University of Alberta

Robust Prediction and Extrapolation Designs, with Applications to Accelerated Life Testing

by

Xiaojian Xu 

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 and Statistical Sciences

Edmonton, Alberta

Fall 2006

Your file Votre référence
ISBN: 978-0-494-23129-6
Our file Notre référence
ISBN: 978-0-494-23129-6

ABSTRACT

We consider the construction of robust prediction and extrapolation designs for a misspecified generalized linear response. Possible violations from homoscedasticity are also taken into account.

Firstly, we find minimax designs and corresponding optimal regression weights in the context of the following problems:

1. For nonlinear ordinary least squares (OLS) estimation with homoscedasticity, determine: (a) a prediction design to minimize the maximum value of the integrated mean squared prediction error (IMSPE), (b) a one-point extrapolation design to minimize the maximum value of the mean squared extrapolation error (MSEE), and (c) a general extrapolation design to minimize the maximum value of the integrated mean squared prediction error (IMSEE), with the maximum being evaluated over the possible departures (from assumed model and homoscedasticity) from the response function;

2. For nonlinear OLS estimation with heteroscedasticity, determine (a), (b), and (c), with the maximum being evaluated over both the departure from assumed response function and that from homoscedasticity;

3. For nonlinear weighted least squares estimation, determine both weights and a design to minimize the maximum IMSPE, MSEE, or IMSEE;

4. Choose weights and design points to minimize the maximum IMSPE, MSEE, or IMSEE, subject to design unbiasedness.

Secondly, we present the construction of robust designs when the data are censored. The minimax designs are found for maximum likelihood estimation in the

context of both prediction and extrapolation problems in case of with or without restraint of design unbiasedness.

Our study extends the previous work of others in three aspects: (i) by considering a nonlinear fitted regression response; (ii) by taking a rather general design and extrapolation spaces; and (iii) most significantly, by dropping all restrictions on the structure of the regressors. Solutions are derived by a nonsmooth optimization technique analytically and given in complete generality. Numerical comparisons indicate that our designs perform well in combining robustness and efficiency and applications to accelerated life testing are highlighted.

Lastly, we discuss the application of our designs to a real life dose-response experiment. We propose two new implementation schemes for approximating a continuous design, and also confirm that one of the matching quantile schemes used in the literature is optimal with respect to certain criteria.

ACKNOWLEDGEMENTS

First of all, I am deeply grateful to my supervisor, Dr. Douglas Wiens for his inspirational instruction, tremendous support and invaluable guidance. It is him who introduced me to the area of robust designs and gave me the impetus to study accelerated life testing. I feel honoured to work with him. I believe that the research experience I have gained during my study with him has had a profound influence and impact upon my graduate education and my academic career.

My sincere thanks are extended to other members of my examining committee: Dr. Edit Gombay, Dr. Byron Schmuland, Dr. Ming Jian Zuo, Dr. N.G.Narasimha Prasad, and Dr. William Notz for their thorough review, insightful comments, and helpful suggestions. I also greatly appreciate Dr. Notz's complimentary comments on my thesis work and his editorial corrections.

I am grateful to Dr. Eric Woolgar, for his encouragement, and understanding during my study especially when I was confused. He always listened patiently and provided support. His inspirational advice and his prompt help with my writing were also greatly appreciated. Many thanks also go to Dr. Ming Jian Zuo for his support upon numerous incidences; to Dr. Michael Li for being my teaching mentor and his advice during my doctoral program; to Dr. Gerald Cliff for the helpful discussion and referential guide on matrix properties.

I also would like to thank Ms. Dona Guelzow for facilitating my doctoral study. Her warm presence made my study enjoyable.

I extend my deepest gratitude and appreciation to many friends for their assistance and encouragement, for their understanding throughout my doctoral study, to mention but a few, Jane and Peter Harr, Liping Liu, and Anne Marie and Paul Milazzo. I am grateful to all of my friends, who gave me precious spiritual support with warm words, emails, calls, and cards which helped me through the ups and downs of this journey.

My words cannot adequately express what I feel from the bottom of my heart. Without the help and support of the people above, this dissertation could not have been accomplished.

This work and the conference travels during my study are supported by Dr. Douglas Wiens, Department of Mathematical and Statistical Department, Province of Alberta Graduate Fellowship, and Faculty of Graduate Studies and Research. I am thankful.

DEDICATION

To my parents:

Xu Ming Zhong and Chang Lin Chuan,

my husband Chunlei, and my daughter Olivia,

who made all of this possible,

for their unending encouragement, for their unfailing support, and for their belief in me.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

In this dissertation we study the construction of optimal regression experimental designs, for both prediction and extrapolation of a response, when the regression model is misspecified. Throughout the dissertation we divide our work into three different types of problems: prediction problems, one-point extrapolation problems, and general extrapolation problems. For a prediction problem, our goal is the response estimation throughout the design space in which the design (support) points can be chosen. For an extrapolation problem, we are interested in extrapolation of a response to an extrapolation space. If such an extrapolation space consists of just one point outside the design space, we call it a one-point extrapolation problem. If the extrapolation space has nonzero Lebesgue measure, we name it a general extrapolation problem. This work contains five chapters. Three of them, Chapter 2, Chapter 3, and Chapter 4, are independent papers which have been prepared for publication. The robust one-point extrapolation designs are studied in Chapter 2 while the robust prediction designs and robust general extrapolation designs are provided in Chapter 3. In addition, we construct the robust optimal designs incorporating censored data in Chapter 4. Chapter 5 contains implementation methods of continuous designs, and applications of the optimal designs constructed in this work.

Most previous work for robust optimal designs has been done for linear regression models; however, this dissertation treats nonlinear regression as well. We take into account not only the possible misspecification in assumed regression response but also the possible departure from assumed homoscedasticity. Robustness of the designs on the parameter-dependency problems in nonlinear regression is also addressed.

This present chapter provides a review of the literature regarding the previous work in direction of (1) optimal prediction designs, (2) optimal extrapolation designs, (3) robust optimal designs against various situations, and (4) robust optimal designs for accelerated life testing (ALT) and censored data. Several sections are presented in this chapter. Section 1.1 outlines the purpose of our study; it presents the motivation of the dissertation that involves an application of life testing and the consideration of nonlinear regression models. Section 1.2 reviews classical optimal designs for estimation, prediction, and extrapolation problems. It also gives the classical optimality criteria and some classical optimal design examples. Section 1.3 reviews the robust optimal designs against various situations. The robust designs, particularly in life testing literature, are briefly introduced. Some examples of the robust optimal designs which lead to our current study are detailed. Finally, Section 1.4 summarizes the results of this dissertation.

1.1 Purpose of the Study

1.1.1 Motivation

Nowadays more and more products have high reliability. The time consumed in life testing for such products at normal conditions is exorbitant. A common approach to this problem is that of acceleration. ALT provides timely information about a product's life by testing a sample of units at higher than usual stress levels and extrapolating through an assumed statistical model to estimate life at a lower stress level anticipated in practice. Since this extrapolation is to one point outside of testing stress space (experimental design space), it is a one-point extrapolation problem.

Fang and Wiens (1999) point out that "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 example, suppose design space is $S = [a, b]$. The experimenter tries to fit a

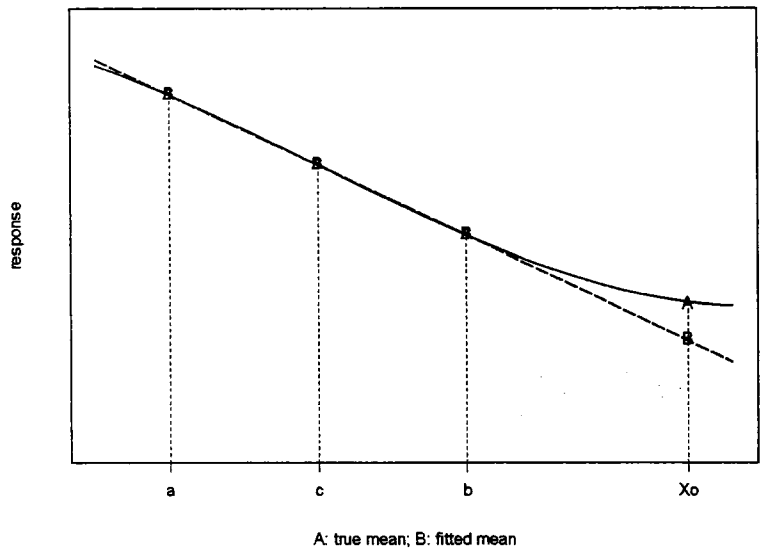

Figure 1: Misspecified simple linear regression. A solid line represents the true regression response while a broken line represents the fitted regression response.

straight line regression model and then extrapolate the fitted line to a point x_0 outside S. There are a few problems that could possibly occur. Firstly, the extrapolation in estimation is biased at extrapolation point x_0 although the fit within S seems good. See Figure 1 for a demonstration of the danger of extrapolation. Secondly, the departure from the assumed model cannot even be detected. Furthermore, there is no information about what kind of departure would occur. Therefore, we are motivated to explore a "good" design which takes into the consideration of possible but unknown departure from the model which one is fitting.

1.1.2 Extension from Linear to Generalized Linear Response

This study also aims to treat nonlinear regression. Let us examine a couple of nonlinear regression examples. First of all, we consider an accelerated failure model, which is commonly used in life testing and survival analysis (Hosmer and Lemeshow, 1998)

for modelling the relationship between lifetime Y and the stress variable x:

$$E(Y|x) = \exp(\theta_0 + \theta_1 x). \tag{1.1}$$

The second example is a dose-response curve for modelling the probability $P(x)$ of showing an adverse effect due to a dose at level x:

$$P(x) = 1 - \exp\left(-\sum_{j=0}^{k} \theta_j x^j\right). \tag{1.2}$$

This model is frequently applied in the field of cancer dose response. Either (1.1) or (1.2) is a monotonic function of a linear function of unknown parameters and known regressors. It is termed the generalized linear regression.

There is a wealth of literature regarding robust optimal regression designs against misspecification in a linear relationship

$$E(Y|\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta} \tag{1.3}$$

between a response variable Y and an experimental vector $\mathbf{x}$, where $\boldsymbol{\theta}$ is the vector of unknown parameters and $\mathbf{z}^T(\mathbf{x}) = (z_1(\mathbf{x}), z_2(\mathbf{x}), ..., z_p(\mathbf{x}))$ is the vector of regressors, depending on $\mathbf{x}$. However, there seem to be relatively few robust designs dealing with misspecification in a nonlinear regression. This present study attempts to extend the previous work on (1.3) to its generalized version

$$E(Y|\mathbf{x}) = h\left(\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}\right)$$

for some monotonic function h.

1.1.3 Address both ALT and Censoring

With regard to life testing, the two most commonly used testing time saving plans are acceleration and censoring, both of which stimulate our study. When life testing runs at the stress levels within the range of product normally in use, our goal is prediction. Namely we are interested in the estimation of mean response throughout the region

of interest. For ALT, in which products are tested at higher than normal usage stress levels, the goal is extrapolation. For extrapolation problems, if one is interested in estimating the lifetime at a certain range of normal usage stress levels, we classify it as a general extrapolation problem; if one is interested in estimating the lifetime at a particular normal usage stress level which is lower than testing stress levels, it is a one-point extrapolation problem. Therefore, three types of designs are considered under this dissertation study to address these three types of problems.

Tests yielding complete data generally take too long to run especially for those products having long life-spans. To save testing time, the test may be censored in order that the result can be analyzed before all units fail. Subsequently, the data obtained from such testing plans will consist of lifetime information on unfailed units (so called "censored data"). Another purpose of censoring is to permit one to analyze the most recent test data while the test is still running. The robust optimal designs incorporating censoring are also a part of our study.

1.1.4 Remove the Restrictions

There is another issue that we have considered in this dissertation. In the robust design literature, due to the intractability of obtaining an explicit design there are often certain restrictions put on the regressors' structure, on the design space, and additionally on the extrapolation space for general extrapolation problems. For instance, the minimax explicit design problems for high degree polynomials are not tractable as indicated in Heo, Schmuland, and Wiens (2001). This dissertation aims to provide a way of constructing a robust optimal design without such restrictions.

1.2 Classical Optimal Designs

1.2.1 Classical Design Problems

Suppose that the linear regression model (1.3) with $\mathbf{x} \in \mathcal{R}^q$, $\boldsymbol{\theta} \in \mathcal{R}^{p+1}$, and $\mathbf{z}(\mathbf{x}) \in \mathcal{R}^{p+1}$ a vector of regressors is considered. In order to estimate $\boldsymbol{\theta}$ and explain certain

aspect of this model, 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

$$y_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\theta} + \varepsilon_i, \quad i = 1, 2, ..., n, \tag{1.4}$$

where the ε_i's are random errors. Two classical assumptions made to this regression model 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 common variance σ^2.

Define $\mathbf{X} = \left(\mathbf{z}^T(\mathbf{x}_1), \mathbf{z}^T(\mathbf{x}_2), ..., \mathbf{z}^T(\mathbf{x}_n)\right)^T$ and $\mathbf{y} = (y_1, y_2, ..., y_n)^T$. Under assumptions A1, A2 and that of $\mathbf{X}^T\mathbf{X}$ being nonsingular, the least squares (LS) estimate $\hat{\boldsymbol{\theta}}_{LS}$ given by

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

is the best estimate of $\boldsymbol{\theta}$ among all linear unbiased estimates, and the covariance matrix of $\hat{\boldsymbol{\theta}}_{LS}$ is the smallest in the sense of the Loewner ordering of nonnegative definite 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. As we know, $\hat{\boldsymbol{\theta}}_{LS}$ has covariance matrix

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

which only depends on $\mathbf{X}$. Hence, the problem of experimental design is to choose the appropriate $\mathbf{X}$ in order that $COV(\hat{\boldsymbol{\theta}}_{LS})$ will be as "small" as possible. The matrix $COV(\hat{\boldsymbol{\theta}}_{LS})$ being "small" means that an appropriate real-valued function of it is small. Different functions will give different minimization criteria, as seen in Section 1.2.3.

1.2.2 Approximate Design Theory

Suppose we plan or are allowed to perform an experiment for the model (1.4) by choosing the observation sites $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ from a design space $S \subset \mathcal{R}^q$. In practice, a design should tell the experimenter what the design support points are and how many subjects should be allocated to each of these points. An implementable design, ξ, must be a discrete probability measure which puts probability $\xi(\mathbf{x}_i)$ at $\mathbf{x}_i$. For a given $r \leq n$, let $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_r$ be the distinct design points, the number of subjects allocated to a particular design point $\mathbf{x}_i$ on design space is then, $n\xi(\mathbf{x}_i)$. Thus, in addition, each $\xi(\mathbf{x}_i)$, $i = 1, 2, ..., r$, should be an integer multiple of n^{-1}. A design with this integer property is called an exact design, but in general the exact design problem is mathematically intractable. The situation is analogous to the much simpler one where we wish to maximize a function defined on the integers. Because of discrete domain, calculus techniques cannot be exploited in this situation. A commonly used device for this simpler problem is to extend the definition of the function to all real numbers and use calculus to find the maximizing real number; and then to argue that the maximum of the function over integers occurs at an integer adjacent to this maximizing real number. This idea is adapted for the design problem and leads to what Kiefer has termed "approximate design theory". Approximate design theory extends the class of designs, allowing it to be any probability distribution on S. Define Ξ to be the set of all probability distributions on S. We seek an optimal design, ξ_*, within Ξ and hope that an exact design which approximates ξ_* will be close to optimal. This approach is adopted by us throughout this work as well.

1.2.3 Optimality Criteria

Let ξ be a design measure, that is, the empirical distribution function of $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$:

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

with $I_{\mathbf{x}}$ representing the point mass 1 at $\mathbf{x}$, for any $\mathbf{x} \in S$. For any $\xi \in \Xi$, define a matrix

$$\mathbf{B}_{\xi} = \int_{S} \mathbf{z}(\mathbf{x})\mathbf{z}^{T}(\mathbf{x})d\xi.$$

It is a nonnegative definite $(p+1) \times (p+1)$ matrix. Since $\mathbf{B}_{\xi} = \frac{1}{n}\left(\mathbf{X}^{T}\mathbf{X}\right)$, we have $COV(\hat{\boldsymbol{\theta}}_{LS}) = \frac{\sigma^2}{n}\mathbf{B}_{\xi}^{-1}$. The classical design problem is to find an optimal design in order that an appropriate scalar function of $\mathbf{B}_{\xi}^{-1}$ will be minimal. Such a scalar function is called an optimality criteria or a loss function.

When we are interested in finding the best estimates for the coefficients in regression function, the optimal design which we are searching for is normally intended to minimize the generalized variance of parameter estimates. We call such design an estimation design. The commonly used loss functions for searching for an estimation design include the determinant, the trace, and the largest eigenvalue of $\mathbf{B}_{\xi}^{-1}$, and these give the D-, A-, and E-optimality criterion respectively. Kiefer (1974) generalizes these criteria in terms of the sum of the certain powers of the eigenvalues of $\mathbf{B}_{\xi}^{-1}$. Dette, Heiligers, and Studden (1995) investigate the geometric structure of a class of minimax optimality criteria containing Kiefer's generalized criteria as special cases. When we are interested in the best estimation of the response function, the design problems are to find the optimal prediction designs. Such designs are found to minimize the variance of predicted response. With $d(\mathbf{x},\xi) = \mathbf{z}^{T}(\mathbf{x})\mathbf{B}_{\xi}^{-1}\mathbf{z}(\mathbf{x})$, for a prediction design the loss function normally is a function of $d(\mathbf{x},\xi)$. The two commonly used ones are maximum: $\max_{x \in S} d(\mathbf{x},\xi)$ and average: $\int_{S} d(\mathbf{x},\xi)d\mathbf{x}$ over design space S. They provide the G-, and Q- (or I-) optimality criterion respectively. For extrapolation designs, the loss functions considered are the maximum and the average of $d(\mathbf{x},\xi)$ over extrapolation space T instead of S. We still call them the G-, and Q- (or I-) optimality criterion. For nonlinear regression, the covariance matrix depends on not only the design but also the unknown parameter values. Therefore,

normally the optimal design minimizing the corresponding loss function for nonlinear regression depends on the unknown parameter values. The corresponding criteria are called D_{θ}-, A_{θ}-, E_{θ}-, G_{θ}, and Q_{θ}- (or I_{θ}-) optimality. For details, see Silvey (1980) who provides us with an elegant reference to the theory of classical optimal designs.

1.2.4 Examples

Example 1: Classical optimal designs for polynomial regression.

Guest (1958) presents the D-optimal design for the polynomial regression. For the p^{th} degree polynomial regression model

$$E(Y|x,\boldsymbol{\theta}) = \theta_0 + \theta_1 x + ... + \theta_p x^p$$

with a parameter vector, $\boldsymbol{\theta}^T = (\theta_0, \theta_1, ..., \theta_p)$, the D-optimal design has $(p+1)$ design points with an equal mass of $1/(p+1)$. These design points are the solution of the following equations

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

where $P_p(x)$ is the p^{th} Legendre polynomial. Let $S = [-1,1]$. For instance, when $p = 3$ the design points are ± 1 and ± 0.447; when $p = 4$ they are 0, ± 1, and ± 0.665.

Example 2: Classical optimal extrapolation designs for polynomial regression.

The G-optimal extrapolation designs for polynomial regression are found by Hoel and Levine (1964). For $S = [-1,1]$, $T = [1,t]$ or $T = \{t\}$ with $t > 1$, and $\mathbf{z}(x) = (1,\ x,\ x^2,\ ...,\ x^p)^T$, the G-optimal extrapolation design is supported by $p+1$ Chebyshev points

$$x_i = -\cos\left(\frac{i\pi}{p}\right),\ i = 0,\ 1, ...,\ p,$$

and the mass on each point is proportional to the absolute value of the corresponding Lagrange interpolation polynomial at this specific point. Namely,

$$\xi_G(x_i) = \frac{|L_i(t)|}{\Sigma_{i=0}^{p}|L_i(t)|}$$

where

$$L_i(t) = \frac{(t-x_0)\ldots(t-x_{i-1})\,(t-x_{i+1})\ldots(t-x_p)}{(x_i-x_0)\ldots(x_i-x_{i-1})\,(x_i-x_{i+1})\ldots(x_i-x_p)}.$$

Example 3: Classical optimal designs for generalized linear regression.

Ford (1976) considers the optimal design problem for nonlinear regression. Suppose that the response Y takes values 0 and 1; the independent variable x is continuous; the parameter vector is $\boldsymbol{\theta}^T = (\theta_0, \theta_1)$, and

$$P(Y=1|x,\boldsymbol{\theta}) = \exp(\theta_0+\theta_1 x)\,/\,[1+\exp(\theta_0+\theta_1 x)]$$

with $\theta_0 > 0$ and $\theta_1 > 0$. This is a generalized linear regression model. Let the design space be $S = [-1, 1]$. Ford has shown that the $D_{\boldsymbol{\theta}}$-optimal design measure is supported on two points, each having probability of 0.5. The support points depend on $\boldsymbol{\theta}$ in the following way.

Let $a \approx 1.5434$ be the positive solution of the equation $e^z = (z+1)/(z-1)$. Also, we make use of the classes below:

$$\begin{aligned}
\mathcal{T}_1 &= \{\boldsymbol{\theta} \mid \theta_1 - \theta_0 \geq a,\ \theta_0 > 0,\ \theta_1 > 0\},\\
\mathcal{T}_2 &= \left\{\boldsymbol{\theta} \,\middle|\, \theta_1 - \theta_0 < a,\ \theta_0 > 0,\ \theta_1 > 0,\ \exp(\theta_0+\theta_1) \leq \frac{\theta_1+1}{\theta_1-1}\right\},\\
\mathcal{T}_3 &= \left\{\boldsymbol{\theta} \,\middle|\, \theta_0 > 0,\ \theta_1 > 0,\ \exp(\theta_0+\theta_1) > \frac{\theta_1+1}{\theta_1-1}\right\}.
\end{aligned}$$

Then, the support points are

(i) $\frac{a-\theta_0}{\theta_1}$ and $\frac{-a-\theta_0}{\theta_1}$, if $\boldsymbol{\theta} \in \mathcal{T}_1$;

(ii) -1 and x^*, where x^* is the solution of $\exp(\theta_0+\theta_1 x) = \frac{2+(x+1)\theta_1}{-2+(x+1)\theta_1}$, if $\boldsymbol{\theta} \in \mathcal{T}_2$;

(iii) -1 and 1, if $\boldsymbol{\theta} \in \mathcal{T}_3$.

The locations of the support points depend on $\boldsymbol{\theta}$. This dependency may be roughly summarized by saying that the more nearly linear is the response curve in the interval $[-1, 1]$, the further towards -1 and $+1$ are the support points pushed.

1.3 Robust Designs

1.3.1 Robustness against Various Situations

The classical optimal designs are sensitive to model assumptions. Box and Draper (1975) summarized fourteen criteria which can be used to judge a design for fitting a regression function. One of these is described as to "be insensitive to wild observations and to violation of the usual theory assumptions." Robust designs are required to have this particular property. In the literature, robust designs are constructed to safeguard against various situations which the experimenter should watch out for. These situations are listed by us as follows:

(1) misspecification in a regression function;

(2) violation of homoscedasticity;

(3) misspecified parameter values which the constructed design depends upon;

(4) violation of independence of observations;

(5) misspecification of an underlying distribution;

(6) wild observations.

For (4), see Wiens and Zhou (1996, 1999), Zhou (2001), and also the references therein. For (5), see Chaloner and Larntz (1992), and more recently Pascual and Montepiedra (2003). For (6), Box and Draper (1975) present a measure of insensitivity to wild observations so that the robust designs with protection from wild observations can be determined by minimizing such a measurement.

Our study addresses the situations of (1), (2), and (3). The literature on these three situations is described in the following three subsections.

1.3.2 Robust Design against (1)

In light of Box and Draper (1959), the optimality criteria using the mean squared error matrix are broadly adopted. This is the sum of one term related to the variance of the estimates of the unknown parameters in the regression model and the other

term related to the bias present because of fitting an incorrect model. Many authors have dealt with the situation wherein the underlying regression model is assumed to be some low degree polynomial, and where the possible departures from this model range from a higher degree polynomial to a very large class of functions. In general, the regression model under the consideration of robustness is assumed to be approximately known. The true model is a contaminated version of a fitted model (1.4). It can be written as

$$y_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\theta} + f(\mathbf{x}_i) + \varepsilon_i, \quad i = 1, 2, ..., n, \tag{1.5}$$

where the random errors are uncorrelated and have homogeneous variances, and f represents an unknown contaminant which usually belongs to a predefined contamination class $\mathcal{F}$. Due to the presence of f, the least squares estimator $\hat{\boldsymbol{\theta}}_{LS}$ is no longer unbiased. Its bias vector and covariance matrix are

$$E\left(\hat{\boldsymbol{\theta}}_{LS}\right) - \boldsymbol{\theta} = \mathbf{B}_\xi^{-1}\mathbf{b}_{f,S}, \text{ and } COV(\hat{\boldsymbol{\theta}}_{LS}) = \frac{\sigma^2}{n}\mathbf{B}_\xi^{-1},$$

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

$$MSE(f,\xi) = \mathbf{B}_\xi^{-1}\mathbf{b}_{f,S}\mathbf{b}_{f,S}^T\mathbf{B}_\xi^{-1} + \frac{\sigma^2}{n}\mathbf{B}_\xi^{-1}.$$

To obtain the optimal designs under model (1.5), the loss functions are naturally chosen by replacing the covariance matrix with the mean squared error matrix $MSE(f,\xi)$. Since $MSE(f,\xi)$ involves the unknown contaminant f, we use $\max_{f\in\mathcal{F}} MSE(f,\xi)$ to safeguard against the worst possible situation. Therefore the determinant, the trace, and the largest eigenvalue of $\max_{f\in\mathcal{F}} MSE(f,\xi)$ give the D-, A-, E-optimality criterion respectively. Let $d_m(\mathbf{x},\xi) = \max_{f\in\mathcal{F}} \mathbf{z}^T(\mathbf{x})MSE(f,\xi)\mathbf{z}(\mathbf{x})$, the loss functions: $\max_{x\in S} d_m(\mathbf{x},\xi)$ and $\int_S d_m(\mathbf{x},\xi)d\mathbf{x}$ provide the G-optimality and Q- (or I-) optimality criterion respectively. Most recently Adewale and Wiens (2006) discuss a new criterion of robust optimal designs. It replaces the maximum of $MSE(f,\xi)$ with the average of $MSE(f,\xi)$ over $\mathcal{F}$.

These robust optimal designs differ in the choice of the contamination class $\mathcal{F}$ and the criterion of optimality. Box and Draper (1959) and Kiefer (1973) considered that the true regression function is a polynomial of possibly higher degree than that of the assumed one. Their attention is on finite dimensional $\mathcal{F}$. There are two major types of less restricted contamination classes, infinite dimensional $\mathcal{F}$, used in the literature. They are:

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

where η_S is a constant and assumed "small" so that the linear term in (1.5) is still dominant; and

$$\mathcal{F}_2 = \left\{ f : \ |f(\mathbf{x})| \leq \varphi(\mathbf{x}) \text{ for any } \mathbf{x} \in S, \int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})d\mathbf{x} = \mathbf{0} \right\},$$

with various assumptions about φ. The first class $\mathcal{F}_1$ is used in Huber (1975, 1981), Wiens (1990, 1991, 1992, 1993, 1994, 1996, 1998) while the second $\mathcal{F}_2$ is used in Marcus and Sacks (1976), Li and Notz (1982), Pesotchinsky (1982), Li (1984), and Liu and Wiens (1997). The first class is so full that the optimal designs obtained are continuous and need to be approximated by a discrete design prior to implementation. However, the second class is too thin, with the result that the robust designs found generally have a small number of support points and thus do not allow exploration of models larger than the fitted one. The review by Chang and Notz (1996) gives a summary of the previous work in this subject. It is still an open problem to define a compromise contamination class to accommodate both implementation convenience and model-exploration enhancement. In an innovative way, Yue and Hickernell (1999) introduce a class $\mathcal{F}$ which is a reproducing kernel Hilbert space admitting a reproducing kernel function defined on $S \times S$; they have shown that under this contamination class, the continuous uniform design on S is still the bias minimizing design.

For some resulting robust optimal designs regarding different optimality criteria,

see Pesotchinsky (1982) and Wiens (1992). Both papers address multiple linear regression, i.e. $\mathbf{z}(\mathbf{x}) = \left(1, \mathbf{x}^T\right)$. Wiens (1992) constructs the robust D-, A-, E-, Q- and G-optimal designs under $\mathcal{F}_1$. Pesotchinsky (1982) was interested in the best estimates for the coefficients in the regression function and obtained the D-, A-, and E-optimal designs under $\mathcal{F}_2$.

1.3.3 Robust Design against both (1) and (2)

To the author's knowledge, Wiens (1998) is the first paper which takes the consideration of robustness against possible violation of homoscedasticity assumption. He assumes the model (1.5) where the random errors, although uncorrelated with mean 0, are possibly heteroscedastic with

$$var\left[\varepsilon\left(\mathbf{x}\right)\right] = \sigma^2 g\left(\mathbf{x}\right),$$

where $g\left(\mathbf{x}\right)$ is an unknown member of a fluctuation class $\mathcal{G}$. Prior to this paper, many authors such as Wong (1992), Wong and Cook (1993), and Dasgupta, Mukhopadhyay, and Studden (1992) constructed optimal designs for heteroscedastic regression models with *known* efficiency functions $\lambda\left(\mathbf{x}\right) = g\left(\mathbf{x}\right)^{-1}$. Wiens (1998) defines

$$\mathcal{G} = \left\{ g : \int_S g\left(\mathbf{x}\right) d\mathbf{x} \leq \int_S d\mathbf{x} \right\},$$

and constructs the robust Q-optimal prediction designs in the presence of both possible variance fluctuation within $\mathcal{G}$ and possible regression contamination within $\mathcal{F}_1$. He considers the following problems:

P1) For ordinary least squares (OLS), determine a design to minimize the maximum of integrated mean squared error (IMSE), $\int_S \mathbf{z}^T(\mathbf{x}) MSE(f, g, \xi) \mathbf{z}(\mathbf{x})$, over f and g;

P2) For weighted least squares (WLS), 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.

The solutions for both P1) and P2) are found for multiple linear regression with no interaction and with a spherical design space. The solution to P3) is given in complete generality without restrictions on regressors and design space. This dissertation will extend these results for P1) and P2) by removing the restrictions on both regressors and design space in Chapter 3.

Fang and Wiens (1999) extend the work of Wiens (1998) to general extrapolation problems. They consider the same model as (1.5), same fluctuation class $\mathcal{G}$, and a similar contamination class to $\mathcal{F}_1$ but add on a condition for the contaminant on the extrapolation space T :

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

where η_T is a constant. T is assumed to have nonzero Lebesgue measure and be disjoint from S. They construct the robust Q-optimal extrapolation designs in the presence of both possible variance fluctuation within $\mathcal{G}$ and possible regression contamination within $\mathcal{F}_1^E$. The corresponding problems to P1)-P3) are addressed. Similarly, for this general extrapolation case, the solutions to both P1) and P2) are found for multiple linear regression with no interaction, with a spherical design space, and with an annular extrapolation space. The solution to P3) is given in complete generality without restrictions on regressors, design space and extrapolation space. This dissertation will extend these results for P1) and P2) by removing the restrictions on regressors' structure, design space, and extrapolation space in Chapter 3. It will also extend these results for P1), P2) and P3) to the one-point extrapolation case, where the extrapolation space has Lebesgue measure zero, in Chapter 2.

1.3.4 Robust Design against (3)

There are a few situations wherein the optimal designs generally depend on unknown parameters which we aim to estimate; such designs are called "locally optimal". These situations are:

(1) For a nonlinear regression model.

(2) For a heteroscedastic linear model with a variance function depending on unknown parameters (for instance, the variance of a response is an exponential or a power function of its mean).

(3) For a linear regression yet with the purpose of estimating a nonlinear aspect, such as the ratio of two parameters.

Several strategies such as a Bayes approach, minimax scheme, and sequential procedures are used to overcome the parameter-dependency of the designs. For (3), we refer to Müller (1995). She uses the maximin efficiency criterion to find the maximin optimal designs which maximize the minimum relative efficiency over a parameter space. For (2), we refer to Dasgupta, Mukhopadhyay and Studden (1992). They assume the parameter vector has a suitable specified prior distribution so that the Bayes optimal designs can be obtained. However, most designs found in the literature which deal with the parameter-dependency problem are constructed for (1).

For nonlinear regression, Ford, Titterington, and Kitsos (1989) present various static and sequential designs for nonlinear models without the consideration of model uncertainty. Sinha and Wiens (2002) have employed notions of robustness in the construction of sequential designs for approximately specified nonlinear models. However, in some applications - ALT for example - sequential designs are not feasible (Ford, Titterington, and Kitsos, 1989); hence our focus in this dissertation is on static designs and therefore the minimax approach (Silvey, 1980) is adopted to find the locally most robust optimal design for the least favourable parameter value over a parameter region.

1.3.5 Robust Designs for ALT and Censored Data

The resulting robust optimal designs seen in the literature vary according to a number of factors. For instance, (1) the contamination class which gives the range of

departures; (2) optimality criterion; (3) estimation method used for estimating the quantity of interest such as least squares estimators (Wiens, 1998), linear estimator (Li and Notz, 1982), M-estimator (Wiens, 1994); (4) the aspect of interest: estimation of the regression parameters, prediction of the response function, or extrapolation; a linear aspect such as the difference between two treatment effects (Heckman, 1987), a nonlinear aspect such as the ratio of two regression parameters (Müller, 1995). Lastly, the design also depends on the limitation of the experiments in practice. Censoring is just such an example. This limitation provides us the special type of data we have to cope with. We call it "limitation" since this issue not only becomes a part of the design at the design stage, but also limits the choice of estimation methods in the inference stage. Designs for censored data would wipe out the possibility of using least squares estimation in addition to the censoring scheme being a part of the design.

Recent work on robust designs for censored data in ALT are reported by Chaloner and Larntz (1992), Pascual and Montepiedra (2002), and Pascual and Montepiedra (2003). These studies emphasize the robustness against misspecification of the underlying distribution and assume that the "true" model belongs to, or is distributed (with a known prior) onto, a set of several known candidates. Both the Bayesian-type approach and the minimax strategy are used. Ginebra and Sen (1998) investigated optimal designs for censored data, which are robust against possibly misspecified parameter values on which the optimal designs depend. The explicit designs obtained in these works are under straight line regression. Chapter 4 of this dissertation will extend these works by considering a more general regression model, allowing more arbitrary uncertainty in the fitted regression response, and by dropping the restrictions on the structure of the regressors.

1.3.6 Examples

Example 1: Robust optimal design for misspecified simple linear regression.

Wiens (1992) presents the explicit D-, A-, E-, Q- and G-optimal designs under $\mathcal{F}_1$ in the case of fitting a plane, with the design space being a sphere of unit volume in $\mathcal{R}^p$. His work extends Huber (1975) in terms of moving from simple linear regression to multiple linear regression, and also with regard to considering all admitted loss functions. Under $\mathcal{F}_1$ when $p = 1$, $S = [-0.5, 0.5]$, Huber (1975) gives the robust Q-optimal design. This design has the density of

$$k(x) = (ax^2 + b)^+, \tag{1.6}$$

where $(z)^+ := \max\{z, 0\}$, and $a > 0$, b depend on the value of $\nu := \frac{\sigma^2}{n\eta_S^2}$. Under the same condition, Wiens (1992) presents the robust D-optimal design with the density having the same form as of (1.6) but with different coefficients, and the robust A-optimal design with the density having a different form of

$$k(x) = (b - \frac{a}{x^2})^+.$$

The quantity ν can be interpreted as the relative importance of variance versus bias in the belief of the experimenter. When ν is small, one places more emphasis on reducing the bias. For the extreme case, $\nu \to 0$, the design reduces the bias alone so it tends to the continuous uniform design. On the other hand, when ν is large, one places more emphasis on minimizing the variance. The extreme case, $\nu \to \infty$, then will be the design minimizing variance alone and tends to the classical optimal design which places half of the observations on each of -0.5 and 0.5. See Figure 2 for the plots of robust D-, and A-optimal designs obtained in Wiens (1992); and the robust Q-optimal design obtained by Huber (1975). We take $\nu = 1$ in these plots.

Example 2: Restricted robust optimal design for misspecified polynomial regression.

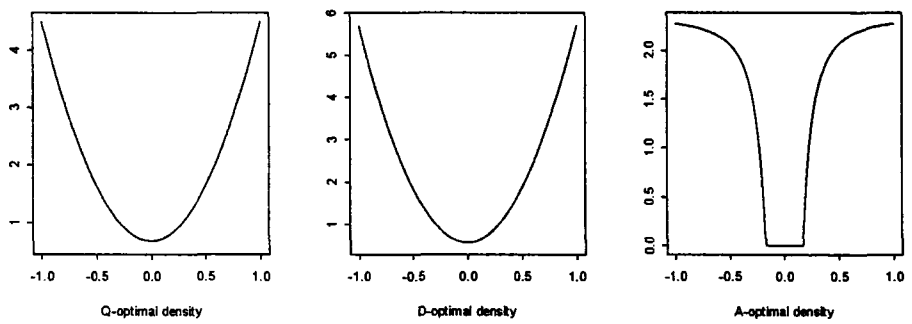

Figure 2: Robust Q-, D-, and A-optimal design densities for misspecified simple linear regression, minimax in contamination space $\mathcal{F}_1$ with $\nu = 1$.

Heo, Schmuland, and Wiens (2001) consider approximately polynomial regression models and illustrate some of the difficulties that can be encountered in the minimax approach without further restriction on the design densities by considering approximate quadratic regression $\mathbf{z}(x) = (1, x, x^2)^T$. They therefore introduce a restricted class of designs to avoid the mathematical and numerical intractability found in the unrestricted minimax theory. They further assume that S is symmetric. For the polynomial regression with $\mathbf{z}(x) = (1, x, ..., x^p)$, they restrict their search for minimax designs within the class of design measures with densities of the form

$$k(\mathbf{x}) = \left(\theta_0 + \sum_{j=1}^{p} \theta_j x^{2j}\right)^+ .$$

The restricted robust Q-, D-, and A-optimal designs are constructed explicitly for both misspecified multiple linear and polynomial regression models. See Figure 3 for restricted robust Q-, D-, and A-optimal designs of quadratic regression. All plots provided use $\nu = 1$.

Example 3: Unrestricted robust optimal design for misspecified polynomial regression.

Shi, Ye, and Zhou (2003) extend the results of Heo, Schmuland, and Wiens (2001), constructing the robust optimal designs by relaxing the restriction on the design

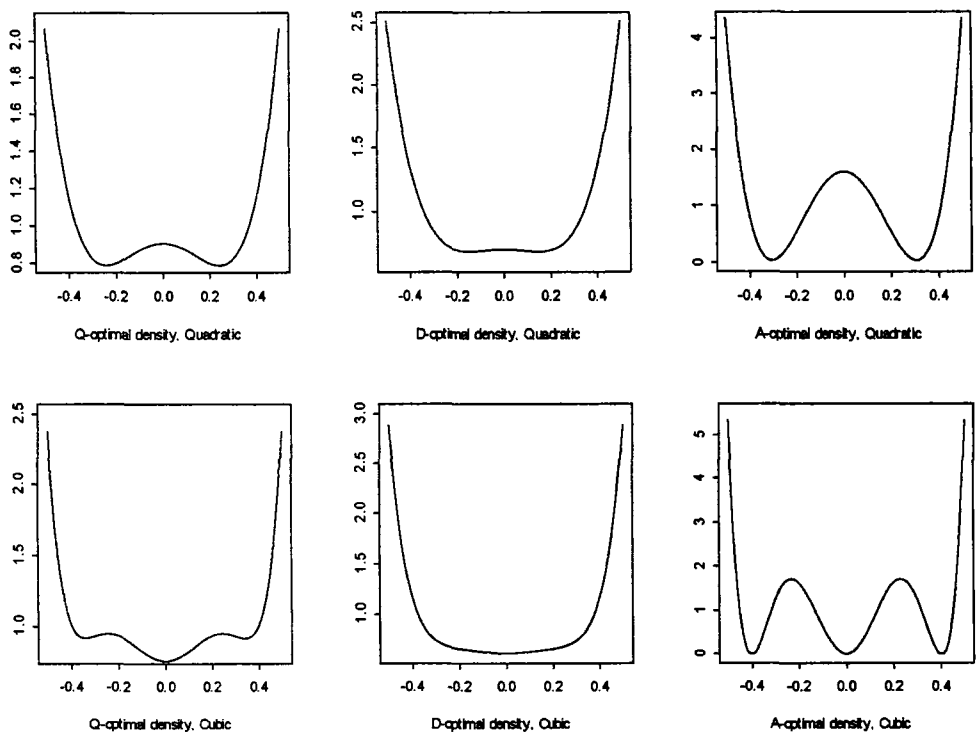

Figure 3: Restricted robust Q-, D-, and A-optimal designs for misspecified polynomial regressions.

density form; furthermore, they extend the results of Huber (1975) and of Wiens (1992) by removing the restrictions on the regressor's structure and on the design space. This paper uses the key tool from nonsmooth optimization theory and makes more explicit and applicable results possible even without any of these restrictions mentioned above. The main result of this paper is that, under $\mathcal{F}_1$, the robust optimal design density has an analytic form

$$k(\mathbf{x}) = \left(\frac{\mathbf{z}^T(\mathbf{x})\mathbf{P}\mathbf{z}(\mathbf{x}) + d}{\mathbf{z}^T(\mathbf{x})\mathbf{Q}\mathbf{z}(\mathbf{x})} \right)^+,$$

where constant matrices $\mathbf{B}$, $\mathbf{D}$, and a constant d are determined by minimizing a specified optimality criterion. These robust optimal designs are constructed for misspecified *linear* regression under ordinary least squares estimation. Chapter 3 of this

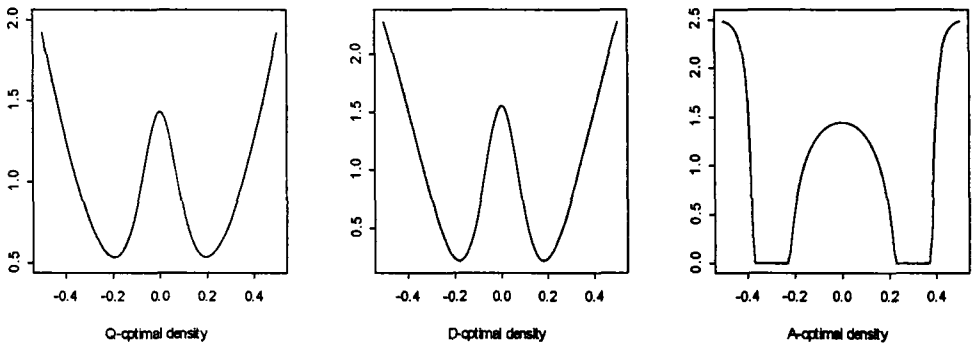

Figure 4: Unrestricted robust Q-, D-, and A-optimal design densities for misspecified quadratic linear regression, minimax in contamination space $\mathcal{F}_1$ with $\nu = 1$.

dissertation will extend their results in three directions: (1) to generalized linear regression; (2) to extrapolation; (3) to weighted least squares estimation.

The explicit and unrestricted robust Q-, D-, and A-optimal design for misspecified polynomial and multiple linear regression are given in Shi, Ye, and Zhou (2003). See Figure 4 for unrestricted robust Q-, D-, and A-optimal designs of quadratic regression. All plots provided use $\nu = 1$.

1.4 Summary of the Results in this Dissertation

The focus of our study is to construct robust designs for the cases of prediction, one-point extrapolation, and general extrapolation. Possibly misspecified nonlinear responses are considered. We assume that our regression model is an approximately known function of a linear regression function. Our designs are robust against the various situations including (1), (2), and (3) as listed in Section 1.3.1. We also deal with different types of data. For complete data, we construct the robust designs when nonlinear (possibly weighted) least squares estimation is used; for censored data, our robust designs assume maximum likelihood estimation is used.

The main results of this dissertation are included in three chapters. In Chapter 2, entitled "Robust Designs for One-Point Extrapolation", we consider the construction of optimal designs for the extrapolation of a regression response to one point outside of the design space. The response function is only approximately specified. As well, we allow for variance heterogeneity. The minimax designs and corresponding optimal regression weights are found in the context of the following problems:

(i) For (nonlinear) ordinary least squares (OLS) estimation with homoscedasticity, determine a design to minimize the maximum value of the mean squared extrapolation error (MSEE), with the maximum being evaluated over the possible departures from the response function;

(ii) For OLS with heteroscedasticity, determine a design to minimize the maximum value of MSEE, with the maximum being evaluated over the departures in both the assumed regression function and the variance homogeneity;

(iii) For (nonlinear) weighted least squares (WLS) estimation, determine both weights and a design to minimize the maximum MSEE;

(iv) Choose both weights and design points to minimize the maximum MSEE, subject to a side condition of unbiasedness.

Solutions to (i)–(iv) are given in complete generality. Numerical comparisons indicate that our designs and weights perform well in combining robustness and efficiency. Applications to accelerated life testing are highlighted.

Chapter 3 is named "Robust Prediction and Extrapolation Designs for Misspecified Generalized Linear Regression Models". In this chapter, we study minimax robust designs for response prediction and extrapolation in biased generalized linear regression models. Minimax designs have been constructed for the following problems:

(i) For OLS estimation with homoscedasticity, determine optimal extrapolation designs to minimize the maximum value of the integrated mean squared extrapolation error (IMSEE);

(ii) For OLS with heteroscedasticity, determine optimal prediction designs to minimize the maximum value of the integrated mean squared prediction error (IMSPE);

(iii) For OLS with heteroscedasticity, determine optimal extrapolation designs to minimize the maximum value of IMSEE;

(iv) For WLS estimation with heteroscedasticity, determine optimal regression weights and minimax prediction designs to minimize the maximum value of IMSPE;

(v) For WLS with heteroscedasticity, determine optimal regression weights and minimax extrapolation designs to minimize the maximum value of IMSEE.

This chapter extends the previous work of others in three aspects: firstly, by considering a nonlinear fitted regression response; secondly, by taking a rather general extrapolation space; finally, and most significantly, by dropping all restrictions on the structure of the regressors.

In Chapter 4, titled "Robust Prediction and Extrapolation Designs for Censored Data", we present the construction of optimal designs for both response prediction and extrapolation with a possibly misspecified generalized linear regression model when the data are censored. The minimax designs are found for maximum likelihood estimation in the context of the following problems:

(i) For prediction, determine a minimax design which minimizes the maximum value of IMSPE, with the maximum value being evaluated over the possible departure from the assumed response function;

(ii) For one-point extrapolation, determine a minimax design which minimizes the maximum MSEE;

(iii) For general extrapolation, determine a minimax design which minimizes the maximum value of IMSEE;

(iv) Determine unbiased minimax design for both prediction and extrapolation problems.

This chapter extends the work on robust designs for complete data in Chapters 2

and 3 by incorporating censoring and maximum likelihood estimation. Solutions are derived by a nonsmooth optimization technique analytically, and are given in complete generality. A typical example in accelerated life testing is also demonstrated.

Chapter 5 provides applications and implementation of the continuous optimal designs that we have constructed in the chapters prior to it. A real life dose-response experiment is discussed. A couple of practical implementation methods are proposed and displayed after other existing approaches are reviewed. We also state that one of the existing implementation schemes is optimal under certain criteria. The comparison between one of our proposed approaches and an existing one shows that the proposed one is fairly close to optimal.

References

Adewale, A., and Wiens (2006). New Criteria for Robust Integer-Valued Designs in Linear Models. *Computational Statistics and Data Analysis,* to appear.

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

Box, G. E. P., and Draper, N. R. (1975). Robust Designs. *Biometrika,* 62(2), 347-352.

Chang, Y. J., and Notz, W. I. (1996). Model Robust Designs. *Handbook of Statistics,* ed. Ghosh, S., and Rao, C. R. Elsevier, 13, 1055-1098.

Chaloner, K. and Larntz, K. (1992). Bayesian design for accelerated life testing. *Journal of Statistical Planning and Inference,* 33, 245-259.

Dasgupta, A., Mukhopadhyay, S., and Studden, W. J. (1992). Compromise Designs in Heteroscedastic Linear Models. *Journal of Statistical Planning and Inference,* 32, 363-384.

Dette, H., Heiligers, B., and Studden, W. J. (1995). Minimax Designs in Linear Regression Models. *Annals of Statistics*, 23, 30-40.

Fang, Z. and Wiens, D. P. (1999). Robust extrapolation designs and weights for biased regression models with heteroscedastic errors. *The Canadian Journal of Statistics*, 27, 751-770.

Ford, I. (1976). *Optimal Static and Sequential Design: A Critical Review.* Unpublished Doctoral thesis, University of Glasgow.

Ford, I., Titterington, D. M., and Kitsos, C. P. (1989). Recent Advances in Nonlinear Experimental Design. *Technometrics*, 31, 49-60.

Ginebra, J. and Sen, A. (1998). Minimax Approach to Accelerated Life Tests. *IEEE Transactions on Reliability*, 47, 261-267.

Guest, P. G. (1958). The Spacing of Observations in Polynomial Regression. *Annals of Mathematical Statistics*, 29, 294-299.

Heckman, N. E. (1987). Robust Design in a Two Treatment Comparison in the Presence of a Covariate. *Journal of Statistical Planning and Inference, 16*, 75-81.

Heo, G., Schmuland, B., and Wiens, D. P. (2001). Restricted Minimax Robust Designs for Misspecified Regression Models. *The Canadian Journal of Statistics*, 29, 117-128.

Hoel, P. G., and Levine, A. (1964). Optimal Spacing and Weighing in Polynomial Prediction. *Annals of Mathematical Statistics*, 35, 1553–1560.

Hosmer, D. W. Jr., and Lemeshow, S. (1998). *Applied Survival Analysis.* Wiley, New York.

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

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

Kiefer, J. (1973). Optimal Designs for Fitting Multiresponse Surface. *Symposium on Multivariate Analysis III*, 287-297. Academic Press, New York.

Kiefer, J. (1974). General Equivalence Theory for Optimum Designs (Approximate Theory). *Annals of Statistics,* 2, 849-879.

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

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

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.

Müller, C. H. (1995). Maxmin Efficient Designs for Estimating Nonlinear Aspects in Linear Models. *Journal of Statistical Planning and Inference,* 44, 117-132.

Pascual, F. G. and Montepiedra, G. (2002). On Minimax Designs when There Are Two Candidate Models. *Journal of Statistical Computation and Simulation,* 72, 841-862.

Pascual, F. G. and Montepiedra, G. (2003). Model-Robust Test Plans with Applications in Accelerated Life Testing. *Technometrics*, 45, 47-57.

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

Shi, P., Ye, J., and Zhou, J. (2003). Minimax Robust Designs for Misspecified Regression Models. *The Canadian Journal of Statistics*, 31, 397-414.

Silvey, S.D. (1980). *Optimal Design.* Chapman and Hall, London.

Sinha, S. and Wiens, D. P. (2002). Robust Sequential Designs for Nonlinear Regression. *The Canadian Journal of Statistics*, 30, 601-618.

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

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

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

Wiens, D. P. (1993). Designs for Approximately Linear Regression: Maximizing the Minimum Coverage Probability of Confidence Ellipsoids. *The 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. (1996) Robust, Sequential Designs for Approximately Linear Models. *The Canadian Journal of Statistics*, 24, 67-79.

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

Wiens, D. P., and Zhou, J. (1996). Minimax Regression Designs for Approximately Linear Models With Autocorrelated Errors. *Journal of Statistical Planning and Inference*, 55, 95-106.

Wiens, D. P., and Zhou, J. (1999). Minimax Designs for Approximately Linear Models With AR(1) Errors. *The Canadian Journal of Statistics*, 27, 781-794.

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 the Royal Statistical Society*, Series B, 55, 871-880.

Yue, R. and Hickernell, F. (1999). Robust Designs for Fitting Linear Models with Misspecification. *Statistica Sinica*, 9, 1053-1069.

Zhou, J. (2001). A Robust Criterion for Experimental Designs for Serially Correlated Observations. *Technometrics*, 43, 462-467.

CHAPTER II

ROBUST DESIGNS FOR ONE-POINT EXTRAPOLATION

Abstract We consider the construction of designs for the extrapolation of a regression response to one point outside of the design space. The response function is an only approximately known function of a specified linear function. As well, we allow for variance heterogeneity. We find minimax designs and corresponding optimal regression weights in the context of the following problems: (P1) For nonlinear least squares estimation with homoscedasticity, determine a design to minimize the maximum value of the mean squared extrapolation error (MSEE), with the maximum being evaluated over the possible departures from the response function; (P2) For nonlinear least squares estimation with heteroscedasticity, determine a design to minimize the maximum value of MSEE, with the maximum being evaluated over both types of departures; (P3) for nonlinear weighted least squares estimation, determine both weights and a design to minimize the maximum MSEE; (P4) Choose weights and design points to minimize the maximum MSEE, subject to a side condition of unbiasedness. Solutions to (P1)–(P4) are given in complete generality. Numerical comparisons indicate that our designs and weights perform well in combining robustness and efficiency. Applications to accelerated life testing are highlighted.

[1]Co-authored with Professor Douglas P. Wiens. Submitted for publication.

2.1 Introduction

In this chapter we study the construction of designs for the extrapolation of regression responses to one point outside of the design space. Such 'one-point extrapolation' designs are of interest in problems of accelerated life testing (ALT), in which products are typically tested at unusual stress levels, with the results then extrapolated to a lower stress level anticipated in practice. Our model is somewhat similar to a generalized linear model, in that the response fitted by the experimenter is a function of a linear function of unknown parameters and known regressors. Our designs are robust in that we allow both for imprecision in the specification of the response, and for possible heteroscedasticity.

Robust designs for extrapolation of a, possibly misspecified, *linear* response were obtained by Fang and Wiens (1999); see also the references therein, in particular Dette and Wong (1996), Draper and Herzberg (1973), Huang and Studden (1988), Huber (1975) and Spruill (1984). The current work goes beyond Fang and Wiens (1999) in two ways - in the move to a generalized linear response as described above, and in our emphasis on extrapolation to a single point, thus allowing for more explicit and applicable results than were previously possible.

For nonlinear regression, Ford, Kitsos and Titterington (1989) present various static and sequential designs for nonlinear models without the consideration of model uncertainty. Sinha and Wiens (2002) have employed notions of robustness in the construction of sequential designs for the nonlinear model. In many ALT applications however, sequential designs are not feasible (Ford, Kitsos, & Titterington, 1989), hence our focus in this chapter on static designs.

Fang and Wiens (1999) point out that "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." With this in mind, we shall depart rather broadly from the usual generalized linear response models:

1. The response is taken to be an approximately known function of a linear function of known regressors and unknown parameters:

$$E(Y|\mathbf{x}) = h(\boldsymbol{\theta}_0^T \mathbf{z}(\mathbf{x})) + n^{-1/2} f(\mathbf{x})$$

for p regressors $\mathbf{z}(\mathbf{x}) = (z_1(\mathbf{x}), z_2(\mathbf{x}), ..., z_p(\mathbf{x}))^T$, depending on a $q-$dimensional vector $\mathbf{x}$ of independent variables. The function h is strictly monotonic, with a bounded second derivative. We assume that $||\mathbf{z}(\mathbf{x})||$ is bounded on S. The response contaminant f represents uncertainty about the exact nature of the regression response and is unknown and arbitrary, subject to certain restrictions. We estimate $\boldsymbol{\theta}$ but not f; this leads to possibly biased extrapolations $\hat{Y}(\mathbf{x}) = h(\hat{\boldsymbol{\theta}}^T \mathbf{z}(\mathbf{x}))$ of $E(Y|\mathbf{x})$. The factor $n^{-1/2}$ is necessary for a sensible asymptotic treatment. It ensures that losses due to bias remain of the same asymptotic order as those due to variance, and is analogous to the requirement of contiguity in the asymptotic theory of hypothesis testing.

2. The experimenter takes n uncorrelated observations $Y_i = Y(\mathbf{x}_i)$, with $\mathbf{x}_i$ freely chosen from a design space S. Our goal is to choose these design points from S in an optimal manner in order to extrapolate the estimates of $E(Y|\mathbf{x})$ to $\mathbf{x}_0$.

3. The observations Y_i are possibly heteroscedastic, with $VAR\{Y(\mathbf{x}_i)\} = \sigma^2 g(\mathbf{x}_i)$ for a function g satisfying conditions given below.

We estimate $\boldsymbol{\theta}$ by nonlinear least squares, possibly weighted with weights $w(\mathbf{x})$. Our loss function is n times the mean squared error of $\hat{Y}(\mathbf{x}_0)$ in estimating $E(Y|\mathbf{x}_0)$. This depends on the design measure $\xi = n^{-1}\Sigma_{i=1}^{n} \delta_{\mathbf{x}_i}$ as well as on w, f and g:

$$MSEE(f, g, w, \xi) = nE\{[\hat{Y}(\mathbf{x}_0) - E(Y|\mathbf{x}_0)]^2\}.$$

We denote unweighted least squares by $w = \mathbf{1}$, and homogeneous variances by $g = \mathbf{1}$. The following problems will be addressed:

(P1) For ordinary least squares (OLS) estimation under homoscedasticity, determine designs to minimize the maximum value, over f, of $MSEE(f, \mathbf{1}, \mathbf{1}, \xi)$.

(P2) For OLS estimation under possible heteroscedasticity, determine designs to minimize the maximum value, over f and g, of $MSEE(f, g, \mathbf{1}, \xi)$.

(P3) For weighted least squares (WLS) estimation, determine designs and weights to minimize the maximum value, over f and g, of $MSEE(f, g, w, \xi)$.

(P4) Choose weights and design points to minimize $\max_{f,g} MSEE(f, g, w, \xi)$, subject to a side condition of unbiasedness.

The rest of this chapter is organized as follows. The designs for P1 are provided in Section 2.4. The designs and weights which constitute solutions to problems P2 and P3 are given in Section 2.5. Those for P4 are given in Section 2.6. Some mathematical preliminaries are detailed in Section 2.2. The maximization part of the minimax designs construction are provided in Section 2.3. Comparisons of these designs are presented in Section 2.7. All proofs are in the Appendix.

2.2 Preliminaries and Notation

We define the 'target' parameter $\boldsymbol{\theta}_0$ to be that which produces the best agreement, in the L_2-sense, between $h(\boldsymbol{\theta}^T \mathbf{z}(\mathbf{x}))$ and $E(Y|\mathbf{x})$:

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

We assume that $\boldsymbol{\theta}_0$ is unique, so that with

$$f_n(\mathbf{x}) = \sqrt{n} \left[E(Y|\mathbf{x}) - h(\boldsymbol{\theta}_0^T \mathbf{z}(\mathbf{x})) \right]$$

and

$$\tilde{\mathbf{z}}(\mathbf{x}) = h' \left(\boldsymbol{\theta}_0^T \mathbf{z}(\mathbf{x}) \right) \mathbf{z}(\mathbf{x})$$

we have $\int_S \tilde{\mathbf{z}}(\mathbf{x}) f_n(\mathbf{x}) d\mathbf{x} = 0$. Where possible we drop the subscript on f.

We shall assume that $f_n = f$ is an unknown member of the class

$$\mathcal{F} = \left\{ f \mid \int_S f^2(\mathbf{x})d\mathbf{x} \leq \eta_S^2 < \infty, |f(\mathbf{x}_0)| \leq \eta_T < \infty, \int_S \bar{\mathbf{z}}(\mathbf{x})f(\mathbf{x})d\mathbf{x} = \mathbf{0} \right\}$$

for positive constants η_S, η_T. The departure from homogeneity of variances is measured by $g(\mathbf{x})$, which is assumed to be an unknown member of the class

$$\mathcal{G} = \left\{ g \mid \int_S g^2(\mathbf{x})d\mathbf{x} \leq \Omega^{-1} := \int_S d\mathbf{x} < \infty \right\}. \tag{2.1}$$

The condition in (2.1) is equivalent to defining $\sigma^2 = \sup_g \left\{ \int_S var^2 \left[\varepsilon(\mathbf{x}) \right] \Omega d\mathbf{x} \right\}^{1/2}$.

To ensure the nonsingularity of a number of relevant matrices, we assume that the regressors and design space satisfy

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

We propose to estimate $\boldsymbol{\theta}_0$ using nonlinear least squares (LS) to fit $E(\hat{Y}|\mathbf{x}) = h(\boldsymbol{\theta}_0^T\mathbf{z}(\mathbf{x}))$ with nonnegative weights $w(\mathbf{x})$.

We make use of the following matrices and vectors:

$$\mathbf{A}_S = \int_S \bar{\mathbf{z}}(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})d\mathbf{x}, \qquad \mathbf{A}_T = \bar{\mathbf{z}}(\mathbf{x}_0)\bar{\mathbf{z}}^T(\mathbf{x_0}),$$

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

$$\mathbf{b}_{f,S} = \int_S \bar{\mathbf{z}}(\mathbf{x})f(\mathbf{x})w(\mathbf{x})\xi\left(d\mathbf{x}\right), \qquad \mathbf{b}_{f,T} = \bar{\mathbf{z}}(\mathbf{x_0})f(\mathbf{x_0}).$$

It follows from (A) that $\mathbf{A}_S$ is non-singular. The LS estimator of $\boldsymbol{\theta}_0$ is

$$\hat{\boldsymbol{\theta}} = \arg\min \sum_{i=1}^{n} [Y_i - h(\boldsymbol{\theta}^T\mathbf{z}(\mathbf{x}))]^2 w(\mathbf{x}_i)$$

and satisfies $\sum_{i=1}^n \dot{\phi}_i(\hat{\boldsymbol{\theta}}) = 0$ for

$$\dot{\phi}_i(\boldsymbol{\theta}) = [Y_i - h(\boldsymbol{\theta}^T\mathbf{z}(\mathbf{x}_i))] \left[h'(\boldsymbol{\theta}^T\mathbf{z}(\mathbf{x}_i)) \right] w(\mathbf{x}_i)\mathbf{z}(\mathbf{x}_i).$$

In addition, we have

$$\begin{aligned}\ddot{\Phi}(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \ddot{\phi}_i(\boldsymbol{\theta}) \\ &= \sum_{i=1}^{n} [Y_i - h(\boldsymbol{\theta}^T \mathbf{z}(\mathbf{x}_i))] \left[h''(\boldsymbol{\theta}^T \mathbf{z}(\mathbf{x}_i))\right] w(\mathbf{x}_i)\mathbf{z}(\mathbf{x}_i)\mathbf{z}^T(\mathbf{x}_i) \\ &\quad - \sum_{i=1}^{n} \left(\left[h'(\boldsymbol{\theta}^T \mathbf{z}(\mathbf{x}_i))\right]^2 w(\mathbf{x}_i)\mathbf{z}(\mathbf{x}_i)\mathbf{z}^T(\mathbf{x}_i).\right.\end{aligned}$$

The information matrix is

$$\mathcal{I}(\boldsymbol{\theta}_0) = \lim_{n\to\infty} E(-\frac{1}{n}\ddot{\Phi}(\boldsymbol{\theta}_0)) = \mathbf{B},$$

since

$$\begin{aligned}&E\left\{\frac{1}{n}\sum_{i=1}^{n}[Y_i - h(\boldsymbol{\theta}_0^T \mathbf{z}(\mathbf{x}_i))] \left[h''(\boldsymbol{\theta}_0^T \mathbf{z}(\mathbf{x}_i))\right] w(\mathbf{x}_i)\mathbf{z}(\mathbf{x}_i)\mathbf{z}^T(\mathbf{x}_i)\right\} \\ = \; & n^{-1/2} \cdot \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{x}_i)) \left[h''(\boldsymbol{\theta}_0^T \mathbf{z}(\mathbf{x}_i))\right] w(\mathbf{x}_i)\mathbf{z}(\mathbf{x}_i)\mathbf{z}^T(\mathbf{x}_i)\end{aligned}$$

is $O\left(n^{-1/2}\right)$ by virtue of our assumptions on f, h and $\mathbf{z}$.

By Taylor's Theorem,

$$\mathbf{0} = \sum_{i=1}^{n} \dot{\phi}_i(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \left\{\dot{\phi}_i(\boldsymbol{\theta}_0) + \ddot{\phi}_i(\bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right\},$$

where $\bar{\boldsymbol{\theta}}$ lies between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. Then

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \left(-\frac{1}{n}\sum_{i=1}^{n} \ddot{\phi}_i(\bar{\boldsymbol{\theta}})\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \dot{\phi}_i(\boldsymbol{\theta}_0)\right).$$

Note that $n^{-1/2}\sum_{i=1}^{n} \dot{\phi}_i(\boldsymbol{\theta}_0)$ is asymptotically normal, with asymptotic mean $\mathbf{b}_{f,S}$ and covariance

$$COV[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \dot{\phi}_i(\boldsymbol{\theta}_0)] = \frac{1}{n}\sum_{i=1}^{n} \left[h'(\boldsymbol{\theta}_0^T \mathbf{z}(\mathbf{x}_i))\right]^2 \sigma^2 g(\mathbf{x}_i)\mathbf{z}(\mathbf{x}_i)\mathbf{z}^T(\mathbf{x}_i)w^2(\mathbf{x}_i) = \sigma^2\mathbf{D}.$$

As in Seber and Wild (1989, §12.2), the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \sim AN(\mathbf{B}^{-1}\mathbf{b}_{f,S}, \sigma^2\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}),$$

and then by the delta method,

$$\sqrt{n}\left(h(\hat{\boldsymbol{\theta}}^T\mathbf{z}(\mathbf{x}_0)) - h(\boldsymbol{\theta}_0^T\mathbf{z}(\mathbf{x}_0))\right) \sim AN\left(\bar{\mathbf{z}}^T(\mathbf{x}_0)\mathbf{B}^{-1}\mathbf{b}_{f,S}, \sigma^2\bar{\mathbf{z}}^T(\mathbf{x}_0)\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}\bar{\mathbf{z}}(\mathbf{x}_0)\right).$$

The loss function $MSEE$ splits into terms due to (squared) extrapolation bias, extrapolation variance, and model misspecification:

$$\begin{aligned} MSEE(f,g,w,\xi) &= nE\left\{[\hat{Y}(\mathbf{x}_0) - E(Y|\mathbf{x}_0)]^2\right\} \\ &= nE\left\{\left[h(\hat{\boldsymbol{\theta}}^T\mathbf{z}(\mathbf{x}_0)) - h(\boldsymbol{\theta}_0^T\mathbf{z}(\mathbf{x}_0)) - \frac{1}{\sqrt{n}}f(\mathbf{x}_0)\right]^2\right\} \\ &= EB(f,w,\xi) + EV(g,w,\xi) + f^2(\mathbf{x}_0) \end{aligned}$$

where the squared extrapolation bias (EB) and extrapolation variance (EV) are

$$\begin{aligned} EB(f,w,\xi) &= n\left\{E\left[h(\hat{\boldsymbol{\theta}}^T\mathbf{z}(\mathbf{x}_0)) - h(\boldsymbol{\theta}_0^T\mathbf{z}(\mathbf{x}_0))\right]\right\}^2 \\ &\quad -2\sqrt{n}f(\mathbf{x}_0)E\left[h(\hat{\boldsymbol{\theta}}^T\mathbf{z}(\mathbf{x}_0)) - h(\boldsymbol{\theta}_0^T\mathbf{z}(\mathbf{x}_0))\right], \end{aligned}$$

and

$$EV(g,w,\xi) = nVAR(\hat{Y}(\mathbf{x}_0)) = nVAR(h(\hat{\boldsymbol{\theta}}^T\mathbf{z}(\mathbf{x}_0))).$$

Asymptotically,

$$\begin{aligned} EB(f,w,\xi) &= \mathbf{b}_{f,S}^T\mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1}\mathbf{b}_{f,S} - 2\mathbf{b}_{f,T}^T\mathbf{B}^{-1}\mathbf{b}_{f,S}, \\ EV(g,w,\xi) &= \sigma^2\bar{\mathbf{z}}^T(\mathbf{x}_0)\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}\bar{\mathbf{z}}(\mathbf{x}_0) = \sigma^2 tr\mathbf{A}_T\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}. \end{aligned}$$

We have defined ξ to be a discrete measure, with atoms of size n^{-1} at the design points (possibly repeated). We now adopt the viewpoint of approximate design theory and allow ξ to be any probability measure on S. One reason for this is that as in Lemma 1 of Wiens (1992), the class $\mathcal{F}$ is so broad that only absolutely continuous measures ξ can have finite maximum loss. Thus, let $k(\mathbf{x})$ be the density of ξ, and define $m(\mathbf{x}) = k(\mathbf{x})w(\mathbf{x})$. Without loss of generality, we 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 satisfies

$$\int_S \frac{m(\mathbf{x})}{w(\mathbf{x})}d\mathbf{x} = 1, \tag{2.2}$$

and

$$\begin{aligned} \mathbf{B} &= \int_S \bar{\mathbf{z}}(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})m(\mathbf{x})d\mathbf{x}, \\ \mathbf{b}_{f,S} &= \int_S \bar{\mathbf{z}}(\mathbf{x})f(\mathbf{x})m(\mathbf{x})d\mathbf{x}. \end{aligned}$$

Various methods for implementing designs with continuous measures are discussed in Heo, Schmuland and Wiens (2001) and the references therein. As an example, a practical implementation for univariate x is to place the n design points at the quantiles $x_i = \xi^{-1}\left(\frac{i-1}{n-1}\right)$.

From the definitions of $\mathbf{B}$, $\mathbf{b}_{f,S}$ and $\mathbf{b}_{f,T}$, we notice that $EB(f, w, \xi)$ depends on (w, ξ) only through m and $EV(g, w, \xi)$ through m and w. Hence, we can optimize over m and w subject to (2.2) rather than over k and w. In the next four sections we exhibit solutions to P1 – P4.

2.3 Maximization over $f \in \mathcal{F}$ and $g \in \mathcal{G}$

In this section we exhibit the maxima of $MSEE$, for fixed functions $m(\mathbf{x})$ and $w(\mathbf{x})$. The minimizing m and w then constitute the solutions to P1 – P4. The maxima are obtained in a manner very similar to that used in Fang & Wiens (1999), and so their derivations are omitted.

Define positive semidefinite matrices

$$\begin{aligned} \mathbf{K} &= \int_S \bar{\mathbf{z}}(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})m^2(\mathbf{x})d\mathbf{x}, \\ \mathbf{G} &= \mathbf{K} - \mathbf{B}\mathbf{A}_S{}^{-1}\mathbf{B} = \int_S \left[\left(m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}_S^{-1}\right)\bar{\mathbf{z}}(\mathbf{x})\right]\left[\left(m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}_S^{-1}\right)\bar{\mathbf{z}}(\mathbf{x})\right]^T d\mathbf{x}, \end{aligned}$$

and constants $r_{T,S} = \eta_T/\eta_S$, reflecting the relative amounts of model response uncertainty in the extrapolation and design space, and $\nu = \sigma^2/\eta_S^2$, representing the relative importance of variance versus bias. In this notation, we have the following theorem.

Theorem 2.1 *The maximum squared extrapolation bias is*

$$\sup_{f\in\mathcal{F}} EB(f, m) = \eta_S^2 \left\{\lambda_m + 2r_{T,S}\sqrt{\lambda_m}\right\},$$

where $\lambda_m = \tilde{\mathbf{z}}^T(\mathbf{x}_0)\mathbf{B}^{-1}\mathbf{G}\mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)$. *The maximum is attained at*

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

where $\mathbf{a}_0 = \mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)/\sqrt{\lambda_m}$.

We obtain Theorems 2.2 and 2.3 from this result. Theorem 2.2 gives the maximum $MSEE$ under homoscedasticity while Theorem 2.3 gives that under heteroscedasticity.

Theorem 2.2 *The maximum mean squared extrapolation error in problem P1 is*

$$\sup_{f \in \mathcal{F}} MSEE(f, 1, 1, m) = \eta_S^2 \left\{ \left(\sqrt{\lambda_m} + r_{T,S} \right)^2 + \nu \tilde{\mathbf{z}}^T(\mathbf{x}_0)\mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0) \right\}, \tag{2.3}$$

attained at f_m.

Theorem 2.3 *Define* $l_m(\mathbf{x}) = \left[\tilde{\mathbf{z}}^T(\mathbf{x})\mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)\right]^2$ *and* $\alpha_m = \int_S [l_m(\mathbf{x})m^2(\mathbf{x})]^{2/3} d\mathbf{x}$. *Then the maximum mean squared extrapolation error in problems P2 - P4 is*

$$\sup_{f \in \mathcal{F}, g \in \mathcal{G}} MSEE(f, g, w, m) = \eta_S^2 \left\{ \begin{array}{c} \left(\sqrt{\lambda_m} + r_{T,S} \right)^2 \\ + \nu \Omega^{-1/2} \left[\int_S \{ w(\mathbf{x}) l_m(\mathbf{x}) m(\mathbf{x}) \}^2 d\mathbf{x} \right]^{1/2} \end{array} \right\}, \tag{2.4}$$

attained at f_m *and*

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

The following theorem, whose proof is very similar to that of Theorem 2.2(a) in Fang & Wiens (1999), gives the minimax weights for fixed $m(\mathbf{x})$.

Theorem 2.4 *For fixed* $m(\mathbf{x})$ *the weights minimizing* $\sup_{f \in \mathcal{F}, g \in \mathcal{G}} MSEE(f, g, w, m)$ *subject to (2.2) are given by*

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

Then $min_w \{ \sup_{f \in \mathcal{F}, g \in \mathcal{G}} MSEE(f, g, w, m) \} = \eta_S^2 \left\{ \left(\sqrt{\lambda_m} + r_{T,S} \right)^2 + \nu \Omega^{-1/2} \alpha_m^{3/2} \right\}$.

2.4 Optimal Designs with Homoscedasticity: Solution to P1

Problem P1 has become that of finding a density $m_*(\mathbf{x})$ which minimizes (2.3). The solution is given by Theorem 2.5, which reduces the problem to a $(2p+1)$-dimensional numerical problem. The generality of our solution to P1, as well as those to P2 and P3, should be compared with the corresponding development in Fang & Wiens (1999). This generality, and the relative simplicity of the solutions, is made possible by our use of a one-point extrapolation region.

Theorem 2.5 *The density $m_*(\mathbf{x})$ minimizing (2.3) for OLS estimation is of the form*

$$m_*(\mathbf{x}) = \left[\frac{\mathbf{z}^T(\mathbf{x})\boldsymbol{\gamma}}{\mathbf{z}^T(\mathbf{x})\boldsymbol{\beta}} + \frac{\lambda}{(\mathbf{z}^T(\mathbf{x})\boldsymbol{\beta})^2}\right]^+,$$

where $(z)^+ = \max(z, 0)$. The $p \times 1$ nonzero vectors $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$ and constant λ satisfy: (i) $\int_S m_*(\mathbf{x})d\mathbf{x} = 1$, and (ii) minimize (2.3).

Example 1. We consider an approximate accelerated failure model in survival analysis (Hosmer & Lemeshow 1998, p. 272). It is a generalized simple linear regression with $\mathbf{z}^T(x) = (1, x)$, $\mu = \theta_0 + \theta_1 x$, and $h(\mu) = e^{\mu}$. By Theorem 2.5, the optimal design density has the form

$$m_*(x) = \left[\frac{a_1 x + a_2}{a_3 x + a_4} + \frac{a_5}{e^{2\theta_1 x}(a_3 x + a_4)^2}\right]^+. \tag{2.5}$$

Note that (2.5) is over-parameterized - if one of $a_1 - a_5$ is nonzero then we can assume that it is unity. The term $e^{2\theta_0}$ has been absorbed into a_5, but m_* still depends on θ_1. To address this issue we adopt a mixture of minimax and local approaches. We start at some $\theta_1 = \theta_1^{(0)}$. The corresponding optimal design density is $m_*^{(0)}(x)$. Then, we maximize (2.3) with $m = m_*^{(0)}$ over an interval containing $\theta_1^{(0)}$ to find the least favourable value of θ_1, say $\theta_1^{(1)}$. We iterate between minimizing over designs and maximizing over θ_1 until attaining convergence, say to θ_1^{LF}. Finally, we employ Theorem 2.5 to

Table 2.1. Numerical values for (2.5)
$S = [0,1]$, $\theta_1 \in [0,2]$, $r_{TS} = 1$, and $x_0 = 1.17$.

ν	a_1	a_2	a_3	a_5	θ_1^{LF}
0.5	0.130	0.421	−0.578	0.736	2
1	0.344	0.000224	−0.778	1.28	2
2	0.173	0.000286	−0.885	1.20	2

construct the 'locally most robust' design density $m_*(x) = \left[\frac{a_1 x + a_2}{a_3 x + a_4} + \frac{a_5}{e^{2\theta_1^{LF} x}(a_3 x + a_4)^2}\right]^+$ corresponding to θ_1^{LF}.

To illustrate the approach, we consider the Class-H insulation data from Nelson (1990, Table 2.1). We transform the temperature variable t used there to our stress variable x with domain of $[0,1]$ via the linear transformation

$$x = \frac{-1.876 + 1000/(t + 273.15)}{0.283}.$$

The least squares estimate for the nominal model is $\hat{\theta}_1 = 0.946$, with standard error 0.0486. The corresponding 99% confidence interval for θ_1 is $(0.814, 1.08)$. Taking the model misspecification into account, we consider a broader region $\theta_1 \in [0,2]$. We use the same extrapolation point $x_0 = 1.17$ as Nelson (*op. cit.*). We carried out the process described above for several values of ν, each time starting at $\theta_1^{(0)} = 0.946$. In each case we obtained $\theta_1^{LF} = 2$. See Table 2.1 for the numerical values of the constants, and Figure 5(a) for plots. As a comparison, Figure 5(b) provides the plots of the locally optimal design densities at $\theta_1 = 0.946$. All plots use $a_4 = 1$ and $r_{T,S} = 1$.

2.5 Optimal Designs with Heteroscedasticity

Our problems P2 and P3 have become the following:

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

$$\begin{aligned} & \eta_S^{-2} \sup_{f \in \mathcal{F}, g \in \mathcal{G}} MSEE(f, g, \mathbf{1}, m) \\ = \; & \left(\sqrt{\lambda_m} + r_{T,S}\right)^2 + \nu \Omega^{-1/2} \left[\int_S \{l_m(\mathbf{x}) m(\mathbf{x})\}^2 \, d\mathbf{x}\right]^{1/2} \end{aligned} \tag{2.6}$$

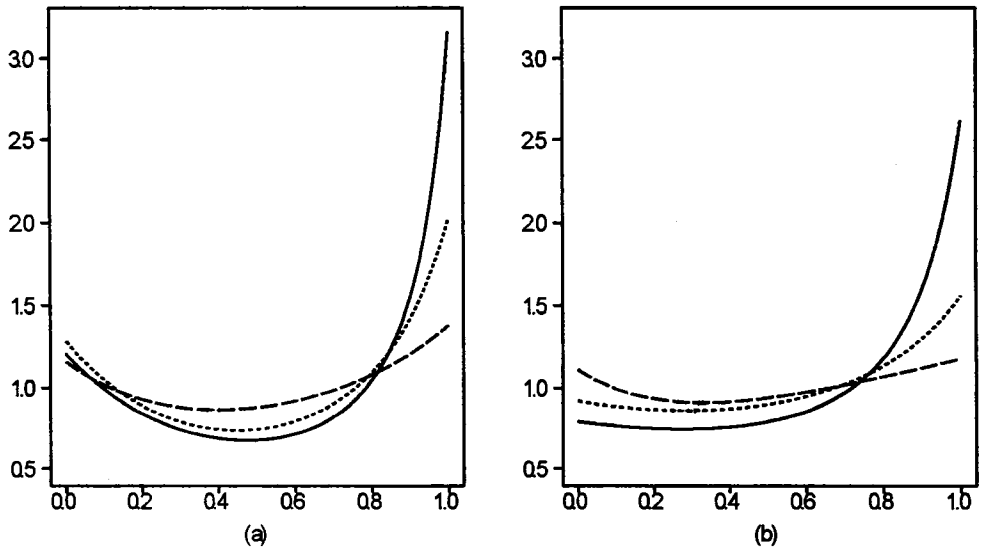

Figure 5: Optimal minimax design densities $m(x) = \left[\frac{a_1x+a_2}{a_3x+a_4} + \frac{a_5}{e^{2\theta_1^* x}(a_3x+a_4)^2}\right]^+$ in Example 1 for $x_0 = 1.17$. (a) locally most robust design densities for $\theta_1 = \theta_1^{LF}$ in $[0, 2]$; (b) locally optimal design densities for $\theta_1 = \theta_1^{(0)} = 0.946$. Each plot uses three values of ν: $\nu = 2$ (solid line), $\nu = 1$ (dotted line), $\nu = 0.5$ (broken line).

with λ_m and $l_m(\mathbf{x})$ as defined in Theorems 2.1 and 2.3 respectively. Then $k_*(\mathbf{x}) = m_*(\mathbf{x})$ is the optimal one-point extrapolation design density for OLS estimation.

(P3) Find a density $m_*(\mathbf{x})$ which minimizes

$$\begin{aligned} &\eta_S^{-2} \sup_{f\in\mathcal{F}, g\in\mathcal{G}} MSEE(f, g, w_m, m) \\ &= \left(\sqrt{\lambda_m}+r_{T,S}\right)^2 + \nu\Omega^{-1/2}\left[\int_S \left\{l_m(\mathbf{x})m^2(\mathbf{x})\right\}^{2/3} d\mathbf{x}\right]^{3/2}. \end{aligned} \tag{2.7}$$

Then the weights

$$w_*(\mathbf{x}) = \alpha_{m_*}\left\{l_{m_*}^2(\mathbf{x})m_*(\mathbf{x})\right\}^{-1/3} I\left[m_*(\mathbf{x}) > 0\right] \tag{2.8}$$

and the density

$$k_*(\mathbf{x}) = \alpha_{m_*}^{-1}\left[l_{m_*}(\mathbf{x})m_*^2(\mathbf{x})\right]^{2/3}, \tag{2.9}$$

with α_{m_*} defined in Theorem 2.3, are optimal for one-point extrapolation with WLS estimation.

2.5.1 Minimax Designs for OLS: Solution to P2

The solution to P2 is provided by Theorem 2.6 below.

Theorem 2.6 *The density $m_*(\mathbf{x})$ minimizing (2.6) for OLS estimation is of the form*

$$m_*(\mathbf{x}) = \frac{\left[\left(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\gamma}\right)\left(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta}\right) + \lambda\right]^+}{\left(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta}\right)^2\left[1 + t\left(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta}\right)^2\right]}.$$

The $p \times 1$ nonzero vectors $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, positive constant t, and constant λ satisfy: (i) $\int_S m_*(\mathbf{x})d\mathbf{x} = 1$, and (ii) minimize (2.6).

Example 2. Consider an approximate polynomial regression model $E(Y|x) \approx \mu = \mathbf{z}^T(x)\boldsymbol{\theta}_0 = \theta_0 + \theta_1 x + ... + \theta_p x^p$, where $\mathbf{z}^T(x) = (1, x, ..., x^p)$ and the design space $S = [-1, 1]$. By Theorem 2.6, the optimal minimax extrapolation design density for OLS has the form

$$k_*(x) = \left[\frac{\left(\sum_{i=0}^{p}\beta_i x^i\right)\left(\sum_{i=0}^{p}\gamma_i x^i\right) + \lambda}{\left(\sum_{i=0}^{p}\beta_i x^i\right)^2 + t\left(\sum_{i=0}^{p}\beta_i x^i\right)^4}\right]^+,$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)^T \neq \mathbf{0}$, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, ..., \gamma_p)^T \neq \mathbf{0}$, and $t > 0$. The minimax design we obtained is ξ_* with density $\xi_*'(x) = k_*(x)$.

For $p = 1$, the minimax optimal design density has the form

$$m_*(x) = \left[\frac{(a_1 x + a_2)(a_3 x + a_4) + a_5}{(a_3 x + a_4)^2 + a_6(a_3 x + a_4)^4}\right]^+ \tag{2.10}$$

where $a_6 > 0$. Figure 6 gives plots of the minimax extrapolation design densities for varying x_0 and ν with $a_4 = 1$ when $r_{T,S} = 1$. A smaller ν (more emphasis on bias) results in the minimax design becoming more uniform, while a larger ν results in a design resembling that which minimizes variance alone. An extrapolation point x_0 closer to one end of the design space leads to more design points being placed on the corresponding side of the design space. As the distance between x_0 and S increases

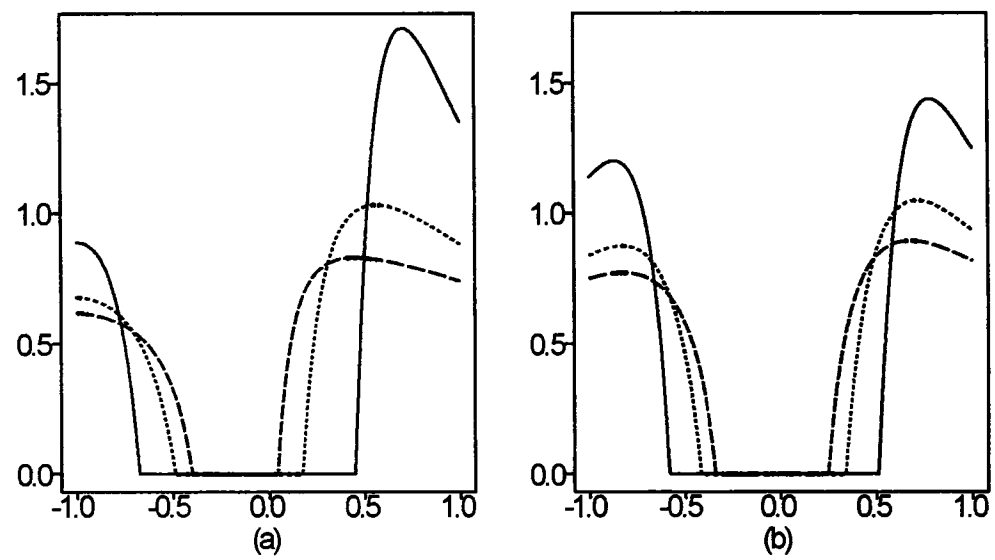

Figure 6: Optimal minimax design densities $m_*(x) = \left[\frac{(a_1x+a_2)(a_3x+1)+a_5}{(a_3x+1)^2+a_6(a_3x+1)^4}\right]^+$ in Example 2 with $p = 1$. (a) $x_0 = 1.5$; (b) $x_0 = 5$. Each plot uses three values of ν: $\nu = 10$ (solid line), $\nu = 1$ (dotted line), $\nu = 0.5$ (broken line).

the design tends to become more symmetric. See Table 2.2 for some numerical values of the constants.

When $p = 2$, the minimax optimal design density has the form

$$m_*(x) = \left[\frac{(a_0 + a_1x + a_2x^2)(b_0 + b_1x + b_2x^2) + c}{(a_0 + a_1x + a_2x^2)^2 + d(a_0 + a_1x + a_2x^2)^4}\right]^+ \tag{2.11}$$

where $d > 0$. See Figure 7 for plots of, and Table 2.3 for numerical values with $a_0 = 1$ for, the minimax extrapolation design densities for varying $x_0 > 1$ and ν when $r_{T,S} = 1$. We observe the same qualitative comparisons as when $p = 1$.

We now compare the minimax design ξ_* found in Example 2 with two common competitors. Let ξ_{HL} be the Hoel-Levine design (Hoel & Levine, 1964) which was derived under the assumption of an exactly correct fitted model. Let ξ_U be the continuous uniform design on $[-1, 1]$.

Table 2.2. Numerical values for m_* of Example 2 with $p = 1$.

x_0	ν	a_1	a_2	a_3	a_5	a_6
1.5	0.25	3.78	0.853	5.96	−0.294	0.00116
	0.5	6.59	1.51	7.83	−2.43	0.00253
	1	14.06	3.18	11.23	−16.52	0.00320
	10	294.38	55.93	30.15	−2758.06	0.00537
	100	1247.25	223.60	39.53	−17826.34	0.00946
5	0.25	26.55	1.84	31.15	−34.92	0.000183
	0.5	54.39	3.72	45.58	−213.28	0.000188
	1	148.64	9.73	80.38	−1712.68	0.000122
	10	751.39	42.48	66.96	−15013.05	0.00124
	100	2138.31	117.57	45.21	−32404.40	0.0150

Table 2.3. Numerical values for m_* of Example 2 with $p = 2$.

x_0	ν	b_0	b_1	b_2	a_1	a_2	c	d
1.5	0.25	0.668	−0.521	−2.37	−0.192	−2.44	−0.102	0.123
	0.5	0.853	−0.644	−2.99	−0.127	−2.30	−0.200	0.280
	1	1.23	−0.858	−4.10	−0.0710	−2.17	−0.396	0.627
	10	8.20	−3.77	−24.49	0.0358	−1.92	−3.83	8.03
	100	77.71	−30.61	−223.96	0.0628	−1.86	−37.47	86.13
5	0.25	0.829	−0.144	−2.54	−0.0323	−2.29	−0.116	0.232
	0.5	1.12	−0.186	−3.40	−0.0191	−2.19	−0.237	0.472
	1	1.69	−0.255	−5.01	−0.0079	−2.11	−0.482	0.949
	10	11.68	−1.20	−31.87	0.0147	−1.92	−4.89	9.64
	100	111.06	−10.02	−296.67	0.0204	−1.88	−48.76	97.34

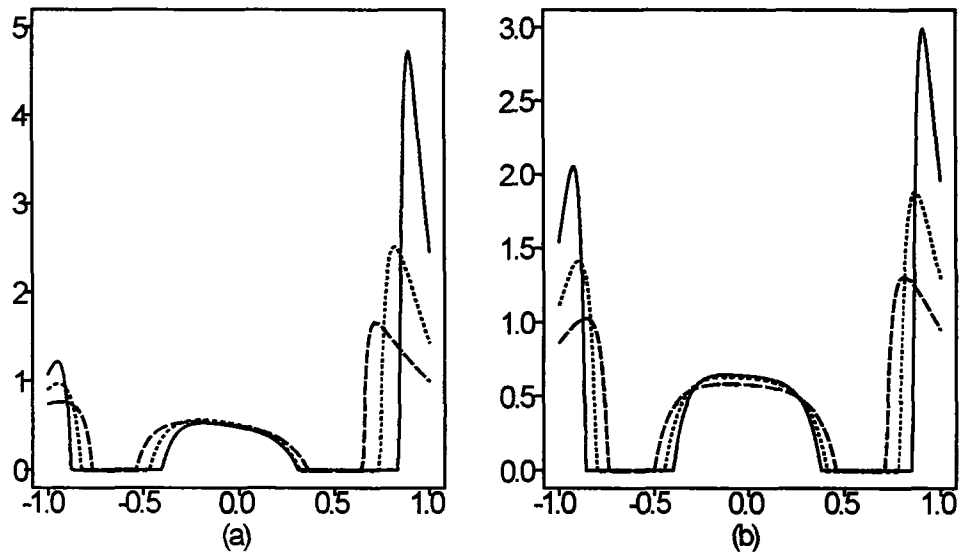

Figure 7: Optimal minimax design densities $m(x) = \left[\frac{(1+a_1x+a_2x^2)(b_0+b_1x+b_2x^2)+c}{(1+a_1x+a_2x^2)^2+d(1+a_1x+a_2x^2)^4}\right]^+$ in Example 2 for $p = 2$. (a) for $x_0 = 1.5$; (b) for $x_0 = 5$. Each plot uses three values of ν: $\nu = 10$ (solid line), $\nu = 1$ (dotted line), $\nu = 0.25$ (broken line).

When $p = 1$, the design points of ξ_{HL} are $x_1 = -1$ and $x_2 = +1$ with mass $\xi_{HL}(-1) = \frac{x_0-1}{2x_0}$ and $\xi_{HL}(1) = \frac{x_0+1}{2x_0}$. In Figure 8(a), we compare the loss for our minimax design ξ_* with that of ξ_{HL} and that of ξ_U when the model is exactly correct, for varying x_0. When the model may contain response contamination and heteroscedastic errors, ξ_{HL} has $sup_{f,g}MSEE = \infty$. Figure 8(b) provides plots of $sup_{f,g}MSEE$ for ξ_U and ξ_* when the model contamination is maximal.

For the calculation of the loss at the nominal model we note that when the fitted model $E(Y|x) = \mathbf{z}^T(x)\boldsymbol{\theta}_0$ is correct and the variances are homogeneous, the OLS estimates are unbiased and the loss is

$$\eta_S^{-2}MSEE(f = \mathbf{0}, g = \mathbf{1}, w = \mathbf{1}, \xi) = \eta_S^{-2}EV(\mathbf{1}, \mathbf{1}, \xi) = \nu\mathbf{z}^T(x_0)\mathbf{B}_*^{-1}\mathbf{z}(x_0),$$

where $\mathbf{B}_* = (b_{ij})_{(p+1)\times(p+1)}$ with element $b_{ij} = \int_{-1}^{1} x^{i+j-2}\xi_*(dx)$. For the minimax design ξ_*, $\sup_{f,g}MSEE$ is given by (2.6). For the uniform design ξ_U, Theorem 2.3

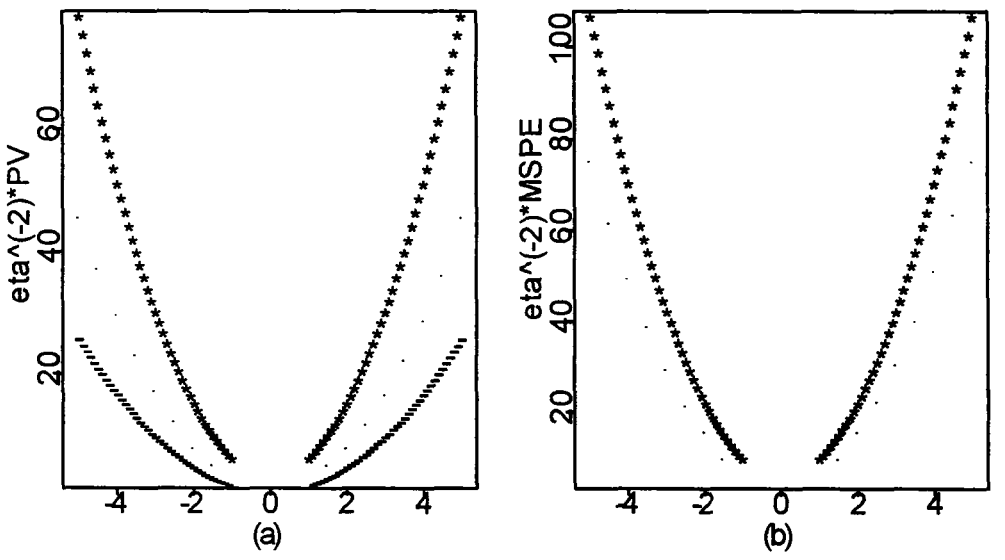

Figure 8: Comparisons of loss for a nominal straight line response; $\nu = 1$, $r_{T,S} = 1$. (a) $\eta_S^{-2} PV$ vs. x_0 (b) $\eta_S^{-2} sup_{f,g} MPSE$ vs. x_0. Each plot uses a dotted line for ξ_*, and an asterisked line for ξ_U. A broken line is used for ξ_{HL} in plot (a).

gives

$$\eta_S^{-2} \sup_{f,g} MSEE(f, g, w = \mathbf{1}, \xi) = {r_{T,S}}^2 + 2\sqrt{2}\nu \left[\int_S \{\mathbf{z}^T(x)\mathbf{A}_S^{-1}\mathbf{z}(x_0)\}^4 dx \right]^{1/2}.$$

For $p = 2$, ξ_{HL} has three design points: $x_1 = -1, x_2 = 0$, and $x_3 = +1$ with mass $\xi_{HL}(-1) = \frac{x_0(x_0-1)}{2(2x_0^2-1)}$, $\xi_{HL}(0) = \frac{(x_0+1)(x_0-1)}{2x_0^2-1}$ and $\xi_{HL}(1) = \frac{x_0(x_0+1)}{2(2x_0^2-1)}$. In Figure 9(a), we compare $\eta_S^{-2} EV$ for our minimax design ξ_* with that of ξ_{HL} and that of ξ_U when the model is exactly correct. Figure 9(b) provides plots of $\eta_S^{-2} sup_{f,g} MSEE$ for design ξ_U and ξ_* when the model contamination is maximal.

2.5.2 Minimax Designs for WLS: Solution to P3

The solution to P3 is provided by Theorem 2.7 below.

Theorem 2.7 *The minimizing $m_*(\mathbf{x})$ in (2.7) for WLS estimation is of the form*

$$m_*(\mathbf{x}) = \frac{\left[\left(\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\gamma} \right) \left(\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta} \right) + \lambda - d(\mathbf{x}) \right]^+}{\left(\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta} \right)^2},$$

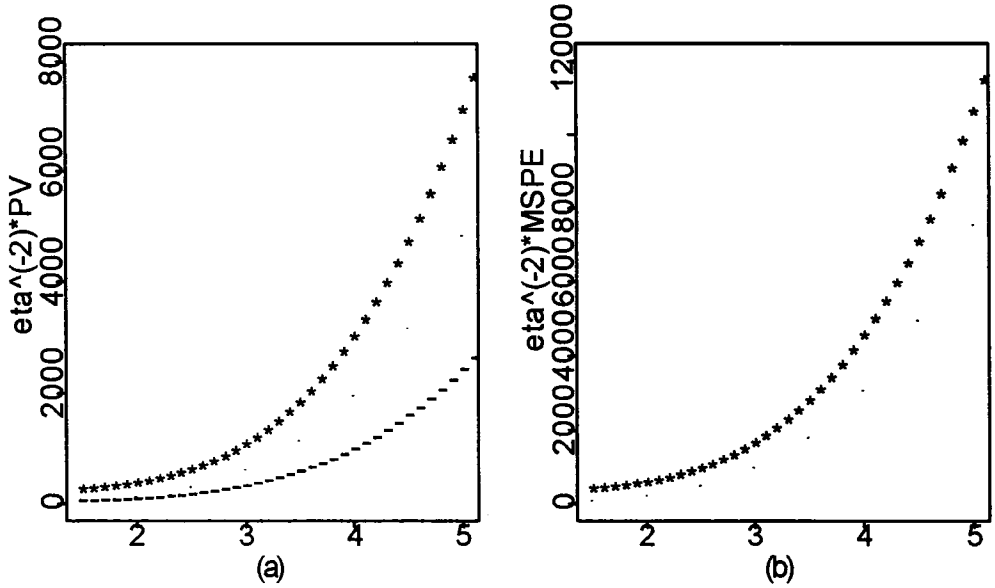

Figure 9: Comparisons of loss for a nominal quadratic response; $\nu = 1$, $r_{T,S} = 1$. (a) $\eta_S^{-2} PV$ vs. x_0 (b) $\eta_S^{-2} sup_{f,g} MPSE$ vs. x_0. Each plot uses a dotted line for ξ_*, and an asterisked line for ξ_U. A broken line is used for ξ_{HL} in plot (a).

where d satisfies the cubic equation

$$d^3 + t(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})^2 d - t(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})^2 \left[(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\gamma}) + \lambda\right] = 0.$$

Explicitly,

$$d(\mathbf{x}) = \left(\frac{t}{2}(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})^2\right)^{1/3} \left[\left\{ \frac{\left[(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\gamma}) + \lambda\right] + \sqrt{\left[(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\gamma}) + \lambda\right]^2 + \frac{4t}{27}(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})^2}} \right\}^{1/3} + \left\{ \frac{\left[(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\gamma}) + \lambda\right] - \sqrt{\left[(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\gamma}) + \lambda\right]^2 + \frac{4t}{27}(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})^2}} \right\}^{1/3} \right].$$

The $p \times 1$ nonzero vectors $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, and constants λ and $t > 0$ satisfy: (i) $\int_S m_*(\mathbf{x})d\mathbf{x} = 1$, (ii) minimize (2.7).

Then, (2.8) and (2.9) provide the optimal one-point extrapolation regression weights and design density for WLS estimation respectively.

Example 3. Consider an approximate polynomial model as in Example 2. By Theorem 2.7, the optimal minimax $m_*(x) = k_*(x)w_*(x)$ for WLS has the form

$$m_*(x) = \frac{\left[\left(\sum_{i=0}^{p}\beta_i x^i\right)\left(\sum_{i=0}^{p}\gamma_i x^i\right) + \lambda - d(x)\right]^+}{\left(\sum_{i=0}^{p}\beta_i x^i\right)^2},$$

where d satisfies

$$d^3 + t\left(\sum_{i=0}^{p}\beta_i x^i\right)^2 d - t\left(\sum_{i=0}^{p}\beta_i x^i\right)^2\left[\left(\sum_{i=0}^{p}\beta_i x^i\right)\left(\sum_{i=0}^{p}\gamma_i x^i\right) + \lambda\right] = 0.$$

The minimax design ξ_* has density $\xi_*'(x) = k_*(x)$ computed from (2.9). The minimax weights $w_*(x)$ are obtained from (2.8). Assuming a non-zero intercept we can, without loss of generality, take $\beta_0 = 1$.

For $p = 1$, the minimax optimal $m_*(\mathbf{x})$ has the form of

$$m_*(x) = \left[\frac{(1 + a_1x)(a_2 + a_3x) + \lambda - d(x)}{(1 + a_1x)^2}\right]^+ \tag{2.12}$$

where

$$d(x) = \left(\frac{t}{2}(1 + a_1x)^2\right)^{1/3}\left[\left\{\frac{[(1 + a_1x)(a_2 + a_3x) + \lambda] + }{\sqrt{[(1 + a_1x)(a_2 + a_3x) + \lambda]^2 + \frac{4t}{27}(1 + a_1x)^2}}\right\}^{1/3} + \left\{\frac{[(1 + a_1x)(a_2 + a_3x) + \lambda] - }{\sqrt{[(1 + a_1x)(a_2 + a_3x) + \lambda]^2 + \frac{4t}{27}(1 + a_1x)^2}}\right\}^{1/3}\right]$$

with $t > 0$ and $a_2^2 + a_3^2 > 0$.

See Table 2.4 for numerical values of the constants. Figure 10 gives plots of the minimax extrapolation design densities for $S = [-1, 1]$ and varying $x_0 > 1$. For $p = 2$, the minimax optimal $m_*(\mathbf{x})$ has the form

$$m_*(x) = \left[\frac{(1 + a_1x + a_2x^2)(a_3 + a_4x + a_5x^2) + \lambda - d(x)}{(1 + a_1x + a_2x^2)^2}\right]^+ \tag{2.13}$$

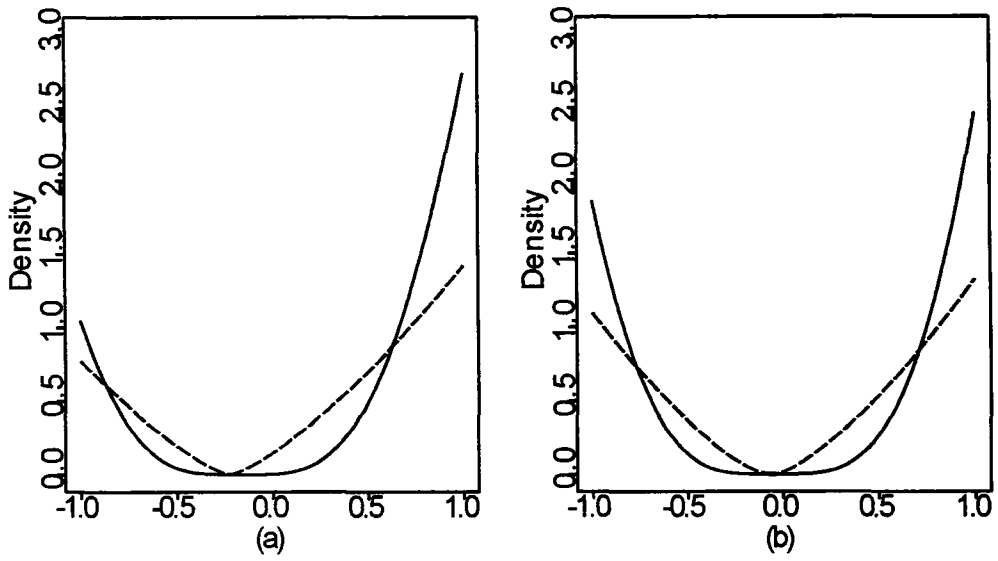

Figure 10: Optimal extrapolation design densities for WLS and simple linear regression: (a) $x_0 = 1.5$; (c) $x_0 = 5$. Each plot uses two values of ν: $\nu = 10$ (solid line), $\nu = 0.5$ (dotted line).

where

$$d(x) = \left(\frac{t}{2}(1 + a_1x + a_2x^2)^2\right)^{1/3} \left[\left\{ \frac{[(1 + a_1x + a_2x^2)(a_3 + a_4x + a_5x^2) + \lambda]}{+\sqrt{[(1 + a_1x + a_2x^2)(a_3 + a_4x + a_5x^2) + \lambda]^2 + \frac{4t}{27}(1 + a_1x + a_2x^2)^2}} \right\}^{1/3} + \left\{ \frac{[(1 + a_1x + a_2x^2)(a_3 + a_4x + a_5x^2) + \lambda]}{-\sqrt{[(1 + a_1x + a_2x^2)(a_3 + a_4x + a_5x^2) + \lambda]^2 + \frac{4t}{27}(1 + a_1x + a_2x^2)^2}} \right\}^{1/3} \right]$$

with $t > 0$ and $a_3^2 + a_4^2 + a_5^2 > 0$.

See Table 2.5 for some numerical values of the constants, and Figure 11 for plots.

Table 2.4. Numerical values for Example 3 with $p = 1$, $r_{T,S} = 1$, $S = [-1, 1]$.

x_0	ν	a_1	a_2	a_3	t	λ
1.5	0.25	0.287	0.0136	0.388	0.0177	0.675
	0.5	−0.0118	0.00215	−0.0124	0.000161	0.541
	1	−0.255	1.70	−0.966	68.73	2.00
	10	8.69	8.78	53.96	8625.44	0.000230
	100	16.09	149.83	1023.40	47926350	−0.139
5	0.25	40.37	0.634	24.34	0.280	0.000618
	0.5	34.69	0.670	21.03	0.227	−0.00118
	1	18.99	1.30	21.97	31.95	−0.000154
	10	1.47	6.51	1.46	108419.4	−0.00368
	100	1.07	7.04	1.04	68740870	−0.367

Table 2.5. Numerical values for Example 3 with $p = 2$, $r_{T,S} = 1$, $S = [-1, 1]$.

x_0	ν	a_1	a_2	a_3	a_4	a_5	t	λ
1.5	0.25	0.469	2.47	0.00250	0.411	2.18	0.0300	0.767
	0.5	0.514	2.83	0.199	0.751	4.18	5.02	1.73
	1	0.533	3.15	0.0207	1.64	9.80	141.91	4.87
	10	−1.25	0.466	0.000150	−1.52	−0.30	0.0000641	0.477
	100	−2.96	1.92	0.888	−1.90	0.617	0.034	−0.00813
5	0.25	0.0665	1.76	0.173	0.122	3.38	10.87	2.19
	0.5	0.0779	2.14	0.288	0.376	10.97	645.82	7.55
	1	0.0878	2.66	0.962	12.87	416.62	36,493,560	285.96
	10	−0.000741	−0.603	0.326	0.210	−11.52	759.74	7.84
	100	−0.00100	−0.617	0.220	0.632	−36.26	21033.71	23.69

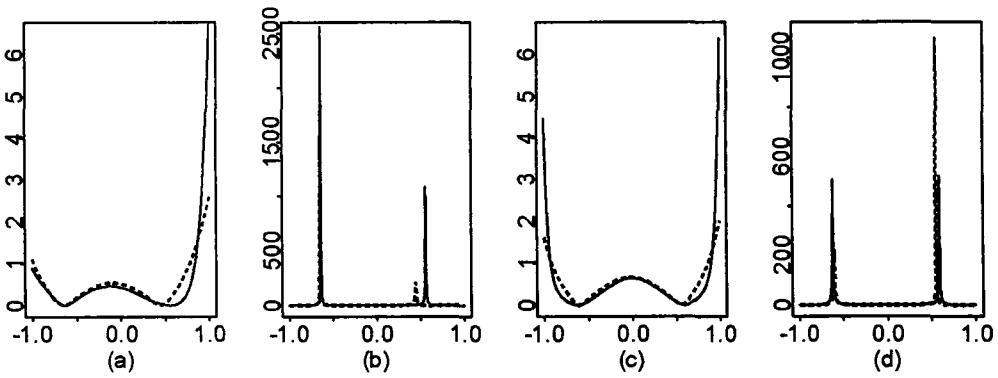

Figure 11: Optimal extrapolation design densities and minimax weights for WLS and quadratic regression: (a) design densities for $x_0 = 1.5$; (b) minimax weights for $x_0 = 1.5$; (c) design densities for $x_0 = 5$; (d) minimax weights for $x_0 = 5$. Each plot uses two values of ν: $\nu = 10$ (solid line), $\nu = 0.1$ (dotted line).

2.6 Optimal Unbiased Designs: Solution to P4

We say that a design/weights pair (ξ, w) is *unbiased* if it satisfies $EB(f, w, \xi) = 0$ for all $f \in \mathcal{F}$, so that $\sup_{f \in \mathcal{F}} EB(f, w, \xi) = 0$. The following theorem, which is essentially Theorem 2.2(b) of Fang & Wiens (1999), gives a necessary and sufficient condition for unbiasedness.

Theorem 2.8 *The pair (w, ξ) is unbiased if and only if*

$$m(\mathbf{x}) \equiv \Omega = \frac{1}{\int_S d\mathbf{x}}.$$

We can construct the optimal unbiased extrapolation design $m_0(\mathbf{x})$ by forcing $\sup_{f \in \mathcal{F}} EB(f, w, \xi) = 0$, and then minimizing $\sup_{g \in \mathcal{G}} EV(g, w_m, \xi)$. Thus let $m_0(\mathbf{x}) \equiv \Omega$, i.e. $k_0(\mathbf{x}) w_0(\mathbf{x}) \equiv \Omega$. From Theorem 2.4, the optimal weight function is

$$w_0(\mathbf{x}) = w_{m_0}(\mathbf{x}) = \Omega \alpha_{m_0} [\tilde{\mathbf{z}}^T(\mathbf{x}) \mathbf{A}_S^{-1} \tilde{\mathbf{z}}(\mathbf{x}_0)]^{-4/3}, \tag{2.14}$$

and the optimal unbiased extrapolation design density is

$$k_0(\mathbf{x}) = \alpha_{m_0}^{-1} [\tilde{\mathbf{z}}^T(\mathbf{x}) \mathbf{A}_S^{-1} \tilde{\mathbf{z}}(\mathbf{x}_0)]^{4/3},$$

with

$$\alpha_{m_0} = \int_S [\tilde{\mathbf{z}}^T(\mathbf{x})\mathbf{A}_S^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)]^{4/3} d\mathbf{x}.$$

The minimax $MSEE$ is

$$\min_{(w,\xi)} \sup_{f\in\mathcal{F},g\in\mathcal{G}} MSEE(f,g,w,m) = \eta_T^2 + \min_{(w,\xi)} \sup_{g\in\mathcal{G}} EV(g,w_m,\xi) = \eta_S^2 \left\{ r_{T,S}^2 + \nu\Omega^{-1/2}\alpha_{m_0}^{3/2} \right\}.$$

We summarize these observations below.

Theorem 2.9 *The density $k_0(\mathbf{x})$ of the optimal unbiased one-point extrapolation design measure ξ_0, and optimal weights w_0, which minimize $\sup_{f\in\mathcal{F},g\in\mathcal{G}} MSEE(f,g,w,\xi)$ subject to $\sup_{f\in\mathcal{F}} EB(f,w,\xi) = 0$ are given by*

$$k_0(\mathbf{x}) = \frac{[\tilde{\mathbf{z}}^T(\mathbf{x})\mathbf{A}_S^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)]^{4/3}}{\int_S [\tilde{\mathbf{z}}^T(\mathbf{x})\mathbf{A}_S^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)]^{4/3} d\mathbf{x}},$$

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

$$\begin{aligned} &\sup_{f\in\mathcal{F},g\in\mathcal{G}} MSEE(f,g,w_0,\xi_0) \\ &= \eta_S^2 \left\{ r_{T,S}^2 + \nu\Omega^{-1/2} \left[\int_S [\tilde{\mathbf{z}}^T(\mathbf{x})\mathbf{A}_S^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)]^{4/3} d\mathbf{x} \right]^{3/2} \right\}, \end{aligned} \tag{2.15}$$

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

Example 4 Consider an approximate log-linear multiple regression model $E(Y|\mathbf{x})$ $\approx \exp(\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}_0) = \exp(\theta_0 + \theta_1 x_1 + ... + \theta_q x_q)$.

Note that the designs provided by Theorem 2.9 for this example depend on $\boldsymbol{\theta}_1 = (\theta_1, ..., \theta_q)^T$ but not on θ_0. As in Example 1 we can find locally most robust designs in a neighbourhood Θ of a starting value $\boldsymbol{\theta}_1^{(0)}$. We first construct the design $k_0(\mathbf{x}, \boldsymbol{\theta}_1^{(0)})$ and weights $\Omega/k_0(\mathbf{x}, \boldsymbol{\theta}_1^{(0)})$ provided by Theorem 2.9. We then find the least favourable $\boldsymbol{\theta}_{1,LF}$ in Θ. From Theorem 2.3, we find that this is equivalent to maximizing

$$\int_S \frac{\{\tilde{\mathbf{z}}^T(\mathbf{x})\mathbf{A}_S^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)\}^4}{k_0^2(\mathbf{x}, \boldsymbol{\theta}_1^{(0)})} d\mathbf{x}$$

over the occurrences of $\boldsymbol{\theta}_1$ in the numerator of the integrand. We then construct the unbiased optimal design for $\boldsymbol{\theta}_{LF}$, and then iterate to convergence.

When $q = 1$, $S = [-0.5, 0.5]$ the unbiased minimax design density is

$$k_0(x) = \begin{cases} \frac{\left\{e^{\theta_{LF}x}[(c-bx_0)+(ax_0-b)x]\right\}^{4/3}}{\int_{-0.5}^{0.5}\left\{e^{\theta_{LF}x}[(c-bx_0)+(ax_0-b)x]\right\}^{4/3}dx}, & \text{when } \theta_{LF} \neq 0, \\ \frac{(1+12x_0x)^{4/3}}{\int_{-0.5}^{0.5}(1+12x_0x)^{4/3}dx}, & \text{when } \theta_{LF} = 0, \end{cases}$$

where

$$a = \frac{\sinh(\theta_{LF})}{\theta_{LF}}, b = \frac{\cosh(\theta_{LF}) - a}{2\theta_{LF}}, c = \frac{0.5\sinh(\theta_{LF}) - 2b}{2\theta_{LF}}.$$

For a simple demonstration of the procedure described above, we take $\Theta = [0.5, 0.7]$ and consider the cases $x_0 = \pm 2$ and $x_0 = \pm 9$. For both $x_0 = 2$ and $x_0 = 9$, the iterates converge to $\theta_{LF} = 0.7$. The unbiased minimax design density at $x_0 = 2$ is

$$k_0(x) = \frac{\{e^{0.7x}(.149 - 2.045x)\}^{4/3}}{\int_{-0.5}^{0.5}\{e^{0.7x}(.149 - 2.045x)\}^{4/3}dx};$$

and that at $x_0 = 9$ is

$$k_0(x) = \frac{\{e^{0.7x}(1.006 - 9.631x)\}^{4/3}}{\int_{-0.5}^{0.5}\{e^{0.7x}(1.006 - 9.631x)\}^{4/3}dx}.$$

When $x_0 = -2$ and $x_0 = -9$, we find $\theta_{LF} = 0.5$. The unbiased minimax design density at $x_0 = -2$ is

$$k_0(x) = \frac{\{e^{0.5x}(.261 - 2.170x)\}^{4/3}}{\int_{-0.5}^{0.5}\{e^{0.5x}(.261 - 2.170x)\}^{4/3}dx};$$

and that at $x_0 = -9$ is

$$k_0(x) = \frac{\{e^{0.5x}(.859 - 9.465x)\}^{4/3}}{\int_{-0.5}^{0.5}\{e^{0.5x}(.859 - 9.465x)\}^{4/3}dx}.$$

The corresponding optimal weights are $w_0(x) = 1/k_0(x)$.

Example 5. Consider an approximate polynomial regression model $E(Y|x) \approx \mathbf{z}^T(x)\boldsymbol{\theta}_0 = \theta_0 + \theta_1 x + ... + \theta_p x^p$ with $S = [-1, 1]$. By Theorem 2.9, the unbiased optimal density is

$$k_0(x) = \frac{[\mathbf{z}^T(x)\mathbf{A}_S^{-1}\mathbf{z}(x_0)]^{4/3}}{\int_S [\mathbf{z}^T(x)\mathbf{A}_S^{-1}\mathbf{z}(x_0)]^{4/3}d\mathbf{x}}$$

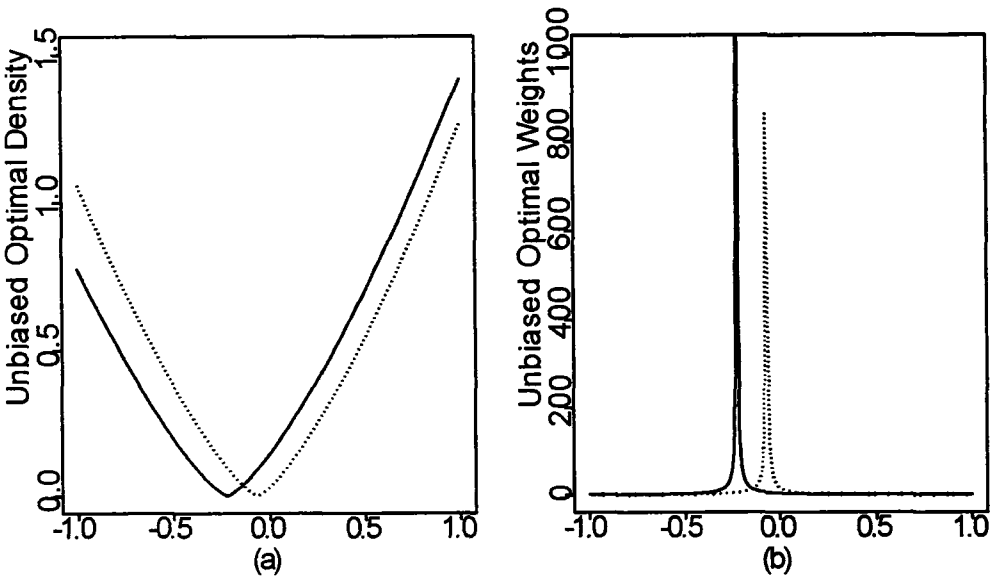

Figure 12: Unbiased optimal densities and weights for SLR: (a) design densities; (b) weights. Each plot uses two values of x_0: $x_0 = 1.5$ (solid line) and $x_0 = 5$ (dotted line).

with optimal weights $w_0(x) = .5/k_0(x)$.

When $p = 1$, we have the design

$$k_0(x) = \frac{3.5x_0(0.5 + 1.5x_0x)^{4/3}}{(0.5 + 1.5x_0)^{7/3} - (0.5 - 1.5x_0)^{7/3}}.$$

See Figure 12 for plots.

When $p = 2$, the design is

$$k_0(x) = \frac{[(1.125 - 1.875x_0^2) + 1.5x_0x + (5.625x_0^2 - 1.875)x^2]^{4/3}}{\int_{-1}^{1}[(1.125 - 1.875x_0^2) + 1.5x_0x + (5.625x_0^2 - 1.875)x^2]^{4/3}\,dx}.$$

For plots, see Figure 13.

2.7 Comparisons and Remarks

In Examples 2 we compared our designs for P2 with two more conventional competing designs ξ_{HL} and ξ_U. In this section, we use the approximate polynomial models ($p = 1, 2$) of these examples to compare the robust minimax designs for P2, P3 and

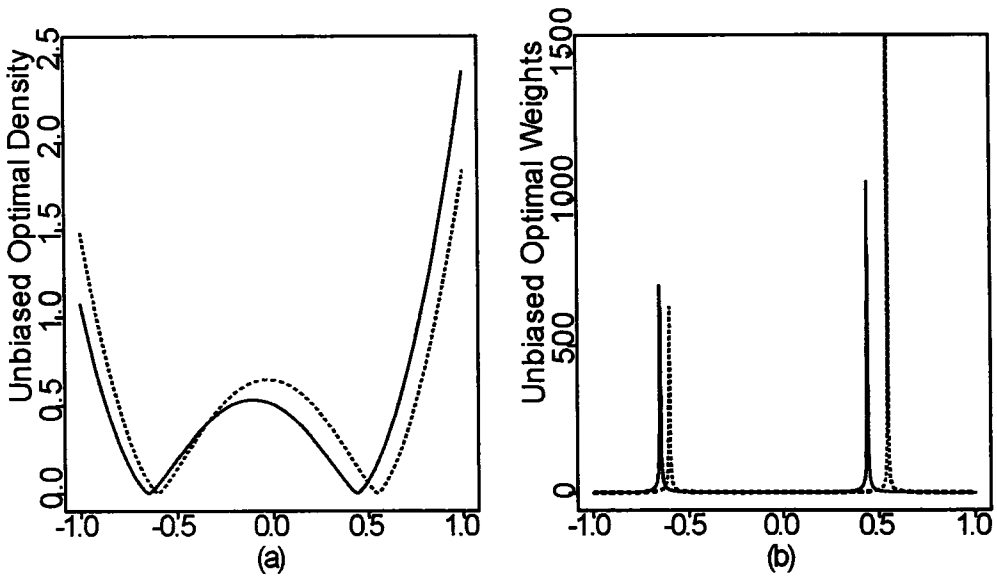

Figure 13: Unbiased optimal design densities and weights for quadratic regression: (a) densities; (b) weights. Each plot uses two values of x_0: $x_0 = 1.5$ (solid line) and $x_0 = 10$ (dotted line).

P4 with each other and again with ξ_{HL} and ξ_U. Let $\xi^{(2)}$, $\xi^{(3)}$, and $\xi^{(4)}$ denote the robust optimal designs that we obtained for P2, P3 and P4 respectively. Table 2.6 gives the comparative values of $\eta_S^{-2} EV$ when there is no contamination and Table 2.7 gives those of $\eta_S^{-2} \sup_{f,g} MSEE$ for ξ_U when there is maximal contamination. Of course $\sup_{f,g} MSEE$ for ξ_{HL} is infinite.

When there is no contamination, we denote by $re_{HL}^{(0)}\left(\xi^{(\cdot)}\right)$ the efficiencies of $\xi^{(2)}$, $\xi^{(3)}$ and $\xi^{(4)}$ relative to ξ_{HL} and by $re_U^{(0)}\left(\xi^{(\cdot)}\right)$ the efficiencies relative to ξ_U. Under maximal contamination we write instead $re_U^{(\max)}\left(\xi^{(\cdot)}\right)$ and $re_{HL}^{(\max)}\left(\xi^{(\cdot)}\right) = \infty$. Table 2.8 provides the relative efficiencies $re_{HL}^{(0)}$ and $re_U^{(0)}$ while Table 2.9 provides the relative efficiencies $re_U^{(\max)}\left(\xi^{(\cdot)}\right)$.

We have provided methods of constructing optimally robust designs for one-point regression extrapolation, taking into account various model uncertainties. The results require extensive numerical work prior to implementation. However, we can give some

Table 2.6. Comparative values of $\eta_S^{-2}EV$
when there is no contamination.

$p=1$						
x_0	ν	$\xi^{(2)}$	$\xi^{(3)}$	$\xi^{(4)}$	ξ_{HL}	ξ_U
1.5	0.25	1.57	1.44	1.44	0.563	1.94
	0.5	2.70	2.88	2.88	1.13	3.88
	1	4.71	5.68	5.76	2.25	7.75
	10	36.47	42.00	57.62	22.5	77.5
	100	347.05	398.60	576.19	225	775
5	0.25	13.77	14.14	14.60	6.25	19
	0.5	24.81	28.27	29.21	12.5	38
	1	45.28	51.06	58.41	25	76
	10	383.80	432.99	584.13	250	760
	100	3726.18	4197.94	5841.26	2500	7600
$p=2$						
x_0	ν	$\xi^{(2)}$	$\xi^{(3)}$	$\xi^{(4)}$	ξ_{HL}	ξ_U
1.5	0.25	7.30	7.96	8.41	3.06	12.27
	0.5	13.16	15.29	16.82	6.13	24.54
	1	23.87	28.82	33.65	12.25	49.08
	10	198.74	261.10	336.49	122.5	490.78
	100	1923.71	2424.7	3364.90	1225	4907.81
5	0.25	1140.24	1240.23	1311.03	600.25	1730.25
	0.5	2112.96	2387.43	2622.06	1200.5	3460.5
	1	3950.39	4564.29	5244.11	2401.00	6921
	10	35092.10	42875.19	52441.14	24010	69210
	100	343816.60	417639.3	524411.4	240100	692100

Table 2.7. Comparative values of $\eta_S^{-2} \sup_{f,g} MSEE$ when there is maximal contamination.

			$p = 1$		
x_0	ν	$\xi^{(2)}$	$\xi^{(3)}$	$\xi^{(4)}$	ξ_U
1.5	0.25	3.50	2.59	2.59	3.78
	0.5	5.58	4.17	4.17	6.56
	1	9.27	7.27	7.35	12.12
	10	66.91	58.24	64.49	112.15
	100	629.58	552.78	635.90	1112.55
5	0.25	22.42	16.65	16.95	26.71
	0.5	40.69	32.30	32.90	52.43
	1	74.91	62.85	64.80	103.86
	10	648.81	575.17	638.96	1029.60
	100	6330.03	5643.09	6380.57	10286.98
			$p = 2$		
x_0	ν	$\xi^{(2)}$	$\xi^{(3)}$	$\xi^{(4)}$	ξ_U
1.5	0.25	19.50	14.71	10.39	23.10
	0.5	30.85	23.61	19.79	45.20
	1	51.63	40.85	38.58	89.40
	10	384.07	336.75	376.77	884.98
	100	3629.60	3211.04	3758.73	8840.79
5	0.25	1833.70	1422.04	1430.95	2586.07
	0.5	3410.07	2780.92	2860.90	5171.13
	1	6405.83	5432.51	5720.81	10341.27
	10	57275.39	52036.29	57199.07	103403.65
	100	561056.9	512051.6	571981.67	1034027.5

Table 2.8. Relative efficiencies $\mathrm{re}_{HL}^{(0)}\left(\xi^{(\cdot)}\right)$ and $\mathrm{re}_{U}^{(0)}\left(\xi^{(\cdot)}\right)$.

$p=1$				
x_0	ν	$\xi^{(2)}$: HL/U	$\xi^{(3)}$: HL/U	$\xi^{(4)}$: HL/U
1.5	0.25	0.36/1.14	0.39/1.35	0.39/1.35
	0.5	0.42/1.44	0.39/1.35	0.39/1.35
	1	0.48/1.65	0.40/1.36	0.39/1.35
	10	0.62/2.13	0.54/1.85	0.39/1.35
	100	0.65/2.23	0.56/1.94	0.39/1.35
5	0.25	0.45/1.38	0.44/1.34	0.43/1.30
	0.5	0.50/1.53	0.44/1.34	0.43/1.30
	1	0.55/1.68	0.49/1.49	0.43/1.30
	10	0.65/1.98	0.58/1.76	0.43/1.30
	100	0.67/2.04	0.59/1.81	0.43/1.30
$p=2$				
x_0	ν	$\xi^{(2)}$: HL/U	$\xi^{(3)}$: HL/U	$\xi^{(4)}$: HL/U
1.5	0.25	0.42/1.68	0.38/1.54	0.36/1.46
	0.5	0.47/1.86	0.40/1.60	0.36/1.46
	1	0.51/2.06	0.43/1.70	0.36/1.46
	10	0.62/2.47	0.47/1.88	0.36/1.46
	100	0.64/2.55	0.51/2.02	0.36/1.46
5	0.25	0.53/1.52	0.48/1.40	0.46/1.32
	0.5	0.57/1.64	0.50/1.45	0.46/1.32
	1	0.61/1.75	0.53/1.52	0.46/1.32
	10	0.68/1.97	0.56/1.61	0.46/1.32
	100	0.70/2.01	0.57/1.66	0.46/1.32

Table 2.9. Relative efficiencies $re_{U}^{(\max)}\left(\xi^{(\cdot)}\right)$

		$p=1$			$p=2$		
x_0	ν	$\xi^{(2)}$	$\xi^{(3)}$	$\xi^{(4)}$	$\xi^{(2)}$	$\xi^{(3)}$	$\xi^{(4)}$
1.5	0.25	1.08	1.46	1.46	1.18	1.57	2.22
	0.5	1.18	1.57	1.57	1.47	1.91	2.28
	1	1.31	1.67	1.65	1.73	2.19	2.32
	10	1.68	1.93	1.74	2.30	2.63	2.35
	100	1.77	2.01	1.75	2.44	2.75	2.35
5	0.25	1.19	1.60	1.58	1.41	1.82	1.81
	0.5	1.29	1.62	1.59	1.52	1.86	1.81
	1	1.39	1.65	1.60	1.61	1.90	1.81
	10	1.59	1.79	1.61	1.81	1.99	1.81
	100	1.63	1.82	1.61	1.84	2.02	1.81

informative guidelines:

1. As v increases, the designs place more emphasis on variance minimization and less on protection from bias. As we would expect, the experimenter should then place relatively more design points closer to the boundary of the design space. With respect to the position of the extrapolation point relative to the design space, the experimenter should place relatively more design points in that segment of the design space which is closer to the extrapolation point, with this prescription becoming more emphatic when the extrapolation point is close to the design space.

2. Compared to designs for variance minimization alone, the designs we have found in this work are substantially more uniform. They can roughly be described as being obtained by replacing the point masses of the variance minimizing designs by uniform densities on regions containing, but not restricted to, these atoms.

3. Under heteroscedasticity the designs for P3 are, as expected, the most efficient. The gains in efficiency are greater when ν is at least moderately large. Particularly for small ν, the numerical simplicity of the designs for P4 make them attractive competitors.

Appendix: Derivations

The proof of Theorem 2.5 is very similar to but simpler than that of Theorem 2.6, and so is omitted.

Proof of Theorem 2.6. We seek a nonnegative function $m(\mathbf{x})$ minimizing (2.6) subject to $\int_S m(\mathbf{x})d\mathbf{x} = 1$. For a Lagrange multiplier s it is necessary and sufficient that m minimize

$$(\sqrt{\lambda_m}+r_{T,S})^2 + \nu\Omega^{-1/2}\left[\int_S \{l_m(\mathbf{x})m(\mathbf{x})\}^2 d\mathbf{x}\right]^{1/2} - 2s\int_S m(\mathbf{x})d\mathbf{x}$$

among all densities, and satisfy the side condition. After a lengthy calculation we obtain the first order condition

$$I := \int_S \{P(\mathbf{x})m(\mathbf{x}) - Q(\mathbf{x}) - u\}\,(m(\mathbf{x}) - m_1(\mathbf{x}))d\mathbf{x} \geq 0 \qquad (2.A.1)$$

for all densities m_1, where

$$P(\mathbf{x}) = \left(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta}\right)^2 \left[1 + t(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})^2\right] \text{ and } Q(\mathbf{x}) = \left(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\gamma}\right)\left(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta}\right),$$

for

$$\begin{aligned}
\boldsymbol{\beta} &= \mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0), \\
t &= \frac{\nu}{2}\Omega^{-1/2}\left(1+\frac{r_{T,S}}{\sqrt{\lambda_m}}\right)^{-1}\left[\int_S \{l_m(\mathbf{x})\,m(\mathbf{x})\}^2\,d\mathbf{x}\right]^{-1/2} > 0, \\
\boldsymbol{\gamma} &= \mathbf{B}^{-1}\left\{\mathbf{K} + \frac{t}{2}\left[\int_S \tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^T(\mathbf{x})l_m(\mathbf{x})\,m^2(\mathbf{x})d\mathbf{x}\right]\right\}\boldsymbol{\beta}, \\
u &= \left(1+\frac{r_{T,S}}{\sqrt{\lambda_m}}\right)^{-1} s.
\end{aligned}$$

To see the consequences of (2.A.1), write S^+ for the subset of S on which $m(\mathbf{x}) > 0$, and $S^0 = S \setminus S^+$. Let $c = \sup_S \{P(\mathbf{x})m(\mathbf{x}) - Q(\mathbf{x}) - u\}$, let $\{\mathbf{x}_j\}$ be a sequence of points in S^+ with $P(\mathbf{x}_j)m(\mathbf{x}_j) - Q(\mathbf{x}_j) - u$ approaching c, and consider a sequence $\{m_{1j}\}$ of point masses at $\mathbf{x}_j$. Then for this sequence (2.A.1) implies

$$\int_{S^+} \{P(\mathbf{x})m(\mathbf{x}) - Q(\mathbf{x}) - u\}\,m(\mathbf{x})d\mathbf{x} \geq c \geq \sup_{S^+} \{P(\mathbf{x})m(\mathbf{x}) - Q(\mathbf{x}) - u\},$$

so that in particular $P(\mathbf{x})m(\mathbf{x}) - Q(\mathbf{x}) - u \equiv c$ on S^+ and $-Q(\mathbf{x}) - u = P(\mathbf{x})m(\mathbf{x}) - Q(\mathbf{x}) - u \leq c$ on S^0. Thus

$$m(\mathbf{x}) = \frac{Q(\mathbf{x}) + u + c}{P(\mathbf{x})},\ \mathbf{x} \in S^+. \qquad (2.A.2)$$

Conversely, if (2.A.2) holds and $Q(\mathbf{x}) + u + c \geq 0$ on S^0 then

$$\begin{aligned}
I &= c\int_{S^+}(m(\mathbf{x}) - m_1(\mathbf{x}))d\mathbf{x} + \int_{S^0}[Q(\mathbf{x}) + u]\,m_1(\mathbf{x})d\mathbf{x} \\
&= c - c\int_{S^+} m_1(\mathbf{x})d\mathbf{x} + \int_{S^0}[Q(\mathbf{x}) + u + c]\,m_1(\mathbf{x})d\mathbf{x} - c\int_{S^0} m_1(\mathbf{x})d\mathbf{x} \\
&= c - c\int_{S} m_1(\mathbf{x})d\mathbf{x} + \int_{S^0}[Q(\mathbf{x}) + u + c]\,m_1(\mathbf{x})d\mathbf{x} \\
&\geq 0,
\end{aligned}$$

satisfying (2.A.1). Thus, in order that (2.A.1) hold, it is necessary and sufficient that (2.A.2) hold, for any c such that the right hand side of (2.A.2) is non-negative throughout S. More generally, m has the form

$$m(\mathbf{x}) = \left(\frac{Q(\mathbf{x}) + u + c}{P(\mathbf{x})}\right)^+ = \frac{\left[\left(\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\gamma}\right)\left(\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta}\right) + \lambda\right]^+}{\left(\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta}\right)^2\left[1 + t(\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})^2\right]}, \tag{2.A.3}$$

with $\lambda = u + c$. Of course $\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda$ and t themselves depend on m. Rather than solve (2.A.3) for m it is simpler merely to choose these constants so as to satisfy $\int_S m(\mathbf{x})d\mathbf{x} = 1$ and minimize (2.6). □

Proof of Theorem 2.7. First, we show that $h_m(\mathbf{x}) = \frac{(\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})(\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\gamma}) + \lambda - d}{(\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})^2}$ minimizes (2.7) subject to $\int_S h_m(\mathbf{x})d\mathbf{x} = 1$ but without the restriction of $h_m(\mathbf{x})$ being nonnegative on S. We introduce a Lagrange multiplier λ. It is sufficient to show that $h_m(\mathbf{x})$ minimizes

$$(\sqrt{\lambda_m} + r_{T,S})^2 + \nu\Omega^{-1/2}\left\{\int_S [l_m(\mathbf{x})\, m^2(\mathbf{x})]^{2/3} d\mathbf{x}\right\}^{3/2} - 2\lambda\int_S m(\mathbf{x})d\mathbf{x}$$

for some constant λ subject to $\int_S h_m(\mathbf{x})(\mathbf{x})d\mathbf{x} = 1$. This is again a simple variational problem. The minimizing $m(\mathbf{x})$ satisfies

$$\int_S \left\{a(\mathbf{x})m^{1/3}(\mathbf{x}) + b(\mathbf{x})m(\mathbf{x}) - c(\mathbf{x})\right\}(m - m_1)d\mathbf{x} = 0,$$

where

$$\begin{aligned}
a(\mathbf{x}) &= \nu\Omega^{-1/2}\left[\int_S l_m^{2/3}(\mathbf{x})\, m^{4/3}(\mathbf{x})d\mathbf{x}\right]^{1/2} l_m^{2/3}(\mathbf{x}), \\
b(\mathbf{x}) &= \left\{1 + \frac{r_{T,S}}{\sqrt{\lambda_m}}\right\} l_m(\mathbf{x}), \\
c(\mathbf{x}) &= \left\{\bar{\mathbf{z}}^{\mathbf{T}}(\mathbf{x})\begin{bmatrix} \left\{1 + \frac{r_{T,S}}{\sqrt{\lambda_m}}\right\}\mathbf{B}^{-1}\mathbf{K}\mathbf{B}^{-1} \\ +\nu\Omega^{-1/2}\left[\int_S l_m^{2/3}(\mathbf{x})\, m^{4/3}(\mathbf{x})d\mathbf{x}\right]^{1/2} \\ \mathbf{B}^{-1}\left\{\int_S \bar{\mathbf{z}}(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x}) l_m^{-1/3}(\mathbf{x})\, m^{4/3}(\mathbf{x})d\mathbf{x}\right\}\mathbf{B}^{-1} \end{bmatrix}\bar{\mathbf{z}}(\mathbf{x}_0)\right\}\left[\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{B}^{-1}\bar{\mathbf{z}}(\mathbf{x}_0)\right] \\
&\quad +\lambda
\end{aligned}$$

for any m_1 with $\int_S m_1(\mathbf{x})d\mathbf{x} = 1$. Therefore, we obtain the minimizing $m(\mathbf{x})$, without the restriction of $m(\mathbf{x}) \geq 0$, as a solution of

$$a(\mathbf{x})m^{1/3}(\mathbf{x}) + b(\mathbf{x})m(\mathbf{x}) - c(\mathbf{x}) = 0. \tag{2.A.4}$$

Let $\tilde{m} = c - bm$. Then, (2.A.4) becomes

$$\tilde{m}^3 = -\frac{a^3}{b}\tilde{m} + \frac{a^3 c}{b}. \tag{2.A.5}$$

Since a and b are positive almost everywhere in S, (2.A.5) has only one real solution. Applying Cardano's formula for cubic equations, we obtain the only real solution for $\tilde{m}$ is

$$d = a\left[\begin{array}{l}\left\{\frac{c}{2b} + \sqrt{\left(\frac{c}{2b}\right)^2 + \left(\frac{a}{3b}\right)^3}\right\}^{1/3} \\ + \left\{\frac{c}{2b} - \sqrt{\left(\frac{c}{2b}\right)^2 + \left(\frac{a}{3b}\right)^3}\right\}^{1/3}\end{array}\right].$$

Then, the minimizing $m(\mathbf{x})$, without restriction of $m(\mathbf{x}) \geq 0$, is:

$$h_m(\mathbf{x}) = \frac{c-d}{b},$$

where λ in c ensures $\int_S h_m(\mathbf{x})d\mathbf{x} = 1$.

The following step is to show the minimizing $m(\mathbf{x})$ with restriction of $m(\mathbf{x}) \geq 0$, is

$$h_m^+(\mathbf{x}) = \left[\frac{c-d}{b}\right]^+.$$

It is sufficient to show that for any nonnegative $h_{m_1}(\mathbf{x})$ in the following form of

$$h_{m_1}(\mathbf{x}) = \begin{cases} h_m(\mathbf{x}), & \text{when } h_m(\mathbf{x}) > 0; \\ \geq 0, & \text{when } h_m(\mathbf{x}) \leq 0, \end{cases}$$

being a density, will give a larger loss than $h_m^+(\mathbf{x})$ does. We claim that $c \leq 0$ almost everywhere in S, when $\frac{c-d}{b} \leq 0$. Then, we have

$$\begin{aligned} & \int_S \left\{a(\mathbf{x})[h_m^+(\mathbf{x})]^{1/3} + b(\mathbf{x})h_m^+(\mathbf{x}) - c\right\}(h_m^+(\mathbf{x})(\mathbf{x}) - h_{m_1}(\mathbf{x}))d\mathbf{x} \\ = & \int_{S\cap h_m(\mathbf{x})\leq 0} \left\{a(\mathbf{x})[h_m^+(\mathbf{x})]^{1/3} + b(\mathbf{x})h_m^+(\mathbf{x}) - c\right\}((h_m^+(\mathbf{x}) - h_{m_1}(\mathbf{x}))d\mathbf{x} \\ = & \int_{S\cap\frac{c-d}{b}\leq 0} ch_{m_1}(\mathbf{x}))d\mathbf{x} \leq 0. \end{aligned}$$

Therefore, $h_m^+(\mathbf{x})$ should give the minimum loss. This gives the minimizing design density $m^*(\mathbf{x})$ with the form of

$$m^*(\mathbf{x}) = \left[\frac{c-d}{b}\right]^+$$

for almost all $\mathbf{x} \in S$.

Now, we prove the claim of $c \leq 0$ almost everywhere in S, when $\frac{c-d}{b} \leq 0$.

The claim is true due to d being a solution of (2.A.5), which means

$$d^3 = -\frac{a^3}{b}d + \frac{a^3 c}{b} = \frac{a^3}{b}(c-d) \leq 0 \text{ almost everywhere, when} \frac{c-d}{b} \leq 0,$$

since $a > 0$ almost everywhere. Due to $b > 0$ almost everywhere, $c \leq d \leq 0$ almost everywhere.

Therefore, the optimal minimax densities are

$$m^*(\mathbf{x}) = \left[\frac{c-d}{b}\right]^+$$

for almost all $\mathbf{x} \in S$, where

$$d = a\left[\begin{array}{l} \left\{\frac{c}{2b} + \sqrt{\left(\frac{c}{2b}\right)^2 + \left(\frac{a}{3b}\right)^3}\right\}^{1/3} \\ + \left\{\frac{c}{2b} - \sqrt{\left(\frac{c}{2b}\right)^2 + \left(\frac{a}{3b}\right)^3}\right\}^{1/3} \end{array}\right],$$

and

$$\begin{aligned} a &= \tilde{t}(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})^{4/3},\ b = (\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})^2, \\ c &= (\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\gamma}) + \lambda, \end{aligned}$$

for some $p \times 1$ vectors $\boldsymbol{\gamma} \neq \mathbf{0}$, $\boldsymbol{\beta} \neq \mathbf{0}$, and constants $\tilde{t} > 0$, λ, which satisfy: (i) $\int_S m^*(\mathbf{x})d\mathbf{x} = 1$, and (ii) minimize (2.7). Let $t = \tilde{t}^3 > 0$, then $d^3 = -\frac{a^3}{b}d + \frac{a^3 c}{b}$ gives

$$d^3 + t(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})^2 d = t(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})^2\left[(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})(\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\gamma}) + \lambda\right].$$

The final evaluation for d and $m^*(\mathbf{x})$ completes the proof. □

Acknowledgements

The research of both authors is supported by the Natural Sciences and Engineering Research Council of Canada.

References

Dette, H. and Wong, W.-K. (1996). Robust optimal extrapolation designs. *Biometrika*, 83, 667-680.

Draper, N. R. and Herzberg, A. (1973). Some designs for extrapolation outside a sphere. *Journal of the Royal Statistical Society Series B,* 35, 268-276.

Dunham, W. (1990). Cardano and the Solution of the Cubic. Ch. 6 (pp. 133-154) in *Journey through Genius: The Great Theorems of Mathematics.* Wiley, New York.

Fang, Z. and Wiens, D. P. (1999). Robust extrapolation designs and weights for biased regression models with heteroscedastic errors. *The Canadian Journal of Statistics,* 27, 751-770.

Ford, I., Kitsos, C. P. and Titterington, D. M. (1989). Recent advances in nonlinear experimental design. *Technometrics,* 31, 49-60.

Heo, G., Schmuland, B., and Wiens, D.P. (2001). Restricted Minimax Robust Designs for Misspecified Regression Models. *The Canadian Journal of Statistics,* 29, 117-128.

Hoel, P. G. and Levine, A. (1964). Optimal Spacing and Weighing in Polynomial Prediction. *Annals of Mathematical Statistics*, 35, 1553–1560.

Hosmer, D. W. Jr. and Lemeshow, S. (1998). *Applied Survival Analysis.* Wiley, New York.

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.

Nelson, W. (1990). *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses.* Wiley, New York.

Seber, G. A. F. and Wild, C. J. (2003). *Nonlinear Regression.* Wiley, New York.

Sinha, S. and Wiens, D. P. (2002). Robust Sequential Designs for Nonlinear Regression. *The Canadian Journal of Statistics*, 30, 601-618.

Spruill, M. C. (1984). Optimal designs for minimax extrapolation. *Journal of Multivariate Analysis*, 15, 52-62.

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

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

CHAPTER III

ROBUST PREDICTION AND EXTRAPOLATION DESIGNS FOR MISSPECIFIED GENERALIZED LINEAR REGRESSION MODELS

Abstract We study minimax robust designs for response prediction and extrapolation in biased generalized linear regression models. We extend previous work of others by considering a nonlinear fitted regression response, by taking a rather general extrapolation space and, most significantly, by dropping all restrictions on the structure of the regressors. Several examples are discussed.

3.1 Introduction

In this chapter, we investigate the construction of robust designs for both prediction and extrapolation of regression responses. In our framework the response fitted by the experimenter is a known function of a linear function of unknown parameters and known regressors. Our designs are robust in that we allow both for imprecision in the specification of the regression response, and for possible heteroscedasticity.

Consider a regression model

$$E(Y|\mathbf{x}) \approx h(\boldsymbol{\theta}^T \mathbf{z}(\mathbf{x})) \tag{3.1}$$

[2]Co-authored with Professor Douglas P. Wiens. To appear in *Journal of Statistical Planning and Inference.*

for a q-dimensional vector $\mathbf{x}$ belonging to a bounded design space S and for p regressors $\mathbf{z}(\mathbf{x}) = (z_1(\mathbf{x}), z_2(\mathbf{x}), ..., z_p(\mathbf{x}))^T$. The function h is strictly monotonic with a bounded second derivative. We assume that $||\mathbf{z}(\mathbf{x})||$ is bounded on S. As indicated in (3.1), the fitted response is typically acknowledged to be only an approximation. The least squares estimates $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ and $\hat{Y} = h(\mathbf{z}^T(\mathbf{x})\hat{\boldsymbol{\theta}})$ of $E(Y|\mathbf{x})$ are possibly biased if the response is misspecified. In this situation, robust designs can play an important role in choosing optimal design points $\mathbf{x}_1$,..., $\mathbf{x}_n \in S$ so that estimates $\hat{\boldsymbol{\theta}}$ and $\hat{Y}$ remain relatively efficient, with small bias caused by the model misspecification.

The true model may be written

$$E(Y|\mathbf{x}) = h(\boldsymbol{\theta}^T\mathbf{z}(\mathbf{x})) + n^{-1/2}f(\mathbf{x}), \tag{3.2}$$

where the contaminant f is unknown but 'small'. This may be viewed as arising from imprecision in the specification of h, or it can arise from a misspecified linear term and a two-term Taylor expansion: $h(\boldsymbol{\theta}^T\mathbf{z}(\mathbf{x}) + \phi(\mathbf{x})) \approx h(\boldsymbol{\theta}^T\mathbf{z}(\mathbf{x})) + h'(\boldsymbol{\theta}^T\mathbf{z}(\mathbf{x}))\phi(\mathbf{x}) = h(\boldsymbol{\theta}^T\mathbf{z}(\mathbf{x})) + n^{-1/2}f(\mathbf{x})$. The factor $n^{-1/2}$ is necessary for an appropriate asymptotic treatment - see Chapter 2.

The experimenter takes n uncorrelated observations $Y_i = Y(\mathbf{x}_i)$, with $\mathbf{x}_i$ freely chosen from a design space S. One possible goal is prediction, or equivalently the estimation of $E(Y|\mathbf{x})$ throughout the region $T = S$. If instead $T \cap S = \phi$, the goal is extrapolation. In this chapter, we discuss both prediction problems and extrapolation problems. We will as well allow for the possibility that observations on Y, although uncorrelated, are heteroscedastic: $var\{Y(\mathbf{x})\} = \sigma^2 g(\mathbf{x})$ for an unknown function within a certain class. We estimate $\boldsymbol{\theta}$ by nonlinear least squares, possibly weighted with nonnegative weights $w(\mathbf{x})$.

For the prediction case, our loss function is n times the integrated mean squared prediction error ($IMSPE$) of $\hat{Y}(\mathbf{x})$ in estimating $E(Y|\mathbf{x})$, $\mathbf{x} \in S$. For extrapolation, the loss is n times the integrated mean squared extrapolation error ($IMSEE$) of $\hat{Y}(\mathbf{x})$ in estimating $E(Y|\mathbf{x})$, $\mathbf{x} \in T$. Both depend on the design measure $\xi =$

$n^{-1}\Sigma_{i=1}^{n} I(\mathbf{x} = \mathbf{x}_i)$ as well as on w, f and g. Formally,

$$IMSPE(f,g,w,\xi) = n\int_S E\left\{\left[\hat{Y}(\mathbf{x}) - E(Y|\mathbf{x})\right]^2\right\} d\mathbf{x},$$
$$IMSEE(f,g,w,\xi) = n\int_T E\left\{\left[\hat{Y}(\mathbf{x}) - E(Y|\mathbf{x})\right]^2\right\} d\mathbf{x}.$$

There is a sizeable literature concerning regression designs for a possibly misspecified linear response. Such designs for homoscedastic errors have been studied by Box and Draper (1959), Huber (1975) and Wiens (1992). Designs for prediction with as well possible heteroscedasticity were obtained by Wiens (1998). For extrapolation with homoscedastic errors see Draper and Herzberg (1973), Huber (1975), Lawless (1984), Spruill (1984), and Dette and Wong (1996) whose extrapolation designs for polynomial responses are robust against misspecification of the degree of the polynomial. In these studies, the goal was extrapolation to one fixed point on or outside the boundary of the design space. Robust designs for extrapolation with possible heteroscedasticity were obtained by Fang and Wiens (1999), and by Chapter 2 of this thesis.

For nonlinear regression, Atkinson and Haines (1996) and Ford, Titterington and Kitsos (1989) present various static and sequential designs for nonlinear models without the consideration of model uncertainty. Sinha and Wiens (2002) also employ notions of robustness in the construction of sequential designs for the nonlinear model. In addition, Chapter 2 discusses the construction of robust designs for a possibly misspecified nonlinear model and for extrapolation of a regression response to one point outside of the design space. The current work goes beyond that of Chapter 2 in that we deal with both prediction and extrapolation and, in the latter case, we allow the extrapolation space T to have nonzero measure. We go beyond Fang and Wiens (1999) in treating nonlinear models. The major advance, though, is perhaps our treatment of essentially unrestricted regressors $\mathbf{z}(\mathbf{x})$. Explicit designs in almost all problems involving misspecified regressors were hitherto restricted to cases in which $\mathbf{z}(\mathbf{x})$ was

well structured - e.g. straight line regression ($\mathbf{z}(x) = (1, x)^T$) or multiple regression without interactions on a spherical design space ($\mathbf{z}(\mathbf{x}) = \left(1, \mathbf{x}^T\right)^T$, $\|\mathbf{x}\| \leq const.$). The improvements in the current work are made possible by our adaptation of recent results of Shi, Ye and Zhou (2003), henceforth referred to as SYZ.

SYZ investigated the analytical form of minimax designs for prediction problems when the function f was an unknown member of the class

$$\mathcal{F} = \left\{ f \left| \int_S \mathbf{z}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \mathbf{0}, \int_S f^2(\mathbf{x}) d\mathbf{x} \leq \eta^2 < \infty \right. \right\}.$$

In our terminology, they considered the case of approximate linearity - $h = \mathbf{1}$, where $\mathbf{1}(\mathbf{x}) = \mathbf{x}$ - and homoscedasticity - $g = \mathbf{1}$. The orthogonality condition in $\mathcal{F}$ ensures that the parameter $\boldsymbol{\theta}$ is uniquely defined in model (3.1). The second condition assures that overall f is not too large.

The class $\mathcal{F}$ is sufficiently rich that any 'design' with finite maximum loss must have a density, and thus must be approximated to make it implementable. Approximation methods are discussed in Heo, Schmuland and Wiens (2001). These can, for instance, take the form of choosing the design points so as to obtain agreement between the $(i-1)/(n-1)$-quantiles $(i = 1, ..., n)$ of the empirical and theoretical design measures, or between the moments to a sufficiently high order. SYZ show that the minimax design densities are of the form

$$m(\mathbf{x}) = \left[\frac{\mathbf{z}^T(\mathbf{x}) \mathbf{P} \mathbf{z}(\mathbf{x}) + d}{\mathbf{z}^T(\mathbf{x}) \mathbf{Q} \mathbf{z}(\mathbf{x})} \right]^+$$

for almost all $\mathbf{x} \in S$, where $c^+ = \max(c, 0)$, for suitable constant symmetric matrices $\mathbf{P}, \mathbf{Q}$ and a constant d. These constants may then be determined numerically.

In this chapter we extend SYZ so as to obtain robust designs for extrapolation and prediction, assuming that the regression response is as at (3.2) and that the errors may be heteroscedastic. If the function h in (3.2) is not the identity then our designs are only locally optimal. They are however still of substantial practical

interest - see reasons for this as listed in Ford, Torsney and Wu (1989) and restated in Ford, Titterington and Kitsos (1992). One typical reason is that where sequential designs can be carried out in batches, the design for the next batch might be a locally optimal design based on the estimates obtained from the previous batch. Allowing for uncertainty in our best guess at a local parameter, we adopt the approach introduced in Chapter 2 to find 'locally most robust' designs which are minimax with respect to a region containing the initial parameters.

We denote unweighted least squares by $w = \mathbf{1}$, homogeneous variances by $g = \mathbf{1}$ and the linear regression problem by $h = \mathbf{1}$. The following problems will be addressed:

(P1) Ordinary least squares (OLS) estimation with homoscedasticity: determine designs to minimize the maximum value, over f, of $IMSEE(f, \mathbf{1}, \mathbf{1}, \xi)$.

(P2) OLS with heteroscedasticity: determine designs to minimize the maximum value, over f and g, of $IMSPE(f, g, \mathbf{1}, \xi)$.

(P3) OLS with heteroscedasticity: determine designs to minimize the maximum value, over f and g, of $IMSEE(f, g, \mathbf{1}, \xi)$.

(P4) Weighted least squares (WLS) estimation with heteroscedasticity: determine designs and weights to minimize the maximum value, over f and g,

of $IMSPE(f, g, w, \xi)$.

(P5) WLS with heteroscedasticity: determine designs and weights to minimize the maximum value, over f and g, of $IMSEE(f, g, w, \xi)$.

The rest of this chapter is organized as follows. The designs for P1 are provided in Section 3.3. Those for P2 and P3 are given in Section 3.4. The designs and weights which constitute the solutions to problems P4 and P5 are given in Section 3.5. Some mathematical preliminaries are detailed in Section 3.2. We present several

examples in Section 3.6, and conclude with a few remarks in Section 3.7. Derivations are provided in an appendix.

3.2 Preliminaries and Notation

The regression models discussed in this chapter are very similar to those in Chapter 2, except that we consider the prediction case as well and allow the extrapolation space to be any space, of positive Lebesgue measure, outside the design space. For the reader's convenience, we briefly describe this model here.

We assume that the contaminant $f(\cdot)$ is an unknown member of

$$\mathcal{F} = \left\{ f \left| \int_S f^2(\mathbf{x})d\mathbf{x} \leq \eta_S^2 < \infty, \int_T f^2(\mathbf{x})d\mathbf{x} \leq \eta_T^2 < \infty, \int_S \tilde{\mathbf{z}}(\mathbf{x})f(\mathbf{x})d\mathbf{x} = \mathbf{0} \right. \right\}, \tag{3.3}$$

where $\mu = \boldsymbol{\theta}^T\mathbf{z}(\mathbf{x})$, $\tilde{\mathbf{z}}(\mathbf{x}) = \left(\frac{dh}{d\mu}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\mathbf{z}(\mathbf{x})$ and η_S, η_T are positive constants. For prediction problems ($T = S$) the second condition in (3.3) merges into the first. The last condition is required in order that $\boldsymbol{\theta}_0$ can be uniquely defined, and in fact arises through the definition

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

together with

$$f_n(\mathbf{x}) = \sqrt{n}\left[E(Y|\mathbf{x}) - h(\boldsymbol{\theta}_0^T\mathbf{z}(\mathbf{x}))\right].$$

Where possible, we drop the subscript on f.

The observations Y_i, although uncorrelated with mean $h(\boldsymbol{\theta}_0^T\mathbf{z}(\mathbf{x}_i)) + n^{-1/2}f(\mathbf{x}_i)$, are possibly heteroscedastic with

$$var\{Y(\mathbf{x}_i)\} = \sigma^2 g(\mathbf{x}_i), \tag{3.4}$$

for a function g satisfying conditions given in Section 3.4.

For extrapolation problems, the only assumptions made about T are that it is disjoint from S and has nonzero Lebesgue measure. 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 $\{\mathbf{x} \in S \cup T : \mathbf{a}^T \bar{\mathbf{z}}(\mathbf{x}) = 0\}$ has Lebesgue measure zero.

We make use of the following matrices and vectors:

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

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

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

It follows from (A) that $\mathbf{A}_S$ is non-singular and that $\mathbf{B}$ is non-singular as well if, as is assumed below, ξ is absolutely continuous. The least squares estimate of $\boldsymbol{\theta}_0$ is

$$\hat{\boldsymbol{\theta}} = \arg\min \sum_{i=1}^{n} [Y_i - h(\boldsymbol{\theta}^T \mathbf{z}(\mathbf{x}))]^2 w(\mathbf{x}_i).$$

The information matrix is

$$\mathcal{I}(\boldsymbol{\theta}_0) = \lim_{n\to\infty} E(-\frac{1}{n}\ddot{\Phi}(\boldsymbol{\theta}_0)) = \mathbf{B},$$

and the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \sim AN(\mathbf{B}^{-1}\mathbf{b}_{f,S}, \sigma^2\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}).$$

For prediction, the loss function $IMSPE$ splits into terms due to bias, variance and model misspecification:

$$\begin{aligned} IMSPE(f,g,w,\xi) &= n\int_S E\left\{\left[\hat{Y}(\mathbf{x}) - E(Y|\mathbf{x})\right]^2\right\} d\mathbf{x} \\ &= n\int_S E\left\{[h(\hat{\boldsymbol{\theta}}^T\mathbf{z}(\mathbf{x})) - h(\boldsymbol{\theta}_0^T\mathbf{z}(\mathbf{x})) - \frac{1}{\sqrt{n}}f(\mathbf{x})]^2\right\} d\mathbf{x} \\ &= IPB(f,w,\xi) + IPV(g,w,\xi) + \int_S f^2(\mathbf{x})d\mathbf{x}, \end{aligned}$$

where the integrated bias (IPB) and integrated variance (IPV) are

$$\begin{aligned} IPB(f,w,\xi) &= n\int_S \left\{E\left[h(\hat{\boldsymbol{\theta}}^T\mathbf{z}(\mathbf{x})) - h(\boldsymbol{\theta}_0^T\mathbf{z}(\mathbf{x}))\right]\right\}^2 d\mathbf{x} \\ &\quad -2\sqrt{n}\int_S f(\mathbf{x})E\left[h(\hat{\boldsymbol{\theta}}^T\mathbf{z}(\mathbf{x})) - h(\boldsymbol{\theta}_0^T\mathbf{z}(\mathbf{x}))\right] d\mathbf{x} \end{aligned}$$

and

$$IPV(g,w,\xi) = n\int_S VAR(\hat{Y}(\mathbf{x}))d\mathbf{x} = n\int_S VAR(h(\hat{\boldsymbol{\theta}}^T\mathbf{z}(\mathbf{x})))d\mathbf{x}.$$

Asymptotically,

$$\begin{aligned} IPB(f,w,\xi) &= \mathbf{b}_{f,S}^T\mathbf{B}^{-1}\mathbf{A}_S\mathbf{B}^{-1}\mathbf{b}_{f,S}, \\ IPV(g,w,\xi) &= \sigma^2 tr(\mathbf{B}^{-1}\mathbf{A}_S\mathbf{B}^{-1}\mathbf{D}). \end{aligned}$$

(The second term in IPB vanishes asymptotically by virtue of the orthogonality condition in the definition of $\mathcal{F}$.)

For extrapolation, the loss function $IMSEE$ decomposes in a similar fashion:

$$\begin{aligned} IMSEE(f,g,w,\xi) &= n\int_T E\left\{\left[\hat{Y}(\mathbf{x})-E(Y|\mathbf{x})\right]^2\right\}d\mathbf{x} \\ &= IEB(f,w,\xi) + IEV(g,w,\xi) + \int_T f^2(\mathbf{x})d\mathbf{x}, \end{aligned}$$

where, asymptotically,

$$\begin{aligned} IEB(f,w,\xi) &= \mathbf{b}_{f,S}^T\mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1}\mathbf{b}_{f,S} - 2\mathbf{b}_{f,T}\mathbf{B}^{-1}\mathbf{b}_{f,S}, \\ IEV(g,w,\xi) &= \sigma^2 tr\left(\mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1}\mathbf{D}\right). \end{aligned}$$

Let $k(\mathbf{x})$ be the density of ξ, and define $m(\mathbf{x}) = k(\mathbf{x})w(\mathbf{x})$. Without loss of generality, we 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 satisfies

$$\int_S \frac{m(\mathbf{x})}{w(\mathbf{x})}d\mathbf{x} = 1, \tag{3.5}$$

and

$$\begin{aligned} \mathbf{B} &= \int_S \bar{\mathbf{z}}(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})m(\mathbf{x})d\mathbf{x}, \\ \mathbf{b}_{f,S} &= \int_S \bar{\mathbf{z}}(\mathbf{x})f(\mathbf{x})m(\mathbf{x})d\mathbf{x}. \end{aligned}$$

From the definitions of $\mathbf{B}$, $\mathbf{b}_{f,S}$ and $\mathbf{b}_{f,T}$, we notice that $IPB(f,w,\xi)$ and $IEB(f,w,\xi)$ rely on (w,ξ) only through m and $IPV(g,w,\xi)$ and $IEV(g,w,\xi)$ through m and w. Hence, we can optimize over m and w subject to (3.5) rather than over k and w.

Although the IEB may be negative,

$$IEB + \int_T f^2(\mathbf{x})d\mathbf{x} = n\int_T \left\{E\left[h(\hat{\boldsymbol{\theta}}^T\bar{\mathbf{z}}(\mathbf{x})) - h(\boldsymbol{\theta}_0^T\bar{\mathbf{z}}(\mathbf{x})) - n^{-1/2}f(\mathbf{x})\right]\right\}^2 d\mathbf{x} \geq 0.$$

We define $r_{T,S} = \eta_T/\eta_S$, reflecting the relative amounts of model response uncertainty in the extrapolation and design spaces and $\nu = \sigma^2/\eta_S^2$ representing the relative importance of variance versus bias. We remark that for prediction our results depend on the unknown parameters only through ν and $\boldsymbol{\theta}_0$ while for extrapolation they depend on the parameters only through $r_{T,S}$, ν and $\boldsymbol{\theta}_0$. In the special case $h = \mathbf{1}$, the results are independent of $\boldsymbol{\theta}_0$.

We also require the definitions $\mathbf{K} = \int_S \bar{\mathbf{z}}(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})m^2(\mathbf{x})d\mathbf{x}$, $\mathbf{G} = \mathbf{K} - \mathbf{B}\mathbf{A}_S^{-1}\mathbf{B}$, $\mathbf{H}_S = \mathbf{B}^{-1}\mathbf{A}_S\mathbf{B}^{-1}$ and $\mathbf{H}_T = \mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1}$.

In the next three sections, we will exhibit solutions to P1-P5.

3.3 Optimal Extrapolation Designs with Homoscedasticity: Solutions to P1

SYZ provide the form of the minimax density for prediction when $h = \mathbf{1}$. In this section, we extend this result to extrapolation and to a general h.

Denote the largest eigenvalue of a matrix $\mathbf{X}$ by $\lambda_{\max}(\mathbf{X})$. As in Theorem 2.1(a) in Fang and Wiens (1999), the maximum extrapolation bias is

$$\sup_{f\in\mathcal{F}} IEB(f, \mathbf{1}, \xi) = \eta_S^2\left[\left(\sqrt{\lambda_{\max}(\mathbf{G}\mathbf{H}_T)} + r_{T,S}\right)^2 - r_{T,S}^2\right] \geq 0.$$

Therefore, the maximum $IMSEE$ is

$$\begin{aligned}\sup_{f\in\mathcal{F}} IMSEE(f, \mathbf{1}, \mathbf{1}, m) &= \eta_S^2\left[\left(\sqrt{\lambda_{\max}(\mathbf{G}\mathbf{H}_T)} + r_{T,S}\right)^2 + \nu\int_S \bar{\mathbf{z}}^T(\mathbf{x})\mathbf{H}_T\bar{\mathbf{z}}(\mathbf{x})m(\mathbf{x})d\mathbf{x}\right] \\ &= \eta_S^2\left[\left(\sqrt{\lambda_{\max}(\mathbf{G}\mathbf{H}_T)} + r_{T,S}\right)^2 + \nu tr\left(\mathbf{B}^{-1}\mathbf{A}_T\right)\right]. \qquad (3.6)\end{aligned}$$

A minimax design is one for which the density m minimizes (3.6). This is an optimization problem with an objective function involving a generally nonsmooth

function $\lambda_{\max}$. Employing nonsmooth optimization theory (Clarke 1983; see SYZ for a useful review), we obtain the following result.

Theorem 3.1 *The minimax design density for extrapolation, when the variances are homogeneous, is of the form*

$$m(\mathbf{x}) = \left[\frac{\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{P}\bar{\mathbf{z}}(\mathbf{x}) + d}{\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{Q}\bar{\mathbf{z}}(\mathbf{x})}\right]^+ \tag{3.7}$$

for almost all $\mathbf{x} \in S$, *for constant symmetric matrices* $\mathbf{P}, \mathbf{Q}\,(\geq \mathbf{0})$ *and a constant* d. *The constants minimize (3.6) and satisfy* $\int_S m(\mathbf{x})d\mathbf{x} = 1$.

Remarks:

1. As in SYZ, in the examples for linear regression in this chapter, we only consider symmetric densities when the structure of the design and extrapolation spaces make this appropriate.

2. The symmetric - in each component of $\mathbf{x}$ - minimax density has the form exhibited in Theorem 3.1 but with the odd functions of these components vanishing. The proof of this is very similar to the proof in Shi (2002) for linear regression.

Example 3.1. For the regression model

$$Y = \theta_0 + \theta_1 x + \theta_2 x^2 + f(x) + \varepsilon,\ x \in [-a, a]$$

with symmetric extrapolation space $[-r_2, -r_1) \cup (r_1, r_2]$ with $0 < a \leq r_1 < r_2$, it is reasonable to restrict to symmetric designs. According to Theorem 3.1, the symmetric optimal design for this model with homoscedasticity is of the form

$$m(x) = \left(\frac{a_1 + a_2x^2 + a_3x^4}{a_4 + a_5x^2 + a_6x^4}\right)^+, \tag{3.8}$$

where a_4 and a_6 are nonnegative. Some computations for this case are shown in Example 6.1.

Example 3.2. For the linear regression model with two interacting regressors

$$Y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{12} x_1 x_2 + f(x_1, x_2) + \varepsilon,$$

with $S = [-a, a] \times [-a, a]$ and $T = [-r, r] \times [-r, r] \setminus S$ $(r > a)$, the minimax designs for prediction were studied by Adewale (2002) who states that the symmetric, exchangeable minimax density is given by

$$m(x_1, x_2) = \left(\frac{a + b(x_1^2 + x_2^2) + c x_1^2 x_2^2}{a' + b'(x_1^2 + x_2^2) + c' x_1^2 x_2^2} \right)^+ .$$

From Theorem 3.1, the minimax symmetric and exchangeable density for extrapolation is also of this form.

Example 3.3. For the nonlinear regression model

$$Y = e^{\theta_0 + \theta_1 x} + f(x) + \varepsilon, \tag{3.9}$$

for which $h(x) = e^x$, we take $S = [0, 1]$ and $T = (1, r]$. The locally most robust extrapolation design density is given by

$$m(x) = \left(\frac{e^{2\theta_1 x}(a_1 + b_1 x + c_1 x^2) + d}{e^{2\theta_1 x}(a_2 + b_2 x + c_2 x^2)} \right)^+ ,$$

where $a_2 \geq 0$, $c_2 \geq 0$ and a_1, b_1, c_1, a_2, b_2, c_2 and d chosen in order to minimize (3.6) subject to $\int_0^1 m(x)dx = 1$. The dependence of the design on θ_1 is an issue which will be addressed in Example 6.2.

3.4 Optimal Prediction and Extrapolation Designs with Heteroscedasticity for OLS: Solutions to P2 and P3

In this and the next section we construct designs which are robust against heteroscedasticity as well as against departures from the fitted response. The heteroscedasticity is governed by $g(\cdot)$ - recall (3.4) - which is assumed to belong to

$$\mathcal{G} = \{g \mid \int_S g^2(\mathbf{x})d\mathbf{x} \leq \Omega^{-1} = \int_S d\mathbf{x} < \infty\}. \tag{3.10}$$

In (3.10), the equality condition is equivalent to defining

$$\sigma^2 = \sup_g \left[\int_S var^2 \{\varepsilon(\mathbf{x})\} \Omega d\mathbf{x} \right]^{1/2}.$$

As in Theorem 1(c) in Wiens (1998) and Theorem 2.1(c) of Fang and Wiens (1999), for OLS the maximum integrated mean square prediction error and extrapolation error are

$$\begin{aligned} &\sup_{f\in\mathcal{F},g\in\mathcal{G}} IMSPE(f,g,\mathbf{1},m) \qquad (3.11)\\ &= \eta_S^2 \left\{ \lambda_{\max}(\mathbf{K}\mathbf{H}_S) + \nu\Omega^{-1/2} \left[\int_S \{\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{H}_S\bar{\mathbf{z}}(\mathbf{x})m(\mathbf{x})\}^2 d\mathbf{x} \right]^{1/2} \right\}, \end{aligned}$$

and

$$\sup_{f\in\mathcal{F},g\in\mathcal{G}} IMSEE(f,g,\mathbf{1},m) = \eta_S^2 \left\{ \begin{array}{c} (\sqrt{\lambda_{\max}(\mathbf{G}\mathbf{H}_T)}+r_{T,S})^2+ \\ \nu\Omega^{-1/2} \left[\int_S \{\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{H}_T\bar{\mathbf{z}}(\mathbf{x})m(\mathbf{x})\}^2 d\mathbf{x}\right]^{1/2} \end{array} \right\} \qquad (3.12)$$

respectively. Therefore problem P2 requires finding a density $m(\cdot)$ which minimizes (3.11) whereas P3 requires finding a density which minimizes (3.12).

Theorem 3.2 *The minimax design densities for both prediction and extrapolation with OLS estimation, when the variances are possibly heterogeneous, have the form*

$$m(\mathbf{x}) = \left[\frac{\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{P}\bar{\mathbf{z}}(\mathbf{x}) + d}{\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{Q}\bar{\mathbf{z}}(\mathbf{x}) + \{\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{U}\bar{\mathbf{z}}(\mathbf{x})\}^2} \right]^+ \qquad (3.13)$$

for almost all $\mathbf{x} \in S$, *for constant symmetric matrices* $\mathbf{P}, \mathbf{Q}(\geq \mathbf{0}), \mathbf{U}(> \mathbf{0})$ *and a constant* d *such that (1)* $\int_S m(\mathbf{x})d\mathbf{x} = 1$ *and (2) for prediction, (3.11) is minimized, while for extrapolation (3.12) is minimized.*

Example 4.1. For the simple linear regression model

$$Y = \theta_0 + \theta_1 x + f(x) + \varepsilon, \qquad (3.14)$$

with $S = [-1, 1]$, the minimax prediction design was studied by Wiens (1998). It was shown there that the minimax symmetric density is given by

$$m(x) = \left(\frac{a + bx^2}{1 + cx^2 + dx^4} \right)^+, \qquad (3.15)$$

a form which now follows as well from Theorem 3.2. Similarly, for extrapolation, Fang and Wiens (1999) derive the form (3.15). More generally, for OLS in the multiple linear regression model

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

with S being a unit hypersphere centred at the origin and $T = \{\mathbf{x} \,|\, 1 < \|\mathbf{x}\| \leq r\}$, Fang and Wiens (1999) obtained conditions under which the minimax symmetric extrapolation design density would be given by

$$m(\mathbf{x}) = \left(\frac{a + b\|\mathbf{x}\|^2}{c + d\|\mathbf{x}\|^2 + e\|\mathbf{x}\|^4} \right)^+.$$

This form now follows, without conditions, from Theorem 3.2 and Remark 2 in Section 3.3.

Example 4.2. For the nonlinear model (3.9) it follows from Theorem 3.2 that the locally optimal robust design density for both prediction and extrapolation is of the form

$$m(x) = \left(\frac{e^{2\theta_1 x}(a_1 + b_1 x + c_1 x^2) + d}{e^{2\theta_1 x}\left[(a_2 + b_2 x + c_2 x^2) + e^{4\theta_1 x}(a_3 + b_3 x + c_3 x^2)^2\right]} \right)^+, \tag{3.16}$$

where $a_2 \geq 0$, $c_2 \geq 0$, $a_3 > 0$ and $c_3 > 0$. When $\theta_1 = 0$, (3.16) can be reduced to

$$m^*(\mathbf{x}) = \left[\frac{a_1 + a_2 x + a_3 x^2}{1 + a_5 x + a_6 x^2 + a_7 x^3 + a_8 x^4} \right]^+,$$

where a_6 and a_8 are positive. The computation of our designs for this model are detailed in Example 6.2.

3.5 Optimal Prediction and Extrapolation Designs with Heteroscedasticity for WLS: Solutions to P4 and P5

In this section we propose to estimate $\boldsymbol{\theta}$ by weighted least squares, and again consider both prediction and extrapolation problems. For prediction we proceed as in Wiens

(1998) and obtain

$$\sup_{f\in\mathcal{F},g\in\mathcal{G}} IMSPE(f,g,w,m) = \eta_S^2 \left\{ \begin{array}{c} \lambda_{\max}(\mathbf{KH}_S)+ \\ \nu\Omega^{-1/2}\left[\int_S \{w(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{H}_S\bar{\mathbf{z}}(\mathbf{x})m(\mathbf{x})\}^2 d\mathbf{x}\right]^{1/2} \end{array} \right\}. \tag{3.17}$$

The weights minimizing (3.17) for fixed $m(\mathbf{x})$, subject to $\int_S \frac{m(\mathbf{x})}{w(\mathbf{x})}d\mathbf{x} = 1$, are, in terms of

$$\alpha_{S,m} = \int_S [\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{H}_S\bar{\mathbf{z}}(\mathbf{x})m^2(\mathbf{x})]^{2/3}d\mathbf{x},$$

given by

$$w_{S,m}(\mathbf{x}) = \alpha_{S,m}[\left\{\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{H}_S\bar{\mathbf{z}}(\mathbf{x})\right\}^2 m(\mathbf{x})]^{-1/3} I\{m(\mathbf{x}) > 0\}. \tag{3.18}$$

Then

$$\min_w \sup_{f\in\mathcal{F},g\in\mathcal{G}} IMSPE(f,g,w,m) = \eta_S^2 \left\{\lambda_{\max}(\mathbf{KH}_S) + \nu\Omega^{-1/2}\alpha_{S,m}^{3/2}\right\} \tag{3.19}$$

and problem P4 becomes that of finding a density $m^*(\mathbf{x})$ which minimizes (3.19). Then the weights $w_{S,m^*}(\mathbf{x})$ obtained from (3.18) and the design density

$$k_*(\mathbf{x}) = \frac{m^*(\mathbf{x})}{w_{S,m^*}(\mathbf{x})} = \alpha_{S,m^*}^{-1}\left\{\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{H}_S\bar{\mathbf{z}}(\mathbf{x})m^{*2}(\mathbf{x})\right\}^{2/3}$$

are optimal for WLS prediction.

For extrapolation we follow Fang and Wiens (1999) and obtain

$$\sup_{f\in\mathcal{F},g\in\mathcal{G}} IMSEE(f,g,w,m) = \eta_S^2 \left\{ \begin{array}{c} (\sqrt{\lambda_{\max}(\mathbf{GH}_T)}+r_{T,S})^2+ \\ \nu\Omega^{-1/2}\left[\int_S \{w(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{H}_T\bar{\mathbf{z}}(\mathbf{x})m(\mathbf{x})\}^2 d\mathbf{x}\right]^{1/2} \end{array} \right\}.$$

In terms of

$$\alpha_{T,m} = \int_S [\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{H}_T\bar{\mathbf{z}}(\mathbf{x})m^2(\mathbf{x})]^{2/3}d\mathbf{x},$$

the minimizing weights are given by

$$w_{T,m}(\mathbf{x}) = \alpha_{T,m}[\left\{\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{H}_T\bar{\mathbf{z}}(\mathbf{x})\right\}^2 m(\mathbf{x})]^{-1/3} I\{m(\mathbf{x}) > 0\}, \tag{3.20}$$

with

$$\min_w \sup_{f\in\mathcal{F},g\in\mathcal{G}} IMSEE(f,g,m) = \eta_S^2 \left\{(\sqrt{\lambda_{\max}(\mathbf{GH}_T)}+r_{T,S})^2 + \nu\Omega^{-1/2}\alpha_{T,m}^{3/2}\right\}, \tag{3.21}$$

and to solve P5 we seek a density $m^*(\mathbf{x})$ which minimizes (3.21). Then the weights $w_{T,m^*}(\mathbf{x})$ obtained from (3.20) and the design density

$$k_*(\mathbf{x}) = \frac{m^*(\mathbf{x})}{w_{T,m^*}(\mathbf{x})} = \alpha_{T,m^*}^{-1} \left\{ \bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{H}_T \bar{\mathbf{z}}(\mathbf{x}) m^{*2}(\mathbf{x}) \right\}^{2/3} \tag{3.22}$$

are optimal for WLS extrapolation.

The following theorem provides the form of $m^*(\mathbf{x})$ for both prediction and extrapolation.

Theorem 3.3 *The minimax densities $m^*(\mathbf{x})$ for both prediction and extrapolation with WLS estimation, when the variances are possibly heterogeneous, are of the form*

$$m^*(\mathbf{x}) = \left[\frac{c(\mathbf{x}) - k(\mathbf{x})}{b(\mathbf{x})} \right]^+, \tag{3.23}$$

where, for constant symmetric matrices $\mathbf{P}, \mathbf{Q} (\geq \mathbf{0}), \mathbf{U} (> \mathbf{0})$ and a constant d we have $b(\mathbf{x}) = \bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{Q} \bar{\mathbf{z}}(\mathbf{x})$, $c(\mathbf{x}) = \bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{P} \bar{\mathbf{z}}(\mathbf{x}) + d$ and

$$k^3 + \frac{a^3}{b} k - \frac{a^3 c}{b} = 0,$$

with $a(\mathbf{x}) = \left\{ \bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{U} \bar{\mathbf{z}}(\mathbf{x}) \right\}^{2/3}$. Explicitly,

$$k = a \left[\left\{ \frac{c}{2b} + \sqrt{\left(\frac{c}{2b}\right)^2 + \left(\frac{a}{3b}\right)^3} \right\}^{1/3} + \left\{ \frac{c}{2b} - \sqrt{\left(\frac{c}{2b}\right)^2 + \left(\frac{a}{3b}\right)^3} \right\}^{1/3} \right]. \tag{3.24}$$

The constants satisfy (1) $\int_S m(\mathbf{x}) d\mathbf{x} = 1$ and (2) minimize (3.19) for prediction, (3.21) for extrapolation.

Example 5.1. For the simple linear regression model (3.14) with $S = [-1, 1]$ and $T = \{x \,|\, 1 < |x| \leq r\}$ we obtain (3.23) with

$$\begin{aligned} c(x) &= a_1 + a_2 x^2, \\ b(x) &= a_3 + a_4 x^2, \\ a(x) &= (a_5 + a_6 x^2)^{2/3} \end{aligned}$$

where $a_3 \geq 0, a_4 \geq 0$, $a_3^2 + a_4^2 > 0$, $a_5 > 0$ and $a_6 > 0$ are determined as in the statement of Theorem 3.3. The minimax weights are obtained from (3.18) and (3.20) with

$$\begin{aligned}
\tilde{\mathbf{z}}^T(\mathbf{x})\mathbf{H}_S\tilde{\mathbf{z}}(\mathbf{x}) &= 2 + \frac{2}{3}x^2\left(\int_{-1}^{1} x^2 m(x)dx\right)^{-2}, \\
\tilde{\mathbf{z}}^T(\mathbf{x})\mathbf{H}_T\tilde{\mathbf{z}}(\mathbf{x}) &= 2(r-1) + \frac{2}{3}(r^3-1)x^2\left(\int_{-1}^{1} x^2 m(x)dx\right)^{-2}.
\end{aligned}$$

Example 5.2. For the nonlinear model (3.9) with $S = [0,1]$ and $T = (1,r]$ $(r > 1)$ we attain (3.23) with

$$\begin{aligned}
c(x) &= a_0 + e^{2\theta_1 x}\left(a_1 + a_2 x + a_3 x^2\right), \\
b(x) &= e^{2\theta_1 x}\left(a_4 + a_5 x + a_6 x^2\right), \\
a(x) &= \left[e^{2\theta_1 x}\left(a_7 + a_8 x + a_9 x^2\right)\right]^{2/3}
\end{aligned}$$

where $a_4 \geq 0, a_6 \geq 0$, $a_4^2 + a_6^2 > 0$, $a_7 > 0$ and $a_9 > 0$ are determined as in Theorem 3.3. Note that the term $e^{2\theta_0}$ has been absorbed into $a_1, ..., a_9$. The minimax weights are derived from (3.18) and (3.20) with

$$\begin{aligned}
\tilde{\mathbf{z}}^T(\mathbf{x})\mathbf{H}_S\tilde{\mathbf{z}}(\mathbf{x}) &= \left(u_1u_3 - u_2^2\right)^{-2} e^{2\theta_1 x}\phi(x; s_1, s_2, s_3, u_1, u_2, u_3), \\
\tilde{\mathbf{z}}^T(\mathbf{x})\mathbf{H}_T\tilde{\mathbf{z}}(\mathbf{x}) &= \left(u_1u_3 - u_2^2\right)^{-2} e^{2\theta_1 x}\phi(x; t_1, t_2, t_3, u_1, u_2, u_3),
\end{aligned}$$

where

$$\begin{aligned}
&\phi(x; s_1, s_2, s_3, u_1, u_2, u_3) \\
&= \left(u_3^2 s_1 - 2u_2u_3s_2 + u_2^2 s_3\right) + 2\left(u_1u_3s_2 - u_1u_2s_3 - u_2u_3s_1 + u_2^2 s_2\right)x \\
&\quad + \left(u_1^2 s_3 - 2u_1u_2s_2 + u_2^2 s_1\right)x^2, \\
s_1 &= \frac{e^{2\theta_1} - 1}{2\theta_1}, \quad s_2 = \frac{e^{2\theta_1} - s_1}{2\theta_1}, \quad s_3 = \frac{e^{2\theta_1} - 2s_2}{2\theta_1}, \\
t_1 &= \frac{e^{2\theta_1 r} - e^{2\theta_1}}{2\theta_1}, \quad t_2 = \frac{re^{2\theta_1 r} - e^{2\theta_1} - t_1}{2\theta_1}, \quad t_3 = \frac{r^2e^{2\theta_1 r} - e^{2\theta_1} - 2t_2}{2\theta_1}, \\
u_1 &= \int_0^1 e^{2\theta_1 x} m^*(x)dx, \quad u_2 = \int_0^1 xe^{2\theta_1 x} m^*(x)dx, \quad u_3 = \int_0^1 x^2 e^{2\theta_1 x} m^*(x)dx.
\end{aligned}$$

Table 3.1. Coefficient values for the density (3.8) with $a_4 = 1$ for Example 6.1.

r	ν	$a1$	a_2	a_3	a_5	$a6$	Loss
1.5	0.1	0.377	78.22	273.86	278.32	149.87	6.79
	0.25	1.16	−6.45	289.98	217.64	27.96	9.50
	1	1.16	−37.05	70.72	−2.60	18.62	19.70
	5	0.804	−13.31	16.85	−3.98	4.18	61.06
	10	0.923	−20.53	24.57	−3.99	3.98	108.49
	100	1.65	−124.34	132.51	−3.50	3.06	853.72
5	0.1	0.524	−1.04	1.55	−1.10	1.59	1302.54
	0.25	0.654	−4.17	5.71	−5.02	6.30	2544.55
	1	0.822	−6.52	8.38	−4.62	5.33	8508.54
	5	1.21	−16.62	19.64	−4.28	4.59	34294.25
	10	1.47	−28.05	32.05	−4.19	4.39	63387.51
	100	2.98	−214.47	226.23	−4.02	4.04	534222.2

3.6 Computations and Examples

Example 6.1. Recall Example 3.1 and (3.8). We take $r_1 = 1$ and denote r_2 by r. If either of a_4 or a_6 is nonzero, we may take it to be unity. We take $a_4 = 1$ and $r_{T,S} = 1$. Some numerical values of the constants are shown in Table 3.1. Figure 14 gives plots of the minimax extrapolation densities for varying r and ν. The designs can be roughly described as replacing those points with masses at -1, 1 and 0 in the variance minimizing designs by more or less uniformly distributed clusters in neighbourhoods of these points. Decreasing ν results in more uniform designs. A larger r (wider extrapolation region) results in more uniformity as well, especially in the central region.

Example 6.2. Recall Example 4.2 and the nonlinear model (3.9) with possible heteroscedasticity. The locally optimal design density for prediction is given by (3.16). See Table 3.2 for the numerical values of the constants in (3.16) and Figure 15(a) for plots. Here we have taken $a_2 = 1$ and $\theta_1 = 1$.

These designs are only locally optimal since they depend on the value of θ_1. To deal with this, we obtain 'locally most robust' designs as in Chapter 2. For this, we

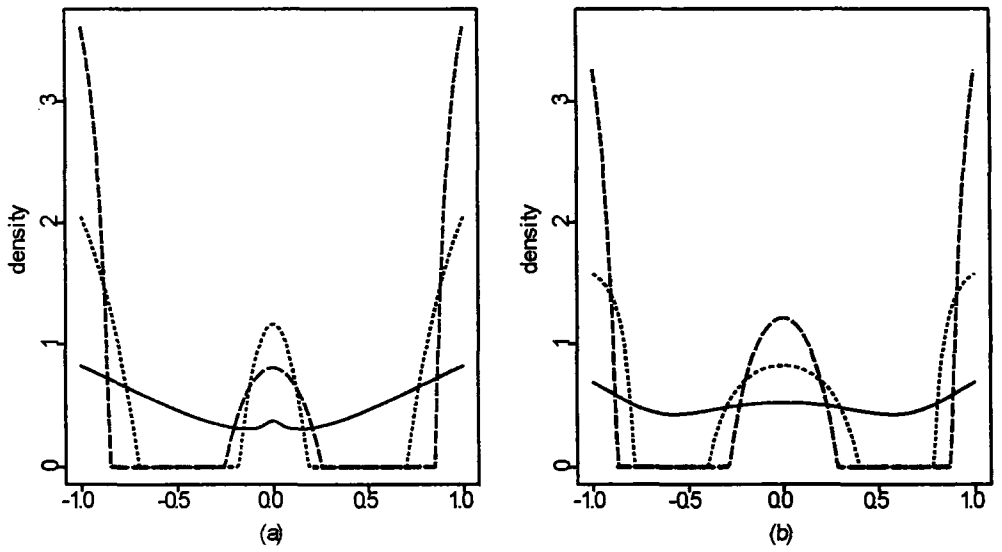

Figure 14: Minimax extrapolation densities $m(x) = \left(\frac{a_1+a_2x^2+a_3x^4}{1+a_5x^2+a_6x^4}\right)^+$ in Example 6.1. (a) $r = 1.5$; (b) $r = 5$. Each plot uses three values of ν: $\nu = 0.1$ (solid line), $\nu = 1$ (dotted line), $\nu = 5$ (broken line).

Table 3.2. Coefficient values for the density (3.16) in Example 6.2 with $a_2 = 1$ and $\theta_1 = 1$.

ν	a_1	b_1	c_1	d	b_2	c_2	a_3	b_3	c_3
0.5	−513.21	495.36	6.99	782.40	3.44	46.96	16.07	−39.74	25.31
1	932.34	−2428.57	1892.85	−322.54	12.23	106.62	25.44	−60.19	38.81
5	2044.47	−4883.06	3303.49	−816.61	7.95	25.45	35.90	−81.08	50.00

take a further maximum of the loss as θ_1 varies over some interval I, and determine the coefficients of $m(x)$ so as to minimize this maximum loss. For $I = [0, 2]$, the locally most robust designs are detailed in Table 3.3 for varying ν. In each case, we found that the least favourable θ_1 within I, say θ_1^{LF}, is 2. See Figure 15(b) for plots. Although, as pointed out in Silvey (1980), local designs tailored for optimality at a least favourable parameter value are sometimes inefficient at distant points, it has been our experience that the designs constructed here do not exhibit a strong dependence on θ_1.

Example 6.3. Recall Example 5.2 and model (3.9) with $S = [0, 1]$. The locally

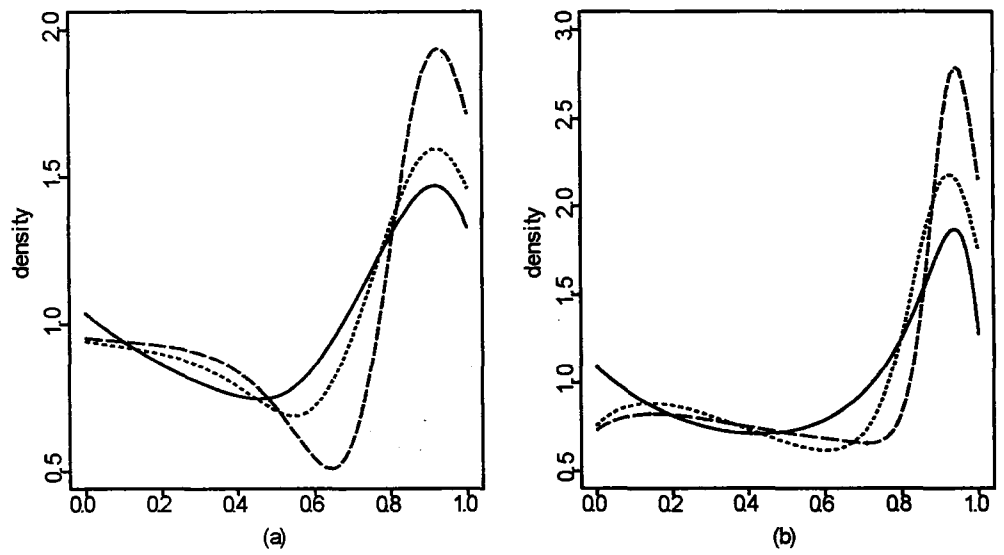

Figure 15: Optimal minimax design densities

$$m(x) = \left(\frac{e^{2\theta_1^* x}(a_1 + b_1 x + c_1 x^2) + d}{e^{2\theta_1^* x}\left[(1 + b_2 x + c_2 x^2) + e^{4\theta_1^* x}(a_3 + b_3 x + c_3 x^2)^2\right]} \right)^+$$

in Example 6.2. (a) Locally optimal design densities for $\theta_1^* = 1$; (b) locally most robust design densities for $\theta_1^* = \theta_1^{LF}$ in $[0, 2]$. Each plot uses three values of ν: $\nu = 0.5$ (solid line), $\nu = 1$ (dotted line), $\nu = 5$ (broken line).

optimal product of density and weights for the prediction problem is given by (3.23). We take $a_4 = 1$. For $\theta_1 = 1$, the numerical values of the constants in (3.23) are given in Table 3.4. See Figure 16 (a) and (b) for plots of the locally optimal design densities and the corresponding optimal regression weights.

For $I = [0.5, 1.5]$, the locally most robust products of density and weights are provided in Table 3.5 for varying ν. In each case, we found that the least favourable

Table 3.3. Coefficient values for the locally most robust density (3.16).

ν	a_1	b_1	c_1	d	b_2	c_2	a_3	b_3	c_3
0.5	24022.10	−28233.18	4874.55	3336.73	1.78	20.0	158.10	−291.2	136.26
1	92029.01	−217292.1	138815.1	−69222.6	1.34	8.32	173.63	−362.9	200.60
5	122521.4	−276528.6	161973.8	−85323.7	9.06	105.1	225.56	−484.0	265.67

Table 3.4. Coefficient values for the locally optimal product (3.23) of density and weights in Example 6.3.

ν	a_0	a_1	a_2	a_3	a_5	a_6	a_7	a_8	a_9
0.5	163.19	582.58	−369.48	211.80	6.18	26.47	18978.42	165.78	1122.10
1	134.09	2710.50	−2352.19	1737.83	11.92	84.98	138645.0	4451.39	65374.78
2	−3269.9	10678.5	−12184.9	9087.9	13.45	81.59	581177.6	20860.9	701666.1

Table 3.5. Coefficient values for the locally most robust product (3.23) of density and weights in Example 6.3.

ν	a_0	a_1	a_2	a_3	a_5	a_6	a_7	a_8	a_9
0.5	454.81	1703.6	−838.72	822.42	13.29	79.77	92972.48	2543.26	13839.7
1	439.79	3903.5	−2821.33	2281.76	0.198	81.49	255487.7	299.60	30479.6
2	−4938.2	7237.6	−6652.43	3692.79	−3.95	74.05	95311.10	3999.84	56657.3

θ_1 within I is 0.5. See Figure 16 (c) and (d) for plots.

3.7 Concluding Remarks

We have derived minimax prediction and extrapolation designs for misspecified generalized linear response models in the following three cases: (i) using OLS estimation under homoscedasticity, (ii) using OLS estimation under possible heteroscedasticity and (iii) using WLS estimation under possible heteroscedasticity. For each case with OLS, we conclude that the minimax extrapolation design density has the same form as that for the corresponding prediction problem. For case (iii), the product of the design density and weights function has the same form for both prediction and extrapolation. These analytic forms are completely general, but contain several constants to be determined numerically.

Chapter 2 have derived minimax designs for extrapolation to a single point. Although the current work has assumed an extrapolation space with positive Lebesgue measure, the designs for one point extrapolation can be derived informally as limits of those in this chapter, as follows.

(1) The minimax one-point extrapolation design density for (i) above was shown

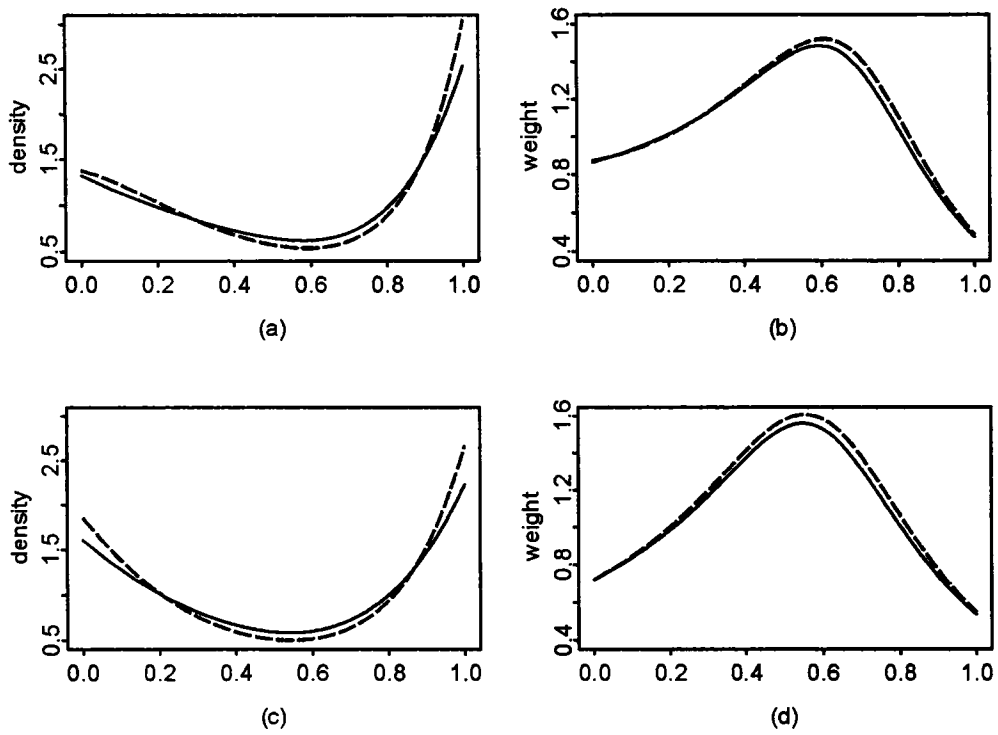

Figure 16: Locally optimal and most robust design densities and corresponding weights for WLS in Example 6.3: (a) locally optimal design densities, (b) optimal weights corresponding to (a); (c) locally most robust design densities, and (d) optimal weights corresponding to (c). Each plot uses two values of ν: $\nu = 0.5$ (solid line) and $\nu = 2$ (broken line).

in Chapter 2 to have the form

$$m(\mathbf{x}) = \left[\frac{\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta}\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\gamma} + d}{\{\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta}\}^2}\right]^+.$$

This is the special case of form (3.7) with $\mathbf{P} = \frac{\boldsymbol{\beta}\boldsymbol{\gamma}^T + \boldsymbol{\gamma}\boldsymbol{\beta}^T}{2}$ and $\mathbf{Q} = \boldsymbol{\beta}\boldsymbol{\beta}^T$.

(2) The minimax one-point extrapolation design density for (ii) above was shown to have the form

$$m(\mathbf{x}) = \left[\frac{\{\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta}\}\{\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\gamma}\} + d}{\{\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta}\}^2 + b\{\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta}\}^4}\right]^+.$$

This is the special case of (3.13) with $\mathbf{P} = \frac{\boldsymbol{\beta}\boldsymbol{\gamma}^T + \boldsymbol{\gamma}\boldsymbol{\beta}^T}{2}$, $\mathbf{Q} = \boldsymbol{\beta}\boldsymbol{\beta}^T$ and $\mathbf{U} = \sqrt{b}\mathbf{Q}$.

(3) The minimax product of design densities and weights for (iii) was shown to

have the form

$$m(\mathbf{x}) = \left[\frac{(\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})(\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\gamma}) + d - c}{(\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})^2}\right]^+,$$

where c satisfies the cubic equation

$$c^3 + b(\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})^2 c = b(\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})^2 \left[(\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta})(\bar{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\gamma}) + d\right]$$

and $b > 0$. This is the special case of (3.23) with $\mathbf{P} = \frac{\boldsymbol{\beta}\boldsymbol{\gamma}^T + \boldsymbol{\gamma}\boldsymbol{\beta}^T}{2}$, $\mathbf{Q} = \boldsymbol{\beta}\boldsymbol{\beta}^T$ and $\mathbf{U} = \sqrt{b}\mathbf{Q}$.

Appendix: Derivations

Proof of Theorem 3.1. In what follows a prime $(\cdot)'$ denotes the Fréchet derivative of $(\cdot)$, $\partial(\cdot)$ is the Clarke generalized gradient of $(\cdot)$ and $N_{m(\mathbf{x})\geq 0}(m)$ is the normal cone of $\{m : m(\mathbf{x}) \geq 0\}$, i.e.

$$N_{m(\mathbf{x})\geq 0}(m) = \left\{\omega(\mathbf{x}) : \int_S \omega(\mathbf{x})(m_1 - m)d\mathbf{x} \leq 0, \text{ for any } m_1(\mathbf{x}) \geq 0\right\}.$$

(See SYZ for basic definitions.)

Define $L_1 = \sup_{f\in\mathcal{F}} IMSEE(f, \mathbf{1}, \mathbf{1}, m)$, given at (3.6), and let $m(\mathbf{x})$ be a density minimizing L_1. (The existence of such a density is established as in Ye and Zhou, 2005.) Then by the non-smooth Lagrange multiplier rule (Clarke 1983, Th. 6.1.1), there exist real numbers $\lambda \geq 0$ and δ, not both zero, such that

$$0 \in \lambda\partial L_1(m) + \delta\left(\int_S m(\mathbf{x})d\mathbf{x} - 1\right)' + N_{m(\mathbf{x})\geq 0}(m). \tag{3.A.1}$$

Note that $\mathbf{G} = \int_S \left[\left\{m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}_S^{-1}\right\}\bar{\mathbf{z}}(\mathbf{x})\right]\left[\left\{m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}_S^{-1}\right\}\bar{\mathbf{z}}(\mathbf{x})\right]^T d\mathbf{x} \geq \mathbf{0}$. We temporarily assume that $\mathbf{G}$ is positive definite. Then as at Theorem 2 of SYZ, the generalized gradient of $\lambda_{\max}(\mathbf{G}\mathbf{H}_T)$ at m is

$$\partial\lambda_{\max}(\mathbf{G}\mathbf{H}_T) = co\left\{\left(\frac{\mathbf{w}^T\mathbf{H}_T\mathbf{w}}{\mathbf{w}^T\mathbf{G}^{-1}\mathbf{w}}\right)' : \mathbf{w} \in M(m)\right\},$$

where

$$M(m) = \left\{\mathbf{w} : \frac{\mathbf{w}^T\mathbf{H}_T\mathbf{w}}{\mathbf{w}^T\mathbf{G}^{-1}\mathbf{w}} = \max_{\|\mathbf{w}\|=1}\frac{\mathbf{w}^T\mathbf{H}_T\mathbf{w}}{\mathbf{w}^T\mathbf{G}^{-1}\mathbf{w}}\right\}$$

and

$$coA = \left\{\sum\lambda_i a_i : \lambda_i \geq 0, \sum\lambda_i = 1, a_i \in A\right\}$$

is the convex hull of set A. From the Chain Rule (Clarke 1983, Th. 2.3.10),

$$\begin{aligned}\partial\left(\sqrt{\lambda_{\max}(\mathbf{G}\mathbf{H}_T)}+r_{T,S}\right)^2 &= \left(1 + \frac{r_{T,S}}{\sqrt{\lambda_{\max}(\mathbf{G}\mathbf{H}_T)}}\right)\partial\lambda_{\max}(\mathbf{G}\mathbf{H}_T) \qquad (3.A.2)\\ &= \left(1 + \frac{r_{T,S}}{\sqrt{\lambda_{\max}(\mathbf{G}\mathbf{H}_T)}}\right) co\left\{\left(\frac{\mathbf{w}^T\mathbf{H}_T\mathbf{w}}{\mathbf{w}^T\mathbf{G}^{-1}\mathbf{w}}\right)' : \mathbf{w} \in M(m)\right\}.\end{aligned}$$

We require the following Fréchet derivatives, which can be calculated as in SYZ:

$$\left(tr\left[\mathbf{B}^{-1}\mathbf{A}_T\right]\right)' = -\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1}\bar{\mathbf{z}}(\mathbf{x}), \tag{3.A.3}$$

$$\left(\frac{\mathbf{w}^T\mathbf{H}_T\mathbf{w}}{\mathbf{w}^T\mathbf{G}^{-1}\mathbf{w}}\right)'_m = \bar{\mathbf{z}}^T(\mathbf{x})\tilde{\mathbf{M}}_{\mathbf{w}}\bar{\mathbf{z}}(\mathbf{x}) + \left\{\bar{\mathbf{b}}_{\mathbf{w}}^T\bar{\mathbf{z}}(\mathbf{x})\right\}^2 m(\mathbf{x}). \tag{3.A.4}$$

In (3.A.4), $\mathbf{w}$ is any vector in $\mathbb{R}$ and $\tilde{\mathbf{M}}_{\mathbf{w}}$ is a $p \times p$ symmetric matrix, $\bar{\mathbf{b}}_{\mathbf{w}}$ a $p \times 1$ vector whose specific values are not important to us.

By (3.A.1), (3.A.2), (3.A.3) and (3.A.4), we have that

$$0 \in \lambda\eta_S^2 \left[\begin{array}{c} \left(1 + \frac{r_{T,S}}{\sqrt{\lambda_{\max}(\mathbf{GH}_T)}}\right) co\left\{\bar{\mathbf{z}}^T(\mathbf{x})\tilde{\mathbf{M}}_{\mathbf{w}}\bar{\mathbf{z}}(\mathbf{x}) + \left(\bar{\mathbf{b}}_{\mathbf{w}}^T\bar{\mathbf{z}}(\mathbf{x})\right)^2 m(\mathbf{x}) : \mathbf{w} \in M(m)\right\} \\ -\nu\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1}\bar{\mathbf{z}}(\mathbf{x}) \end{array} \right] + \delta + N_{m(\mathbf{x})\geq 0}(m). \tag{3.A.5}$$

Let $\mathbf{M}_{\mathbf{w}} = \left(1 + \frac{r_{T,S}}{\sqrt{\lambda_{\max}(\mathbf{GH}_T)}}\right)\tilde{\mathbf{M}}_{\mathbf{w}}$, $\mathbf{b}_{\mathbf{w}} = \left(1 + \frac{r_{T,S}}{\sqrt{\lambda_{\max}(\mathbf{GH}_T)}}\right)^{1/2}\bar{\mathbf{b}}_{\mathbf{w}}$ and note that $\left(\int_S m(\mathbf{x})d\mathbf{x} - 1\right)'_m = 1$. Then (3.A.5) becomes

$$0 \in \lambda\eta_S^2\left[co\left\{\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{M}_{\mathbf{w}}\bar{\mathbf{z}}(\mathbf{x}) + \left(\mathbf{b}_{\mathbf{w}}^T\bar{\mathbf{z}}(\mathbf{x})\right)^2 m(\mathbf{x}) : \mathbf{w} \in M(m)\right\} - \nu\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1}\bar{\mathbf{z}}(\mathbf{x})\right] + \delta + N_{m(\mathbf{x})\geq 0}(m).$$

It can be shown, as in the proof of Theorem 1 of SYZ, that $\lambda \neq 0$.

By the definition of convex hull, there exists a positive integer N, nonnegative scalars $\lambda_1, ..., \lambda_N$ with $\lambda_1 + ... + \lambda_N = 1$, $\mathbf{w}_i \in M(m) \subset \mathbf{R}^p$ and $\varepsilon \in N_{m(\mathbf{x})\geq 0}(m)$ such that

$$\begin{aligned} 0 &= \lambda\eta_S^2 \left[\begin{array}{c} \sum_{i=1}^N \lambda_i\{\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{M}_{\mathbf{w}_i}\bar{\mathbf{z}}(\mathbf{x}) + \left(\mathbf{b}_{\mathbf{w}_i}^T\bar{\mathbf{z}}(\mathbf{x})\right)^2 m(\mathbf{x})\} \\ -\nu\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1}\bar{\mathbf{z}}(\mathbf{x}) \end{array} \right] + \delta + \varepsilon \\ &= \lambda\eta_S^2 \sum_{i=1}^N \lambda_i \left(\mathbf{b}_{\mathbf{w}_i}^T\bar{\mathbf{z}}(\mathbf{x})\right)^2 m(\mathbf{x}) - \\ &\quad \lambda\eta_S^2\bar{\mathbf{z}}^T(\mathbf{x})\left[\nu\mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1} - \sum_{i=1}^N \lambda_i\mathbf{M}_{\mathbf{w}_i}\right]\bar{\mathbf{z}}(\mathbf{x}) + \delta + \varepsilon. \end{aligned}$$

Consequently, there exists a constant symmetric matrix

$$\mathbf{P} = \lambda\eta_S^2 \left[\nu \mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1} - \sum_{i=1}^{N} \lambda_i \mathbf{M}_{\mathbf{w}_i} \right],$$

a constant positive semi-definite matrix $\mathbf{Q} = \lambda\eta_S^2 \sum_{i=1}^{N} \lambda_i \mathbf{b}_{\mathbf{w}_i}\mathbf{b}_{\mathbf{w}_i}^T$ and a constant $d = -\delta$ such that

$$0 = \bar{\mathbf{z}}^T(\mathbf{x})\mathbf{Q}\bar{\mathbf{z}}(\mathbf{x})m(\mathbf{x}) - \bar{\mathbf{z}}^T(\mathbf{x})\mathbf{P}\bar{\mathbf{z}}(\mathbf{x}) - d + \varepsilon. \tag{3.A.6}$$

From Proposition 3 of SYZ we see that $\varepsilon = 0$ almost everywhere on $\{\mathbf{x} \in S : m(\mathbf{x}) > 0\}$ and hence

$$\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{Q}\bar{\mathbf{z}}(\mathbf{x})m(\mathbf{x}) - \bar{\mathbf{z}}^T(\mathbf{x})\mathbf{P}\bar{\mathbf{z}}(\mathbf{x}) - d = 0$$

for all $\mathbf{x}$ such that $m(\mathbf{x}) > 0$. Since $\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{Q}\bar{\mathbf{z}}(\mathbf{x}) = \sum_{i=1}^{N} \lambda_i \left\{ \mathbf{b}_{\mathbf{w}_i}^T \bar{\mathbf{z}}(\mathbf{x}) \right\}^2 > 0$ for almost all $\mathbf{x} \in S$ we obtain

$$m(\mathbf{x}) = \frac{\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{P}\bar{\mathbf{z}}(\mathbf{x}) + d}{\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{Q}\bar{\mathbf{z}}(\mathbf{x})}$$

for almost all $\mathbf{x} \in S$ such that $m(\mathbf{x}) > 0$. For those $\mathbf{x} \in S$ such that $m(\mathbf{x}) = 0$ we apply Proposition 3 of SYZ again to infer that $\varepsilon \leq 0$ a.e. and hence by (3.A.6), $\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{P}\bar{\mathbf{z}}(\mathbf{x}) + d \leq 0$. Consequently,

$$\frac{\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{P}\bar{\mathbf{z}}(\mathbf{x}) + d}{\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{Q}\bar{\mathbf{z}}(\mathbf{x})} \leq 0$$

for almost all $\mathbf{x} \in S$ such that $m(\mathbf{x}) = 0$, and (3.7) follows in the case that $\mathbf{G}$ is positive definite. This unnecessary assumption may now be dropped by arguing in the same manner as in the proof of Theorem 1 in SYZ. □

Proof of Theorem 3.2. We give the proof only for extrapolation, that for prediction being similar but simpler. Define $l_m(\mathbf{x}) = \bar{\mathbf{z}}^T(\mathbf{x})\mathbf{H}_T\bar{\mathbf{z}}(\mathbf{x})$. Then from (3.12) we seek a density $m(\cdot)$ minimizing

$$L_2 \stackrel{def.}{=} \sup_{f\in\mathcal{F}, g\in\mathcal{G}} IMSEE(f, g, \mathbf{1}, m) = \eta_S^2 \left\{ \begin{array}{c} \left(\sqrt{\lambda_{\max}(\mathbf{G}\mathbf{H}_T)} + r_{T,S}\right)^2 + \\ \nu\Omega^{-1/2} \left[\int_S \{l_m(\mathbf{x})m(\mathbf{x})\}^2 d\mathbf{x}\right]^{1/2} \end{array} \right\}.$$

We again initially assume $\mathbf{G} > 0$. As in the preceding proof there exist real numbers $\lambda \geq 0$ and δ, not both zero, such that

$$0 \in \lambda \partial L_2(m) + \delta \left(\int_S m(\mathbf{x}) d\mathbf{x} - 1 \right)' + N_{m(\mathbf{x}) \geq 0}(m), \tag{3.A.7}$$

where the last two terms $\left(\int_S m(\mathbf{x}) d\mathbf{x} - 1 \right)'$ and $N_{m(\mathbf{x}) \geq 0}(m)$ are the same as those in the proof of Theorem 3.1. Note that L_1 and L_2 differ only in their variance terms. Using

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

$$\left(\int_S \{ \{ l_m(\mathbf{x}) m(\mathbf{x}) \}^2 d\mathbf{x} \right)^{-1/2} \left\{ \begin{array}{c} l_m^2(\mathbf{x}) m(\mathbf{x}) - \\ 2 \bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{H}_T \left(\int_S l_m(\mathbf{x}) m^2(\mathbf{x}) \bar{\mathbf{z}}(\mathbf{x}) \bar{\mathbf{z}}^T(\mathbf{x}) d\mathbf{x} \right) \mathbf{B}^{-1} \bar{\mathbf{z}}(\mathbf{x}) \end{array} \right\}$$

in the evaluation of (3.A.7) we obtain

$$\begin{aligned} 0 \in \lambda \eta_S^2 & \left\{ \begin{array}{c} co \left\{ \bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{M}_{\mathbf{w}} \bar{\mathbf{z}}(\mathbf{x}) + \left(\mathbf{b}_{\mathbf{w}}^T \bar{\mathbf{z}}(\mathbf{x}) \right)^2 m(\mathbf{x}) : \mathbf{w} \in M(m) \right\} + \\ \nu \Omega^{-1/2} \left(\int_S \{ l_m(\mathbf{x}) m(\mathbf{x}) \}^2 d\mathbf{x} \right)^{-1/2} \\ \left\{ \begin{array}{c} l_m^2(\mathbf{x}) m(\mathbf{x}) - \\ 2 \bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{H}_T \left(\int_S l_m(\mathbf{x}) m^2(\mathbf{x}) \bar{\mathbf{z}}(\mathbf{x}) \bar{\mathbf{z}}^T(\mathbf{x}) d\mathbf{x} \right) \mathbf{B}^{-1} \bar{\mathbf{z}}(\mathbf{x}) \end{array} \right\} \end{array} \right\} \\ & + \delta + N_{m(\mathbf{x}) \geq 0}(m). \end{aligned}$$

As in the proof of Theorem 3.1, $\lambda \neq 0$.

Employing the definition of convex hull we assert the existence of a positive integer N, nonnegative scalars $\lambda_1, ..., \lambda_N$ with $\lambda_1 + ... + \lambda_N = 1$, $\mathbf{w}_i \in M(m) \subset \mathbf{R}^p$ and $\varepsilon \in N_{m(\mathbf{x}) \geq 0}(m)$ such that

$$\begin{aligned} 0 = \lambda \eta_S^2 & \left[\begin{array}{c} \sum_{i=1}^N \lambda_i \left\{ \bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{M}_{\mathbf{w}_i} \bar{\mathbf{z}}(\mathbf{x}) + \left(\mathbf{b}_{\mathbf{w}_i}^T \bar{\mathbf{z}}(\mathbf{x}) \right)^2 m(\mathbf{x}) \right\} + \\ \nu \Omega^{-1/2} \left(\int_S \{ l_m(\mathbf{x}) m(\mathbf{x}) \}^2 d\mathbf{x} \right)^{-1/2} \\ \left\{ \begin{array}{c} l_m^2(\mathbf{x}) m(\mathbf{x}) - \\ 2 \bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{H}_T \left(\int_S l_m(\mathbf{x}) m^2(\mathbf{x}) \bar{\mathbf{z}}(\mathbf{x}) \bar{\mathbf{z}}^T(\mathbf{x}) d\mathbf{x} \right) \mathbf{B}^{-1} \bar{\mathbf{z}}(\mathbf{x}) \end{array} \right\} \end{array} \right] \\ & + \delta + \varepsilon. \end{aligned}$$

Consequently, there exists a symmetric matrix

$$\mathbf{P} = \eta_S^2 \left[\nu \Omega^{-1/2} \left(\int_S \{l_m(\mathbf{x}) m(\mathbf{x})\}^2 d\mathbf{x} \right)^{-1/2} \{\mathbf{P}_1 + \mathbf{P}_1^T\} - \sum_{i=1}^{N} \lambda_i \mathbf{M}_{\mathbf{w}_i} \right]$$

with $\mathbf{P}_1 = \mathbf{H}_T \left(\int_S l_m(\mathbf{x}) m^2(\mathbf{x}) \bar{\mathbf{z}}(\mathbf{x}) \bar{\mathbf{z}}^T(\mathbf{x}) d\mathbf{x} \right) \mathbf{B}^{-1}$, a positive semi-definite matrix

$$\mathbf{Q} = \eta_S^2 \sum_{i=1}^{N} \lambda_i \mathbf{b}_{\mathbf{w}_i} \mathbf{b}_{\mathbf{w}_i}^T,$$

a positive definite matrix

$$\mathbf{U} = \eta_S \nu^{1/2} \left(\Omega \int_S \{l_m(\mathbf{x}) m(\mathbf{x})\}^2 d\mathbf{x} \right)^{-1/4} \mathbf{H}_T,$$

and scalars

$$\varepsilon = \frac{\varepsilon_0}{\lambda}, \; d = -\frac{\delta}{\lambda}$$

such that

$$0 = \left[\bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{Q} \bar{\mathbf{z}}(\mathbf{x}) + \left\{ \bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{U} \bar{\mathbf{z}}(\mathbf{x}) \right\}^2 \right] m(\mathbf{x}) - \bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{P} \bar{\mathbf{z}}(\mathbf{x}) - d + \varepsilon. \tag{3.A.8}$$

The proof is now completed in a manner essentially identical to that of Theorem 3.1. □

Proof of Theorem 3.3. Again we give the proof only for extrapolation. In a manner very similar to that in the preceding two proofs we find that there exists a symmetric matrix $\mathbf{P}$, a positive semi-definite matrix $\mathbf{Q}$, a positive definite matrix $\mathbf{U}$ and a constant d such that on the set where $m(\mathbf{x}) > 0$,

$$\bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{Q} \bar{\mathbf{z}}(\mathbf{x}) m(\mathbf{x}) + \left\{ \bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{U} \bar{\mathbf{z}}(\mathbf{x}) \right\}^{2/3} m^{1/3}(\mathbf{x}) - \bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{P} \bar{\mathbf{z}}(\mathbf{x}) - d = 0.$$

Therefore the minimizing $m(\mathbf{x})$ is a solution to

$$a(\mathbf{x}) m^{1/3}(\mathbf{x}) + b(\mathbf{x}) m(\mathbf{x}) - c(\mathbf{x}) = 0, \tag{3.A.9}$$

where $a(\mathbf{x}) = \left\{ \bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{U} \bar{\mathbf{z}}(\mathbf{x}) \right\}^{2/3}$, $b(\mathbf{x}) = \bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{Q} \bar{\mathbf{z}}(\mathbf{x})$ and $c(\mathbf{x}) = \bar{\mathbf{z}}^T(\mathbf{x}) \mathbf{P} \bar{\mathbf{z}}(\mathbf{x}) + d$. Let $\tilde{m} = c - bm$. Then, (3.A.9) becomes

$$\tilde{m}^3 + \frac{a^3}{b} \tilde{m} - \frac{a^3 c}{b} = 0. \tag{3.A.10}$$

Since a and b are positive almost everywhere in S, (3.A.10) has only one real solution. Applying Cardano's formula for cubic equations (Dunham 1990), we obtain $\tilde{m}(\mathbf{x}) = k(\mathbf{x})$, where k is as at (3.24). Thus

$$m(\mathbf{x}) = \frac{c(\mathbf{x}) - k(\mathbf{x})}{b(\mathbf{x})}$$

on the set where $m(\mathbf{x}) > 0$. The rest of the proof is now essentially identical to that of Theorem 3.1. □

Acknowledgements

The research of both authors is supported by the Natural Sciences and Engineering Research Council of Canada. We thank Professors Jane Ye and Julie Zhou, University of Victoria, and Dr. Peilin Shi, University of British Columbia, for helpful discussions.

References

Adewale, A. (2002), "Robust Designs for Approximate Regression Models with Two Interacting Regressors," unpublished M.Sc. thesis, University of Alberta.

Atkinson, A.C., and Haines, L.M. (1996), "Designs for Nonlinear and Generalized Linear Models," in: S. Ghosh and C. R. Rao, eds., *Handbook of Statistics*, Vol. 13, 437-475.

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

Clarke, F.H. (1983), *Optimization and Nonsmooth Analysis*, Wiley, New York.

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

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

Dunham, W. (1990), "Cardano and the Solution of the Cubic," Ch. 6 (pp. 133-154) in *Journey through Genius: The Great Theorems of Mathematics,* Wiley, New York.

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

Ford, I., Titterington, D.M., and Kitsos, C.P. (1989), "Recent Advances in Nonlinear Experimental Design," *Technometrics*, 31, 49-60.

Ford, I., Torsney, B., and Wu, C.F.J. (1992), "The Use of a Canonical Form in the Construction of Locally Optimal Designs for Non-Linear Problems," *Journal of the Royal Statistical Society B*, 54, 569-583.

Heo, G., Schmuland, B., and Wiens, D.P. (2001); "Restricted Minimax Robust Designs for Misspecified Regression Models," *The Canadian Journal of Statistics*, 29, 117-128.

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.

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.

Shi, P. (2002), "Minimax Robust Designs for Misspecified Regression Models," unpublished Ph.D. thesis, University of Victoria.

Shi, P., Ye, J., and Zhou, J. (2003), "Minimax Robust Designs for Misspecified Regression Models," *The Canadian Journal of Statistics*, 31, 397-414.

Silvey, S.D. (1980). *Optimal Design*, Chapman and Hall, London.

Sinha, S. and Wiens, D.P. (2002), "Robust Sequential Designs for Nonlinear Regression," *The Canadian Journal of Statistics*, 30, 601-618.

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

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 the American Statistical Association*, 93, 1440-1450.

Ye, J., and Zhou, J. (2005), "Existence and Symmetry of Minimax Regression Designs," to appear in Journal of Statistical Planning and Inference.

CHAPTER IV

ROBUST PREDICTION AND EXTRAPOLATION DESIGNS FOR CENSORED DATA

Abstract This chapter presents the construction of robust designs for a possibly misspecified generalized linear regression model when the data are censored. The minimax designs and unbiased designs are found for maximum likelihood estimation in the context of both prediction and extrapolation problems. We extend preceding work of robust designs for complete data in Chapter 2 and 3 by incorporating censoring and maximum likelihood estimation. It also broadens former work of robust designs for censored data from others by considering both nonlinearity and much more arbitrary uncertainty in the fitted regression response, and by dropping all restrictions on the structure of the regressors. Solutions are derived by a nonsmooth optimization technique analytically and given in full generality. A typical example in accelerated life testing is also demonstrated.

4.1 Introduction

We investigate the construction of designs for both prediction and extrapolation of a regression response incorporating censored data. Such designs are of interest in problems of life testing where there are two commonly used testing-time saving plans: censoring and acceleration. In this chapter, we will address both.

Generally speaking, tests yielding complete data take too long to run especially for those products having long life-spans. To save time, the testing results can be

analyzed before all units fail. The data then consist of lifetime information on unfailed units, so called censored data. If such information is ignored, the estimates of mean lifetime would be definitely underestimated. Another purpose of censoring is that one can analyze the most recent test data while the test is still running.

When life testing runs at the stress levels within the range that the product would be normally used, the goal is prediction, namely the estimation of the mean response throughout the region of interest. For accelerated life testing (ALT), in which products are tested at higher than normal usage stress levels, the goal is extrapolation. For an extrapolation problem, if one is interested in estimating the mean response at a particular normal usage stress level which is lower than testing stress levels, we call it a one-point extrapolation problem; if one's interest falls into estimating that at a certain range of normal usage stress levels, we call it a general extrapolation problem.

There is considerable literature regarding robust regression designs for a possibly misspecified *linear* response when observations are complete. For prediction problems, those with homoscedastic errors have been studied by Box and Draper (1959), Huber (1975), and Wiens (1992) and those with possible heteroscedasticity were obtained by Wiens (1998). For one-point extrapolation problems, those with homoscedasticity were investigated by Draper and Herzberg (1973), Huber (1975), Lawless (1984), Spruill (1984); those with possible heteroscedasticity were observed in Chapter 2. For general extrapolation problems, those with homoscedasticity were discussed in Chapter 3; those with heteroscedasticity were studied by Fang and Wiens (1999) and Chapter 3.

For *nonlinear* regression problems without the consideration of model uncertainty, Atkinson and Haines (1996), and Ford, Titterington, and Kitsos (1989) presented various static and sequential designs; for those considering robustness, Sinha and Wiens (2002) provided the construction of sequential designs which were robust against model uncertainty, and Chapters 2 and 3 discuss the construction of static designs

which are robust against possibly misspecification in nonlinear models when observations are complete.

Recent work on robust designs with censored data in ALT are reported by Chaloner and Larntz (1992), Pascual and Montepiedra (2002), and Pascual and Montepiedra (2003), to name a few. These studies emphasize the robustness against model misspecification on the underlying distribution and assume that the "true" model belongs to, or is distributed, with a known prior, onto a set of several known candidates. Both Bayesian-type approaches and minimax strategies were used. Ginebra and Sen (1998) investigated optimal designs, which are robust against possibly misspecified parameter values on which the optimal designs depend. The explicit designs obtained in those works are under straight line regression.

This present work focuses on the robustness against possible misspecification in regression models that describe the behaviour of mean responses in relation to the explanatory variables. Such misspecification generates a bias in the estimation of the mean response. We assume that the "true" model involves an unknown member of a certain contamination class but may not be the assumed one. This work broadens the previous work of robust designs with censoring from others by considering both nonlinearity and much more arbitrary uncertainty in the fitted regression response and by dropping all restrictions on the structure of regressors. It also extends previous work on robust designs for such model settings for complete data by incorporating censoring and maximum likelihood estimation (MLE).

We consider a singly censored design with a specified censoring time for each stress level. The underlying distribution is assumed to be normal. For the lifetime, the underlying distribution is usually considered to be normal after the observations are transformed. For example, suppose some product's lifetime is possibly a lognormal distribution. In such case, we take the logarithm of all observations and censoring times, then carry out the regression method on the data after such transformations.

Let Y be the random variable, for instance, (transformed) lifetime of the product or material; let ζ be the (correspondingly transformed) censoring time which is constant for a fixed stress level; and let $\mathbf{x}$ represent the stress level and be a q-dimensional vector belonging to a bounded design space S.

We consider

$$Y(\mathbf{x}_i) = E(Y|\mathbf{x}_i) + \varepsilon_i,$$

with design points $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ freely chosen from S, where the ε_i's are uncorrelated and identically distributed with a normal distribution and a common variance σ^2. The mean response is regarded as being an only approximately known function of a linear function of a given $p \times 1$ regressor vector $\mathbf{z}^T(\mathbf{x})$, and unknown parameters:

$$E(Y|\mathbf{x}) \approx h\left(\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}_0\right).$$

The function h is strictly monotonic with a bounded second derivative. We assume $\left\|\mathbf{z}^T(\mathbf{x})\right\|$ is bounded on S. The "correct" vector $\boldsymbol{\theta}_0$ of regression parameters may be defined by

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

Then after introducing $f_n(\mathbf{x}) = \sqrt{n}[E(Y|\mathbf{x}) - h\left(\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}_0\right)]$, we obtain

$$E(Y|\mathbf{x}) = h(\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}_0) + n^{-1/2} f_n(\mathbf{x}). \tag{4.1}$$

Whenever it is clear from the context, we drop the subscript n on f_n. The contaminant f is unknown but relatively "small". Such misspecification may be generated by a transformation of the data for the purpose of enhancing normality. It may also be "viewed as arising from imprecision in the specification of h, or it can arise from a misspecified linear term and a two-term Taylor expansion" as discussed in Chapter 3.

We let $\bar{\mathbf{z}}(\mathbf{x}) = \left(\frac{dh}{d\mu}|_{\theta=\theta_0}\right) \mathbf{z}(\mathbf{x})$ with $\mu = \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}$. By the definition of $\boldsymbol{\theta}_0$, we have

$$\int_S \bar{\mathbf{z}}^T(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = 0. \tag{4.2}$$

To ensure that $\boldsymbol{\theta}_0$ is well-defined, we need both (4.2) and $\int_S \tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^T(\mathbf{x})d\mathbf{x}$ being non-singular which holds from Assumption (A) in Section 4.2. Let $\hat{\boldsymbol{\theta}}$ be the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}_0$ obtained from the censored data. We define the loss functions for the following cases:

(1) For prediction problems, where we attempt to estimate the mean response $E(Y|\mathbf{x})$ on the entire design space S, let

$$I_1 = n\int_S E[h\left(\mathbf{z}^T(\mathbf{x})\hat{\boldsymbol{\theta}}\right) - E(Y|\mathbf{x})]^2 d\mathbf{x};$$

(2) For extrapolation problems, we take

$$I_2 = n\int_T E[h\left(\mathbf{z}^T(\mathbf{x})\hat{\boldsymbol{\theta}}\right) - E(Y|\mathbf{x})]^2 \omega\left(d\mathbf{x}\right),$$

where $\omega\left\{\mathbf{x}_0\right\} = 1$ in the case of one-point extrapolation when we estimate $E(Y|\mathbf{x}_0)$, at $\mathbf{x}_0 \notin S$, by $\hat{Y}(\mathbf{x}_0) = h\left(\mathbf{z}^T(\mathbf{x_0})\hat{\boldsymbol{\theta}}\right)$; and ω is a Lebesgue measure in the case of general extrapolation with extrapolation region T assuming $\omega\left(T\right) > 0$, and $T \cap S = \phi$, where we intend to extrapolate $E(Y|\mathbf{x})$ to the entire region T .

These loss functions depend on the design measure $\xi = n^{-1}\Sigma_{i=1}^{n} I\left(\mathbf{x} = \mathbf{x}_i\right)$, where I is the indicator function, as well as on f. The following problems will be addressed in this chapter sequentially:

P1: Determine designs to minimize the maximum value of I_1 over f;

P2: Determine designs to minimize the maximum value of I_2 over f for one-point extrapolation;

P3: Determine designs to minimize the maximum value of I_2 over f for general extrapolation;

P4, P5, and P6: Determine unbiased designs in the context of prediction, one-point extrapolation, and general extrapolation respectively.

We assume that the contaminant f involved in P1-P6 varies within certain specified contamination classes.

The rest of this chapter is arranged as follows: Some mathematical preliminaries and notation are detailed in Section 4.2; the designs for P1-P3 are presented in Section 4.3; the designs for P4-P6 are delineated in Section 4.4; the computation of the resulting designs has been demonstrated using a typical ALT example in Section 4.5; and derivations of all theorems in this chapter are provided in an appendix.

4.2 Preliminaries and Notation

For any observation $y(\mathbf{x})$ at stress level $\mathbf{x}$, we define an indicator function, $c = c(y|\mathbf{x})$ in terms of the censoring time $\zeta(\mathbf{x})$ by letting it be 1 when failure occurs and 0 when an observation is censored, i.e.

$$c(y|\mathbf{x}) = \begin{cases} 1, \text{ when } y(\mathbf{x}) \leq \zeta(\mathbf{x}), \\ 0, \text{ when } y(\mathbf{x}) > \zeta(\mathbf{x}). \end{cases}$$

Let ϕ and Φ be the standard normal density and cumulative distribution function respectively. Under the fitted regression model $E(Y|\mathbf{x}) = h\left(\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}\right)$, the log likelihood of the ith individual observation (y_i, c_i) at stress level $\mathbf{x}_i$ is

$$\begin{aligned} l &= c_i \left\{ -\ln(\sigma) - \frac{1}{2}\ln(2\pi) - \frac{1}{2}\left(\frac{y_i - h\left(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\theta}\right)}{\sigma} \right)^2 \right\} + \\ & \quad (1-c_i)\left\{ \ln\left[1 - \Phi\left(\frac{\zeta(\mathbf{x}_i) - h\left(\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\theta}\right)}{\sigma} \right) \right] \right\}. \end{aligned}$$

Denote the standardized observation and censoring time at a specified stress level $\mathbf{x}$ by $w(\mathbf{x})$ and $\tau(\mathbf{x})$, i.e. $w(\mathbf{x}) = \frac{y(\mathbf{x}) - h\left(\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}_0\right)}{\sigma}$ and $\tau(\mathbf{x}) = \frac{\zeta(\mathbf{x}) - h\left(\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}_0\right)}{\sigma}$. Then we have

$$\begin{aligned} \left.\frac{\partial l_i}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} &= \frac{\bar{\mathbf{z}}^T(\mathbf{x}_i)}{\sigma}\left\{ c_i w_i + (1-c_i)\frac{\phi(\tau_i)}{1-\Phi(\tau_i)} \right\}, \text{ and} \\ \left.\frac{\partial^2 l_i}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} &= -\frac{\bar{\mathbf{z}}(\mathbf{x}_i)\bar{\mathbf{z}}^T(\mathbf{x}_i)}{\sigma^2}\left\{ c_i + (1-c_i)\left[\left(\frac{\phi(\tau_i)}{1-\Phi(\tau_i)} \right)^2 - \frac{\tau_i \phi(\tau_i)}{1-\Phi(\tau_i)} \right] \right\}, \end{aligned}$$

where $\tau_i = \tau(\mathbf{x}_i)$, $c_i = c(\mathbf{x}_i)$ and $w_i = w(\mathbf{x}_i)$. Both expressions above involve two random variables: c_i and $c_i w_i$. The following derives the expectations of, variances of, and covariance between these two variables based on the "true" model.

We notice that c_i and w_i have the following distributions:

$$\begin{aligned} c_i &\sim bin(1, P_i) \text{ with } P_i = P(c_i = 1), \\ w_i &\sim N\left(\frac{f(\mathbf{x}_i)}{\sqrt{n}\sigma}, 1\right); \end{aligned}$$

and their product is: $c_i w_i = \begin{cases} w_i, & \text{when } w_i \le \tau_i, \\ 0, & \text{when } w_i > \tau_i. \end{cases}$.

According to (4.1), we obtain

$$\begin{aligned} E(c_i) &= \Phi(\tau_i) - \phi(\tau_i)\frac{f(\mathbf{x}_i)}{\sqrt{n}\sigma} + o\left(\frac{f(\mathbf{x}_i)}{\sqrt{n}}\right), \\ E(c_i w_i) &= -\phi(\tau_i) + [\Phi(\tau_i) - \tau_i\phi(\tau_i)]\frac{f(\mathbf{x}_i)}{\sqrt{n}\sigma} + o\left(\frac{f(\mathbf{x}_i)}{\sqrt{n}}\right), \\ Var(c_i) &= \Phi(\tau_i) - \Phi^2(\tau_i) + \phi(\tau_i)[2\Phi(\tau_i) - 1]\frac{f(\mathbf{x}_i)}{\sqrt{n}\sigma} + o\left(\frac{f(\mathbf{x}_i)}{\sqrt{n}}\right), \\ Var(c_i w_i) &= [\Phi(\tau_i) - \tau_i\phi(\tau_i) - \phi^2(\tau_i)] \\ &\quad + \phi(\tau_i)\left[2\Phi(\tau_i) - 2\tau_i\phi(\tau_i) - \tau_i^2 + 2\right]\frac{f(\mathbf{x}_i)}{\sqrt{n}\sigma} + o\left(\frac{f(\mathbf{x}_i)}{\sqrt{n}}\right), \\ Cov(c_i, c_i w_i) &= -\phi(\tau_i)[1 - \Phi(\tau_i)] \\ &\quad - \left\{\phi^2(\tau_i) + [1 - \Phi(\tau_i)][\tau_i\phi(\tau_i) - \Phi(\tau_i)]\right\}\frac{f(\mathbf{x}_i)}{\sqrt{n}\sigma} + o\left(\frac{f(\mathbf{x}_i)}{\sqrt{n}}\right). \end{aligned}$$

Hence,

$$E\left(\frac{\partial l_i}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) = a(\tau_i)\frac{f(\mathbf{x}_i)\bar{\mathbf{z}}^T(\mathbf{x}_i)}{\sqrt{n}\sigma^2} + o\left(\frac{f(\mathbf{x}_i)}{\sqrt{n}}\right),$$

$$\begin{aligned} &E\left(-\frac{\partial^2 l_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) \\ &= \frac{\bar{\mathbf{z}}(\mathbf{x}_i)\bar{\mathbf{z}}^T(\mathbf{x}_i)}{\sigma^2}\left\{a(\tau_i) + \left(b(\tau_i) + a(\tau_i)\left[\left(\frac{d^2h}{d\mu^2}\right)\left(\frac{dh}{d\mu}\right)^{-2}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\frac{f(\mathbf{x}_i)}{\sqrt{n}\sigma}\right\} + o\left(\frac{f(\mathbf{x}_i)}{\sqrt{n}}\right), \end{aligned}$$

and

$$\begin{aligned} &Cov\left(\frac{\partial l_i}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) \\ &= \frac{\bar{\mathbf{z}}(\mathbf{x}_i)\bar{\mathbf{z}}^T(\mathbf{x}_i)}{\sigma^2}\left\{a(\tau_i) + \left[2\phi(\tau_i) - \tau_i^2\phi(\tau_i) + \frac{\phi^3(\tau_i)}{\{1 - \Phi(\tau_i)\}^2}\right]\frac{f(\mathbf{x}_i)}{\sqrt{n}\sigma}\right\} + o(\frac{f(\mathbf{x}_i)}{\sqrt{n}}), \end{aligned}$$

where $a(\tau(\mathbf{x})) = \Phi(\tau) - \tau\phi(\tau) + \frac{\phi^2(\tau)}{1-\Phi(\tau)}$ and $b(\tau) = \frac{\phi^3(\tau)}{\{1-\Phi(\tau)\}^2} - \frac{\tau\phi^2(\tau)}{1-\Phi(\tau)} - \phi(\tau)$. We note that $a(\tau(\mathbf{x}))$ is a function of $\mathbf{x}, \boldsymbol{\theta_0}$, and σ, but the dependency on $\mathbf{x}, \boldsymbol{\theta_0}$, and σ is only through τ. For readability, we use $a(\tau)$ when we emphasize its dependence on τ and $a(\mathbf{x})$ otherwise with

$$a(\mathbf{x}) = \Phi\left(\frac{\zeta(\mathbf{x}) - h\left(\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}_0\right)}{\sigma}\right) + \frac{\phi^2\left(\frac{\zeta(\mathbf{x})-h\left(\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}_0\right)}{\sigma}\right)}{1-\Phi\left(\frac{\zeta(\mathbf{x})-h\left(\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}_0\right)}{\sigma}\right)}$$
$$-\frac{\zeta(\mathbf{x}) - h\left(\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}_0\right)}{\sigma}\phi\left(\frac{\zeta(\mathbf{x}) - h\left(\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}_0\right)}{\sigma}\right).$$

To avoid trivialities and to make sure of the nonsingularity of a number of relevant matrices, we assume that the design space S and extrapolation space T satisfy

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

We assume

$$\int_S f^2(\mathbf{x})d\mathbf{x} \leq \eta_S^2 < \infty, \tag{4.3}$$

for a positive constant η_S, and also define the following matrices and vectors:

$$\mathbf{A}_S = \textstyle\int_S \bar{\mathbf{z}}(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})d\mathbf{x}, \quad \mathbf{A}_0 = \bar{\mathbf{z}}(\mathbf{x}_0)\bar{\mathbf{z}}^T(\mathbf{x}_0),$$

$$\mathbf{A}_T = \textstyle\int_T \bar{\mathbf{z}}(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})d\mathbf{x}, \quad \mathbf{B} = \textstyle\int_S a(\mathbf{x})\bar{\mathbf{z}}(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})\xi(d\mathbf{x}),$$

$$\mathbf{H}_S = \mathbf{B}^{-1}\mathbf{A}_S\mathbf{B}^{-1}, \qquad \mathbf{H}_0 = \mathbf{B}^{-1}\mathbf{A}_0\mathbf{B}^{-1},$$

$$\mathbf{H}_T = \mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1}, \qquad \mathbf{b}_{f,S} = \textstyle\int_S a(\mathbf{x})\bar{\mathbf{z}}(\mathbf{x})f(\mathbf{x})\xi(d\mathbf{x}),$$

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

It follows from (A) stated above that $\mathbf{A}_S$, $\mathbf{A}_T$ are nonsingular and that $\mathbf{B}$ is also nonsingular whenever ξ does not place mass on sets of Lebesgue measure zero. This

requirement turns out to be necessary since ξ has to be absolutely continuous due to (4.3) as discussed later in this section.

By virtue of our assumption on f, $\mathbf{z}$ and h and the definition of $\mathbf{B}$ and $\mathbf{b}_{f,S}$, we obtain the following results. The asymptotic information matrix of $\boldsymbol{\theta}_0$ is

$$\begin{aligned}\mathbf{I}(\boldsymbol{\theta}_0) &= \lim_{n\longrightarrow\infty}\frac{1}{n}E\left(-\sum_{i=1}^{n}\frac{\partial^2 l_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\\ &= \frac{1}{\sigma^2}\int_S a(\mathbf{x})\bar{\mathbf{z}}(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})\xi(d\mathbf{x})\\ &= \frac{1}{\sigma^2}\mathbf{B}.\end{aligned}$$

The asymptotic expectation of the score function evaluated at $\boldsymbol{\theta}_0$ is

$$\begin{aligned}\bar{\mathbf{b}}_f(\boldsymbol{\theta}_0) &= \frac{1}{\sqrt{n}}\lim_{n\longrightarrow\infty}\frac{1}{n}E\left(\sqrt{n}\sum_{i=1}^{n}\frac{\partial l_i}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\\ &= \frac{1}{\sqrt{n}\sigma^2}\int_S a(\mathbf{x})\bar{\mathbf{z}}(\mathbf{x})f(\mathbf{x})\xi(d\mathbf{x})\\ &= \frac{1}{\sqrt{n}\sigma^2}\mathbf{b}_{f,S}.\end{aligned}$$

And the asymptotic variance-covariance matrix of the score function evaluated at $\boldsymbol{\theta}_0$ is

$$\begin{aligned}\mathbf{C}(\boldsymbol{\theta}_0) &= \lim_{n\longrightarrow\infty}\frac{1}{n}\sum_{i=1}^{n}Cov\left(\frac{\partial l_i}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\\ &= \frac{1}{\sigma^2}\int_S a(\mathbf{x})\bar{\mathbf{z}}(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})\xi(d\mathbf{x})\\ &= \frac{1}{\sigma^2}\mathbf{B}.\end{aligned}$$

Since the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ is a root of the score function which can be expanded around $\boldsymbol{\theta}_0$ as

$$\frac{1}{n}\left(\sum_{i=1}^{n}\frac{\partial l_i}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)+\frac{1}{n}\left(-\sum_{i=1}^{n}\frac{\partial^2 l_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+O\left((\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\right),$$

we have

$$(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\approx\frac{1}{n}\left(\sum_{i=1}^{n}\frac{\partial l_i}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\mathbf{I}^{-1}(\boldsymbol{\theta}_0).$$

Consequently, the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is then

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \sim AN(\mathbf{B}^{-1}(\boldsymbol{\theta}_0)\mathbf{b}_{f,S}(\boldsymbol{\theta}_0), \sigma^2\mathbf{B}^{-1}(\boldsymbol{\theta}_0)).$$

We denote

$$\begin{aligned}
\mathcal{F}_1 &= \{f|\ \int_S \bar{\mathbf{z}}(\mathbf{x})f(\mathbf{x})d\mathbf{x} = 0,\ \int_S f^2(\mathbf{x})d\mathbf{x} \le \eta_S^2 < \infty\},\\
\mathcal{F}_2 &= \{f|\ \int_S \bar{\mathbf{z}}(\mathbf{x})f(\mathbf{x})d\mathbf{x} = 0,\ \int_S f^2(\mathbf{x})d\mathbf{x} \le \eta_S^2 < \infty,\ |f(\mathbf{x}_0)| \le \eta_0 < \infty\},\ \text{and}\\
\mathcal{F}_3 &= \{f|\ \int_S \bar{\mathbf{z}}(\mathbf{x})f(\mathbf{x})d\mathbf{x} = 0,\ \int_S f^2(\mathbf{x})d\mathbf{x} \le \eta_S^2 < \infty,\ \int_T f^2(\mathbf{x})d\mathbf{x} \le {\eta_T}^2 < \infty\},
\end{aligned}$$

for positive constants η_0 and η_T. For the regression model (4.1), we assume that the contamination function $f(\mathbf{x})$ is an unknown member of one of the classes above. In fact, since the contamination classes above are so full, ξ has to have a density in order to guarantee $\sup_{f\in\mathcal{F}_j} I_i$, with $(i,j) = (1,1),(2,2)$, or $(2,3)$, is finite. This can be established by modifying the proof of Lemma 1 of Wiens (1992). In practice, ξ must be discrete. A consequence is that the optimal design obtained must be approximated to make them implementable. Approximation methods are discussed in Heo, Schmuland, and Wiens (2001) and also in Chapter 5 of this thesis.

Let $k(\mathbf{x})$ be the density of $\xi(\mathbf{x})$, then we have

$$\begin{aligned}
\mathbf{B} &= \int_S a(\mathbf{x})k(\mathbf{x})\bar{\mathbf{z}}(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})d\mathbf{x},\\
\mathbf{b}_{f,S} &= \int_S a(\mathbf{x})k(\mathbf{x})\bar{\mathbf{z}}(\mathbf{x})f(\mathbf{x})d\mathbf{x}.
\end{aligned}$$

We also define $\mathbf{K} = \int_S a^2(\mathbf{x})k^2(\mathbf{x})\bar{\mathbf{z}}(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})d\mathbf{x}$ and $\mathbf{G} = \mathbf{K} - \mathbf{B}\mathbf{A}_S^{-1}\mathbf{B}$. There are two facts we mark:

(1) $\mathbf{G}$ is positive semidefinite since, for any vector $\mathbf{c}$,

$$\mathbf{c}^T\mathbf{G}\mathbf{c} = \int_S \{\mathbf{c}^T[k(\mathbf{x})a(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}_S^{-1})\bar{\mathbf{z}}(\mathbf{x})]\}^2 d\mathbf{x} \ge 0;$$

(2) a is a nonnegative function since

$$a(\tau)[1 - \Phi(\tau)] = [\Phi(\tau) - \tau\phi(\tau)][1 - \Phi(\tau)] + \phi^2(\tau)$$

with $\Phi(\tau) - \tau\phi(\tau) > 0$.

4.3 Minimax Designs for Censored Data

In this section, we investigate the optimal designs that minimize the maximum value of the loss, over f, in the following three cases:

(P1) prediction problems with $f \in \mathcal{F}_1$;

(P2) one-point extrapolation problems with $f \in \mathcal{F}_2$; and

(P3) general extrapolation problems with $f \in \mathcal{F}_3$.

4.3.1 Minimax Designs for Prediction: Solutions to P1

The loss function for Problem P1 is

$$\begin{aligned} I_1 &= n \int_S E[h\left(\mathbf{z}^T(\mathbf{x})\hat{\boldsymbol{\theta}}\right) - h\left(\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}_0\right) - n^{-1/2} f(\mathbf{x})]^2 d\mathbf{x} \\ &= \mathbf{b}_{f,S}^T \mathbf{H}_S^{-1} \mathbf{b}_{f,S} + \sigma^2 tr(\mathbf{A}_S \mathbf{B}^{-1}) + \int_S f^2(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

Let $\nu := \sigma^2/\eta_S^2$ represent the relative importance of variance versus bias. We define $\lambda_k^{(1)}$ to be the largest solution to $|\mathbf{G} - \lambda \mathbf{H}_S| = 0$ and $\mathbf{c}_1$ to be any vector satisfying $(\mathbf{G}\mathbf{H}_S^{-1}\mathbf{G} - \lambda_k^{(1)}\mathbf{G})\mathbf{c} = 0$, and

$$\mathbf{c}^T \mathbf{G} \mathbf{c} = 1. \tag{4.4}$$

Given fixed $k(\mathbf{x})$, the "max" part of the minimax solution is presented in Theorem 4.1.

Theorem 4.1 *The maximum of I_1 is*

$$\sup_{f \in \mathcal{F}_1} I_1(f, \xi) = \eta_S^2 [\lambda_k^{(1)} + 1 + \nu tr(\mathbf{A}_S \mathbf{B}^{-1})], \tag{4.5}$$

attained at

$$f_k(\mathbf{x}) = \eta_S \, \tilde{\mathbf{z}}^T(\mathbf{x}) \{a(\mathbf{x}) k(\mathbf{x}) \mathbf{I} - \mathbf{A}_S^{-1} \mathbf{B}\} \mathbf{c}_1.$$

Problem P1 has becomeone of finding a density $k(\mathbf{x})$ that minimizes (4.5). The following theorem provides the analytical form of such minimax design density.

Theorem 4.2 *The design density $k(\mathbf{x})$ minimizing (4.5) for prediction is of the form*

$$k(\mathbf{x}) = \frac{\left[a(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{P}\bar{\mathbf{z}}(\mathbf{x}) + d\right]^+}{a^2(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{Q}\bar{\mathbf{z}}(\mathbf{x})}, \tag{4.6}$$

where $(w)^+ = \max(w, 0)$, for some constant symmetric matrix $\mathbf{P}$, a positive semi-definite matrix $\mathbf{Q}$, and a constant d that minimize (4.5) and satisfy $\int_S k(\mathbf{x})d\mathbf{x} = 1$.

The following presents two models with different censoring plans. These two models will serve as two typical examples throughout this section and the sections hereafter in the context of all prediction, one-point, and general extrapolation problems. Prior to introducing these models, we first describe the definitions of the two censoring plans involved in these upcoming models: (1) Time (Type I) censoring is where the data are censored at a predefined time; (2) Failure (Type II) censoring is where the data are censored after a prespecified number of failures. Detailed information about these and other types of censoring can be found in Nelson (1990).

Model 1: We suppose that the experimenter plans a design under the assumed regression model:

$$E(Y|x) = \theta_0 + \theta_1 x,$$

and employs time censoring. The data are collected at a fixed time $\zeta(x) \equiv \zeta$ for all test units at all stress levels. Note that for life testing, Y and x stand for the transformed lifetime and stress respectively. Such transformations are sometimes employed for the purpose of enhancing both linearity and normality simultaneously. So, it is sensible to consider that the regression model assumed above is approximately true.

Model 2: For the nonlinear regression model:

$$E(Y|x) \approx h(\theta_0 + \theta_1 x),$$

where $h(z) = e^z$, we suppose that the failure censoring is planned with a constant expected proportion of failures at all stress levels. We assume $\tau(x) \equiv \tau$. Namely, the experimenter expects that about $100 \times \Phi(\tau)\%$ of the units fail at each stress level. So, $a(\tau)$ remains constant as well from its definition.

For both Model 1 and 2, we take $S = [b_1, b_2]$.

Example 1. According to Theorem 4.2, the locally optimal robust prediction design for Model 1 is given by

$$k(x) = \left[\frac{a_1 + a_2 x + a_3 x^2}{a(\tau)(a_4 + a_5 x + a_6 x^2)} + \frac{d}{a^2(\tau)(a_4 + a_5 x + a_6 x^2)}\right]^+,$$

where $\tau = \frac{\zeta - \theta_0 - \theta_1 x}{\sigma}$, a_4 and a_6 are nonnegative and satisfy $4a_4 a_6 \geq a_5^2$. In addition, $a_1 - a_6$ and d are selected in order to minimize (4.5) subject to $\int_0^1 k(x)dx = 1$.

Example 2. For Model 2, as a result of Theorem 4.2, the locally optimal design density for prediction is of the form

$$k(x) = \left[\frac{a_1 + a_2 x + a_3 x^2}{a_4 + a_5 x + a_6 x^2} + \frac{d}{e^{2\theta_1 x}(a_4 + a_5 x + a_6 x^2)}\right]^+,$$

where $a_4 \geq 0$, $a_6 \geq 0$, and $4a_4 a_6 \geq a_5^2$. Besides, $a_1 - a_6$ and d are determined so as to minimize (4.5) subject to $\int_0^1 k(x)dx = 1$. The dependence of the design on θ_1 makes such a design only locally optimal. This issue will be addressed in Section 4.5.

4.3.2 Minimax Designs for One-Point Extrapolation: Solutions to P2

The loss function for Problem P2 is

$$\begin{aligned} I_2 &= nE[h(\mathbf{z}^T(\mathbf{x}_0)\hat{\boldsymbol{\theta}}) - h\left(\mathbf{z}^T(\mathbf{x}_0)\boldsymbol{\theta}_0\right) - n^{-1/2} f(\mathbf{x}_0)]^2 \\ &= \mathbf{b}_{f,S}^T \mathbf{H}_0^{-1} \mathbf{b}_{f,S} - 2f(\mathbf{x}_0)\bar{\mathbf{z}}^T(\mathbf{x}_0)\mathbf{B}^{-1}\mathbf{b}_{f,S} + \sigma^2 \bar{\mathbf{z}}^T(\mathbf{x}_0)\mathbf{B}^{-1}\bar{\mathbf{z}}(\mathbf{x}_0) + f^2(\mathbf{x}_0). \end{aligned}$$

Let $r_{\mathbf{x}_0,S} := \eta_0/\eta_S$ represent the relative amount of model response uncertainty at the extrapolation point and within the design space. We define $\lambda_k^{(2)} = \bar{\mathbf{z}}^T(\mathbf{x}_0)\mathbf{B}^{-1}\mathbf{G}\mathbf{B}^{-1}\bar{\mathbf{z}}(\mathbf{x}_0)$, and $\mathbf{c}_2 = \frac{\mathbf{B}^{-1}\bar{\mathbf{z}}(\mathbf{x}_0)}{\sqrt{\lambda_k^{(2)}}}$. For a fixed $k(\mathbf{x})$, the maximization part of the minimax solutions to Problem P2 is given by Theorem 4.3.

Theorem 4.3 *The maximum of I_2 is*

$$\sup_{f\in\mathcal{F}_2} I_2 = \eta_S^2[(\sqrt{\lambda_k^{(2)}}+r_{\mathbf{x}_0,S})^2 + \nu\tilde{\mathbf{z}}^T(\mathbf{x}_0)\mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)], \tag{4.7}$$

attained at

$$f_k(\mathbf{x}) = \begin{cases} \eta_S\tilde{\mathbf{z}}^T(\mathbf{x})[a(\mathbf{x})k(\mathbf{x})\mathbf{I}-\mathbf{A}_S^{-1}\mathbf{B}]\mathbf{c}_2, & \mathbf{x}\in S, \\ -\eta_0, & \mathbf{x}=\mathbf{x}_0. \end{cases}$$

Problem P2 has become that of finding a density that minimizes (4.7). The minimax solution for P2 is presented in Theorem 4.4 below.

Theorem 4.4 *The design density $k(\mathbf{x})$ minimizing (4.7) for one-point extrapolation is*

$$k(\mathbf{x}) = \left[\frac{\mathbf{z}^T(\mathbf{x})\boldsymbol{\alpha}}{a(\mathbf{x})\mathbf{z}^T(\mathbf{x})\boldsymbol{\beta}} + \frac{\lambda}{a^2(\mathbf{x})[\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta}]^2}\right]^+$$

for some $p\times 1$ vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and constant λ which satisfy: (i) $\int_S k(\mathbf{x})d\mathbf{x}=1$, (ii) minimize (4.7).

Example 3. Recall Model 1. Suppose that the estimation extrapolates to one point x_0: either greater than b_2 or less than b_1. According to Theorem 4.4, the locally optimal robust one-point extrapolation design for this model is given by

$$k(x) = \left[\frac{a_1x+a_2}{a(\tau)(a_3x+a_4)} + \frac{a_5}{\{a(\tau)(a_3x+a_4)\}^2}\right]^+,$$

and a_1-a_5 are chosen in order to minimize (4.7) subject to $\int_0^1 k(x)dx=1$.

Example 4. For Model 2 with either $x_0>b_2$ or $x_0<b_1$, the locally optimal design density for one-point extrapolation is given by

$$k(x) = \left[\frac{a_1x+a_2}{a_3x+a_4} + \frac{a_5}{e^{2\theta_1x}(a_3x+a_4)^2}\right]^+, \tag{4.8}$$

where a_1-a_5 are again selected by minimizing (4.7) subject to $\int_0^1 k(x)dx=1$. The computation of the numerical values for a_1-a_5 in this design will be presented in Section 4.5.

4.3.3 Minimax Designs for General Extrapolation: Solutions to P3

The loss function for Problem P3 is

$$\begin{aligned} I_2 &= n\int_T E[h\left(\mathbf{z}^T(\mathbf{x})\hat{\boldsymbol{\theta}}\right) - h\left(\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}_0\right) - n^{-1/2}f(\mathbf{x})]^2 d\mathbf{x} \\ &= \mathbf{b}_{f,S}^T\mathbf{H}_T\mathbf{b}_{f,S} - 2\mathbf{b}_{f,T}^T\mathbf{B}^{-1}\mathbf{b}_{f,S} + \sigma^2 tr(\mathbf{A}_T\mathbf{B}^{-1}) + \int_T f^2(\mathbf{x})d\mathbf{x}. \end{aligned}$$

We denote $r_{T,S} := \eta_T/\eta_S$ for the relative amount of model response uncertainty in the extrapolation and design spaces. We also denote $\lambda_k^{(3)}$ to be the largest solution to $|\mathbf{G} - \lambda\mathbf{H}_T| = 0$ and let $\mathbf{c}_3$ be any vector satisfying $(\mathbf{G}\mathbf{H}_T\mathbf{G} - \lambda_k^{(3)}\mathbf{G})\mathbf{c} = 0$ and (4.4). The maximum of I_2 is given by Theorem 4.5.

Theorem 4.5 *The maximum of I_2 is*

$$\sup_{f\in\mathcal{F}_3} I_2(f,\xi) = \eta_S^2[(\sqrt{\lambda_k^{(3)}} + r_{T,S})^2 + \nu tr(\mathbf{A}_T\mathbf{B}^{-1})], \tag{4.9}$$

attained at

$$f_k(\mathbf{x}) = \begin{cases} \eta_S\bar{\mathbf{z}}^T(\mathbf{x})[a(\mathbf{x})k(\mathbf{x})\mathbf{I} - \mathbf{A}_S^{-1}\mathbf{B}]\mathbf{c}_3, & \mathbf{x}\in S, \\ -\frac{\eta_T\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{B}^{-1}\mathbf{G}\mathbf{c}_3}{\sqrt{\lambda_k^{(3)}}}, & \mathbf{x}\in T. \end{cases}$$

Problem P3 now is to find a density that minimizes (4.9). The following theorem gives the optimal minimax design density for the general extrapolation problem (P3), which has the same form as (4.6) for P1.

Theorem 4.6 *The design density $k(\mathbf{x})$ minimizing (4.9) for general extrapolation is*

$$k(\mathbf{x}) = \frac{[a(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{P}\bar{\mathbf{z}}(\mathbf{x}) + d]^+}{a^2(\mathbf{x})\bar{\mathbf{z}}^T(\mathbf{x})\mathbf{Q}\bar{\mathbf{z}}(\mathbf{x})}$$

for some constant symmetric matrix $\mathbf{P}$, a positive semi-definite matrix $\mathbf{Q}$, and a constant d that minimize (4.9) and satisfy $\int_S k(\mathbf{x})d\mathbf{x} = 1$.

Example 5. For Model 1 and in the context of general extrapolation with extrapolation space $T = [t_1, t_2]\backslash[b_1, b_2]$, as a result of Theorem 4.6, the locally optimal robust extrapolation design is of the form

$$k(x) = \left[\frac{(a_1 + a_2x + a_3x^2)}{(a_4 + a_5x + a_6x^2)} + \frac{d}{a(\tau)(a_4 + a_5x + a_6x^2)}\right]^+,$$

where a_4 and a_6 are nonnegative and satisfy $4a_4a_6 \geq a_5^2$. Moreover, $a_1 - a_6$ and d are chosen to minimize (4.9) subject to $\int_0^1 k(x)dx = 1$.

Example 6. With Model 2 in the context of general extrapolation with extrapolation space T defined as in Example 5, by Theorem 4.6 the locally optimal design density for general extrapolation is given by

$$k(x) = \left[\frac{e^{2\theta_1 x}(a_1 + a_2x + a_3x^2) + d}{e^{2\theta_1 x}(a_4 + a_5x + a_6x^2)}\right]^+,$$

where $a_4 \geq 0$, $a_6 \geq 0$, $4a_4a_6 \geq a_5^2$, and $a_1 - a_6$, as well as d are selected so as to minimize (4.9) conditional on $\int_0^1 k(x)dx = 1$.

4.4 Unbiased Designs for Censored Data: Solutions to P4-P6

We say that a design/censoring pair (ξ, ζ) is unbiased if it satisfies

$$E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}_0 \quad \text{for all } f \in \mathcal{F}_i,\ i = 1, 2, \text{ or } 3$$

so that $\sup_f IB_i(f, \xi, \zeta) = 0$ for $i = 1, 2$, or 3, where we define

$$IB_1(f, \xi, \zeta) = \mathbf{b}_{f,S}^T \mathbf{H}_S^{-1} \mathbf{b}_{f,S}$$

for P4,

$$IB_2(f, \xi, \zeta) = \mathbf{b}_{f,S}^T \mathbf{H}_0^{-1} \mathbf{b}_{f,S} - 2f(\mathbf{x}_0)\bar{\mathbf{z}}^T(\mathbf{x}_0)\mathbf{B}^{-1}\mathbf{b}_{f,S}$$

for P5, and

$$IB_3(f, \xi, \zeta) = \mathbf{b}_{f,S}^T \mathbf{H}_T \mathbf{b}_{f,S} - 2\mathbf{b}_{f,T}^T \mathbf{B}^{-1} \mathbf{b}_{f,S}$$

for P6.

P4, P5, or P6 involves finding the design such that the maximum, over f, of its matching bias: $\sup_f IB_i(f, \xi, \zeta)$ for $i = 1, 2$, or 3 is zero. Let $\alpha = \{\int_S a^{-1}(\mathbf{x})d\mathbf{x}\}^{-1}$. The following theorem offers a necessary and sufficient condition for unbiasedness, the unbiased designs and the resulting losses for P4, P5, and P6 respectively.

Theorem 4.7 *(a) The design $k(\mathbf{x})$ is unbiased if and only if*

$$a(\mathbf{x})k(\mathbf{x}) \equiv \alpha;$$

(b) The unbiased design density is

$$k(\mathbf{x}) = \alpha a^{-1}(\mathbf{x});$$

(c) The corresponding losses under unbiased designs are as follows:

(i) for prediction,

$$I_1 = \eta_S^2 + \alpha p \sigma^2;$$

(ii) for one-point extrapolation at $\mathbf{x}_0$,

$$I_2 = \eta_T^2 + \alpha\sigma^2 \bar{\mathbf{z}}^T(\mathbf{x}_0)\mathbf{A}_S^{-1}\bar{\mathbf{z}}(\mathbf{x}_0);$$

(iii) for general extrapolation, $I_2 = \eta_T^2 + \alpha\sigma^2 tr(\mathbf{A}_T\mathbf{A}_S^{-1})$.

We notice that the unbiased designs for the cases of prediction, one-point and general extrapolations are the same.

Example 7. With Model 1, as stated in Theorem 4.7, the locally unbiased design is of the form

$$k(x) = \{\int_S a^{-1}(x)dx\}^{-1} a^{-1}(x),$$

where

$$a(x) = \Phi\left(\frac{\zeta - \theta_0 - \theta_1 x}{\sigma}\right) - \frac{\zeta - \theta_0 - \theta_1 x}{\sigma}\phi(\frac{\zeta - \theta_0 - \theta_1 x}{\sigma}) + \frac{\phi^2(\frac{\zeta-\theta_0-\theta_1 x}{\sigma})}{1 - \Phi(\frac{\zeta-\theta_0-\theta_1 x}{\sigma})}.$$

Example 8. For Model 2, the unbiased robust design is uniform with density $k(x) = \{\int_S dx\}^{-1}$ since $a(x)$ is constant in this model.

4.5 Computation

In this section, we demonstrate the computation of numerical values of the constants in our constructed designs using one typical ALT example: Model 2 with one-point extrapolation.

As indicated in Example 4, the locally optimal design density for one-point extrapolation is given by (4.8). We assume $S = [0, 1]$ and $x_0 > 1$. We also let τ be a predefined constant which is essentially the standard normality quantile corresponding to the expected percentage of failures. For instance, if the experimenter plans to employ Type II censoring and expects 70% of the units to fail at each stress level, $\tau = \Phi^{-1}(0.7) = 0.525$. Then, $a(\tau)$ remains constant given by $a(\tau = 0.525) = 0.921$.

For Model 2, since the constant term $a(\tau)$ has been cancelled out in the first term of (4.7), it is only involved in the second term of (4.7). Let $\rho = v/a(\tau)$. For prespecified model parameters, the design varies when x_0, $r_{\mathbf{x}_0,S}$, or ρ changes. All of x_0, $r_{\mathbf{x}_0,S}$ and ρ can be determined by the experimenter. We also notice that the term $e^{2\theta_0}$ has been cancelled out in both the first and second terms of (4.7). However, for fixed x_0, $r_{\mathbf{x}_0,S}$, and ρ, the optimal design depends still on the value of θ_1. Therefore, it is only locally optimal. To deal with this issue, we search for 'locally most robust' designs as discussed in Chapter 2. To do this, firstly we obtain the local optimal design for an initial value of θ_1; secondly, for this optimal design, we take a further maximum of the loss as θ_1 varies over some interval I and record the least favourable θ_1 (written θ_1^{LF}); thirdly, we determine the coefficients of $k(x)$ so as to minimize this maximum loss for θ_1^{LF}. We repeat this procedure until θ_1^{LF} converges.

To illustrate the approach described above, we consider the Class-B insulation data from Nelson (1990, Table 4.1 of Chapter 3). Those data are collected from a singly time-censored ALT conducted using a uniform design on four specified testing levels. The acceleration stress is temperature. The intention of this experiment is to estimate the lifetime for electric motors at the normal usage temperature of 130^0C.

Table 4.1. Numerical values for (4.8)
$S = [0,1]$, $\theta_1 = 0.987$, $r_{\mathbf{x}_0,S} = 1$, and $x_0 = 1.98$.

ρ	a_1	a_2	a_3	a_5
0.5	−0.190	0.0003	−0.730	1.131
1	−0.420	0.002	−0.806	1.146
2	−0.676	0.00003	−0.858	1.196

The failures obtained from this test are at stress levels ranging from 170^0C to 220^0C. The Arrhenius-lognormal model is fitted for those data and the MLEs for the model parameters are computed by Nelson (1990). The transformation of the lifetime used in this example is the logarithm and that of stress t is 1000 times the inverse of the absolute temperature in degrees Kelvin, i.e., $x' = \frac{1000}{t+273.16}$. For simplicity, we transform x' to our stress variable x with domain of $[0,1]$ through the linear transformation $x = \frac{-2.028+x'}{0.229}$. Under such transformations, the MLE of θ_1 for the nominal model is $\hat{\theta}_1 = 0.987$ and the corresponding 99% confidence interval for θ_1 is $(0.730, 1.243)$. Taking the model misspecification into account, we consider an even broader region $\theta_1 \in I = [0.5, 1.5]$. We apply the same extrapolation point $x_0 = 1.98$ as employed in Nelson (1990), which is equivalent to the normal usage temperature: $t_0 = 130^0C$.

It should be noted that in (4.8), if one of $a_1 - a_5$ is nonzero, then we can assume that it is 1. In the following computation we take $a_4 = 1$. For locally optimal designs when $\theta_1 = 0.987$, see Table 4.1 for the numerical values of the constants in (4.8) with various ρ and Figure 17(a) for the plots. For locally most robust designs, we carry out the process described above for $I = [0.5, 1.5]$ and several ρ, each time starting at $\theta_1 = 0.987$. The locally most robust designs are detailed in Table 4.2. In each case, we find that the least favourable θ_1 within I is 1.5. See Figure 17(b) for the plots. All plots use $a_4 = 1$ and $r_{\mathbf{x}_0,S} = 1$.

We note that for the extreme case, when the data are complete: $\tau \longrightarrow \infty$, we have $a(\tau) = 1$. In this case, the results obtained in Section 4.3 degenerate into the exact optimal robust designs for complete data which are presented in Chapter 2 and

Table 4.2. Numerical values for (4.8)
$S = [0,1]$, $\theta_1 \in [0.5, 1.5]$, $r_{\mathbf{x}_0,S} = 1$, and $x_0 = 1.98$.

ρ	a_1	a_2	a_3	a_5	θ_1^{LF}
0.5	0.049	0.001	−0.783	1.311	1.5
1	−0.142	0.002	−0.855	1.349	1.5
2	−0.318	0.00003	−0.897	1.417	1.5

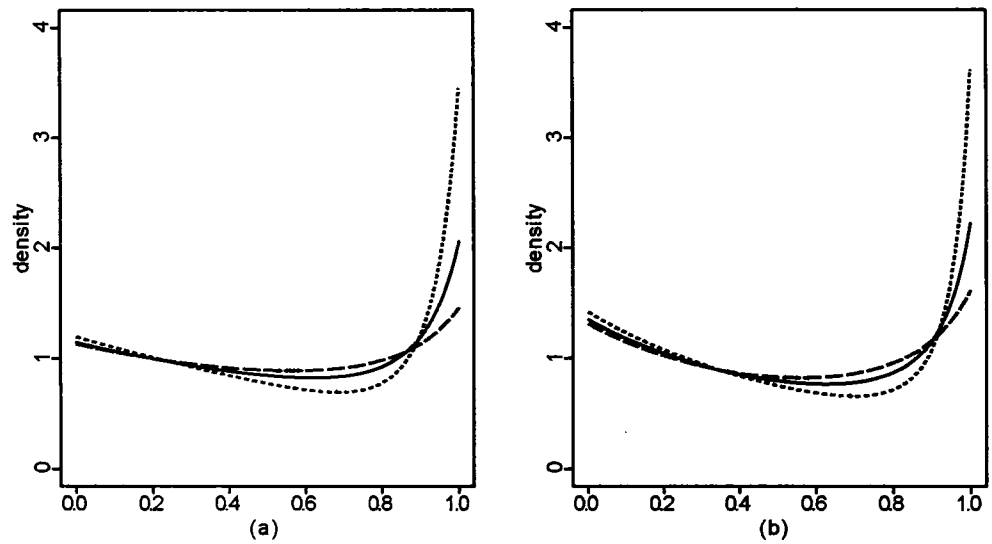

Figure 17: Optimal design densities $k(x) = \left[\frac{a_1 x + a_2}{a_3 x + a_4} + \frac{a_5}{e^{2\theta_1 x}(a_3 x + a_4)^2}\right]^+$ for $S = [0,1]$, $r_{\mathbf{x}_0,S} = 1$, and $x_0 = 1.98$. (a) locally optimal design densities for $\theta_1 = 0.987$; (b) locally most robust design densities for θ_1 within $[0.5, 1.5]$. Each plots uses three values of $\rho : \rho = 1$ (solid line), $\rho = 0.5$ (broken line), $\rho = 2$ (dotted line).

3.

Appendix: Derivations

Proof of Theorem 4.1. Since the term $\sigma^2 tr(\mathbf{A}_S \mathbf{B}^{-1})$ in I_1 does not involve f, the maximization problem becomes to maximize $\mathbf{b}_{f,S}^T \mathbf{H}_S^{-1} \mathbf{b}_{f,S} + \int_S f^2(\mathbf{x}) d\mathbf{x}$ over f. We obtain

$$\max_f \left(\mathbf{b}_{f,S}^T \mathbf{H}_S^{-1} \mathbf{b}_{f,S}\right) = \eta_S^2 \lambda_k^{(1)},$$

attained at $f_k(\mathbf{x}) = \eta_S \, \bar{\mathbf{z}}^T(\mathbf{x})\{k(\mathbf{x})a(\mathbf{x})\mathbf{I} - \mathbf{A}^{-1}\mathbf{B}\}\mathbf{c}_1$ in a way akin to that used in

Theorem 1 of Wiens (1992), and therefore the derivations are omitted here. Theorem 4.1 follows immediately from $\int_S f_k^2(\mathbf{x})d\mathbf{x} = \eta_S^2$. □

The result of Theorem 4.2 is obtained in a manner very similar to but simpler than that used in Theorem 3.1 of Chapter 3, and so its proof is omitted.

Theorem 4.3 follows Theorem 2.1 of Chapter 2 immediately.

Proof of Theorem 4.4. We look for a nonnegative function $k(\mathbf{x})$ minimizing (4.7) subject to $\int_S k(\mathbf{x})d\mathbf{x} = 1$. We introduce a Lagrange multiplier t. It is sufficient to show that $k(\mathbf{x})$ minimizes

$$(\sqrt{\lambda_k^{(2)}} + r_{\mathbf{x}_0,S})^2 + \nu\tilde{\mathbf{z}}^T(\mathbf{x}_0)\mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0) - 2t\int_S k(\mathbf{x})d\mathbf{x}$$

among all density functions. After some protracted calculation, we obtain the first order condition

$$\int_S \{P(\mathbf{x})k(\mathbf{x}) - Q(\mathbf{x}) - u\}(k - k_1)d\mathbf{x} \geq 0$$

for all densities k_1, where

$$P(\mathbf{x}) = \left[a(\mathbf{x})\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta}\right]^2 \text{ and } Q(\mathbf{x}) = \left[\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\gamma}\right]\left[a(\mathbf{x})\tilde{\mathbf{z}}^T(\mathbf{x})\boldsymbol{\beta}\right] \text{ with}$$

$$\begin{aligned}
\boldsymbol{\beta} &= \mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0), \\
\boldsymbol{\gamma} &= \left[\mathbf{B}^{-1}\mathbf{K} + \frac{\nu}{2}\left(1 + \frac{r_{\mathbf{x}_0,S}}{\sqrt{\lambda_k^{(2)}}}\right)^{-1}\mathbf{I}\right]\boldsymbol{\beta}, \\
u &= \left(1 + \frac{r_{\mathbf{x}_0,S}}{\sqrt{\lambda_k^{(2)}}}\right)^{-1} t.
\end{aligned}$$

The proof now can be completed in the same way as for Theorem 2.6 of Chapter 2. □

Proof of Theorem 4.5. The term $\nu tr(\mathbf{A}_T\mathbf{B}^{-1})$ in I_2 does not involve f, so the maximization problem becomes maximizing $\mathbf{b}_{f,S}^T\mathbf{H}_T\mathbf{b}_{f,S} - 2\mathbf{b}_{f,T}^T\mathbf{B}^{-1}\mathbf{b}_{f,S} + \int_T f^2(\mathbf{x})d\mathbf{x}$ over f. We obtain

$$\max_f \left(\mathbf{b}_{f,S}^T\mathbf{H}_T\mathbf{b}_{f,S} - 2\mathbf{b}_{f,T}^T\mathbf{B}^{-1}\mathbf{b}_{f,S}\right) = \eta_S^2\left[(\sqrt{\lambda_k^{(3)}} + r_{T,S})^2 - r_{T,S}^2\right],$$

attained at

$$f_k(\mathbf{x}) = \begin{cases} \eta_S \bar{\mathbf{z}}^T(\mathbf{x})[a(\mathbf{x})k(\mathbf{x})\mathbf{I} - \mathbf{A}_S^{-1}\mathbf{B}]\mathbf{c}_3, & \mathbf{x} \in S, \\ -\eta_T \bar{\mathbf{z}}^T(\mathbf{x})\mathbf{B}^{-1}\mathbf{G}\mathbf{c}_3/\sqrt{\lambda_k^{(3)}}, & \mathbf{x} \in T, \end{cases}$$

in a way essentially identical to that used in Theorem 2.1 (a) of Fang and Wiens (1999). Theorem 4.5 follows immediately from $\int_S f_k^2(\mathbf{x})d\mathbf{x} = \eta_S^2$. □

The proof of Theorem 4.6 is very similar to that of Theorem 3.1 in Chapter 3, so is omitted. □

Proof of Theorem 4.7. According to the results of Theorem 4.1, 4.3, and 4.5, we have got

$$\begin{aligned} \sup_{f \in \mathcal{F}_1} IB_1(f,k) &= \eta_S^2 \lambda_k^{(1)}, \\ \sup_{f \in \mathcal{F}_2} IB_2(f,k) &= \eta_S^2[(\sqrt{\lambda_k^{(2)}} + r_{\mathbf{x}_0,S})^2 - r_{\mathbf{x}_0,S}^2], \\ \sup_{f \in \mathcal{F}_3} IB_3(f,k) &= \eta_S^2[(\sqrt{\lambda_k^{(3)}} + r_{T,S})^2 - r_{T,S}^2]. \end{aligned}$$

For the unbiased designs, we then have

$$\begin{aligned} \sup_{f \in \mathcal{F}_i} IB_i(f,k) &= 0 \Leftrightarrow \lambda_k^{(i)} = 0 \Leftrightarrow \mathbf{G} = \mathbf{0} \\ &\Leftrightarrow \left[a(\mathbf{x})k(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}_S^{-1}\right]\mathbf{z}(\mathbf{x}) = 0 \text{ a. e.} \end{aligned}$$

for each $i = 1$, 2, or 3. We find that $a(\mathbf{x})k(\mathbf{x})$ is a constant almost everywhere on S in a manner essentially identical to that in the proof of Theorem 2.2 (b) in Fang and Wiens (1999). This fact together with $\int_S k(\mathbf{x})d\mathbf{x} = 1$, completes the proof of (a) and Part (b), (c) follow (a) immediately. □

References

Atkinson, A. C., and Haines, L.M. (1996). Designs for Nonlinear and Generalized Linear Models. In: *Handbook of Statistics*, eds. Ghosh, S. and Rao, C. R., Vol. 13, 437-475.

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

Chaloner, K. and Larntz, K. (1992). Bayesian design for accelerated life testing. *Journal of Statistical Planning and Inference*, 33, 245-259.

Clarke, F.H. (1983). *Optimization and Nonsmooth Analysis*. Wiley, New York.

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

Fang, Z. and Wiens, D. P. (1999). Robust extrapolation designs and weights for biased regression models with heteroscedastic errors. *The Canadian Journal of Statistics*, 27, 751-770.

Ford, I., Titterington, D. M., and Kitsos, C. P. (1989). Recent Advances in Nonlinear Experimental Design. *Technometrics*, 31, 49-60.

Ginebra, J. and Sen, A. (1998). Minimax Approach to Accelerated Life Tests. *IEEE Transactions on Reliability*, 47, 261-267.

Heo, G., Schmuland, B., and Wiens, D. P. (2001). Restricted Minimax Robust Designs for Misspecified Regression Models. *The Canadian Journal of Statistics*, 29, 117-128.

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

Lawless, J. F. (1984). Some problems concerning experimental designs for extrapolation. In: *Topics in statistics*, ed. Dwivedi, T. D., Montreal: Concordia University Press, 357-366

Nelson, W. (1990). *Accelerated testing - statistical models, test plans, and data analyses*. Wiley, New York.

Pascual, F. G. and Montepiedra, G. (2002). On Minimax Designs when There Are Two Candidate Models. *Journal of Statistical Computation and Simulation,* 72, 841-862.

Pascual, F. G. and Montepiedra, G. (2003). Model-Robust Test Plans with Applications in Accelerated Life Testing. *Technometrics*, 45, 47-57.

Shi, P., Ye, J., and Zhou, J. (2003). Minimax Robust Designs for Misspecified Regression Models. *The Canadian Journal of Statistics*, 31, 397-414.

Sinha, S. and Wiens, D. P. (2002). Robust Sequential Designs for Nonlinear Regression. *The Canadian Journal of Statistics*, 30, 601-618.

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

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 the American Statistical Association*, 93, 1440-1450.

CHAPTER V

APPLICATION AND IMPLEMENTATION

This chapter discusses the application of our designs, constructed in Chapter 2, to a real life dose-response experiment. We propose a couple of new implementation schemes which are utilized to approximate a robust design having a density, which in practice is not implementable. We also confirm that one of the matching quantile schemes used in the literature is optimal with respect to certain criteria. Several examples are demonstrated for using the implementation schemes proposed. In addition, a comparison between a proposed approach and an existing one is given.

5.1 Case Study

5.1.1 Introduction

In recent years, the public has become increasingly aware of the presence of harmful chemicals in our environment. Many people express concerns about pesticide residue or other foreign substances in food, contaminants in drinking water, and toxic pollutants in the air. How can we determine which of these potential hazards deserve our attention? Dose-response assessment is employed to tackle such problems. Dose-response assessment is the process of characterizing the relation between the dose of a chemical and the incidence of an adverse health effect in exposed populations, and of estimating the incidence of the effect as a function of human exposure to the chemical. 'Dose' is commonly used to indicate the amount of the chemical while 'response' refers to the effect of the chemical once administered. Generally, increasing the dose will result in an increase in the incidence of an adverse effect. Our intention is to estimate the effects of chemicals at low exposure levels using a dose-response

regression model. Krewski and Brown (1981) provide a guide to the literature for carcinogenic risk assessment. Crump (1979) reveals that since direct estimates of effect associated with very low levels of exposure often need very large numbers of experimental subjects, such estimates are inevitably based on the downward extrapolation of the results acquired at relatively high dose levels using a moderate number of subjects. Fang (1999) also points out that for the low-dose extrapolation problem it is important to keep an eye on possible violations of the assumed *linear* model in constructing the optimal design.

The experimenter takes a set of dose levels and runs an experiment with a prespecified number of subjects at each dose level. The number of subjects responding at each dose level is recorded. Let x be the dose level administrated to the experimental subjects, such as animals, and $P(x)$ be the probability of success in showing the response at dose level x. The dose-response curve illustrating the relation between x and $P(x)$ is fitted. This curve can then be employed to estimate the life risk of a subject exposed to a very low dose level, x_0. As indicated in Chapter 2, this is a one-point extrapolation problem. Since reducing sample sizes is important in an experiment of the life risk assessment, the optimal designs can be helpful for improving the efficiency. However, due to the estimation being extrapolated to x_0, any misspecification in a regression model adopted in obtaining the optimal design will endanger its optimality. We therefore take into account any possible departures from the model and seek an optimal robust design for such a one-point extrapolation problem.

5.1.2 Approximate Generalized Linear Regression Model

To demonstrate the application of a robust design constructed in Chapter 2 for the one-point extrapolation, we revisit the vinyl chloride data cited by Hoel and Jennrich (1979). These data (Table 5.1) are recorded from an experiment for the purpose

Table 5.1. Vinyl chloride data

Doses	Animals tested	Observed responses
0	58	0
50	59	1
250	59	4
500	59	7

of estimating the probability of developing cancer after an animal is exposed to a carcinogenic material - vinyl chloride. The dose unit for vinyl chloride is *ppm* (parts per million, 1 $ppm = 2.6\ mg/m^3$). The commonly used model by many researchers in the field of cancer dose response is a generalized linear function,

$$P(x) = 1 - \exp\left(-\sum_{j=0}^{k} \alpha_j x^j\right). \tag{5.1}$$

The maximum likelihood fit of $P(x)$ to the vinyl chloride data using (5.1) identified by Guess, Crump, and Peto (1977) is

$$P(x) = 1 - \exp(-\alpha x) \tag{5.2}$$

with $\alpha = 0.000267377$.

Hoel and Jennrich (1979) reconsidered this problem and presented the optimal design for the assumed model (5.2) and a target value of $\alpha = 0.000267377$ with $x_0 = 0.5$. However, there are no observations at any dose levels between 0 and 50 and only one observation between 50 and 250. Consequently, within $(0, 50)$ we cannot detect any possible departures from the assumed model. In addition, within $(0, 250)$, we can neither validate the model nor possibly obtain any information on how the model would depart from that which is assumed. Therefore, we suppose that (5.2) is only approximate for the vinyl chloride data, namely

$$P(x) \approx 1 - \exp(-\alpha x), \tag{5.3}$$

with an unknown contamination bounded in $\mathcal{L}_2$-norm.

5.1.3 Optimal Minimax Design for Low-dose Extrapolations

According to Theorem 2.6 in Chapter 2, the locally optimal robust design for (5.3) has the form of

$$k(x) = \left[\frac{a + bx^2e^{-2\alpha x}}{x^2e^{-2\alpha x}\left(1 + cx^2e^{-2\alpha x}\right)}\right]^+.$$

We consider the design space as the range of dose levels where subject toxic response is observed. For this case, $S = [50, 500]$. We take $r_{x_0,S} = 1$, which means that the amount of model uncertainty at the extrapolation point and that in the design space are regarded as equal. We also take $\nu = 1$, to indicate that we consider the importance of bias reduction and that of variance minimization as being the same. Our goal is to estimate the probability of observing a toxic response at dose level $x_0 = 0.5$. For $\alpha = 0.00053475$, the optimal minimax design ξ_k has a density of

$$k(x) = \frac{8.269 + 13.837x^2e^{-2\alpha x}}{x^2e^{-2\alpha x}\left(1 + 0.2669x^2e^{-2\alpha x}\right)}. \tag{5.4}$$

See Figure 18 for its plot. As indicated in Hoel and Jennrich (1979), the optimal design obtained is very robust against misspecification in the assumed parameter value within a moderate parameter region containing it, although locally optimal. The implementation of this design is provided in Section 5.2.2.

5.2 Implementations

5.2.1 Introduction

In the preceding section of the present chapter and in the previous three chapters, we have obtained a number of robust designs for various cases which turned out to have densities. Such designs with densities are prevalent in the literature. See, for instance, Huber (1975), and Wiens (1990, 1992). In addition, Heo (1998) listed quite a few such continuous designs. Wiens (1992) showed that if the contamination class is an L_2-type of space, e.g. Class $\mathcal{F}_1$ in Chapter 1, any optimal design minimizing the

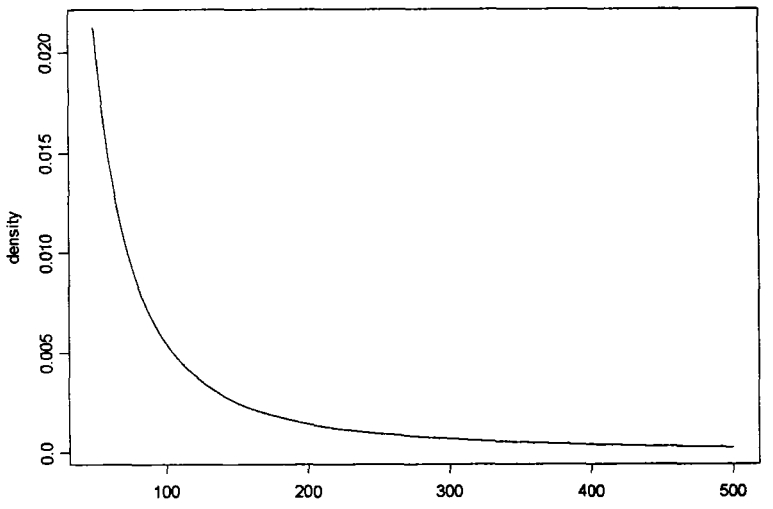

Figure 18: Optimal minimax design density for the dose-response example.

maximum loss function over such contamination space has to be absolutely continuous. These designs with densities are not implementable. In practice, a design should tell the experimenter what the design support points are and how many subjects should be allocated to each of these points. An implementable design, ξ, must be a discrete probability measure which puts probability $\xi(x_i)$ at x_i. For a given $r \leq n$, let $x_1, x_2, ..., x_r$ be the distinct design points, the number of subjects allocated to a particular design point x_i on design space is then, $n\xi(x_i)$. In addition, each $\xi(\mathbf{x}_i)$, $i = 1, 2, ..., r$, should be an integer multiple of $\frac{1}{n}$. A design with this integer property is called an exact design.

A few implementation approaches have been introduced in the literature. Wiens (1992) suggested a randomized design. That is, the design points are randomly chosen from an optimal design density. Wiens and Zhou (1996) presented a more systematic approach of matching quantiles, that places an equal number of observations at the quantiles of an optimal design density. This type of approach has been used in both

Heo (1998) and Fang (1999). Its properties are also discussed in Section 5.2.4 of this chapter. Matching moments is another method that chooses design points such that the empirical moments match up as closely as possible with the theoretical moments, obtained from the optimal density, to a sufficiently high order. This approach has been seen in Heo, Schmuland and Wiens (2001) and Adewale (2002).

An innovation for obtaining an exact design was presented by Fang and Wiens (2000), who redefined the problem by considering a finite design space and applying the simulated annealing algorithm. See also most recently in Adewale and Wiens (2006).

In the rest of this section, we propose two other types of approximation schemes: equally spaced design and histogram design. We also verify that one of the matching quantile approximation methods is optimal under the defined criteria.

5.2.2 Equally Spaced Design

An equally spaced design is referred to as a design whose support points are evenly dispersed within the design space. We recall the matching quantile approach that is based on uniform y-axis partitioning. Commonly, this leads to a design with non-equally spaced support points but with an equal number of observations at each support point. We now propose to partition the x-axis instead, obtaining an equally spaced design with an unequal (normally) number of observations on its design points. In many fields, particularly in life testing, such designs are traditionally applied. For instance, the best traditional designs, the Meeker-Hahn designs (Nelson, 1990), and the model-robust 4:2:1 design (Pascual and Montepiedra, 2003) are all equally spaced. Equally spaced designs are often preferable when adjustment for stress levels is not as simple as adjustment for the number of test subjects at each level. In reality, the experimenter will shift from a traditional to an optimal design provided there is a minimum change/cost in experimental equipment/material settings. Therein lies the

motivation for the equally spaced implementation scheme.

A given number (r) of design support points (stress/dose levels) are positioned equidistantly within the design space, while the design space itself is divided into r equal subspaces. On each subspace, to determine the proportion of subjects assigned to the design point within we use the integral of the density over this subspace. It is easy to show that such a design tends weakly to the optimal design as r tends to infinity. It should be noted that in order to get the exact design, this approximation usually needs to be rounded. For efficient rounding, we refer to Pukelsheim and Rieder (1992).

The following gives an implementation of the optimal design (5.4) using the equally spaced approach. We take $r = 10$. To obtain an equally spaced design over the design space of $[50, 500]$, the interval is divided into ten equal parts: $[50, 95)$, $[95, 140)$, ..., and $[455, 500]$. The ten dose levels can be chosen at 50, 100, ..., and 500. We still use the same number of test subjects as in Guess *et al. (op. cit.)*: $n = 235$. The allocation of these test subjects can be calculated as

$$n_{50} = 235 \int_{50}^{95} \frac{8.269 + 13.837x^2e^{-2\alpha x}}{x^2e^{-2\alpha x}\left(1 + 0.2669x^2e^{-2\alpha x}\right)}dx \approx 119,$$

$$n_{100} = 235 \int_{95}^{140} \frac{8.269 + 13.837x^2e^{-2\alpha x}}{x^2e^{-2\alpha x}\left(1 + 0.2669x^2e^{-2\alpha x}\right)}dx \approx 44, ...,$$

and

$$n_{500} = 235 \int_{455}^{500} \frac{8.269 + 13.837x^2e^{-2\alpha x}}{x^2e^{-2\alpha x}\left(1 + 0.2669x^2e^{-2\alpha x}\right)}dx \approx 3.$$

Then, when $r = 10$, the resulting design requires 119 animals at a dose level of 50, 44 at 95, ..., and 3 at 500, namely as follows:

dose levels	50	100	150	200	250	300	350	400	450	500
allocation	119	44	23	14	10	7	6	5	4	3.

5.2.3 Histogram Design

We propose using a histogram type of approach to approximate an optimal design having a density. This approach is inspired by Professor Douglas Wiens' seminar

presentation at the University of Alberta in November, 2004. We have noticed that compared with classical optimal designs (minimizing variance alone), optimal robust designs against model misspecifications (minimizing mean square error) are obviously more uniform with mass spread over the design space, yet heavily loaded near those design points obtained for minimizing variance alone. Such a pattern has been found in, for example, the robust designs constructed by Fang and Wiens (1999, 2000), and those achieved in the previous chapters by us. We decompose the design space into a number of subspaces (the bases of the bars in a histogram). Then, we assign the relative frequency (the height of a histogram bar) on each subspace to be the average density over this subspace. We note that such a design tends weakly to the optimal design as the number of subspaces tends to infinity.

In addition, the density of an optimal design is often much higher in a neighborhood of each of the points, than elsewhere; these design points would have been used for minimizing variance alone. Consequently, we take the number of support points for the classical minimizing variance design as the number of subspaces with nonzero relative frequencies in the histogram design.

5.2.3.1 Examples using histogram designs

In this subsection, we focus on the implementation of the robust optimal designs constructed in Example 2, Chapter 2. This example presents the resulting optimal design densities for approximate pth degree polynomial regression models in the case of one-point extrapolation. For $p = 1$ (linear), the optimal design densities are exhibited in Figure 6 in Chapter 2, which indicates that there are only two subintervals in the design space from which we may obtain observations for each of those optimal designs. Figure 7 in Chapter 2 gives the optimal design densities for $p = 2$ (quadratic). This figure indicates that there are only three subintervals over which observations can be made for each optimal design. The remaining of design intervals have densities

Table 5.2. Histogram designs for Example 2 of Chapter 2 with $p = 1$.

x_0	ν	S_1	h_1	$A_1(\%)$	S_2	h_2	$A_2(\%)$
1.5	0.5	$[-1, -0.401]$	0.480	28.8	$[0.044, 1]$	0.745	71.2
	1	$[-1, -0.488]$	0.510	26.2	$[0.173, 1]$	0.893	73.8
	10	$[-1, -0.675]$	0.661	21.5	$[0.451, 1]$	1.43	78.5
5	0.5	$[-1, -0.339]$	0.641	42.3	$[0.249, 1]$	0.769	57.7
	1	$[-1, -0.419]$	0.717	41.7	$[0.341, 1]$	0.884	58.3
	10	$[-1, -0.582]$	0.986	41.0	$[0.511, 1]$	1.21	59.0

of zero. To determine the subintervals of these designs, we solve

$$\frac{(a_1x + a_2)(a_3x + 1) + a_5}{(a_3x + 1)^2 + a_6(a_3x + 1)^4} = 0$$

for x when $p = 1$, where the values of $a_1 - a_3$, a_5, and a_6 for diverse ν are listed in Table 2.2 of Chapter 2. We solve

$$\frac{(1 + a_1x + a_2x^2)(b_0 + b_1x + b_2x^2) + c}{(1 + a_1x + a_2x^2)^2 + d(1 + a_1x + a_2x^2)^4} = 0$$

for x when $p = 2$, where the values of a_1, a_2, b_0, b_1, b_2, c, and d for various ν are given in Table 2.3 of Chapter 2.

The histogram designs are presented in Table 5.2 for $p = 1$ and in Table 5.3 for $p = 2$. In these tables, S_i and h_i denote the base and the height of the histogram bar for each subinterval respectively. A_i stands for the portion of subjects assigned to the corresponding subinterval. A result of such approximation using the histogram approach is exhibited in Figure 19. It should be noted that this scheme might not be so obvious for finding the appropriate subspaces without a 'natural' divider, as is the case in these examples. Nevertheless, it may serve as a complement to the existing approaches when the pattern of an optimal design permits.

5.2.3.2 Obtaining exact designs

These histogram designs are still not exact designs that the experimenter can manipulate in practice. We use each of the portions obtained in Tables 5.2 and 5.3 to

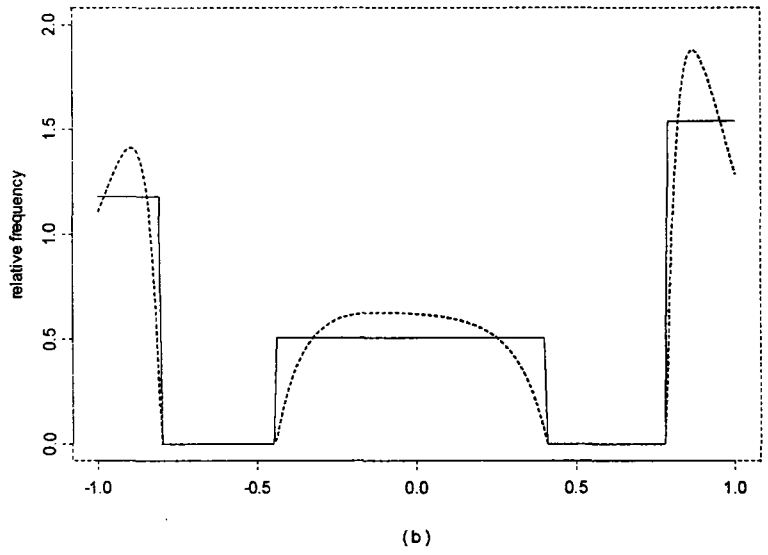

Figure 19: Histogram approximation for Example 2 of Chapter 2 with $p = 2$.

Table 5.3. Histogram designs for Example 2 of Chapter 2 with $p = 2$.

x_0	ν	S_1	h_1	$A_1(\%)$	S_2	h_2	$A_2(\%)$	S_3	h_3	$A_3(\%)$
1.5	0.5	$[-1,-0.80]$	0.71	14.3	$[-0.50,0.34]$	0.43	36.0	$[0.69,1]$	1.61	49.7
	1	$[-1,-0.83]$	0.74	13.5	$[-0.47,0.32]$	0.43	35.2	$[0.74,1]$	1.93	51.3
	10	$[-1,-0.88]$	1.01	12.3	$[-0.41,0.31]$	0.41	29.2	$[0.84,1]$	3.59	58.5
5	0.5	$[-1,-0.78]$	1.03	23.0	$[-0.47,0.43]$	0.49	44.3	$[0.75,1]$	1.32	32.7
	1	$[-1,-0.80]$	1.18	23.1	$[-0.44,0.41]$	0.51	43.4	$[0.78,1]$	1.54	33.5
	10	$[-1,-0.86]$	1.72	23.5	$[-0.40,0.38]$	0.53	41.0	$[0.85,1]$	2.44	35.5

assign experimental subjects uniformly over its corresponding subinterval. The design points within each subinterval, $S_i = [a_i, b_i]$, can be assigned as

$$x_j = a_i + \frac{b_i - a_i}{2n_i}(2j - 1)$$

for $j = 1, 2, ..., n_i$ with n_i being a rounded integer of nA_i. For a demonstration with $n = 10$, the design points are displayed in Figure 20 for $p = 1$ and Figure 21 for $p = 2$.

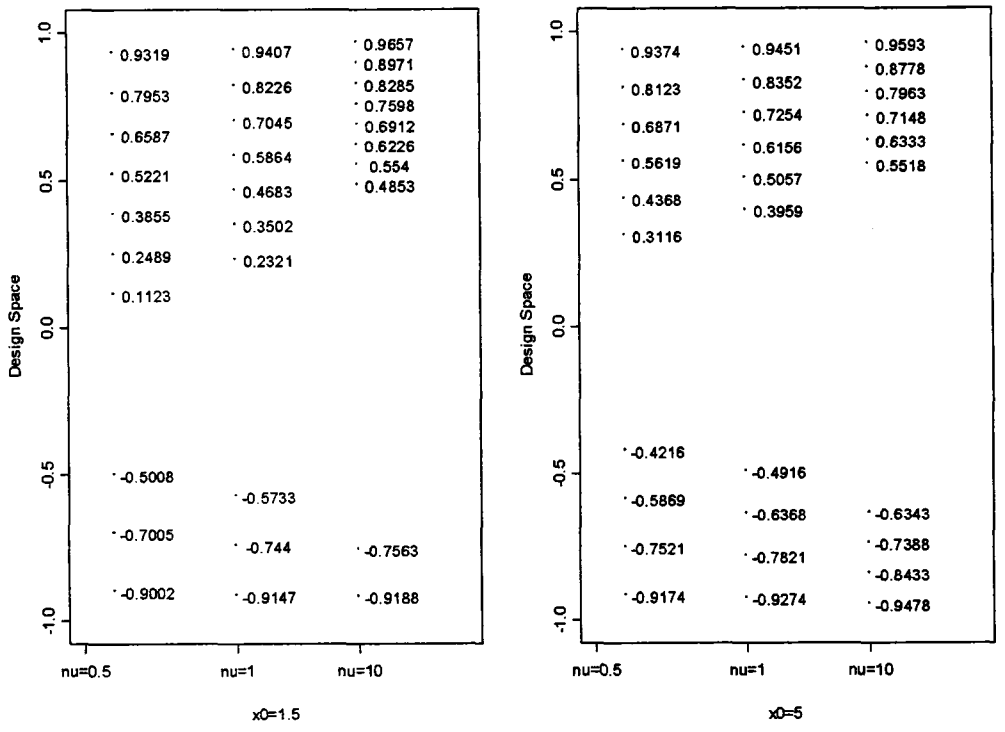

Figure 20: Exact designs using histogram scheme for Example 2 of Chapter 2 with $p = 1$.

5.2.4 Optimal Approximation

There are two kinds of matching quantile approaches appearing in the literature. In the first, used in Wiens and Zhou (1996), one observation is placed at each of the following quantiles:

$$x_i^{(1)} = F_{\xi_0}^{-1}(\frac{i-0.5}{n}),\ i = 1,\ 2, ...,\ n, \tag{5.5}$$

where ξ_0 is an optimal design with a density, and F_{ξ_0} is the cumulative distribution function of ξ_0. In the second, used in Heo *et al. (op. cit.)*, one observation is placed at

$$x_i^{(2)} = F_{\xi_0}^{-1}(\frac{i-1}{n-1}), i = 1, 2, ..., n.$$

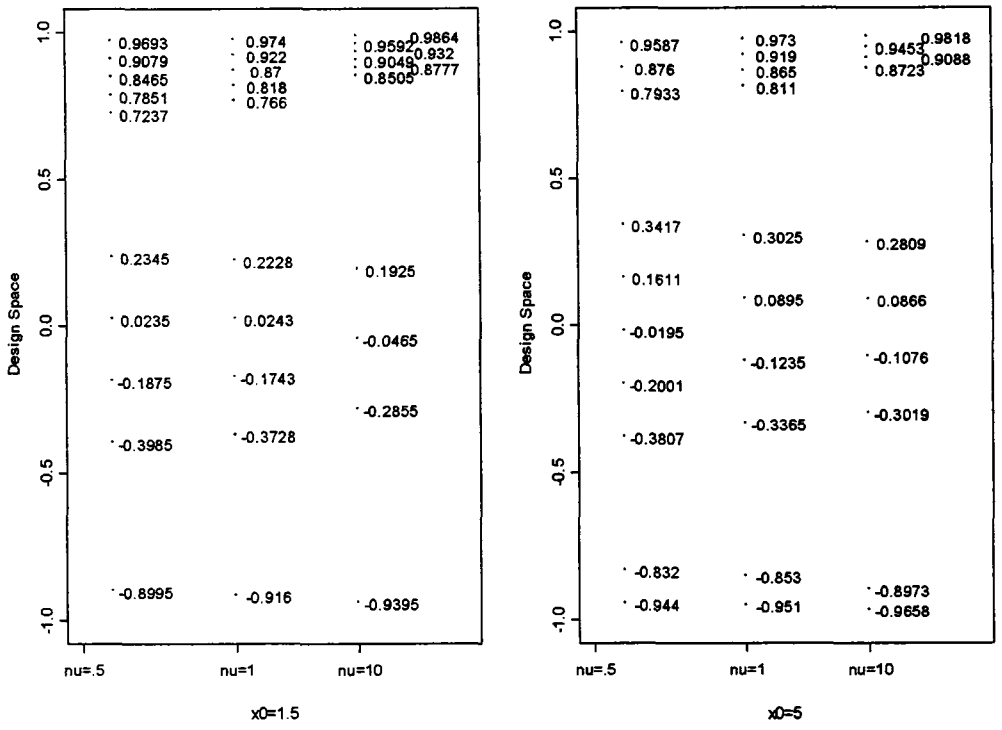

Figure 21: Exact designs using histogram scheme for Example 2 of Chapter 2 with $p = 2$.

The first approach provides a sample from ξ_0 with a smaller Kolmogorov-Smirnov statistic, since

$$\max_{1\le i\le n}\left|F_{\xi_0}\left(x_i^{(1)}\right)-\frac{t_i}{n}\right|=\frac{1}{2n},$$

and

$$\max_{1\le i\le n}\left|F_{\xi_0}\left(x_i^{(2)}\right)-\frac{t_i}{n}\right|=\frac{1}{n},$$

where t_i is the number of design points less than x_i. For a fixed number of experimental subjects n, the following theorem shows that (5.5) is optimal under certain criteria.

Theorem 5.1 *The first matching quantile approach provided by (5.5) offers an optimal approximation which minimizes the overall "distance" between the optimal design*

ξ_0 *and the resulting one when this distance is defined as*

$$D = \int_S \left| F_{\xi_0}(x) - F_\xi(x) \right|^m dx,$$

for any positive m*, among all the designs of the form of* $\xi(x) = \frac{1}{n}\sum_{i=1}^{n} I_{x_i}(x)$*, where* I_x *is an indicator function with pointmass* 1 *at* x*.*

To illustrate this, we put the x_i in increasing order, i.e. $x_1 < x_2 < ... < x_n$. Let $S = [a, b]$, we then have

$$\begin{aligned} D &= \int_a^{x_1} \left| F_{\xi_0}(x) - F_\xi(x) \right|^m dx + \int_{x_n}^{b} \left| F_{\xi_0}(x) - F_\xi(x) \right|^m dx \\ &\quad + \sum_{i=1}^{n-1} \int_{x_i}^{x_{i+1}} \left| F_{\xi_0}(x) - F_\xi(x) \right|^m dx. \end{aligned}$$

To minimize D, by taking the derivative with respect to each x_i, we find that the minimizing x_i is a solution of the following equation

$$\left| F_{\xi_0}(x) - \frac{i}{n} \right|^m = \left| F_{\xi_0}(x) - \frac{i+1}{n} \right|^m, \tag{5.6}$$

for $i = 1, 2, ..., n$. Theorem 5.1 follows because (5.5) satisfies (5.6) for $m > 0$.

The matching quantile methods are generally problematic for the case that the design space has a higher dimension, so other methods still may be viable alternatives. Figures 22 and 23 provide the resulting exact design points using (5.5) for Example 2 of Chapter 2. For this example, the exact designs obtained from the histogram approximation and those by using (5.5) are very close to each other even with n as small as 10. A comparison is demonstrated in Figure 24.

References

Adewale, A. (2002). *Robust Designs for Approximate Regression Models with Two Interacting Regressors.* Unpublished Master's thesis, University of Alberta.

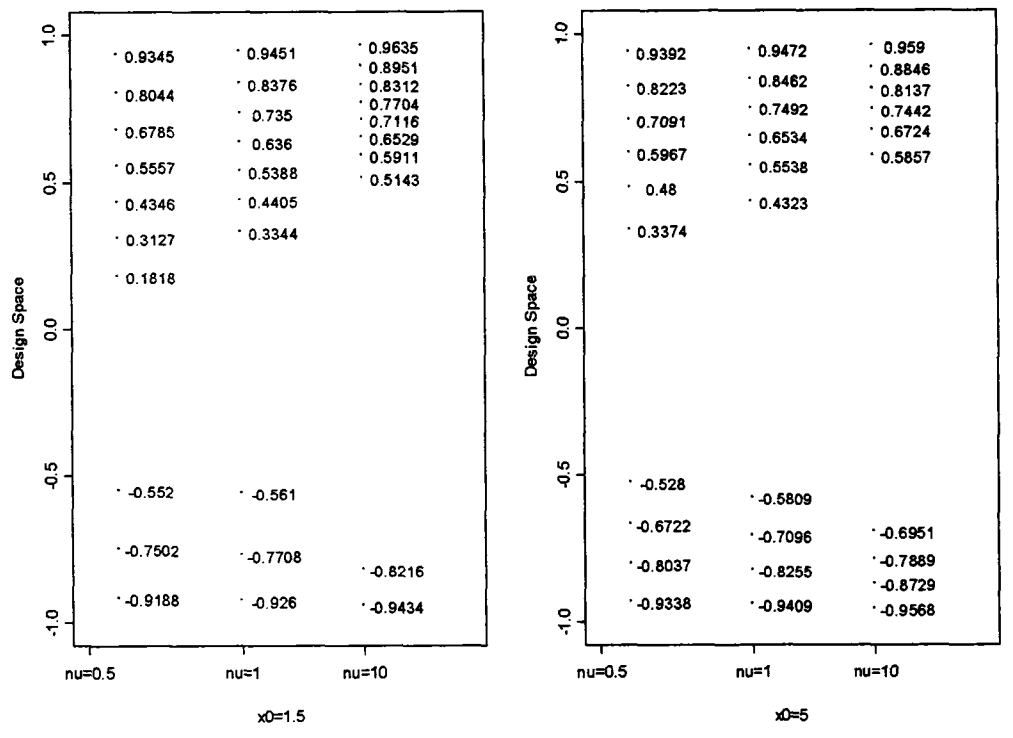

Figure 22: Exact designs using (5.5) for Example 2 of Chapter 2 with $p = 1$.

Adewale, A., and Wiens (2006). New Criteria for Robust Integer-Valued Designs in Linear Models. *Computational Statistics and Data Analysis*, to appear.

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

Fang, Z. (1999). *Robust Extrapolation Designs for Linear Models.* Unpublished Doctoral thesis, University of Alberta.

Fang, Z., and Wiens, D. P. (1999). Robust extrapolation designs and weights for biased regression models with heteroscedastic errors. *The Canadian Journal of Statistics*, 27, 751-770.

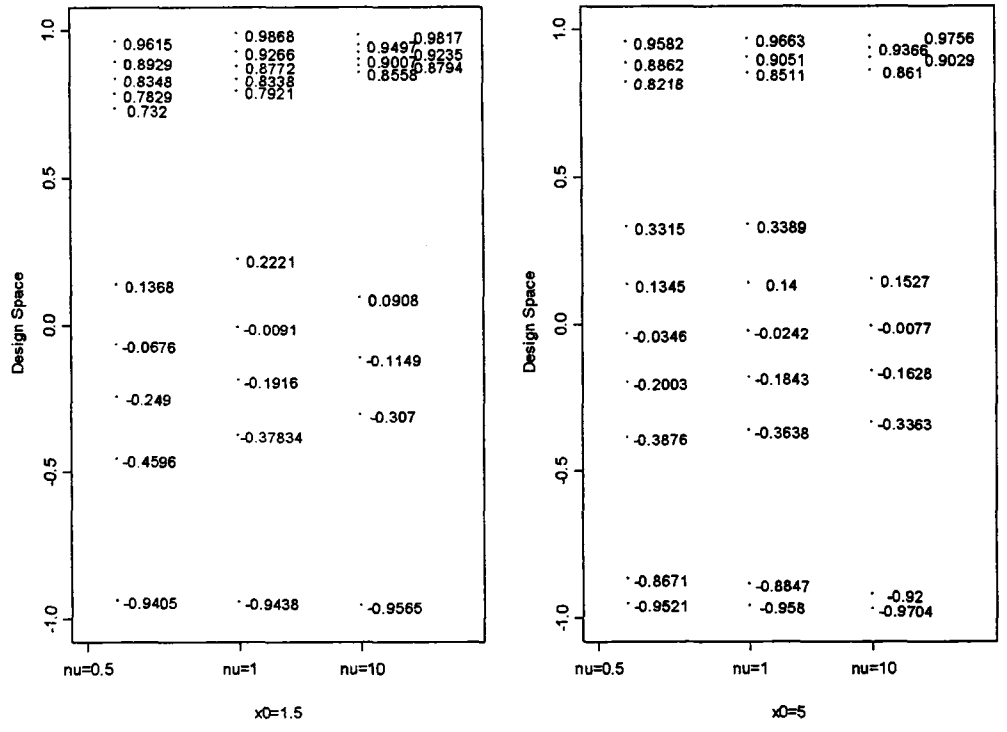

Figure 23: Exact designs using (5.5) for Example 2 of Chapter 2 with $p = 2$.

Fang, Z., and Wiens, D. P. (2000). Integer-Valued, Minimax Robust Designs for Estimation and Extrapolation in Heteroscedastic, Approximately Linear Models. *Journal of the American Statistical Association*, 95, 807-818.

Guess, H., Crump, K., and Peto, R. (1977). Uncertainty estimates for low dose rate extrapolation of animal carcinogenicity data. *Cancer Research*, 37, 3475-83.

Heo, G. (1998). *Optimal Designs for Approximately Polynomial Regression Models.* Unpublished Doctoral thesis, University of Alberta.

Heo, G., Schmuland, B., and Wiens, D.P. (2001). Restricted Minimax Robust Designs for Misspecified Regression Models. *The Canadian Journal of Statistics*, 29, 117-128.

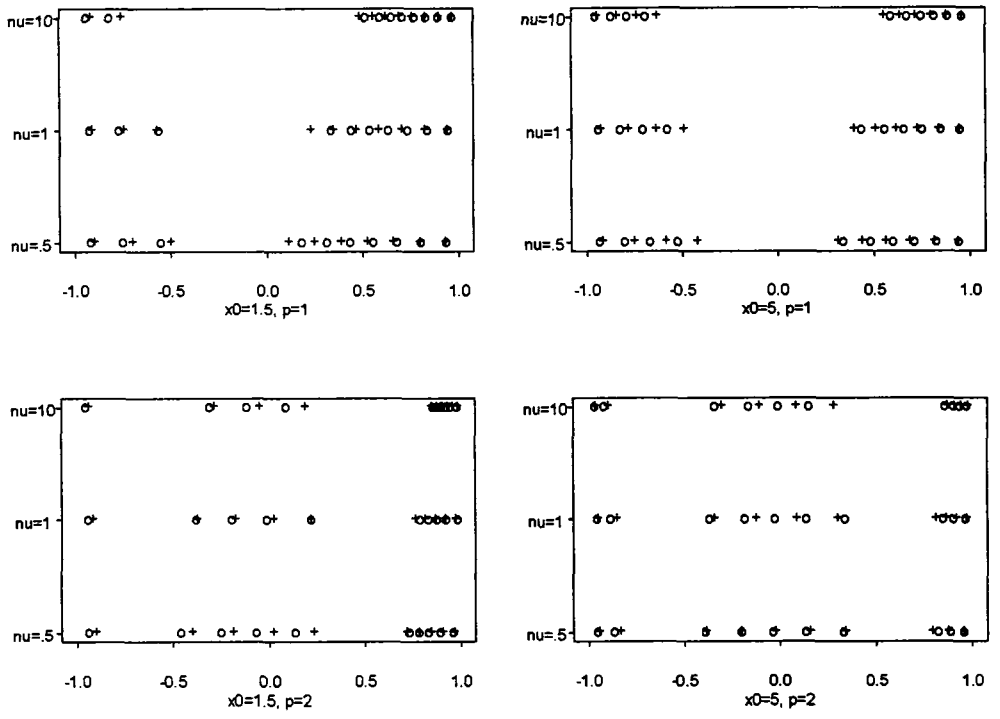

Figure 24: Compare the exact designs using both approaches for Example 2 of Chapter 2 with $\nu = 0.5$, 1, and 10. (a) $x_0 = 1.5$ when $p = 1$, (b) $x_0 = 5$ when $p = 1$, (c) $x_0 = 1.5$ when $p = 1$, and (d) $x_0 = 5$ when $p = 2$. Each plot uses "+" for the design points of histogram designs and "o" for the design points obtained from (5.5).

Hoel, P. G., and Jennrich, R. I. (1979). Optimal designs for dose response experiments in cancer research. *Biometrika,* 66, 2, 307-316.

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

Krewski, D., and Brown, C. (1981). Carcinogenic risk assessment: a guide to the literature. *Biometrics,* 37, 353-366.

Nelson, W. (1990). *Accelerated testing - statistical models, test plans, and data analyses.* Wiley, New York.

Pascual, F. G. and Montepiedra, G. (2003). Model-Robust Test Plans with Applications in Accelerated Life Testing. *Technometrics*, 45, 47-57.

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

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

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

Wiens, D. P., and Zhou, J. (1996). Minimax Regression Designs for Approximately Linear Models With Autocorrelated Errors. *Journal of Statistical Planning and Inference*, 55, 95-106.