Integer-Valued, Minimax Robust Designs for Estimation and Extrapolation in Heteroscedastic, Approximately Linear Models
Author(s): Zhide Fang and Douglas P. Wiens
Source: *Journal of the American Statistical Association*, Vol. 95, No. 451 (Sep., 2000), pp. 807-818
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: https://www.jstor.org/stable/2669464
Accessed: 03-02-2019 01:51 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

American Statistical Association, Taylor & Francis, Ltd. are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association*

This content downloaded from 142.104.240.194 on Sun, 03 Feb 2019 01:51:42 UTC
All use subject to https://about.jstor.org/terms

Integer-Valued, Minimax Robust Designs for Estimation and Extrapolation in Heteroscedastic, Approximately Linear Models

Zhide FANG and Douglas P. WIENS

We present our findings on a new approach to robust regression design. This approach differs from previous investigations into this area in three respects: the use of a finite design space, the use of simulated annealing to carry out the numerical minimization problems, and in our search for integer-valued, rather than continuous, designs. We present designs for the situation in which the response is thought to be approximately polynomial. We also discuss the cases of approximate first- and second-order multiple regression. In each case we allow for possible heteroscedasticity and also obtain minimax regression weights. The results are extended to cover extrapolation of the regression response to regions outside of the design space. A case study involving dose–response experimentation is undertaken. The optimal robust designs, which protect against bias as well as variance, can be roughly described as being obtained from the classical variance-minimizing designs by replacing replicates with clusters of observations at nearby but distinct sites.

KEY WORDS: Bioassay; Carcinogen; Dose response; Efficient rounding; Finite design space; Fisher consistency; Logistic model; Polynomial regression; Probit model; Quota rounding; Second-order design; Simulated annealing; Weighted least squares.

1. INTRODUCTION

In this article we present our findings on a new approach to robust regression design. As was done by Box and Draper (1959), Li and Notz (1982), Marcus and Sacks (1976), Pesotchinsky (1982), and Wiens (1992, 1998), to name but a few contributors, we seek robustness against departures from the assumed linear response function and/or from the assumption of homoscedasticity. Our model allows for such departures of a very general form.

One way in which this work differs from that of the aforementioned authors is in our use of a finite design space. To our knowledge, robust regression designs under this restriction have previously been investigated only by Li (1984). The number of possible design points may be arbitrarily large, so that finiteness is not a practical restriction. The mathematical problem is, however, reduced to a finite-dimensional one, resulting in a considerable simplification both analytically and numerically.

The formulation of the problem of Li (1984) led to a linear programming problem sufficiently involved as to limit the investigation to the case of straight line regression over an interval. In contrast, we use an easily implemented simulated annealing algorithm that yields (near-) optimal solutions with no restrictions on the fitted model or on the structure of the design space. A general discussion of simulated annealing algorithms, with an application to a nonlinear design problem, has been given by Bohachevsky, Johnson, and Stein (1986). Haines (1987) and Meyer and Nachtsheim (1988) took this approach in constructing optimal designs assuming the fitted (linear, homoscedastic) model to be exactly correct.

The final departure from previous work lies in our search for integer-valued, rather than continuous, designs. Previous approaches have generally resulted in "designs" that are arbitrary and possibly continuous probability functions $\xi(\mathbf{x})$ on the design space. The number of observations allocated to a particular design point $\mathbf{x}_i$ is then $n\xi(\mathbf{x}_i)$, with the experimenter left to approximate this quantity when, as is typically the case, it is not an integer. In two of the three classes of problems considered here, the annealing algorithm gives exact integer-valued designs; that is, $n\xi(\mathbf{x}_i)$ is an integer. In the remaining case, although the designs need not be exact, we use and compare two methods of obtaining integer-valued approximations to $n\xi(\mathbf{x}_i)$.

We begin with a discussion of our model of approximate linearity and possible heteroscedasticity. We follow with a description of the design problems for which we seek solutions. In Section 2 we reduce these problems to numerical minimization problems. In Sections 3–5 we give the annealing algorithms, and examples of the minimum loss designs, for polynomial regression over an interval and for first- or second-order multiple regression over a multidimensional cube. In Section 6 we extend our results to the case of extrapolation, and apply them to a case study involving dose–response experimentation. In Section 7 we give conclusions and guidelines based on the observation that the optimal robust designs, which protect against bias and heteroscedasticity, can be roughly described as being obtained from the classical, (homogeneous) variance-minimizing designs by replacing replicates by clusters of observations at nearby but distinct sites.

Zhide Fang is Visiting Professor, Department of Mathematics, University of New Orleans, Lakefront, New Orleans, LA 70148 (E-mail: *zfang@math.uno.edu*). Douglas Wiens is Professor of Statistics, Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1 (E-mail: *doug.wiens@ualberta.ca*). This research is supported by the Natural Sciences and Engineering Research Council of Canada. The authors acknowledge the helpful comments of the editor, an associate editor, and a referee in the preparation of this article.

© 2000 American Statistical Association
Journal of the American Statistical Association
September 2000, Vol. 95, No. 451, Theory and Methods

This content downloaded from 142.104.240.194 on Sun, 03 Feb 2019 01:51:42 UTC
All use subject to https://about.jstor.org/terms

1.1 Statistical Model

The statistical model under consideration is initially formulated simply as a general regression model with additive errors, $Y = E[Y|\mathbf{x}] + \varepsilon$. Possible design points $\mathbf{x}_i$ $(i = 1, \ldots, N)$ are specified in advance and generally will be equally spaced over the region to be explored by the experimenter. At these points, uncorrelated observations Y_{ij} may be taken. The experimenter is to allocate $n_i \geq 0$ observations to $\mathbf{x}_i$, with $\sum_{i=1}^N n_i = n$ specified in advance. The design problem is to choose $n_1, \ldots, n_N$ in an optimal manner.

The mean response $E[Y|\mathbf{x}]$ is thought to be well approximated by a linear combination $\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}$ of p regressors $(z_1(\mathbf{x}), \ldots, z_p(\mathbf{x}))^T = \mathbf{z}(\mathbf{x})$. We assume throughout this article that the number of distinct locations at which $n_i > 0$ exceeds p. As is common in robustness problems in which the population being sampled may be incorrectly specified, one must decide beforehand exactly what is being estimated. Somewhat akin to the imposition of Fisher consistency in robust estimation problems, we define the parameter vector so as to minimize the average squared error of the approximation $E[Y|\mathbf{x}] \approx \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}$,

$$\boldsymbol{\theta} = \arg\min_{\mathbf{t}} \frac{1}{N} \sum_{i=1}^N \{E[Y|\mathbf{x}_i] - \mathbf{z}^T(\mathbf{x}_i)\mathbf{t}\}^2. \tag{1}$$

Write $f(\mathbf{x}) = E[Y|\mathbf{x}] - \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}$, so that the model becomes

$$Y_{ij} = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\theta} + f(\mathbf{x}_i) + \varepsilon_{ij}, \qquad i = 1, \ldots, N; \qquad j = 1, \ldots, n_i, \tag{2}$$

where, by virtue of (1),

$$\frac{1}{N} \sum_{i=1}^N \mathbf{z}(\mathbf{x}_i) f(\mathbf{x}_i) = \mathbf{0}. \tag{3}$$

To bound the bias of an estimate $\hat{\boldsymbol{\theta}}$, we also assume that

$$\frac{1}{N} \sum_{i=1}^N f^2(\mathbf{x}_i) \leq \eta^2 \tag{4}$$

for a given constant η^2.

We allow for possibly heteroscedastic errors, $\mathrm{var}[\varepsilon(\mathbf{x}_i)] = \sigma^2 g_i$, where $g_i = g(\mathbf{x}_i)$ for $g \in \mathcal{G}$ satisfying

$$\frac{1}{N} \sum_{i=1}^N g^2(\mathbf{x}_i) \leq 1. \tag{5}$$

The special case $\mathcal{G} = \{\mathbf{1}\}$ corresponds to homoscedastic errors.

The estimation method is least squares, possibly weighted with weights $w_i = w(\mathbf{x}_i)$ for $w \in \mathcal{W}$. The special case $\mathcal{W} = \{\mathbf{1}\}$ corresponds to ordinary least squares (OLS).

As a loss function, we take the average mean squared error (AMSE) I of $\hat{Y}(\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\hat{\boldsymbol{\theta}}$ as an estimate of $E[Y|\mathbf{x}]$,

$$I = \frac{1}{N} \sum_{i=1}^N E[(\hat{Y}(\mathbf{x}_i) - E[Y|\mathbf{x}_i])^2]. \tag{6}$$

This corresponds to the classical notion of I-optimality ("integrated variance"), termed Q-optimality by Fedorov (1972), with respect to which Studden (1977) characterized designs for exact polynomial regression. Huber (1975) used this loss function to obtain continuous robust designs for approximate straight line regression; Wiens (1990) extended Huber's theory to the case of multiple, approximately linear regression.

1.2 Design Problems

By a *design,* we mean a probability distribution $\{p_i\}_{i=1}^N$ on the design space $\mathcal{S} = \{\mathbf{x}_i\}_{i=1}^N$; if $p_i = n_i/n$ for *integers* $n_i \geq 0$, then we say that the design is *integer valued.* Such a design is implemented by assigning n_i observations to $\mathbf{x}_i$. We consider the following problems:

P1. Take $\mathcal{G} = \{\mathbf{1}\}, \mathcal{W} = \{\mathbf{1}\}$ and find an integer-valued design that minimizes the maximum, over f, value of the loss. Such a design is to be used with ordinary least squares estimation when the errors are assumed to be homoscedastic.

P2. Take $\mathcal{W} = \{\mathbf{1}\}$ and find an integer-valued design that minimizes the maximum over f and g, value of the loss. Such a design is to be used with OLS estimation when the errors may be heteroscedastic.

P3. Find an integer-valued design and weights that minimize the maximum, over f and g, value of the loss. Such a design is to be used with weighted least squares (WLS) estimation when the errors may be heteroscedastic. It turns out that we must be content with an approximate solution to this problem.

2. PRELIMINARIES

Let $\{p_i = n_i/n\}_{i=1}^N$ be an integer-valued design on $\mathcal{S}$. Without loss of generality, we assume the weights to be normed in such a way that $\sum_{i=1}^N p_i w_i = 1$ and define $m_i = p_i w_i$; then $\{m_i\}_{i=1}^N$ is a probability distribution on $\mathcal{S}$ satisfying

$$\sum_{i=1}^N \frac{m_i}{w_i} = 1. \tag{7}$$

For OLS, $m_i \equiv p_i$.

Let $\mathbf{Z}$ be the $N \times p$ matrix with rows $\mathbf{z}^T(\mathbf{x}_1), \ldots, \mathbf{z}^T(\mathbf{x}_N)$, and define $\mathbf{f} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_N))^T$. Let $\mathbf{G}, \mathbf{W}$, and $\mathbf{M}$ be $N \times N$ diagonal matrices with diagonal elements $\{g_i\}, \{w_i\}$, and $\{m_i\}$. Assume that $\mathbf{Z}$ is of full rank p and let $\mathbf{Z} = \mathbf{U}_{N\times p}\boldsymbol{\Lambda}_{p\times p}\mathbf{V}_{p\times p}^T$ be the singular value decomposition, with $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}_p$ and $\boldsymbol{\Lambda}$ diagonal and invertible. Augment $\mathbf{U}$ by $\tilde{\mathbf{U}}_{N\times N-p}$ in such a way that $[\mathbf{U} \vdots \tilde{\mathbf{U}}]_{N\times N}$ is orthogonal. Then by (3) and (4), we have that there is an $N - p \times 1$ vector $\mathbf{c}$, with $\|\mathbf{c}\| \leq 1$, satisfying

$$\mathbf{f} = \eta\sqrt{N}\tilde{\mathbf{U}}\mathbf{c}. \tag{8}$$

Finally, define $p \times p$ matrices $\mathbf{M}_j = \mathbf{U}^T\mathbf{M}^j\mathbf{U}$ for $j = 1, 2$ and

$$l_i = (\mathbf{U}\mathbf{M}_1^{-2}\mathbf{U}^T)_{ii}, \qquad i = 1, \ldots, N.$$

This content downloaded from 142.104.240.194 on Sun, 03 Feb 2019 01:51:42 UTC
All use subject to https://about.jstor.org/terms

In this notation, the WLS estimate

$$\hat{\boldsymbol{\theta}} = \left(\sum_{i=1}^{N} n_i w_i \mathbf{z}(\mathbf{x}_i)\mathbf{z}^T(\mathbf{x}_i)\right)^{-1} \sum_{i=1}^{N}\sum_{j=1}^{n_i} w_i \mathbf{z}(\mathbf{x}_i) Y_{ij},$$

has bias vector $\mathbf{d} := E[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}]$ and covariance matrix $\mathbf{C} := \mathrm{cov}[\hat{\boldsymbol{\theta}}]$ given by

$$\begin{aligned}\mathbf{d} &= \left(\sum_{i=1}^{N} n_i w_i \mathbf{z}(\mathbf{x}_i)\mathbf{z}^T(\mathbf{x}_i)\right)^{-1} \sum_{i=1}^{N} n_i w_i \mathbf{z}(\mathbf{x}_i) f(\mathbf{x}_i) \quad (9)\\ &= (\mathbf{Z}^T\mathbf{M}\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{M}\mathbf{f}\\ &= \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\mathbf{f} \quad (10)\end{aligned}$$

and

$$\begin{aligned}\mathbf{C} &= \sigma^2 \left(\sum_{i=1}^{N} n_i w_i \mathbf{z}(\mathbf{x}_i)\mathbf{z}^T(\mathbf{x}_i)\right)^{-1} \sum_{i=1}^{N} n_i w_i^2 g_i \mathbf{z}(\mathbf{x}_i)\mathbf{z}^T(\mathbf{x}_i)\\ &\quad \times \left(\sum_{i=1}^{N} n_i w_i \mathbf{z}(\mathbf{x}_i)\mathbf{z}^T(\mathbf{x}_i)\right)^{-1}\\ &= \frac{\sigma^2}{n}\,(\mathbf{Z}^T\mathbf{M}\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{M}\mathbf{W}\mathbf{G}\mathbf{Z}(\mathbf{Z}^T\mathbf{M}\mathbf{Z})^{-1}\\ &= \frac{\sigma^2}{n}\mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\mathbf{W}\mathbf{G}\mathbf{U}\mathbf{M}_1^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{V}^T. \quad (11)\end{aligned}$$

Our results depend on η^2 and σ^2 only through $\nu := \sigma^2/(n\eta^2)$, which can be chosen by the experimenter to reflect his view of the relative importance of variance versus bias. From (6), using (2), (10), (11), and finally (8), the AMSE $I = I(f, g, w, m)$ is

$$\begin{aligned}I(f,g,w,m) &= \frac{1}{N}\sum_{i=1}^{N}(E[\hat{Y}(\mathbf{x}_i)] - \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\theta})^2\\ &\quad + \frac{1}{N}\sum_{i=1}^{N}\mathrm{var}\,[\hat{Y}(\mathbf{x}_i)] + \frac{1}{N}\sum_{i=1}^{N} f^2(\mathbf{x}_i)\\ &= \frac{1}{N}\,\mathbf{d}^T\mathbf{Z}^T\mathbf{Z}\mathbf{d} + \frac{1}{N}\,\mathrm{tr}\,(\mathbf{Z}\mathbf{C}\mathbf{Z}^T) + \frac{1}{N}\,\mathbf{f}^T\mathbf{f}\\ &= \frac{1}{N}\,\mathbf{f}^T\mathbf{M}\mathbf{U}\mathbf{M}_1^{-2}\mathbf{U}^T\mathbf{M}\mathbf{f}\\ &\quad + \frac{\sigma^2}{nN}\sum_{i=1}^{N} m_i w_i g_i l_i + \frac{1}{N}\,\mathbf{f}^T\mathbf{f}\\ &= \eta^2\left\{\mathbf{c}^T(\mathbf{I} + \tilde{\mathbf{U}}^T\mathbf{M}\mathbf{U}\mathbf{M}_1^{-2}\mathbf{U}^T\mathbf{M}\tilde{\mathbf{U}})\mathbf{c}\right.\\ &\quad \left. + \frac{\nu}{N}\sum_{i=1}^{N} m_i w_i g_i l_i\right\}. \quad (12)\end{aligned}$$

Designs Solving Problems P1, P2, and P3

Problems P1–P3 require us to evaluate the maximum, over f and g, loss and to then minimize this maximum loss over $\{w_i\}_{i=1}^N$ and $\{p_i\}_{i=1}^N$. The minimizations can equivalently be done over $\{w_i\}_{i=1}^N$ and $\{m_i\}_{i=1}^N$ subject to (7). The maximizations over f and g, and the minimization over $\{w_i\}_{i=1}^N$, can then be evaluated algebraically. This leaves only one numerical minimization over $\{m_i\}_{i=1}^N$.

Fix a design $\{m_i\}_{i=1}^N$ on $\mathcal{S}$. Lemma 1, which is required for the maximization of (12) over f, follows from the observation that the nonzero eigenvalues of $\tilde{\mathbf{U}}^T\mathbf{M}\mathbf{U}\mathbf{M}_1^{-1} \cdot \mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\tilde{\mathbf{U}}$ are the same as those of

$$\begin{aligned}&\mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}\tilde{\mathbf{U}} \cdot \tilde{\mathbf{U}}^T\mathbf{M}\mathbf{U}\mathbf{M}_1^{-1}\\ &\quad = \mathbf{M}_1^{-1}\mathbf{U}^T\mathbf{M}(\mathbf{I} - \mathbf{U}\mathbf{U}^T)\mathbf{M}\mathbf{U}\mathbf{M}_1^{-1}\\ &\quad = \mathbf{M}_1^{-1}\mathbf{M}_2\mathbf{M}_1^{-1} - \mathbf{I}.\end{aligned}$$

Lemma 1. The maximum, over vectors $\mathbf{c}$ with $\|\mathbf{c}\| \le 1$, of $\mathbf{I} + \tilde{\mathbf{U}}^T\mathbf{M}\mathbf{U}\mathbf{M}_1^{-2}\mathbf{U}^T\mathbf{M}\tilde{\mathbf{U}}$ is the largest eigenvalue λ_m of $\mathbf{M}_1^{-1}\mathbf{M}_2\mathbf{M}_1^{-1}$. The maximizing $\mathbf{c}$ is the unit eigenvector of $\tilde{\mathbf{U}}^T\mathbf{M}\mathbf{U}\mathbf{M}_1^{-2}\mathbf{U}^T\mathbf{M}\tilde{\mathbf{U}}$ belonging to its maximum eigenvalue $\lambda_m - 1$.

From Lemma 1 and (12), we immediately obtain Theorem 1.

Theorem 1. For OLS estimation with homoscedastic errors, $(\mathcal{G} = \{\mathbf{1}\}, \mathcal{W} = \{\mathbf{1}\})$ we have

$$\max_f I(f, \mathbf{1}, \mathbf{1}, m) = \eta^2\left\{\lambda_m + \frac{\nu}{N}\sum_{i=1}^{N} m_i l_i\right\}, \quad (13)$$

and so the minimax design for P1 has $\{p_i\}_{i=1}^N = \{m_i\}_{i=1}^N$, where $\{m_i\}_{i=1}^N$ minimizes (13). For computational purposes, we note that $\sum_{i=1}^N m_i l_i = \mathrm{tr}\,\mathbf{M}_1^{-1}$. The least favorable response contaminant $\mathbf{f}$ is determined by (8), where $\mathbf{c}$ is as in Lemma 1.

To obtain the maximum loss under heteroscedasticity, we must maximize (12) over g; this is a straightforward application of the Cauchy–Schwarz inequality.

Lemma 2. The maximum value of $\sum_{i=1}^N m_i w_i g_i l_i$ subject to (5) is $\sqrt{N\sum_{i=1}^N m_i^2 w_i^2 l_i^2}$, attained with $g_i \propto m_i w_i l_i$.

Lemma 2, together with Lemma 1 and (12), give

$$\max_{f,g} I(f, g, w, m) = \eta^2\left\{\lambda_m + \frac{\nu}{\sqrt{N}}\left(\sum_{i=1}^{N} m_i^2 w_i^2 l_i^2\right)^{1/2}\right\}, \quad (14)$$

from which Theorem 2 is immediate.

Theorem 2. For OLS estimation $(\mathcal{W} = \{\mathbf{1}\})$ with heteroscedastic errors, we have

$$\max_{f,g} I(f, g, \mathbf{1}, m) = \eta^2\left\{\lambda_m + \frac{\nu}{\sqrt{N}}\left(\sum_{i=1}^{N} m_i^2 l_i^2\right)^{1/2}\right\}, \quad (15)$$

This content downloaded from 142.104.240.194 on Sun, 03 Feb 2019 01:51:42 UTC
All use subject to https://about.jstor.org/terms

and so the minimax design for P2 has $\{p_i\}_{i=1}^N = \{m_i\}_{i=1}^N$, where $\{m_i\}_{i=1}^N$ minimizes (15).

The maximum loss (14) for WLS can be minimized over the weights with the aid of a Lagrange multiplier. The result of this minimization is given by Lemma 3.

Lemma 3. The minimum value of $\sum_{i=1}^N (m_i w_i l_i)^2$ subject to (7) is $(\sum_{i=1}^N m_i^{4/3} l_i^{2/3})^3$, attained with $w_i \propto m_i^{-1/3} l_i^{-2/3}$ whenever $m_i > 0$.

Lemmas 1, 2, and 3, together with (12), now give Theorem 3.

Theorem 3. For WLS estimation with heteroscedastic errors, we have

$$\min_w \max_{f,g} I(f,g,w,m) = \eta^2 \left\{ \lambda_m + \frac{\nu}{\sqrt{N}} \left(\sum_{i=1}^N m_i^{4/3} l_i^{2/3} \right)^{3/2} \right\}. \quad (16)$$

The minimax design $\{p_i\}_{i=1}^N$ for P3 has $p_i \propto m_i^{4/3} l_i^{2/3}$, where $\{m_i\}_{i=1}^N$ minimizes (16). The least favorable variances satisfy $g_i \propto \sqrt{p_i}$ and the optimum weights satisfy $w_i \propto m_i/p_i$ whenever $m_i > 0$.

The motivation for considering P3 is that, as seen in (12), the loss has two distinct components. One—the bias component—arises purely from the effect of f. The other—the variance component—arises purely from the effect of g. It thus seems plausible that one could reduce one component through an appropriate choice of design points, and the other through a judicious choice of weights. This is what we attempt to do through the minimization of (16); see Section 4.3 for a numerical indication that the venture can be quite successful.

An interesting special case of Theorem 3 is $\nu = 0$, in which case the design and weights minimize the maximum bias measure $\lambda_m - 1$. We can in any event impose unbiasedness by requiring that $m_i \equiv N^{-1}$, resulting in the absolute minimum $\lambda_m = 1$. In this case the l_i are proportional to the diagonal elements h_{ii} of the hat matrix $\mathbf{H} = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T$, and we have Corollary 1.

Corollary 1. The design with $p_i \propto h_{ii}^{2/3}$ and $w_i \propto p_i^{-1}$ minimizes the maximum AMSE, subject to the side condition that $E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$ for all f.

3. EXACT DESIGNS FOR P1 AND P2; POLYNOMIAL RESPONSE

3.1 Description of the Simulated Annealing Algorithm

We consider approximate polynomial regression; that is, $\mathbf{z}(x) = (1, x, x^2, \ldots, x^{p-1})^T$, on $[-1, 1]$. For simplicity, we assume that one of (n, N) is a multiple of the other. If $n < N$, then we also require that they have the same parity. We take equally spaced design points,

$$x_i = -1 + \frac{2(i-1)}{N-1}, \qquad i = 1, \ldots, N. \quad (17)$$

A simulated annealing algorithm, in general, consists of the following:

1. A description of the initial state of the process; that is, of the starting vector of allocations $\mathbf{n} = (n_1, \ldots, n_N)$
2. A scheme by which subsequent states are generated
3. A criterion according to which these subsequent states are accepted or rejected.

If $n > N$, then the initial state is the uniform design, with $n_i = n/N$ for $i = 1, \ldots, N$. If $n \le N$, then this vector of frequencies assigned to $x_1, \ldots, x_N$ starts with the vector formed by repeating the vector $(1, 0, \ldots, 0)$ (with $N/n - 1$ 0's) $[n/2]$ times. This is followed by the same vector with the order of its elements reversed. If N is odd, then also a vector $(0, \ldots, 0, 1, 0, \ldots, 0)$ of length N/n is inserted in the middle. Thus in either case the initial design is symmetric and at least close to uniform. We impose symmetry on the designs largely for its intuitive appeal. However, we remark that searches for better, possibly asymmetric, designs have yielded no improvements.

To generate new designs, first define $\mathbf{v}$ to be the $[N/2] \times 1$ vector consisting of the initial segment $(n_1, \ldots, n_{[N/2]})$ of the current allocation vector. Define

$$J_+ = \{i | v_i > 0\}, J_0 = \{i | v_i = 0\}$$

with cardinalities $j_+ \ge 1$ and j_0. Generate a Bernoulli random variable,

$$B = \begin{cases} 1, & \text{with probability } \frac{j_0}{j_0 + j_+} \\ 0, & \text{with probability } \frac{j_+}{j_0 + j_+}. \end{cases}$$

If $j_+ \ge 2$, then choose two indices (t_1, t_2) from J_+, at random and without replacement, and (if $B = 1$) pick an index t_0 from J_0, at random. Replace $\mathbf{v}$ by the vector $\tilde{\mathbf{v}}$ whose elements are those of $\mathbf{v}$ except for

$$\begin{aligned} \tilde{v}_{t_0} &= v_{t_0} + B, \qquad \tilde{v}_{t_1} = v_{t_1} - 1, \qquad \text{and} \\ \tilde{v}_{t_2} &= v_{t_2} + 1 - B. \end{aligned} \quad (18)$$

If $j_+ = 1$, then pick t_0 from J_0 at random, let t_1 be the element of J_+, and replace (18) by

$$\tilde{v}_{t_0} = v_{t_0} + 1, \qquad \tilde{v}_{t_1} = v_{t_1} - 1.$$

If N is even, then let $\tilde{\mathbf{n}} = (\tilde{n}_1, \ldots, \tilde{n}_N) = (\tilde{v}_1, \ldots, \tilde{v}_{N/2}, \tilde{v}_{N/2}, \ldots, \tilde{v}_1)$, thus preserving symmetry. If N is odd, then a further Bernoulli experiment is simulated, with probability $1/N$ of "success." If a success is obtained, then, with probability $1/2, n_{[N/2]+1}$ (the frequency assigned to 0) is increased by 2, with these taken randomly and symmetrically from the remaining n_i. With the remaining probability $1/2, n_{[N/2]+1}$ is reduced by 2, with these allocated randomly and symmetrically to the remaining n_i. If $n_{[N/2]+1} < 2$, then this step is omitted. Then $\tilde{\mathbf{n}}$ is constructed as earlier, with also the inclusion of the new frequency $\tilde{n}_{[N/2]+1}$.

To accept or reject $\tilde{\mathbf{n}}$ as the next state, first evaluate the loss $I = I(\tilde{\mathbf{n}})$. If $I(\tilde{\mathbf{n}}) < I(\mathbf{n})$, then $\tilde{\mathbf{n}}$ is accepted and iterations continue. If $\Delta I = I(\tilde{\mathbf{n}}) - I(\mathbf{n}) > 0$, then $\tilde{\mathbf{n}}$ is accepted with probability $\exp(-\Delta I/T)$, where T is a user-chosen parameter. Following Haines (1987), we initially choose T

This content downloaded from 142.104.240.194 on Sun, 03 Feb 2019 01:51:42 UTC
All use subject to https://about.jstor.org/terms

in such a way that the acceptance rate is at least 50%; this parameter is then progressively decreased. As discussed by Press, Flannery, Teukolsky, and Vetterling (1989), we decrease T by a factor of .9 after each 100 iterations.

The timing parameters were decided on partly by trial and error. Decreasing T too quickly—in particular, implementing a pure descent algorithm by setting $T = 0$—often caused the process to become trapped in a local minimum. We satisfied ourselves that the final minimum was global by running the algorithm several times, with a variety of timing parameters. All computations were carried out using S-PLUS; the relevant code is available from the authors on request.

3.2 Examples

We first discuss some limiting cases, where it is possible to compare our results with the known solutions to closely related problems. If $\nu = 0$, so that only bias is a consideration, then by (3) and (9), the best continuous design has $p_i \equiv N^{-1}$; that is, is uniform on $\{x_i\}_{i=1}^N$. With our restriction to integer-valued designs, exact uniformity is not attainable unless n is a multiple of N, and in such cases our annealing algorithm converges to the uniform design. Otherwise, the final designs are approximately uniform, with the approximations serving to bring the total number of observations up to n.

As $\nu \to \infty$, one anticipates that the designs should approach their classically optimal counterparts that minimize variance alone. We have considered in particular the case of approximate cubic regression ($p = 4$) with $n = 20$ and $N = 40$. With a finite design space, there is no previous theory that applies. But because our equally spaced design points (17) discretize the interval $[-1, 1]$, Studden's (1977) results for continuous I-optimal designs over this interval may be taken as a guide. Studden found that the I-optimal design places mass .1545 and .3455 at ± 1 and $\pm .447$. Our algorithm results in a design placing 3 of the 20 observations at each of ± 1 and 7 at each of $\pm .436$ if the errors are homoscedastic (problem P1). The points $\pm .436$ are the nearest, in our design space, to $\pm .447$, so that our algorithm attains the closest approximation to Studden's solution. For heteroscedastic errors (problem P2), our design is very similar to that in Figure 2(b), the only difference being that the sites $\pm .231$ are moved to $\pm .538$.

Figures 1 and 2 illustrate examples of designs obtained using $\nu = 10$ and $\eta^2 = 1$. We have found that the designs change only slowly with changes in ν. Figure 1 displays case P1; Figure 2 displays case P2. A message illustrated in these figures, and obtained by us quite generally, is that in the presence of bias and heteroscedasticity, making new observations at nearby sites seems to be preferable to replication at the same site. These sites are located close to those at which replicates would be taken if bias and heteroscedasticity were ignored and homogeneous variance alone were minimized. The end result is sometimes a near-uniformity, which then also supports the remark of Box and Draper (1959, p. 622) that "the optimal design in typical situations in which both variance and bias occur is very nearly the same as would be obtained if *variance were ignored completely* and the experiment designed so as to *minimize bias alone*."

We have quantified the cost, in terms of lost efficiency, of the robustness given by these designs. Our measure of this efficiency is the ratio of the variance component $[= N^{-1}\sum_{i=1}^N m_i g_i l_i$ in (12)$]$ of the loss when the design is constructed assuming the fitted response to be correct ($f(\cdot) \equiv 0$), to this term for the minimax design. For P1, this resulted in an efficiency of 96.7%. We first compared the minimax design for P2 to the design derived assuming both $f(\cdot) \equiv 0$ and homoscedastic errors, and with both designs evaluated under these assumptions; this gave an efficiency of 90.0%. When we compared this minimax design to the design derived assuming $f(\cdot) \equiv 0$ but heteroscedastic errors, and with both designs then evaluated at their least favorable variance functions, the resulting efficiency was 99.7%. It thus appears that the robustness is obtained quite inexpensively.

4. APPROXIMATE DESIGNS FOR P3; POLYNOMIAL RESPONSE

For problem P3, the transition from the probabilities $\{m_i\}_{i=1}^N$ to the probabilities $\{p_i\}_{i=1}^N$ makes it very difficult to obtain exact integer-valued designs. One could of course apply the methods of the previous section to obtain probabilities m_i with the nm_i being integers, but there is no guarantee that this property would be retained by the p_i. We thus obtain designs without the integer restriction, then discuss ways in which these can be approximated by integer-valued designs.

4.1 Description of the Simulated Annealing Algorithm

We describe the algorithm for N even; in the odd case,

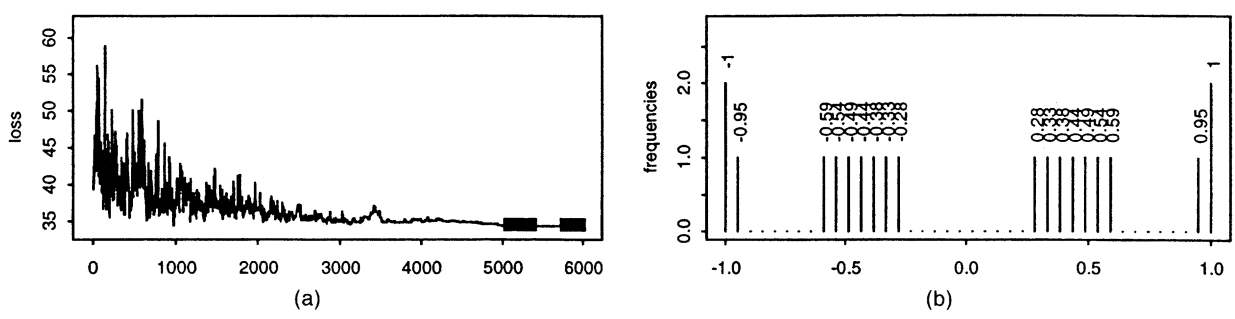

*Figure 1. Exact Integer-Valued Design for Approximate Cubic Regression and Case P1 With n = 20, N = 40, and ν = 10. (a) Loss l(**n**) versus iteration number. Vertical bars above horizontal axis indicate occurrences of minimax design, with loss l = 34.28. (b) Design points and frequencies.*

This content downloaded from 142.104.240.194 on Sun, 03 Feb 2019 01:51:42 UTC
All use subject to https://about.jstor.org/terms

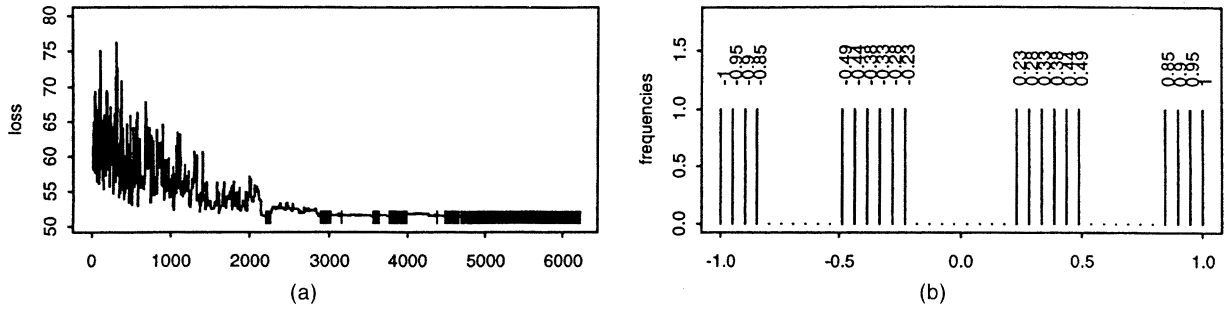

Figure 2. Exact Integer-Valued Design for Approximate Cubic Regression and Case P2 with n = 20, N = 40, and ν = 10. The setup is the same as for Figure 1. Minimax loss is I = 51.41.

an adjustment analogous to that in the previous section is made. We take the same initial state as in Section 3.1, define $\mathbf{v}$ to be the $N/2 \times 1$ vector consisting of the initial segment $(m_1, \ldots, m_{N/2})$ of the current vector $\mathbf{m}$ of probabilities, and define J_+, J_0, and B as before. The next possible state is generated as follows. If $B = 1$ or if $j_+ = 1$, then we pick indices $t_0 \in J_0$ and $t_1 \in J_+$ at random and proceed as in Section 3.1, but replace (18) by

$$\tilde{v}_{t_0} = v_{t_1}, \qquad \tilde{v}_{t_1} = v_{t_0}. \tag{19}$$

If $B = 0$, then we pick indices t_1 and $t_2 \in J_+$, and replace (19) by

$$\tilde{v}_{t_1} = v_{t_1} + \frac{U}{n}, \qquad \tilde{v}_{t_2} = v_{t_2} - \frac{U}{n},$$

where U is a random variable uniformly distributed over [0, 1]. For all other indices, put $v_i = \tilde{v}_i$. Because these $\tilde{v}_i$ may now not be probabilities, we next replace them by $\psi(\tilde{v}_i)/(2\sum_{i=1}^{N/2}\psi(\tilde{v}_i))$ for $i = 1, \ldots, N/2$, where $\psi(v) = \min(\max(v, 0), .5)$. The $\{\tilde{v}_i\}_{i=1}^{N/2}$ so obtained are then in [0, .5] and sum to .5. Finally, let $\tilde{\mathbf{m}} = (\tilde{m}_1, \ldots, \tilde{m}_N) = (\tilde{v}_1, \ldots, \tilde{v}_{N/2}, \tilde{v}_{N/2}, \ldots, \tilde{v}_1)$. The criterion for acceptance or rejection of this and subsequent states is as in Section 3.1.

4.2 Approximation Methods

Once the minimizing probabilities $\{m_i\}_{i=1}^N$ are determined by annealing, the probabilities $\{p_i\}_{i=1}^N$ are determined as in Theorem 3. The problem is now to approximate the allocations $n_i = np_i$ by integers $\hat{n}_i$ in a suitable fashion. We have implemented and compared two approximation methods:

1. The *minimum norm* method (termed the *quota* method in Pukelsheim 1993), as used by Kiefer (1971). This minimizes the l_p norm between the $\{n_i\}$ and the $\{\hat{n}_i\}$, for any p. It is implemented by first rounding down the n_i to their integer parts $[n_i]$, and then distributing the discrepancy $n - \sum[n_i]$ among those x_i for which the fractional parts $n_i - [n_i]$ are the greatest. If the discrepancy is odd, then one more observation is allocated to 0 before this process is carried out.

2. The *multiplier,* or *efficient rounding,* method proposed by Pukelsheim and Rieder (1992). In this method, slightly modified here to preserve symmetry, one first computes frequencies $\check{n}_i = \lceil (n - .5l)p_i \rceil$, where $\lceil \cdot \rceil$ denotes rounding *up* to the next integer and l is the number of $n_i > 0$. Then one loops until the discrepancy $n - \sum \check{n}_i$ is 0, either increasing a frequency $\check{n}_i$ that attains $\check{n}_i/n_i = \min_{n_i>0} \check{n}_i/n_i$ to $\check{n}_i + 1$ or decreasing an $\check{n}_i$ that attains $(\check{n}_i - 1)/n_i = \max_{n_i>0}(\check{n}_i - 1)/n_i$ to $\check{n}_i - 1$. As in the minimum norm method, if the discrepancy is odd, then a preliminary adjustment is made before the looping begins: $\check{n}_{[N/2]+1}$ is decreased by 1 if it is positive, and increased by 1 otherwise. The looping is then applied only to the remaining frequencies.

These increases and decreases are made sequentially and hence are not uniquely determined. In our implementation we have first increased those for which the support points are largest in absolute value, and first decreased those for which the support points are smallest in absolute value. The net effect is to move relatively more mass toward the extremes of the design space.

After obtaining the integer allocations $\hat{n}_i$ by one of these methods, we compute weights w_i as in Lemma 3, and then compute the loss by evaluating (14) at w_i and $m_i = \hat{n}_i w_i / n$.

Lest the reader think that we have missed something obvious, we remark that we investigated the option of computing both integer-valued approximations after each iteration of the annealing algorithm and basing the choice of the next state on the minimum of the two losses associated with these integer designs. This approach typically gave quick convergence to a design with loss about one unit greater than that of the optimum, but then gave no further progress. The reason appears to be that due to the rounding, small changes in the $\{m_i\}$ very often resulted in the same values $\{n_i\}$. There was then no change in the loss and no reason to change states.

Our numerical studies have indicated that the two approximation methods often yield identical or almost identical results. Of course the designer can, and probably should, compute selections of both methods in each application. We have also noticed that the losses of the integer-valued designs can be lower when nonoptimal allocations np_i are being approximated. Indeed, this is implied by the fact that the approximate solution to P2 in Figure 3 does not coincide with the exact integer-valued solution in Figure 2.

4.3 Examples

The methods of this section can be applied to problems P1 and P2 as well, thus affording a means of assessing the quality of the approximation methods. For the model and

This content downloaded from 142.104.240.194 on Sun, 03 Feb 2019 01:51:42 UTC
All use subject to https://about.jstor.org/terms

parameters of Section 3.2, both approximation methods resulted in the same design for P1 as there, illustrated in Figure 1. The loss associated with this design is very close—34.28 versus 34.03—to that of the non–integer-valued design being approximated. For P2, both approximation methods again resulted in the same design, whose loss is about 3% greater than that of the exact integer-valued design. This loss for the exact design in turn exceeds that of the non–integer-valued design by about 3%.

We have obtained weights and designs for approximate cubic regression and problem P3, using the same parameters as were used in Section 3.2 (see Fig. 3(b)–(d)). In this case the multiplier method resulted in a slightly higher loss than the minimum norm method.

Plots (not shown) of the least favorable $\mathbf{f}$ reveal that for all three problems, m_i is roughly proportional to $|f_i|$. In P2 we find that g_i is increasing in $|x_i|$ within each cluster of design points. In P3, however, g_i is decreasing in $|x_i|$ within each such cluster, as is w_i. In this latter case are we then designing for a particularly unlikely contingency, at the cost of protection against more realistic departures from homoscedasticity? To give a partial answer to this question, we assessed the performance of our designs assuming that the true variance function was known to be of the form $g_i^* \propto 1+c|x_i|^d$, normed to satisfy (5) with equality. We computed the maximum (over f) loss of the exact designs for P1 and P2 and the two approximate designs—both with the minimax weights and with optimal weights $w_i^* \propto 1/g_i^*$—for P3. These maxima are displayed in Table 1. Note that with the minimum norm approximation, the minimax weights in fact result in a smaller loss than do the optimal weights. This is presumably due to the bias reduction effected by the minimax weights, as anticipated in the discussion following Theorem 3.

5. DESIGNS FOR FIRST- AND SECOND-ORDER MULTIPLE REGRESSION

In this section we outline a method by which the theoretical and computational methods of the preceding sections can be adapted to the q-variate approximate regression model, and discuss the qualitative features of the resulting designs. For these models, $\mathbf{x} = (x_1, \ldots, x_q)^T$ and $\mathbf{z}(\mathbf{x})$ has elements $1, x_1, \ldots, x_q$ and possibly second order terms $x_i x_j$ $(1 \le i \le j \le q)$. Thus $p = q + 1$ for first-order models, and $p = (q+1)(q+2)/2$ for second-order models. As design space, we take the q-fold Cartesian product $\mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_q$, where each $\mathcal{S}_j$ consists of N_0 equally spaced points in $[-1, 1]$, as at (17) with N replaced by N_0. Thus $N = N_0^q$. There being no a priori reason to prefer one axis to another, we require that the designs be exchangeable and symmetric in each variable. Such designs can be generated by symmetrically choosing n_0 points on the x_1-axis and then forming the q-fold Cartesian product of these points with themselves, whence $n = n_0^q$. We assume that n_0 and N_0 are restricted in the same way as were n and N in Sections 3 and 4. The algorithms of those sections may then be used to choose the n_0 points, so that the computational complexity does not increase with the dimensionality.

Our simulations with $q = 2$ have led to the following observations. Recall that the classical designs that mini-

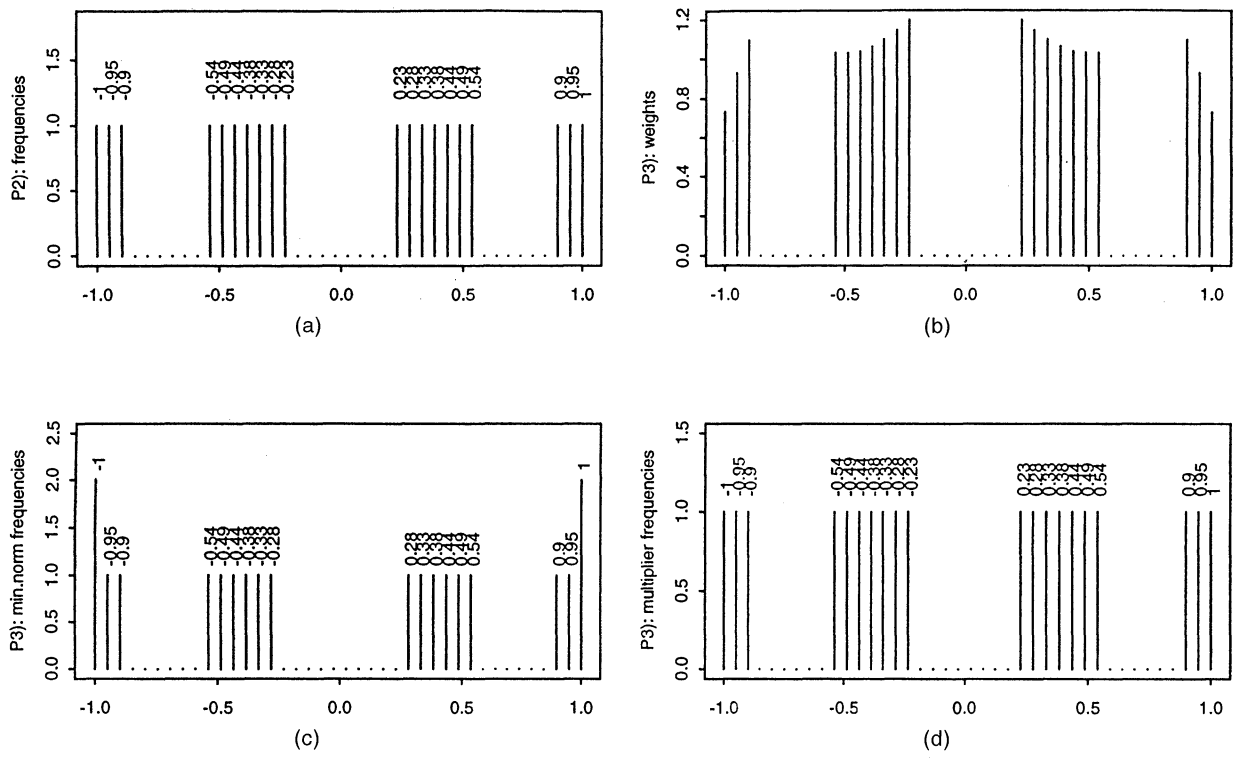

Figure 3. Integer-Valued Approximations, Using the Annealing Scheme and Approximation Methods of Section 4, to the Minimax Designs; n = 20, N = 40, and ν = 10. For P1, both approximations result in the exact design of Figure 1. (a) P2, both approximations, loss l = 52.58. (b) P3, regression weights. (c) P3, minimum norm approximation, loss l = 52.03. (d) P3, multiplier approximation, loss l = 52.99. The losses associated with the non—integer-valued designs being approximated are P1, l = 34.03; P2, l = 49.83; and P3, l = 49.20.

This content downloaded from 142.104.240.194 on Sun, 03 Feb 2019 01:51:42 UTC
All use subject to https://about.jstor.org/terms

Table 1. Losses Associated With Variance Functions $g_i^* \propto 1 + c|x_i|^d$ Under Various Design and Weight Combinations

				P3, minimum norm		P3, multiplier	
c	d	P1	P2	Minimax weights	Optimal weights	Minimax weights	Optimal weights
1	2	36.44	37.75	36.12	36.49	38.71	36.65
1	.5	35.89	37.66	36.49	36.52	38.03	37.81
5	2	36.64	36.58	34.45	35.49	38.80	33.00
5	.5	36.76	38.03	36.83	37.05	38.79	37.50

mize variance alone are formed, in the Cartesian manner described earlier, from sites at $x_1 = \pm 1$ for the first-order model, and as well $x_1 = 0$ for the second-order model. The robust designs for the second-order model move mass away from the corners of the design space and into more central sites. They also do this, but to a lesser extent, for the first-order model. Furthermore, relative to the variance-minimizing designs, the robust designs replace replicates with clusters. For instance, if $n_0 = 7$, we find that the robust first-order designs with $N_0 = 21$ are (for each of P1, P2, and P3) given by the Cartesian product of $(0, \pm.8, \pm.9, \pm 1)$ with itself. Thus there are clusters of nine points in each corner, with as well three points near the middle of each axis and one center point. The second-order designs are instead formed from $(0, \pm.1, \pm.9, \pm 1)$, with four points clustered in each corner, six points in the middle of each axis, and nine points near the center. The solution for P3 gives very little weight to the center point of the first-order design, and weights all other points approximately equally in the other cases. We also note that the rounding mechanisms used for P3 can sometimes slightly alter the Cartesian product property.

Calculations such as those described at the end of Section 3 gave efficiencies of 90.6%, 90.6%, and 100% for the first-order designs, and 93.3%, 93.3%, and 100% for the second-order designs. The robust designs thus seem very sensible. Their emphasis on near, rather than exact, replication protects against model bias and allows for the estimation of alternate models; the clustering makes them fairly efficient when the fitted model is in fact correct.

6. EXTRAPOLATION DESIGNS

In this section we consider designs for the extrapolation of the estimates of the mean response, determined from observations made within the design space $\mathcal{S}$, to an extrapolation region $\mathcal{T}$ disjoint from $\mathcal{S}$. Extrapolation is an inherently risky procedure, exacerbated by an overreliance on model assumptions; for this reason, robustness against model violations is particularly important in such applications.

Designs for extrapolation of polynomials, assuming a correctly specified response, were studied by Hoel and Levine (1964) and Kiefer and Wolfowitz (1964a,b). Studden (1971) studied such problems for multivariate polynomial models. Dette and Wong (1996) and Spruill (1984) constructed extrapolation designs for polynomial regression, robust against various misspecifications of the degree of the polynomial. Huber (1975) obtained designs for extrapolation of a response, assumed to have a bounded derivative of a certain order but to be otherwise arbitrary, to one point outside of the design interval. These results were corrected and extended by Huang and Studden (1988). Draper and Herzberg (1973) extended the methods of Box and Draper (1959) to extrapolation under response uncertainty. In their approach, one estimates a first-order model but designs with the possibility of a second-order model in mind; the goal is extrapolation to one fixed point outside of the spherical design space. In earlier work (Fang and Wiens 1999), we obtained approximate (i.e., continuous) designs robust against departures from linearity and homoscedasticity similar to those entertained in this article. The goal there was extrapolation to a region of positive Lebesgue measure.

Our model is as described by (1)–(5) for $\mathbf{x} \in \mathcal{S}$. For $\mathbf{x} \in \mathcal{T}$, it is given by $Y = \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta} + f_T(\mathbf{x}) + \varepsilon$, where the function f_T is constrained only by its $L_2(\mu)$ norm,

$$\int_{\mathcal{T}} f_T^2(\mathbf{x})\mu(d\mathbf{x}) \le \eta_T^2, \tag{20}$$

for some measure μ and a given constant η_T^2. Important special cases are $\mathcal{T}$, an interval; μ, Lebesgue measure; and $\mathcal{T}$, a point at which μ places unit mass. As a loss function, we take the maximum, over all f_T satisfying (20), value of the integrated mean squared prediction error (IMSPE),

$$I_T(f, g, w, m) = \max_{f_T} \int_{\mathcal{T}} E[(\hat{Y}(\mathbf{x}) - E[Y|\mathbf{x}])^2]\mu(d\mathbf{x}).$$

Define a $p \times p$ matrix, $\mathbf{A}_T = \int_{\mathcal{T}} \mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})\mu(d\mathbf{x})$, of rank $q \le p$. Let $\mathbf{A}_T^{1/2}$ be a $p \times q$ square root of $\mathbf{A}_T$ and define $\mathbf{Q}_{p\times q} = \boldsymbol{\Lambda}^{-1}\mathbf{V}^T\mathbf{A}_T^{1/2}$. Let

$$\tilde{l}_i = (\mathbf{U}\mathbf{M}_1^{-1}\mathbf{Q}\mathbf{Q}^T\mathbf{M}_1^{-1}\mathbf{U}^T)_{ii}, \qquad i = 1, \ldots, N,$$

and define $r_{T,S} = \eta_T/(\sqrt{N}\eta)$. Using this notation, we first calculate that

$$\begin{aligned} I_T(f, g, w, m) &= \max_{f_T} \left\{ \operatorname{tr}(\mathbf{C}\mathbf{A}_T) + \mathbf{d}^T\mathbf{A}_T\mathbf{d} - 2\int_{\mathcal{T}} \mathbf{d}^T\mathbf{z}(\mathbf{x})f_T(\mathbf{x})\mu(d\mathbf{x}) \right. \\ &\qquad \left. + \int_{\mathcal{T}} f_T^2(\mathbf{x})\mu(d\mathbf{x}) \right\}. \end{aligned} \tag{21}$$

The maximum is achieved by requiring the sign of $f_T(\mathbf{x})$ to be opposite that of $\mathbf{d}^T\mathbf{z}(\mathbf{x})$, by requiring equality in (20), and by then maximizing $|\int_{\mathcal{T}} \mathbf{d}^T\mathbf{z}(\mathbf{x})f_T(\mathbf{x})\mu(d\mathbf{x})|$ by applying the Cauchy–Schwarz inequality. The maximizing f_T is given by $f_T(\mathbf{x}) = -\eta_T\mathbf{d}^T\mathbf{z}(\mathbf{x})/\sqrt{\mathbf{d}^T\mathbf{A}_T\mathbf{d}}$. Substituting this into (21), and evaluating the trace, gives

$$I_T(f, g, w, m) = N\eta^2 \left\{ \left(\sqrt{\frac{\mathbf{d}^T\mathbf{A}_T\mathbf{d}}{N\eta^2}} + r_{T,S} \right)^2 + \frac{\nu}{N} \sum_{i=1}^{N} m_i w_i g_i \tilde{l}_i \right\}. \tag{22}$$

6.1 Extrapolation Designs Solving Problems P1, P2, and P3

Proceeding as in Section 2.1, we find that the maximum

This content downloaded from 142.104.240.194 on Sun, 03 Feb 2019 01:51:42 UTC
All use subject to https://about.jstor.org/terms

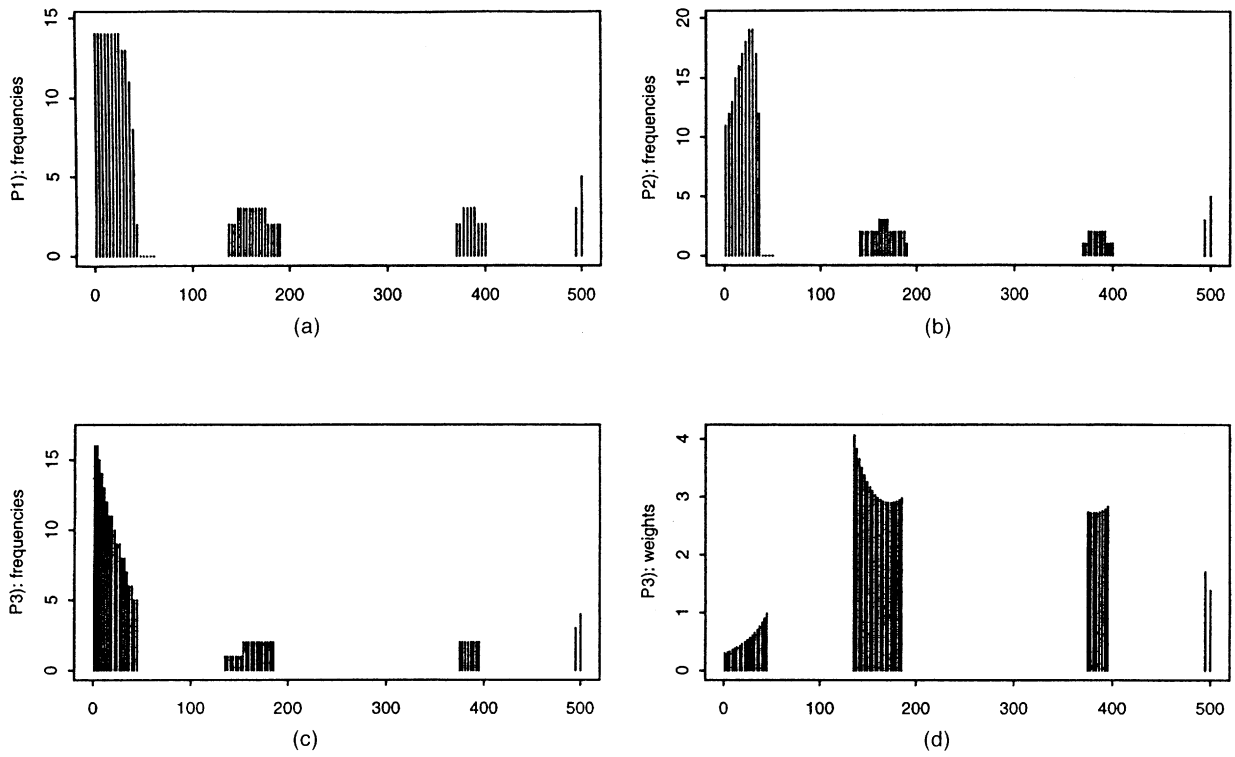

Figure 4. Designs for Extrapolation From [1, 500] to {.5}; n = 235, N = 47, ν = 10, and $r_{T,S}$ = 1. (a), (b), (c) Designs for P1, P2, and P3 (minimum norm approximation). (d) Weights for P3.

value of $\mathbf{d}^T\mathbf{A}_T\mathbf{d}$, over functions f satisfying (3) and (4), is $N\eta^2\lambda_{m,T}$, where $\lambda_{m,T}$ is the largest characteristic root of the $q \times q$ matrix $\mathbf{Q}^T(\mathbf{M}_1^{-1}\mathbf{M}_2\mathbf{M}_{1\tilde{}}^{-1} - \mathbf{I})\mathbf{Q}$. Lemmas 2 and 3 remain valid with l_i replaced by $\tilde{l}_i$. The analog of (14) thus becomes

$$\max_{f,g} I_T(f,g,w,m)$$

$$= N\eta^2\left\{(\sqrt{\lambda_{m,T}} + r_{T,S})^2 + \frac{\nu}{\sqrt{N}}\left(\sum_{i=1}^N m_i^2 w_i^2 \tilde{l}_i^2\right)^{1/2}\right\}, \quad (23)$$

and the following results are immediate.

Theorem 4. For OLS estimation with homoscedastic errors, $(\mathcal{G} = \{\mathbf{1}\}, \mathcal{W} = \{\mathbf{1}\})$, we have

$$\max_f I_T(f,\mathbf{1},\mathbf{1},m)$$

$$= N\eta^2\left\{(\sqrt{\lambda_{m,T}} + r_{T,S})^2 + \frac{\nu}{N}\sum_{i=1}^N m_i\tilde{l}_i\right\}, \quad (24)$$

and so the minimax extrapolation design for P1 has $\{p_i\}_{i=1}^N = \{m_i\}_{i=1}^N$, where $\{m_i\}_{i=1}^N$ minimizes (24). For computational purposes, we note that $\sum_{i=1}^N m_i\tilde{l}_i = \mathrm{tr}(\mathbf{Q}^T\mathbf{M}_1^{-1}\mathbf{Q})$.

Theorem 5. For OLS estimation $(\mathcal{W} = \{\mathbf{1}\})$ with heteroscedastic errors, we have

$$\max_{f,g} I_T(f,g,\mathbf{1},m)$$

$$= N\eta^2\left\{(\sqrt{\lambda_{m,T}} + r_{T,S})^2 + \frac{\nu}{\sqrt{N}}\left(\sum_{i=1}^N m_i^2\tilde{l}_i^2\right)^{1/2}\right\}, \quad (25)$$

and so the minimax extrapolation design for P2 has $\{p_i\}_{i=1}^N = \{m_i\}_{i=1}^N$, where $\{m_i\}_{i=1}^N$ minimizes (25).

Theorem 6. For WLS estimation with heteroscedastic errors, we have

$$\min_w \max_{f,g} I_T(f,g,w,m)$$

$$= N\eta^2\left\{(\sqrt{\lambda_{m,T}} + r_{T,S})^2 + \frac{\nu}{\sqrt{N}}\left(\sum_{i=1}^N m_i^{4/3}\tilde{l}_i^{2/3}\right)^{3/2}\right\}. \quad (26)$$

The minimax extrapolation design $\{p_i\}_{i=1}^N$ for P3 has $p_i \propto m_i^{4/3}\tilde{l}_i^{2/3}$, where $\{m_i\}_{i=1}^N$ minimizes (26). The least favorable variances satisfy $g_i \propto \sqrt{p_i}$, and the optimum weights satisfy $w_i \propto m_i/p_i$ whenever $m_i > 0$.

Corollary 2. The design with $p_i \propto ((\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{A}_T(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T)_{ii})^{2/3}$ and $w_i \propto p_i^{-1}$ minimizes the maximum IMSPE, subject to the side condition that $E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$ for all f.

6.2 Case Study

Consider the following extrapolation problems for bioassays or dose–response experiments. Let $P(x)$ be the probability of a particular response when a drug or carcinogen is administered at dose x. At various levels of x, one

This content downloaded from 142.104.240.194 on Sun, 03 Feb 2019 01:51:42 UTC
All use subject to https://about.jstor.org/terms

observes the proportion p_x of subjects exhibiting the response, and transforms to the p_x-quantile $Y = G^{-1}(p_x)$ for a suitable distribution G. If G is the logistic distribution, then one obtains the logit model; G as the normal distribution gives the probit model. The regression function $E[Y|x] = E[G^{-1}(p_x)]$ is then approximated by $G^{-1}(P(x))$. Because $P(x)$ is unknown, a further approximation, $E[Y|x] \approx \zeta(x)$, is often made, where $\zeta(x)$ is a polynomial, typically of low degree. Of course, var$[Y|x]$ will also vary with x, due to the nature of the data as proportions and to the transformation. The model of Section 1.1 would then seem to be quite appropriate.

In the "low-dose" problem, it is difficult or impossible to observe Y near $x = 0$, or the error variance increases markedly as $x \to 0$. Either of these situations leads to the extrapolation of estimates computed from data observed at, say, $x \in [a, b]$ $(a > 0)$ to estimate $E\ [Y|x = t]$ for small nonnegative values of $t < a$. A related problem is that of estimating the excess probability $P(x) - P(0)$ of a subject exhibiting the response on continuous exposure at dose x. A third problem involves determination of a "virtually safe dose" (Cornfield 1977) below which the excess probability will be less than a specified quantity.

Hoel and Jennrich (1979) obtained optimal designs for these problems, assuming $G^{-1}(P(x)) = -\log(1 - P(x))$ to be an exact polynomial in x and assuming the variance function, derived by the delta method, to be exact in finite samples. This variance function depends on the unknown parameters and so was estimated to determine the design by inserting the estimates from a prior experiment. Krewski, Bickis, Kovar, and Arnold (1986) considered designs for low-dose problems assuming that $E[Y|x]$ was exactly linear in $\ln x$. Lawless (1984) obtained designs that minimize the MSPE of $\hat{Y}_{|x=0}$, for various trial values of $E\ [Y|x = 0] - \zeta(0)$. Of course, this difference is unknown; our approach is to model it [by $f_T(0)$] in such a way as to allow a minimax treatment. Lawless (1984) reached qualitative conclusions very similar to ours, remarking that "in extrapolation problems, a slight degree of model inadequacy quickly wipes out advantages that minimum variance designs possess when the model is exactly correct."

An example considered by Hoel and Jennrich (1979) concerns an experiment discussed by Guess, Crump, and Peto (1977) in which a cubic polynomial is to be fitted to data from $n = 235$ responses to various doses x, with x ranging over [1, 500]. The extrapolation region is the point $\mathcal{T} = \{.5\}$. We first obtained a design for problem P1 in this situation, with design space comprising $N = 3n = 705$ points equally spaced over [1, 500] and a measure μ placing unit mass at .5. This resulted in four clusters of points, with each point allocated one observation. The clusters ranged over [1, 60], [137, 190], [370, 400], and [494, 500]. Because some replication is necessary for experimentation of this type, we then reran the program with a restricted design space of $N = n/5 = 47$ points, roughly uniformly distributed over these intervals. This yielded the final design, illustrated in Figure 4(a) with more details in Table 2. We carried out a similar process for P2 and for P3. In all cases we took $\nu = 10, r_{T,S} = 1$, and $\eta^2 = 1$.

For each case, we carried out 100 runs, each consisting of 100 iterations of the algorithms described in Sections 3.1 and 4.1. We repeated this procedure several times to ensure that a minimum had been reached. The computations were carried out using one 360 MHz CPU on a Sun UltraSPARC-II workstation; typical CPU times for these sets of 100 runs were 494, 594, and 504 seconds for P1, P2, and P3 (both approximations).

The resulting designs are illustrated in Figures 4(b) and 4(c), with weights for the latter design given in Figure 4(d). For comparison, the design of Hoel and Jennrich (1979), for a particular prior reflecting a "background effect" (i.e., $P(0) > 0$), placed 63, 125, 35, and 12 observations at $x = 1, 82.6, 342$, and 500. Such a design does not allow for testing the model for lack of fit nor for estimating polynomial responses of degree greater than three. The design of Huang and Studden (1988) for extrapolation of a response assumed to have a bounded fourth derivative can be obtained from their theorem 4.4. It depends on a parameter ρ similar to our ν; for $\rho = 10$, the design places mass .988, .008, .003, and .001, yielding frequencies 232, 2, 1, and 0 at sites 1, 125.75, 375.3, and 500. Thus even the cubic response is not estimable without an adjustment of the design.

We compared the variances of the prediction at $x = .5$ from the designs for P1, P2, and P3 to those of the Hoel–Jennrich design. In each case $g_i \propto P(x_i)/(n_i(1 - P(x_i)))$ was estimated by plugging in the estimated response function $P(x) = 1 - \exp(-.01 - .00026377x)$ obtained by Guess et al. (1977) and used in this way by Hoel and Jennrich (1979). The resulting efficiencies of the robust designs

Table 2. Extrapolation Designs for the Case Study of Section 6.2

Case	*Ranges (frequencies) of blocks of design points*				*Loss*
Extrapolation to $\mathcal{T} = \{.5\}$					
P1	[1, 43] (159)	[137, 190] (60)	[370, 400] (22)	[494, 500] (8)	89.39
P2	[1, 36] (169)	[140, 190] (39)	[370, 400] (19)	[494, 500] (8)	114.43
P3[a]	[1, 45] (181)	[135, 185] (31)	[375, 395] (16)	[495, 500] (7)	85.47[a]
Extrapolation to $\mathcal{T} = [0, .5]$					
P1	[1, 51] (144)	[130, 203] (54)	[359, 410] (27)	[490, 500] (10)	73.54
P2	[1, 43] (157)	[139, 194] (46)	[358, 402] (23)	[492, 500] (9)	83.43
P3[b]	[1, 49] (183)	[138, 191] (31)	[364, 410] (14)	[495, 500] (7)	68.90[b]

[a] Minimum norm approximation; loss for multiplier approximation was 85.92.
[b] Minimum norm approximation; loss for multiplier approximation was 69.87.

This content downloaded from 142.104.240.194 on Sun, 03 Feb 2019 01:51:42 UTC
All use subject to https://about.jstor.org/terms

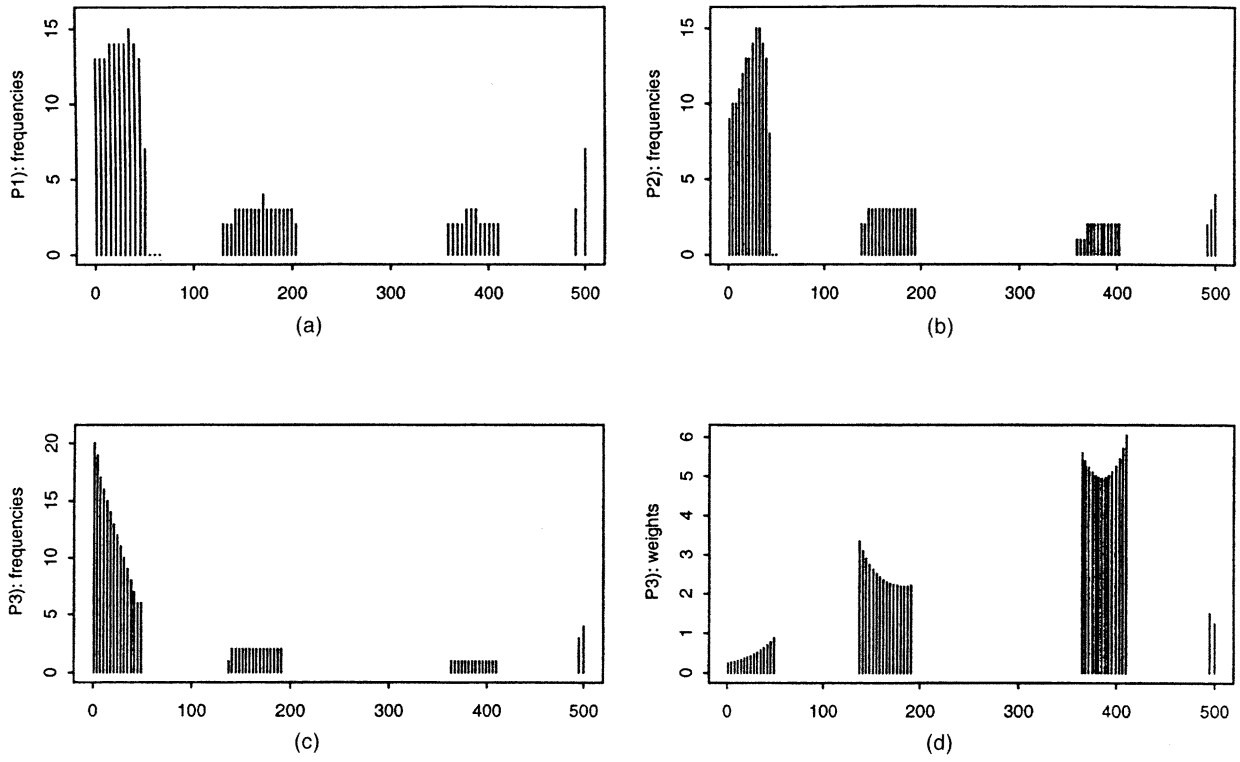

Figure 5. Designs for Extrapolation From [1, 500] to [0, .5]; n = 235, N = 47, ν = 10, and $r_{T;S}$ = 1. (a), (b), (c) Designs for P1, P2, and P3 (minimum norm approximation). (d) Weights for P3.

were 32.3%, 37.0%, and 13.4%. These indicate that the premium paid for the robustness is high, if one has confidence in the cubic model and in the assumed variance function. An indication of the amount of insurance obtained for this premium is given by comparing the prediction variances under homoscedasticity: $g_i \propto 1/n_i$, which gives efficiencies of 109.8%, 124.0%, and 64.7%. If instead g_i is altered by taking a prior response function $P(x) = 1 - \exp(-.1 - .00026377x)$, corresponding to a different prior estimate of the background effect, then the efficiencies are 87.3%, 99.0%, and 45.9%.

We repeated the process just described with the change $\mathcal{T} = [0, .5], \mu =$ Lebesgue measure, for extrapolation to a low-dose region. These designs and weights are given in Figure 5. Relative to those in the design for a one-point extrapolation space, the clusters of sites are somewhat more spread out. Note also that some of the designs continue to call for a number of unreplicated observations; the experimenter would presumably want to group these in some manner, to give replicates. A possibility is to run the programs once again, with a design space containing fewer possible sites in these regions.

For both types of extrapolation regions, the algorithms for P3 did not improve on the initial allocations $m_i \equiv N^{-1}$. These designs then agree with those given by Corollary 2. For larger values of ν, this was no longer the case.

7. CONCLUSIONS AND GUIDELINES

We have presented integer-valued regression designs, and in some cases regression weights, that are at least nearly optimal for various estimation and extrapolation problems. The designs are robust against an incorrectly specified response function and against possible error heteroscedasticity. The use of a finite design space and of a simulated annealing algorithm has greatly facilitated the construction of the designs. In particular, these methods, combined in some cases with the rounding methods used, have allowed us to present exact (i.e. integer-valued) designs in situations where only approximate (continuous) designs were previously obtainable.

A recurring message has been that the designs that protect against these very general forms of model bias and error heteroscedasticity may be approximated by taking the (homogeneous) variance-minimizing designs, which typically have replicates at p sites, and replacing these replicates by clusters of observations at nearby but distinct sites. A judicious application of this maxim alone should enable the experimenter to obtain improved, if not completely optimal, designs in those cases in which the classical design problem has been solved. In less structured problems, the optimal designs can be obtained by applying simulated annealing, as in Sections 3–5 of this article, to the minimization problems outlined in Section 2.

[Received April 1999. Revised January 2000.]

REFERENCES

Bohachevsky, I. O., Johnson, M. E., and Stein, M. L. (1986), "Generalized Simulated Annealing for Function Optimization," *Technometrics*, 28, 209–217.

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*,

This content downloaded from 142.104.240.194 on Sun, 03 Feb 2019 01:51:42 UTC
All use subject to https://about.jstor.org/terms

54, 622–654.

Cornfield, J. (1977), "Carcinogenic Risk Assessment," *Science*, 198, 693–699.

Dette, H., and Wong, W. K. (1996), "Robust Optimal Extrapolation Designs," *Biometrika*, 83, 667–680.

Draper, N. R., and Herzberg, A. (1973), "Some Designs for Extrapolation Outside a Sphere," *Journal of the Royal Statistical Society*, Ser. B, 35, 268–276.

Fang, Z., and Wiens, D. P. (1999), "Robust Extrapolation Designs and Weights for Biased Regression Models With Heteroscedastic Errors," *The Canadian Journal of Statistics*, 27, 751–770.

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

Guess, H., Crump, K., and Peto, R. (1977), "Uncertainty Estimates for Low-Dose Extrapolation of Animal Carcinogenicity Data," *Cancer Research*, 37, 3475–3483.

Haines, L. M. (1987), "The Application of the Annealing Algorithm to the Construction of Exact Optimal Designs for Linear-Regression Models," *Technometrics*, 29, 439–447.

Hoel, P. G., and Levine, A. (1964), "Optimal Spacing and Weighting in Polynomial Prediction," *Annals of Mathematical Statistics*, 35, 1533–1560.

Hoel, P. G., and Jennrich, R. I. (1979), "Optimal Designs for Dose–Response Experiments in Cancer Research," *Biometrika*, 66, 307–316.

Huang, M. N. L., and Studden, W. J. (1988), "Model Robust Extrapolation Designs," *Journal of Statistical Planning and Inference*, 18, 1–24.

Huber, P. J. (1975), "Robustness and Designs," in *A Survey of Statistical Design and Linear Models*, ed. J. N. Srivastava, Amsterdam: North-Holland, pp. 287–303.

Kiefer, J. C. (1971), "The Role of Symmetry and Approximation in Exact Design Optimality," in *Statistical Decision Theory and Related Topics, Proceedings of a Symposium, Purdue University 1971*, eds. S. S. Gupta and J. Yackel, New York: Academic Press, pp. 109–118.

Kiefer, J., and Wolfowitz, J. (1964a), "Optimum Extrapolation Designs, I," *Annals of the Institute of Statistical Mathematics*, 16, 79–108.

——— (1964b), "Optimum Extrapolation Designs II," *Annals of the Institute of Statistical Mathematics*, 16, 295–303.

Krewski, D., Bickis, M., Kovar, J., and Arnold, D. L. (1986), "Optimal Experimental Designs for Low-Dose Extrapolation. I. The Case of Zero Background," *Utilitas Mathematicas*, 29, 245–262.

Lawless, J. F. (1984), "Some Problems Concerning Experimental Designs for Extrapolation," in *Topics in Applied Statistics*, ed. T. D. Dwivedi, Montreal: Concordia University Press, pp. 357–366.

Li, K. C. (1984), "Robust Regression Designs When the Design Space Consists of Finitely Many Points," *The Annals of Statistics*, 12, 269–282.

Li, K. C., and Notz, W. (1982), "Robust Designs for Nearly Linear Regression," *Journal of Statistical Planning and Inference*, 6, 135–151.

Marcus, M. B., and Sacks, J. (1976), "Robust Designs for Regression Problems," in *Statistical Theory and Related Topics II*, eds. S. S. Gupta and D. S. Moore, New York: Academic Press, pp. 245–268.

Meyer, R. K., and Nachtsheim, C. J. (1988), "Constructing Exact D-Optimal Experimental Designs by Simulated Annealing," *American Journal of Mathematical and Management Sciences*, 3 & 4, 329–359.

Pesotchinsky, L. (1982), "Optimal Robust Designs: Linear Regression in R^k," *The Annals of Statistics*, 10, 511–525.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1989), *Numerical Recipes: The Art of Scientific Computing*, Cambridge, U.K.: Cambridge University Press.

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

Pukelsheim, F., and Rieder, S. (1992), "Efficient Rounding of Approximate Designs," *Biometrika*, 79, 763–770.

Spruill, M. C. (1984), "Optimal Designs for Minimax Extrapolation," *Journal of Multivariate Analysis*, 15, 52–62.

Studden, W. J. (1971), "Optimal Designs for Multivariate Polynomial Extrapolation," *Annals of Mathematical Statistics*, 42, 828–832.

——— (1977), "Optimal Designs for Integrated Variance in Polynomial Regression," in *Statistical Decision Theory and Related Topics II*, eds. S. S. Gupta and D. S. Moore, New York: Academic Press, pp. 411–420.

Wiens, D. P. (1990), "Robust, Minimax Designs for Multiple Linear Regression," *Linear Algebra and Its Applications, Second Special Issue on Linear Algebra and Statistics*, 127, 327–340.

——— (1992), "Minimax Designs for Approximately Linear Regression," *Journal of Statistical Planning and Inference*, 31, 353–371.

——— (1998), "Minimax Robust Designs and Weights for Approximately Specified Regression Models With Heteroscedastic Errors," *Journal of the American Statistical Association*, 93, 1440–1450.

This content downloaded from 142.104.240.194 on Sun, 03 Feb 2019 01:51:42 UTC
All use subject to https://about.jstor.org/terms