UNIVERSITY OF ALBERTA

Robust Extrapolation Designs for Linear Models

By

Zhide Fang ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**

in

STATISTICS

DEPARTMENT OF MATHEMATICAL SCIENCES

EDMONTON, ALBERTA

FALL 1999

0-612-46835-6

UNIVERSITY OF ALBERTA

LIBRARY RELEASE FORM

NAME OF AUTHOR: **Zhide Fang**

TITLE OF THESIS: **Robust Extrapolation Designs for Linear Models**

DEGREE: **Doctor of Philosophy**

YEAR THIS DEGREE GRANTED: **1999**

Department of Mathematical Sciences

University of Alberta

Edmonton, Canada T6G 2G1

Date: Aug. 16, 1999

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommended to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Robust Extrapolation Designs for Linear Models** submitted by Mr. Zhide Fang in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics.

Dr. D.P. Wiens (Supervisor)

Dr. B.A. Schmuland (Chair)

Dr. D.H. Kelker

Dr. E. Gombay

Dr. Michael R.W. Dawson

BS.

Dr. William Welch (External Reader)
University of Waterloo

Date: 99-8-10

ABSTRACT

We construct optimal designs for the extrapolation of regression responses. These designs are robust against small departures from both the assumed linear regression model and the assumption of homoscedasicity in the errors. We derive the robust extrapolation designs for different kinds of model-violation classes.

First, we assume the model departure is bounded in $\mathcal{L}_2$-norm. With the assumption of homoscedasicity in the errors, we obtain the minimax extrapolation designs, the bounded variance designs and the bounded bias designs for multiple linear regression without interaction terms in Chapter 2. When the errors are heteroscedastic, we construct the minimax extrapolation designs for ordinary Least Squares estimation and the minimax extrapolation designs and weights for weighted Least Squares estimation for multiple linear regression with no interactions in Chapter 3. We also exhibit optimal unbiased extrapolation designs and weights for general regression models. All the designs in these two chapters are approximate, i.e. absolutely continuous because the contaminant class is so full. By assuming polynomial fitted model and design interval as $[-1, 1]$, we find that the limiting density of optimal unbiased extrapolation design has the support points of classical D-optimal design as its modes when the extrapolation region is sufficiently "large". Methods for implementing the designs are discussed.

Second, with a rather thin model-violation class and the assumption of homosceda-

sicity in the errors, we seek robust optimal designs for extrapolation of polynomial regression response. For straightline and quadratic fitted models, we observe that robust extrapolation designs turn out to be the minimum veriance extrapolation designs and coincide with the D_1-optimal designs when the extrapolation region is sufficiently "large". For cubic or higher degree fitted models, the designs can be constructed numerically.

Finally, we discuss the possible applications of our designs to the low-dose extrapolation problems using a real data set in the literature.

Dedicated

to my mother,

my brothers and sisters - Zhixin, Zhiyong, Yanzhen, Yanyu,

my wife Catherine and my daughter Stephanie.

ACKNOWLEDGEMENTS

It is one of the milestones in my life to study in the Department of Mathematical Sciences at the University of Alberta, which provides congenial research environment. Especially, I am deeply grateful to my supervisor, Dr. D. P. Wiens, for his tremendous support, constant encouragement, invaluable guidance and friendship. It is him who brings me into the area of Robust Statistics.

I wish to acknowledge with thanks the members of my examining committee Dr. M. Dawson, Dr. E. Gombay, Dr. D. Kelker, Dr. B. Schmuland, Dr. W. Welch for their thorough review and helpful suggestions.

I like to thank the academic staff of the Statistics Centre, Department of Mathematical Sciences for their great help in my study. The staff of the Mathematical Sciences general office are very nice and patient.

I thank my colleagues, Dr. Giseon Heo, Dr. Alwell Oyet and Dr. Julie Zhou, with whom I shared valuable discussions and good humor.

I would also like to thank Dr. D. P. Wiens, Dr. P. Hooper, the Department of Mathematical Sciences and the University of Alberta for their generous financial support and travel grants.

Table of Contents

Chapter One

Introduction

In this dissertation we study the construction of optimal designs for the extrapolation of regression responses when bias is present and the errors are possibly heteroscedastic. The work consists of three papers which have been prepared for publication. Each chapter, from Chapter 2 to Chapter 4, is an independent paper. Chapter 5 contains applications and conclusions. This chapter presents a literature review of (robust) optimal (extrapolation) designs and the motivation for the dissertation.

Section 1 gives the setup of the classical linear regression model and the model assumptions. The classical design problems are introduced. In Subsection 1.1, we present the commonly used criteria of optimality. The classical optimal designs corresponding to different kinds of criteria are given in Example 1 and Example 2. In Subsection 1.2, we discuss the necessity of considering extrapolation in some practical situations. Some optimal designs for extrapolation in the literature follow. Section 2 explains what the robust designs provide protection against and why we need to construct these designs. Different kinds of model-violation classes are introduced. This section is also divided into two subsections - reviews and examples of the robust optimal designs and robust extrapolation designs. The robust optimal designs for extrapolation studied in this dissertation are summarized in Section 3.

1 Classical design problems

In real life, we encounter many relations among variables. For example, how the son's height depends on his parents' average height; how the strength of steel relates to the amounts of some elements and how chemical components (carbon, manganese, copper, nickel, chromiun and molybdenum) affect the physical properties of rebar, etc. Suppose a linear relationship

$$E[Y|\mathbf{x}] = \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}$$

exists between a response variable Y and an experimental vector $\mathbf{x} \in \mathcal{R}^q$, where $\mathbf{z}(\mathbf{x}) \in \mathcal{R}^{p+1}$ is a vector of regressors and $\boldsymbol{\theta} \in \mathcal{R}^{p+1}$ is a vector of unknown parameters. In order to estimate $\boldsymbol{\theta}$ and explain certain aspect of this relationship, measurements on Y are to be made for each of n points $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$, which need not be distinct. Because of the experimental errors, the observations $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n)$ follow a statistical linear model

$$y_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\theta} + \varepsilon_i, \quad i = 1, 2, ..., n, \tag{1}$$

where the ε_i are uncontrollable random errors. Two classical assumptions made to model (1) are

A1. The regression response $E[Y|\mathbf{x}] = \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}$ is exactly correct.

A2. The errors ε_i are uncorrelated and have variance σ^2.

Define $\mathbf{X} = (\mathbf{z}(\mathbf{x}_1), \mathbf{z}(\mathbf{x}_2), ..., \mathbf{z}(\mathbf{x}_n))^T$ and $\mathbf{y} = (y_1, y_2, ...y_n)^T$. Under assumptions A1 and A2, the Least Squares (LS) estimate $\hat{\boldsymbol{\theta}}_{LS}$, given by

$$\hat{\boldsymbol{\theta}}_{LS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y},$$

is the best linear unbiased estimate of $\boldsymbol{\theta}$ in that among all linear unbiased estimates, the covariance matrix of $\hat{\boldsymbol{\theta}}_{LS}$ is smallest in the sense of Loewner ordering of symmetric matrices. Furthermore, we can remove the restriction on linearity if the random errors are assumed to be normally distributed. This implies that $\hat{\boldsymbol{\theta}}_{LS}$ is efficient. But, as we know, $\hat{\boldsymbol{\theta}}_{LS}$ has covariance matrix

$$COV(\hat{\boldsymbol{\theta}}_{LS}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1},$$

which only depends on $\mathbf{X}$, the so-called "model matrix". Hence, the problem of experimental design is to choose the appropriate $\mathbf{X}$ in order that $COV(\hat{\boldsymbol{\theta}}_{LS})$ be as small as possible. We will outline different criteria for optimally choosing $\mathbf{X}$ in the following subsection.

1.1 Optimal designs and examples

Suppose we plan to perform an experiment whose outcome can be described by (1). Let $S \subset \mathcal{R}^q$ be the design space, from which the observation sites $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ are to be chosen. Let ξ be a design measure, that is, the empirical distribution function of $\{\mathbf{x}_i\}_{i=1}^n$:

$$\xi(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\delta_{\mathbf{x}_i}(\mathbf{x}),$$

with $\delta_{\mathbf{x}}$ representing the pointmass 1 at $\mathbf{x}$, for any $\mathbf{x} \in S$. Our aim is to find an optimal design, where the criteria of optimality will be given in the next part of this subsection. In practice, ξ must be a discrete probability measure, with $\xi(\mathbf{x}_i)$, $i = 1, 2, ..., n$, being an integral multiple of n^{-1}. A design with this integral property is called an exact design. But the exact design problem is very difficult and mathematically intractable. A common way to handle this difficulty is to extend the class of designs, allowing ξ to be any probability distribution on S. Define Ξ to be the set of all probability distributions on S. We seek an optimal design $\xi^* \in \Xi$ and hope that an exact design which approximates ξ^* will be close to optimal. This is the viewpoint of the *approximate design theory*, which will be adopted throughout this dissertation.

Now, for any $\xi \in \Xi$, define a matrix

$$\mathbf{B}_\xi = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})d\xi,$$

then $\mathbf{B}_\xi \in NND(p+1)$, the set of all nonnegative definite $(p+1)\times(p+1)$ matrices, and is called the information matrix of ξ. Since $\mathbf{B}_\xi = \frac{1}{n}X^TX$, we have $COV(\hat{\boldsymbol{\theta}}_{LS}) = \frac{\sigma^2}{n}\mathbf{B}_\xi^{-1}$. The optimal designs are determined by minimizing appropriate scalar functions of the covariance matrix, that is, the scalar functions of $\mathbf{B}_\xi^{-1}$, which are called optimality criteria or loss functions. Denote by Ψ the loss function. The most popular design criteria used in the literature are:

a). D-optimality: $\Psi(\mathbf{B}_\xi^{-1}) = det(\mathbf{B}_\xi^{-1})$, the determinant criterion.

b). A-optimality: $\Psi(\mathbf{B}_\xi^{-1}) = \frac{1}{k+1} trace(\mathbf{B}_\xi^{-1})$, the average-variance criterion.

c). E-optimality: $\Psi(\mathbf{B}_\xi^{-1}) = \lambda_{max}(\mathbf{B}_\xi^{-1})$, the maximum-eigenvalue criterion.

Given $d(\mathbf{x}, \xi) = \mathbf{z}^T(\mathbf{x})\mathbf{B}_\xi^{-1}\mathbf{z}(\mathbf{x})$, $\mathbf{x} \in S$, we have two other criteria:

d). G-optimality: $\Psi(\mathbf{B}_\xi^{-1}) = max_{\mathbf{x}\in S} d(\mathbf{x}, \xi)$, the global criterion.

e). Q-optimalty (or I-optimality): $\Psi(\mathbf{B}_\xi^{-1}) = \int_S d(\mathbf{x}, \xi) d\mathbf{x}$, the average-variance of the estimate of the response surface criterion.

An elegant result of Kiefer and Wolfowitz (1960) shows the equivalence of D-optimality and G-optimality. Followings are some examples of optimal designs in the literature. For details, we refer to Fedorov (1972), Studden (1977) and Pukelsheim (1993).

Example 1. With $q = 1$, $S = [-1, 1]$ and $\mathbf{z}(x) = (1, x, ..., x^p)^T$– the p^{th}-degree polynomial regression model, the D-optimal design assigns equal weight $(1/(p+1))$ to the $(p+1)$ designs points $\{x_i\}_{i=0}^p$ which are the roots of the equation

$$(1 - x^2)\frac{\partial}{\partial x} P_p(x) = 0,$$

with $P_p(x)$ being the p^{th}-degree Legendre polynomial. For example, these points are $< \pm 1, \pm 0.447 >$ when $p = 3$ and $< \pm 1, \pm 0.655, 0 >$ when $p = 4$. The A-optimal designs and the E-optimal designs also have $(p+1)$ support points. The analytic construction of these two kinds of designs are possible only in the simplest cases. But it is possible to give a numerical construction by using iterative procedures. Here, we just present some special cases. When $p=3$, $\xi_A(\pm 1)=0.151$, $\xi_A(\pm 0.464)=0.349$ and $\xi_E(\pm 1)=0.127$, $\xi_E(\pm 0.5)=0.373$. When $p=4$, $\xi_A(\pm 1)=0.105$, $\xi_A(\pm 0.667)=0.25$, $\xi_A(0)=0.29$ and $\xi_E(\pm 1)=0.093$, $\xi_E(\pm 0.707)=0.248$, $\xi_E(0)=0.318$. The Q-optimal design points consist of the two end-points of design interval and $(p-1)$ interior points which are the roots of the $(p-1)^{th}$-degree polynomial Q_{p-1}, where Q_0, Q_1, Q_2, ... are the polynomials orthogonal with respect to the measure $(1-x^2)dx$. After the design points $\{x_i\}_{i=0}^p$ are calculated, the masses are given by

$$\xi_Q(x_i)=\frac{(\int_S l_i^2(x)dx)^{1/2}}{\sum_{i=0}^p(\int_S l_i^2(x)dx)^{1/2}}, \quad i=0,1,...,p,$$

where $l_i(x)=\Pi_{j\neq i}\frac{(x-x_j)}{(x_i-x_j)}$ is the Lagrange interpolation polynomial.

Example 2. Let $S=[0,1]^p$, and $\mathbf{z}(\mathbf{x})=(x_1,x_2,...,x_p)^T$—the multiple linear model without interaction over the unit square. For simplicity, we only give the optimal designs for $p=2$ (For arbitrary p, we refer to Chapter 14 of Pukelsheim (1993)). The D-optimal design and A-optimal design have the same sets of design support points $\{(1,1)^T, (1,0)^T, (0,1)^T\}$. But the masses are different: $\xi_D((1,1)^T)=\xi_D((1,0)^T)=\xi_D((0,1)^T)=\frac{1}{3}$; $\xi_A((1,1)^T)=1-\frac{4}{3+\sqrt{3}}$, $\xi_A((1,0)^T)=\xi_A((0,1)^T)=\frac{2}{3+\sqrt{3}}$. The E-optimal design is supported by two points $\{\ (1,0)^T),\ (0,1)^T\}$, with equal mass on each point.

1.2 Optimal extrapolation designs and examples

There are situations where the response may be observed at points in a set S (the design space), with S being different from the set T on which the fitted values are of interest. For example, when estimating the potential carcinogenic risks because of exposure to environmental chemicals, Crump (1979) pointed out that direct estimation of the risk at very low levels of exposure would be difficult or impossible. This requires the extrapolation of estimates obtained from the data observed from relatively high levels of exposure. In connection with this extrapolation, the design problem is referred to as the extrapolation design problem.

Under the regression setup (1) and assumptions A1, A2, the Least Squares predictor of $E[Y|\mathbf{x}]$, denoted by $\hat{Y}(\mathbf{x})$, is the minimum variance unbiased linear estimator for any $\mathbf{x} \in T$, with variance

$$Var(\hat{Y}(\mathbf{x})) = \sigma^2 \mathbf{z}^T(\mathbf{x})(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{z}(\mathbf{x}) = \frac{\sigma^2}{n} d(\mathbf{x}, \xi).$$

Hence, if LS prediction is adopted, the extrapolation design problem is to choose ξ minimizing some scalar function of $d(\mathbf{x}, \xi)$, for example, minimizing $max_{\mathbf{x}\in T}(d(\mathbf{x}, \xi))$ or $\int_T d(\mathbf{x}, \xi) d\mathbf{x}$. These are still called the G-optimality or Q-optimality problems.

With $S = [-1, 1]$, $T = [1,\ t]$ or $T = \{t\}$, $t > 1$, and $\mathbf{z}(x) = (1, x, x^2, ..., x^p)^T$, Hoel and Levine (1964) show that the G-optimal extrapolation design is supported by the

Chebyshev oscillation points

$$x_i = -cos(\frac{i\pi}{p}), \qquad i = 0, 1, ..., p, \tag{2}$$

with masses $\xi_G(x_i) \propto l_i(t)$, the Lagrange interpolation polynomial. Kiefer and Wolfowitz (1965) extend the results to cover nonpolynomial regression problems, where the regressors $z_0(\mathbf{x})$, $z_1(\mathbf{x})$, $\cdots$, $z_p(\mathbf{x})$ consist of a Chebyshev system. Spruill (1990) considers the Q-optimal extrapolation design and concludes by numerical comparison that the design whose support is the Chebyshev oscillation points (2) and whose masses $\xi_Q(x_i)$ are proportional to $(\int_T l_i^2(x)dx)^{1/2}$ is very nearly optimal.

2 Robust design problems

The experimenter builds regression models using his knowledge of the experimental phenomena. The classical design and extrapolation design discussed above are obtained by minimizing the variances of related estimates under the assumption that the model is capable of providing a perfect representation for the relationship between the regression response and experimental levels. However, the Least Squares estimator is biased when the model is only approximately correct. Thus, three questions arise: a). Are the classical designs and extrapolation designs still optimal when bias is present in the model? b). Are the model departures so significant that we can assume a new model before constructing the optimal designs and extrapolation designs? c). Can the classical designs provide any opportunity to check the model adequacy?

G. N. P. Box and N. Draper might be the pioneers who considered these problems. In their paper of 1959, they make clear the dangers of using the classical designs when the model is not exactly correct. By fitting a straightline linear regression model when the true response is quadratic, they conclude that "The optimal design in typical situations in which both variance and bias occur is very near the same as would be obtained if variance were ignored completely and the experiment designed so as to minimize the bias alone." (pp. 622, Box and Draper (1959)). In the straightline regression situation, Huber (1981) points out that "deviations from linearity that are too small to be detected are already large enough to tip the balance away from the 'optimal' designs, which assume exact linearity and put observations on the extreme points of the observable range, toward the 'naive' ones which distribute the observations more or less evenly over the entire design space." Meanwhile, Lawless (1984) reaches the conclusion 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."

In the p^{th}-degree polynomial model, we find that the classical designs and extrapolation designs take observations on $(p+1)$ sites. This make it impossible to check the model adequacy by testing the hypothesis H_0 : the model is p^{th}-degree polynomial vs H_1 : the model is p_1^{th}-degree polynomial $(p_1 > p)$, because taking observations in only $(p+1)$ sites makes the coefficients of the $(p+1)^{th}, ..., (p_1)^{th}$-degree terms not estimable.

These three questions make it necessary to construct optimal designs and extrapolation designs, while allowing small departures from model assumptions A1) and A2). These are called robust (extrapolation) designs in the literature. Generally speaking, robust (extrapolation) designs are those which are not sensitive to small departures from model assumptions.

From the practical point of view, any model built to relate a response variable y and explanatory variables $\mathbf{x}$ can only be assumed to be approximately true. This approximation causes the violation of A1). Besides, although the random errors are still uncorrelated, their variances might not be homogeneous and might depend on the observation sites. This violates A2). In the literature, series of papers obtained the optimal (extrapolation) designs which are robust against various violations of model assumptions. We will give a review of these designs in the next two subsections and then outline the designs obtained in this dissertation.

2.1 Robust optimal designs

In an experiment whose outcome can only be presented approximately by a linear regression model (1), the above discussion makes it necessary to introduce a bias term to the model when constructing the optimal designs. Hence, the true model can be phrased as follows,

$$y_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\theta} + f(\mathbf{x}_i) + \varepsilon_i, \qquad i = 1, 2, ..., n, \tag{3}$$

where the random errors ε_i are uncorrelated and have homogeneous variances, and f is unknown and represents the model contamination. Usually, f belongs to a pre-defined class $\mathcal{F}$. The Least Squares estimator $\hat{\boldsymbol{\theta}}$ is no longer unbiased and has bias vector and covariance matrix

$$E[\hat{\boldsymbol{\theta}}] - \boldsymbol{\theta} = \mathbf{B}_\xi^{-1}\mathbf{b}_{f,S}, \qquad COV[\hat{\boldsymbol{\theta}}] = \frac{\sigma^2}{n}\mathbf{B}_\xi^{-1},$$

where $\mathbf{b}_{f,S} = \frac{1}{n}\sum \mathbf{z}(\mathbf{x}_i)f(\mathbf{x}_i) = \int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})d\xi$. Hence, the mean squared error matrix of $\hat{\boldsymbol{\theta}}$ is given by

$$\begin{aligned} MSE(f,\xi) &= E\left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\right] \\ &= \frac{\sigma^2}{n}\mathbf{B}_\xi^{-1} + \mathbf{B}_\xi^{-1}\mathbf{b}_{f,S}\mathbf{b}_{f,S}^T\mathbf{B}_\xi^{-1}. \end{aligned}$$

Therefore, when constructing the optimal designs under model (3), it is natural to define the loss functions by replacing the argument $\mathbf{B}_\xi^{-1}$ (or covariance matrix of $\hat{\boldsymbol{\theta}}$) of the loss functions given in Section 1.2 by $MSE(f,\xi)$, obtaining $\Psi(MSE(f,\xi))$. Then a robust optimal design is obtained by minimizing

$$max_{f\in\mathcal{F}}\Psi(MSE(f,\xi)). \tag{4}$$

With different choices of Ψ given in Section 1.2, these designs are called robust D-, A-, E-, G-, and Q-optimal design. The maximum is taken here because the model contamination function f is unspecified and we want to safeguard against the worst situation. After the milestone paper of Box and Draper (1959), many authors have worked on design construction for the model (3). These robust optimal designs differ in the choice of the class $\mathcal{F}$, the design space S, the regressors $\mathbf{z}(\mathbf{x})$ and the loss functions Ψ.

Huber (1975, 1981) and Wiens (1990, 1991, 1992, 1993, 1994, 1998) consider the class

$$\mathcal{F}_1 = \left\{ f : \quad \int_S f^2(\mathbf{x}) d\mathbf{x} \leq \eta_S^2, \quad \int_S \mathbf{z}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \mathbf{0} \right\}, \tag{5}$$

where η_S is assumed "small". The first condition in (5) is required in order that the contamination function f be "small", so the linear part of (3) is still the dominant term and the errors due to bias do not swamp those due to variance. The second condition is to guarantee that the true $\boldsymbol{\theta}$ is uniquely defined and is made without loss of generality. To see this, suppose a linear function $\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}_0$ of the regressors is the best approximation of an experimental outcome, $E[Y|\mathbf{x}]$, in the sense that

$$\boldsymbol{\theta}_0 = argmin_{\boldsymbol{\theta}} \left\{ \int_S (E[Y|\mathbf{x}] - \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta})^2 d\mathbf{x} \right\}.$$

Define

$$f(\mathbf{x}) = E[Y|\mathbf{x}] - \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}_0,$$

then we have the model (3) and $f(\mathbf{x})$ satisfies

$$\int_S \mathbf{z}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \mathbf{0},$$

under the regularity condition that $\int_S \mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})d\mathbf{x}$ is positive definite. This is the second condition in (5).

It is realized, and proved as Lemma 1 of Wiens (1992), that $max_{f\in\mathcal{F}_1}\Psi(MSE(f,\xi))$ will be infinite if ξ is a discrete probability measure. Thus, designs robust against (5) must have a density.

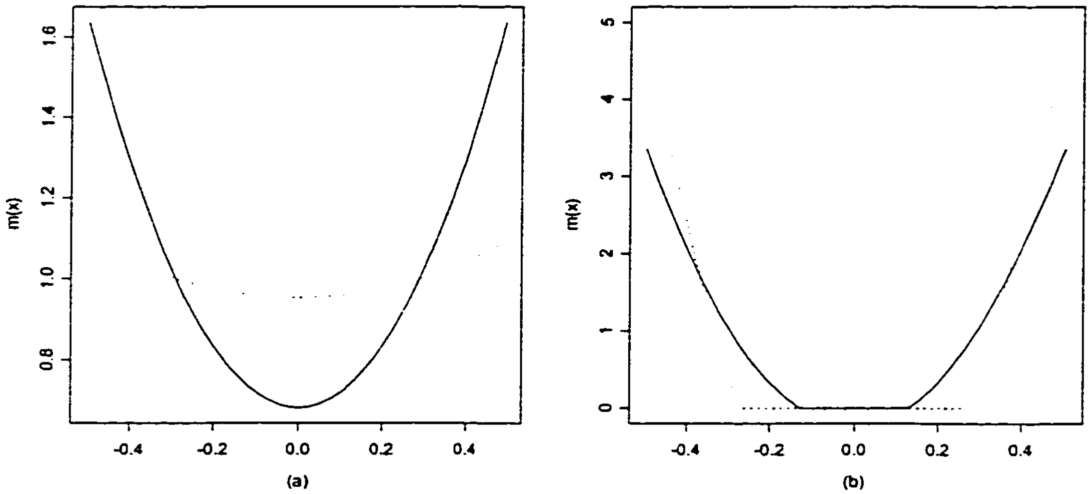
By using the loss function of Q-optimality, Huber (1975) obtains robust optimal designs when $\mathbf{z}(x) = (1, x)^T$, $S = [-0.5, 0.5]$. The robust design ξ_H has density

$$m_H(x) = (ax^2 + b)^+,$$

where a, b depend on the parameter $\nu := \sigma^2/(n\eta_S^2)$. This parameter ν can be interpreted as representing the relative importance of variance versus bias, in the mind of the experimenter. As $\nu \to 0$ bias completely dominates the problem, and ξ_H tends to the uniform continuous design ξ_U which has density $m_U(x) \equiv 1$ on S. As $\nu \to \infty$ the bias term disappears from the model (3), so it becomes a 'pure variance' problem, and ξ_H tends to the classical optimal design which takes half of the observations on each of the extreme points of S. See Figure 1 for some examples of $m_H(x)$.

Wiens (1990) considers the same problem for multiple linear regression, with $\mathbf{z}(\mathbf{x}) = (1, x_1, ..., x_p)^T$ and S a sphere of unit volume. When the design space is a p-dimensional cube, he gives robust Q-optimal designs for the special case of $k = 2$, a bivariate model, with $\mathbf{z}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)^T$. Furthermore, Wiens (1992) constructs the robust D-, A-, E-, G-, and Q-optimal designs for multiple linear regression. These

Figure 1: Huber's robust design density $m(x)$. (a) Small ν : $\nu = 1$ (solid line) and $\nu = 0.1$ (dotted line); (b) Large ν : $\nu = 10$ (solid line) and $\nu = 100$ (dotted line).



robust optimal designs are all absolutely continuous and their density functions depend on $\mathbf{x}$ through $||\mathbf{x}|| := \sqrt{x_1^2 + \cdots + x_p^2}$, which gives a way to implement these designs by first drawing n values of $||\mathbf{x}||$ from the density of $U := ||\mathbf{x}||$, and then choosing $\mathbf{x}_i$ from the uniform distribution on $||\mathbf{x}|| = u_i$, for each i. See Wiens (1992) for details. Wiens (1993) derives robust designs which maximize the minimum coverage probability of confidence ellipsoids. Robust designs for M-estimators are given in Wiens (1994).

Different from the class (5), another typical class of model-violation, defined as

$$\mathcal{F}_2 = \left\{ f : \quad | f(\mathbf{x}) | \leq \phi(\mathbf{x}), \quad \forall \mathbf{x} \in S, \quad \int_S \mathbf{z}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \mathbf{0} \right\}, \tag{6}$$

with various assumptions being made about ϕ, is used in Marcus and Sacks (1976), Sacks and Ylvisaker (1978), Pesotchinsky (1982), Li and Notz (1982), Li (1984) and Liu and Wiens (1997).

With the multiple linear regression setup,

$$y = \theta_0 + \sum_{j=1}^{p} x_j \theta_j + f(\mathbf{x}) + \varepsilon, \tag{7}$$

Li and Notz (1982) assume $\phi(\mathbf{x}) \equiv c$, a positive constant, for any $\mathbf{x} \in S \subset \mathcal{R}^p$. They seek robust optimal designs by minimizing the maximum value of the weighted mean squared error:

$$max_{f \in \mathcal{F}_2} E[\sum_{j=0}^{p} \tau_j (\hat{\theta}_j - \theta_j)^2],$$

where the τ_j are known constants. The robust optimal designs have support points on the boundary of S when $\hat{\theta}_j$ is the Least Squares estimate. If S is a p-dimensional cube, then the robust optimal design puts equal mass on each of the 2^p corners of S. If S is the $(p-1)$-dimensional simplex,

$$S = \left\{ \mathbf{x} = (x_1, ..., x_p)^T : \sum_{j=1}^{p} x_j = 1, \quad x_j \geq 0 \; \textit{for all } j \right\},$$

and if $\theta_0 = \tau_0 = 0$, $\tau_1 = \cdots = \tau_p = 1$, the optimal design assigns equal mass on each of the p extreme points of S.

Pesotchinsky (1982) considers the same model as (7) and requires $\phi(\mathbf{x}) = c\varphi(\mathbf{x})$, with $c > 0$ and $\varphi(\mathbf{x})$ being a convex function of $||\mathbf{x}||^2$. The author restricts attention to the robust D-optimality, E-optimality and A-optimality criteria. Denote by Ξ the class of all designs over S and by $\Xi(m)$ the class of all symmetric designs ξ with fixed $E_\xi(x_i^2) = m$, $i = 1, ..., p$. Let $m_0 = argmin_{m>0}\{(m^{-1}\sigma^2/nc^2)^p(\sigma^2/nc^2 + \varphi^2(\sqrt{mp}))\}$. For D-optimality, Pesotchinsky shows that

a). Any symmetric design $\xi^* \in \Xi(m)$ supported only by the points of the sphere $S_{\sqrt{pm}}$ of radius $\sqrt{pm}$ is robust D-optimal in $\Xi(m)$.

b). If S is compact and symmetric (namely, invariant under permutations and changes of signs of the coordinates) and $d_S = max_{\mathbf{x}\in S}||\mathbf{x}|| = R < \sqrt{pm_0}$, or if $m_0 = \infty$, then there is at least one symmetric robust D-optimal design in $\Xi(R^2/p)$ which is robust D-optimal over S in the class Ξ. Conversely, if S is symmetric compact, then the robust D-optimal designs over S are symmetric.

Unlike in the case of D-optimality, Pesotchinschy does not establish the overall optimality of symmetric designs for the robust A-optimality and E-optimality criteria. However, he proves that the robust A- and E-optimal symmetric designs are unique and with some common conditions, the robust A- and E-optimal designs in Ξ over S are symmetric ones. The results can be outlined as

a). In a class of all symmetric designs, the robust A-optimal (E-optimal) design is a uniform continuous measure on a sphere of radius $\sqrt{\nu_1 p}$ ($\sqrt{\nu_\infty p}$), with ν_1 and ν_∞ defined on pp. 521, Pesotchinsky (1982).

b). Let S contain the sphere of radius $\sqrt{m_1 p}$ (or $\sqrt{m_\infty p}$). Then, if $m_1 \leq a_p$ (or $m_\infty \leq \mu_\infty$), the robust A-optimal (or E-optimal) design in Ξ over S is a symmetric one (uniform on the sphere of radius $\sqrt{m_1 p}$ (or $\sqrt{m_\infty p}$)), where $m_1, m_\infty, \mu_\infty$ are defined on p. 521 and a_p on pp. 514, Pesotchinsky (1982).

Liu and Wiens (1997) study the approximate polynomial regression model

$$y = \sum_{j=0}^{p-1} \theta_j x^j + x^p f(x) + \varepsilon, \tag{8}$$

where $x \in S = [-1,\ 1]$ and $f \in \mathcal{F}_2$, with ϕ being a continuous, even function on S. In this paper, the robust D-optimal designs are given for arbitrary given functions ϕ when $p = 2$ (straightline regression is fitted) and $p = 3$ (quadratic regression is fitted) and $\phi(x) \equiv$ constant when $p \geq 4$. These designs are similar to the classical D-optimal designs in that they have p support points.

The two classes $\mathcal{F}_1$ and $\mathcal{F}_2$ describe the possible violations of model assumption A1. The class $\mathcal{F}_1$ will be used Chapters 2 and 3 while $\mathcal{F}_2$ is adopted in Chapter 4 of this dissertation. One violation of model assumption A2 is that the random errors are possibly heteroscedastic although they are still uncorrelated. Wiens (1998) constructs the robust Q-optimal designs in the presence of both heteroscedasticity and $\mathcal{F}_1$. He considers the following problems

P1) For ordinary Least Squares, seek a minimax design which minimizes the maximum, over both model-violations, value of the integrated mean squared error (IMSE) of the fitted values.

P2) For weighted Least Squares, determine both weights and a design to minimize the maximum IMSE.

P3) Choose weights and design points to minimize the maximum IMSE, subject to a side condition of unbiasedness.

He finds the solutions to P1) and P2) for multiple linear regression with no interaction and a spherical design space. The solution to P3) is given with no restrictions on the regressors and design space. For polynomial regression, the modes of the robust optimal design are exactly the support points of the classical D-optimal design. In this sense, the robust optimal design is a smoothed version of this classical design. We will extend these results to the extrapolation design problem in Chapter 3.

2.2 Robust optimal extrapolation designs

Draper and Herzberg (1973) might have been the first to use the "variance plus bias" methods of Box and Draper (1959) to make design recommendations for robust extrapolation problems. They fit a first-order regression model and choose the design to guard against the possibility of bias from the second-order terms, when S is a q-dimensional ball and T is the line segment from a point t to the closest point of S. The designs are limited to a certain subclass of designs, of which the support points include the intersection point of S and T and the q-simplex points. The robust optimal extrapolation design is given by minimizing the integrated mean squared error of the Least Squares estimator over T. Huber (1975) deals with the extrapolation problem on $S = [0, \infty]$, $T = \{t\}$ with $t < 0$, and assumes an approximately polynomial model (3), with f in a class (for a constant $\tau > 0$),

$$\mathcal{F}_3 = \left\{ f : \ f \, is \ (p+1) \text{ times } \textit{differentiable and} \ \mid f^{(p+1)}(x) \mid \leq \tau, \ x \in [t, \ \infty] \right\}.$$

By using the *MSE* of a linear estimate of the response surface as loss function, he concludes that the robust optimal extrapolation design sits on $(p+1)$ points: $0 = x_1 < \cdots < x_{p+1}$, which after the addition of another point x_{p+2} constitute the set of Chebychev points of order $(p+1)$ in $[0, x_{p+2}]$. Huang and Studden (1988) point out an error in Huber's proof and give an example, but prove that Huber's conclusion is true if the designs considered are limited to exactly $(p+1)$-point designs. Furthermore, they give the formula to calculate the value of x_{p+2} and extend the results to the case $S = [-1, 1]$. With $S = [a, b]$ and $T = \{c\}$, $c > b$, Spruill (1984) has found the robust optimal extrapolation design for the same model as Huber (1975), with $\mathcal{F}$ being a Sobolev space, that is,

$$\mathcal{F}_4 = \left\{ f : \ f^{(p)}(x) \text{ is absolutely continuous and } \int_a^c (f^{(p+1)}(x))^2 dx \leq \tau^2 \right\}.$$

The designs depend on the choice of value $(n\tau^2)^{-1}$. He gives a formula for finding the robust optimal locations when $p = 1$ (approximately straightline regression) and numerical results when $p \geq 2$, but no general formula. Also, Huang (1990) considers the partially linear regression model when the departure satisfies a Lipschitz condition. Dette and Wong (1996) construct robust extrapolation designs when there is uncertainty in the degree of polynomial model, and they propose a new class of optimality criteria rather than using the minimax procedure.

3 Summary of results in this dissertation

Chapter 2 consists of a paper entitled "Robust Optimal Designs for Extrapolation Outside a Hypersphere When Bias is Present." When the model is approximately linear and the contamination term is bounded in $\mathcal{L}_2$ norm (see (3) and (5)), we construct three kinds of robust optimal extrapolation designs for the general multiple regression model without interaction terms: 1) The minimax extrapolation design which minimizes the maximum of the Integrated Mean Squared Prediction Error (*IMSPE*) of the fitted values over the extrapolation space, with the maximum being evaluated over the departure from the model; 2) The bounded variance design which minimizes the maximum of the Integrated Squared Prediction Bias (*ISPB*) subject to bounding the Integrated Squared Prediction Variance (*IPV*); 3) The bounded bias design which minimizes *IPV* subject to bounding the maximum of *ISPB*.

In Chapter 3, entitled "Robust Extrapolation Designs and Weights for Biased Regression Models With Heteroscedastic Errors," we consider the construction of designs for extrapolation, allowing both for possible heteroscedasticity in the errors and for imprecision in the specification of the response function, with the contamination term satisfying the same conditions as in Chapter 2. Three problems are tackled in this paper: 1) For ordinary Least Squares estimation, determine a design to minimize the maximum value of *IMSPE*, with the maximum being evaluated over both types of departure; 2) For weighted Least Squares estimation, determine both weights and a design to minimize the maximum *IMSPE*. 3) Choose weights and design points to minimize the maximum *IMSPE*, subject to a side condition of unbiasedness. Solutions to 1) and 2) are given for multiple linear regression with no interactions, a spherical design space and an annular extrapolation space. For 3) the solution is given in complete generality; as an example we consider polynomial regression. We remark that the optimal designs for moderately large symmetric and one-sided extrapolation regions are, for practical purposes, identical, but when the extrapolation regions are small, the optimal designs for one-sided extrapolation place appreciably more mass on the side of design region closer to the extrapolation region.

The designs given in Chapters 2 and 3 are absolutely continuous measures because the model-violation class used is so full. These designs have to be approximated before implementation. We discuss possible approximations in each paper. Other ways to implement these designs will be the subject of further research.

By considering a rather thin model-violation class, we construct robust optimal designs for extrapolation in the polynomial regression setup (8) in Chapter 4, entitled "Robust Extrapolation Designs for Biased Polynomial Models." The robust extrapolation designs are identical to the minimum variance extrapolation designs for linear and quadratic fitted models. For cubic or higher degree fitted models, the designs vary with the choice of $\phi(x)$ and the value of σ^2/n. These designs are discrete and implementable, but have shortcomings: they afford no opportunity to assess the fit of the model and have infinite maximum risk in (5).

The author recommends the designs given in the Chapters 2 and 3. Our attitude is similar to that of Wiens (1992), who states that "an approximation to a design which is robust against more realistic alternatives is preferable to an exact solution in a neighbourhood which is unrealistically sparse."

References

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

Crump, K. S. (1979). Dose Response Problems in Carcinogenesis. *Biometrics*, **35**, 157 – 167.

Dette, H. and Wong, W.K. (1996). Robust Optimal Extrapolation Designs. *Biometrika* 83, 3, 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.

Herzberg, A. and Cox, D. (1972). Some Optimal Designs for Interpolation and Extrapolation. *Biometrika,* 59, 551–561.

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

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

Huang, M. N. L. (1990). Optimal Extrapolation Designs for a Partly Linear Model. *Computational Statistics and Data Analysis*, **10**, 109 – 115.

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

Huber, P. J. (1981). Robust Statistics. Wiley, New York.

Kiefer and Wolfowitz, J. (1960). The Equivalence of Two Extremum Problems. *Canadian Journal of Mathematics*, **12**, 363 – 366.

Kiefer, J. and Wolfowitz, J. (1965). On a Theorem of Hoel and Levine on Extrapolation. *Annals of Mathematical Statistics*, 36, 1627–1655.

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. *Annal 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.

Liu S. X. and Wiens, D. P. (1997). Robust Designs for Approximately Polynomial Regression. *Journal of Statistical Planning and Inference*, **64**, 369 – 381.

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

Pesotchinsky, L. (1982). Optimal Robust Designs: Linear Regression in $\mathcal{R}^k$. *Annals of Statistics*, **10**, 511 – 525.

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

Sacks J. and Ylvisaker D. (1978). Linear Estimation for Approximately Linear Models. *Annals of Statistics*, **6**, 1122 – 1137.

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

Spruill, M. C. (1990). Good Designs for Polynomial Extrapolation. *Journal of Statistical Planning and Inference*, **26**, 149 – 159.

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

Wiens, D. P. (1990). Robust Minimax Designs for Multiple Linear Regression. *Linear Algebra and its Applications*, **127**, 327 – 340.

Wiens, D. P. (1991). Designs for Approximately Linear Regression: Two Optimality Properties of Uniform Designs. *Statistics & Probability Letters*, **12**, 217 – 221.

Wiens, D. P. (1992). Minimax Designs for Approximately Linear Regression. *Journal of Statistical Planning and Inference*, **31**, 355 – 371.

Wiens, D. P. (1993). Designs for Approximately Linear Regression: Maximizing the Minimum Coverage Probability of Confidence Ellipsoids. *Canadian Journal of Statistics*, **21**, 59 – 70.

Wiens, D. P. (1994). Robust Designs for Approximately Linear Regression: M-estimated Parameters. *Journal of Statistical Planning and Inference*, **40**, 135 – 160.

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

Chapter Two

Robust Optimal Designs for Extrapolation Outside a Hypersphere When Bias is Present[1]

Abstract

For regression problems where the response may be observed at points in a set S which is different from the set T on which the fitted values are of interest, we consider designs for an approximately linear model $E(Y|\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta} + f(\mathbf{x})$, $\mathbf{x} \in S$, where $f(\mathbf{x})$ is a non-linear disturbance restricted only by its $\mathcal{L}_2$ norm. Specifically, we exhibit solutions to the following problems: P1) Determine a design to minimize the maximum of the Integrated Mean Squared Prediction Error (*IMSPE*) of the fitted values over T, with the maximum being evaluated over the departures from the model. P2) After splitting the *IMSPE* into two parts: Integrated Prediction Variance (*IPV*) and Integrated Squared Prediction Bias (*ISPB*), we seek the design which minimizes the maximum *ISPB* subject to bounding *IPV*. P3) We choose a design which minimizes *IPV* subject to bounding the maximum *ISPB*. We find that the forms of the optimal designs are quite sensitive to the volume of the extrapolation space T.

[1] Submitted for publication

1 Introduction

Consider the regression model given by

$$Y(\mathbf{x}_i) = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\theta} + \varepsilon_i \qquad i = 1, \ldots, n, \tag{1}$$

where the regressors $\{z_j(\mathbf{x}_i)\}_{j=0}^{p}$ are linearly independent, $\mathbf{x}_i \in S \subset \mathcal{R}^q$, $\boldsymbol{\theta} \in \mathcal{R}^{p+1}$ and the ε_i are additive, uncorrelated errors with common variance σ^2. The extrapolation problem is to predict the value of $\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}$, $\mathbf{x} \in T \subset \mathcal{R}^q$ $(S \cap T = \emptyset)$ and in connection with this prediction the design problem is to choose the design points $\mathbf{x}_i$ from S in some optimal manner.

When the model (1) is exactly correct, the extrapolation design problem is considered in a series of papers: Kiefer and Wolfowitz (1964a, 1964b, 1965), Hoel and Levine (1964), Studden (1971) and Herzberg and Cox (1972). Box and Draper (1959) pointed out that it would become dangerous to design a regression experiment which assumes that (1) is exactly correct. Therefore, it is reasonable to introduce a bias term in (1). The corresponding model can be defined now by

$$Y(\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta} + f(\mathbf{x}) + \varepsilon, \tag{2}$$

where $f(\mathbf{x})$ is an unspecified contamination function from some class $\mathcal{F}$.

Some authors have constructed optimal extrapolation designs for versions of (2), in which they choose different kinds of contamination classes, the design spaces and the extrapolation spaces, the regressors and the loss functions used. Draper and

Herzberg (1973) handle the design problem for extrapolation with bias in which first-order regression is considered and the bias comes from the second-order terms, when S is a q-dimensional ball and T is the line segment from a point t to the closest point of S. They use the Least Squares technique to estimate the model parameters. But, the designs are limited to a certain subclass of designs over which the integrated mean squared error over T is minimized. Karson, Manson and Hader (1969) have treated the same problem in a different way. They propose choosing a class of linear estimators which minimize the integrated squared bias over T, then choosing the optimal design to minimize the resulting variance. With $S = [0, \infty]$ and $T = \{t\}$, $t < 0$, Huber (1975) suggests (for some constant $\tau > 0$)

$$\mathcal{F} = \{f : f \text{ is } (p+1) \text{ times differentiable and } |f^{(p+1)}(x)| < \tau, \; x \in [t, \infty]\},$$

Huber finds that by using *MSE* as loss function, the optimal extrapolation design sits on $(p+1)$ points: $0 = x_1 < \cdots < x_{p+1}$, which after the addition of another point x_{p+2} constitute the set of Chebyshev points of order $(p+1)$ in $[0, x_{p+2}]$. Huang and Studden (1988) give a counter example to Huber's proof, but confirm Huber's result. In their paper, the designs considered are limited to exactly $(p+1)$-point designs. As well, Spruill (1984) constructs the extrapolation design when $\mathcal{F}$ is a Sobolev space and Huang (1990) treats the partially linear model when the departure satisfies a Lipschitz condition. Dette and Wong (1996) construct extrapolation designs when there is uncertainty in the degree of polynomial, and they propose a new class of optimality criteria for extrapolation.

In this article we address the construction of extrapolation designs for model (2) when the contamination class $\mathcal{F}$ is an $\mathcal{L}_2$-neighbourhood, specified below. We choose the Integrated Mean Squared Prediction Error (*IMSPE*) over T of the fitted value $\hat{Y}(\mathbf{x})$ as loss function. The following problems are considered:

P*1)*. Determine a design to minimize the maximum value of *IMSPE* over T, with the maximum being evaluated over f.

P*2)*. After splitting the *IMSPE* into two parts: Integrated Prediction Variance (*IPV*) and Integrated Squared Prediction Bias (*ISPB*), we seek the design which minimizes the maximum of *ISPB*, subject to bounding *IPV*.

P*3)*. We construct designs which minimize the *IPV* subject to bounding the maximum, over f, value of $ISPB$.

The organization of this paper is as follows. In Section 2, we define precisely the regression model to be considered, and give the maximum of *IMSPE* over T. Solutions to P1) - P3) are given in Section 3. Two simulation studies are given in Section 4. All proofs are given in the Appendix.

2 PRELIMINARIES AND NOTATION

In general, we assume that S and T are two non-intersecting subsets of $\mathcal{R}^q$, and the i-th observation $y_i \in \mathcal{R}$ can be described by the model

$$Y_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\theta} + f(\mathbf{x}_i) + \varepsilon_i, \qquad i = 1, ..., n, \tag{3}$$

where $\mathbf{x}_i \in S$ with unit volume, $\boldsymbol{\theta} \in \mathcal{R}^{p+1}$, and $\mathbf{z}(\mathbf{x})$ is a $(p+1)$-vector of real valued functions on $S \cup T$. The response error $f(\mathbf{x})$ belongs to the class (for some known positive constants η_S and η_T):

$$\mathcal{F} = \{f : \int_S f^2(\mathbf{x})d\mathbf{x} \leq \eta_S^2; \int_T f^2(\mathbf{x})d\mathbf{x} \leq \eta_T^2; \int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})d\mathbf{x} = 0\}, \tag{4}$$

in which the last condition is without loss of generality, and ensures the identifiability of $\boldsymbol{\theta}$. The requirement that T and S be disjoint needs not exclude the application of our results to interpolation problem, i.e. the case $T \subset S$, as long as design points are not to be chosen from within T. One can then replace S by $S \backslash T$. The ε_i are additive, uncorrelated errors with common variance σ^2. As to the predicted value, we choose the Least Squares estimator.

Let ξ be the design measure, i.e. the empirical distribution of $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. Define matrices $\mathbf{A}_T$, $\mathbf{A}_S$, $\mathbf{B}_\xi$ and vectors $\mathbf{b}_{f,\xi}$, $\mathbf{b}_{f,T}$ as follows:

$$\mathbf{A}_T = \int_T \mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})d\mathbf{x}, \qquad \mathbf{A}_S = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})d\mathbf{x},$$

$$\mathbf{B}_\xi = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})d\xi(\mathbf{x}), \qquad \mathbf{b}_{f,\xi} = \int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})d\xi(\mathbf{x}),$$

$$\mathbf{b}_{f,T} = \int_T \mathbf{z}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}.$$

Then the Least Squares estimator of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = \mathbf{B}_\xi^{-1} \int_S \mathbf{z}(\mathbf{x}) y(\mathbf{x}) d\xi(\mathbf{x})$, with bias vector and covariance matrix

$$E[\hat{\boldsymbol{\theta}}] - \boldsymbol{\theta} = \mathbf{B}_\xi^{-1} \mathbf{b}_{f,\xi}, \qquad COV[\hat{\boldsymbol{\theta}}] = \frac{\sigma^2}{n} \mathbf{B}_\xi^{-1}.$$

We predict $E[Y|\mathbf{x}]$ for $\mathbf{x} \in T$ by $\hat{Y}(\mathbf{x}) = \mathbf{z}^T(\mathbf{x}) \hat{\boldsymbol{\theta}}$, and consider the resulting *IMSPE*

$$IMSPE(f, \xi) = \int_T E[\hat{Y}(\mathbf{x}) - E(Y|\mathbf{x})]^2 d\mathbf{x}.$$

The *IMSPE* can be split into two parts:

$$IMSPE(f, \xi) = IPV(\xi) + ISPB(f, \xi),$$

where

$$\begin{aligned} IPV(\xi) &= \int_T E[\ \hat{Y}(\mathbf{x}) - E(\hat{Y}(\mathbf{x}))\ ]^2 d\mathbf{x} = \frac{\sigma^2}{n} tr(\mathbf{A}_T \mathbf{B}_\xi^{-1}), \\ ISPB(f, \xi) &= \int_T (\ E[\hat{Y}(\mathbf{x})] - E[Y|\mathbf{x}]\ )^2 d\mathbf{x} \\ &= \mathbf{b}_{f,\xi}^T \mathbf{B}_\xi^{-1} \mathbf{A}_T \mathbf{B}_\xi^{-1} \mathbf{b}_{f,\xi} - 2\mathbf{b}_{f,T} \mathbf{B}_\xi^{-1} \mathbf{b}_{f,\xi} + \int_T f^2(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

We find that $IMSPE(f, \xi)$ depends on f only through $ISPB(f, \xi)$, that is, f does not affect the value of $IPV(\xi)$. So evaluating the maximum value of $IMSPE(f, \xi)$ is equivalent to evaluating the maximum value of $ISPB(f, \xi)$.

We adopt the methods of approximate design theory, and allow ξ to be any distribution function on S. Then it can be shown by modifying Lemma 1 of Wiens

(1992) that the optimal extrapolation designs are absolutely continuous, and so must be approximated to be made implementable.

Let $\nu := \frac{\sigma^2}{n\eta_S^2}$ and $r_{T,S} := \eta_T/\eta_S$. Then ν can be viewed as the measurement of the relative importance of variance and bias, and $r_{T,S}$ gives the relative amounts of model response uncertainty in the extrapolation and design spaces. For fixed ξ, with density function $m(\mathbf{x})$, the following theorem gives the maximum of $IMSPE(f,\xi)$, over f.

Theorem 1 *Define* $\mathbf{K} = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})m^2(\mathbf{x})d\mathbf{x}$, $\mathbf{G} = \mathbf{K} - \mathbf{B}_\xi\mathbf{A}_S^{-1}\mathbf{B}_\xi$, *and* $\mathbf{H} = \mathbf{B}_\xi\mathbf{A}_T^{-1}\mathbf{B}_\xi$. *Let* λ_m *be the largest solution to* $|\mathbf{G} - \lambda\mathbf{H}| = 0$ *and let the vector* $\mathbf{a}_0$ *be defined by*

$$(\mathbf{G}\mathbf{H}^{-1}\mathbf{G} - \lambda_m\mathbf{G})\mathbf{a}_0 = \mathbf{0} \textit{ and } \mathbf{a}_0^T\mathbf{G}\mathbf{a}_0 = 1.$$

Then we have

$$\begin{aligned} max_{f\in\mathcal{F}} ISPB(f,\xi) &= \eta_S^2(\sqrt{\lambda_m} + r_{T,S})^2, \\ max_{f\in\mathcal{F}} IMSPE(f,\xi) &= \eta_S^2[\nu tr(\mathbf{A}_T\mathbf{B}_\xi^{-1}) + (\sqrt{\lambda_m} + r_{T,S})^2], \end{aligned}$$

which are attained at the following least favourable function

$$f_m(\mathbf{x}) = \begin{cases} \eta_S\mathbf{z}^T(\mathbf{x})[m(\mathbf{x})\mathbf{I} - \mathbf{A}_S^{-1}\mathbf{B}_\xi]\mathbf{a}_0 & \textit{if } \mathbf{x} \in S, \\ -\eta_S r_{T,S}\mathbf{z}^T(\mathbf{x})\mathbf{B}_\xi^{-1}\mathbf{G}\mathbf{a}_0/\sqrt{\lambda_m} & \textit{if } \mathbf{x} \in T. \end{cases}$$

Based on Theorem 1, our problems can be rewritten as:

P1). Find a density $m_1(\mathbf{x})$ to minimize $\nu tr(\mathbf{A}_T\mathbf{B}_\xi^{-1}) + (\sqrt{\lambda_m} + r_{T,S})^2$.

P2). Seek a density $m_2(\mathbf{x})$ to minimize $\eta_S^2\ (\sqrt{\lambda_m}+r_{T,S})^2$, subject to $\eta_S^2 \nu tr(\mathbf{A}_T \mathbf{B}_\xi^{-1}) \leq d$, for some constant d.

P3). Choose a density $m_3(\mathbf{x})$ which minimizes $\eta_S^2 \nu tr(\mathbf{A}_T \mathbf{B}_\xi^{-1})$, subject to $\eta_S^2\ (\sqrt{\lambda_m}+ r_{T,S})^2 \leq k^2$, for some constant k^2.

3 OPTIMAL EXTRAPOLATION DESIGNS

In this section we construct optimal extrapolation designs for multiple linear regression without interaction: $\mathbf{z}^T(\mathbf{x}) = (1,\ \mathbf{x}^T)$, with S being a p-dimensional sphere of unit volume:

$$S = \{\mathbf{x}|\ \|\mathbf{x}\| \leq \beta_S := \frac{[\Gamma(\frac{p}{2}+1)]^{\frac{1}{p}}}{\sqrt{\pi}}\},$$

and T chosen as

$$T = \{\mathbf{x}|\ \beta_S \leq \|\mathbf{x}\| \leq \beta_T\}.$$

Since S and T are symmetric about the origin, we restrict to densities $m(\mathbf{x})$ with identical, symmetric marginals and seek the minimax extrapolation design density within this class. See Wiens (1992) for a discussion of this point.

Define a volume measure δ_p and three types of second moments γ, γ_S, γ_T by

$$\begin{aligned} \delta_p &= \int_T d\mathbf{x} = \frac{\beta_T^p}{\beta_S^p} - 1, & \gamma = \int_S x_1^2 m(\mathbf{x}) d\mathbf{x}, \\ \gamma_S &= \int_S x_1^2 d\mathbf{x} = \frac{\beta_S^2}{(p+2)}\ , & \gamma_T = \int_T x_1^2 d\mathbf{x} = \frac{\beta_T^{p+2} - \beta_S^{p+2}}{(p+2)\beta_S^p}, \end{aligned}$$

then

$$\mathbf{A}_S = 1 \oplus \gamma_S \mathbf{I}_p, \qquad \mathbf{A}_T = \delta_p \oplus \gamma_T \mathbf{I}_p, \qquad \mathbf{B}_\xi = 1 \oplus \gamma \mathbf{I}_p,$$
$$\mathbf{K} = (\int_S m^2(\mathbf{x})d\mathbf{x}) \oplus (\int_S x_1^2 m^2(\mathbf{x})d\mathbf{x})\mathbf{I}_p, \qquad \frac{\gamma_T}{\gamma_S} = \delta_{p+2}.$$

The maximum eigenvalue λ_m in Theorem 1 is found to be $\lambda_m = max(\lambda_m^{(0)}, \lambda_m^{(1)})$, with

$$\lambda_m^{(0)} = \delta_p(\int_S m^2(\mathbf{x})d\mathbf{x} - 1), \qquad \lambda_m^{(1)} = \delta_{p+2}(\frac{\gamma_S}{\gamma^2}\int_S x_1^2 m^2(\mathbf{x})d\mathbf{x} - 1), \tag{5}$$

and then

$$max_{f\in\mathcal{F}} IMSPE(f,\xi) = \eta_S^2[\nu(\delta_p + p\frac{\gamma_T}{\gamma}) + (\sqrt{\lambda_m} + r_{T,S})^2]. \tag{6}$$

We exhibit the solutions to P1) - P3) in the following subsections.

3.1 Minimax extrapolation design: Solution to P1)

Theorem 1 shows that the maximum of $IMSPE(f,\xi)$, over f, is attained at f_m, so the problem P1) is to find an optimal extrapolation design density which minimizes $IMSPE(f_m,\xi)$. We first find m_* minimizing (6) with $\lambda_m = \lambda_m^{(i)}$ ($i = 0$ or 1). Then if m_* satisfies $\lambda_{m_*}^{(i)} \geq \lambda_{m_*}^{(1-i)}$, it is the required minimax density.

Let $\rho = \frac{\gamma}{\gamma_S}$. It can be shown that: $\rho = \frac{1}{p\gamma_S}\int_S ||\mathbf{x}||^2 m(\mathbf{x})d\mathbf{x} \leq \frac{1}{p\gamma_S}\int_S \beta_S^2 m(\mathbf{x})d\mathbf{x} = (p+2)/p$ and $\rho = 1$ when $m(\mathbf{x}) \equiv 1$, the continuous uniform design. If ξ is a design such that $\rho < 1$, then

$$max_{f\in\mathcal{F}} IMSPE(f,\xi) > \eta_S^2[\nu(\delta_p + p\frac{\gamma_T}{\gamma_S}) + r_{T,S}^2] = max_{f\in\mathcal{F}} IMSPE(f,\xi_U),$$

where ξ_U represents the continuous uniform design, because $\lambda_{\xi_U} = 0$. So we only need to consider $\rho \in [1,\ (p+2)/p]$.

Before stating the theorem, define functions for $b \in [0,\ 1]$:

$$k_p(b) = \int_{\sqrt{b}}^{1} u^{p-1}(u^2 - b)du; \qquad g_p(b) = \frac{1}{(p+2)k_p(b)} - \frac{b}{p+2}\frac{k_{p-2}(b)}{k_p^2(b)} - 1;$$

$$J_p(b) = \frac{p(1-b^2) - 2b^2(1-b^p)}{p+2}; \qquad h_p(b) = \frac{p}{p+2}\frac{J_{p+2}(b)}{J_p^2(b)} - \frac{b^2}{J_p(b)} - 1.$$

Theorem 2 *(i). Let b_1 be the solution to*

$$\nu p \delta_{p+2} \frac{\partial}{\partial b}\left(\frac{p}{p+2}\frac{k_{p-2}(b)}{k_p(b)}\right) + \frac{\partial}{\partial b}\left(\sqrt{\delta_{p+2} g_p(b)} + r_{T,S}\right)^2 = 0. \tag{7}$$

Then if b_1 satisfies

$$\delta_{p+2} g_p(b_1) \geq \delta_p g_{p-2}(b_1), \tag{8}$$

we have the minimax extrapolation design density

$$m_1(\mathbf{x}) = a_1[1 - b_1(\frac{\beta_s}{\|\mathbf{x}\|})^2]^+,$$

with a_1 determined by

$$a_1 = \frac{1}{p k_{p-2}(b_1)}.$$

It can be shown that $\rho = \frac{p+2}{p}\frac{k_p(b_1)}{k_{p-2}(b_1)} \in [1,\ \frac{p+2}{p}]$ and that (8) holds for $\nu \in [\nu_4(\delta_p), \infty]$, for some $\nu_4(\delta_p) > 0$.

(ii). a) Define $\rho_0 = \sqrt{\frac{\delta_{p+2}}{\delta_p}\frac{(p+2)^3}{p(p+4)(p+6)}}$ and $\kappa = \sqrt{\delta_p \frac{p(p+4)}{4}}$. Let ρ_1 be the root of the equation

$$p\nu\delta_{p+2} = 2\kappa\rho_1^2[\kappa(\rho_1 - 1) + r_{T,S}]. \tag{9}$$

Then if

$$\rho_0 \leq \rho_1 \leq 1 + \frac{4}{p(p+4)}, \tag{10}$$

we have the optimal extrapolation design density:

$$m_1(\mathbf{x}) = 1 + \frac{p+4}{4}(\rho_1 - 1)(\frac{\|\mathbf{x}\|^2}{\gamma_S} - p).$$

It can be shown that when δ_p is small, (10) is true for $\nu \in [\nu_1(\delta_p),\ \nu_2]$, where ν_2 is calculated by (9) when $\rho_1 = 1 + \frac{4}{p(p+4)}$.

b) Let c_1 be the root of the equation

$$\nu p \delta_{p+2} \frac{\partial}{\partial c}(\frac{J_p(c)}{J_{p+2}(c)}) + \frac{\partial}{\partial c}(\sqrt{\delta_p h_p(c)} + r_{T,S})^2 = 0. \tag{11}$$

Then if c_1 satisfies

$$\delta h_p(c_1) \geq \delta_{p+2} h_{p+2}(c_1), \tag{12}$$

we have the optimal extrapolation design density

$$m_1(\mathbf{x}) = \frac{1}{J_p(c_1)}[(\frac{\|\mathbf{x}\|}{\beta_S})^2 - c_1^2]^+.$$

It can be shown that $\rho = \frac{J_{p+2}(c_1)}{J_p(c_1)} \in [1 + \frac{4}{p(p+4)},\ \frac{p+2}{p}]$, and that when δ_p is small, (12) holds for $c_1 \in [0,\ c_1(\delta_p)]$, i.e. for $\nu \in [\nu_2,\ \nu_3(\delta_p)]$, where ν_2 is calculated by (11) when $c_1 = 0$.

Remarks

1. The assumptions on $m(\mathbf{x})$, together with a convexity argument as in Fang and Wiens (1999), imply that the design density depends on $\mathbf{x}$ only through $\|\mathbf{x}\|$. Conditions (8), (10), (12) ensure that λ_{m_1} is indeed the larger of the two eigenvalues in (5).

2. The minimax density depends on the unknown parameters through ν, δ_p and the ratio $r_{T,S}$. Although we do not have the solution for some intervals of ν, the numerical results in Table 1 show that $\nu_4(\delta) \rightarrow 0$ as $\delta_p \rightarrow \infty$ and the limit is approached very fast. So when δ_p becomes larger, our solution in Theorem 2 (i) is optimal for almost all ν.

3. As to the implementation of these designs, we may use the fact that under the density $m(\mathbf{x})$, $\mathbf{x}/\|\mathbf{x}\|$ and $U := \|\mathbf{x}\|$ are independently distributed, with $\mathbf{x}/\|\mathbf{x}\|$ having a uniform distribution over the surface of the unit sphere. Hence a possible implementation can be as follows. Let $H(u)$ be the cumulative distribution function of U, i.e. $H(u) = \int_0^u (pu^{p-1}/\beta_S^p) m_1(u) du$. We obtain $[\sqrt{n}]$ values:

$u_i = H^{-1}((i-1)/([\sqrt{n}]-1))$, $i = 1, ..., [\sqrt{n}]$. Then we have $[\sqrt{n}]$ design points

$$\mathbf{x}_{ij} = u_i \cdot \begin{pmatrix} sin(\theta_{1i} + \frac{\pi(j-1)}{\sqrt{n}}) \\ cos(\theta_{1i} + \frac{\pi(j-1)}{\sqrt{n}}) \cdot sin(\theta_{2i} + \frac{\pi(j-1)}{\sqrt{n}}) \\ cos(\theta_{1i} + \frac{\pi(j-1)}{\sqrt{n}}) \cdot cos(\theta_{2i} + \frac{\pi(j-1)}{\sqrt{n}}) \cdot sin(\theta_{3i} + \frac{\pi(j-1)}{\sqrt{n}}) \\ \cdots\cdots\cdots\cdots \\ cos(\theta_{1i} + \frac{\pi(j-1)}{\sqrt{n}}) \cdots cos(\theta_{(p-2)i} + \frac{\pi(j-1)}{\sqrt{n}}) \cdot sin(\theta_{(p-1)i} + \frac{2\pi(j-1)}{\sqrt{n}}) \\ cos(\theta_{1i} + \frac{\pi(j-1)}{\sqrt{n}}) \cdots cos(\theta_{(p-2)i} + \frac{\pi(j-1)}{\sqrt{n}}) \cdot cos(\theta_{(p-1)i} + \frac{2\pi(j-1)}{\sqrt{n}}) \end{pmatrix},$$

where $j = 1, ..., [\sqrt{n}]$, equally spaced over $||\mathbf{x}|| = u_i$, for each i. And we choose $(n - [\sqrt{n}]^2)$ points at the origin. The angles θ_{ri} $(r = 1, ...p-2)$ are uniformly distributed over $(-\frac{\pi}{2}, -\frac{\pi}{2} + \frac{\pi}{\sqrt{n}})$ and $\theta_{(p-1)i}$ is uniformly distributed over $(-\pi, -\pi + \frac{2\pi}{\sqrt{n}})$. All the angles θ_{ri} $(r = 1, ...p-2)$ and $\theta_{(p-1)i}$ are independent. The rationale of this implementation is that the empirical distribution of the chosen design points tends weakly to the distribution of the optimal design.

Example 1. For the straightline regression (SLR) situation, i.e. $p = 1$, we have exhibited some numerical values of the parameters in Table 2. We assume that $r_{T,S} = 1$. Based on these values, we have the following observations about the optimal $m_1(x)$:

(1). As $\nu \longrightarrow \infty$ (i.e. $\eta_S = 0$, variance dominant), $m_1(x)$ corresponds to the design with all mass at $|x| = \pm\frac{1}{2}$, which coincides with the result of Hoel and Levine

(1964). But $m_1 \equiv 1$ becomes optimal as $\delta_p \to \infty$ and $\nu \to 0$ (bias dominant case). See Figure 1(a).

(2). In practice, we should assume that $\eta_S = O(n^{-\frac{1}{2}})$ in order that the errors due to variance and to bias remain of the same order of magnitude, that is, ν is relatively small but should be bounded away from zero. This is the so-called "typical situation", which is to be anticipated as one in which the experimenter would probably be prepared to tolerate or accept biases that are of the same order roughly as variances. In this situation, we use $m_1(x) = (ax^2 + b)^+$ if δ_p is small while we adopt $m_1(x) = a[1 - b(\frac{\beta_\xi}{x})^2]^+$ if δ_p is large. See Figure 1(b).

(3). Table 2 implies that for fixed ν, the minimax design density puts more mass on an area closer to T when δ_p becomes large. So the minimax design moves the design points closer to T when T is large.

In Figure 2, we compare the behaviour of the minimax extrapolation design with that of other typical designs: the classical extrapolation design, which has design points $< -0.5, 0.5 >$, and the continuous uniform design. When the fitted model $E[Y|\mathbf{x}] = \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}$ is exactly correct (no contamination), the loss is

$$IMSPE(f = 0, \xi) = IPV(\xi) = \frac{\sigma^2}{n} tr(A_T B_\xi^{-1}) = \frac{\sigma^2}{n}(\delta_p + \frac{(1 + \delta_p)^{1+2/p} - 1}{12\gamma}),$$

where γ is the second moment of ξ. Figure 2(a) gives the plot of $(\frac{\sigma^2}{n})^{-1} IPV(\xi)$ vs δ_p for three designs. When the model is only approximately linear, the bias term is

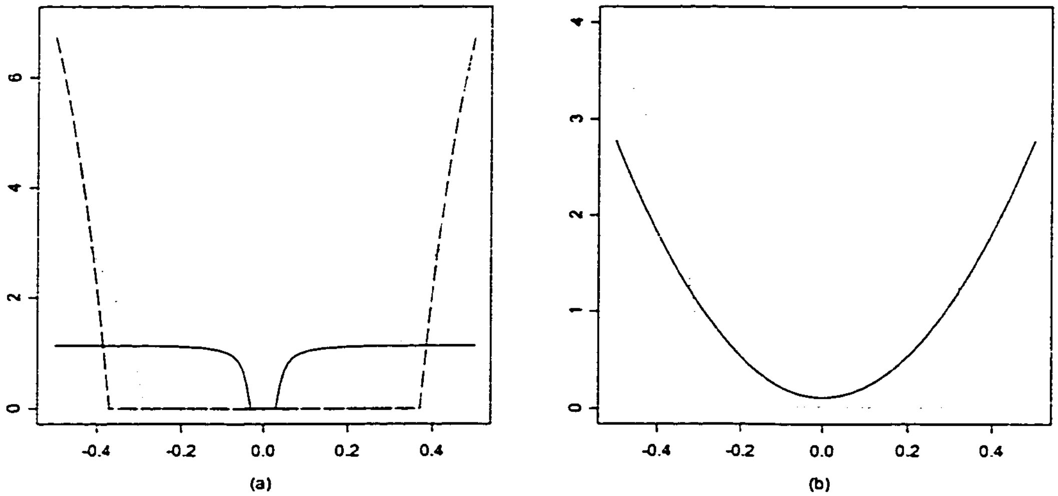
Table 1. Values of $\nu_k(\delta_p)$ when p=1 (values in parentheses) and p=2: Case (i) holds for $\nu > \nu_4(\delta_p)$; Case (ii) $a)$ holds for $\nu_1(\delta_p) \leq \nu \leq \nu_2(\delta_p)$; Case (ii) $b)$ holds for $\nu_2(\delta_p) \leq \nu \leq \nu_3(\delta_p)$.

δ_p	$\nu_1(\delta_p)$	$\nu_2(\delta_p)$	$\nu_3(\delta_p)$	$\nu_4(\delta_p)$
0.0625	4.96(8.0)	6.83(11.11)	477.7(449.2)	500.2(493.1)
0.125	3.65(6.07)	4.94(7.96)	121.5(107.2)	132.9(129)
0.25	2.76(4.93)	3.53(5.5)	31.15(23.95)	37.08(34.8)
0.5				11.16(9.76)
1				3.7(2.82)
5				0.4(0.15)
10				0.16(0.04)

Table 2. The constants for $m_1(x)$: Letters in parentheses indicate which case holds.

δ_1	ν	a_1	b_1	c_1	ρ_1	δ_1	ν	a_1	b_1
0.0625	8.503(*ii. a*))				1.67	1	5.268(i)	4	0.25
	20.5(*ii. b*))			0.455			20.5(i)	11.657	0.5
	521.84(i)	143.5	0.84				∞(i)	∞	1
	∞(i)	∞	1						
0.125	7.09(*ii. a*))				1.72	5	0.578(i)	1.754	0.06
	20.5(*ii. b*))			0.551			5.268(i)	5.28	0.319
	521.84(i)	173.7	0.854				20.5(i)	14.52	0.544
	∞(i)	∞	1				∞(i)	∞	1
0.25	5.257(*ii. a*))				1.71	10	0.047(i)	1.14	0.004
	20.5(*ii. b*))			0.625			5.268(i)	5.5	0.329
	521.84(i)	201.3	0.864				20.5(i)	14.98	0.55
	∞(i)	∞	1				∞(i)	∞	1

Figure 1: Minimax extrapolation design densities for SLR. (a) Case (i). $\delta_1 = 10$: $\nu = 0.047$ (solid line), $\nu = 5.268$ (dotted line), $\nu = 20.5$ (broken line); (b) $\nu = 5.257$: $\delta_1 = 0.25$ (solid line, Case (ii) a)), $\delta_1 = 5$ (dotted line, Case (i)).



included in the loss. We compare the *IMSPE* instead of the *IPV*. While the classical extrapolation design has infinite $max_f IMSPE(f, \xi)$ we plot $\eta_S^{-2} max_f IMSPE(f, \xi)$ for the minimax extrapolation design and the continuous uniform design in Figure 2(b). We have chosen $p = 1$, $\nu = 3$ and $r_{T,S} = 1$. We find that the use of the minimax extrapolation design results in a significant reduction of loss in the presence of contamination. When there is no contamination in the model, the performance of the minimax extrapolation design is close to that of the classical extrapolation design.

3.2 Bounded variance design: Solution to P2)

For some given bound d, P2) may be phrased as looking for

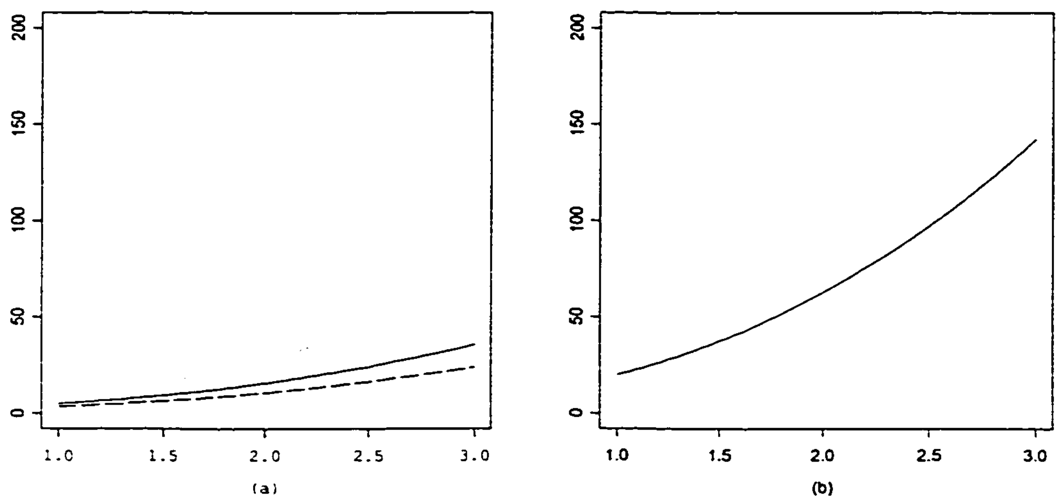
$$\begin{aligned} m_2(\mathbf{x}) &= argmin\{\eta_S^2(\sqrt{\lambda_m} + r_{T,S})^2 : \int_S m(\mathbf{x})d\mathbf{x} = 1,\ \eta_S^2\nu(\delta_p + p\frac{\gamma_T}{\gamma}) \leq d\} \\ &= argmin\{\lambda_m : \int_S m(\mathbf{x})d\mathbf{x} = 1,\ \int_S \|\mathbf{x}\|^2 m(\mathbf{x})d\mathbf{x} \geq d_1\}, \end{aligned}$$

where $d_1 = p^2\gamma_T/(d/\eta_S^2\nu - \delta_p) > 0$. The technique to solve this problem is the same as that in Section 3.1.

Theorem 3 *If $d_1 \in [0, \frac{p}{p+2}\beta_S^2]$, then the continuous uniform design is the optimal bounded variance design for extrapolation, that is, $m_2(\mathbf{x}) \equiv 1$ is the solution to P2).*

Combining this theorem and the fact that $\int_S \|\mathbf{x}\|^2 m(\mathbf{x})d\mathbf{x} \leq max_{\mathbf{x}\in S}||\mathbf{x}||^2 = \beta_S^2$, we only need to consider the case when $d_1 \in (\frac{p}{p+2}\beta_S^2, \beta_S^2]$. With $\lambda_m = \lambda_m^{(1)}$, we have

Figure 2: (a) Integrated Prediction Variance vs δ_p when there is no contamination: Minimax (solid line), classical design (broken line) and uniform design (dotted line). (b) $max_f IMSPE$ vs δ_p when there exists contamination in model: minimax (solid line) and uniform design (dotted line).



the following lemma.

Lemma 4 *The function*

$$argmin\{\delta_{p+2}(\frac{\gamma_S}{\gamma^2}\int_S x_1^2 m^2(\mathbf{x})d\mathbf{x} - 1) : \int_S m(\mathbf{x})d\mathbf{x} = 1, \int_S \|\mathbf{x}\|^2 m(\mathbf{x})d\mathbf{x} \geq d_1\}$$

is equal to

$$argmin\{\int_S \|\mathbf{x}\|^2 m^2(\mathbf{x})d\mathbf{x} : \int_S m(\mathbf{x})d\mathbf{x} = 1, \int_S \|\mathbf{x}\|^2 m(\mathbf{x})d\mathbf{x} = d_1\}.$$

This lemma changes the inequality in P2) into an equality, so it greatly simplifies the minimization problem.

Theorem 5 *For any fixed δ_p, if $d_1 \in [d_1(\delta_p), \beta_S^2]$, the solution to P2) has the form*

$$m_2(\mathbf{x}) = a[1 - b\frac{\beta_s^2}{\|\mathbf{x}\|^2}]^+,$$

where a, b are determined by

$$\begin{aligned} a &= \frac{1}{pk_{p-2}(b)}, \\ d_1 &= apk_p(b)\beta_S^2 . \end{aligned}$$

The minimax IMSPE is

$$\eta_S^{-2} max_{f\in\mathcal{F}} IMSPE(f, \xi_2) = \nu(\delta_p + p^2\frac{\gamma_T}{d_1}) + (\sqrt{\delta_{p+2}g_p(b)} + r_{T,S})^2. \quad (13)$$

The bound $d_1(\delta_p)$ depends on δ_p and is obtained by

$$\begin{aligned} d_1(\delta_p) &= \frac{k_p(b)}{k_{p-2}(b)}\beta_S^2, \\ \delta_{p+2}g_p(b) &= \delta_p g_{p-2}(b). \end{aligned}$$

Table 3. Some typical values of $d_1(\delta_p)$

p	δ_p	$d_1(\delta_p)$	p	δ_p	$d_1(\delta_p)$	p	δ_p	$d_1(\delta_p)$
1	1	0.1479	2	1	0.2394	3	1	0.3164
	10	0.1008		10	0.1738		10	0.2487
	∞	0.0833		∞	0.1592		∞	0.2309

It can be shown that when $\delta_p \to \infty$, $d_1(\delta_p) \to \frac{p}{p+2}\beta_S^2$, the lower bound of d_1.

Remarks

1. We only present the solution for large d_1 when δ_p is given. But the limit in the theorem is approached quite rapidly (see Table 3), so our solution will hold for almost all d_1 when δ_p becomes large.

2. Table 4 gives the efficiencies of the bounded variance design ξ_2 relative to other designs: the continuous uniform design (ξ_U) and the design (ξ_B) which places all mass on the boundary of design space. When the model does not include any contamination term, the *OLS* estimate is unbiased. This efficiency is defined by

$$rel(\xi) = \frac{IPV(f=0,\xi)}{IPV(f=0,\xi_2)} = \frac{\delta_p + p\frac{\gamma_{\mathcal{I}}}{\gamma}}{\delta_p + p^2\frac{\gamma_{\mathcal{I}}}{d_1}},$$

where γ represents the second moments of ξ. When the true response is only

Table 4. Relative efficiencies of ξ_2 relative to the uniform design (ξ_U) and the discrete design (ξ_B).

p	d_1	δ_p	$re1(\xi_U)$	$re1(\xi_B)$	$re2(\xi_U)$	$re2(\xi_B)$
1	0.15	1	1.64	0.68	1.35	∞
		10	1.79	0.61	1.63	∞
2	0.25	1	1.45	0.83	1.16	∞
		10	1.53	0.8	1.36	∞
3	0.35	1	1.42	0.93	1.02	∞
		10	1.47	0.92	1.18	∞

partially linear, in contrast, this efficiency is defined by

$$re2(\xi) = \frac{max_{f\in\mathcal{F}} IMSPE(f,\xi)}{max_{f\in\mathcal{F}} IMSPE(f,\xi_2)},$$

where $max_{f\in\mathcal{F}} IMSPE(f,\xi_2)$ is calculated by (13) and $\eta_S^{-2} max_{f\in\mathcal{F}} IMSPE(f,\xi_U)$ $= \nu(\delta_p + p\delta_{p+2}) + r_{T,S}^2$. In Table 4, we assume that $\nu = 3$ and $r_{T,S} = 1$.

3.3 Bounded bias design: Solution to P3)

Problem P3) is to seek the design density

$$m_3(\mathbf{x}) = argmax\{\frac{1}{p}\int_S \|\mathbf{x}\|^2 m(\mathbf{x})d\mathbf{x} : \int_S m(\mathbf{x})d\mathbf{x} = 1,\ \lambda_m \le k_1\},$$

with a given value $k_1 = (\frac{k-\eta_T}{\eta_S})^2 \ge 0$.

To solve this problem, we first find

$$m_{31}(\mathbf{x}) = argmax\{\frac{1}{p}\int_S \|\mathbf{x}\|^2 m(\mathbf{x})d\mathbf{x} : \int_S m(\mathbf{x})d\mathbf{x} = 1,\ \delta_{p+2}(\frac{\gamma_S}{p\gamma^2}\int_S \|\mathbf{x}\|^2 m^2(\mathbf{x})d\mathbf{x} - 1) \leq k_1\},$$

and

$$m_{32}(\mathbf{x}) = argmax\{\frac{1}{p}\int_S \|\mathbf{x}\|^2 m(\mathbf{x})d\mathbf{x} : \int_S m(\mathbf{x})d\mathbf{x} = 1,\ \delta_p(\int_S m^2(\mathbf{x})d\mathbf{x} - 1) \leq k_1\}.$$

Then we claim that $m_3(\mathbf{x}) = m_{31}(\mathbf{x})$ if

$$\delta_p(\int_S m_{31}^2(\mathbf{x})d\mathbf{x} - 1) \leq k_1, \tag{14}$$

or $m_3(\mathbf{x}) = m_{32}(\mathbf{x})$ if

$$\delta_{p+2}(\frac{\gamma_S}{p\gamma^2}\int_S \|\mathbf{x}\|^2 m^2(\mathbf{x})d\mathbf{x} - 1) \leq k_1. \tag{15}$$

To establish the claim, note that if (14) holds, then $\lambda_{m_{31}} \leq k_1$. And for any density $m_*(\mathbf{x})$ satisfying $\lambda_{m_*} \leq k_1$, the definition of $m_{31}(\mathbf{x})$ shows that $\int_S \|\mathbf{x}\|^2 m_*(\mathbf{x})d\mathbf{x} \leq \int_S \|\mathbf{x}\|^2 m_{31}(\mathbf{x})d\mathbf{x}$. This implies that $m_3(\mathbf{x}) = m_{31}(\mathbf{x})$. The case when (15) holds is similar.

Now, to find $m_{31}(\mathbf{x})$, let $k_2 := \frac{p(k_1+\delta_{p+2})}{\delta_{p+2}\gamma_S} = \frac{p(k_1+\delta_{p+2})}{\gamma_T}$ $(\geq \frac{p\delta_{p+2}}{\gamma_T} = \frac{p}{\gamma_S} = \frac{p(p+2)}{\beta_S^2})$ and define

$$C_\gamma = \{m(\mathbf{x}) : \int_S m(\mathbf{x})d\mathbf{x} = 1,\ \int_S \|\mathbf{x}\|^2 m(\mathbf{x})d\mathbf{x} = p\gamma,\ \int_S \|\mathbf{x}\|^2 m^2(\mathbf{x})d\mathbf{x} \leq k_2\gamma^2\}.$$

Lemma 6 *(i) If $C_\gamma \neq \emptyset$, then there exists a, b such that*

$$a[1 - b(\frac{\beta_S}{\|\mathbf{x}\|})^2]^+ \in C_\gamma.$$

(ii) For any fixed value k_2 $(\geq \frac{p(p+2)}{\beta_S})$, *the largest* γ *such that* $C_\gamma \neq \emptyset$ *is determined by*

$$\gamma = ak_p(b)\beta_S^2, \tag{16}$$

$$a = \frac{1}{pk_{p-2}(b)}, \tag{17}$$

$$k_2 = \frac{p(p+2)(g_p(b)+1)}{\beta_S^2}. \tag{18}$$

We note that given k_2*, we calculate the value of* b *by (18) and then the largest value of* γ *is given by (16) and (17):* $\gamma = (k_p(b)\beta_S^2)/(pk_{p-2}(b))$. *These* γ*,* a*, and* b *satisfy* $\int_S a[1-b(\frac{\beta_S}{\|\mathbf{x}\|})^2]^+ d\mathbf{x} = 1$ *and* $\int_S \|\mathbf{x}\|^2 a[1-b(\frac{\beta_S}{\|\mathbf{x}\|})^2]^+ d\mathbf{x} = p\gamma$.

For any k_2, let γ, a, and b be the values given by (16), (17) and (18). Then for any density $m_*(\mathbf{x})$ with second moment $\gamma_* = \frac{1}{p}\int_S \|\mathbf{x}\|^2 m_*(\mathbf{x})d\mathbf{x}$ and $\int_S \|\mathbf{x}\|^2 m_*^2(\mathbf{x})d\mathbf{x} \leq k_2\gamma_*^2$, we have $\gamma \geq \gamma_*$ by Lemma 6 (*ii*). This implies that

$$\int_S \|\mathbf{x}\|^2 a[1-b(\frac{\beta_S}{\|\mathbf{x}\|})^2]^+ d\mathbf{x} \geq \int_S \|\mathbf{x}\|^2 m_*(\mathbf{x})d\mathbf{x}.$$

Hence, we conclude that

$$m_{31}(\mathbf{x}) = a[1-b(\frac{\beta_S}{\|\mathbf{x}\|})^2]^+, \tag{19}$$

with a, b given by (17), (18) for any fixed $k_2(\geq \frac{p(p+2)}{\beta_S})$, i.e. fixed $k_1(\geq 0)$.

Theorem 7 *Suppose that* $k_2 \geq k_2(\delta_p)$, *determined by the equations*

$$k_2(\delta_p) = \frac{p(p+2)(g_p(b)+1)}{\beta_S^2},$$

$$\delta_p g_{p-2}(b) = \delta_{p+2} g_p(b),$$

Table 5. The values of $k_1(\delta)$ and $k_2(\delta)$.

δ	p	$k_1(\delta)$	$k_2(\delta)$	p	$k_1(\delta)$	$k_2(\delta)$	p	$k_1(\delta)$	$k_2(\delta)$
1	1	0.842	13.444	2	0.819	31.99	3	0.809	53.48
10		0.776	12.007		0.612	25.26		0.507	39.35
∞		0	12		0	25.13		0	38.98

for any given δ_p. Then the solution to P3) is $m_3(\mathbf{x}) = m_{31}(\mathbf{x})$, defined by (17), (18) and (19). The minimax IMSPE is

$$\eta_S^{-2} max_{f \in \mathcal{F}} IMSPE(f, \xi_3) = \nu(\delta_p + \frac{p^2}{p+2}\delta_{p+2}\frac{k_{p-2}(b)}{k_p(b)}) + (\sqrt{\delta_{p+2} g_p(b)} + r_{T,S})^2.$$

Note that $k_2(\delta_p) \to \frac{p(p+2)}{\beta_S^2}$ as $\delta_p \to \infty$.

In Table 5, we give some typical values of $k_2(\delta_p)$ and related $k_1(\delta_p)$, for $p = 1, 2, 3$. We find that $k_2(\infty)$ (or $k_1(\infty)$) can be reached very quickly. The efficiencies of design ξ_3, which has density $m_3(\mathbf{x})$, relative to ξ_U and ξ_B are given in the Table 6.

4 Comparisons

Example 2 *(Linear regression).* Consider a regression model as at (3) and (4) with $\mathbf{z}(x) = (1, x)^T, -0.5 \le x \le 0.5$, normally distributed errors with $\sigma^2 = 1$, and sample size $n = 16$. As model contamination function $f(x)$, we take the quadratic Legendre polynomial, $f(x) \propto (12x^2 - 1)$,with normalization $\int_S f^2(x)dx = \eta_S^2 = 1/8$, so that

Table 6. Relative efficiencies of ξ_3 relative to uniform design (ξ_U) and discrete design (ξ_B).

p	k_2	δ	$re1(\xi_U)$	$re1(\xi_B)$	$re2(\xi_U)$	$re2(\xi_B)$
1	14	1	1.69	0.7	1.34	∞
		10	1.86	0.63	1.66	∞
2	34	1	1.45	0.82	1.11	∞
		10	1.52	0.79	1.3	∞
3	54	1	1.31	0.86	1.04	∞
		10	1.35	0.84	1.17	∞

$\nu = 0.5$. The extrapolation region has volume $\delta = 5$. We compare five designs: designs from P1), P2), P3), the continuous uniform design (U) and the classical design (C) which has eight observations at each of ± 0.5. For the four continuous designs four observations are made at each of $\xi^{-1}((i-1)/(\sqrt{n}-1)), i = 1, 2, 3, 4$.

Values of *IPSB*, *IPV*, and *IMSPE* are given in the Table 7. All three robust designs performed better than U and C by comparing the values of *IMSPE*. The classical design has highest bias although its variance is lowest. The three robust designs, especially P1), greatly reduce the bias. Meanwhile, the variances of the robust designs are much lower than that of the uniform design.

Is this model inadequacy likely to be detected? To answer this question, we fit a

Table 7. *ISPB, IPV* and *IMSPE* for the designs of

Example 2 and Example 3; Incorrect fitted response

	Linear response			*bivariate response*		
design	*ISPB*	*IPV*	*IMSPE*	*ISPB*	*IPV*	*IMSPE*
P1)	0.697	7.123	7.82	0.091	3.389	3.48
P2)	1.957	5.458	7.415	0.116	3.319	3.435
P3)	2.232	5.255	7.487	0.198	3.158	3.356
U	0.318	8.375	8.693	0.024	3.813	3.837
C	3.036	4.792	7.828	1.127	2.5	3.627

quadratic response $\theta_0 + \theta_1 x + \theta_2 x^2$ and test the significance of θ_2. A size 0.05 t-test of $H_0 : \theta_2 = 0$ $vs.$ $H_1 : \theta_2 \neq 0$ was carried out. The classical design is excluded from consideration because it does not have enough sites for fitting a quadratic response. The powers, based on $20,000$ simulations are given in Table 8. The same $320,000$ standard normal values are simulated and used in each of the four designs. The powers are low in view of the fact that the parametric form of the alternative hypothesis is correctly specified.

Example 3 *(Multiple Regression With Two Regressors).* Assume the model defined by (3) and (4) has two regressors: $p = 2$. Then $\beta_S = 0.5642$. Assume $\delta = 5$. We compare the designs (P1, P2, P3) with the continuous uniform (U) and

Table 8. The powers of tests in Examples 2 and 3

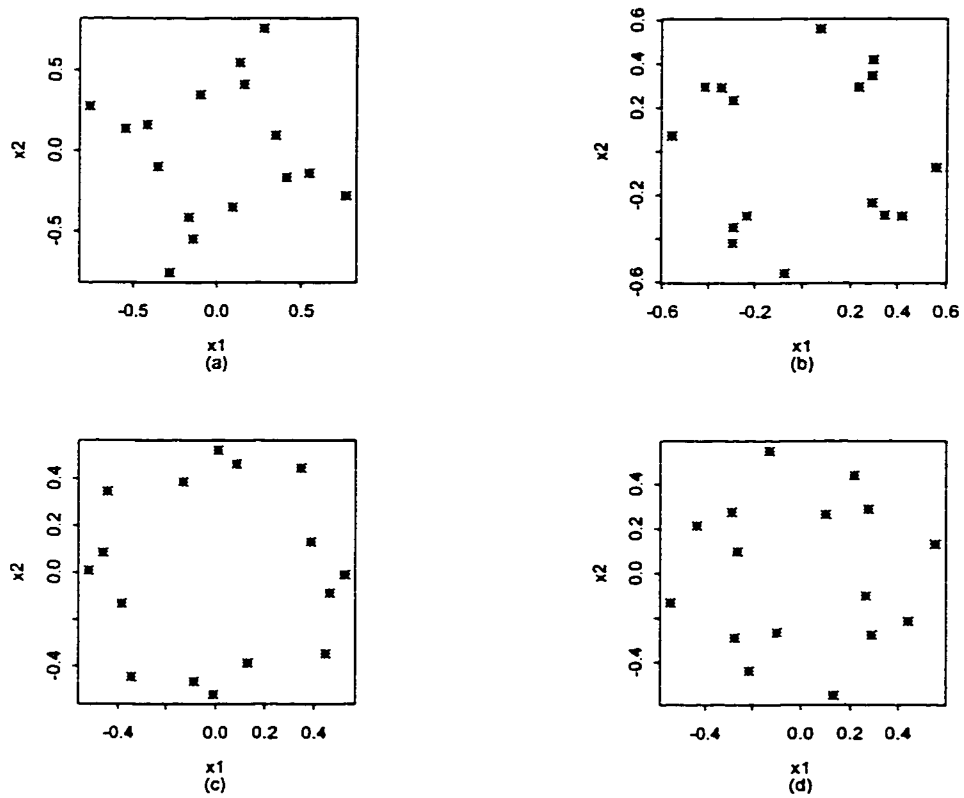
design	*linear response*	*bivariate response*
P1)	0.322	0.102
P2)	0.089	0.098
P3)	0.072	0.083
U	0.497	0.134

classically optimal (C) designs. As model contamination function, we use $f(\mathbf{x}) \propto [2(x_1^2+x_2^2)/\beta_S^2-1]$, with normalization $\int_S f^2(\mathbf{x})dx_1dx_2 = \eta_S^2 = \frac{1}{8}$. Let the sample size $n = 16$, then $\nu = 0.5$.

Design C consists of n points equally spaced over the boundary $||\mathbf{x}|| = \beta_S$. It is classically optimal in the sense of minimizing the integrated variance (IPV) when the model is correct. To implement the four continuous designs, apply Remark 3 of Section 3.1 with $p = 2$. See Figure 3.

The values of *ISPB*, *IPV* and *IMSPE* in Table 7 have the same performance measures as in Example 2. We test the significance of the second-order terms, that is, we test $H_0 : \theta_{11} = \theta_{22} = \theta_{12} = 0$. The test is the extra-sum-of-squares F-test. The powers are again based on $20,000$ $N(0,1)$ simulated errors and are given in Table 8. We find that the powers for P1) and U are much lower than those in Example 2 while the powers are approximately the same for P2) and P3) in both of Examples 2 and

Figure 3: Design points of Example 3: (a) $P1$); (b) $P2$); (c) $P3$); (d) U.



3. These powers indicate that the model departure used in these simulations is not likely to be detected. This strongly suggests that we should anticipate and address such departures at the design stage.

APPENDIX: DERIVATIONS

Proof of Theorem 1. First, we assume $\mathbf{G} > \mathbf{0}$, i.e. $\mathbf{G}$ is positive definite.

If $f \in \mathcal{F}$ is such that $\int_S f^2(\mathbf{x})d\mathbf{x} < \eta_S^2$, $\int_T f^2(\mathbf{x})d\mathbf{x} < \eta_T^2$, then we can define a function cf as $c_S f$ on S, $c_T f$ on T where $|c_S| > 1$, $|c_T| > 1$ and their signs are chosen such that $-2\mathbf{b}_{cf,T}^T\mathbf{B}_\xi^{-1}\mathbf{b}_{cf,\xi} = -2c_T c_S \mathbf{b}_{f,T}^T\mathbf{B}_\xi^{-1}\mathbf{b}_{f,\xi} > -2\mathbf{b}_{f,T}^T\mathbf{B}_\xi^{-1}\mathbf{b}_{f,\xi}$. Thus $ISPB(cf,\xi) > ISPB(f,\xi)$. Hence, it is sufficient to evaluate the maximum value of $ISPB(f,\xi)$, over f, subject to $\int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})d\mathbf{x} = \mathbf{0}$, $\int_S f^2(\mathbf{x})d\mathbf{x} = \eta_S^2$, $\int_T f^2(\mathbf{x})d\mathbf{x} = \eta_T^2$.

Given any $f \in \mathcal{F}$, define

$$h_f(\mathbf{x}) = \begin{cases} s_f \mathbf{z}^T(\mathbf{x})[m(\mathbf{x})\mathbf{I} - \mathbf{A}_S^{-1}\mathbf{B}_\xi]\mathbf{H}^{-1}\mathbf{b}_{f,\xi} & \text{if } x \in S, \\ t_f \mathbf{z}^T(\mathbf{x})\mathbf{B}_\xi^{-1}\mathbf{b}_{f,\xi} & \text{if } x \in T, \end{cases}$$

with $s_f^2 = \eta_S^2/\mathbf{b}_{f,\xi}^T\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,\xi}$, $t_f^2 = \eta_T^2/\mathbf{b}_{f,\xi}^T\mathbf{H}^{-1}\mathbf{b}_{f,\xi}$; and $s_f = \pm\sqrt{s_f^2}$, $t_f = \pm\sqrt{t_f^2}$ chosen such that $\mathbf{b}_{h_f,T}^T\mathbf{B}_\xi^{-1}\mathbf{b}_{h_f,\xi} \leq 0$. Then

$$i)\ h_f(\mathbf{x}) \in \mathcal{F}; \quad and \quad ii)\ ISPB(h_f,\xi) \geq ISPB(f,\xi).$$

In fact, $i)$ can be shown by verifying $\int_S h_f^2(\mathbf{x})d\mathbf{x} = \eta_S^2$, $\int_T h_f^2(\mathbf{x})d\mathbf{x} = \eta_T^2$, $\int_S \mathbf{z}(\mathbf{x})h_f(\mathbf{x})d\mathbf{x} = \mathbf{0}$. To prove $ii)$, we note that

(1) $\mathbf{b}_{h_f,T} = t_f \mathbf{A}_T \mathbf{B}_\xi^{-1} \mathbf{b}_{f,\xi}$ and $\mathbf{b}_{h_f,\xi} = s_f \mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,\xi}$,

(2) $\int_S f(\mathbf{x})h_f(\mathbf{x})d\mathbf{x} = s_f \mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{b}_{f,\xi}$ and $\int_T f(\mathbf{x})h_f(\mathbf{x})d\mathbf{x} = t_f \mathbf{b}_{f,\xi}^T \mathbf{B}_\xi^{-1}\mathbf{b}_{f,T}$.

Thus

$$\begin{aligned} ISPB(h_f,\xi) &= \mathbf{b}_{h_f,\xi}^T \mathbf{H}^{-1}\mathbf{b}_{h_f,\xi} + 2|\mathbf{b}_{h_f,T}^T \mathbf{B}_\xi^{-1}\mathbf{b}_{h_f,\xi}| + \eta_T^2 \\ &= s_f^2 \mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,\xi} + 2|s_f||t_f|\mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,\xi} + \eta_T^2. \end{aligned}$$

By the first equality of (2) and the Cauchy-Schwarz inequality, we have

$$s_f^2(\mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{b}_{f,\xi})^2 \le \int_S f^2(\mathbf{x})d(\mathbf{x}) \int_S h_f^2(\mathbf{x})d(\mathbf{x}) \le \eta_S^4,$$

so the definition of s_f gives

$$\eta_S^2 \ge \frac{(\mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{b}_{f,\xi})^2}{\mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,\xi}}.$$

Similarly,

$$\eta_T^2 \ge \frac{(\mathbf{b}_{f,\xi}^T \mathbf{B}_\xi^{-1}\mathbf{b}_{f,T})^2}{\mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{b}_{f,\xi}}.$$

Hence

$$\begin{aligned} &ISPB(h_f,\xi) \\ = \ & \frac{\eta_S^2(\mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}b_{f,\xi})}{\mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,\xi}} + \frac{2\eta_S\eta_T(\mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}b_{f,\xi})}{(\mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,\xi})^{1/2}(\mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{b}_{f,\xi})^{1/2}} + \eta_T^2 \\ \ge \ & \frac{(\mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{b}_{f,\xi})^2}{(\mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,\xi})^2}\mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,\xi} + 2|\mathbf{b}_{f,\xi}^T \mathbf{B}_\xi^{-1}\mathbf{b}_{f,T}| + \eta_T^2. \end{aligned}$$

This implies that $ISPB(h_f,\xi) \ge ISPB(f,\xi)$ if

$$(\mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{b}_{f,\xi})(\mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,\xi}) \ge (\mathbf{b}_{f,\xi}^T \mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,\xi})^2. \tag{A.1}$$

In fact, with $\mathbf{w} = \mathbf{H}^{-\frac{1}{2}}\mathbf{b}_{f,\xi}$ and $\mathbf{J} = \mathbf{H}^{-\frac{1}{2}}\mathbf{G}\mathbf{H}^{-\frac{1}{2}}$, (A.1) becomes

$$(\mathbf{w}^T\mathbf{w})(\mathbf{w}^T\mathbf{J}^2\mathbf{w}) \geq (\mathbf{w}^T\mathbf{J}\mathbf{w})^2. \tag{A.2}$$

This is true by Cauchy-Schwarz inequality and (A.1) follows.

Therefore, we can restrict to $f \in \mathcal{F}$ of the form

$$f(\mathbf{x};\mathbf{a},\mathbf{c}) = \begin{cases} \mathbf{z}^T(\mathbf{x})[m(\mathbf{x})\mathbf{I} - \mathbf{A}_S^{-1}\mathbf{B}_\xi]\mathbf{a} & \text{if } x \in S, \\ \mathbf{z}^T(\mathbf{x})\mathbf{B}_\xi^{-1}\mathbf{c} & \text{if } x \in T, \end{cases}$$

where $\mathbf{a}$ and $\mathbf{c}$ are arbitrary vectors satisfying

$$\eta_S^2 = \int_S f^2(\mathbf{x};\mathbf{a},\mathbf{c})d\mathbf{x} = \mathbf{a}^T\mathbf{G}\mathbf{a}, \qquad \eta_T^2 = \int_T f^2(\mathbf{x};\mathbf{a},\mathbf{c})d\mathbf{x} = \mathbf{c}^T\mathbf{H}^{-1}\mathbf{c}.$$

Subject to these conditions we maximize

$$\begin{aligned} ISPB(f,\xi) &= \mathbf{b}_{f,\xi}^T\mathbf{H}^{-1}\mathbf{b}_{f,\xi} - 2\mathbf{b}_{f,T}^T\mathbf{B}_\xi^{-1}\mathbf{b}_{f,\xi} + \eta_T^2 \\ &= \mathbf{a}^T\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{a} - 2\mathbf{c}^T\mathbf{H}^{-1}\mathbf{G}\mathbf{a} + \eta_T^2. \end{aligned}$$

This is maximized by

$$\mathbf{c} = -\frac{\eta_T\mathbf{G}\mathbf{a}}{\|\mathbf{H}^{-1/2}\mathbf{G}\mathbf{a}\|},$$

and then

$$\begin{aligned} ISPB(f,\xi) &= \mathbf{a}^T\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{a} + 2\eta_T\frac{\mathbf{a}^T\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{a}}{\|\mathbf{H}^{-1/2}\mathbf{G}\mathbf{a}\|} + \eta_T^2 \\ &= (\sqrt{\mathbf{a}^T\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{a}} + \eta_T)^2. \end{aligned}$$

With $\mathbf{a}_0 = \mathbf{a}/\eta_S$, we are then to maximize

$$ISPB(f,\xi) = (\eta_S\sqrt{\mathbf{a}_0^T\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{a}_0} + \eta_T)^2,$$

subject to $\mathbf{a}_0^T\mathbf{G}\mathbf{a}_0 = 1$. This is a standard eigenvalue problem. If λ_m is the largest solution to

$$|\mathbf{G}\mathbf{H}^{-1}\mathbf{G} - \lambda\mathbf{G}| = 0,$$

that is

$$|\mathbf{G} - \lambda\mathbf{H}| = 0,$$

then the maximizer $\mathbf{a}_0$ will be defined by

$$(\mathbf{G}\mathbf{H}^{-1}\mathbf{G} - \lambda_m\mathbf{G})\mathbf{a}_0 = \mathbf{0},$$

normalized to satisfy $\mathbf{a}_0^T\mathbf{G}\mathbf{a}_0 = 1$. This proves the theorem when $\mathbf{G} > 0$.

If a design density $m(\mathbf{x})$ is such that $\mathbf{G} \geq \mathbf{0}$, but $|\mathbf{G}| = 0$, we can approximate it as follows. Take any density $m_1(\mathbf{x})$, with corresponding matrix $\mathbf{G}_1 > 0$. Let $m_t(\mathbf{x}) = (1-t)m(\mathbf{x}) + tm_1(\mathbf{x})$, then $m_t(\mathbf{x})$ is a density. Define $\mathbf{G}_t$ as the matrix corresponding to $\mathbf{G}$ when the density is $m_t(\mathbf{x})$ and $p(t) = |\mathbf{G}_t|$. Then $p(t)$ is a polynomial in $t \in [0,\ 1]$ and $p(0) = 0$, $p(1) = |\mathbf{G}_1| > 0$. So $p(t)$ is non-constant and $p(t) \geq 0$. This implies that $p(t) > 0$ for all sufficiently small $t > 0$. Now, the theorem holds for $\mathbf{G}_t$. By letting $t \to 0$, we show that the theorem holds for $\mathbf{G}$. □

Proof of Theorem 2. (i) Based on Theorem 1, the maximum $IMSPE(f,\xi)$ is

$$max_{f\in\mathcal{F}}IMSPE(f,\xi) = \eta_S^2[\nu(\delta_p + p\frac{\gamma_T}{\gamma}) + (\sqrt{\lambda_m} + r_{T,S})^2],$$

where

$$\lambda_m = max\{\delta_p(\int_S m^2(\mathbf{x})d\mathbf{x} - 1);\ \ \delta_{p+2}(\frac{\gamma_S}{p\gamma^2}\int_S \|\mathbf{x}\|^2 m^2(\mathbf{x})d\mathbf{x} - 1)\}.$$

Under the stated conditions, we show that $m_1(\mathbf{x})$ minimizes

$$\Phi(m) = \eta_S^2[\nu(\delta_p + p\frac{\gamma_T}{\gamma}) + (\sqrt{\delta_{p+2}(\frac{\gamma_S}{p\gamma^2}\int_S \|\mathbf{x}\|^2 m^2(\mathbf{x})d\mathbf{x} - 1)} + r_{T,S})^2],$$

and that

$$\lambda_{m_1} = \delta_{p+2}(\frac{\gamma_S}{p\gamma^2}\int_S \|\mathbf{x}\|^2 m_1^2(\mathbf{x})d\mathbf{x} - 1).$$

We first minimize $\Phi(m)$ for fixed γ, then minimize over γ, i.e. we first minimize $\int_S \|\mathbf{x}\|^2 m^2(\mathbf{x})d\mathbf{x}$ for fixed γ, then minimize the resulting $\Phi(m)$ over γ.

At the first stage, we have

$$m_1(\mathbf{x}) = a[1 - b(\frac{\beta_S}{\|\mathbf{x}\|})^2]^+,$$

with Lagrange multipliers a, b determined by $\int_S m_1(\mathbf{x})d\mathbf{x} = 1$, $\int_S \|\mathbf{x}\|^2 m_1(\mathbf{x})d\mathbf{x} = p\gamma$.

If $b < 0$, then $m_1(\mathbf{x}) = a[1 - b(\frac{\beta_S}{\|\mathbf{x}\|})^2]$. When $p < 3$, we have $\int_S m_1(\mathbf{x})d\mathbf{x} = \infty$. This is a contradiction. When $p \geq 3$, we have $a - \frac{pab}{p-2} = 1$, $p\gamma = ap\gamma_S - ab(p+2)\gamma_S$. This implies $\frac{\gamma}{\gamma_S} = 1 + \frac{4ab}{p(p-2)} < 1$, which is useless to us.

Thus, we only need to consider the case $b \in [0,\ 1]$. Then a, b satisfy

$$a = \frac{1}{pk_{p-2}(b)}, \tag{A.3}$$

$$\frac{\gamma}{\gamma_S} = a(p+2)k_p(b). \tag{A.4}$$

So $m_1(\mathbf{x})$ can be parameterized by $b \in [0,\ 1]$, with a, γ determined by (A.3), (A.4).

Since

$$\int_S \|\mathbf{x}\|^2 m_1^2(\mathbf{x})d\mathbf{x} = a\int_S \|\mathbf{x}\|^2 m_1(\mathbf{x})d\mathbf{x} - ab\beta_S^2\int_S m_1(\mathbf{x})d\mathbf{x} = ap\gamma - ab\beta_S^2,$$

$$\int_S m_1^2(\mathbf{x})d\mathbf{x} = \int_0^{\beta_S} (a[1 - b\beta_S^2/u^2]^+)^2 \frac{pu^{p-1}}{\beta_S^p} du = \int_{\sqrt{b}}^1 a^2(v^2 - b)^2 pv^{p-5} dv$$
$$= pa^2 k_{p-2}(b) - pa^2 b k_{p-4}(b),$$

then by (A.3) and (A.4), we have

$$\delta_{p+2}(\frac{\gamma_S}{p\gamma^2} \int_S \|\mathbf{x}\|^2 m_1^2(\mathbf{x})d\mathbf{x} - 1) = \delta_{p+2} g_p(b),$$

$$\delta_p(\int_S m_1^2(\mathbf{x})d\mathbf{x} - 1) = \delta_p g_{p-2}(b),$$

$$\Phi(m_1) = \eta_S^2\{\nu[\delta_p + p\delta_{p+2}\frac{pk_{p-2}(b)}{(p+2)k_p(b)}] + [\sqrt{\delta_{p+2}g_p(b)} + r_{T,S}]^2\}.$$

Therefore, the definition of b_1 indicates that b_1 is the minimizer of $\Phi(m_1)$ and it follows that

$$m_1(\mathbf{x}) = a_1[1 - b_1(\frac{\beta_S}{\|\mathbf{x}\|})^2]^+$$

with a_1 determined by (A.3) (take $b = b_1$), minimizes $\Phi(m)$ unconditionally. If (8) holds then $\lambda_{m_1} = \delta_{p+2}(\frac{\gamma_S}{p\gamma^2} \int_S \|\mathbf{x}\|^2 m_1^2(\mathbf{x})d\mathbf{x} - 1)$ and it implies that $m_1(\mathbf{x})$ is the minimax extrapolation design density. Now, we have the following facts:

1). The function $l_1(\delta_p) = \frac{\delta_{p+2}}{\delta_p} = \frac{(1+\delta_p)^{1+\frac{2}{p}} - 1}{\delta_p}$ is an increasing function of $\delta_p \in (0, \infty)$, with $\lim_{\delta_p \to 0} l_1(\delta_p) = 1 + \frac{2}{p}$.

2). For $b \in [0, 1]$, the function $l_2(b) = \frac{g_{p-2}(b)}{g_p(b)}$ is decreasing, with $\lim_{b \to 1} l_2(b) = 1 + \frac{2}{p}$ and $\lim_{b \to 0} l_2(b) = \infty$.

3). The function $l_3(b) = (-1)\frac{\partial}{\partial b}[\sqrt{\delta_{p+2}g_p(b)} + r_{T,S}]^2 / (p\delta_{p+2}\frac{\partial}{\partial b}[\frac{pk_{p-2}(b)}{(p+2)k_p(b)}])$ is increasing, with $lim_{b \to 1} l_3(b) = \infty$.

And 1) and 2) imply that for fixed δ_p, (8) holds when b is large. Hence, by 3), (8) is true when $\nu \in [\nu_4(\delta),\ \infty]$, for some $\nu_4(\delta) > 0$.

(ii). Similar to (i), under the stated conditions, we show that $m_1(\mathbf{x})$ minimizes

$$\Psi(m) = \eta_S^2[\nu(\delta_p + p\frac{\gamma_T}{\gamma}) + (\sqrt{\delta_p(\int_S m^2(\mathbf{x})d\mathbf{x} - 1)} + r_{T,S})^2],$$

and that

$$\lambda_{m_1} = \delta_p(\int_S m_1^2(\mathbf{x})d\mathbf{x} - 1).$$

First fix ρ, i.e. fix $\rho = \frac{\gamma}{\gamma_S}$, then the minimizer of $\Psi(m)$ has the form:

$$m_{1,\rho}(x) = (a||\mathbf{x}||^2 + b)^+,$$

with Lagrange's multipliers $a,\ b$ determined by $\int_S m_{1,\rho}(\mathbf{x})d\mathbf{x} = 1,\ \int_S ||\mathbf{x}||^2 m_{1,\rho}(\mathbf{x})d\mathbf{x} = p\gamma$. The forms of the minimax extrapolation design depend on the sign of b.

a). $a, b > 0$, then $1 < \rho < \frac{(p+2)^2}{p(p+4)}$ and $m_{1,\rho}(x) = 1 + \frac{p+4}{4}(\rho - 1)(\frac{||\mathbf{x}||^2}{\gamma_S} - p)$. Therefore,

$$\begin{aligned}
&\delta_{p+2}[\frac{\gamma_S}{p\gamma^2}\int_S ||\mathbf{x}||^2 m_{1,\rho}^2(\mathbf{x})d\mathbf{x} - 1] = \delta_{p+2}[\frac{(p+2)^3}{4(p+6)}(\frac{1}{\rho} - 1)^2],\\
&\delta_p[\int_S m_{1,\rho}^2(x)dx - 1] = \delta_p\frac{p(p+4)}{4}(\rho - 1)^2,\\
&\Psi(m_{1,\rho}) = \eta_S^2[\nu(\delta_p + p\delta_{p+2}\frac{1}{\rho}) + (\kappa(\rho - 1) + r_{T,S})^2].
\end{aligned}$$

So, by the definition of ρ_1, $m_1(\mathbf{x}) = 1 + \frac{p+4}{4}(\rho_1 - 1)(\frac{||\mathbf{x}||^2}{\gamma_S} - p)$ minimizes $\Psi(m)$ unconditionally. If (10) holds, then $\lambda_{m_1} = \delta_p(\int_S m_1^2(\mathbf{x})d\mathbf{x} - 1)$ and it follows $m_1(\mathbf{x})$ is the solution to P1). The remaining statement of i) follows as in (i) by the facts:

1). ρ_0 increases with δ_p and $\lim_{\delta_p \to 0} \rho_0 = (\frac{(p+2)^4}{p^2(p+4)(p+6)})^{\frac{1}{2}} < (1 + \frac{4}{p(p+4)})$;

2). ν increases with ρ.

b). $a > 0,\ b < 0$. Define $c \in (0,\ 1)$ by $c^2\beta_S^2 = -b/a$. Then $\frac{(p+2)^2}{p(p+4)} \leq \rho = \frac{J_{p+2}(c)}{J_p(c)} \leq \frac{p+2}{p}$ and $m_{1,\rho}(x) = \frac{1}{J_p(c)}[(\frac{\|\mathbf{x}\|}{\beta_S})^2 - c^2]^+$, described by the parameter c. Thus,

$$\delta_{p+2}[\frac{\gamma_S}{p\gamma^2} \int_S \|\mathbf{x}\|^2 m_{1,\rho}^2(\mathbf{x})d\mathbf{x} - 1] = \delta_{p+2}h_{p+2}(c)$$

$$\delta_p[\int_S m_{1,\rho}^2(\mathbf{x})d\mathbf{x} - 1] = \delta_p h_p(c);$$

$$\Psi(m_{1,\rho}) = \eta_S^2[\nu(\delta_p + p\delta_{p+2}\frac{J_p(c)}{J_{p+2}(c)}) + (\sqrt{\delta_p h_p(c)} + r_{T,S})^2]$$

Hence, if c_1 is the root of (11), then $m_1(\mathbf{x}) = \frac{1}{J_p(c_1)}[(\frac{\|\mathbf{x}\|}{\beta_S})^2 - c_1^2]^+$ minimizes $\Psi(m)$ unconditionally. If (12) holds then $\lambda_{m_1} = \delta_p(\int_S m_1^2(\mathbf{x})d\mathbf{x} - 1)$ and it follows that $m_1(\mathbf{x})$ is the solution to P1). The remaining part of *b*) is similar to *a*). □

Proof of Theorem 3: In the light of the fact: $\lambda_m = 0$ and $\int_S \|\mathbf{x}\|^2 m(\mathbf{x})d\mathbf{x} = p\gamma_S$ when $m(\mathbf{x}) \equiv 1$.

Proof of Lemma 4: Let $m_{21}(\mathbf{x})$ be the solution to the second minimization, then

$$m_{21} = a[1 - b\frac{\beta_s^2}{\|\mathbf{x}\|^2}]^+$$

with $a,\ b \in (0,\ 1)$ determined by $\int_S m_{21}(\mathbf{x})d\mathbf{x} = 1$, $\int_S \|\mathbf{x}\|^2 m_{21}(\mathbf{x})d\mathbf{x} = d_1$, i.e. (note that $\beta_S^2 = (p+2)\gamma_S$)

$$a = \frac{1}{pk_{p-2}(b)}, \tag{A.5}$$

$$d_1 = apk_p(b)\beta_S^2 = ak_p(b)p(p+2)\gamma_S, \tag{A.6}$$

Hence, we have (note that $p\gamma = d_1$ here):

$$\begin{aligned} \frac{\int_S \|\mathbf{x}\|^2 m_{21}^2(\mathbf{x})d\mathbf{x}}{p\gamma^2} &= \frac{p}{d_1^2}[ad_1 - ab(p+2)\gamma_S] \\ &= \frac{g_p(b)+1}{\gamma_s} \end{aligned}$$

which is an increasing function of b. So it is an increasing function of d_1 by the fact that d_1 increases with b. This proves the Lemma. □

Proof of Theorem 5: We want to show :

1) $m_2(\mathbf{x}) = argmin\{\delta_{p+2}(\frac{\gamma_S}{p\gamma^2}\int_S \|\mathbf{x}\|^2 m^2(\mathbf{x})d\mathbf{x} - 1) : \int_S m(\mathbf{x})d\mathbf{x} = 1,\ \int_S \|\mathbf{x}\|^2 m(\mathbf{x})d\mathbf{x} \geq d_1\}$,

2) $\lambda_{m_2} = \delta_{p+2}(\frac{\gamma_S}{p\gamma^2}\int_S \|\mathbf{x}\|^2 m_2^2(\mathbf{x})d\mathbf{x} - 1)$.

As in Lemma 4, we have 1), and that

$$\begin{aligned} \delta_{p+2}(\tfrac{\gamma_S}{p\gamma^2}\textstyle\int_S \|\mathbf{x}\|^2 m^2(\mathbf{x})d\mathbf{x} - 1) = \delta_{p+2}g_p(b) &= ((1+\delta_p)^{1+2/p} - 1)g_p(b), \\ \delta_p[\textstyle\int_S m_2^2(\mathbf{x})d\mathbf{x} - 1] = \delta_p g_{p-2}(b). \end{aligned}$$

Thus, 2) is true if $\frac{(1+\delta_p)^{1+2/p}-1}{\delta_p} \geq \frac{g_{p-2}(b)}{g_p(b)}$. But the latter holds for $d_1 \in [d_1(\delta_p),\ (p+2)\gamma_S]$, with $d_1(\delta_p)$ determined as in the statement of the Theorem, in light of the facts: $\frac{g_{p-2}(b)}{g_p(b)}$ decreases with b and hence decreases with d_1; $\lim_{b\to 1}\frac{g_{p-2}(b)}{g_p(b)} = 1 + \frac{2}{p} \leq \frac{(1+\delta_p)^{1+2/p}-1}{\delta_p}$. This complete the proof. □

Proof of Lemma 6: For any γ, if $C_\gamma \neq \emptyset$, then there exists a density $m_*(\mathbf{x})$, such that

$$\int_S m_*(\mathbf{x})d\mathbf{x} = 1,\ \int_S \|\mathbf{x}\|^2 m_*(\mathbf{x})d\mathbf{x} = p\gamma,\ \int_S \|\mathbf{x}\|^2 m_*^2(\mathbf{x})d\mathbf{x} \leq k_2\gamma^2.$$

But, we have parameters a, b, such that

$$\begin{aligned} & argmin\{\int_S \|\mathbf{x}\|^2 m^2(\mathbf{x})d\mathbf{x} : \int_S m(\mathbf{x})d\mathbf{x} = 1, \int_S \|\mathbf{x}\|^2 m(\mathbf{x})d\mathbf{x} = p\gamma\} \\ = & a[1 - b(\frac{\beta_S}{\|\mathbf{x}\|})^2]^+, \end{aligned}$$

with a, b determined by (16), (17). So $\int_S \|\mathbf{x}\|^2(a[1-b(\frac{\beta_S}{\|\mathbf{x}\|})^2]^+)^2 d\mathbf{x} \leq \int_S \|\mathbf{x}\|^2 m_*^2(\mathbf{x})d\mathbf{x}$ $\leq k_2\gamma^2$. This proves (i). Since

$$\begin{aligned} \frac{\int_S \|\mathbf{x}\|^2(a[1-b(\frac{\beta_S}{\|\mathbf{x}\|})^2]^+)^2 d\mathbf{x}}{\gamma^2} &= \frac{1}{\gamma^2}[pa\gamma - ab(p+2)\gamma_S] \\ &= \frac{p(g_p(b)+1)}{\gamma_S}, \end{aligned}$$

is an increasing function of b, it increases with γ. Hence, (ii) is true. □

Proof of Theorem 7: Simple calculations show that

$$\delta_p(\int_S m_{31}^2(x)dx - 1) = \delta_p g_{p-2}(b),$$

with b determined by (18), i.e.

$$\frac{p(k_1^2 + \delta_{p+2})}{\gamma_T} = k_2 = \frac{p(g_p(b)+1)}{\gamma_S}.$$

Thus

$$k_1^2 = \delta_{p+2} g_p(b),$$

and (14) can be rewritten as:

$$\frac{g_{p-2}(b)}{g_p(b)} \leq \frac{(1+\delta_p)^{1+p/2} - 1}{\delta_p}.$$

The monotonicity of these functions implies the theorem. □

REFERENCES

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

Dette, H. and Wong, W.K. (1996). Robust Optimal Extrapolation Designs. *Biometrika* 83, 3, 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*, in press.

Herzberg, A. and Cox, D. (1972). Some Optimal Designs for Interpolation and Extrapolation. *Biometrika,* 59, 551 – 561.

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

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

Huang, M. N. L. (1990). Optimal Extrapolation Designs for a Partly Linear Model. *Computational Statistics and Data Analysis*, 10, 109 – 115.

Huber, P. J. (1975). Robustness and Designs. In: *A Survey of Statistical Design and Linear Models*, North-Holland, Amsterdam, 287 – 303.

Karson, N. J., Manson, A. R. and Hader,R. J. (1969). Minimum Bias Estimation and Experimental Design for Response Surfaces. *Technometrics*, 11, 461 – 475.

Kiefer, J. and Wolfowitz, J. (1964a). Optimum Extrapolation Designs (I). *Ann. Inst. Statist. Math.*, 16, 79 – 108.

Kiefer, J. and Wolfowitz, J. (1964b). Optimum Extrapolation Designs (II). *Ann. Inst. Statist. Math.*, 16, 295 – 303.

Kiefer, J. and Wolfowitz, J. (1965). On a Theorem of Hoel and Levine on Extrapolation. *Annals of Mathematical Statistics*, 36, 1627 – 1655.

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

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

Wiens, D. P. (1992). Minimax Designs for Approximately Linear Regression. *Journal of Statistical Planning and Inference*, 31, 355 – 371.

Chapter Three

Robust Extrapolation Designs and Weights for Biased Regression Models With Heteroscedastic Errors[1]

Abstract

In this article we consider the construction of designs for the extrapolation of regression responses, allowing both for possible heteroscedasticity in the errors and for imprecision in the specification of the response function. We find minimax designs and correspondingly optimal estimation weights in the context of the following problems: 1) For ordinary least squares estimation, determine a design to minimize the maximum value of the Integrated Mean Squared Prediction Error *(IMSPE)*, with the maximum being evaluated over both types of departure; 2) For weighted least squares estimation, determine both weights and a design to minimize the maximum *IMSPE*; 3) Choose weights and design points to minimize the maximum *IMSPE*, subject to a side condition of unbiasedness. Solutions to 1) and 2) are given for multiple linear regression with no interactions, a spherical design space and an annular extrapolation space. For 3) the solution is given in complete generality; as one example we consider polynomial regression. Applications to a dose response problem for bioassays are discussed. Numerical comparisons including a simulation study indicate that, as well

[1]Co-authored with D.P. Wiens. To appear in The Canadian Journal of Statistics.

as being easily implemented, the designs/weights for 3) perform as well as those for 1) and 2), and outperform some common competitors, for moderate but undetectable amounts of model bias.

1 Introduction

In this article we study the construction of designs for the extrapolation of regression responses, in the presence of both possible error heteroscedasticity and an approximately and possibly incorrectly specified response function. Design problems for *estimation* in the face of response uncertainty, but for *homoscedastic* errors, have been studied by Box and Draper (1959), Huber (1975), Pesotchinsky (1982), Wiens (1992) and others; Wiens (1998) allows also for heteroscedastic errors. Designs under error heteroscedasticity, assuming the fitted response to be exactly correct, were considered by Wong (1992) and Wong and Cook (1993); both of these papers assumed a *known* variance structure. Designs for extrapolation of polynomials, again assuming a correctly specified response, were studied by Kiefer and Wolfowitz (1964a,b) and Hoel and Levine (1964). Studden (1971) studied such problems for multivariate polynomial models. Spruill (1984) and Dette and Wong (1996) constructed extrapolation designs for polynomial regression, robust against various misspecifications of the degree of the polynomial. 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. 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).

Extrapolation to regions outside of that in which observations are taken is of course an inherently risky procedure, and is made even more so by an over-reliance on stringent model assumptions. For such reasons we shall take rather broad classes of departures from the usual linear model:

1. The response is taken to be only *approximately* linear in the regressors; *viz.*

$$E(Y|\mathbf{x}) = \boldsymbol{\theta}^T \mathbf{z}(\mathbf{x}) + f(\mathbf{x}) \tag{1}$$

   for a p-dimensional vector $\mathbf{z}$ of regressors, depending on a q-dimensional vector $\mathbf{x}$ of independent variables. The response contaminant f represents uncertainty about the exact nature of the regression response and is unknown and arbitrary, subject to certain restrictions detailed in Section 2. One estimates $\boldsymbol{\theta}$ but not f, leading possibly to biased estimation of $E(Y|\mathbf{x})$ and consequently to biased predictions. The experimenter is to take n uncorrelated observations $Y_i = E(Y|\mathbf{x}_i) + \varepsilon(\mathbf{x}_i)$, with $\mathbf{x}_i$ freely chosen from a design space S. The goal is to extrapolate the estimates of $E(Y|\mathbf{x})$ to a given region T disjoint from S.

2. The random errors, although uncorrelated with mean zero, are possibly heteroscedastic: $\mathrm{VAR}\{\varepsilon(\mathbf{x})\} = \sigma^2 g(\mathbf{x})$ for a function g satisfying assumptions given in Section 2.

As an optimality criterion we take an analogue of the classical notion of Q-optimality: the supremum, over f and g, of the Integrated Mean Squared Prediction Error (*IMSPE*) of the predicted values $\hat{Y}(\mathbf{x})$, with the integration being over the extrapolation region T, is to be minimized by an appropriate choice of design. The following problems will be addressed:

P1) For *ordinary* least squares *(OLS)* estimation, determine designs to minimize the maximum, over f and g, value of the *IMSPE.*

P2) For *weighted* least squares *(WLS)* estimation, determine designs and weights to minimize the maximum *IMSPE.*

P3) Choose weights and design points to minimize the maximum *IMSPE*, subject to a side condition of unbiasedness.

As a possible application, consider the following extrapolation problem for bioassays. Let $P(x)$ be the probability of a particular response when a drug is administered at dose x. At various levels of x one 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 such as the logistic. The regression function is then modelled as

$E(Y|x) \approx G^{-1}(P(x))$. Since $P(x)$ is unknown, $E(Y|x)$ is often approximated by a low-degree polynomial $\zeta(x)$. 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, $x \in [a, 1]$ $(a > 0)$ to estimate $E(Y|x = 0)$. Krewski *et al* (1986) consider designs for such problems assuming that $E(Y|x)$ is exactly linear in $\ln x$. Lawless (1984) takes an approach closer to ours, obtaining designs which 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; the approach of the current article is to model it (by $f(0)$ in (1)) in such a way as to open the door to a minimax treatment. Another point of departure of our approach from that of Lawless (1984) or Huber (1975) is that although our treatment does not allow the case $T = \{0\}$ (or any other extrapolation space of Lebesgue measure zero), it *does* treat the case of an interval T, i.e. extrapolation to a *range* of values near $x = 0$. This is particularly significant if the problem is to determine a 'virtually safe dose' (Cornfield 1977).

Despite these differences, Lawless (1984) reaches 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."

The designs and weights which constitute solutions to problems P1), P2) and P3) are given in Sections 3, 4 and 5 respectively. Those for P1) and P2) are theoretically and numerically rather complex, and our solutions are restricted to situations exhibiting considerable structure. In contrast, the solution to P3) is given in complete generality and turns out to be computationally straightforward. We apply the solution to P3) to the dose response problem described above. A comparative study accompanied by concluding remarks and recommendations is given in Section 6. Some mathematical preliminaries are detailed in Section 2, where we reduce each of P1) - P3) to a single minimization over a class of densities. Proofs for Section 2 are postponed to the appendix.

2 PRELIMINARIES AND NOTATION

For the regression model described in Section 1 we shall assume that the contamination function $f(\mathbf{x})$ in (1) is an unknown member of the class

$$\mathcal{F} = \{f \mid \int_S f^2(\mathbf{x})d\mathbf{x} \leq \eta_S^2 < \infty, \int_T f^2(\mathbf{x})d\mathbf{x} \leq \eta_T^2 < \infty, \int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})d\mathbf{x} = \mathbf{0}\}, \quad (2)$$

for positive constants η_S and η_T. The random errors $\epsilon(\mathbf{x}_i)$ satisfy

$$VAR\{\epsilon(\mathbf{x}_i)\} = \sigma^2 g(\mathbf{x}_i), \int_S g^2(\mathbf{x})d\mathbf{x} \leq \Omega^{-1} := \int_S d\mathbf{x} < \infty. \quad (3)$$

The last condition of (2) is required in order that the true parameter $\boldsymbol{\theta}$ be uniquely defined, and then $\boldsymbol{\theta} := \arg\min_{\mathbf{t}} \int_S \{\mathbf{t}^T\mathbf{z}(\mathbf{x}) - E(Y|\mathbf{x})\}^2 d\mathbf{x}$. One can instead *start* with this definition of $\boldsymbol{\theta}$, then *define* $f(\mathbf{x}) = \boldsymbol{\theta}^T\mathbf{z}(\mathbf{x}) - E(Y|\mathbf{x})$, thus obtaining the last condition of (2) as a natural consequence of the definition of the parameter being estimated. The other conditions of (2) are needed to ensure that errors due to estimation and prediction bias remain bounded. The conditions of (3) are equivalent to defining $\sigma^2 = sup_g[\int_S VAR^2\{\epsilon(\mathbf{x})\}\Omega d\mathbf{x}]^{\frac{1}{2}}$.

At the outset the only assumptions made about T are that it be disjoint from S and that the integrals in (2) exist; special cases will be considered in Sections 3 to 5. The requirement that T and S be disjoint need not exclude the application of our results to interpolation problems, i.e. the case $T \subset S$, as long as design points are not to be chosen from within T. One can then replace S by $S\backslash T$. The cases in which $S \subset T$, or in which S and T are merely overlapping, may be handled similarly. If design points may be chosen from within T then $f(\mathbf{x})$ is defined for values $\mathbf{x} \in S \cap T$ and our method of maximizing the loss over $\mathcal{F}$ fails.

We remark that for P1) and P2) our results depend on the unknown parameters only through $\nu := \sigma^2/(n\eta_S^2)$ and $r_{T,S} := \eta_T/\eta_S$; for P3) no knowledge whatsoever is required of these parameters. One can interpret ν as representing the relative importance of bias versus variance, in the mind of the experimenter. As $\nu \rightarrow 0$ bias completely dominates the problem, whereas $\nu \rightarrow \infty$ results in a 'pure variance' problem. Similarly, the choice of $r_{T,S}$ reflects the relative amounts of model response uncertainty in the extrapolation and design spaces. In our simulation study for this article we made the rather arbitrary choice $\nu = \Omega$ and the intuitively appealing choice

$r_{T,S} = 1$; a perhaps equally appealing choice of $r_{T,S}$ is the ratio of the volume of T to that of S. The qualitative aspects of the results did not change when other choices of ν and $r_{T,S}$ were made.

To avoid trivialities, and to ensure the nonsingularity of a number of relevant matrices, we assume that the design and extrapolation spaces satisfy

A) For each $\mathbf{a} \neq \mathbf{0}$, the set $\{\mathbf{x} \in S \cup T : \mathbf{a}^T\mathbf{z}(\mathbf{x}) = 0\}$ has Lebesgue measure zero.

We propose to estimate $\boldsymbol{\theta}$ by least squares, possibly weighted with non-negative weights $w(\mathbf{x})$. Let ξ be the design measure, i.e. $\xi = n^{-1}\sum_{i=1}^{n}\delta_{\mathbf{x}_i}$, where $\delta_{\mathbf{x}}$ is point mass at $\mathbf{x}$. Define matrices and vectors

$$\mathbf{A}_T = \int_T \mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})d\mathbf{x}, \qquad \mathbf{A}_S = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})d\mathbf{x},$$

$$\mathbf{B} = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})w(\mathbf{x})\xi(d\mathbf{x}), \quad \mathbf{D} = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})w^2(\mathbf{x})g(\mathbf{x})\xi(d\mathbf{x}),$$

$$\mathbf{b}_{f,S} = \int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})w(\mathbf{x})\xi(d\mathbf{x}), \quad \mathbf{b}_{f,T} = \int_T \mathbf{z}(\mathbf{x})f(\mathbf{x})d\mathbf{x}.$$

It follows from A) that $\mathbf{A}_T$ and $\mathbf{A}_S$ are nonsingular, and that $\mathbf{B}$ is non-singular if ξ does not place mass on sets of Lebesgue measure zero. As discussed below, this sufficient requirement turns out to be necessary as well.

The *WLS* estimator of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \mathbf{B}^{-1} \cdot n^{-1}\sum_{i=1}^{n}\mathbf{z}(\mathbf{x}_i)w(\mathbf{x}_i)Y_i = \mathbf{B}^{-1}\int_S \mathbf{z}(\mathbf{x})w(\mathbf{x})y(\mathbf{x})\xi(d\mathbf{x}),$$

with bias vector and covariance matrix

$$E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta} = \mathbf{B}^{-1}\mathbf{b}_{f,S}, \quad COV(\hat{\boldsymbol{\theta}}) = \frac{\sigma^2}{n}\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}.$$

Note that $\sigma^2/n = \eta_S^2\nu$; we shall henceforth use the latter expression since it will generally appear together with functions of the bias.

We predict $E(Y|\mathbf{x})$ for $\mathbf{x} \in T$ by $\hat{Y}(\mathbf{x}) = \hat{\boldsymbol{\theta}}^T\mathbf{z}(\mathbf{x})$ and consider the resulting Integrated Mean Squared Prediction Error. The *IMSPE* splits into terms due to prediction bias, prediction variance, and model misspecification:

$$\begin{aligned} IMSPE(f,g,w,\xi) &= \int_T E[\{\hat{Y}(\mathbf{x}) - E(Y|\mathbf{x})\}^2]d\mathbf{x} \\ &= IPB(f,w,\xi) + IPV(g,w,\xi) + \int_T f^2(\mathbf{x})d\mathbf{x}, \end{aligned}$$

where, with $\mathbf{H} := \mathbf{B}\mathbf{A}_T^{-1}\mathbf{B}$, the Integrated Prediction Bias (*IPB*) and Integrated Prediction Variance (*IPV*) are

$$\begin{aligned} IPB(f,w,\xi) &= \int_T [E\{\hat{Y}(\mathbf{x}) - \boldsymbol{\theta}^T\mathbf{z}(\mathbf{x})\}]^2 d\mathbf{x} - 2\int_T E\{\hat{Y}(\mathbf{x}) - \boldsymbol{\theta}^T\mathbf{z}(\mathbf{x})\}f(\mathbf{x})d\mathbf{x} \\ &= \mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{b}_{f,S} - 2\mathbf{b}_{f,T}^T\mathbf{B}^{-1}\mathbf{b}_{f,S}, \qquad (4) \\ IPV(g,w,\xi) &= \int_T VAR\{\hat{Y}(\mathbf{x})\}d\mathbf{x} = \eta_S^2\nu\int_S \mathbf{z}^T(\mathbf{x})\mathbf{H}^{-1}\mathbf{z}(\mathbf{x})w^2(\mathbf{x})g(\mathbf{x})\xi(d\mathbf{x}). \end{aligned}$$

In contrast to the decomposition of *IMSE* for *estimation* into positive summands, the *IPB* may be negative. However, $IPB + \int_T f^2(\mathbf{x})d\mathbf{x} \geq 0$.

In practice ξ must be discrete, with atoms consisting of integral multiples of n^{-1} at the design points. We adopt the viewpoint of *approximate* design theory and allow

ξ to be any probability measure on S. It then turns out that the optimal extrapolation designs are not discrete. In fact, to guarantee that either of $\sup_f IPB(f, w, \xi)$ or $\sup_g IPV(g, w, \xi)$ be finite, it is necessary that ξ have a density. This can be established by modifying the proof of Lemma 1 of Wiens (1992). A consequence is that the optimal extrapolation designs must be approximated to make them implementable. This can be carried out by placing the design points at an appropriate number of quantiles of ξ.

Let $k(\mathbf{x})$ be the density of ξ, and define $m(\mathbf{x}) = k(\mathbf{x})w(\mathbf{x})$. Without loss of generality, assume that the mean weight is $\int_S w(\mathbf{x})\xi(d\mathbf{x}) = 1$. Then $m(\mathbf{x})$ is also a density on S which for fixed weights satisfies

$$\int_S \frac{m(\mathbf{x})}{w(\mathbf{x})} d\mathbf{x} = 1. \tag{5}$$

From the definitions of $\mathbf{B}$ and $\mathbf{b}_{f,S}$ we see that $IPB(f, w, \xi)$ depends on (w, ξ) only through m and $IPV(g, w, \xi)$ through m and w. Hence, we can optimize over m and w subject to (5) rather than over w and k.

Given fixed $m(\mathbf{x})$ and $w(\mathbf{x})$, the 'max' parts of the minimax solutions are given by Theorem 1. Before stating this we define matrices $\mathbf{K} = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})m^2(\mathbf{x})d\mathbf{x}$ and $\mathbf{G} = \mathbf{K} - \mathbf{B}\mathbf{A}_S^{-1}\mathbf{B}$. We define λ_m to be the largest solution to $|\mathbf{G} - \lambda\mathbf{H}| = 0$ and let $\mathbf{a}_0$ be any vector satisfying $(\mathbf{G}\mathbf{H}^{-1}\mathbf{G} - \lambda_m\mathbf{G})\mathbf{a}_0 = \mathbf{0}$ and $\mathbf{a}_0^T\mathbf{G}\mathbf{a}_0 = 1$. Define also $l_m(\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\mathbf{H}^{-1}\mathbf{z}(\mathbf{x})$ and $\alpha_m = \int_S\{l_m(\mathbf{x})m^2(\mathbf{x})\}^{2/3}d\mathbf{x}$.

Theorem 1 *a) Maximum Integrated Prediction Bias is*

$$\sup_{f \in \mathcal{F}} IPB(f, w, \xi) = \eta_S^2 \{(\sqrt{\lambda_m} + r_{T,S})^2 - r_{T,S}^2\} \geq 0,$$

attained at

$$f_m(\mathbf{x}) = \begin{cases} \eta_S \mathbf{z}^T(\mathbf{x})\{m(\mathbf{x})I - \mathbf{A}_S^{-1}\mathbf{B}\}\mathbf{a}_0 & \mathbf{x} \in S, \\ -\eta_T \mathbf{z}^T(\mathbf{x})\mathbf{B}^{-1}\mathbf{G}\mathbf{a}_0/\sqrt{\lambda_m} & \mathbf{x} \in T. \end{cases}$$

b) Maximum Integrated Prediction Variance is

$$sup_g IPV(g, w, \xi) = \eta_S^2 \nu \Omega^{-1/2} \left[\int_S \{w(\mathbf{x}) l_m(\mathbf{x}) m(\mathbf{x})\}^2 d\mathbf{x} \right]^{1/2},$$

attained at $g_{m,w}(\mathbf{x}) \propto w(\mathbf{x}) l_m(\mathbf{x}) m(\mathbf{x})$.

c) Maximum Integrated Mean Squared Prediction Error is

$$\sup_{f,g} IMSPE(f, g, w, \xi) = \eta_S^2 \left\{ \left(\sqrt{\lambda_m} + r_{T,S}\right)^2 + \nu \Omega^{-1/2} \left[\int_S \{w(\mathbf{x}) l_m(\mathbf{x}) m(\mathbf{x})\}^2 d\mathbf{x} \right]^{1/2} \right\}.$$

Note that the least favourable contaminant is in fact *linear* (in $\mathbf{z}$) on T and that f_m also maximizes $IPB + \int_T f^2(\mathbf{x}) d\mathbf{x}$ (since $\int_T f_m^2(\mathbf{x}) d\mathbf{x} = \eta_T^2$).

We say that a design/weights pair (ξ, w) is *unbiased* if it satisfies

$$E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta} \text{ for all } f \in \mathcal{F},$$

so that $sup_f IPB(f, w, \xi) = 0$. The following theorem gives the minimax weights for fixed $m(\mathbf{x})$, and a necessary and sufficient condition for unbiasedness.

Theorem 2 *a) For fixed $m(\mathbf{x})$ the weights minimizing $sup_g IPV(g, w, \xi)$ subject to (5) are given by*

$$w_m(\mathbf{x}) = \alpha_m \{l_m^2(\mathbf{x}) m(\mathbf{x})\}^{-1/3} I\{m(\mathbf{x}) > 0\}.$$

Then $sup_g IPV(g, w_m, \xi) = \eta_S^2 \nu \Omega^{-1/2} \alpha_m^{3/2}$.

b) The pair (ξ, w) is unbiased if and only if $m(\mathbf{x}) \equiv \Omega$.

In view of Theorems 1 and 2, our problems in this article can be rewritten as follows:

P1) Find a density $m_*(\mathbf{x})$ which minimizes

$$\eta_S^{-2} \sup_{f,g} IMSPE(f, g, \mathbf{1}, \xi) = \left(\sqrt{\lambda_m} + r_{T,S}\right)^2 + \nu \Omega^{-1/2} \left[\int_S \{l_m(\mathbf{x}) m(\mathbf{x})\}^2 d\mathbf{x}\right]^{1/2}. \tag{6}$$

Then $k_*(\mathbf{x}) = m_*(\mathbf{x})$ is the optimal extrapolation design density for *OLS* estimation.

P2) Find a density $m_*(\mathbf{x})$ which minimizes

$$\eta_S^{-2} \sup_{f,g} IMSPE(f, g, w_m, \xi) = \left(\sqrt{\lambda_m} + r_{T,S}\right)^2 + \nu \Omega^{-1/2} \alpha_m^{3/2}.$$

Then the weights $w_*(\mathbf{x}) = m_*(\mathbf{x})/k_*(\mathbf{x})$ and the design density $k_*(\mathbf{x}) = \alpha_{m_*}^{-1} \{m_*^2(\mathbf{x}) l_{m_*}(\mathbf{x})\}^{2/3}$ are optimal for *WLS* estimation.

P3) Find weights $w_0(\mathbf{x}) \propto l_m(\mathbf{x})^{-2/3}$, satisfying (5) with $m(\mathbf{x}) \equiv \Omega$. Then $w_0(\mathbf{x})$ and the design density $k_0(\mathbf{x}) = \Omega / w_0(\mathbf{x})$ are optimal in that they minimize $\sup_{f,g} IMSPE(f, g, w, \xi)$, subject to the side condition of unbiasedness.

Note that we have multiplied the quantities to be minimized by η_S^{-2}; this is without loss of generality and makes our results dependent only on the parameters ν and $r_{T,S}$.

3 MINIMAX EXTRAPOLATION DESIGNS FOR OLS

For P1) and P2) we consider only multiple linear regression without interactions, i.e. $\mathbf{z}^T(\mathbf{x}) = (1, \mathbf{x}^T)$, with S being a q-dimensional sphere of unit radius centered at the origin. We take an annular extrapolation space: $T = \{\mathbf{x} \mid 1 < \|\mathbf{x}\| \leq \beta\}$. There being no reason to give preference to one coordinate of $\mathbf{x}$ over another, we restrict to densities $m(\mathbf{x})$ with identical, symmetric marginals. Then $\mathbf{A}_S = \Omega^{-1}(1 \oplus (q+2)^{-1}\mathbf{I}_q)$ where $\Omega = \Gamma(1 + q/2)/\pi^{q/2}$, and $\mathbf{B} = 1 \oplus \gamma\mathbf{I}_q$ where $\gamma := \int_S x_1^2 m(\mathbf{x}) d\mathbf{x}$. Define parameters

$$\tau_0 = \int_T d\mathbf{x}, \quad \kappa_0 = \Omega,$$

$$\tau_1 = \frac{\int_T x_1^2 d\mathbf{x}}{q\gamma^2}, \quad \kappa_1 = \Omega q(q+2)\gamma^2.$$

We calculate that

$$\int_T x_1^2 d\mathbf{x} = \frac{\beta^{q+2} - 1}{\Omega(q+2)},$$

yielding $\tau_i = (\beta^{q+2i} - 1)/\kappa_i$ for $i = 0, 1$, and that $\mathbf{A}_T = \tau_0 \oplus \int_T x_1^2 d\mathbf{x} \cdot \mathbf{I}_q$.

We find that $l_m(\mathbf{x}) = l(\|\mathbf{x}\|; \gamma)$, where $l(u; \gamma) := \tau_0 + \tau_1 q u^2$ depends on the design only through γ. The maximum eigenvalue λ_m in (6) is found to be $\lambda_m =$

$\max(\lambda_m^{(0)}, \lambda_m^{(1)})$, where

$$\lambda_m^{(i)} := \tau_i \left\{ \int_S ||\mathbf{x}||^{2i} m^2(\mathbf{x}) d\mathbf{x} - \kappa_i \right\}.$$

We must consider two cases. For $u \in [0,1]$ and $i = 0, 1$ define

$$h_i(u;\gamma) = \frac{a_i \nu (b_i + u^2)^+}{(1 + r_{T,S} c_i) u^{2i} + d_i \nu l^2(u;\gamma)},$$

where the constants $a_i = a_i(\gamma) > 0$, $b_i = b_i(\gamma)$, $c_i = c_i(\gamma) > 0$ and $d_i = d_i(\gamma) > 0$ satisfy

$$\int_0^1 \frac{qu^{q-1}}{\Omega} h_i(u;\gamma) du = 1, \tag{7}$$

$$\int_0^1 \frac{qu^{q-1}}{\Omega} u^2 h_i(u;\gamma) du = q\gamma, \tag{8}$$

$$c_i^2 \tau_i \left\{ \int_0^1 \frac{qu^{q-1}}{\Omega} u^{2i} h_i^2(u;\gamma) du - \kappa_i \right\} = 1, \tag{9}$$

$$2 d_i \tau_i \left\{ \int_0^1 l^2(u;\gamma) q u^{q-1} h_i^2(u;\gamma) du \right\}^{1/2} = 1. \tag{10}$$

We denote these by Case 0 ($i = 0$) and Case 1 ($i = 1$). It turns out that for fixed ν, Case 0 holds for small values of β, Case 1 for large values of β. The precise relationship between ν and β has not been determined.

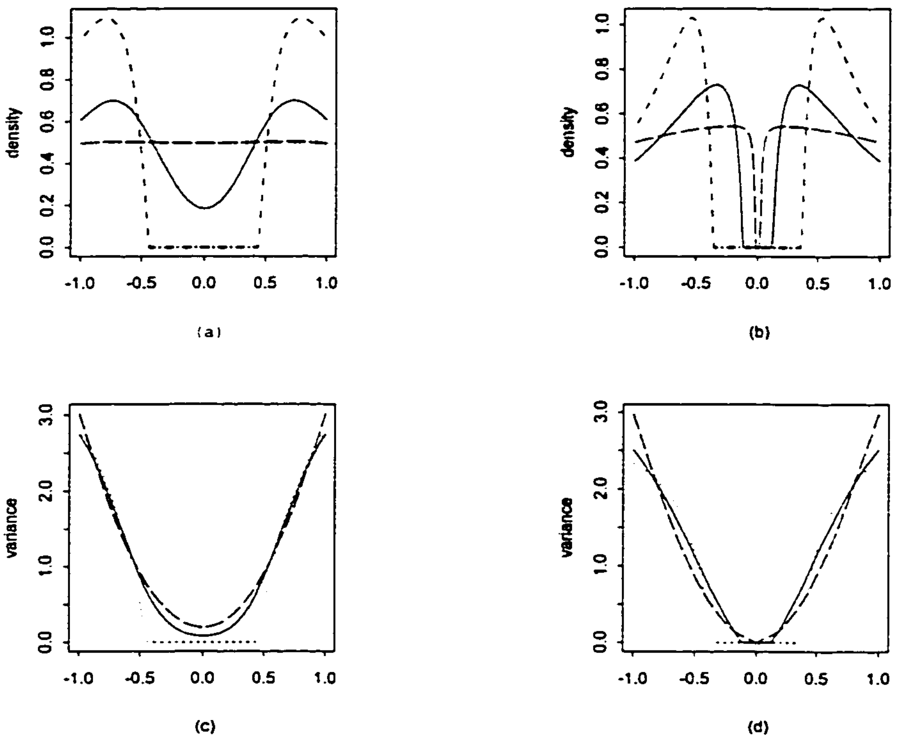
Theorem 3 *Minimax extrapolation designs for OLS. For $i = 0, 1$ define*

$$\gamma_i = argmin_{\gamma \geq 0} \left[\left\{ c_i(\gamma)^{-1} + r_{T,S} \right\}^2 + \frac{\nu}{2\Omega d_i(\gamma) \tau_i} \right]. \tag{11}$$

If the inequality

$$E_i\{ U^{2(1-i)} h_i(U;\gamma_i) \} \leq \frac{c_i(\gamma_i)^{-2}}{\tau_{1-i}} + \kappa_{1-i} \tag{12}$$

Figure 1: Optimal extrapolation design densities and least favourable variances for *OLS* and *SLR*. (a) Design densities, $\beta = 1.5$; (b) Design densities, $\beta = 5$; (c) Least favourable variances, $\beta = 1.5$; (d) Least favourable variances, $\beta = 5$. Each plot uses three values of ν: $\nu = .25$ (broken line), $\nu = 1$ (solid line), $\nu = 100$ (dotted line).



holds, where $E_i\{\cdot\}$ denotes expectation with respect to the density $(qu^{q-1}/\Omega)h_i(u;\gamma_i)$ and where κ_1 and τ_1 are evaluated at $\gamma = \gamma_1$, then the minimax (for OLS) extrapolation design density is

$$k_*(\mathbf{x}) = m_*(\mathbf{x}) = h_i(\|\mathbf{x}\|;\gamma_i).$$

Minmax IMSPE is

$$\sup_{f,g} IMSPE(f,g,w=\mathbf{1},\xi_*) = \eta_S^2\left[\ \left\{c_i(\gamma_i)^{-1} + r_{T,S}\right\}^2 + \frac{\nu}{2\Omega d_i(\gamma_i)\tau_i}\right]. \tag{13}$$

Remarks:

1. We sketch the proof of Theorem 3 for $i = 0$; that for $i = 1$ is similar. We first find m_0 minimizing (6) with $\lambda_m = \lambda_m^{(0)}$. Then if m_0 satisfies $\lambda_{m_0}^{(0)} \geq \lambda_{m_0}^{(1)}$, it is the required minimax density. For fixed γ and $\lambda_m = \lambda_m^{(0)}$ the loss (6) is a convex functional of m which remains fixed under orthogonal transformations of $\mathbf{x}$. By averaging over the orthogonal group we find that the minimizing m_0 is spherically symmetric. A standard variational argument shows that $m_0(\mathbf{x})$ is of the form $h_0(\|\mathbf{x}\|;\gamma)$ for appropriately chosen constants a_0-d_0. The integrand in (7) is the density of $U = \|\mathbf{x}\|$, equation (8) fixes $\gamma = E(U^2)/q$, equation (9) states that $c_0^{-2} = \lambda_m^{(0)}$ and equation (10) expresses the first order variational condition that h_0 be a stationary point. These equations allow (6) to be expressed as a function of γ alone; a further minimization over γ then results in (11). Condition (12) ensures that $\lambda_{m_0}^{(0)}$ $(= c_0(\gamma_0)^{-2}) \geq \lambda_{m_0}^{(1)}$.

Table 1. Constants for $m_*(\mathbf{x})$ of Theorem 3;

$q = 1$ (SLR) and $r_{T,S} = 1$.

β	ν	a	b	c	d	γ
1.5	.25*	15.26	14.52	110.56	.140	.334
	.5*	14.97	.203	4.61	.159	.417
	1*	11.57	.079	3.80	.163	.423
	10*	7.07	−.163	1.53	.184	.579
	100*	6.80	−.195	1.45	.185	.596
	∞*	6.77	−.198	1.44	.185	.597
5	.25†	7.01	−.000	2.15	4.04e−6	.334
	.5†	2.51	−.005	.771	4.35e−6	.339
	1†	1.48	−.015	.445	4.62e−6	.343
	10†	.983	−.096	.296	9.01e−6	.436
	100†	.981	−.132	.280	1.12e−5	.475
	∞†	.973	−.134	.277	1.13e−5	.476

*Case 0; †Case 1.

2. For the numerical work equations (10) (for $i = 0, 1$) were first eliminated by using them, in the presence of (7) to (9), to express d_i in terms of a_i-c_i:

$$d_i^{-1} = 4\Omega\tau_i^2 \left[a_i(b_i + q\gamma) - \frac{1 + r_{T,S}c_i}{\nu} \left\{ \left(c_i^2\tau_i\right)^{-1} + \kappa_i \right\} \right]. \tag{14}$$

The remaining equations were then solved. Finally (11) was minimized and (12) verified. See Table 1 for some numerical values of the constants in the case $q = 1$ - straight-line regression (SLR) - with $r_{T,S} = 1$. Figure 1 gives plots of the minimax extrapolation design densities for varying β and ν.

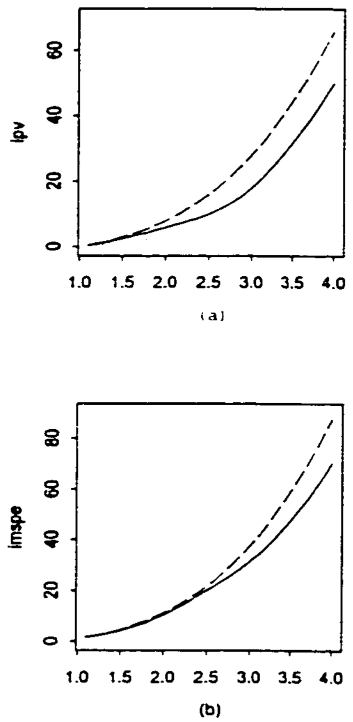
3. To implement these designs, we may use the fact that under the density $m_*(\mathbf{x})$, $\mathbf{x}/||\mathbf{x}||$ and $U = ||\mathbf{x}||$ are independently distributed, with $\mathbf{x}/||\mathbf{x}||$ being uniformly distributed over the surface of the unit sphere. A possible implementation is then as follows. Let H_* be the cumulative distribution function of U. Choose r_n design points uniformly distributed over each of the annuli $||\mathbf{x}|| = H_*^{-1}(i/[n/r_n])$, $i = 1, \ldots, [n/r_n]$, and $(n - r_n[n/r_n])$ points at the origin.

4. When the fitted model $E(Y|\mathbf{x}) = \boldsymbol{\theta}^T\mathbf{z}(\mathbf{x})$ is correct and the variances are homogeneous, the OLS estimate is unbiased and the loss is

$$IMSPE(\mathbf{0}, \mathbf{1}, \mathbf{1}, \xi) = IPV(\mathbf{1}, \mathbf{1}, \xi) = \eta_S^2\nu E\{l(||\mathbf{x}||; \gamma)\} = \eta_S^2\nu\left(\tau_0 + \frac{q}{\gamma}\int_T x_1^2 d\mathbf{x}\right),$$

where γ is the second moment of ξ. In Figure 2(a) we compare the loss for our minimax SLR design ξ_* with that of the two-point (± 1) design ξ_1, constructed

Figure 2: (a) Integrated Prediction Variance vs. β; (b) $sup_{f,g}IMSPE$ vs.β for three designs: ξ_* (minimax, $\nu = \Omega$; solid lines), ξ_1 (2-point; dotted line), ξ_2 (uniform; broken lines); all for *OLS* and *SLR*.



by Hoel and Levine (1964) under the assumption of an exactly correct fitted model, and of the continuous uniform design ξ_2.

When the model may contain response contamination and heteroscedastic errors, ξ_1 has $\sup_{f,g} IMSPE = \infty$. Figure 2(b) gives plots of $\sup_{f,g} IMSPE$ for the uniform design and for ξ_*. For the minimax design $\sup_{f,g} IMSPE$ is given by (13). For the uniform design Theorem 1c) gives

$$\sup_{f,g} IMSPE(f,g,\mathbf{1},\xi_2)$$

$$= \eta_S^2 \left\{ r_{T,S}^2 + \nu \left[\{\tau_0 + q^2(q+2)\gamma^2\tau_1\}^2 + \frac{4q}{q+4}\{q(q+2)\gamma^2\tau_1\}^2 \right]^{1/2} \right\}, \quad (15)$$

with τ_1 evaluated at γ. We have used $\nu = \Omega = .5$ and $\eta_S = \eta_T = 1$ in Figure 2. For this value of ν the minimax design is close to the uniform and the efficiencies relative to ξ_1, when the model is correct, are rather low. For larger values of ν these relative efficiencies are somewhat higher.

4 MINIMAX EXTRAPOLATION DESIGNS AND WEIGHTS FOR WLS

We consider the same multiple linear regression model, spherical design space and extrapolation space as in the previous section. We again consider two cases. For $u \in [0,1]$ and $i = 0,1$ define $h_i(u;\gamma)$ to be the (sole) real root of

$$\frac{1 + r_{T,S}c_i}{\nu} u^{2i} h_i(u;\gamma) + \left\{ \frac{l^2(u;\gamma)h_i(u;\gamma)}{4\Omega^2 d_i \tau_i^4} \right\}^{1/3} - a_i(b_i + u^2)^+ = 0,$$

i.e. $h_i^{1/3}(u;\gamma) = z^{1/3}(u) - \left[\nu\left\{l^2(u;\gamma)/\left(4\Omega^2 d_i \tau_i^4\right)\right\}^{1/3} / \left\{3(1+r_{T,S}c_i)u^{2i}\right\}\right] z^{-1/3}(u)$, where

$$z(u) = \frac{\nu}{2(1+r_{T,S}c_i)u^{2i}}\left[a_i(b_i+u^2)^+ + \sqrt{\{a_i(b_i+u^2)^+\}^2 + \frac{\nu l^2(u;\gamma)}{27\Omega^2 d_i \tau_i^4 (1+r_{T,S}c_i)u^{2i}}}\right].$$

The constants $a_i = a_i(\gamma) > 0$, $b_i = b_i(\gamma)$, $c_i = c_i(\gamma) > 0$ and $d_i = d_i(\gamma) > 0$ are determined by (7), (8), (9) and (14).

The following result is established in a manner similar to that used for Theorem 3.

Theorem 4 *Minimax extrapolation designs and weights for WLS. For $i = 0, 1$ define*

$$\gamma_i = argmin_{\gamma \geq 0}\left[\{c_i(\gamma)^{-1} + r_{T,S}\}^2 + \frac{\nu}{4\Omega d_i(\gamma)\tau_i}\right].$$

If the inequality (12) holds, then the minimax (for WLS) extrapolation design density $k_*(\mathbf{x})$ and weights $w_{m_*}(\mathbf{x})$ are given by

$$\begin{aligned} k_*(\mathbf{x}) &= \left\{4\Omega^{1/2} d_i(\gamma_i)\tau_i h_i^2(||\mathbf{x}||;\gamma_i) l(||\mathbf{x}||;\gamma_i)\right\}^{2/3}, \\ m_*(\mathbf{x}) &= h_i(||\mathbf{x}||;\gamma_i),\ w_*(\mathbf{x}) = m_*(\mathbf{x})/k_*(\mathbf{x}). \end{aligned}$$

Minmax IMSPE is

$$\sup_{f,g} IMSPE(f,g,w_*,\xi_*) = \eta_S^2\left[\left\{c_i(\gamma_i)^{-1} + r_{T,S}\right\}^2 + \frac{\nu}{4\Omega d_i(\gamma_i)\tau_i}\right]. \quad (16)$$

Table 2 gives some typical values of the constants and Figure 3 shows plots of the minimax design densities and weights, both for $q = 1$ and $r_{T,S} = 1$.

Figure 3: Optimal extrapolation design densities and minimax weights for *WLS* and *SLR*. (a) Design densities,$\beta = 1.5$; (b) Design densities, $\beta = 5$; (c) Minimax weights, $\beta = 1.5$; (d) Minimax weights, $\beta = 5$. Each plot uses three values of ν: $\nu = .25$ (broken line), $\nu = 1$ (solid line), $\nu = 100$ (dotted line).

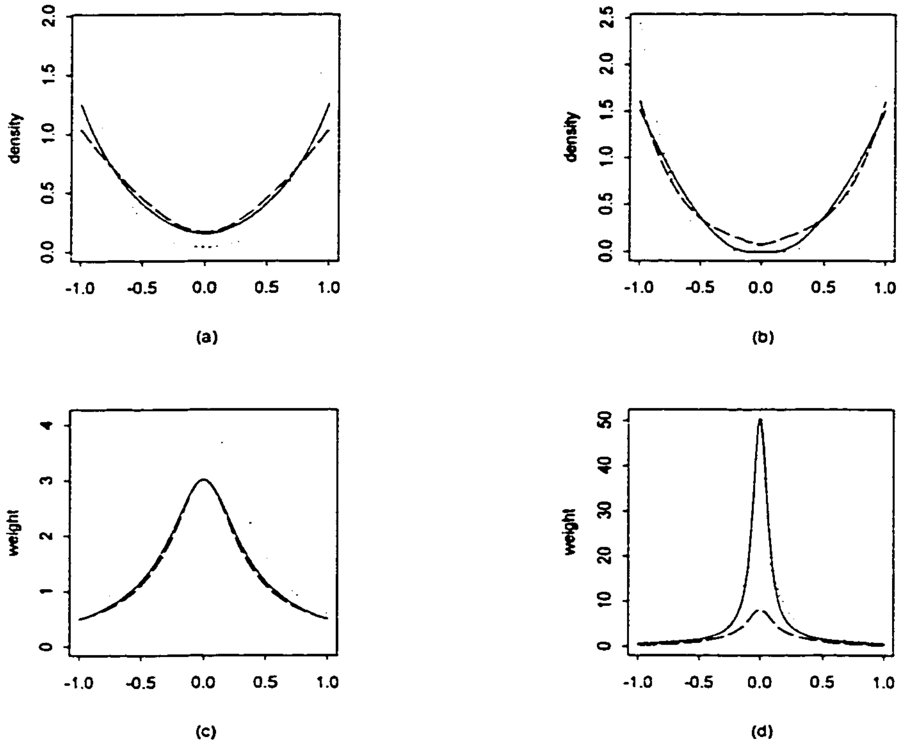


Table 2. Constants for $m_*(\mathbf{x})$ of Theorem 4;

$q = 1$ (SLR) and $r_{T,S} = 1$.

β	ν	a	b	c	d	γ
1.5	.25*	21.82	20.69	225.84	.096	.336
	.5*	11.69	7.46	84.91	.097	.336
	1*	11.15	.865	16.04	.099	.359
	10†	1.39	.122	2.51	.016	.485
	100†	1.18	.161	1.96	.017	.508
	∞†	1.16	.166	1.91	.017	.511
5	.25*	60.76	.233	3.35	2.95e−5	.349
	.5†	3.44	.005	.875	5.26e−6	.425
	1†	2.26	.005	.467	7.07e−6	.485
	10†	1.38	.027	.247	9.68e−6	.559
	100†	1.34	.027	.215	1.06e−5	.585
	∞†	1.32	.029	.215	1.06e−5	.585

*Case 0; †Case 1.

We have computed the efficiencies of ξ_* relative to other designs ξ, also symmetric with identical marginals and with second moment γ. When the fitted model $E(Y|\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}$ is correct and the variances are homogeneous, this relative efficiency is

$$rel(\xi) = \frac{IPV(g=\mathbf{1}, w=\mathbf{1}, \xi)}{IPV(g=\mathbf{1}, w=w_*, \xi_*)} = \frac{\tau_0 + \frac{q}{\gamma}\int_T x_1^2 d\mathbf{x}}{(4\sqrt{\Omega}d_i\tau_i)^{-2/3}\int_0^1 (qu^{q-1}/\Omega)h_i^{2/3}(u;\gamma_i)l^{1/3}(u;\gamma_i)du}.$$

Table 3 gives some representative values of $rel(\xi)$ for $\xi = \xi_1$, with all mass on the boundary of S, and $\xi = \xi_2$, the continuous uniform design. Also given are values of

$$re2(\xi) = \frac{\sup_{f,g} IMSPE(f, g, \mathbf{1}, \xi)}{\sup_{f,g} IMSPE(f, g, w_*, \xi_*)}, \tag{17}$$

which measures the efficiency of (ξ_*, w_*) relative to another design ξ, with constant weights, when the true response is only partially linear and the variances are heteroscedastic. The denominator of (17) is (16). For ξ_1 the numerator is ∞, for ξ_2 it is given by (??). As before, we take $\nu = \Omega$ and $r_{T,S} = 1$. The numbers in Table 3 show the appreciable gains to be enjoyed when ξ_* is employed in the presence of contamination and heteroscedasticity.

5 OPTIMAL UNBIASED EXTRAPOLATION DESIGNS

In this section we make no *apriori* restrictions (beyond assumption A)) on the design density, design space or extrapolation space. Note that if $m_0(\mathbf{x}) \equiv \Omega$, we have $\mathbf{B} = \Omega\mathbf{A}_S$. The following result is then an immediate consequence of Theorem 2.

Table 3. Relative efficiencies re1 (no contamination) and re2 (maximal contamination) of ξ_* of Theorem 4, with optimal weights w_* and $\nu = \Omega$, versus the design ξ_1 with all mass on $||\mathbf{x}|| = 1$ and the uniform design ξ_2, both with constant weights.

β	q	$re1(\xi_1)$	$re1(\xi_2)$	$re2(\xi_1)$	$re2(\xi_2)$
1.5	1*	.517	1.15	∞	1.27
	2*	.609	1.07	∞	1.16
	3*	.671	1.04	∞	1.10
5	1†	.495	1.40	∞	1.57
	2†	.567	1.11	∞	1.24
	3†	.649	1.07	∞	1.14

*Case 0; †Case 1.

Theorem 5 *The density $k_0(\mathbf{x})$ of the optimal extrapolation design measure ξ_0, and optimal weights $w_0(\mathbf{x})$, which minimize $\sup_{f,g} IMSPE(f,g,w,\xi)$ subject to $\sup_f IPB(f,w,\xi) = 0$, are given by*

$$k_0(\mathbf{x}) = \frac{\{\mathbf{z}^T(\mathbf{x})\mathbf{A}_S^{-1}\mathbf{A}_T\mathbf{A}_S^{-1}\mathbf{z}(\mathbf{x})\}^{2/3}}{\int_S\{\mathbf{z}^T(\mathbf{x})\mathbf{A}_S^{-1}\mathbf{A}_T\mathbf{A}_S^{-1}\mathbf{z}(\mathbf{x})\}^{2/3}d\mathbf{x}} \tag{18}$$

and $w_0(\mathbf{x}) = \Omega/k_0(\mathbf{x})$. Minmax IMSPE is

$$\sup_{f,g} IMSPE(f,g,w_0,\xi_0) = \eta_S^2\left\{r_{T,S}^2 + \nu\Omega^{-1/2}\left[\int_S\{\mathbf{z}^T(\mathbf{x})\mathbf{A}_S^{-1}\mathbf{A}_T\mathbf{A}_S^{-1}\mathbf{z}(\mathbf{x})\}^{2/3}d\mathbf{x}\right]^{3/2}\right\},$$

attained at $g_0(\mathbf{x}) = w_0(\mathbf{x})^{-1/2}$.

Example 1. Consider the multiple linear regression model, design space and extrapolation space of Sections 3 and 4. The optimal unbiased extrapolation design density is

$$k_0(\mathbf{x}) \propto \left\{1 + (q+2)\frac{\beta^{q+2}-1}{\beta^q-1}\|\mathbf{x}\|^2\right\}^{2/3}.$$

See Table 4 for relative efficiencies, with $\nu = \Omega$ and $r_{T,S} = 1$. These efficiencies are at most only marginally lower than those of (ξ_*, w_*) of Section 4.

Example 2. In this example there is insufficient structure to allow for a tractable treatment via P1) or P2), but (18) is easily evaluated. The regression response is as in Example 1, but the design space is the q-dimensional cube $S = [-1,1]^q$ and the extrapolation region is the possibly asymmetric perimeter $T = [-\beta_1, \beta_2]^q \backslash S$, where $\beta_1, \beta_2 \geq 1$. One of β_1, β_2 may be unity, for one-sided extrapolation. We find that

$$k_0(\mathbf{x}) \propto \left\{\left(1 + 3\mu_1\sum_{i=1}^q x_i\right)^2 + 9\left(\mu_2 - \frac{1}{3\mu_3^q}\right)\|\mathbf{x}\|^2 - \frac{1}{\mu_3^q}\right\}^{2/3},$$

Table 4. Relative efficiencies re1 (no contamination) and re2 (maximal contamination) of ξ_0 of Example 1, with optimal weights w_0 and $\nu = \Omega$, versus the design ξ_1 with all mass on $||\mathbf{x}|| = 1$ and the uniform design ξ_2, both with constant weights.

β	q	$re1(\xi_1)$	$re1(\xi_2)$	$re2(\xi_1)$	$re2(\xi_2)$
1.5	1	.514	1.14	∞	1.27
	2	.607	1.07	∞	1.16
	3	.671	1.04	∞	1.10
5	1	.441	1.22	∞	1.54
	2	.562	1.10	∞	1.23
	3	.640	1.06	∞	1.13

where $\mu_1 = (\beta_2 - \beta_1)/2$, $\mu_2 = (\beta_2 + \beta_1)^2/12$ and $\mu_3 = (\beta_2 + \beta_1)/2$. For symmetric extrapolation $\beta_1 = \beta_2$ and $\mu_1 = 0$.

For the dose response problem discussed in Section 1, if a linear approximation to $E(Y|x)$ is taken then the design density is

$$\frac{2}{1-a}k_0\left((x - \frac{1+a}{2})(\frac{2}{1-a})\right),\ a \leq x \leq 1;\ \mu_1 = \frac{a}{2},\ \mu_2 = \frac{a^2}{12}.$$

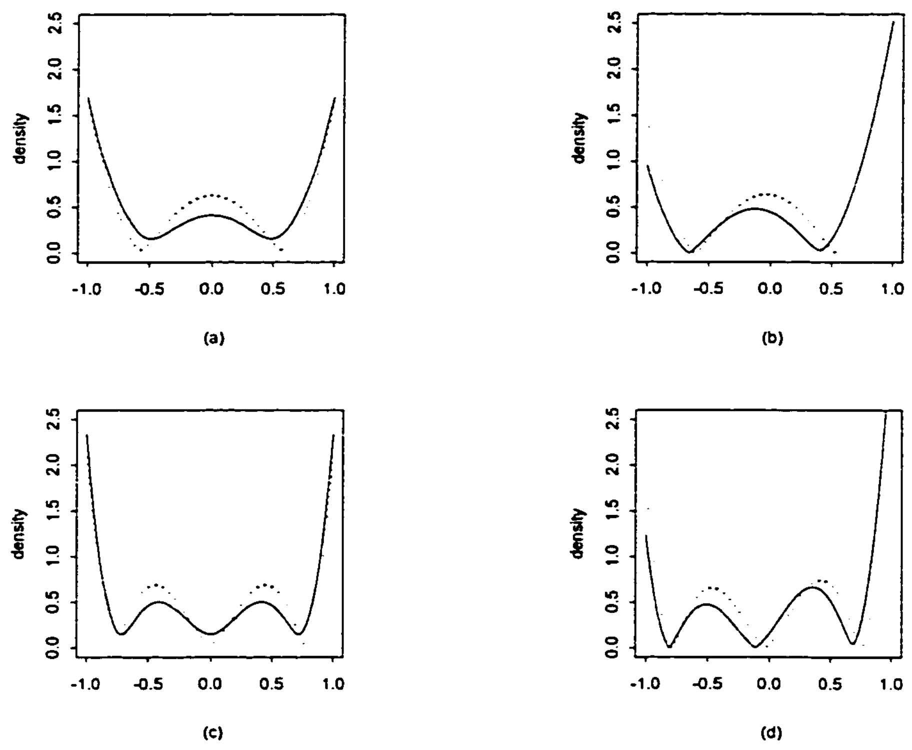
If instead a polynomial approximation is thought more appropriate then the design density is obtainable by applying a similar linear transformation to x in Example 3 below. In either case, a suitable implementation would consist of taking an appropriate number of replicates at each of a number of quantiles of $\xi_0(\cdot)$. The number of replicates vs. the number of quantiles would likely be determined by the requirements of the particular problem under investigation.

Example 3. Polynomial regression. Take $\mathbf{z}^T(x) = (1, x, \ldots, x^{p-1})$, corresponding to polynomial regression of degree $p-1$, on $S = [-1, 1]$. To evaluate (18) it is convenient to first express $\mathbf{z}(x)$ in terms of the Legendre polynomials. Denote by $P_m(x)$ the m^{th} degree Legendre polynomial, normalized by $\int_{-1}^{1} P_m^2(x)dx = (m + .5)^{-1}$. For instance $P_0(x) = 1,\ P_1(x) = x,\ P_2(x) = (3x^2 - 1)/2,\ P_3(x) = (5x^3 - 3x)/2$. We then find that

$$\mathbf{z}^T(x)\mathbf{A}_S^{-1}\mathbf{A}_T\mathbf{A}_S^{-1}\mathbf{z}(x) = \sum_{0 \leq i,j \leq p-1} \alpha_{ij} P_i(x) P_j(x),$$

where $\alpha_{ij} = (i + .5)(j + .5) \int_T P_i(x)P_j(x)dx$.

Figure 4: Optimal unbiased extrapolation design densities in biased quadratic and cubic polynomial models: (a) quadratic model, symmetric extrapolation region; (b) quadratic model, one-sided extrapolation region; (c) cubic model, symmetric extrapolation region; (d) cubic model, one-sided extrapolation region. Each plot uses two values of β: $\beta = 1.5$ (solid line), $\beta = 5$ (dotted line).



Denote the density (18) by $k_{p-1}(x;\beta)$. When T is symmetric, i.e. $T = [-\beta,\ \beta]\backslash S$, we find

$$
\begin{aligned}
k_2(x;\beta) &\propto \{5\beta^3(\beta+1)(3x^2-1)^2 - \beta(\beta+1)(5x^4-22x^2+5) + 4(1-2x^2+5x^4)\}^{2/3}, \\
k_3(x;\beta) &\propto \{175\beta^5(\beta+1)x^2(3-5x^2)^2 - 5\beta^3(\beta+1)(595x^6-963x^4+369x^2-9) + \\
&\quad 5\beta(\beta+1)(140x^6-177x^4+90x^2-9) + 4(175x^6-165x^4+45x^2+9)\}^{2/3}.
\end{aligned}
$$

When $T = [1,\ \beta]$ is one-sided, we find

$$
\begin{aligned}
k_2(x;\beta) &\propto \{5\beta^4(3x^2-1)^2 + 5\beta^3(3x-1)(x+1)(3x^2-1) - \beta^2(5x^4-30x^3-22x^2+ \\
&\quad 10x+5) - \beta(x+1)(5x^3-15x^2-7x+5) + \\
&\quad 2(10x^4+5x^3-4x^2+x+2)\}^{2/3}, \\
k_3(x;\beta) &\propto \{175\beta^6x^2(5x^2-3)^2 + 175\beta^5x(x+1)(5x^2-2x-1)(5x^2-3) - \\
&\quad 5\beta^4(595x^6-525x^5-963x^4+490x^3+369x^2-105x-9) - \\
&\quad 5\beta^3(x+1)(595x^5-385x^4-578x^3+258x^2+111x-9) + \\
&\quad 5\beta^2(140x^6-210x^5-177x^4+320x^3+90x^2-102x-9) + \\
&\quad 5\beta(x+1)(140x^5-35x^4-142x^3+48x^2+42x-9) + \\
&\quad 700x^6+525x^5-660x^4-470x^3+180x^2+165x+36\}^{2/3}.
\end{aligned}
$$

For both symmetric and one-sided extrapolation regions,

$$
k_{p-1}(x;\infty) = \frac{\left\{P_{p-1}^2(x)\right\}^{2/3}}{\int_{-1}^{1}\left\{P_{p-1}^2(x)\right\}^{2/3}dx}. \tag{19}
$$

Table 5. Relative efficiencies re1 (no contamination) and re2 (maximal contamination) of ξ_0 of Example 3, with optimal weights w_0, versus the uniform design ξ_2 and the p-point design ξ_3, both unweighted.

		$T = [-\beta, \beta] \setminus S$				$T = (1, \beta]$			
β	$p-1$	re1(ξ_2)	re1(ξ_3)	re2(ξ_2)	re2(ξ_3)	re1(ξ_2)	re1(ξ_3)	re2(ξ_2)	re2(ξ_3)
1.5	1	1.14	.514	1.27	∞	1.35	.380	1.42	∞
	2	1.23	.387	1.78	∞	1.51	.329	2.27	∞
	3	1.32	.389	2.17	∞	1.53	.340	2.89	∞
5	1	1.25	.441	1.54	∞	1.30	.424	1.61	∞
	2	1.29	.455	1.76	∞	1.32	.452	1.83	∞
	3	1.29	.471	1.86	∞	1.32	.469	1.93	∞

Remarks:

1. The limit in (19) is approached quite rapidly and we find that for moderately large β the symmetric and one-sided design densities are, for practical purposes, identical. In contrast (see Figure 4), for small β and one-sided extrapolation the optimal designs place appreciably more mass on that side of S closer to the extrapolation region.

   For large p one can combine (19) with the asymptotic expansion

$$(p-1)P_{p-1}^2(x) = \frac{1}{\pi\sqrt{1-x^2}} + \frac{U_{2(p-1)}(x)}{\pi} + O(p^{-1/2}), \tag{20}$$

   where $U_{2(p-1)}(x) = \sin((2p-1)\arccos x)/\sin(\arccos x)$ is Chebyshev's polynomial of the second kind. The right-hand side of (20) is a density whose first term is the limiting density of the D-optimal design, as $p \to \infty$.

2. The modes of $k_{p-1}(x;\infty)$ are at ± 1 and at the critical points of $P_{p-1}(x)$. Recall that these are precisely the support points of the classical D-optimal design, which minimizes estimation variance alone. Thus $k_{p-1}(x;\infty)$ may be viewed as a smoothed version of the D-optimal design. Efficiencies relative to the continuous uniform design ξ_2 and Hoel and Levine's (1964) extrapolation design ξ_3, with $\nu = \Omega$ and $r_{T,S} = 1$, are given in Table 5. Note however that both ξ_3 and the D-optimal design have only as many design points as parameters, so that there is no opportunity to assess the fit of the model.

Table 6. Comparative values of

$\eta_S^{-2} \max_{f,g} IMSPE$ for the

designs of Section 6.

ν	*P1*	*P2*	*P3*	*U*	*HL*
0.25	2.8	2.3	2.3	2.8	∞
0.5	4.6	3.6	3.6	4.6	∞
1	7.7	6.2	6.2	8.1	∞
10	57.1	48.4	53.1	72.5	∞
100	543	482	522	716	∞

6 COMPARISONS

We have carried out a simulation study for a regression model as at (1) - (3) with $\mathbf{z}(x) = (1, x)^T$ $(-1 \leq x \leq 1)$, normally distributed errors with $\sigma^2 = 1$, and sample size $n = 20$. We took $r_{T,S} = 1$ and $T = [-\beta, \beta] \backslash S$ with $\beta = 1.5$. Designs solving problems P1), P2) and P3) were constructed and compared with the continuous uniform design ("U") and the two-point design ("HL") of Hoel and Levine (1964). Table 6 gives some values of $\eta_S^{-2} \max_{f,g} IMSPE$. In preparing this table it was assumed that *P2* and *P3* would be used with the correspondingly robust weights. Note that by this measure of maximum loss, the unbiased design *P3* performs as well as *P2*, for moderate values of ν.

To compare the relative performances against particular types of departures, we then chose a quadratic response: $f(x) \propto P_2(x)$ with the normalization $\int_S f^2(x)dx = 1/5$, and the variance function $g(x) \propto (1+x^2)^\alpha$ $(\alpha = 0, 2)$, with the normalization $\int_S g^2(x)dx = \Omega^{-1}$. Designs *P1* and *P2* employed $\nu = \Omega = 0.5$. Note that then $\int_S f^2(x)dx = 2\eta_S^2$; this choice was made to further test the robustness of *P1* and *P2*. For the continuous designs the design points were placed at the quantiles $\xi^{-1}((i-1)/(n-1))$ $(i = 1, ..., n)$ of the design measures. When using *WLS*, the weights used for *P1* were generated from Theorem 4 in the same way as those for *P2*. The uniform design weights were generated from Theorem 5 in the same way as those for *P3*.

For *HL*, with 10 points at each of ± 1, weighting has no effect. The other design points and weights were as follows:

P1:	±.095	±.265	±.398	±.506	±.600	±.684	±.764	±.842	±.920	±1.00
weights:	2.49	1.71	1.24	.976	.811	.698	.612	.544	.487	.438
P2:	±.133	±.344	±.484	±.591	±.680	±.757	±.825	±.886	±.942	±1.00
weights:	2.62	1.58	1.15	.929	.792	.697	.628	.575	.532	.493
P3:	±.148	±.353	±.489	±.595	±.682	±.759	±.827	±.889	±.947	±1.00
weights:	2.59	1.57	1.15	.934	.800	.705	.636	.583	.539	.504
U:	±.053	±.158	±.263	±.368	±.474	±.579	±.684	±.789	±.895	±1.00
weights:	2.20	1.84	1.43	1.10	.867	.701	.581	.491	.421	.367

Table 7 gives values of $\eta_S^{-2}IPB$, $\eta_S^{-2}IPV$ and $\eta_S^{-2}IMSPE$ for both *OLS* and *WLS*

Table 7. *IPB*, *IPV* and *IMSPE* for the simulations of Section 6; heteroscedastic errors and contaminated response function. Values of *IPV* under homoscedasticity in parentheses.

	OLS			*WLS*		
Design	$\eta_S^{-2}IPB$	$\eta_S^{-2}IPV$	$\eta_S^{-2}IMSPE$	$\eta_S^{-2}IPB$	$\eta_S^{-2}IPV$	$\eta_S^{-2}IMSPE$
P1	−.22	.30 (.23)	1.09	.11	.27 (.27)	1.38
P2	−.33	.28 (.20)	.96	−.02	.25 (.25)	1.23
P3	−.33	.28 (.20)	.95	−.04	.25 (.24)	1.22
U	−.07	.35 (.26)	1.28	.29	.30 (.33)	1.59
HL	−.86	2.52 (1.29)	2.66	−.86	2.52 (1.29)	2.66

fits. All three robust designs performed substantially better than did *U* or *HL*; *P2* and *P3* in particular did well both with and without weights. Note however that when used without weights this good performance was attained at the cost of a substantial negative *IPB*. When used with the optimal weights, *P2* and *P3* virtually eliminated this bias. Design *P3* enjoys the additional advantage of requiring no particular assumptions on the design space or fitted response function.

Faced with data reflecting the departures modeled by these simulations, would a statistician see evidence of the inadequacy of the linear model? To answer this we fit a quadratic response $\theta_0 + \theta_1 x + \theta_2 x^2$ and carried out size .05 t-tests of $H_0 : \theta_2 = 0$ *vs.* $H_1 : \theta_2 \neq 0$. Both *OLS* and *WLS* fits were compared. The powers, based on 20,000 simulations, are presented in Table 8. The same 400,000 simulated normal errors were used in each of the four design cases. Note that for HL the quadratic model cannot be fitted and the power is zero.

A message to be gotten from the powers in Table 8 is that the, common and realistic, response departure used in these simulations is not likely to be detected, even when its parametric form is specified exactly by the alternative hypothesis. We view this as a powerful argument in favour of anticipating and addressing such departures at the design stage.

Table 8. Power of t-test of quadratic vs. linear response for homoscedastic ($\alpha = 0$) and heteroscedastic ($\alpha = 2$) errors.

	OLS		*WLS*	
Design	$\alpha = 0$	$\alpha = 2$	$\alpha = 0$	$\alpha = 2$
P1	0.29	0.31	0.29	0.33
P2	0.29	0.28	0.32	0.33
P3	0.29	0.27	0.32	0.32
U	0.31	0.37	0.20	0.29

7 CONCLUSIONS AND GUIDELINES

We have given methods of designing for regression extrapolation, in the face of model uncertainties and possible heteroscedasticity, under a number of optimality criteria. The results tend to be somewhat complex and in some cases require extensive numerical work prior to implementation. They do however admit a number of informal and heuristic guidelines.

- In general, and as one would expect, the experimenter should place relatively more design points closer to the boundary between the design space S and the extrapolation space T, either as the volume of T increases relative to that of S, or as the emphasis on variance minimization versus bias minimization increases (as expressed by increasing values of ν).

- Notwithstanding the previous point, relative to designs for variance minimization alone the designs of this article are substantially more uniform, with mass spread throughout S rather than only at extreme points near T. This allows both for bias minimization and for the testing of alternative models.

- The unbiased designs of Section 5 are numerically less demanding than those of the preceding sections, although not completely without computational requirements. In line with (18) the general prescription is for the designer to place mass at points $\mathbf{x}$ proportional to values of $t(\mathbf{x}) := \{\mathbf{z}^T(\mathbf{x})\mathbf{A}_S^{-1}\mathbf{A}_T\mathbf{A}_S^{-1}\mathbf{z}(\mathbf{x})\}^{2/3}$; the appropriate regression weights are then inversely proportional to this quantity. This requires a study of $t(\mathbf{x})$ for the particular design and extrapolation spaces under consideration. Some intuition can be gained from the explicit expressions in Examples 1 and 2; the latter in particular illustrates the manner in which the relative magnitude of $t(\mathbf{x})$ varies as T changes. As in Example 3, it can be convenient to transform to orthogonal regressors, so that $\mathbf{A}_S$ becomes a diagonal matrix.

A relevant problem concerns the manner in which the desired number n of observations is to be apportioned between design sites and replicates. We have recommended placing the former at quantiles of the optimal design densities; the determination of the number of such quantiles is the subject of further research.

APPENDIX: DERIVATIONS

Proof of Theorem 1: *a):* First note that we can assume that the inequalities in (2) are in fact equalities. For, if $f \in \mathcal{F}$ is such that $\int_S f^2(\mathbf{x})d\mathbf{x} < \eta_S^2$ or $\int_T f^2(\mathbf{x})d\mathbf{x} < \eta_T^2$, then we define a function $cf \in \mathcal{F}$ as being $c_S f$ on S and $c_T f$ on T, where $|c_S| \geq 1$, $|c_T| \geq 1$ and the sign of $c_T c_S$ is chosen so that $-2\mathbf{b}_{cf,T}^T\mathbf{B}^{-1}\mathbf{b}_{cf,S} = -2c_T c_S \mathbf{b}_{f,T}^T\mathbf{B}^{-1}\mathbf{b}_{f,S} \geq -2\mathbf{b}_{f,T}^T B^{-1}\mathbf{b}_{f,S}$. Then $IPB(cf,\xi) \geq IPB(f,\xi)$. Hence it is sufficient to evaluate the maximum value of $IPB(f,\xi)$ under the conditions $\int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})d\mathbf{x} = \mathbf{0}$, $\int_S f^2(\mathbf{x})d\mathbf{x} = \eta_S^2$, $\int_T f^2(\mathbf{x})d\mathbf{x} = \eta_T^2$.

Note that

$$\mathbf{G} = \int_S [\{m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}_S^{-1}\}\mathbf{z}(\mathbf{x})][\{m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}_S^{-1}\}\mathbf{z}(\mathbf{x})]^T dx \geq \mathbf{0}. \tag{A.1}$$

We temporarily assume that $\mathbf{G}$ is positive definite. Given any $f \in \mathcal{F}$, define

$$h_f(x) = \begin{cases} s_f \mathbf{z}^T(\mathbf{x})\{m(\mathbf{x})\mathbf{I} - \mathbf{A}_S^{-1}\mathbf{B}\}\mathbf{H}^{-1}\mathbf{b}_{f,S}, & x \in S, \\ t_f \mathbf{z}^T(\mathbf{x})\mathbf{B}^{-1}\mathbf{b}_{f,S}, & x \in T, \end{cases}$$

with $s_f^2 = \eta_S^2/\mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S}$, $t_f^2 = \eta_T^2/\mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{b}_{f,S}$ and $s_f = \pm\sqrt{s_f^2}$, $t_f = \pm\sqrt{t_f^2}$ chosen so that $\mathbf{b}_{h_f,T}^T\mathbf{B}^{-1}\mathbf{b}_{h_f,S} \leq 0$. Then we claim that *i)* $h_f(x) \in \mathcal{F}$ and *ii)* $IPB(h_f,\xi) \geq IPB(f,\xi)$. The verification of *i)* is straightforward. For *ii)* we note that $\mathbf{b}_{h_f,T} = t_f \mathbf{A}_T \mathbf{B}^{-1}\mathbf{b}_{f,S}$, that $\mathbf{b}_{h_f,S} = s_f \mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S}$, and that

$$\int_S f(\mathbf{x})h_f(\mathbf{x})d\mathbf{x} = s_f \mathbf{b}_{f,S}^T \mathbf{H}_{f,S}^{-1}\mathbf{b}_{f,S}, \quad \int_T f(\mathbf{x})h_f(\mathbf{x})d\mathbf{x} = t_f \mathbf{b}_{f,S}^T \mathbf{B}^{-1}\mathbf{b}_{f,T}. \qquad \text{(A.2)}$$

Evaluating (4) gives $IPB(h_f,\xi) = s_f^2 \mathbf{b}_{f,S}^T \mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S} + 2|s_f||t_f|\mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S}$. By the first equality of (A.2) and the Cauchy-Schwarz inequality, we have $s_f^2(\mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{b}_{f,S})^2 \leq \int_S f^2(\mathbf{x})d\mathbf{x}\int_S h_f^2(\mathbf{x})d\mathbf{x} \leq \eta_S^4$, so that the definition of s_f gives $\eta_S^2 \geq (\mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{b}_{f,S})^2/\mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S}$. Similarly, $\eta_T^2 \geq (\mathbf{b}_{f,S}^T\mathbf{B}^{-1}\mathbf{b}_{f,T})^2/\mathbf{b}_{f,S}\mathbf{H}^{-1}\mathbf{b}_{f,S}$. Hence

$$\begin{aligned} IPB(h_f,\xi) &= \frac{\eta_S^2(\mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S})}{\mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S}} + \frac{2\eta_S\eta_T(\mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S})}{\sqrt{(\mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S})(\mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{b}_{f,S})}} \\ &\geq \frac{(\mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{b}_{f,S})^2}{(\mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S})^2}\mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S} + 2|\mathbf{b}_{f,S}^T\mathbf{B}^{-1}\mathbf{b}_{f,T}|, \end{aligned}$$

and so $IPB(h_f,\xi) \geq IPB(f,\xi)$ if $(\mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{b}_{f,S})(\mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S}) \geq (\mathbf{b}_{f,S}^T\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S})^2$, an inequality whose verification is again straightforward.

We can now restrict to $f \in \mathcal{F}$ of the same form as h_f, i.e.

$$f(\mathbf{x};\mathbf{a},\mathbf{c}) = \begin{cases} \mathbf{z}^T(\mathbf{x})\{m(\mathbf{x})\mathbf{I} - \mathbf{A}_S^{-1}\mathbf{B}\}\mathbf{a} & x \in S, \\ \mathbf{z}^T(\mathbf{x})\mathbf{B}^{-1}\mathbf{c} & x \in T, \end{cases}$$

where $\mathbf{a}$ and $\mathbf{c}$ satisfy $\eta_S^2 = \int_S f^2(\mathbf{x};\mathbf{a},\mathbf{c})d\mathbf{x} = \mathbf{a}^T\mathbf{G}\mathbf{a}$, $\eta_T^2 = \int_T f^2(\mathbf{x};\mathbf{a},\mathbf{c})d\mathbf{x} = \mathbf{c}^T\mathbf{H}^{-1}\mathbf{c}$. Subject to these conditions we are to maximize $IPB(f,\xi) = \mathbf{a}^T\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{a} - 2\mathbf{c}^T\mathbf{H}^{-1}\mathbf{G}\mathbf{a}$. The maximizing $\mathbf{c}$ is $\mathbf{c} = -\eta_T\mathbf{G}\mathbf{a}/||\mathbf{H}^{-1/2}\mathbf{G}\mathbf{a}||$ and then $IPB(f,\xi) = (\sqrt{\mathbf{a}^T\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{a}} + \eta_T)^2 - \eta_T^2$. With $\mathbf{a}_0 = \mathbf{a}/\eta_S$, we are then to maximize $\mathbf{a}_0^T\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{a}_0$ subject to $\mathbf{a}_0^T\mathbf{G}\mathbf{a}_0 = 1$. This is a standard eigenvalue problem. If λ_m is the largest solution to $|\mathbf{G}\mathbf{H}^{-1}\mathbf{G} - \lambda\mathbf{G}| = 0$, i.e. $|\mathbf{G} - \lambda\mathbf{H}| = 0$, then the maximizing $\mathbf{a}_0$ is a solution to $(\mathbf{G}\mathbf{H}^{-1}\mathbf{G} - \lambda_m\mathbf{G})\mathbf{a}_0 = \mathbf{0}$, normalized to satisfy $\mathbf{a}_0^T\mathbf{G}\mathbf{a}_0 = 1$. A final evaluation of $IPB(f, \xi)$ now completes the proof of *a)* when $\mathbf{G} > \mathbf{0}$.

If the design density $m(\mathbf{x})$ is such that $\mathbf{G} = \mathbf{G}(m) \geq \mathbf{0}$ but $|\mathbf{G}| = 0$, we proceed as follows. Take any density $m_1(\mathbf{x})$ for which the corresponding matrix $\mathbf{G}(m_1) > \mathbf{0}$. Put $m_t(\mathbf{x}) = (1-t)m(\mathbf{x}) + tm_1(\mathbf{x})$ and define $p(t) = |\mathbf{G}(m_t)|$. Then $p(t)$ is a polynomial in $t \in [0,1]$ with $p(0) = 0$ and $p(1) > 0$, so that $p(t)$ is non-constant and non-negative on $[0,1]$. Thus $p(t) > 0$ for all sufficiently small $t > 0$. Now apply *a)* of the theorem to $\mathbf{G}(m_t)$ and let $t \to 0$, to see that the result holds in the general case.

b): By the Cauchy-Schwarz inequality we have

$$\int_S w(\mathbf{x})g(\mathbf{x})l_m(\mathbf{x})m(\mathbf{x})d\mathbf{x} \leq \left[\int_S \{w(\mathbf{x})l_m(\mathbf{x})m(\mathbf{x})\}^2 d\mathbf{x}\right]^{1/2} \left\{\int_S g^2(\mathbf{x})d\mathbf{x}\right\}^{1/2}$$

and *b)* follows. Part *c)* follows from *a)* and *b)*. □

Proof of Theorem 2: Part *a)* is a straightforward variational problem. For *b)*, note that by Theorem 1a) and (A.1) we have

$$\sup_f IPB(f, w, \xi) = 0 \Longleftrightarrow \lambda_m = 0 \Longleftrightarrow \mathbf{G} = \mathbf{0} \Longleftrightarrow (m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}_S^{-1})\mathbf{z}(\mathbf{x}) = 0 \ a.e.$$

Thus $m(\mathbf{x})$ is an eigenvalue of $\mathbf{B}\mathbf{A}_S^{-1}$ if $\mathbf{z}(\mathbf{x}) \neq \mathbf{0}$, so that on $S_0 := \{\mathbf{x} \in S : \mathbf{z}(\mathbf{x}) \neq \mathbf{0}\}$, $m(\mathbf{x})$ can assume at most p distinct values. Decompose S_0 as $S_0 = \bigcup_{i=1}^s S_i$, with $s \leq p$ and $m(\mathbf{x}) \equiv \alpha_i$ on S_i. For any S_i with positive Lebesgue measure the relationship $(\alpha_i\mathbf{I} - \mathbf{B}\mathbf{A}_S^{-1})\mathbf{z}(\mathbf{x}) \equiv \mathbf{0}$, together with assumption A), forces $\alpha_i\mathbf{I} = \mathbf{B}\mathbf{A}_S^{-1}$, so that at most one set S_i can have positive measure. Thus $m(\mathbf{x})$ is almost everywhere constant on S_0, hence on S itself since, again by A), $S \setminus S_0$ is of measure zero. □

ACKNOWLEDGEMENTS

We are grateful for the thorough review, by an anonymous referee, of a previous version of this paper, and for the helpful comments of this referee and the Associate Editor and Editor.

REFERENCES

Box, G.E.P., and Draper, N.R. (1959). A Basis for the Selection of a Response Surface Design. *Journal of American Statistical Association,* 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 Royal Statistical Society, Ser. B*, 35, 268 – 276.

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

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., and Wolfowitz, J. (1964a). Optimum Extrapolation Designs, I. *Ann. Inst. Statist. Math.*, 16, 79 – 108.

Kiefer, J., and Wolfowitz, J. (1964b). Optimum Extrapolation Designs II. *Ann. Inst. Statist. Math.*, 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 Mathematics*, 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.

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

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.

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

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

Wong, W.-K. (1992). A Unified Approach to the Construction of Minimax Designs. *Biometrika*, 79, 611 – 619.

Wong, W.-K., and Cook, R.D. (1993). Heteroscedastic G-optimal Designs. *Journal of Royal Statistical Society, Ser. B*, 55, 871 – 880.

Chapter Four

Robust Extrapolation Designs for Approximate Polynomial Models[1]

Abstract

In this article we consider the construction of optimal designs for extrapolation in the polynomial regression setup, allowing for imprecision in the specification of the response function. We adopt a minimax approach, which determines an optimal design to minimize the maximum value of the Integrated Mean Squared Prediction Error (IMSPE), with the maximum being evaluated over the departures from the model. It turns out that in straight-line and quadratic regression, the minimax extrapolation designs are the same as the Minimum Integrated Variance (MIV) extrapolation designs, which assume that the fitted models are correct. But for cubic or higher degree polynomial regression, numerical comparisons show that the robust extrapolation designs perform better than the MIV extrapolation designs and the uniform design.

[1]Revision requested by the Journal of Statistical Planning and Inference.

1 Introduction

In this article we are concerned with the univariate polynomial regression setup, with a possible contaminant term included in the model. Specifically, by polynomial regression, we mean that an experiment is performed whose outcome has the form

$$Y(x_i) = \sum_{j=0}^{p} \beta_j x_i^j + \varepsilon_i, \quad i = 1, 2, ..., n, \tag{1}$$

where $x_i \in S \subset \mathcal{R}^1$ and $\{\varepsilon_i\}_{i=1}^n$ are uncorrelated random variables with mean 0 and variance σ^2. The coefficients, $\boldsymbol{\beta} = (\beta_0, ..., \beta_p)^T$, are unknown. If the model (1) exactly describes the character of the experimental outcome, then the Least Squares (LS) estimate $\hat{\boldsymbol{\beta}}$ is unbiased. Hence, for any region T ($T \cap S = \phi$), we predict $Y(x)$, $x \in T$, by $\hat{Y}(x) = \sum_{j=0}^{p} \hat{\beta}_j x^j$. The extrapolation design problem, considered in this article, is to choose the experimental points x_i in an optimal way.

As described by Kiefer (1959), an experimental design is a discrete probability measure ξ on S. That is, if n observations $\{x_i\}_{i=1}^n$ are to be taken, then ξ is the empirical distribution function of $\{x_1, ..., x_n\}$. These x_i do not need to be distinct. Suppose that ξ puts mass p_v on x_v where the $np_v = n_v$ are integers and $v = 1, ..., r$, where r represents the number of distinct design points. Then the experimenter takes n uncorrelated observations, with n_v observations at x_v.

Let $\mathbf{z}(x) = (1, x, ..., x^p)^T$. If the model (1) is correct, then the traditional measures of design optimality are usually based on the information matrix

$$\mathbf{M} = \int_S \mathbf{z}(x)\mathbf{z}^T(x)d\xi.$$

When constructing the optimal design for extrapolation to one point $\{t\}$, Hoel and Levine (1964) seek the design which minimizes the variance of the LS estimate of $Y(t)$: $d(t,\xi) = \mathbf{z}^T(t)\mathbf{M}^{-1}\mathbf{z}(t)$. However, as King and Wong (1998) point out, in some situations the point or points which are interesting to predict are not known before the experiment is performed. For this reason, we are concerned with constructing the optimal design for predicting the response surface over any region T. The I-optimal (or Q-optimal) design for extrapolation minimizes the integrated prediction variance: $\int_T \mathbf{z}^T(x)\mathbf{M}^{-1}\mathbf{z}(x)dx$.

But unfortunately, as mentioned by Box and Draper (1959) and followed by Draper and Herzberg (1973), Huber (1975, 1981), Wiens(1992), Fang and Wiens (1999) and others, using extrapolation designs becomes too risky in the situations when the "true" regression response function is only approximated by a polynomial. For such reasons we shall introduce a bias term in the linear model:

$$Y(x_i) = \sum_{j=0}^{p} \beta_j x_i^j + x^{p+1} f(x) + \varepsilon_i, \quad i = 1, 2, ..., n, \tag{2}$$

where $f(x)$ is an unknown function from some class $\mathcal{F}$. A motivation for considering this model is to use a Taylor expansion as an approximation to the "true" regression response function. The contamination function can be treated as the remainder of the

Taylor expansion. One estimates $\boldsymbol{\beta}$ but not f, leading to biased estimation of $E[Y(x)]$ and consequently to biased predictions. This requires that one should consider the bias term as well as the variance term when constructing the design. Therefore, as an optimality criterion we take an analogue of the classical notion of I-optimality: the maximum, over f, of the Integrated Mean Squared Prediction Error (*IMSPE*) of the predicted value $\hat{Y}(x)$, with the integration being over the extrapolation region T,

$$IMSPE(f,\xi) = \int_T E[\hat{Y}(x) - E(Y(x))]^2 dx,$$

is to be minimized by an appropriate choice of design.

For estimation, Liu and Wiens (1997) construct the optimal robust designs for the same model as (2). It turns out that these designs are supported on $(p+1)$ points. Pesotchinsky (1982) obtained minimax designs under departures similar to those in (2), with the fitted response surface being a hyperplane. No work has been done on optimal robust designs for extrapolation under model (2). In this paper, we attempt to fill this gap in the literature. We give the robust designs for extrapolation for linear ($p = 1$), quadratic ($p = 2$) and cubic ($p = 3$) regression. One can similarly obtain robust extrapolation designs when $p > 4$ and p is fixed.

2 Preliminaries

For the regression model (2) described in Section 1, we assume the design interval $S = [-1, 1]$ and the extrapolation region $T = [-t, t]\backslash S$, where $t > 1$. We shall assume that the contamination function f is unknown but satisfies the following conditions:

A1). The function $f(x)$ is continuous, and $|f(x)| \leq \phi(x)$ on $[-t, t]$, where $\phi(x)$ is a continuous, even function which is positive when $x \neq 0$.

A2). The function $l(x) := x\phi(\sqrt{x})$ is a convex function of $x \in [0, t]$, with $l(0) = 0$.

Let $\mathcal{F}$ be the class of functions f which have the above two properties. The continuity of f at $x = 0$ is required in order that the true parameter $\boldsymbol{\beta}$ in (2) be uniquely defined. In fact, if $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ satisfy

$$\boldsymbol{\beta}_1^T \mathbf{z}(x) + x^{p+1} f_1(x) = \boldsymbol{\beta}_2^T \mathbf{z}(x) + x^{p+1} f_2(x),$$

then

$$|(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)^T \mathbf{z}(x)| \leq 2|x|^{p+1}\phi(x). \tag{3}$$

Let $x \longrightarrow 0$, we have $|\beta_{10} - \beta_{20}| \leq 0$, that is, $\beta_{10} = \beta_{20}$. Hence, (3) can be rewritten as

$$\left|\sum_{j=1}^{p}(\beta_{1j} - \beta_{2j})x^{j-1}\right| \leq 2|x|^p\phi(x).$$

Let $x \longrightarrow 0$, we have $\beta_{11} = \beta_{21}$. By repeating the above process, we show that $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.

A1) and A2) together imply the following lemma which is useful for the construction of designs.

Lemma 1 *(a). The function $l(x)$ is monotone increasing on $[0, t]$.*

(b). The function $g_1(x) := l(x^2) = x^2\phi(x)$ is convex for $x \in [0, t]$; $g_2(x) := x\phi(x)$ is monotone increasing on $[0, t]$.

Proof. (a) is obvious. The first part of (b) follows the fact that the composition of positive, increasing, convex functions is convex. For the second part, let $0 \leq x_1 \leq x_2 \leq t$. The convexity of $g_1(x)$ implies that

$$\frac{x_1^2\phi(x_1)}{x_1} \leq \frac{x_2^2\phi(x_2)}{x_2},$$

which implies the monotonicity of $g_2(x)$. ■

We propose to estimate $\boldsymbol{\beta}$ by Least Squares. Define matrices

$$\mathbf{A} = \int_T \mathbf{z}(x)\mathbf{z}^T(x)dx, \quad \mathbf{B}_\xi = \int_S \mathbf{z}(x)\mathbf{z}^T(x)d\xi(x), \quad \mathbf{b}_{f,\xi} = \int_S \mathbf{z}(x)x^{p+1}f(x)d\xi(x).$$

To avoid trivialities, assume that $\mathbf{B}_\xi$ is nonsingular. Then the LS estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = \mathbf{B}_\xi^{-1}\int_S \mathbf{z}(x)y(x)d\xi(x)$, with bias vector and covariance matrix,

$$E[\hat{\boldsymbol{\beta}}] - \boldsymbol{\beta} = \mathbf{B}_\xi^{-1}\mathbf{b}_{f,\xi}, \quad COV[\hat{\boldsymbol{\beta}}] = \frac{\sigma^2}{n}\mathbf{B}_\xi^{-1}.$$

We predict $E[Y|x]$ for $x \in T$ by $\hat{Y}(x) = \hat{\boldsymbol{\beta}}^T\mathbf{z}(x)$ and consider the resulting Integrated Mean Squared Prediction Error. The *IMSPE* splits into terms due to Integrated Prediction Variance (*IPV*) and Integrated Prediction Bias (*IPB*):

$$\begin{aligned} IMSPE(f,\xi) &= \int_T E[\hat{Y}(x) - E(Y|x)]^2 dx \\ &= IPV(\xi) + IPB(f,\xi), \end{aligned}$$

where

$$IPV(\xi) = \int_T \mathbf{z}^T(x)COV[\hat{\boldsymbol{\beta}}]\mathbf{z}(x)dx = \frac{\sigma^2}{n}tr(\mathbf{B}_\xi^{-1}\mathbf{A})$$

$$
\begin{aligned}
IPB(f,\xi) &= \int_T [\mathbf{z}^T(x)(E[\hat{\boldsymbol{\beta}}] - \boldsymbol{\beta}) - x^{p+1}f(x)]^2 dx \\
&= \int_T [\mathbf{z}^T(x)\mathbf{B}_{\boldsymbol{\xi}}^{-1}\mathbf{b}_{f,\xi} - x^{p+1}f(x)]^2 dx.
\end{aligned}
$$

Our goal is to find an optimal design ξ_0 which minimizes $max_{f\in\mathcal{F}}IMSPE(f,\xi)$. In practice, ξ must have jumps consisting of integral multiples of n^{-1} at each design point. For mathematical convenience, we shall drop this restriction and allow $n_v = np_v$ be any positive real number if $\xi(x_v) = p_v$. Therefore we shall adopt the *approximate* design theory. That is, if n_v is not an integer, it would be necessary to choose an integer close to n_v for the actual number of observations at x_v. See Pukelsheim (1993).

The loss $IMSPE(f,\xi)$ depends on f only through $IPB(f,\xi)$. To evaluate the maximum of $IMSPE(f,\xi)$ over $f \in \mathcal{F}$, define $\mathbf{b}_i(\xi) = \int_S \mathbf{z}(x)x^{p+1}(signx)^i\phi(x)d\xi$ and $K_i(\xi) = \int_T[\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_i(\xi) - x^{p+1}(signx)^i\phi(x)]^2dx$. Then $K_i(\xi) = IPB((signx)^i\phi(x),\xi)$, $i = 1, 2$.

Theorem 2 *For any ξ, we have*

$$
max_{f\in\mathcal{F}}IPB(f,\xi) = \begin{cases} K_2(\xi) & \text{if } \phi(0) > 0, \\ max\{K_1(\xi), K_2(\xi)\} & \text{if } \phi(0) = 0. \end{cases}
$$

Proof. It is obvious that $\mathcal{F}$ is a convex set. In light of the fact that the functional $H(f) := \mathbf{z}^T(x)\mathbf{B}_{\boldsymbol{\xi}}^{-1}\mathbf{b}_{f,\xi} - x^{p+1}f(x)$ is linear, $IPB(f,\xi)$ is convex on $\mathcal{F}$. Thus, to maximize $IPB(f,\xi)$ we only need to consider the extremal f_0 which satisfies $|f_0(x)| \equiv$

$\phi(x)$. In fact, if there exists x_1 such that $|f_0(x_1)| < \phi(x_1)$, then we can construct two functions $f_1, f_2 \in \mathcal{F}$ such that $f_0(x) = \frac{1}{2}f_1(x)+\frac{1}{2}f_2(x)$. This implies that $IPB(f_0, \xi) \leq max\{IPB(f_1, \xi), IPB(f_2, \xi)\}$, so that f_0 can not be the maximizer.

In order that such an extremal $f_0(x)$ be continuous, we must have $f_0(x) = \phi(x)$ or $f_0(x) = -\phi(x)$ if $\phi(0) > 0$, while we could still have $f_0(x) = (signx)\phi(x)$ or $f_0(x) = -(signx)\phi(x)$ if $\phi(0) = 0$. The theorem is proved by noting that $IPB(-f, \xi) = IPB(f, \xi)$. ■

Before discussing the optimal designs for extrapolation, we introduce the concept of orthogonal polynomials and one of their properties . This material is from Dette and Studden (1997). A sequence of polynomials $\{P_i(x)\}_{i\geq 0}$, where $P_i(x)$ is of exact degree i, is said to be orthogonal with respect to the measure μ on the interval $[a, b]$ if

$$\int_a^b P_i(x)P_j(x)d\mu(x) = 0, \quad i \neq j.$$

The polynomials are called *monic orthogonal polynomials* with respect to the measure μ if the leading coefficient of $P_i(x)$ is 1 (for all $i \geq 0$).

Lemma 3 *The zeros of the orthogonal polynomials $\{P_i(x)\}_{i\geq 0}$, with respect to a measure μ on the interval $[a, b]$, are real and distinct and are located in the interior of the interval $[a, b]$.*

The lemma is taken from Lemma 2.1.3 of Dette and Studden (1997) (see also Theorem 3.3.1 of Szegö (1975)). It is essential in the proof of the Theorem 5 in Section 3.

3 Optimal designs

Since $S = [-1, 1]$ and $T = [-t, t]\backslash S$ are both symmetric, we will limit our consideration to the class of all symmetric designs. An explanation for this is given as follows. Consider the transformation:

$$\pi : x \longrightarrow \pi(x) = -x, \qquad x \in S.$$

Define $f_\pi(x) = f(\pi(x))$ for all $f \in \mathcal{F}$. Then $\pi S = S$ and $f_\pi \in \mathcal{F}$. Motivated by the discussion in Kiefer (1959), we restrict to the class of invariant designs ξ such that

$$IMSPE(f_\pi, \xi) = IMSPE(f, \xi), \qquad \textit{for all } f.$$

This requires that ξ be symmetric.

Let Ξ_s be the class of all symmetric designs on $[-1, 1]$. For any $\xi \in \Xi_s$, we know that to make $\boldsymbol{\beta}$ estimable ξ has to give positive probability to at least $(p + 1)$ points. The following Theorem 5 shows that we can confine ourselves to the $(p + 1)$-point symmetric design when $\phi(x) \equiv const$. By $\Xi_s(p)$, we mean the class of all symmetric designs which have only $(p + 1)$ support points.

Lemma 4 *Suppose $\phi(x) \equiv d > 0$. Define $\boldsymbol{\alpha}_0 := \mathbf{B}_\xi^{-1}\mathbf{b}_2(\xi)/d$, for any $\xi \in \Xi_s$.*

(a) If $(p+1)$ is even, then $(\boldsymbol{\alpha}_0)_2 = (\boldsymbol{\alpha}_0)_4 = \cdots = (\boldsymbol{\alpha}_0)_{p+1} = 0$.

(b) If $(p+1)$ is odd, then $(\boldsymbol{\alpha}_0)_1 = (\boldsymbol{\alpha}_0)_3 = \cdots = (\boldsymbol{\alpha}_0)_{p+1} = 0$.

Proof. Define

$$g(\boldsymbol{\alpha}) = \int_S (x^{p+1} - \boldsymbol{\alpha}^T \mathbf{z}(x))^2 d\xi.$$

Then $g(\alpha)$ is minimized uniquely by $\boldsymbol{\alpha}_0$. Let P be a $(p+1)\times(p+1)$ diagonal matrix, with $(i,i)^{th}$ element being $(-1)^{i-1}$. Then $\mathbf{z}(-x) = P\mathbf{z}(x)$.

If $(p+1)$ is even, then $g(\boldsymbol{\alpha}) = g(P\boldsymbol{\alpha})$. So the uniqueness of the minimizer implies that $\boldsymbol{\alpha}_0 = P\boldsymbol{\alpha}_0$. Hence, for even index i, we have

$$(\boldsymbol{\alpha}_0)_i = (P\boldsymbol{\alpha}_0)_i = -(\boldsymbol{\alpha}_0)_i,$$

that is, $(\boldsymbol{\alpha}_0)_i = 0$. (a) is proven. The proof of (b) is similar. ∎

Theorem 5 *Suppose $\phi(x) \equiv d > 0$. Then there exists $\xi_0 \in \Xi_s(p)$ such that*

$$\xi_0 = argmin_{\xi\in\Xi_s}\{max_{f\in\mathcal{F}} IMSPE(f,\xi)\}.$$

Proof. Consider the class of designs which are the members of Ξ_s with fixed even moments $E_\xi(X^{2i})$, $i = 0, 1, ..., p-1$. It is known that in this class there is a design belonging to $\Xi_s(p)$ which maximizes the moment $E_\xi(X^{2p})$ (See the proof of Theorem 2.6 of Liu and Wiens (1997)).

Since $\phi(x) \equiv d > 0$, then, by Theorem 2, we have the maximum value as

$$max_{f \in \mathcal{F}} IMSPE(f, \xi) = \int_T \mathbf{z}^T(x) \mathbf{B}_\xi^{-1} \mathbf{z}(x) dx + d^2 \int_T [\mathbf{z}^T(x) \mathbf{B}_\xi^{-1} \mathbf{b}_2(\xi)/d - x^{p+1}]^2 dx.$$

If for any $\xi \in \Xi_s$, with fixed $E_\xi(X^{2i})$, $i = 0, 1, ..., p-1$, $max_{f \in \mathcal{F}} IMSPE(f, \xi)$ is a decreasing function of $E_\xi(X^{2p})$, then minimizing $max_{f \in \mathcal{F}} IMSPE(f, \xi)$ is equivalent to maximizing $E_\xi(X^{2p})$. From the above paragraph, to prove the theorem, we only need to prove that $max_{f \in \mathcal{F}} IMSPE(f, \xi)$ is a decreasing function of $E_\xi(X^{2p})$ when $E_\xi(X^{2i})$, $i = 0, 1, ..., p-1$, are fixed. It is sufficient to show that $\int_T \mathbf{z}^T(x) \mathbf{B}_\xi^{-1} \mathbf{z}(x) dx$ and $\int_T [\mathbf{z}^T(x) \mathbf{B}_\xi^{-1} \mathbf{b}_2(\xi)/d - x^{p+1}]^2 dx$ simultaneously decrease as $E_\xi(X^{2p})$ increases.

In fact, define $\mathbf{e}_j \in \mathcal{R}^{p+1}$, with 1 as j^{th} element and 0 as other elements, $j = 1, 2, ..., p+1$. Define $u = E_\xi(X^{2p})$ and functions $h_1(u) = \mathbf{z}^T(x) \mathbf{B}_\xi^{-1} \mathbf{z}(x)$, $h_2(u) = \mathbf{z}^T(x) \mathbf{B}_\xi^{-1} \mathbf{b}_2(\xi)/d$ for any $x \in T$. Then

$$\frac{\partial h_1(u)}{\partial u} = -\mathbf{z}^T(x) \mathbf{B}_\xi^{-1} \mathbf{e}_{p+1} \mathbf{e}_{p+1}^T \mathbf{B}_\xi^{-1} \mathbf{z}(x) \leq 0.$$

This indicates that $h_1(u)$ decreases with u, for any $x \in T$. Hence $\int_T \mathbf{z}^T(x) \mathbf{B}_\xi^{-1} \mathbf{z}(x) dx$ decreases with $E_\xi(X^{2p})$.

Next, we first show that $h_2(u)$ is an increasing function. With $c_i = E_\xi X^i$ and $\mathbf{b}_2(\xi)/d = (c_{p+1}, ..., c_{2p}, c_{2p+1})^T$, we have

$$\frac{\partial h_2(u)}{\partial u} = -\mathbf{z}^T(x) \mathbf{B}_\xi^{-1} \mathbf{e}_{p+1} \mathbf{e}_{p+1}^T \mathbf{B}_\xi^{-1} \mathbf{b}_2(\xi)/d + \mathbf{z}^T(x) \mathbf{B}_\xi^{-1} \mathbf{e}_p/d.$$

The first term vanishes because $\mathbf{e}_{p+1}^T \mathbf{B}_\xi^{-1} \mathbf{b}_2(\xi)/d$ is just $(\boldsymbol{\alpha}_0)_{p+1}$ in Lemma 4 and is zero. Now, we want to show $\mathbf{z}^T(x) \mathbf{B}_\xi^{-1} \mathbf{e}_p \geq 0$. Define polynomials $Q_{p-1}(x) =$

$\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{e}_p$, then

$$\begin{aligned} Q_{p-1}(x) &= \mathbf{e}_1^T\mathbf{B}_\xi^{-1}\mathbf{e}_p + x\mathbf{e}_2^T\mathbf{B}_\xi^{-1}\mathbf{e}_p + \cdots + x^{p-1}\mathbf{e}_p^T\mathbf{B}_\xi^{-1}\mathbf{e}_p + x^p\mathbf{e}_{p+1}^T\mathbf{B}_\xi^{-1}\mathbf{e}_p \\ &= \frac{1}{|\mathbf{B}_\xi|}\begin{vmatrix} c_0 & 0 & c_2 & \cdots & c_{p-2} & 1 & c_p \\ 0 & c_2 & 0 & \cdots & c_{p-1} & x & c_{p+1} \\ & & \cdots & \cdots & \cdots & & \\ c_p & c_{p+1} & c_{p+2} & \cdots & c_{2p-2} & x^p & c_{2p} \end{vmatrix} \end{aligned}$$

has degree $(p-1)$ because $\mathbf{e}_{p+1}^T\mathbf{B}_\xi^{-1}\mathbf{e}_p = 0$, where the second equality comes from the relation between $\mathbf{B}_\xi^{-1}$ and the adjoint matrix of $\mathbf{B}_\xi$ and from the expression of a determinant in terms of cofactors. Now,

$$\begin{aligned} \int_{-1}^{1} x^i Q_{p-1}(x)d\xi(x) &= \frac{1}{|\mathbf{B}_\xi|}\begin{vmatrix} c_0 & 0 & c_2 & \cdots & c_{p-2} & c_i & c_p \\ 0 & c_2 & 0 & \cdots & c_{p-1} & c_{1+i} & c_{p+1} \\ & & \cdots & \cdots & \cdots & & \\ c_p & c_{p+1} & c_{p+2} & \cdots & c_{2p-2} & c_{p+i} & c_{2p} \end{vmatrix} \\ &= 0, \end{aligned}$$

where $i = 0, 1, \cdots, p-2$, since two columns in the determinant are equal. So $\{Q_{p-1}(x)\}_{p\geq 1}$, with $Q_0(x) \equiv 1$, are orthogonal polynomials with respect to ξ on $[-1, 1]$. By Lemma 3, we conclude that $Q_{p-1}(x)$ has $(p-1)$ real zeros belonging to $(-1, 1)$. Hence $Q_{p-1}(x) > 0$, *i.e.* $\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{e}_p > 0$, for $x > 1$, since $\mathbf{e}_p^T\mathbf{B}_\xi^{-1}\mathbf{e}_p > 0$. This implies that $h_2(u) = \mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_2(\xi)/d$ is an increasing function of u.

We note that as a polynomial of x, $\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_2(\xi)/d$ only has odd-power terms when p is even, and even-power terms and constant term when p is odd. Hence, $\int_T[\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_2(\xi)/d - x^{p+1}]^2dx = 2\int_1^t[\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_2(\xi)/d - x^{p+1}]^2dx$. So to prove that $\int_T[\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_2(\xi)/d - x^{p+1}]^2dx$ decreases as $E_\xi X^{2p}$ increases, it is sufficient to show that $\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_2(\xi)/d < x^{p+1}$, for $x > 1$. Define $W_{p+1}(x) = x^{p+1} - \mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_2(\xi)/d$, $p \geq 1$, then $W_{p+1}(x)$ is monic polynomial and

$$W_{p+1}(x) = x^{p+1} - \frac{1}{|\mathbf{B}_\xi|}\{\begin{vmatrix} 1 & c_1 & \cdots & c_p \\ x & c_2 & \cdots & c_{p+1} \\ & \cdots & \cdots & \\ x^p & c_{p+1} & \cdots & c_{2p} \end{vmatrix} c_{p+1} + \begin{vmatrix} c_0 & 1 & \cdots & c_p \\ c_1 & x & \cdots & c_{p+1} \\ & \cdots & \cdots & \\ c_p & x^p & \cdots & c_{2p} \end{vmatrix} c_{p+2} +$$

$$\cdots + \begin{vmatrix} c_0 & \cdots & 1 & c_p \\ c_1 & \cdots & x & c_{p+1} \\ & \cdots & \cdots & \\ c_p & \cdots & x^p & c_{2p} \end{vmatrix} c_{2p} + \begin{vmatrix} c_0 & c_1 & \cdots & 1 \\ c_1 & c_2 & \cdots & x \\ & \cdots & \cdots & \\ c_p & c_{p+1} & \cdots & x^p \end{vmatrix} c_{2p+1}\}.$$

Since $\int_{-1}^1 x^i W_{p+1}(x)d\xi(x) = c_{p+i+1} - c_{p+i+1} = 0, i = 1, 2, ..., p$, $\{W_{p+1}(x)\}_{p\geq 1}$ are monic orthogonal polynomials with respect to ξ on $[-1, 1]$. In light of Lemma 3, $W_{p+1}(x)$ has $(p+1)$ real zeros belonging to $(-1, 1)$. So $W_{p+1}(x) > 0$, *i.e.* $\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_2(\xi)/d < x^{p+1}$ when $x > 1$. Hence $\int_T[\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_2(\xi)/d - x^{p+1}]^2dx$ decreases with $E_\xi X^{2p}$. This proves the theorem. ■

In the following subsections, we find that when $p = 1, 2$, the result of the above theorem still holds for arbitrary $\phi \in \mathcal{F}$. When the observations are taken at exactly

$(p+1)$ distinct x-values for model (2), the Least Squares estimator of the response function can be expressed as follows (assume $\xi(x_v) = p_v = \frac{n_v}{n} > 0$, $v = 0, 1, ...p$):

$$\hat{Y}(x) = \sum_{v=0}^{p} l_v(x)\bar{y}_v,$$

where $l_v = \Pi_{j\neq v}\frac{(x-x_j)}{(x_v-x_j)}$ is the Lagrange interpolation polynomial and $\bar{y}_v$ is the average of the observed values taken at x_v. Thus, $IMSPE(f,\xi)$ can be rewritten

$$\begin{aligned} IMSPE(f,\xi) &= \int_T VAR[\hat{Y}(x)]dx + \int_T [E(\hat{Y}(x)) - E(Y(x)]^2 dx \\ &= \frac{\sigma^2}{n}\sum_{v=0}^{p}\frac{\int_T l_v^2(x)dx}{p_v} + \int_T [E(\hat{Y}(x)) - E(Y(x)]^2 dx, \end{aligned}$$

where the second term does not depend on p_v since $E(\bar{y}_v) = n_v E(Y(x_v))/n_v = E(Y(x_v))$ does not depend on p_v. Hence, for any chosen set of design supports, the following theorem yields the corresponding masses p_v's which minimize $IMSPE(f,\xi)$.

Theorem 6 *For any symmetric design* ξ*:* $\xi(x_v) = p_v$*, with* $p_v \in (0,1)$ *and* $\sum_{v=0}^{p} p_v = 1$*, the choice*

$$p_v = \frac{[\int_T l_v^2(x)dx]^{1/2}}{\sum_{i=0}^{p}[\int_T l_i^2(x)dx]^{1/2}}, \qquad v = 0, 1, ...p \tag{4}$$

will minimize $IMSPE(f,\xi)$.

Proof. For any design ξ, the minimizers in (4) are obtained by using standard calculus techniques. (See also Lemma 1 of Hoel and Levine (1964)). Now assume ξ has supports $\{\pm x_i,\ i = 1, ..., \frac{p+1}{2}$, when p odd$\}$ or $\{\pm x_i, 0,\ i = 1, ..., \frac{p}{2}$, when p even$\}$, we claim that $p_i = p_{-i}$, with p_{-i} being the mass on $-x_i$. In fact, when p is odd,

$$\begin{aligned}
&\int_T l_i^2(x)dx \\
= &\int_T [\frac{(x+x_{\frac{p+1}{2}})...(x+x_1)(x-x_1)...(x-x_{i-1})(x-x_{i+1})...(x-x_{\frac{p+1}{2}})}{(x_i+x_{\frac{p+1}{2}})...(x_i+x_1)(x_i-x_1)...(x_i-x_{i-1})(x_i-x_{i+1})...(x_i-x_{\frac{p+1}{2}})}]^2 dx \\
= &\int_T [\frac{(-x+x_{\frac{p+1}{2}})...(-x+x_1)(-x-x_1)...(-x-x_{i-1})(-x-x_{i+1})...(-x-x_{\frac{p+1}{2}})}{(x_i+x_{\frac{p+1}{2}})...(x_i+x_1)(x_i-x_1)...(x_i-x_{i-1})(x_i-x_{i+1})...(x_i-x_{\frac{p+1}{2}})}]^2 dx \\
= &\int_T [\frac{(x+x_{\frac{p+1}{2}})...(x+x_{i+1})(x+x_{i-1})...(x+x_1)(x-x_1)...(x-x_{\frac{p+1}{2}})}{(-x_i+x_{\frac{p+1}{2}})...(-x_i+x_{i+1})(-x_i-x_{i-1})...(-x_i+x_1)(-x_i-x_1)...(-x_i-x_{\frac{p+1}{2}})}]^2 dx \\
= &\int_T l_{-i}^2(x)dx.
\end{aligned}$$

This holds as well when p even. This, together with (3), establishes the claim. ■

With these masses, the variance term of $IMSPE(f, \xi)$ becomes

$$IPV(\xi) = \frac{\sigma^2}{n}[\sum_{v=0}^{p}(\int_T l_v^2(x)dx)^{1/2}]^2. \tag{5}$$

Therefore, when constructing the optimal designs, it is sufficient to consider the symmetric design ξ with masses given by (4) if we confine ourselves to designs $\xi \in \Xi_s(p)$. We will find the robust extrapolation designs for special values of p in the following subsections. The conclusions obtained in Subsections 3.1 and 3.2 imply that the optimal robust extrapolation designs (when $p = 1, 2$) are the same as the designs which only minimize the integrated variance. But they are different when $p \geq 3$.

3.1 Linear regression ($p = 1$).

From past experience, the experimenter suspects that the response function is only approximately linear in x. Then, for any $\xi \in \Xi_s$ and ϕ satisfying A1) and A2), we have

$$\mathbf{B}_\xi = \begin{pmatrix} 1 & 0 \\ 0 & E_\xi(X^2) \end{pmatrix}, \qquad \mathbf{A} = \begin{pmatrix} 2(t-1) & 0 \\ 0 & 2(t^3-1)/3 \end{pmatrix},$$

$$\mathbf{b}_1(\xi) = \begin{pmatrix} 0 \\ E_\xi(X^3 sign(X)\phi(X)) \end{pmatrix}, \qquad \mathbf{b}_2(\xi) = \begin{pmatrix} E_\xi(X^2\phi(X)) \\ 0 \end{pmatrix}.$$

Hence,

$$\begin{aligned} K_1(\xi) &= \int_T [x^2 sign(x)\phi(x) - \frac{xE_\xi(X^3 sign(X)\phi(X))}{E_\xi(X^2)}]^2 dx \\ &= 2\int_1^t [x^2\phi(x) - \frac{xE_\xi(X^3 sign(X)\phi(X))}{E_\xi(X^2)}]^2 dx, \\ K_2(\xi) &= \int_T [x^2\phi(x) - E_\xi(X^2\phi(X))]^2 dx = 2\int_1^t [x^2\phi(x) - E_\xi(X^2\phi(X))]^2 dx. \end{aligned}$$

The following theorem extends the result in Theorem 5 to the situation when ϕ is not a constant.

Theorem 7 *For $p = 1$ and any ϕ satisfying A1) and A2), design $\xi_0 : \xi_0(\pm 1) = 0.5$ is the optimal robust extrapolation design, that is,*

$$\xi_0 = argmin_{\xi \in \Xi_s}\{max_{f \in \mathcal{F}} IMSPE(f, \xi)\}.$$

Proof. For any $\xi \in \Xi_s$, Theorem 2 gives

$$max_{f\in\mathcal{F}} IMSPE(f,\xi) = \frac{\sigma^2}{n}[2(t-1) + \frac{2}{3}(t^3-1)/E_\xi(X^2)] + max\{K_1(\xi), K_2(\xi)\}.$$

By Lemma 1, we can show that (1) $x^2\phi(x) \geq E_\xi(X^2\phi(X))$ and $x^2\phi(x) \geq \frac{xE_\xi(X^3 sign(X)\phi(X))}{E_\xi(X^2)}$ for all $x \in [1,t]$; (2) if x_m is the largest value of the design points of ξ and $\xi^*(\pm x_m) = \frac{1}{2}$, then $E_\xi(X^2) \leq x_m^2 = E_{\xi^*}(X^2)$ and $E_\xi(X^2\phi(X)) \leq x_m^2\phi(x_m) = E_{\xi^*}(X^2\phi(X))$; (3)

$$\frac{xE_{\xi^*}(X^3 sign(X)\phi(X))}{E_{\xi^*}(X^2)} = xx_m\phi(x_m) \geq x_m^2\phi(x_m) = E_{\xi^*}(X^2\phi(X)), \quad \text{for all } x \in [1,t].$$

Therefore, $max\{K_1(\xi), K_2(\xi)\} \geq K_2(\xi) \geq K_2(\xi^*) = max\{K_1(\xi^*), K_2(\xi^*)\}$ and it follows that

$$max_{f\in\mathcal{F}} IMSPE(f,\xi) \geq max_{f\in\mathcal{F}} IMSPE(f,\xi^*).$$

So we only need to consider the two-point symmetric design ξ^*. Now, with arbitrary $x_m \in (0,1]$, we have

$$max_{f\in\mathcal{F}} IMSPE(f,\xi^*) = \frac{\sigma^2}{n}[2(t-1)+\frac{2}{3}(t^3-1)/E_{\xi^*}(X^2)]+2\int_1^t [x^2\phi(x)-E_{\xi^*}(X^2\phi(X))]^2dx.$$

Similarly, Lemma 1 gives that $E_{\xi^*}(X^2\phi(X)) \leq 1^2\phi(1)$ and $E_{\xi^*}(X^2) \leq 1$, where equalities hold when $\xi^*(\pm 1) = 0.5$. Hence, $\xi_0 : \xi_0(\pm 1) = 0.5$ is the optimal robust extrapolation design. ■

A by-product of the above proof is that ξ_0 is the minimum integrated variance (IPV (ξ)) design, too.

Kiefer and Wolfowitz (1964) constructed the optimal designs for extrapolation by minimizing $D(\xi) = max_{x\in S\cup T}\{\mathbf{z}^T(x)\mathbf{M}^{-1}\mathbf{z}(x)\}$ when the model (1) is true, where $\mathbf{M}$ is the information matrix. They found that ξ_0 was the optimal design for extrapolation in straightline regression.

3.2 Quadratic regression ($p = 2$).

As in the straightline regression case, the following Theorem 8 indicates that the conclusion of Theorem 4 holds without the restriction of $\phi(x) = const.$ With $\mathbf{z}(x) = (1, x, x^2)^T$ and a symmetric ξ, we have

$$\mathbf{b}_1(\xi) = \begin{pmatrix} E_\xi(X^3 sign(X)\phi(X)) \\ 0 \\ E_\xi(X^5 sign(X)\phi(X)) \end{pmatrix}, \quad \mathbf{b}_2(\xi) = \begin{pmatrix} 0 \\ E_\xi(X^4\phi(X)) \\ 0 \end{pmatrix},$$

$$\mathbf{B}_\xi = \begin{pmatrix} 1 & 0 & E_\xi(X^2) \\ 0 & E_\xi(X^2) & 0 \\ E_\xi(X^2) & 0 & E_\xi(X^4) \end{pmatrix}.$$

Theorem 8 *For $p = 2$ and any ϕ satisfying A1) and A2), there exist a three-point design $\xi_0 \in \Xi_s(2)$, such that*

$$\xi_0 = argmin_{\xi \in \Xi_s}\{max_{f \in \mathcal{F}} IMSPE(f, \xi)\}.$$

Proof. Assume design $\xi : \xi(\pm x_i) = p_i/2$ and $\xi(0) = p_0$, $i = 1, ..., m$, where p_0 might be equal to zero but $p_i > 0$ for $i > 0$ and $\sum_{i=0}^{m} p_i = 1$, then

$$\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_2(\xi) = \frac{E_\xi(X^4\phi(X))}{E_\xi(X^2)}x \leq \frac{E_\xi(X^2)x_m^2\phi(x_m)}{E_\xi(X^2)}x = xx_m^2\phi(x_m).$$

Now for design $\xi^* : \xi^*(\pm x_m) = p$ and $\xi^*(0) = 1 - 2p$, $p > 0$, such that $E_\xi(X^2) = E_{\xi^*}(X^2)$, we obtain

$$\begin{aligned}
\mathbf{z}^T(x)\mathbf{B}_{\xi^*}^{-1}\mathbf{b}_1(\xi^*) &= \frac{-E_\xi(X^4)E_\xi(X^3 sign(X)\phi(X)) + E_\xi(X^2)E_\xi(X^5 sign(X)\phi(X))}{[E_\xi(X^2)]^2 - E_\xi(X^4)} + \\
&\quad \frac{E_\xi(X^2)E_\xi(X^3 sign(X)\phi(X)) - E_\xi(X^5 sign(X)\phi(X))}{[E_\xi(X^2)]^2 - E_\xi(X^4)}x^2 \\
&= x^2 x_m \phi(x_m), \\
\mathbf{z}^T(x)\mathbf{B}_{\xi^*}^{-1}\mathbf{b}_2(\xi^*) &= \frac{E_\xi(X^4\phi(X))}{E_\xi(X^2)}x = \frac{2px_m^4\phi(x_m^4)}{2px_m^2}x = xx_m^2\phi(x_m).
\end{aligned}$$

Hence,

$$\begin{aligned}
K_1(\xi^*) &= \int_T [x^2 x_m\phi(x_m) - x^3 sign(x)\phi(x)]^2 dx = 2\int_1^t [x^2 x_m \phi(x_m) - x^3\phi(x)]^2 dx, \\
K_2(\xi^*) &= \int_T [xx_m^2\phi(x_m) - x^3\phi(x)]^2 dx = 2\int_1^t [xx_m^2\phi(x_m) - x^3\phi(x)]^2 dx.
\end{aligned}$$

The monotonicity of the function $x\phi(x)$ implies that $x^3\phi(x) \geq x^2 x_m\phi(x_m) \geq xx_m^2\phi(x_m)$ for all $x \in [1, t]$. This yields

$$max\{K_1(\xi), K_2(\xi)\} \geq K_2(\xi) \geq K_2(\xi^*) = max\{K_1(\xi^*), K_2(\xi^*)\},$$

which implies that ξ^* has smaller bias term than ξ.

We claim that $tr(\mathbf{B}_\xi^{-1}\mathbf{A}) \geq tr(\mathbf{B}_{\xi^*}^{-1}\mathbf{A})$. In fact, for fixed $E_\xi(X^2)$, $tr(\mathbf{B}_\xi^{-1}\mathbf{A}) = \int_T \mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{z}(x)dx$ is a decreasing function of $E_\xi(X^4)$. Now by $E_\xi(X^2) = E_{\xi^*}(X^2)$, *viz,*

$$2px_m^2 = \sum_{i=1}^{m} p_i x_i^2,$$

we have

$$E_{\xi^*}(X^4) = 2px_m^4 = \sum_{i=1}^{m} p_i x_i^2 x_m^2 > \sum_{i=1}^{m} p_i x_i^4 = E_\xi(X^4),$$

which implies that $tr(\mathbf{B}_\xi^{-1}\mathbf{A}) \geq tr(\mathbf{B}_{\xi^*}^{-1}\mathbf{A})$. The claim is proven.

From the above arguments, to find ξ in order to minimize $max_{f\in\mathcal{F}} IMSPE(f,\xi)$, it is sufficient to confine ourselves to three-point design $\xi \in \Xi_s(3) : \xi(\pm x_m) = p$ and $\xi(0) = 1-2p$, with any $x_m \in (0,1]$ and $p > 0$. Under this circumstance,

$$\begin{aligned} max_{f\in\mathcal{F}} IPB(f,\xi) &= 2\int_1^t [x^2 x_m \phi(x_m) - x^3\phi(x)]^2 dx \\ &\geq 2\int_1^t [x^2 1\phi(1) - x^3\phi(x)]^2 dx, \end{aligned}$$

where the equality holds when $x_m = 1$. In light of Theorem 6 and (5), we have

$$\begin{aligned} IPV(\xi) &= \frac{\sigma^2}{n}\left(2[\int_T (\frac{(x+x_m)x}{2x_m^2})^2 dx]^{1/2} + [\int_T (\frac{(x+x_m)(x-x_m)}{x_m^2})^2 dx]^{1/2}\right)^2 \\ &= \frac{\sigma^2}{n}\left(2[\int_1^t (\frac{(x+x_m)x}{2x_m^2})^2 dx + \int_1^t (\frac{(x-x_m)x}{2x_m^2})^2 dx]^{1/2} + [\int_T (\frac{x^2-x_m^2}{x_m^2})^2 dx]^{1/2}\right)^2 \\ &= \frac{\sigma^2}{n}\left(2[\int_1^t \frac{(2x^2+2x_m^2)x^2}{4x_m^4} dx]^{1/2} + [\int_T (\frac{x^2-x_m^2}{x_m^2})^2 dx]^{1/2}\right)^2 \\ &\geq \frac{\sigma^2}{n}\left(2[\int_1^t \frac{(2x^2+2)x^2}{4} dx]^{1/2} + [\int_T (\frac{x^2-1}{1})^2 dx]^{1/2}\right)^2, \end{aligned}$$

where the equality holds when $x_m = 1$. This shows that the three-point design $\xi_0 : \xi_0(\pm 1) = p$ and $\xi_0(0) = 1-2p$ simultaneously minimizes $max_{f\in\mathcal{F}} IPB(f,\xi)$ and $IPV(\xi)$, where p is determined by Theorem 6. The proof is completed. ■

The value of p varies with t. To determine p, a simple calculation yields

$$\begin{aligned} \int_1^t \frac{(2x^2+2)x^2}{4} dx &= \frac{1}{10}t^5 + \frac{1}{6}t^3 - \frac{4}{15}, \\ \int_T (x^2-1)^2 dx &= \frac{2}{5}t^5 - \frac{4}{3}t^3 + 2t - \frac{16}{15}. \end{aligned}$$

Then,

$$p = \frac{(\frac{1}{10}t^5 + \frac{1}{6}t^3 - \frac{4}{15})^{1/2}}{2(\frac{1}{10}t^5 + \frac{1}{6}t^3 - \frac{4}{15})^{1/2} + (\frac{2}{5}t^5 - \frac{4}{3}t^3 + 2t - \frac{16}{15})^{1/2}}.$$

We exhibit some typical t values and the corresponding values of p:

$$\begin{pmatrix} t: & 1.5 & 5 & 10 & 100 \\ p: & 0.37 & 0.26 & 0.25 & 0.25 \end{pmatrix}.$$

As the proof of Theorem 8 points out, this design ξ_0 minimizes $IPV(\xi)$ as well, that is, the minimum integrated variance design is model robust for quadratic regression. Kiefer and Wolfowitz (1964) gave the optimal design for extrapolation in quadratic regression as: $\xi_{K-W}(0) = 1/2 - 3/8t^2$ and $\xi_{K-W}(\pm 1) = 1/4 + 3/16t^2$. Numerically, $\xi_{K-W}(\pm 1) < \xi_0(\pm 1)$, but the difference is very small when t becomes large. Note that the two designs come from two different criteria.

We find that $p \to 0.25$ as $t \to \infty$. This means that we take the same number of observations at the origin as the end points of design interval when the extrapolation region becomes very large. This design was employed for a most powerful test for distinguishing quadratic regression from straightline linear regression (Huber, 1981, page 249). It is also the so-called D_1-design which minimizes the variance of the

Least Squares estimate of the highest coefficient in the quadratic regression (Dette and Studden, 1997, page 141).

3.3 Cubic regression ($p = 3$).

Theorem 5 implies that the robust extrapolation design only has four support points when $p = 3$ and $\phi(x) \equiv const$. Although we do not know whether the conclusion of this theorem is true when $\phi(x) \neq const$, we still like to confine ourselves to four-point designs because in some situations, people might be interested in the design with as few observations as possible.

Theorem 9 *When $p = 3$ and $\xi \in \Xi_s(3)$, we have, for any ϕ satisfying A1) and A2),*

$$max_{f \in \mathcal{F}} IPB(f, \xi) = K_2(\xi)$$

Proof. Assume the four-point symmetric design $\xi : \xi(\pm x_1) = 0.5 - p$, and $\xi(\pm x_2) = p$, with $p > 0$, then by the definition of the moment of a discrete random variable, we have,

$$\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_1(\xi) = \frac{x x_2^3 \phi(x_2)(x^2 - x_1^2) - x x_1^3 \phi(x_1)(x^2 - x_2^2)}{x_2^2 - x_1^2} \geq 0,$$

and

$$\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_2(\xi) = \frac{x_2^4 \phi(x_2)(x^2 - x_1^2) - x_1^4 \phi(x_1)(x^2 - x_2^2)}{x_2^2 - x_1^2} \geq 0.$$

Now, we claim that $\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_2(\xi) \leq \mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_1(\xi) \leq x^4\phi(x)$ for all $x \in [1, t]$. If these inequalities hold, then, by the definitions of $K_1(\xi)$ and $K_2(\xi)$, we have $K_1(\xi) \leq$

$K_2(\xi)$, *i.e.* $max_{f\in\mathcal{F}} IPB(f,\xi) = K_2(\xi)$ in this case. In fact,

$$\begin{aligned}
& (x_2^2 - x_1^2)[\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_1(\xi) - \mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_2(\xi)] \\
= & \{xx_2^3\phi(x_2)(x^2 - x_1^2) - xx_1^3\phi(x_1)(x^2 - x_2^2)\} - \{x_2^4\phi(x_2)(x^2 - x_2^2) - x_1^4\phi(x_1)(x^2 - x_1^2)\} \\
= & (x - x_2)(x - x_1)\{x(x_2^3\phi(x_2) - x_1^3\phi(x_1)) + x_1x_2(x_2^2\phi(x_2) - x_1^2\phi(x_2))\} \\
\geq & 0,
\end{aligned}$$

where the inequality is because of the monotonicity of function $x\phi(x)$. Hence, $\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_2(\xi) \leq \mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_1(\xi)$. Similarly, we have

$$(x_2^2 - x_1^2)\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_1(\xi) \leq \{x^2x_2^2\phi(x_2)(x^2 - x_1^2) - x^2x_1^2\phi(x_1)(x^2 - x_2^2)\}.$$

But, by the convexity of $g_1(x) = x\phi(\sqrt{x})$, we obtain

$$\frac{x_2^2\phi(x_2) - x_1^2\phi(x_1)}{x_2^2 - x_1^2} \leq \frac{x^2\phi(x) - x_1^2\phi(x_1)}{x^2 - x_1^2},$$

that is

$$x_2^2\phi(x_2)(x^2 - x_1^2) \leq x^2\phi(x)(x_2^2 - x_1^2) + x_1^2\phi(x_1)(x^2 - x_2^2).$$

Hence,

$$\begin{aligned}
& x^2x_2^2\phi(x_2)(x^2 - x_1^2) - x^2x_1^2\phi(x_1)(x^2 - x_2^2) \\
\leq & x^2[x^2\phi(x)(x_2^2 - x_1^2) + x_1^2\phi(x_1)(x^2 - x_2^2) - x_1^2\phi(x_1)(x^2 - x_2^2)] \\
= & x^4\phi(x)(x_2^2 - x_1^2).
\end{aligned}$$

Therefore, $\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_1(\xi) \leq x^4\phi(x)$. Thus the claim is true and the proof is completed. ■

Next, we want to locate the design points, that is, to find x_1, x_2 (where $x_1 < x_2$) in order to minimize $G(x_1, x_2) = max_{f\in\mathcal{F}} IMSPE(f,\xi)$, while the corresponding variance term is $IPV(x_1, x_2)$ (given by (5), since the design only has four support points and Theorem 6 applies) and the bias term $IPB(x_1, x_2) = K_2(\xi)$.

For any fixed $x_1 \in (0, 1)$, we have

$$\begin{aligned}
\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_2(\xi) &= -\frac{x_1^2x_2^2(x_2^2\phi(x_2) - x_1^2\phi(x_1))}{x_2^2 - x_1^2} + \frac{x_2^4\phi(x_2) - x_1^4\phi(x_1)}{x_2^2 - x_1^2}x^2 \\
&= (x^2 - x_1^2)x_2^2\phi(x_2) + x_1^4\phi(x_1) + x_1^2(x^2 - x_1^2)\frac{x_2^2\phi(x_2) - x_1^2\phi(x_1)}{x_2^2 - x_1^2} \\
&\leq (x^2 - x_1^2)1\phi(1) + x_1^4\phi(x_1) + x_1^2(x^2 - x_1^2)\frac{1\phi(1) - x_1^2\phi(x_1)}{1 - x_1^2},
\end{aligned}$$

then

$$\begin{aligned}
IPB(x_1, x_2) &= \int_T [\mathbf{z}^T(x)\mathbf{B}_\xi^{-1}\mathbf{b}_2(\xi) - x^4\phi(x)]^2 dx \\
&\geq IPB(x_1, 1),
\end{aligned}$$

where the equality holds when $x_2 = 1$.

For the variance part,

$$\begin{aligned}
IPV(x_1, x_2) &= \frac{\sigma^2}{n}\{2(\int_T [\frac{(x + x_1)(x - x_1)(x - x_2)}{(-x_2 + x_1)(-x_2 - x_1)(-2x_2)}]^2 dx)^{1/2} + \\
&\quad 2(\int_T [\frac{(x + x_2)(x - x_2)(x - x_1)}{(-x_1 + x_2)(-x_1 - x_2)(-2x_1)}]^2 dx)^{1/2}\}^2 \\
&= \frac{\sigma^2}{n}\frac{1}{(x_2^2 - x_1^2)^2}\{[\int_1^t \frac{(x^2 - x_1^2)^2(2x^2 + 2x_2^2)}{x_2^2}dx]^{1/2} + \\
&\quad [\int_1^t \frac{(x^2 - x_2^2)^2(2x^2 + 2x_1^2)}{x_1^2}dx]^{1/2}\}^2 \\
&\geq IPV(x_1, 1),
\end{aligned}$$

where the equality holds when $x_2 = 1$. So in order that ξ be the optimal robust design for extrapolation, it must have ± 1 as its support points. It remains to choose x_1 to minimize $G(x_1, 1) = IPB(x_1, 1) + IPV(x_1, 1)$. Obviously, the choice will depend on the values of $\frac{\sigma^2}{n}$, t and on ϕ. With several typical ϕ's, we give out the values of x_1 and the corresponding masses in the following examples by the help of computer.

Example 1. When $\phi(x) \equiv 0$, the cubic model is correct and the bias term in the loss function disappears. In this case, only the variance is considered to find the optimal extrapolation design. Then the choice of x_1 only depends on t, not on $\frac{\sigma^2}{n}$. Table 1 gives the positive design points and the corresponding masses on these points. We find that the design points are quite close to the Chebychev points $\{\pm 0.5, \pm 1\}$, which is the optimal design for extrapolation to one point when the cubic model is exactly correct (Hoel and Levine (1964)).

Table 1. The positive design points and masses when $\phi(x) \equiv 0$.

t	1.5	5	10	100
(x_1, x_2)	$(0.475, 1)$	$(0.497, 1)$	$(0.499, 1)$	$(0.5, 1)$
(p_1, p_2)	$(0.238, 0.262)$	$(0.327, 0.173)$	$(0.332, 0.168)$	$(0.333, 0.167)$

As in straightline and quadratic regression, when extrapolating to a symmetric region, Kiefer and Wolfowitz (1964) constructed the optimal design for extrapolation in cubic regression: $\xi_{K-W}(\pm 1/2 \mp 1/16t^2) = 1/3 - 1/9t^2$ and $\xi_{K-W}(\pm 1) = 1/6 + 1/9t^2$. These values are very close to those in Table 1.

Example 2. When $\phi(x) \neq 0$, the cubic regression model is only an approximation of the model response. In this case, the values of x_1 depend on $\frac{\sigma^2}{n}$ as well as t. Table 2 and Table 3 give the positive design points and masses when $\phi(x) \equiv 1$ and $\phi(x) = |x|$.

Table 2. The positive design points and masses when $\phi(x) \equiv 1$.

$\frac{\sigma^2}{n}$	t	1.5	5	10	100
100	(x_1, x_2)	(0.475, 1)	(0.497, 1)	(0.499, 1)	(0.5, 1)
	(p_1, p_2)	(0.238, 0.262)	(0.327, 0.173)	(0.332, 0.168)	(0.333, 0.167)
10	(x_1, x_2)	(0.476, 1)	(0.497, 1)	(0.5, 1)	(0.5, 1)
	(p_1, p_2)	(0.237, 0.263)	(0.327, 0.173)	(0.333, 0.167)	(0.333, 0.167)
1	(x_1, x_2)	(0.479, 1)	(0.504, 1)	(0.507, 1)	(0.508, 1)
	(p_1, p_2)	(0.237, 0.263)	(0.325, 0.175)	(0.33, 0.17)	(0.332, 0.168)
0.1	(x_1, x_2)	(0.516, 1)	(0.575, 1)	(0.582, 1)	(0.584, 1)
	(p_1, p_2)	(0.232, 0.268)	(0.311, 0.189)	(0.314, 0.186)	(0.316, 0.184)
0.01	(x_1, x_2)	(0.724, 1)	(0.799, 1)	(0.805, 1)	(0.806, 1)
	(p_1, p_2)	(0.221, 0.279)	(0.274, 0.226)	(0.276, 0.224)	(0.277, 0.223)

Table 3. The positive design points and masses when $\phi(x) = |x|$.

$\frac{\sigma^2}{n}$	t	1.5	5	10	100
100	(x_1, x_2)	$(0.475, 1)$	$(0.497, 1)$	$(0.5, 1)$	$(0.5, 1)$
	(p_1, p_2)	$(0.238, 0.262)$	$(0.327, 0.173)$	$(0.333, 0.167)$	$(0.333, 0.167)$
10	(x_1, x_2)	$(0.476, 1)$	$(0.501, 1)$	$(0.508, 1)$	$(0.602, 1)$
	(p_1, p_2)	$(0.237, 0.263)$	$(0.326, 0174)$	$(0.33, 0.17)$	$(0.312, 0.188)$
1	(x_1, x_2)	$(0.485, 1)$	$(0.546, 1)$	$(0.601, 1)$	$(0.832, 1)$
	(p_1, p_2)	$(0.236, 0.264)$	$(0.317, 0.183)$	$(0.311, 0.189)$	$(0.273, 0.227)$
0.1	(x_1, x_2)	$(0.585, 1)$	$(0.777, 1)$	$(0.832, 1)$	$(0.928, 1)$
	(p_1, p_2)	$(0.225, 0.275)$	$(0.277, 0.223)$	$(0.272, 0.228)$	$(0.259, 0.241)$
0.01	(x_1, x_2)	$(0.822, 1)$	$(0.907, 1)$	$(0.928, 1)$	$(0.968, 1)$
	(p_1, p_2)	$(0.225, 0.275)$	$(0.26, 0.24)$	$(0.259, 0.241)$	$(0.254, 0.246)$

Remarks:

1. These tables show that when σ^2/n becomes large (variance dominant case), the designs tend to the design obtained by assuming $\phi(x) = 0$. That is, we can construct the design by minimizing the integrated variance only if we know in advance that the sampling variation is very large. When σ^2/n becomes small, the model bias does have an affect on the choice of the design. And when the degree of $\phi(x)$ becomes higher, i.e. the bias of the model becomes more "serious"

(in the extrapolation region), the design points move closer to the end-points of the design interval for every fixed t value.

2. For a fixed value of σ^2/n, the value of x_1 increases as t increases. This makes sense, as Lawless (1984, page 7) points out, because fitting a cubic model based on observations at $x_2 = \pm 1$ and x_1 closer to the end-point of the design interval will reduce the prediction bias. On the other hand, the closer x_1 is to the end-point, the smaller the prediction variance is, so that the optimal extrapolation design is a compromise between having small variance and small bias.

3. Let ξ_0 be the optimal robust design, let ξ_U be the four-point uniform design on $[-1, 1]$ with $\xi_U(\pm 1) = \xi_U(\pm 1/3) = 1/4$ and let ξ_V be the minimum integrated variance design given by Example 1. In Table 4, we give the efficiencies of ξ_0 relative to ξ_U and ξ_V. When the cubic model is correct , the LS estimate is unbiased. The efficiency of ξ_0 relative to ξ is defined by

$$re1(\xi) = \frac{IPV(\xi)}{IPV(\xi_0)},$$

with $IPV(\xi_U) = \frac{\sigma^2}{n}\sum_{v=0}^{3}(4\int_T L_v^2(x)dx)$, while $IPV(\xi_0)$ and $IPV(\xi_V)$ are given by (5). When the true response is only approximated by the cubic polynomial, the LS estimate becomes biased and the relative efficiency should be defined by

$$re2(\xi) = \frac{max_{f\in\mathcal{F}} IMSPE(f,\xi)}{max_{f\in\mathcal{F}} IMSPE(f,\xi_0)}.$$

From Table 4, one has the impression that whether there exists a contaminant term in the regression or not, ξ_0 is approximately as efficient as ξ_V when σ^2/n becomes large. Both of them are more efficient than the uniform design. But when σ^2/n is small and there is a contaminant term in the regression, ξ_0 is preferable.

Table 4. Relative efficiencies of minimax design

ξ_0 (when $\phi(x) \equiv 1$) versus the uniform design

and minimum variance design, with $t = 1.5$

σ^2/n	$re1(\xi_U)$	$re1(\xi_V)$	$re2(\xi_U)$	$re2(\xi_V)$
100	1.39	1	1.14	1
10	1.39	1	1.14	1
1	1.39	1	1.14	1
0.1	1.41	0.99	1.19	1.05
0.01	0.99	0.66	1.66	1.46

4. To compare the relative performances against uncertainty of the degree of the underlying polynomial, we consider the situation in this subsection when the contaminant term is $x^4 f(x) = dx^4$, with different values of d. In other words, we fit a cubic model, aiming to safeguard against a possible quartic model. In order to illustrate the effect of using a polynomial to approximate the true, possibly non-linear, model, we add two cases in Table 5: $f(x) = exp(-|x|/5)$

and $f(x) = exp(|x|/5)$. Table 5 gives the *IMSPE* values corresponding to the different designs we consider. Here, we assume $\sigma^2/n = 0.01$, $t = 1.5$ and ξ_0 is the minimax design given by Table 2.

Table 5. The values of IPV, IPB, IMSPE and $f(x)$

	ξ_U			ξ_V			ξ_0		
$f(x)$	*IPV*	*IPB*	*IMSPE*	*IPV*	*IPB*	*IMSPE*	*IPV*	*IPB*	*IMSPE*
0.1	0.38	0.02	0.40	0.33	0.01	0.34	0.50	0.01	0.51
0.5	0.38	0.39	0.77	0.33	0.35	0.68	0.50	0.24	0.74
1	0.38	1.57	1.95	0.33	1.38	1.71	0.50	0.95	1.45
5	0.38	39.28	39.66	0.33	34.55	34.88	0.50	23.69	24.19
10	0.38	157.11	157.49	0.33	138.22	138.55	0.50	94.76	95.26
$e^{-\|x\|/5}$	0.38	0.73	1.11	0.33	0.63	0.96	0.50	0.41	0.91
$e^{\|x\|/5}$	0.38	3.27	3.65	0.33	2.92	3.25	0.50	2.08	2.58

The results in Table 5 are to be expected, with ξ_0 performing well when d is large, and less well as d becomes small. In all cases, ξ_0 substantially reduces the bias. This indicates that when the departure reaches a certain magnitude, the model bias affect the choice of the designs for extrapolation and the robust extrapolation design ξ_0 is preferable.

REFERENCES

Box, G.E.P., and Draper, N.R. (1959). A basis for the selection of a response surface design. *Journal of American Statistical Association,* 54, 622 – 654.

Dette, H. and Studden W. J. (1997). The theory of canonical moments with applications in Statistics, Probability, and Analysis. Wiley, New York.

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, to appear.

Hoel, P.G. and Levine, A. (1964). Optimum spacing and weighting in polynomial prediction. *Annals of Mathematical Statistics.* 35, 1553 – 1560.

Huber, J. (1975). Robustness and designs. *In: J. N. Srivastava, Ed.,* A Survey of Statistical Design and Linear Models. North Holland, Amsterdam, pp. 287 – 303.

Huber, J. (1981). Robust statistics. Wiley, New York.

Kiefer, J. (1959). Optimal experimental designs (with discussion). *Journal of the Royal Statistical Society, Ser. B,* 21, 272 – 319.

Kiefer, J. and Wolfowitz J. (1964). Optimum extrapolation and interpolation designs, I. *Ann. Inst. Statist. Math.,* 16, 79 – 108.

King, J. and Wong, W. K. (1998). Optimal minimax designs for prediction in heteroscedastic models. *Journal of Statistical Planning and Inference,* 69, 371 – 383.

Lawless, J. F. (1984). Some problems concerning experimental designs for extrapolations. In: *Topics in Applied Statistics,* pp. 357 – 366. T. D. Dwivedi, Ed. Concordia University Press.

Liu, S. X. and Wiens, D. P. (1997). Robust designs for approximately polynomial regression. *Journal of Statistical Planning and Inference,* 64, 369 – 381.

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

Wiens D. P. (1992). Minimax designs for approximately linear regression. *Journal of Statistical Planning and Inference*, 31, 355 − 371.

Szegö. G. (1975). Orthogonal polynomials. *American Mathematical Society Colloquium Publications, Volume XXIII.*

Chapter Five

Applications

In the field of health risk assessment, more and more statisticians reveal their interests in the estimation of toxic effects of environmental chemicals at low exposure levels. See Krewski and Brown (1981) for a guide to the literature. It is pointed out by Crump (1979) that since direct estimates of effects associated with very low levels of exposure often require very large numbers of experimental subjects, such estimates are of necessity based on the downward extrapolation of results obtained at relatively high dose levels using a moderate number of subjects. Usually, people use a set of pre-selected dose levels and run experiments with equal numbers of subjects being tested at those levels. The numbers of subjects showing the response are recorded. Statisticians apply these data to fit the dose-response curve $P(x)$, where $P(x)$ represents the probability of subjects responding at dose x. Then we can apply this curve to estimate the life risk of a subject exposed to a very low dose level or inversely, to estimate a dose level x_0 such that $P(x_0)$ does not exceed a specific low value. The value of x_0 is called the virtually safe dose (VSD) (Cornfield (1977)).

Suppose that n_x subjects are allocated to the dose x and define $n := \sum_{\text{all } x \text{ used}} n_x$, the total number of subjects used in the experiment. Let R_x be the number of subjects showing the response at dose x, then R_x is binomially distributed with the number of trials n_x and the probability of success $P(x)$. At dose x, the ratio $p_x = R_x/n_x$ is obtained and transformed to the p_x-quantile

$$Y_x = F^{-1}(p_x)$$

for some cumulative distribution function F, in order to build the statistical model. Different choices of $F(x)$ give different models. The typical ones in the literature are the probit model (Finney, 1971) with $F(x)$ the standard normal distribution function, the logistic model (Ashton, 1972), with

$$F(x) = \frac{1}{1 + exp(-x)},$$

and the extreme value model (Gumbel, 1958), with

$$F(x) = 1 - exp\{-exp(x)\}.$$

Since

$$E(p_x) = P(x), \qquad Var(p_x) = \frac{P(x)(1 - P(x))}{n_x},$$

then the regression model can be defined as

$$E[Y|x] \approx F^{-1}(P(x)), \tag{1}$$

and

$$Var[Y|x] \approx \left(\frac{\partial F^{-1}}{\partial P(x)}\right)^2 \frac{P(x)(1 - P(x))}{n_x}. \tag{2}$$

The function $P(x)$ is unknown, so we have to consider a further approximation of $F^{-1}(P(x))$ in order to use the above model. $F^{-1}(P(x))$ is usually approximated by a polynomial. Thus, we obtain the model function

$$E[Y|x] \approx \theta_0 + \theta_1 x + \cdots + \theta_p x^p, \tag{3}$$

and the variance as (2). Then, with the (weighted) least squares estimates $\hat{\theta}_j$ of θ_j, $j = 0, 1, \cdots, p$, the estimate of $P(x)$ is given by

$$\hat{P}(x) = F\left(\sum_{j=0}^{p} \hat{\theta}_j x^j\right).$$

1 Approximate Polynomial Models

When $F(x)$ is the logistic distribution function, the logits of the proportions (p_x) of subjects responding at dose x is

$$Y_x = log\left(\frac{p_x}{1-p_x}\right).$$

To illustrate the fact that a polynomial is only an approximation to the logistic model, we use the first data set of Table 1 of Janardan (1995). The experiment was run to test the liver Hepatoma response of mouse to DDt. The dose units are ppm.

Table 1. Toxic Response Data

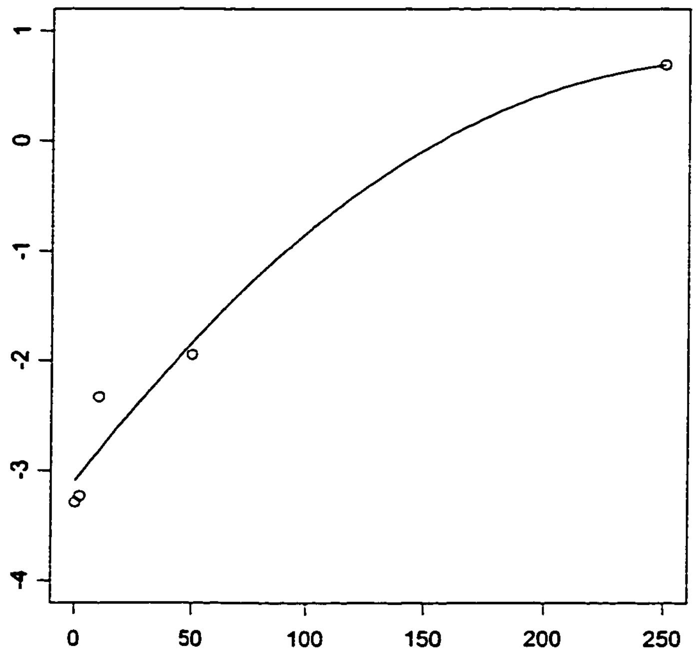
Doses	Animals tested	Animals with toxic response	Expected numbers from quadratic model
0	111	4	4.8
2	105	4	4.8
10	124	11	7
50	104	13	14.3
250	90	60	59.9

Figure 1 gives the plot of logits versus dose levels. The fit of logits to doses using the quadratic model gives

$$E[Y|x] = -3.091831 + 0.02757783x - 0.0000498x^2,$$

with the (residual sum of squares) $RSS = 0.3281413$. The last column of Table 1 gives the expected numbers of animals responding from the fitted model. These fitted numbers are not very close to the observed numbers except for the fitted number at dose = 250. Furthermore, we do not have observations at doses between 50 and 250. So we can not detect any possible departures from the assumed polynomial model in the interval $(50, 250)$. This requires us to construct the optimal design, with an eye on the possible violation of the assumed model.

Figure 1: Plot of logits vs doses: fit the data by quadratic model.



The above example and the approximate natures in (1), (2) and (3) make it reasonable to assume that the regression function of logits to doses is an approximate polynomial only. So (3) can be rewritten as

$$E[Y|x] = \theta_0 + \theta_1 x + \cdots + \theta_p x^p + f(x), \tag{4}$$

with f being unknown and belonging to some class. Thus, the model used in this dissertation seems quite appropriate for the low-dose extrapolation problems.

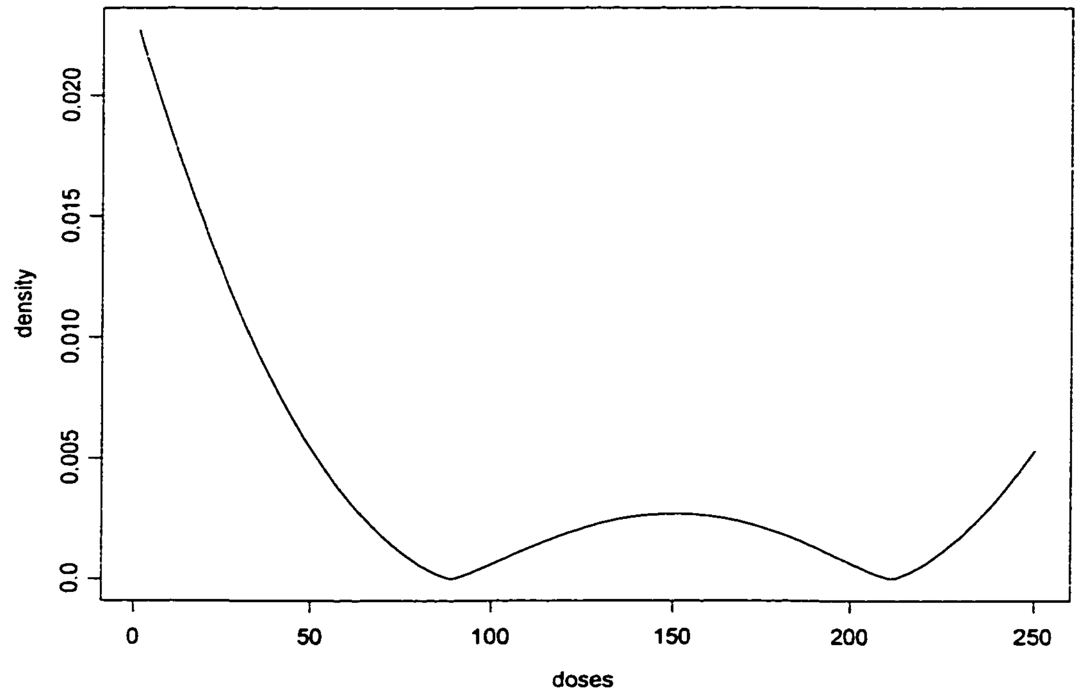
Some authors construct optimal designs for low-dose extrapolation by assuming that the model (3) is exactly correct, that is, ignoring the approximate nature of (3). See Little and Jebe (1975), Meeker and Hahn (1977) and Hoel and Jennrich (1979). Krewski, Bickis, Kovar and Arnold (1986) obtained optimal designs for low-dose problems with the assumption that $E[Y|x]$ was exactly linear in $\ln x$. However, these designs will be non-robust if the polynomial is only an approximation of $E[Y|x]$, because they ignore completely the effects of bias.

2 Optimal design for low-dose extrapolations

Lawless (1984) applied the model (4) to the low-dose problem. He assumed the model was nearly linear with a slight amount of curvature, constructing the designs which minimize the mean squared error of $\hat{Y}|_{x=0}$, for some specific model contamination function $f(x)$. In reality, however, we can not know what $f(x)$ exactly is. This suggests that we use the minimax approach to handle the design problem. In Chapters 2 and 3 of this dissertation, we assumed that $f(x)$ is bounded in $\mathcal{L}_2$-norm. We have two reasons to make this assumption. One is that the $\mathcal{L}_2$-class includes the typical model departures we might encounter in practical situations. The other is that the $\mathcal{L}_2$-class has good analytical properties which are useful for solving optimization problems.

In order to illustrate the application of the designs we obtained in Chapter 3 to the low-dose extrapolation problem, we assume an approximately quadratic model

Figure 2: Optimal design density of DDt data



to the DDt data, with design space as interval $[1, 250]$ and extrapolation point as $t = 0.5$. The solution to P3) gives the design ξ_0, with density

$$k_0(x) = [0.00353899 - 0.0001129439x + (1.28 \times 10^{-6})x^2 - (10^{-9})x^3 + (9.99 \times 10^{-12})x^4]^{2/3},$$

which is plotted in Figure 2.

To implement this design, one can choose the modes $< 1, 150.2, 250 >$ of the density as the sampling sites. But this choice does not allow us to test the model adequacy. We do not recommend it for robustness reasons. Another approach discussed in the dissertation fixes the number r of sampling sites and determine these sites by $K_0^{-1}(\frac{i-1}{r-1})$, $i = 1, 2, ..., r$, where $K_0(x)$ is the cumulative distribution function of ξ_0. Then one makes equal numbers of observations at these sites as experimenters did in the animal experiments described previously. For the DDt experiment with $n = 534$ mice, if we test the animal responses at $r = 6$ dose levels, then these levels given by our design are

$$< 1, 10.79, 23.59, 43.6, 142.79, 250 >,$$

which includes one point in the interval $(50, 250)$. We do not have a unique way for choosing r. Common choices are a factor of n or the value of $[\sqrt{n}]$. The problem of choosing an optimal value of r remains open.

3 Conclusion

The model function we build to fit the experimental data is typically only approximately correct. And the model in the extrapolation region might differ from that in the design region. When the model is inadequate, the estimation of the regression response is biased. So the designs obtained by minimizing the variance do not make sense and it will be dangerous to apply these designs to estimate the response function.

The final point is that the minimum-variance designs always make observations at $(p + 1)$ sites only when the assumed model is a p-degree polynomial, and thus we can not check the validity of the model. Although the robust designs we give are continuous, thus complex to implement, the design densities have mass spread throughout the design interval. This suggests that we make observations on enough sites so that we can reduce the bias of estimation and test the adequacy of the model used.

References

Ashton, W. D. (1972). The Logit Transformation. *Griffin, London.*

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

Crump, K. S. (1979). Dose Response Problems in Carcinogenesis. *Biometrics* 35, 157 – 167.

Finney, D. J. (1971). Probit Analysis (3rd edition). *Cambridge University Press, London.*

Gumbel, E. J. (1958). Statistics of Extremes. *Columbia University Press, New York.*

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

Janardan, K. G. (1995). Environmental Risk Assessment Model for Low Dose Extrapolation. *Communications in Statistics - Theory and Methods* 24(10), 2621-2633.

Krewski, D. and Brown, C. (1981). Carcinogenic Risk Assessment: A Guide to the Literature. *Biometrics* 37, 353 – 366.

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 Mathematics* 29, 245-262.

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

Little, R. E. and Jebe, E. H. (1975). Statistical Design of Fatigue Experiments. *Applied Science Publishers, London.*

Meeker Jr., W. Q. and Hahn G. J. (1977). Asymptotically Optimum Over-stress Tests to Estimate the Survival Probability at a Condition With a Low Expected Failure Probability (with discussion). *Techometrics* 19, 381 – 404.