf{M}_A(\xi): \xi \in \Xi, (M_B(\xi)^{-1})_{44} \leq 32\} \subseteq \text{NND}(3).$$

Each of these sets is convex and compact, which are the only two properties called for by the equivalence theorem.

As pointed out in Section 6, all the optimality criteria considered are information functions ϕ. As a substitute for the notion of a gradient, or a subgradient, it proves advantageous to introduce the *polar information function* ϕ^∞ by defining, for any nonnegative matrix $\mathbf{D}$,

$$\phi^\infty(\mathbf{D}) = \inf_{C>0} \frac{\text{trace } \mathbf{CD}}{\phi(\mathbf{C})},$$

where the notation $\mathbf{C} > 0$ designates positive definiteness of $\mathbf{C}$. For example, the polar functions of the matrix means ϕ_p on NND(s) are known to be $\phi_p^\infty = s\phi_q$, where the numbers p and q are conjugate, $p + q = pq$.

For the sake of simplicity, we assume that the moment matrix $\mathbf{M} = \mathbf{M}(\xi)$ checked for optimality is positive definite.

Theorem. Let the set $\mathcal{M}$ of moment matrices be convex and compact and let the optimality criterion ϕ be an information function.

Then a design ξ with positive definite moment matrix $\mathbf{M} = \mathbf{M}(\xi)$ in $\mathcal{M}$ maximizes the criterion ϕ over $\mathcal{M}$ if and only if there exists a nonnegative solution $\mathbf{N}$ of the equation

$$\phi(\mathbf{M})\phi^\infty(\mathbf{N}) = \text{trace } \mathbf{MN} = 1 \tag{A.1}$$

that satisfies

$$\text{trace } \mathbf{AN} \leq 1 \quad \text{for all} \quad \mathbf{A} \in \mathcal{M}. \tag{A.2}$$

For a proof see Pukelsheim (1980). Equation (A.1) relates to the polar function ϕ^∞ and is called the *polarity equation.* Inequality (A.2) requires the matrix $\mathbf{N}$ to be normal to the set $\mathcal{M}$ at $\mathbf{M}$, and is called the *normality inequality.*

We have already mentioned that the matrix means ϕ_p have polars proportional to ϕ_q. This and related results make it usually easy to solve the polarity equation (A.1) and display the solution(s) $\mathbf{N}$ in terms of $\mathbf{M}$.

The point is to verify the normality inequality (A.2). For a third-degree model, the set $\mathcal{M}$ is generated by the rank one matrices $f(x)f(x)'$ with $x \in [-1, 1]$, where $f(x)$ is the power vector $(1, x, x^2, x^3)'$. Hence the left side of the normality inequality (A.2) turns into a polynomial $P(x)$,

$$\text{trace } f(x)f(x)'\mathbf{N} = f(x)'\mathbf{N}f(x) = (1, x, x^2, x^3)\mathbf{N}\begin{pmatrix}1\\x\\x^2\\x^3\end{pmatrix} = P(x).$$

Thus (A.2) boils down to calculating the polynomial $P(x)$ that comes with the optimality candidate $\mathbf{M}$ and checking whether on $[-1, 1]$ it is bounded by 1. In our exposition, from Section 3.2 on we have directly supplied the polynomial P that, because it is bounded on $[-1, 1]$ by 1, establishes optimality. A unified approach for obtaining the matrix $\mathbf{N}$ that determines P was given by Pukelsheim (1993). Individual approaches for each setting discussed here also exist and are found in the literature references in Section 1.

[*Received July 1991. Revised April 1992.*]

REFERENCES

Atkinson, A. C. (1972), "Planning Experiments to Detect Inadequate Regression Models," *Biometrika,* 59, 275–293.

Atkinson, A. C., and Cox, D. R. (1974), "Planning Experiments for Discriminating Between Models" (with discussion), *Journal of the Royal Statistical Society,* Ser. B, 36, 321–334.

Atkinson, A. C., and Fedorov, V. V. (1975a), "The Design of Experiments for Discriminating Between Two Rival Models," *Biometrika,* 62, 57–70.

——— (1975b), "Optimal Design: Experiments for Discriminating Between Several Models," *Biometrika,* 62, 289–303.

Balinski, M. L., and Young, H. P. (1982), *Fair Representation,* New Haven, CT: Yale University Press.

Box, G. E. P., and Draper, N. R. (1959), "A Basis for the Selection of a Response Surface Design," *Journal of the American Statistical Association,* 54, 622–654.

Bunke, H., and Bunke, O. (1986), *Statistical Inference in Linear Models. Statistical Methods of Model Building* (Volume I), Chichester, UK: John Wiley.

Chaloner, K. (1984), "Optimal Bayesian Experimental Design for Linear Models," *The Annals of Statistics,* 12, 283–300, Corrigendum, 13, 836.

Cook, R. D., and Nachtsheim, C. J. (1982), "Model Robust, Linear-Optimal Designs," *Technometrics,* 24, 49–54.

Covey-Crump, P. A. K., and Silvey, S. D. (1970), "Optimal Regression Designs with Previous Observations," *Biometrika,* 57, 551–566.

Dette, H. (1990), "A Generalization of D- and D_1-Optimal Designs in Polynomial Regression," *The Annals of Statistics,* 18, 1784–1804.

Fedorov, V. V. (1972), *Theory of Optimal Experiments.* New York: Academic Press.

Gaffke, N. (1981). "Some Classes of Optimality Criteria and Optimal Designs for Complete Two-Way Layouts," *The Annals of Statistics,* 9, 893–898.

Hunter, W. G., and Reiner, A. M. (1965), "Designs for Discriminating Between Two Rival Models," *Technometrics,* 7, 307–323.

Kiefer, J. C. (1959), "Optimum Experimental Designs" (with discussion), *Journal of the Royal Statistical Society,* Ser. B, 21, 272–304.

——— (1974), "General Equivalence Theory for Optimum Designs (Approximate Theory)," *The Annals of Statistics,* 2, 849–879.

Kiefer, J. C., and Wolfowitz, J. (1959), "Optimum Designs in Regression Problems," *The Annals of Mathematical Statistics,* 30, 271–294.

——— (1960), "The Equivalence of Two Extremum Problems," *Canadian Journal of Mathematics,* 12, 363–366.

Lau, T.-S. (1988), "D-Optimal Designs on the Unit Q-Ball," *Journal of Statistical Planning and Inference,* 19, 299–315.

Lau, T.-S., and Studden, W. J. (1985), "Optimal Designs for Trigonometric and Polynomial Regression Using Canonical Moments," *The Annals of Statistics,* 13, 383–394.

Läuter, E. (1974), "Experimental Design in a Class of Models," *Mathematische Operationsforschung und Statistik,* 5, 379–398.

——— (1976), "Optimal Multipurpose Designs for Regression Models," *Mathematische Operationsforschung und Statistik,* 7, 51–68.

Lee, C. M.-S. (1987), "Constrained Optimal Designs for Regression Models," *Communications in Statistics, Part A—Theory and Methods,* 16, 765–783.

——— (1988), "Constrained Optimal Designs," *Journal of Statistical Planning and Inference,* 18, 377–389.

Lim, Y. B., and Studden, W. J. (1988), "Efficient D_s-Optimal Design for Multivariate Polynomial Regression on the Q-Cube," *The Annals of Statistics,* 16, 1225–1240.

Pukelsheim, F. (1980), "On Linear Regression Designs Which Maximize Information," *Journal of Statistical Planning and Inference,* 4, 339–364.

——— (1993), *Optimal Design of Experiment,* New York: John Wiley.

Stigler, S. M. (1971), "Optimal Experimental Design for Polynomial Regression," *Journal of the American Statistical Association,* 66, 311–318.

Studden, W. J. (1968), "Optimal Designs on Tchebycheff Points," *The Annals of Mathematical Statistics,* 39, 1435–1447.

——— (1982), "Some Robust-Type D-Optimal Designs in Polynomial Regression," *Journal of the American Statistical Association,* 77, 916–921.

Welch, W. J. (1982), "Branch-and-Bound Search for Experimental Designs Based on D-Optimality and Other Criteria," *Technometrics,* 24, 41–48.

Wynn, H. P. (1977), "Optimum Designs for Finite Populations Sampling," in *Statistical Decision Theory and Related Topics II,* eds. S. S. Gupta and D. S. Moore, New York: Academic Press, pp. 471–478.

——— (1982), "Optimum Submeasures with Application to Finite Population Sampling," in *Statistical Decision Theory and Related Topics III* (Vol. 2), eds. S. S. Gupta and J. O. Berger, New York: Academic Press, pp. 485–495.