UNIVERSITY OF ALBERTA

Robust Experimental Designs for Linear Models

BY

Julie Zhou ©

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

in

STATISTICS

DEPARTMENT OF MATHEMATICAL SCIENCES

EDMONTON, ALBERTA

FALL 1995

ISBN 0-612-06318-6

UNIVERSITY OF ALBERTA

LIBRARY RELEASE FORM

NAME OF AUTHOR: **Julie Zhou**

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

DEGREE: **Doctor of Philosophy**

YEAR THIS DEGREE GRANTED: **1995**

Julie Zhou

Department of Mathematical Sciences

Lakehead University

Thunder Bay, Ontario P7B 5E1

CANADA

Date: August 1, 1995

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

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

D. P. WIENS (Supervisor)

P. M. HOOPER

E. GOMBAY

J. D. WILSON

A. HERZBERG (Queen's University)

Date: 24 July 1995

ABSTRACT

This thesis concerns the construction of robust designs for linear and approximately linear models with correlated errors. These designs are robust against small departures from both the assumed linear regression response and the usual assumption of uncorrelated errors. A minimax approach and an infinitesimal approach are used to derive these designs. The use of the infinitesimal approach is new in robust design theory.

Using the minimax approach, we discover that a design which is asymptotically (minimax) optimal for uncorrelated errors retains its optimality under autocorrelation if the design points are a random sample, or a random permutation, of points from this distribution. Furthermore, minimax designs are obtained for approximately linear models when the errors follow a first order autoregressive process. The results suggest that the design points must follow certain patterns which depend on the sign of the autocorrelation parameter.

For the infinitesimal approach, we define the *Change-of-Bias Function* (CBF) and the *Change-of-Variance Function* (CVF) to be the Gateaux derivatives of the mean squared error of the estimated response, in the direction of a contaminating response function and in the direction of autocorrelation structure respectively. Then robust designs minimize the mean squared error of the estimator at the ideal model, subject to a robustness constraint formulated in term of CBF and/or CVF. Specific examples are considered, and new robust designs are developed.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Douglas P. Wiens, with my deepest appreciation, for his invaluable guidance and friendly support.

I also thank Dr. P. M. Hooper, Dr. E. Gombay, Dr. J.D. Wilson and Dr. A. Herzberg for their careful consideration and helpful suggestions.

I thank the Department of Statistics and Applied Probability for providing a congenial research environment. I thank all Professors who have helped me during my study at the University of Alberta. I am also grateful to Dr. Douglas P. Wiens and the University of Alberta for their financial support.

Table of Contents

Chapter 1. Introduction 1

1. Classical design problem 2
2. Examples of classical optimal designs 4
3. Why do we need robust designs? 6
4. Robust designs for approximately linear models 8
5. Designs robust against autocorrelation of errors 13
6. Robust designs derived in this thesis 17
7. Application 19
8. Conclusion 24

References 25

Chapter 2. Minimax Regression Designs for Approximately Linear Models with Autocorrelated Errors 28

1. Introduction 28
2. Derivations 32
3. Examples 41

References 45

Chapter 3. Minimax Designs for Approximately Linear Models with AR(1) Errors 49

1. Introduction 49
2. Asymptotics 53

3. Minimax designs 56

4. Appendix: Proof of Theorem 3.1 61

References 63

Chapter 4. Robust Designs Based on the Infinitesimal Approach 66

1. Introduction 66

2. The CVF and CBF for designs 69

2.1 Change-of-Variance Function 71

2.2 Change-of-Bias Function 72

3. Optimality Criteria 73

3.1 V-robustness 73

3.2 B-robustness 74

3.3 M-robustness 74

4. V-robust designs 74

5. B-robust designs 85

6. M-robust designs 88

References 90

Chapter One

Introduction

This dissertation concerns the construction of robust regression designs for linear and approximately linear models with correlated errors. The work is presented in three papers which have been prepared for publication. From Chapter Two to Chapter Four, each chapter is an independent paper. This chapter gives a brief review of robust designs in the literature and summarizes the entire thesis.

We begin with a short introduction to classical regression design problems in Section 1. This introduction serves as a motivation which helps readers understand the robust design problems presented in this dissertation. For a more detailed presentation of classical regression designs we refer to, for example, Pukelsheim (1993), Fedorov (1972) or Box (1987). Following this introduction, two examples of classical optimal designs are presented in Section 2, which leads to a systematic discussion of the necessity of studying robust designs in Section 3. In the literature, two types of robust designs (designs robust against small departures from the assumed regression response and designs robust against autocorrelation of errors) have been investigated. After reviewing these designs in Section 4 and 5, we summarize the robust designs studied in this thesis in Section 6. Examples of practical situations in which our robust designs can be applied are given in Section 7. Finally, further research topics in this area are discussed.

1 Classical design problem

For simplicity, a single variable design problem is considered first. Suppose there are two related variables x and y; x is an explanatory variable, y is a response variable, and

$$E[y|x] = \theta_0 + \theta_1 x, \quad x \in [a, b],$$

where θ_0 and θ_1 are unknown parameters. In practice, we often want to estimate θ_0 and θ_1 so that we can predict the y value at a certain x value or make some other inferences about y or $E[y|x]$.

Estimates of θ_0 and θ_1 are usually computed from observations on y obtained through experimentation. Suppose that $x_1, x_2, \cdots, x_n$ are the n points at which we observe $y_1, y_2, \cdots, y_n$. Because of the experimental errors, the observations (x_1, y_1), $(x_2, y_2), \cdots, (x_n, y_n)$ follow a statistical linear model,

$$y_i = \theta_0 + \theta_1 x_i + \epsilon_i, \quad i = 1, \cdots, n, \tag{1.1}$$

where ϵ_i are the experimental errors with mean 0 and variance σ^2.

Let $\hat{\boldsymbol{\theta}}_{LS} = (\hat{\theta}_0, \hat{\theta}_1)^T$ be the Least Squares estimate in (1.1):

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

where

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

How efficient is $\hat{\boldsymbol{\theta}}_{LS}$? From regression analysis, if the errors are uncorrelated, then

$$E[\hat{\boldsymbol{\theta}}_{LS}] = \boldsymbol{\theta},$$

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

To appraise the efficiency of the design, the right measurement is the covariance matrix of $\hat{\boldsymbol{\theta}}_{LS}$. The "smaller" the covariance matrix is, the more efficient the estimate. We notice that the covariance matrix only depends on X. So the choice of the x_i's determines the efficiency of $\hat{\boldsymbol{\theta}}_{LS}$. Therefore, in this case, the design problem is to choose an "optimal" set of x_i's before the experiment in such a way that we have the most efficient estimate (i.e. $COV[\hat{\boldsymbol{\theta}}_{LS}]$ is minimized). The x_i's are called the design points.

In general, the classical design problem with multiple variables can be summarized as follows. We consider the regression model:

$$y_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\theta} + \epsilon_i, \qquad i = 1, \cdots, n; \tag{1.2}$$

$$\mathbf{x}_i \in \mathcal{S} \subset \mathcal{R}^q,\ \mathbf{z}(\mathbf{x}_i) \in \mathcal{R}^p,\ \boldsymbol{\theta} \in \mathcal{R}^p;$$

$$E[\epsilon_i] = 0,\ \ Var[\epsilon_i] = \sigma^2,$$

where $\mathcal{S}$ is a given design space and $\mathbf{z}(\mathbf{x})$ is a given function of $\mathbf{x}$. For instance, if $\mathbf{z}(\mathbf{x}) = (1, \mathbf{x}^T)^T$, then (1.2) is the usual multiple linear regression model; if $x \in \mathcal{R}^1$ and $\mathbf{z}(x) = (1, x, x^2, \cdots, x^{p-1})^T$, then (1.2) is the polynomial regression model.

Furthermore, two classical assumptions are made for model (1.2):

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

A2: The errors ϵ_i are uncorrelated.

Based on A1 and A2, the classical design problem is

$$\min_{\mathbf{x}_1, \cdots, \mathbf{x}_n} \mathcal{L}\left(COV(\hat{\boldsymbol{\theta}})\right) \tag{1.3}$$
$$s.t. \;\; \mathbf{x}_i \in \mathcal{S}, \quad i = 1, \cdots, n,$$

where $\hat{\boldsymbol{\theta}}$ is an unbiased estimate (not necessarily the LS estimate) in (1.2) and $\mathcal{L}$ is a scalar function (loss function). The function $\mathcal{L}$ can be the determinant, the trace or some other function of positive semidefinite matrices, monotonic increasing with respect to the usual ordering by positive definiteness. This design problem has been studied extensively in the literature, and many optimal designs are obtained for various linear models and loss functions. See Fedorov (1972) and Pukelsheim (1993).

2 Examples of classical optimal designs

Example 2.1. Let us consider the simple linear regression model

$$y_i = \theta_0 + \theta_1 x_i + \epsilon_i, \quad x_i \in \mathcal{S}, \quad i = 1, \cdots, n.$$

Without loss of generality, the design space $\mathcal{S}$ is assumed to be $[-0.5, 0.5]$. We use the Least Squares method to estimate θ_0 and θ_1. Let $\mathcal{L} = Determinant$, so that the optimal design is called the D-optimal design. Suppose ξ is the empirical distribution function of $x_1, \cdots, x_n$ and that n is even. Solving (1.3), one gets the D-optimal design

$$\xi_*(-0.5) = \xi_*(0.5) = \frac{1}{2}.$$

The answer means that half of the optimal design points have to be placed at -0.5 and another half at 0.5.

Any implementable and discrete optimal design for finitely many design points is called an exact design, such as ξ_* in Example 2.1. However, sometimes continuous optimal designs are constructed. These designs must be approximated by discrete designs in practice.

Example 2.2. We consider the linear model with two explanatory variables:

$$y_i = \theta_0 + \theta_1 x_{1,i} + \theta_2 x_{2,i} - \epsilon_i, \quad i = 1, \cdots, n.$$

$$(x_{1,i}, x_{2,i}) \in \mathcal{S} = [-0.5, 0.5] \times [-0.5, 0.5].$$

Suppose that ξ is the empirical distribution function of $(x_{1,1}, x_{2,1}), \cdots, (x_{1,n}, x_{2,n})$, then one can show that the D-optimal design is

$$\xi_*(-0.5, -0.5) = \xi_*(-0.5, 0.5) = \xi_*(0.5, -0.5) = \xi_*(0.5, 0.5) = \frac{1}{4}.$$

This 2^2 factorial design is very useful in practice, see Montgomery (1991).

3 Why do we need robust designs?

We notice that the classical optimal designs in Example 2.1 and Example 2.2 place all mass at extreme points of the design spaces. In fact this is a general phenomenon of the classical optimal designs, which is problematic if the two assumptions A1 and A2 (on page 4) are not exactly true. If A1 and A2 hold, the optimal designs yield the most efficient Least Squares estimates consistent with the phenomenon. However if there are any violations of A1 and/or A2, some caution is called for. In many cases, three major reasons indicate that the classical optimal designs should not be used at all.

First we consider the case in which A1 is violated, i.e. the linear model (1.2) is not exactly correct. This is the model departure from linearity. In this situation, the Least Squares estimator is biased, and the classical optimal designs usually yield large bias (comparing with the variance). The classical optimal designs which minimize the variance alone are not optimal anymore because of the contribution of the bias. Box and Draper (1959) made apparent the dangers of designing a regression experiment under A1. They found that very small departures from A1 can eliminate any supposed gains arising from the use of a design which minimizes variance alone, see Wiens (1992). Huber (1981) points out that "deviations from linearity that are too small to be detected are already large enough to tip the balance away from the 'optimal' designs, which assume exact linearity and put observations on the extreme points of the observable range, toward the 'naive' ones which distribute the observations more or less evenly over the entire design space."

Also, when A1 is violated we cannot do model adequacy testing by using the classical optimal designs. Since we only make observations on the extreme points of $\mathcal{S}$, no information is available in the interior of $\mathcal{S}$. As in Example 2.1, optimal design requires observations only at two points: -0.5 and 0.5. Through observations at two points we can only fit a straight line. Any curvature in the design space $\mathcal{S}$ cannot be detected.

In the case in which A2 is violated, i.e. the errors are correlated, the model departure from uncorrelated errors occurs. In this situation, the exact correlation structure of the errors is usually unknown, so that $COV(\hat{\boldsymbol{\theta}})$ is unknown. The classical optimal designs which minimize $\mathcal{L}\left(COV(\hat{\boldsymbol{\theta}})\right)$ for uncorrelated errors do not minimize the true value of $\mathcal{L}\left(COV(\hat{\boldsymbol{\theta}})\right)$.

Therefore, there is a need to study optimal designs under possible small violations of A1 and/or A2. These designs are called robust designs. In general, robust designs are those which are not sensitive to small departures from model assumptions. More discussion of the need to robustify classical designs can be found in Wiens (1990, 1992, 1994).

In practice, the relationship between a response variable y and explanatory variables $\mathbf{x}$ is usually only approximately modeled. This often results in the violation of A1. The violation of A2 can be caused by, for example, serial or spatial correlation or repeated measures. With respect to different kinds of model violations (departures), different kinds of robust designs can be studied. In the next two sections, some robust designs in the literature are briefly reviewed. In Section 6, we outline the types of robust designs derived in this dissertation.

4 Robust designs for approximately linear models

When A1 fails, the linear model (1.2) is not exactly correct. This is the model departure from linearity. In this case, the approximately linear models are introduced,

$$y_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\theta} + f(\mathbf{x}_i) + \epsilon_i, \qquad i = 1, \cdots, n; \tag{4.1}$$

where f is an unknown disturbance function and belongs to a certain class $\mathcal{F}$, and ϵ_i are still uncorrelated. So departures from linearity are modeled in (4.1). Then a robust design for (4.1) is usually obtained by minimizing $\max_{f\in\mathcal{F}} \mathcal{L}(MSE(\sqrt{n}\hat{\boldsymbol{\theta}}))$, for some scalar loss function $\mathcal{L}$.

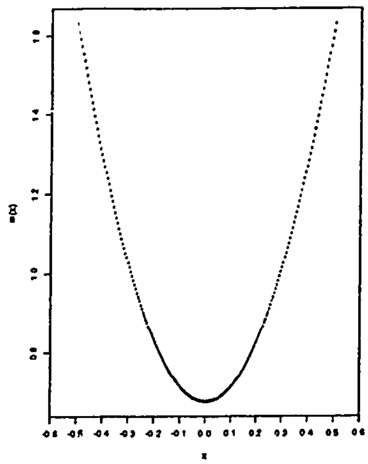
Huber (1975) and Wiens (1990, 1992, 1993, 1994) take

$$\mathcal{F} = \{\ f \mid ||f||_2 = \left(\int_S f^2(\mathbf{x})d\mathbf{x}\right)^{1/2} \leq \eta,\ \int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})d\mathbf{x} = 0\ \}, \tag{4.2}$$

where the radius η is assumed known and "small". The first condition in (4.2) says that the disturbance function f is small, so the linear term $\mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\theta}$ is still the leading term in (4.1). The second condition ensures the identifiability of $\boldsymbol{\theta}$ in the model as long as that $\int_S \mathbf{z}\mathbf{z}^T d\mathbf{x}$ is non-singular.

In Huber (1975), robust designs are obtained for the situation of $p = 1$, $\mathbf{z}(x) = (1, x)^T$, $\mathcal{S} = [-0.5, 0.5]$, and the loss function being the integrated mean squared

Figure 1: Optimal density function $m(x)$ when $\frac{\sigma^2}{n\eta^2} = 1$



error of $\hat{y}(x)$:

$$\begin{aligned}\mathcal{L}(MSE(\sqrt{n}\hat{\theta})) &= n \cdot \int_S E[(\hat{y}(x) - E[y|x])^2]dx \\ &= trace(MSE(\sqrt{n}\hat{\theta}) \cdot A), \quad with \quad A = \int_S \mathbf{z}(x)\mathbf{z}^T(x)dx.\end{aligned}$$

The optimal robust design has density function

$$m(x) = a(x^2 + b)^+, \quad a > 0,$$

where a and b depend on the ratio $\frac{\sigma^2}{n\eta^2}$. When $\frac{\sigma^2}{n\eta^2} \to 0$, $m(x) \to 1$—the uniform distribution on $[-0.5, 0.5]$. When $\frac{\sigma^2}{n\eta^2} \to \infty$, $M(x)$ (distribution function corresponding to $m(x)$) converges to the classical optimal design ξ_* in Example 2.1. Figure 1 is a plot for $m(x)$ when $\frac{\sigma^2}{n\eta^2} = 1$, $a = 3.810$ and $b = 0.178$.

This robust design and the classical optimal design in Example 2.1 are similar in a sense that both designs put heavy weight towards the boundary of $\mathcal{S}$. But the robust design puts some weight in the interior of the design space, which allows us to do model adequacy testing. So in practice, if we suspect any nonlinearity of the regression response, the robust design should be applied.

In Wiens (1990, 1992, 1993, 1994), robust designs are obtained for various estimators, various linear models and different loss functions. Wiens (1990) studies robust designs for multiple linear regression with $\mathbf{z}(\mathbf{x}) = (1, \mathbf{x}^T)^T$, $\mathbf{x} \in \mathcal{R}^p$ and $\mathbf{z}(\mathbf{x}) = (1,\ x_1,\ x_2,\ x_1 \cdot x_2)^T$ in (4.1), with $\mathcal{L}$ =integrated mean squared error. Wiens (1992) derives robust designs for (4.1) for various loss functions corresponding to the classical D-, A-, E-, Q- and G-optimality criteria. They are

1. $\mathcal{L}$ = Determinant;

2. $\mathcal{L}$ = Trace;

3. $\mathcal{L}$ = The largest characteristic root;

4. $\mathcal{L}$ = Integrated mean squared errors;

5. $\mathcal{L}$ = The maximum (over $\mathcal{S}$) MSE of $\hat{y}(\mathbf{x})$.

In Wiens (1993), robust designs are derived by maximizing the minimum coverage probability of confidence ellipsoids. Robust designs for M-estimators are considered in Wiens (1994).

Marcus and Sacks (1976), Sacks and Ylvisaker (1978), Pesotchinsky (1982), Li and Notz (1982), Li (1984) and Liu and Wiens (1994) take

$$\mathcal{F} = \{\ f \mid |f(\mathbf{x})| \leq \phi(\mathbf{x}),\ \ \forall\, \mathbf{x} \in \mathcal{S}\ \}, \tag{4.3}$$

with various assumptions being made about ϕ. The optimal designs constructed in these papers appear to be sensitive to the assumed form of ϕ.

In Pesotchinsky (1982), the following model is considered,

$$y_i = \theta_0 + \sum_{j=1}^{p} \theta_j x_{j,i} + f(\mathbf{x}_i) + \epsilon_i,$$

and

$$\mathcal{F} = \{ \ f \mid |f(\mathbf{x})| \leq \phi(\mathbf{x}), \ \ \mathbf{x} \in \mathcal{S} \ \},$$

where $\phi(\mathbf{x})$ is a convex function of $||\mathbf{x}||^2 = (x_1^2 + \cdots + x_p^2)$. The Least Squares estimate is used. The loss function is

$$\begin{aligned} \mathcal{L} = \Phi_k &= \left[\frac{1}{p+1} tr(MSE^k) \right]^{1/k} \\ &= \left\{ \frac{1}{p+1} \sum_{j=0}^{p} \lambda_j^k \right\}^{1/k}, \quad 0 < k < \infty, \end{aligned}$$

where $\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_p$ denote the eigenvalues of MSE. Then

$\Phi_\infty \propto \lambda_{max}\{MSE\}$ — E-optimality;

$\Phi_1 \propto$ Trace — A-optimality;

$\Phi_0 \propto$ Determinant — D-optimality.

Pesotchinsky limits his consideration to the class $\Xi(m)$ of all symmetric designs ξ with fixed $E_\xi(x_i^2) = m$. Then any symmetric design $\xi_0 \in \Xi(m)$ supported only by the points of sphere S_R of radius $R = \sqrt{mp}$ is D-optimal in $\Xi(m)$. Unlike the D-optimal design, A- and E-optimal symmetric designs are unique and correspond to the uniform continuous measures on appropriate spheres.

Li and Notz (1982) take $\mathcal{F}$ to be a subclass of measurable functions with respect to Lebesque measure on $\mathcal{S}$,

$$\mathcal{F} = \{ \ f \mid \ |f(\mathbf{x})| \leq c, \ \ c > 0, \ \ \forall \, \mathbf{x} \in \mathcal{S} \ \}$$

for the following approximately linear model,

$$y_i = \theta_0 + \boldsymbol{\theta}_1^T \mathbf{x}_i + f(\mathbf{x}_i) + \epsilon_i, \ \ \mathbf{x}_i \in \mathcal{S} \subset \mathcal{R}^k.$$

The optimality criterion is minimax weighted MSE:

$$\min_{\xi} \sup_{f \in \mathcal{F}} E[\beta_0^2(\hat{\theta}_0 - \theta_0)^2 + \sum_{i=1}^{k} \beta_i^2(\hat{\theta}_i - \theta_i)^2],$$

where the β_i are specified constant. If $\mathcal{S}$ is the $(k-1)$ dimensional simplex, i.e.

$$\mathcal{S} = \{ \ (x_1, \cdots, x_k)^T \in \mathcal{R}^k, \ \ \sum_{i=1}^{k} x_i = 1, \ \ x_i \geq 0 \ \ for \ \ all \ \ i \ \},$$

and $\beta_0 = 0, \ \ \beta_1 = \cdots = \beta_k = 1$, then the optimal design puts point mass $1/k$ on each of the k extreme points of $\mathcal{S}$, i.e. the points $\mathbf{x}(1) = (1, 0, \cdots, 0)^T, \ \ \mathbf{x}(2) = (0, 1, 0, \cdots, 0)^T, \ \ \cdots, \ \ \mathbf{x}(k) = (0, \cdots, 0, 1)^T$. If $\mathcal{S} = [-1, 1]^k$, then the optimal design puts point mass $1/2^k$ on each of the 2^k corners of the cube $\mathcal{S}$. This kind of robust designs still has support on the boundary of $\mathcal{S}$.

The class $\mathcal{F}$ in (4.2) is adopted for approximately linear models in this thesis, since the motivation of (4.2) is very clear. Suppose that $E[y|\mathbf{x}]$ is an approximately linear function of $\boldsymbol{\theta}_0$:

$$E[y|\mathbf{x}] \approx \mathbf{z}^T(\mathbf{x}) \boldsymbol{\theta}_0,$$

where $\boldsymbol{\theta}_0$ gives the "best" approximation in the sense of minimizing

$$\int_S \{E[y|\mathbf{x}] - \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}\}^2 d\mathbf{x}.$$

Define

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

then we have the model (4.1) and $f(\mathbf{x})$ satisfies

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

This is the second condition in (4.2). The first condition in (4.2) is to balance the bias and the variance. In one occasion, (4.3) is used for one example in Chapter Four.

5 Designs robust against autocorrelation of errors

When errors $\epsilon_1, \cdots, \epsilon_n$ in (1.2) are correlated, we denote by P_n the autocorrelation matrix. If P_n is known, then the optimal design can be derived by minimizing the covariance matrix $COV(\hat{\boldsymbol{\theta}})$. Here we assume that A1 is true. In most cases, however, P_n is unknown or known only up to certain forms, such as weak stationarity or the first order moving average (MA(1)) autocorrelation. Then we cannot minimize the covariance matrix $COV(\hat{\boldsymbol{\theta}})$ to get optimal designs, since $COV(\hat{\boldsymbol{\theta}})$ contains the unknown parameter P_n. Instead we seek robust designs which safeguard against a range of autocorrelation structures.

In Constantine (1989), robust designs are derived for the linear model $E[\mathbf{y}] = X\boldsymbol{\theta}$ and the covariance matrix $COV[\mathbf{y}] = V = I + M$, where I is the identity matrix and $M = (m_{ij})$, with $m_{i,i+1} = m_{i+1,i} = \rho_i$, and $m_{ij} = 0$ otherwise. The criterion is to maximize the trace of the inverse of the covariance matrix $tr\left((COV(\hat{\boldsymbol{\theta}}))^{-1}\right) = tr\left(X^T V^{-1} X\right)$, here $\hat{\boldsymbol{\theta}} = \left(X^T V^{-1} X\right)^{-1} X^T V^{-1}\mathbf{y}$ is the BLUE (Best Linear Unbiased Estimate). Constantine starts with an optimal, or efficient, design under the usual model with uncorrelated errors and modifies that according to the sign of the correlation that may be introduced at each step of an experiment. The modification does not change the optimality of the design under the uncorrelated model, but it increases its efficiency if nonzero correlations are indeed present. Using a linear approximation to a matrix expression $X^T V^{-1} X$, he finds (assuming all ρ_i's are equal to ρ) that a robust efficient design matrix has as many sign changes as possible within each column of X if ρ is positive, and, on the other hand, requires as few such changes as possible if ρ is negative.

In Berenblut and Webb (1974), robust designs are derived for the linear model $\mathbf{y} = X\boldsymbol{\theta} + \epsilon$, where X is the design matrix, the covariance matrix of errors is $\sigma^2 V(\rho)$, ρ is the parameter of first-order autocorrelated process, and $V(0) = I$ is the identity matrix corresponding to uncorrelated errors. Both non-stationary and stationary first-order autocorrelated error processes are considered. The criterion is to minimize the determinant of the covariance matrix of the BLUE. They first determine the set of optimal designs which minimize $det((X^T X)^{-1})$—the determinant of the covariance matrix of the BLUE for independent errors, and then they select a subset which minimize $det((X^T V^{-1}(\rho)X)^{-1})$—the determinant of the covariance matrix of BLUE for $V(\rho)$.

So in the literature, robust designs for correlated observations are usually obtained through two steps.

Step 1. Find a class $\mathcal{C}(\xi)$ of designs which are optimal for uncorrelated errors;

Step 2. Find a optimal design $\xi^* \in \mathcal{C}(\xi)$ (usually by a proper ordering of the design points $\mathbf{x}_i$) which minimizes the covariance matrix of the estimate under correlation.

Therefore these robust designs are still classical optimal designs. Robustness against autocorrelation of errors is achieved by proper ordering of classical optimal design points $\mathbf{x}_i$; see also Kiefer and Wynn (1981) and Jenkins and Chanmugam (1962).

In a time series context several authors -- Sacks and Ylvisaker (1966, 1968), Bickel and Herzberg (1979), Bickel, Herzberg and Schilling (1981) and Ylvisaker (1987) - have studied the problem of determining time points t_i at which to observe

$$Y(t) = \sum_{j=1}^{k} \theta_j z_j(t) + \epsilon(t), \tag{5.1}$$

under various assumptions on the process $\{\epsilon(t)\}$.

In Bickel and Herzberg (1979), the error process in (5.1) is assumed to be weakly stationary:

$$E[\epsilon(t_i)] = 0,$$

$$Var[\epsilon(t_i)] = \sigma^2,$$

$$Corr(Y(t_i), Y(t_j)) = \gamma\rho(t_i - t_j),$$

where $-T \leq t_1 \leq t_2 \leq \cdots \leq t_n \leq T$, $0 \leq \gamma \leq 1$ and ρ is the correlation function of a nondegenerate stationary process. Then the autocorrelation matrix of the errors is $U = (\gamma\rho(t_i - t_j) + (1-\gamma)\delta_{ij})$. Bickel and Herzberg establish an asymptotic theory of the variance-covariance matrix Σ of the least squares estimates to study the effect of dependence of the observations in experimental design, where $\Sigma = \sigma^2 \left(Z^T Z\right)^{-1} Z^T U Z \left(Z^T Z\right)^{-1}$. Robust designs are constructed for simple cases of model (5.1), i.e. estimation of location ($k = 1,\ z_1(t) = 1$), regression through the origin ($k = 1,\ z_1(t) = t$) and simple linear regression ($k = 2,\ z_1(t) = 1,\ z_2(t) = t$). Two optimality criteria are considered:

1. Minimize $n\Sigma$ over designs $(t_1, \cdots, t_n)$;

2. Minimize $n\Sigma$ over designs $(t_1, \cdots, t_n)$ subject to $n\left(Z^T Z\right)^{-1} \leq \lambda T^{-2}\mathbf{I}, \quad (\lambda > 1)$.

In this thesis, we use these and other approaches to derive robust designs against autocorrelation of errors. We do not confine ourselves to the class of classical optimal designs to search for robust designs. By introducing the Change-of-Variance function into design theory, we have derived robust designs which minimize the determinant of the covariance matrix for uncorrelated errors subject to a robustness constraint.

So far we have discussed robust designs which are robust against only one kind of model departure, i.e. violation of A1 or A2. If there is a violation of both A1 and

A2 in the linear model (1.2), then how do we construct robust designs? This is a major topic studied in this thesis. Both the minimax approach and the infinitesimal approach are used to construct robust designs.

6 Robust designs derived in this thesis

In the second chapter of this thesis, titled "Minimax Regression Designs for Approximately Linear Models with Autocorrelated Errors", we study the construction of regression designs, when the random errors are autocorrelated. Our model of dependence assumes that the spectral density $g(w)$ of the error process is of the form

$$g(w) = (1 - \alpha)g_0(w) + \alpha g_1(w),$$

where $g_0(w)$ is uniform (corresponding to uncorrelated errors), $\alpha \in [0, 1)$ is fixed, and $g_1(w)$ is arbitrary. We consider regression responses which are exactly, or only approximately, linear in the parameters.

Our main results are that *a design which is asymptotically (minimax) optimal for uncorrelated errors retains its optimality under autocorrelation if the design points are a random sample, or a random permutation, of points from this distribution.* Our results are then a partial extension of those of Wu (1981), on the robustness of randomized experimental designs, to the field of regression design.

In the third chapter, titled "Minimax Designs for Approximately Linear Models with AR(1) Errors", we obtain designs for linear models under two main departures from the classical assumptions: *i*) the response is taken to be only *approximately* linear, and *ii*) the errors are not assumed to be independent, but to instead follow a first order autoregressive process. These designs have the property that they minimize the maximum Integrated Mean Squared Error of the estimated response, with the maximum taken over a class of departures from strict linearity, and over all autoregressive parameters ρ, $|\rho| < 1$, of fixed sign. The same design is optimal for the Best Linear Unbiased Estimate and for the Ordinary Least Squares Estimate.

In Chapter four, titled "Robust Designs Based on the Infinitesimal Approach", we introduce an *infinitesimal* approach to the construction of robust designs for linear models. These designs are robust against small departures from the assumed linear regression response and/or small departures from the usual assumption of uncorrelated errors. Subject to satisfying a robustness constraint, they minimize the mean squared error of the estimator at the ideal model. The robustness constraint is formulated in terms of boundedness of the Gateaux derivative of the mean squared error of the estimated response, in the direction of a contaminating response function or autocorrelation structure. These notions are closely related to those of V-robustness and B-robustness of estimators, as formulated by Hampel et al (1986). Specific examples are considered. If the aforementioned bounds are sufficiently large, then the classically optimal designs, which minimize variance alone at the ideal model, meet our robustness criteria. Otherwise, new designs are obtained.

7 Application

We consider the yield-density model in agriculture from Seber and Wild (1989, p360). The model is used for quantifying the relationship between the density of crop planting and the crop yield. Because of competition between plants for resources, the yield per plant tends to decrease with increasing density of planting. The agronomist tends to be interested in yield-density curves for prediction, for finding the density of planting to maximize yield, and for the comparison of relationships under different conditions. Let w denote the yield per unit area, x the density of planting, and $y = w/x$ the average yield per plant if all plants survived. The most commonly used model is Shinozaki and Kira (1956):

$$E[y|x] = (\alpha + \beta x)^{-1}, \quad x > 0, \tag{7.1}$$

where α and β are unknown parameters.

Let $[a, b]$ be an interval of interest, over which $E[y|x]$ is approximately linear. The nonlinear model (7.1) can be linearized by the following two methods.

Method 1. Let $y' = y^{-1}$, $x' = \frac{x - \frac{a+b}{2}}{b-a}$, $\theta_0 = \alpha + \frac{\beta}{2}(a+b)$ and $\theta_1 = (b-a)\beta$, then (7.1) gives

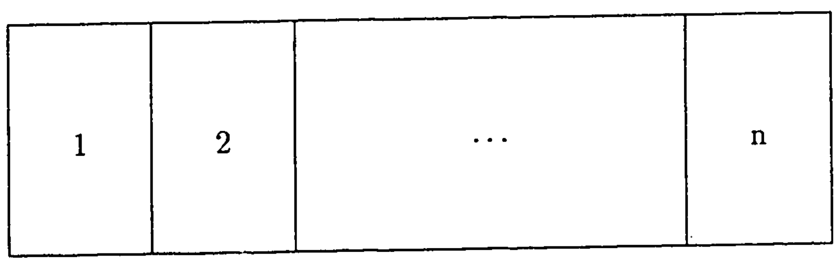
$$E[y'|x'] \approx \theta_0 + \theta_1 x', \quad x' \in [-0.5, 0.5]. \tag{7.2}$$

Method 2. Let $x_0 = \frac{a+b}{2}$ and $h = \frac{b-a}{2}$, then (7.1) becomes

$$\begin{aligned} E[y|x] &= (\alpha + \beta x_0)^{-1} - \beta(\alpha + \beta x_0)^{-2}(x - x_0) + f(x), \quad x \in [a, b], \\ &= \theta_0 + \theta_1 x' + f(x'), \quad x' \in [-0.5, 0.5], \end{aligned} \tag{7.3}$$

where $\theta_0 = (\alpha + \beta x_0)^{-1}$, $\theta_1 = -2\beta h(\alpha + \beta x_0)^{-2}$ and $x' = \frac{x - x_0}{2h}$.

Since (7.2) is only approximately true, it (ignore the $'$ on y) can also be written in the form of (7.3). Now suppose we are going to do an experiment on a piece of land. The purpose of this experiment is to estimate the relationship between the density of crop planting and the crop yield for a given crop of interest. In order to do so, we first divide the piece of land into n plots at the beginning of the planting season, such as the following.



On each plot i, we plant certain numbers of plants of the crop (this is the density of crop planting x_i). When harvesting the crop, we record the total yield w_i for each plot i. Then $y_i = w_i/x_i$ is the average yield per plant (if all plants survived) corresponding to the crop density x_i. Now based on the n observations $(x_1, y_1), \cdots, (x_n, y_n)$, we can estimate θ_0 and θ_1 in (7.3). So the experimental design problem is to choose optimal x_i's at the beginning to plant the crop such that the estimates for θ_0 and θ_1 are the most efficient. It is usual that there exist certain correlations between the crop yields for the neighbouring plots, which are called the spatial correlations. The spatial correlation between two plots depends on the distance between the two plots. Usually the closer the two plots are, the stronger the correlation is. In the following examples, we give robust designs for density x on [a,b] according to various spatial correlations of the crop yield.

Example 7.1. Suppose we only know that the experimental errors are weakly stationary. Then, from Theorem 2.5 in Chapter Two, the minimax design for x' is attained by randomly sampling design points from ξ_*. From Huber (1975), ξ_* has density function

$$m(x') = \frac{1}{(1-2c)^2}\left(1 - \frac{c^2}{x'^2}\right)^+, \quad x' \in [-0.5, 0.5],$$

where c depends on the ratio $\frac{\sigma^2}{n\eta^2}$. When $\frac{\sigma^2}{n\eta^2} \to 0$, $m(x') \to 1$, which is the uniform density function on $[-0.5, 0.5]$. In this case, design points may be chosen as follows. Let M be the distribution function corresponding to m. Select n points $M^{-1}(\frac{i-0.5}{n})$, $i = 1, \cdots, n$, whose distribution function tends weakly to M. Then we take a random permutation of these points. A set of design points selected in such a manner for $n = 16$ when $\frac{\sigma^2}{n\eta^2} = 1$, $m(x') = 5.358(1 - \frac{0.081}{x'^2})^+$, and $0.284 \le |x'| \le 0.5$ is:

$$\langle 0.390, -0.491, -0.413, -0.390, -0.363, 0.473, -0.473, -0.435,$$
$$0.435, -0.325, 0.325, 0.363, -0.455, 0.491, 0.455, 0.413\rangle.$$

Then the design points for x are computed from $x_i = (b-a)x'_i + \frac{a+b}{2}$.

Example 7.2. Suppose that the experimental errors follow a first order autoregressive process with parameter ρ. Then, from Theorem 3.1 in Chapter Three, the minimax design ξ_* for x' has density function

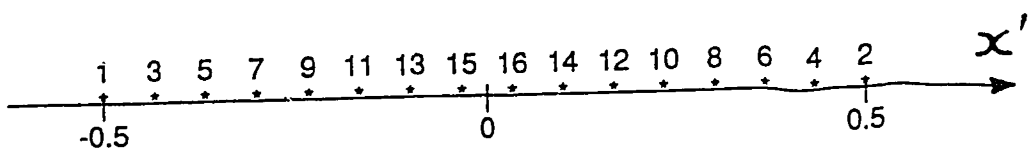
$$m(x') = c(x'^2 + d)^+, \quad x' \in [-0.5, 0.5],$$

where c and d depend on the ratio $\frac{\sigma^2}{n\eta^2}$. In the limiting case as $\frac{\sigma^2}{n\eta^2} \to 0$, the design ξ_* converges to the uniform distribution on $[-0.5, 0.5]$. When $\rho > 0$ (presumably the case), a set of $n = 16$ design points for x' is:

$$\langle -0.5, 0.5, -0.433, 0.433, -0.367, 0.367, -0.3, 0.3,$$

$$-0.233, 0.233, -0.167, 0.167, -0.1, 0.1, -0.033, 0.033 \rangle.$$

A plot of these design points is the following.



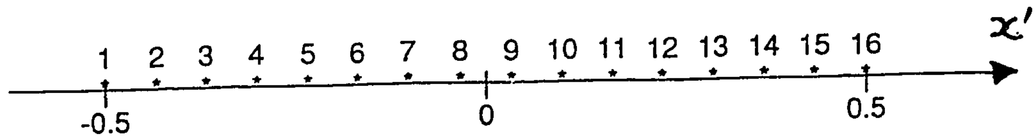
These designs points alternate between negative values and positive values.

When $\rho < 0$, a set of $n = 16$ design points for x' is:

$$\langle -0.5, -0.433, -0.367, -0.3, -0.233, -0.167, -0.1,$$

$$-0.033, 0.033, 0.1, 0.167, 0.233, 0.3, 0.367, 0.433, 0.5 \rangle.$$

A plot of these design points is the following.



The first half of these design points are negative, and the another half are positive. So in this example the order of design points is very important and only depends on the sign of ρ.

Example 7.3. If we assume that the experimental errors follow a first order moving average process with positive correlation, then the M-robust design (robust against small departures from both assumed regression response and uncorrelated errors) can be applied. For some appropriate robustness bounds on the Change-of-Variance Function and the Change-of-Bias Function, design points should be chosen according to Proposition 6.1 in Chapter Four. The optimal design ξ^* is computed from Example 5.1 in Chapter Four, which, for instance, has density function

$$g_0(x') = 12x'^2, \quad x' \in [-0.5, 0.5].$$

Now we select n points $G_0^{-1}(\frac{i-0.5}{n})$ whose distribution function tends weakly to G_0. These design points should be arranged in such a way that $\frac{\sum_{i=1}^{n-1} x_i' x_{i+1}'}{\sum_{i=1}^{n} x_i'^2} < 0$. For $n = 16$, a set of design points is

$$\langle -0.489, 0.489, -0.467, 0.467, -0.441, 0.441, -0.413, 0.413,$$
$$-0.380, 0.380, -0.339, 0.339, -0.286, 0.286, -0.198, 0.198 \rangle.$$

If we are only interested in a design robust against autocorrelation of errors, then the most V-robust design can be used. From Theorem 4.5 in Chapter Four, the most V-robust design is

$$x'_i = 0.5 \cdot (-1)^{i+1} \frac{\sin\frac{\pi i}{n+1}}{\sin\frac{\pi(\frac{n}{2}+1)}{n+1}}.$$

For $n = 16$, design points are

$$\langle 0.092, -0.181, 0.264, -0.338, 0.401, -0.449, 0.483, -0.5,$$

$$0.5, -0.483, 0.449, -0.401, 0.338, -0.264, 0.181, -0.092\rangle.$$

8 Conclusion

In this thesis, we have studied robust designs when errors are correlated. But there is more work to be done in the future study.

We only focus on the case of straight line regression in Chapter Three. For multiple regression, minimax designs can also be studied when errors follow a first order autoregressive process (AR(1)). Of course, one can also consider the construction of minimax designs for other kinds of error processes such as AR(p) or MA(p).

In Chapter Four, the infinitesimal approach is introduced to derive robust designs. Special cases are considered there. Further study is needed to construct more designs robust against autocorrelation of errors.

References

Berenblut, I.I. and G.I. Webb (1974). Experimental design in the presence of autocorrelated errors. *Biometrika* **61**, 427-437.

Bickel, P.J. and A.M. Herzberg (1979). Robustness of design against autocorrelation in time I: Asymptotic theory, optimality for location and linear regression. *Ann. Statist.* **7**, 77-95.

Bickel, P.J., A.M. Herzberg and M.F. Schilling (1981). Robustness of design against autocorrelation in time II: Optimality, theoretical and numerical results for the first-order autoregressive process. *J. Amer. Statist. Assoc.* **76**, 870-877.

Box, G.E.P. (1987). *Empirical Model-Building and Response Surfaces.* Wiley, New York.

Box, G.E.P. and N. Draper (1959). A basis for the selection of a response surface design. *J. Amer. Statist. Assoc.* **54**, 622-654.

Constantine, G.M. (1989). Robust designs for serially correlated observations. *Biometrika* **76**, 245-251.

Cox, D.R. (1951). Some systematic experimental designs. *Biometrika* **38**, 312-323.

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

Hampel, F.R., R. Ronchetti, R.J. Rousseeuw and W. Stahel (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

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

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

Jenkins, G.M. and J. Chanmugam (1962). The estimation of slope when the errors are autocorrelated. *J.R. Statist. Soc.* **B 24**, 199-214.

Kiefer, J. and H.P. Wynn (1981). Optimum balanced block and Latin squares designs for correlated observations. *Ann. Statist.* **9**, 737-757.

Kiefer, J. and H.P. Wynn (1983). Autocorrelation-robust design of experiments. In: T. Leonard and C.-F. Wu, Ed., *Scientific Inference, Data Analysis and Robustness.* Academic Press, New York, 279-299.

Li, K.C. (1984). Robust regression designs when the design space consists of finitely many points. *Ann. Statist.* **12**, 269-282.

Li, K.C. and W. Notz (1982). Robust designs for nearly linear regression. *J. Statist. Plann. Inference* **6**, 135-151.

Liu, S.X. and D.P. Wiens (1994). Robust designs for approximately polynomial regression. Preprint.

Marcus, M.B. and J. Sacks (1976). Robust designs for regression problems. In: S.S. Gupta and D.S. Moore, Eds., *Statistical Decision Theory and Related Topics II.* Academic Press, New York, 245-268.

Montgomery, D.C. (1991). *Design and Analysis of Experiments*, Wiley, New York.

Pesotchinsky, L. (1982). Optimal robust designs: Linear regression in R^k. *Ann. Statist.* **10**, 511-525.

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

Sacks, J. and D. Ylvisaker (1966). Design for regression problems with correlated errors. *Ann. Math. Statist.* **37**, 66-89.

Sacks, J. and D. Ylvisaker (1968). Design for regression problems with correlated errors: many parameters. *Ann. Math. Statist.* **39**, 49-69.

Sacks, J. and D. Ylvisaker (1978). Linear estimation for approximately linear models. *Ann. Statist.* **6**, 1122-1137.

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

Shinozaki, K. and T.Kira (1956). Intraspecific competition among higher plants. VII. Logistic theory of the C-D effect. *J. Inst. Polytech.* Osaka City Univ. D7, 35-72.

Wiens, D.P. (1990). Robust minimax designs for multiple linear regression. *Linear Algebra Appl.* **127**, 327-340.

Wiens, D.P. (1992). Minimax designs for approximately linear regression. *J. Statist. Plann. Inference* **31**, 353-371.

Wiens, D.P. (1993). Designs for approximately linear regression: Maximizing the minimum coverage probability of confidence ellipsoids. *Cdn. J. Statist.* **21**, 59-70.

Wiens, D.P. (1994). Robust designs for approximately linear regression: M-estimated parameters. *J. Statist. Plann. Inference* **40**, 135-160.

Wu, C.-F. (1981). On the robustness and efficiency of some randomized designs. *Ann. Statist.* **9**, 1168-1177.

Ylvisaker, D. (1987). Prediction and design. *Ann. Statist.* **15**, 1-19.

Chapter Two

Minimax Regression Designs for Approximately Linear Models with Autocorrelated Errors

1 Introduction

In this paper we study optimal designs for regression models under certain departures from the classical assumptions. The usual formulation of the fixed-regressors linear regression model, which we write as

$$Y_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\theta} + \varepsilon_i \quad i = 1, \ldots, n; \tag{1.1}$$

$$\mathbf{x}_i \in S \subseteq \mathbb{R}^q, \mathbf{z}(\mathbf{x}_i) \in \mathbb{R}^p, \boldsymbol{\theta} \in \mathbb{R}^p;$$

$$E[\boldsymbol{\varepsilon}] = \mathbf{0}, COV[\boldsymbol{\varepsilon}] = \sigma^2 \boldsymbol{I} \tag{1.2}$$

employs the following assumptions.

[0]A version of this chapter has been submitted for publication. Douglas P. Wiens and Julie Zhou, 1994. *Journal of Statistical Planning and Inference.*

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

II) The errors ε_i are uncorrelated.

The optimal design problem is then to choose values $\mathbf{x}_i$ of the independent variables in such a way as to minimize some scalar valued function of the covariance matrix of the Ordinary Least Squares estimate $\hat{\boldsymbol{\theta}}$.

Beginning with Box and Draper (1959), numerous attempts have been made to relax I) above. A possible alternative assumes only that $E[Y|\mathbf{x}]$ is approximately linear:

$$E[Y|\mathbf{x}] \approx \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}_0$$

where $\boldsymbol{\theta}_0$ minimizes

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

Then with

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

the model (1.1) becomes instead

$$Y_i = \mathbf{z}^T(\mathbf{x}_i)\boldsymbol{\theta}_0 + n^{-1/2} f(\mathbf{x}_i) + \varepsilon_i, \tag{1.3}$$

$$\int_S \mathbf{z}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \mathbf{0}. \tag{1.4}$$

If $f(\mathbf{x})$ above is non-zero then the regression estimates are typically biased. In order that the bias not dominate the variance, $f(\mathbf{x})$ is constrained by a condition such as

$$\int_S f^2(\mathbf{x}) d\mathbf{x} \le \eta^2 \tag{1.5}$$

for some constant η. The optimality problem is then based on the minimization of a function of the Mean Squared Error matrix. See Huber (1975), Wiens (1992, 1993) for details and special cases. For other approaches to the weakening of I) see Marcus and Sacks (1976), Sacks and Ylvisaker (1978), Pesotchinsky (1982), Li and Notz (1982), Li (1984), Notz (1989) and Liu (1994).

The literature concerning optimal design theory under departures from II) is somewhat more sparse. In a time series context several authors – Sacks and Ylvisaker (1966, 1968), Bickel and Herzberg (1979), Bickel, Herzberg and Schilling (1981) and Ylvisaker (1987) – have studied the problem of determining time points t_i at which to observe

$$Y(t) = \sum_j \theta_j z_j(t) + \varepsilon(t),$$

under various assumptions on the process $\{\varepsilon(t)\}$. Constantine (1989) obtained designs for a special case of (1.1) under a lag-one serial correlation model for the errors. See also Cox (1951) and Kiefer and Wynn (1981, 1983).

Here, we study designs both for (1.1) and for (1.3), with (1.2) replaced by

$$E[\boldsymbol{\varepsilon}] = \mathbf{0}, COV[\boldsymbol{\varepsilon}] = \sigma^2 \boldsymbol{P} \tag{1.6}$$

where $\boldsymbol{P}$ is a positive semi-definite Toeplitz matrix with unit diagonal, i.e. the autocorrelation matrix of a weakly stationary process. Thus, if

$$\rho(s) := E[\varepsilon_1 \varepsilon_{1+s}]/\sigma^2$$

then $\boldsymbol{P}_{ij} = \rho(|i-j|)$. We assume throughout that

$$\sum_{s=-\infty}^{\infty} |\rho(s)| < \infty. \tag{1.7}$$

In Section 2 we obtain the asymptotic form of the covariance matrix of $\sqrt{n}\widehat{\boldsymbol{\theta}}$, under (1.1) and (1.6). We then consider a broad class of departures from (1.2) under which

$$\boldsymbol{P} = (1-\alpha)\boldsymbol{I} + \alpha\boldsymbol{Q} \tag{1.8}$$

for an arbitrary (subject to (1.7)) autocorrelation matrix $\boldsymbol{Q}$ and fixed $\alpha \in [0, 1)$. For any scalar valued function L of covariance matrices, monotonic in that

$$\boldsymbol{V}_1 \leq \boldsymbol{V}_2 \text{ (w.r.t. positive definiteness)} \Longrightarrow L(\boldsymbol{V}_1) \leq L(\boldsymbol{V}_2) \tag{1.9}$$

we consider the problem of choosing a design to minimize the maximum loss L, with the maximum evaluated subject to (1.8).

We show that this problem has the following asymptotic solution. Suppose that the corresponding optimal design problem under (1.2) has an asymptotic solution described by a particular design measure, i.e. a probability measure ξ_0 on the design space S. Then *a minimax strategy under (1.8) consists of randomly sampling design points from* ξ_0. *Alternatively, one may take a random permutation of design points whose empirical distribution function (e.d.f.) tends weakly to* ξ_0. The latter strategy typically amounts to designing the experiment for uncorrelated errors, and then randomizing the order of implementation.

We also show that if ξ_* is a minimax design for (1.3) under (1.2), where the maximum is evaluated over f satisfying (1.4) and (1.5), then ξ_* retains its optimality under (1.6) - (1.8) as well, if the design points are randomly sampled from ξ_* or if they constitute a random permutation whose e.d.f. tends weakly to ξ_*.

It has long been argued that the proper use of randomization in experimental design is a source of robustness against model inadequacies. Wu (1981) gave this notion a formal treatment and rigorous justification. The present work can be viewed as a partial extension to the field of regression design.

2 Derivations

We first list several assumptions which are made on the sequence of designs.

Assume

(A1) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{z}(\mathbf{x}_i) = 0.$

For $0 \le s \le n-1$ define matrices

$$\boldsymbol{B}_n(s) = \frac{1}{n}\sum_{i=1}^{n-s} \mathbf{z}(\mathbf{x}_i)\mathbf{z}^T(\mathbf{x}_{i+s})$$

and define $\boldsymbol{B}_n(s) = \boldsymbol{B}_n^T(-s)$ for $s < 0$. Assume

(A2) For each $s, \boldsymbol{B}_n(s)$ tends to a limit $\boldsymbol{B}(s)$ as $n \to \infty$, and $\boldsymbol{B}(0)$ is positive definite.

For the approximately linear model, i.e. if $\eta > 0$ in (1.5), we define

$$\boldsymbol{b}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}(\mathbf{x}_i)f(\mathbf{x}_i)$$

and assume

(A3) There exists $\lim_{n\to\infty} \boldsymbol{b}_n =: \boldsymbol{b}_f$.

Note that **(A3)** imposes no restriction on the exactly linear model ($\eta = 0$), since then $\boldsymbol{b}_n = \boldsymbol{b}_f = \mathbf{0}$.

By (1.7), there exists a symmetric spectral density $g(\omega)$ on $[-\pi, \pi]$ satisfying

$$\begin{aligned} \rho(s) &= \int_{-\pi}^{\pi} e^{is\omega} g(\omega)d\omega = \int_{-\pi}^{\pi} \cos(s\omega)g(\omega)d\omega, \\ g(\omega) &= (2\pi)^{-1} \sum_{s=-\infty}^{\infty} \rho(s)e^{-is\omega} = (2\pi)^{-1} \sum_{s=-\infty}^{\infty} \rho(s)\cos(s\omega). \end{aligned} \tag{2.1}$$

Under a set of assumptions implied by **(A1)** and **(A2)**, it is proven in Grenander and Rosenblatt (1957) that there exists a regression spectral distribution function $\boldsymbol{H}(\omega)$ ($\omega \in [-\pi, \pi]$) – a symmetric $p\times p$ matrix whose increments $\boldsymbol{H}(\omega_2) - \boldsymbol{H}(\omega_1)(\omega_1 < \omega_2)$ are positive semi-definite – satisfying

$$\boldsymbol{B}(s) = \int_{-\pi}^{\pi} e^{is\omega} d\boldsymbol{H}(\omega). \tag{2.2}$$

Lemma 2.1. *Define* $\boldsymbol{C}_n = COV[\sqrt{n}\widehat{\boldsymbol{\theta}}]$ *and* $\boldsymbol{M}_n =$ MSE $[\sqrt{n}\widehat{\boldsymbol{\theta}}]$. *Under assumptions* **(A1)** *–* **(A3)** *there exist* $\lim_{n\to\infty} \boldsymbol{C}_n =: \boldsymbol{C}$ *and* $\lim_{n\to\infty} \boldsymbol{M}_n =: \boldsymbol{M}$. *The limit matrices have representations*

$$\boldsymbol{C} = \sigma^2 \boldsymbol{B}^{-1}(0)[\sum_{s=-\infty}^{\infty} \rho(s)\boldsymbol{B}(s)]\boldsymbol{B}^{-1}(0), \tag{2.3}$$

$$C = 2\pi\sigma^2 B^{-1}(0)\Big[\int_{-\pi}^{\pi} g(-\omega)dH(\omega)\Big]B^{-1}(0), \tag{2.4}$$

$$M = C + B^{-1}(0)b_f b_f^T B^{-1}(0). \tag{2.5}$$

Proof: Let Z be the $n \times p$ design matrix with i^{th} row $\mathbf{z}^T(\mathbf{x}_i)$. Let n be large enough that $B_n(0) = \frac{1}{n}Z^TZ$ is non-singular. By standard regression theory,

$$\begin{aligned} C_n &= \sigma^2(\frac{1}{n}Z^TZ)^{-1}(\frac{1}{n}Z^TPZ)(\frac{1}{n}Z^TZ)^{-1} \\ &= \sigma^2 B_n^{-1}(0)\Big[\sum_{|s|\leq(n-1)} \rho(s)B_n(s)\Big]B_n^{-1}(0). \end{aligned} \tag{2.6}$$

Let $\|\cdot\|$ denote the Euclidean norm. Then

$$\|B_n(s)\| \leq \operatorname{tr}(B_n(0)) \leq 2 \operatorname{tr}B(0) \tag{2.7}$$

for sufficiently large n. This, the Dominated Convergence Theorem and (1.7) applied to (2.6) yield (2.3). Now substitute (2.2) into (2.3), use (1.7) and the Dominated Convergence Theorem to interchange the summation and integration, then apply (2.1) to obtain (2.4). A simple calculation using (1.4) gives

$$M_n = C_n + B_n^{-1}(0)b_n b_n^T B_n^{-1}(0),$$

whence (**A2**) and (**A3**) yield (2.5). □

Suppose now that the error process $\{\varepsilon_i\}$ is the sum of two uncorrelated processes $\{\varepsilon_i'\}$ and $\{\varepsilon_i''\}$, where $\{\varepsilon_i'\}$ is white noise with variance $(1-\alpha)\sigma^2$ for $\alpha \in [0,1)$, and

$\{\varepsilon_i''\}$ is weakly stationary with variance $\alpha\sigma^2$ and absolutely summable autocorrelation function $\rho_1(s)$. Then (1.6) and (1.8) hold, with $\boldsymbol{Q}_{ij} = \rho_1(|i-j|)$. Furthermore,

$$\rho(s) = (1-\alpha)\rho_0(s) + \alpha\rho_1(s), \text{ where } \rho_0(s) = \begin{cases} 1, & s = 0; \\ 0, & s \neq 0. \end{cases}$$

Denote by $g_1(\omega)$ the spectral density corresponding to $\rho_1(s)$, and let $g_0(\omega) = (2\pi)^{-1}I(|\omega| \leq \pi)$. Then $\rho(s)$ has spectral density

$$g(\omega) = (1-\alpha)g_0(\omega) + \alpha g_1(\omega).$$

See Bickel and Herzberg (1979) and Samarov (1987) for other instances of this model for $\{\varepsilon_i\}$. From Lemma 2.1,

$$\boldsymbol{C} = \sigma^2\{(1-\alpha)\boldsymbol{B}^{-1}(0) + \alpha\boldsymbol{B}^{-1}(0)\Big[\sum_{s=-\infty}^{\infty}\rho_1(s)\boldsymbol{B}(s)\Big]\boldsymbol{B}^{-1}(0)\}. \tag{2.8}$$

A consequence of the following lemma is that $\boldsymbol{C}$ is non-singular for all $\alpha \in [0,1)$.

Lemma 2.2. *If $\rho_1(s)$ is absolutely summable then the matrix $\sum_{s=-\infty}^{\infty}\rho_1(s)\boldsymbol{B}(s)$ is positive semi-definite.*

Proof: In the same manner that (2.2) led to (2.4) we obtain

$$\sum_{s=-\infty}^{\infty}\rho_1(s)\boldsymbol{B}(s) = 2\pi\int_{-\pi}^{\pi} g_1(-\omega)d\boldsymbol{H}(\omega),$$

so that for any $p \times 1$ vector $\boldsymbol{a}$,

$$\boldsymbol{a}^T\Big[\sum_{s=-\infty}^{\infty}\rho_1(s)\boldsymbol{B}(s)\Big]\boldsymbol{a} = 2\pi\int_{-\pi}^{\pi} g_1(-\omega)d(\boldsymbol{a}^T\boldsymbol{H}(\omega)\boldsymbol{a}) \geq 0.$$ □

Grenander and Rosenblatt (1957) give conditions under which the LSE $\hat{\boldsymbol{\theta}}$ is fully efficient. In particular, these hold if

$$\boldsymbol{H}(\omega) = (2\pi)^{-1}\omega \boldsymbol{B}(0) \tag{2.9}$$

in which case it is obvious from (2.2) that $\boldsymbol{B}(s) = \mathbf{0}$ for $s \neq 0$. If the design points $\mathbf{x}_i$ constitute a random sample from a particular design distribution, then (2.9) holds. In this case however the effect on $COV[\sqrt{n}\hat{\boldsymbol{\theta}}]$ of the sampling variation must be taken into account. It is not clear that the conclusions of Lemma 2.1 continue to hold.

Lemma 2.3 below shows that the effect of this sampling variation is asymptotically negligible. For this lemma we assume that design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are randomly sampled from a distribution function ξ on S. Define

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

$$\boldsymbol{b}_f(\xi) = \int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})d\xi(\mathbf{x}).$$

Replace assumptions **(A1)** – **(A3)** by:

(B1) The design points $\{\mathbf{x}_i\}_{i=1}^n$ are randomly chosen from ξ, and are uncorrelated with $\{\varepsilon_i\}_{i=1}^n$.

(B2) $E_\xi[\mathbf{z}(\mathbf{x})] = \mathbf{0}$.

(B3) $\boldsymbol{B}(\xi)$ is positive definite.

(B4) The eigenvalues of $\boldsymbol{B}_n(0)$ are bounded above, and away from zero, as $n \to \infty$.

(B5) $f(\mathbf{x})$ is bounded on S.

Note that by **(B1)**, **(B2)** and the Strong Law of Large Numbers,

$$\boldsymbol{B}_n(0) \longrightarrow \boldsymbol{B}(\xi)(\text{a.s.}), \tag{2.10}$$

$$\boldsymbol{B}_n(s) \longrightarrow \mathbf{0}(\text{a.s.}) \text{ for } s \neq 0, \tag{2.11}$$

$$\boldsymbol{b}_n \longrightarrow \boldsymbol{b}_f(\xi)(\text{a.s.}). \tag{2.12}$$

Then **(B3)** and (2.10) imply that **(B4)** holds with probability one as $n \to \infty$.

Lemma 2.3. *Assume* (**B1**)*–*(**B5**)*. Let* $\boldsymbol{C}_n$ *and* $\boldsymbol{M}_n$ *be as in Lemma 2.1. Then the conclusions of Lemma 2.1 hold, with* $\boldsymbol{B}(0) = \boldsymbol{B}(\xi), \boldsymbol{b}_f = \boldsymbol{b}_f(\xi)$ *and* $\boldsymbol{H}(\omega)$ *as in (2.9); i.e.*

$$\boldsymbol{C}_n \longrightarrow \boldsymbol{C}_\xi := \sigma^2 \boldsymbol{B}^{-1}(\xi), \tag{2.13}$$

$$\boldsymbol{M}_n \longrightarrow \boldsymbol{M}_{f,\xi} := \boldsymbol{C}_\xi + \boldsymbol{B}^{-1}(\xi)\boldsymbol{b}_f(\xi)\boldsymbol{b}_f^T(\xi)\boldsymbol{B}^{-1}(\xi). \tag{2.14}$$

Proof: Define random matrices $\boldsymbol{D}_n$ and random vectors $\boldsymbol{d}_n$ by

$$\begin{aligned} \boldsymbol{D}_n &= \boldsymbol{B}_n^{-1}(0)\Big[\sum_{|s|\leq(n-1)} \rho(s)\boldsymbol{B}_n(s)\Big]\boldsymbol{B}_n^{-1}(0), \\ \boldsymbol{d}_n &= \boldsymbol{B}_n^{-1}(0)\boldsymbol{b}_n. \end{aligned}$$

By conditioning on $\{\mathbf{x}_i\}_{i=1}^n$ and using the orthogonality asserted in (**B1**) we obtain

$$\boldsymbol{C}_n = \sigma^2 E[\boldsymbol{D}_n],\ \boldsymbol{M}_n = \boldsymbol{C}_n + E[\boldsymbol{d}_n \boldsymbol{d}_n^T]. \tag{2.15}$$

We claim:

i) $\boldsymbol{D}_n \longrightarrow \boldsymbol{B}^{-1}(\xi)$ (a.s.),

ii) $\boldsymbol{d}_n \longrightarrow \boldsymbol{B}^{-1}(\xi)\boldsymbol{b}_f(\xi)$ (a.s.),

iii) $\|\boldsymbol{D}_n\|, \|\boldsymbol{d}_n\|$ are bounded.

By iii) we may take limits inside the expectation signs in (2.15); applying i) and ii) will then yield (2.13) and (2.14).

For i), let $\delta > 0$ be arbitrary and let N be large enough that $\sum_{s \geq N} |\rho(s)| < \delta$. Denote by $\lambda_{1,n}$ and $\lambda_{p,n}$ the largest and smallest eigenvalues of $\boldsymbol{B}_n(0)$. Then for $n > N$ we have, using (2.7),

$$\begin{aligned} \|\boldsymbol{D}_n - \boldsymbol{B}_n^{-1}(0)\| &\leq 2 \sum_{1 \leq s \leq N} |\rho(s)|\, \|\boldsymbol{B}_n^{-1}(0)\boldsymbol{B}_n(s)\boldsymbol{B}_n^{-1}(0)\| \\ &+ 2 \sum_{N < s < n} |\rho(s)|\, \|\boldsymbol{B}_n^{-1}(0)\|^2 \operatorname{tr}(\boldsymbol{B}_n(0)). \end{aligned} \tag{2.16}$$

The first sum above tends to 0 (a.s.) by (2.10) and (2.11). The second is bounded by

$$2p^3(\lambda_{1,n}/\lambda_{p,n}^2) \sum_{N < s < n} |\rho(s)| < K\delta$$

for some constant K independent of n, by (**B4**). Letting first n, then N tend to ∞ in (2.16) gives i).

Claim ii) above is an immediate consequence of (2.10) and (2.12), using (**B3**). For iii), note that

$$\|\boldsymbol{D}_n\| \leq \|\boldsymbol{B}_n^{-1}(0)\| \sum_{s=-\infty}^{\infty} |\rho(s)| \leq p^{1/2}/\lambda_{p,n} \sum_{s=-\infty}^{\infty} |\rho(s)|,$$

which is bounded by (**B4**). Also

$$\|\boldsymbol{d}_n\| \leq \|\boldsymbol{B}_n^{-1}(0)\|(\ \mathrm{tr}\boldsymbol{B}_n(0))^{1/2}(\frac{1}{n}\sum_{i=1}^{n} f^2(\mathbf{x}_i))^{1/2},$$

which is bounded as above, using also (**B5**). □

Remark: Lemma 2.3 continues to hold with (**B1**) replaced by

(**B1′**) The design points are a random permutation $\{\mathbf{x}_{\pi(i)}\}_{i=1}^{n}$ of $\{\mathbf{x}_i\}_{i=1}^{n}$, where the e.d.f. ξ_n of the $\{\mathbf{x}_i\}$ tends weakly to ξ.

To see this, note that the proof of Lemma 2.3 goes through essentially unchanged under (**B1′**), with (2.11) replaced by

$$E_\pi[\boldsymbol{B}_n(s)] \to \mathbf{0} \quad for\ s \neq 0.$$

This follows from

$$\begin{aligned} E_\pi[\boldsymbol{B}_n(s)] &= \frac{1}{n}\sum_{i=1}^{n-s} E_\pi[\mathbf{z}_{\pi(i)}\mathbf{z}_{\pi(i+s)}^T] \\ &= \frac{n-s}{n-1}E_{\xi_n}[\mathbf{z}(\mathbf{x})]E_{\xi_n}[\mathbf{z}^T(\mathbf{x})] - \frac{n-s}{n(n-1)}\boldsymbol{B}_n(0), \end{aligned}$$

where the second equality is obtained by first conditioning on $\pi(i)$.

Now consider the problem of choosing a design measure ξ to minimize the supremum - over all absolutely summable $\rho_1(s)$ - value of $L(C) = L(C;\rho_1,\xi)$, where L is monotonic in the sense of (1.9). Suppose that ξ_0 minimizes $L(C;\rho_0,\xi) = L(\sigma^2 B^{-1}(\xi))$, i.e. is an optimal design under (1.2). We then have

Theorem 2.4. *An asymptotically minimax design in the exactly linear model (1.1) is attained by randomly sampling design points from ξ_0, if this design measure and the sampling mechanism satisfy (***B1***) or (***B1'***) and (***B2***)-(***B4***).*

Proof: For any design measure ξ, we have

$$\sup L(C;\rho_1,\xi) \geq L(C;\rho_0,\xi) = L(\sigma^2 B^{-1}(\xi)) \geq L(\sigma^2 B^{-1}(\xi_0)) = \sup L(C;\rho_1,\xi_0),$$

where the last equality follows from the randomization. □

For the approximately linear model (1.3), denote the loss $L(M)$ by $L(M;\rho_1,f;\xi)$ and consider the problem of minimizing the supremum, over all absolutely summable $\rho_1(s)$ *and all f satisfying (1.4) and (1.5)*, of $L(M;\rho_1,f;\xi)$. Suppose that ξ_* minimizes the supremum, over f, of $L(M;\rho_0,f;\xi) = L(M_{f,\xi})$, i.e. is a minimax design for the approximately linear model under (1.2).

Theorem 2.5. *An asymptotically minimax design in the approximately linear model (1.3) is attained by randomly sampling design points from the design measure ξ_*,* *provided that this design measure and the sampling mechanism satisfy* (**B1**) *or* (**B1'**) *and* (**B2**)*–*(**B4**)*, and that* $\sup_f L(M; \rho_0, f; \xi_*)$ *is attained at a bounded* f*.*

The proof of Theorem 2.5 is very similar to that of Theorem 2.4 and so is omitted. Note that (**B5**) is required to hold only when sampling from ξ_* – only the least favourable f need be bounded. This is typically satisfied – see the examples in Huber (1975) and Wiens (1992, 1993).

Remark: Assumptions (**A1**) and (**B2**) each preclude fitting a response with an intercept. For an intercept model however, one may write the design matrix in partitioned form as $[\mathbf{1}|Z]$, and the parameter vector as $(\theta_0, \boldsymbol{\theta}_1^T)^T$. Then if this Z satisfies the above assumptions our optimality results continue to hold for the estimation of $\boldsymbol{\theta}_1$, since the columns of Z are orthogonal to $\mathbf{1}$.

3 Examples

Example 1. We consider the approximately linear regression model with $\mathbf{z}^T(\mathbf{x}) = (1, x)$ and $\boldsymbol{\theta}_0 = (\theta_0, \theta_1)^T$ in (1.3). Without loss of generality, the design space $\mathcal{S}$ is taken as $[-0.5, 0.5]$. The optimality criterion is minimax MSE of $\hat{\theta}_1$:

$$\min_{\xi} \max_{f, \rho_1} E(\hat{\theta}_1 - \theta_1)^2.$$

From Huber (1975), the minimax design for the approximately linear model with uncorrelated errors has density function

$$m_0(x) = \frac{1}{(1-2a)^2}\left(1 - \frac{a^2}{x^2}\right)^+, \quad 0 \le a < 1, \quad -0.5 \le x \le 0.5,$$

where a depends on the ratio $\frac{\sigma^2}{\eta^2}$. When $\frac{\sigma^2}{\eta^2} \to 0$, $a \to 0$ and $m_0(x) \to 1$. Then from Theorem 2.5, an asymptotically minimax design for the correlated error model is attained by randomly sampling design points from $m_0(x)$. The design points may be chosen as follows. Let M_0 be the distribution function corresponding to m_0. Select n points $M_0^{-1}(\frac{i-0.5}{n})$, $i = 1, \cdots, n$, whose empirical distribution function tends weakly to M_0. Now take a random permutation of these points. A set of design points chosen in such a manner for $n = 16$ when $a = 0$ is:

$$< 0.218, 0.0[illegible], -0.281, -0.156, -0.406, 0.469, -0.094, 0.344,$$

$$0.031, -0.344, -0.031, 0.219, 0.406, -0.219, 0.156, -0.469 > .$$

Example 2. We consider the approximately linear regression model with $\mathbf{z}^T(\mathbf{x}) = (1, x_1, x_2)$ and $\boldsymbol{\theta}_0 = (\theta_0, \boldsymbol{\theta}_1)^T$ in (1.3). The design space $\mathcal{S}$ is the sphere of unit area: $\{ \mathbf{x} \mid ||\mathbf{x}|| \le r = 1/\sqrt{\pi} \}$. The loss function is the determinant of the MSE matrix of $\hat{\boldsymbol{\theta}}_1$, so that the optimality criterion is

$$\min_{\xi} \max_{f, \rho_1} det\left(E[(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1)(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1)^T]\right).$$

From Wiens (1992), the minimax design for uncorrelated errors is the spherically symmetric design ξ_* in which $||\mathbf{x}||$ has density function

$$g_0(u) = 2\pi a u \left(1 - \frac{b}{\pi u^2}\right)^+, \quad a > 0, \quad 0 \le b < 1, \quad 0 \le u \le r,$$

where a and b depend on the ratio $\frac{\sigma^2}{\eta^2}$. By spherical symmetry, $\frac{\mathbf{x}}{||\mathbf{x}||}$ is uniformly distributed over the boundary of the unit circle, independently of $||\mathbf{x}||$. Then from Theorem 2.5, an asymptotically minimax design for the correlated error model is attained by randomly sampling design points from ξ_*. As in Example 1 we take a random permutation of design points whose empirical distribution function tends weakly to ξ_*. This may be done as follows.

Define a probability distribution by

$$p_0 = 1 - \frac{[\sqrt{n}]^2}{n}, \quad p_i = \frac{[\sqrt{n}]}{n}, \quad i = 1, 2, \cdots, [\sqrt{n}].$$

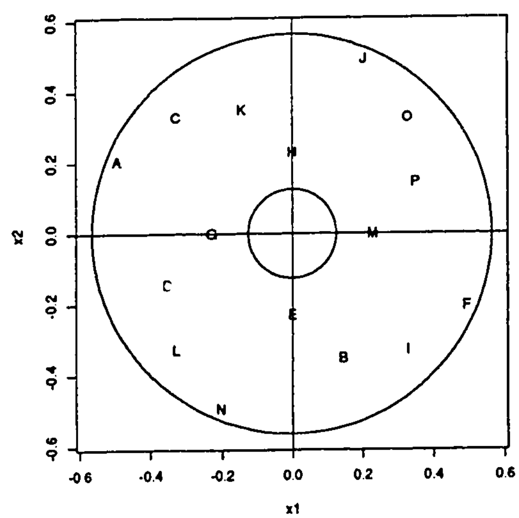
Define

$$a_{-1} = 0, \quad a_i = G_0^{-1}\left(\sum_{j=0}^{i} p_j\right), \quad i = 0, 1, \cdots, [\sqrt{n}],$$

where G_0 is the distribution function corresponding to the optimal density function g_0. Divide the design space $\mathcal{S} = \{\, \mathbf{x} \mid ||\mathbf{x}|| \le r \,\}$ into $[\sqrt{n}] + 1$ annuli $A_0, A_1, \cdots, A_{[\sqrt{n}]}$ with

$$A_i = \{\, \mathbf{x} \mid ||\mathbf{x}|| \in (a_{i-1}, a_i] \,\},$$

Figure 1: Design points for Example 2. Order of implementation is alphabetical.



so that

$$P_{G_0}(\mathbf{x} \in A_i) = p_i.$$

In each A_i we select np_i points equally spaced over $||\mathbf{x}|| = \frac{a_{i-1}+a_i}{2}$. We then have $n - [\sqrt{n}]^2$ points in A_0 and $[\sqrt{n}]$ points in each A_i for $i = 1, \cdots, [\sqrt{n}]$. It is easy to verify that the empirical distribution of the design points tends weakly to ξ_*.

Figure 1 gives a set of random design points for $n = 16$ when $\frac{\sigma^2}{\eta^2} = .01$, $g_0(u) = 7.8477(u - 0.0159/u)^+, \quad 0.126 \leq u \leq 0.564$.

Remark: The implementation of the design of Example 2 requires that the model be transformed in such a way that a spherical design space becomes appropriate. In practice, a transformation to a square design space may be less problematical. For robust minimax designs in the approximately linear model with $\mathbf{z}^T(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$ and $\mathcal{S} = [-0.5, 0.5] \times [-0.5, 0.5]$, see Wiens(1990).

References

Bickel, P.J. and A.M. Herzberg (1979). Robustness of design against autocorrelation in time I: Asymptotic theory, optimality for location and linear regression. *Ann. Statist.* **7**, 77–95.

Bickel, P.J., A.M. Herzberg and M.F. Schilling (1981). Robustness of design against autocorrelation in time II: Optimality, theoretical and numerical results for the first-order autoregressive process. *J. Amer. Statist. Assoc.* **76**, 870–877.

Box, G.E.P. and N. Draper (1959). A basis for the selection of a response surface design. *J. Amer. Statist. Assoc.* **54**, 622–654.

Constantine, G.M. (1989). Robust designs for serially correlated observations. *Biometrika* **76**, 245–251.

Cox, D.R. (1951). Some systematic experimental designs. *Biometrika* **38**, 312–323.

Grenander, U. and M. Rosenblatt (1957). *Statistical Analysis of Stationary Time Series.* Wiley, New York.

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

Kiefer, J. and H. Wynn (1981). Optimum balanced block and latin square designs for correlated observations. *Ann. Statist.* **9**, 737–757.

Kiefer, J. and H. Wynn (1983). Autocorrelation-robust design of experiments. In: T. Leonard and C.-F. Wu, Ed., *Scientific Inference, Data Analysis and Robustness.* Academic Press, New York, 279–299.

Li, K.C. (1984). Robust regression designs when the design space consists of finitely many points. *Ann. Statist.* **12**, 269–282.

Li, K.C. and W. Notz (1982). Robust designs for nearly linear regression. *J. Statist. Plann. Inference* **6**, 135–151.

Liu, S.X. (1994). Some topics in robust estimation and experimental design. Ph.D. thesis, University of Calgary.

Marcus, M.B. and J. Sacks (1976). Robust designs for regression problems. In: S.S. Gupta and D.S. Moore, Eds., *Statistical Decision Theory and Related Topics II.* Academic Press, New York, 245–268.

Notz, W. (1989). Optimal designs for regression models with possible bias. *J. Statist. Plann. Inference* **22**, 43–54.

Pesotchinsky, L. (1982). Optimal robust designs: linear regression in R^k. *Ann. Statist.* **10**, 511–525.

Sacks, J. and D. Ylvisaker (1966). Designs for regression problems with correlated errors. *Ann. Math. Statist.* **37**, 66–89.

Sacks, J. and D. Ylvisaker (1968). Designs for regression problems with correlated errors; many parameters. *Ann. Math. Statist.* **39**, 49–69.

Sacks, J. and D. Ylvisaker (1978). Linear estimation for approximately linear models. *Ann. Statist.* **6**, 1122–1137.

Samarov, A.M. (1987). Robust spectral regression. *Ann. Statist.* **15**, 99–111.

Wiens, D.P. (1992). Minimax design for approximately linear regression. *J. Statist. Plann. Inference.* **31**, 353–371.

Wiens, D.P. (1993). Designs for approximately linear regresssion: Maximizing the minimum coverage probability of confidence ellipsoids. *Cdn. J. Statist.* **21**, 59–70.

Wu, C.-F. (1981). On the robustness and efficiency of some randomized designs. *Ann. Statist.* **9**, 1168–1177.

Ylvisaker, D. (1987). Prediction and design. *Ann. Statist.* **15**, 1–19.

Chapter Three

Minimax Designs for Approximately Linear Models with AR(1) Errors

1 Introduction

In this paper we study optimal - in the minimax sense - designs for linear models under two main departures from the classical assumptions: *i*) the response is taken to be only *approximately* linear, and *ii*) the errors are not assumed to be independent, but to instead follow a first order autoregressive (AR(1)) process.

In Wiens and Zhou (1994), minimax designs were studied for the approximately linear model with errors obeying a very general model of dependence. A main result of that paper is that a design distribution which is asymptotically (minimax) optimal for uncorrelated errors retains its optimality under autocorrelation if the design points are *randomly sampled* from this distribution.

In the aforementioned paper the error process was allowed to vary over a neighbourhood of the uncorrelated error process and was otherwise assumed only to be

[0] A version of this chapter has been submitted for publication. Douglas P. Wiens and Julie Zhou, 1995. *Biometrika.*

weakly stationary. If additional information is available, one can improve on the designs described above.

We consider the following approximately linear model with AR(1) errors.

$$y_i = \theta_0 + \mathbf{x}_i^T \boldsymbol{\theta}_1 + n^{-1/2} f(\mathbf{x}_i) + \epsilon_i, \quad i = 1, \ldots, n,$$

$$\mathbf{x}_i \in \mathcal{S} \subset \mathcal{R}^k, \quad \int_S d\mathbf{x} = 1, \quad \theta_0 \in \mathcal{R}^1, \quad \boldsymbol{\theta}_1 \in \mathcal{R}^k; \tag{1.1}$$

$$E[\epsilon] = \mathbf{0}, \quad COV[\epsilon] = \sigma^2 P_n, \quad P_n(i,j) = \rho^{|i-j|}, \quad -1 < \rho < 1; \tag{1.2}$$

$$f \in \mathcal{F} = \{ \, f \mid \int_S f^2(\mathbf{x}) d\mathbf{x} \le \eta^2, \int_S f(\mathbf{x}) d\mathbf{x} = 0, \int_S \mathbf{x} f(\mathbf{x}) d\mathbf{x} = \mathbf{0} \, \}. \tag{1.3}$$

The purpose of this paper is to derive minimax designs for the model described by (1.1), (1.2) and (1.3), with loss taken to be Integrated Mean Squared Error. This is in analogy to the classical Q-optimality criterion.

For exactly linear models - i.e. $f \equiv 0$ in (1.1) - design problems for (1.2) (or other models of dependence) have been studied by Bischoff (1992,1993), Berenblut and Webb (1974), Constantine (1989), Jenkins and Chanmugam (1962), Kiefer and Wynn (1981, 1984) and Pukelsheim (1993).

The class $\mathcal{F}$ of disturbance functions was used in Huber (1975), Wiens (1991, 1992, 1993, 1994), and Wiens and Zhou (1994). The rationale for using this model of departures from linearity is discussed in Wiens (1992). The normalization of the volume of the design space $\mathcal{S}$ in (1.1), and the $\mathcal{L}_2$-orthogonality of the regressors to the disturbance f in (1.3), may be assumed without any loss of generality. The orthogonality then ensures that the regression parameters are well-defined. The $n^{-1/2}$ rate of decrease of the effect of $f(\mathbf{x})$ on the response is necessary in order that errors due to variance and to bias remain of the same order of magnitude.

For (1.1) and (1.3) with uncorrelated errors, robust (minimax) designs were studied by Huber (1975), and Wiens (1991, 1992, 1993, 1994) using various estimators and various loss functions. Other robust designs for approximately linear models with uncorrelated errors can be found in Box and Draper (1959), Li and Notz (1982), Pesotchinsky (1982) and Liu and Wiens (1994).

Let $\hat{\boldsymbol{\theta}}$ be an estimate of $\boldsymbol{\theta} = (\theta_0, \boldsymbol{\theta}_1^T)^T$. Then the normalized Integrated Mean Squared Error for the response $E[y|\mathbf{x}] = \theta_0 + \mathbf{x}^T\boldsymbol{\theta}_1 + n^{-1/2}f(\mathbf{x})$ is

$$
\begin{aligned}
IMSE_n &= n\int_S E[(\hat{\theta}_0 + \hat{\boldsymbol{\theta}}_1^T\mathbf{x} - E[y|\mathbf{x}])^2]d\mathbf{x} \\
&= \int_S (1, \mathbf{x}^T) MSE_n (1, \mathbf{x}^T)^T d\mathbf{x} + \int_S f^2(\mathbf{x})d\mathbf{x} \\
&= trace(MSE_n \cdot A) + \int_S f^2(\mathbf{x})d\mathbf{x}, \qquad (1.4)
\end{aligned}
$$

where

$$MSE_n = E[(\sqrt{n}\hat{\boldsymbol{\theta}} - \sqrt{n}\boldsymbol{\theta})(\sqrt{n}\hat{\boldsymbol{\theta}} - \sqrt{n}\boldsymbol{\theta})^T] \qquad (1.5)$$

is the Mean Squared Error matrix and $A = \int_S (1 \ \ \mathbf{x}^T)^T(1 \ \ \mathbf{x}^T)d\mathbf{x}$.

We shall focus on the case $k = 1$ (straight line regression). The estimate $\hat{\boldsymbol{\theta}}$ is taken to be either the Best Linear Unbiased Estimate (BLUE) or the Ordinary Least Squares

Estimate (OLSE). In the former case, both *best* and *unbiased* refer only to properties of the estimator in an exactly linear model with a correctly specified autocorrelation process. The experimenter intends to fit a straight line, and anticipates AR(1) errors. The autoregression parameter ρ may be estimated from the data through, e.g. the Cochran-Orcutt procedure. In the latter case, ρ is not estimated. In either case, should the assumption of exact linearity not be realized, the experimenter desires a design that will afford a measure of protection against the consequent increases in IMSE. We note that in order to implement the designs constructed here, only the *sign* of ρ need be known.

As limiting cases we obtain 'most bias-robust' designs as $\frac{\sigma^2}{\eta^2} \rightarrow 0$. Let $\{u_i : i = 1 \ldots n\}$ be points equally spaced over $\mathcal{S}$ and increasing in value. If $\rho < 0$, then this design has $x_i = u_i$, i.e. the experiment is carried out at equispaced, increasing locations. If $\rho > 0$, the design instead has $< x_1, x_2, x_3, x_4, \ldots > = < u_1, u_n, u_2, u_{n-1}, \ldots >$, so that the design points alternate in sign (relative to the centre of $\mathcal{S}$) but decrease (weakly) in absolute magnitude. If instead $\frac{\sigma^2}{\eta^2} \rightarrow \infty$ then the designs of Jenkins and Chanmugam (1962) and Constantine (1989), with all mass at the endpoints of the design interval and with the number of sign changes minimized or maximized as above, are recovered.

In Section 2 below, we obtain the asymptotic values of $IMSE_n$ for the two types of estimates. In Section 3, we construct minimax designs and give strategies to implement these designs.

2 Asymptotics

We use the notation

$$\mathbf{x} = (x_1, \ldots, x_n)^T, \quad \mathbf{1}_n = (1, \ldots, 1)^T, \quad Z = (\mathbf{1}_n, \mathbf{x}),$$

$$\mathbf{y} = (y_1, \ldots, y_n)^T, \quad \mathbf{f} = (f(x_1), \ldots, f(x_n))^T.$$

We first consider the BLUE. Assume that ρ is consistently estimated so that, asymptotically,

$$\hat{\boldsymbol{\theta}}_{BLU} = \left(Z^T P_n^{-1} Z\right)^{-1} Z^T P_n^{-1} \mathbf{y},$$

where

$$P_n^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 & 0 \\ & & & \cdots & & \\ 0 & 0 & 0 & \cdots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}.$$

Then from (1.5),

$$\begin{aligned} MSE_n &= \left(\frac{Z^T P_n^{-1} Z}{n}\right)^{-1} \left(\frac{Z^T P_n^{-1} \mathbf{f}}{n}\right) \left(\frac{Z^T P_n^{-1} \mathbf{f}}{n}\right)^T \left(\frac{Z^T P_n^{-1} Z}{n}\right)^{-1} \\ &\quad + \sigma^2 \left(\frac{Z^T P_n^{-1} Z}{n}\right)^{-1}. \end{aligned}$$

Without loss of generality, the design space $\mathcal{S}$ is taken as $[-0.5, 0.5]$. Then

$$A = \int_{\mathcal{S}} (1\ x)^T (1\ x) dx = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{12} \end{pmatrix}. \tag{2.1}$$

We denote by ξ_n the empirical distribution function of the paired design points (x_i, x_{i+1}), $i = 1, \ldots, n-1$, i.e.

$$\xi_n(v_1, v_2) = \frac{1}{n-1} \sum_{i=1}^{n-1} I_{(x_i \le v_1, x_{i+1} \le v_2)}, \quad v_1,\ v_2 \in \mathcal{S}.$$

The following regularity assumptions are used for the asymptotics.

A1: $\lim_{n\to\infty} \sum_{i=1}^{n} x_i/n = 0$;

A2: ξ_n converges weakly to a design measure ξ on $\mathcal{S} \times \mathcal{S}$;

A3: $|f(x)|$ is bounded on $\mathcal{S}$.

A4: $\frac{Z^T P_n Z}{n}$ converges to a limit $Q(\xi, \rho)$ as $n \to \infty$.

Define

$$\begin{aligned}
b_1(f, \xi, \rho) &= (1-\rho)^2 \int_{\mathcal{S}\times\mathcal{S}} f(x) d\xi(x, v), \\
b_2(f, \xi, \rho) &= \int_{\mathcal{S}\times\mathcal{S}} [(1+\rho^2) x f(x) - \rho f(x) v - \rho x f(v)] d\xi(x, v), \\
r(\xi, \rho) &= \int_{\mathcal{S}\times\mathcal{S}} [(1+\rho^2) x^2 - 2\rho x v] d\xi(x, v), \\
\mathbf{b}(f, \xi, \rho) &= (b_1(f, \xi, \rho), b_2(f, \xi, \rho))^T, \\
B(\xi, \rho) &= \begin{pmatrix} (1-\rho)^2 & 0 \\ 0 & r(\xi, \rho) \end{pmatrix}.
\end{aligned}$$

Straightforward calculations using assumption A1 give

$$\frac{Z^T P_n^{-1} Z}{n} = \frac{1}{1-\rho^2} B(\xi_n, \rho) + o(1),$$

so that A2 and the boundedness of $|(1+\rho^2)x^2 - 2\rho x v|$ on $\mathcal{S}$ give

$$\lim_{n\to\infty} \frac{Z^T P_n^{-1} Z}{n} = \frac{1}{1-\rho^2} B(\xi, \rho). \tag{2.2}$$

In a completely analogous manner, using A3, we obtain

$$\lim_{n\to\infty} \frac{Z^T P_n^{-1} \mathbf{f}}{n} = \frac{1}{1-\rho^2} \mathbf{b}(f, \xi, \rho). \tag{2.3}$$

We define $IMSE(f,\xi,\rho) = \lim_{n\to\infty} IMSE_n$ and $MSE(f,\xi,\rho) = \lim_{n\to\infty} MSE_n$. Combining (2.2) and (2.3), we have

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

Then from (1.4) and (2.1) we obtain

$$\begin{aligned} IMSE_{BLU}(f,\xi,\rho) &= trace\left(MSE(f,\xi,\rho) \cdot \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{12} \end{pmatrix} \right) + \int_S f^2(\mathbf{x})d\mathbf{x} \\ &= \frac{(b_1(f,\xi,\rho))^2}{(1-\rho)^4} + \frac{(b_2(f,\xi,\rho))^2}{12r^2(\xi,\rho)} \\ &\quad + \sigma^2\left(\frac{1+\rho}{1-\rho} + \frac{1-\rho^2}{12r(\xi,\rho)}\right) + \int_S f^2(\mathbf{x})d\mathbf{x}. \end{aligned} \tag{2.4}$$

Similarly, using A4,

$$\begin{aligned} IMSE_{OLS}(f,\xi,\rho) &= (b_1(f,\xi,0))^2 + \frac{(b_2(f,\xi,0))^2}{12r^2(\xi,0)} \\ &\quad + \sigma^2\left(\frac{1+\rho}{1-\rho} + \frac{Q_{22}(\xi,\rho)}{12r^2(\xi,0)}\right) + \int_S f^2(\mathbf{x})d\mathbf{x}. \end{aligned} \tag{2.5}$$

3 Minimax designs

We shall redefine the loss as

$$IMSE'(f,\xi,\rho) = IMSE(f,\xi,\rho) - \sigma^2\frac{1+\rho}{1-\rho} - \int_S f^2(\mathbf{x})d\mathbf{x}. \qquad (3.1)$$

The subtraction of the first term - the variance of the intercept estimate, which cannot be controlled by the design - ensures that the maximum loss is finite in the case $\rho \geq 0$. In the case $\rho \leq 0$ this step is without loss of generality, since in this case both $IMSE'(f,\xi,\rho)$, when evaluated at the optimal design, and $\sigma^2\frac{1+\rho}{1-\rho}$ are maximized at $\rho = 0$. The subtraction of the second term is completely without loss of generality, since $max_{f\in\mathcal{F}}IMSE(f,\xi,\rho)$ is always attained at an f for which $\int_S f^2(\mathbf{x})d\mathbf{x} = \eta^2$.

A minimax design is then a distribution function ξ^* which minimizes the maximum, over $f \in \mathcal{F}$ and all $|\rho| < 1$ of fixed sign, of $IMSE'(f,\xi,\rho)$. Denote by M_1 and M_2 the marginal distributions of ξ:

$$M_1(x_1) = \xi(x_1, 0.5), \qquad M_2(x_2) = \xi(0.5, x_2).$$

Note that

$$\begin{aligned} \xi_n(x,0.5) - \xi_n(0.5,x) &= \frac{1}{n-1}(I_{x_1\leq x} - I_{x_n \leq x}) \\ &\rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned}$$

It follows that $M_1(x) = M_2(x)$ at any continuity point of $M_1(x)$ and $M_2(x)$, $x \in [-0.5, 0.5]$.

We now restrict to the class $\mathcal{C}(\xi)$ of design measures which have absolutely continuous marginal distributions. This condition of absolute continuity of the marginals is required in order that $\sup_{f \in \mathcal{F}} IMSE'(f, \xi, \rho)$ be finite - see Wiens (1992). From the above observation, $M_1(x) = M_2(x)$ for all $x \in [-0.5, 0.5]$. Let $m(x)$ be the common density function.

Before giving the solutions to the problems $\min_{\xi \in \mathcal{C}(\xi)} \max_{f,\rho} IMSE'(f, \xi, \rho)$, we define

$$\begin{aligned} r_0(m) &= \int_S x^2 m(x) dx, \\ \mathcal{L}_*(m) &= \eta^2 \int_S (m(x) - 1)^2 dx + \frac{\sigma^2}{12 r_0(m)}. \end{aligned}$$

Let $m_0(x)$ be a density function on $\mathcal{S}$ which minimizes $\mathcal{L}_*(m)$:

$$\mathcal{L}_*(m) \geq \mathcal{L}_*(m_0) \;\; \textit{for all densities } m(x) \textit{ on } \mathcal{S}. \tag{3.2}$$

We state r main result in the following theorem, whose proof is given in the Appendix.

Theorem 3.1 *Fix the sign of ρ. Suppose that ξ^* has marginal density $m_0(x)$ satisfying (3.2). If as well, for any $f \in \mathcal{F}$,*

$$\int_{S \times S} xv \, d\xi^*(x, v) = -sign(\rho) \int_S x^2 m_0(x) dx, \tag{3.3}$$

$$\begin{aligned} \int_{S \times S} x f(v) d\xi^*(x, v) &= \int_{S \times S} v f(x) d\xi^*(x, v) \\ &= -sign(\rho) \int_S x f(x) m_0(x) dx, \end{aligned} \tag{3.4}$$

then ξ^* is a minimax design for the BLUE. If

$$Q_{22}(\xi^*, \rho) \leq Q_{22}(\xi^*, 0) \text{ for all } \rho \text{ of the given sign,} \tag{3.5}$$

then ξ^* is a minimax design for the OLSE.

The density m_0 defined by (3.2) was obtained by Huber (1975) and is of the form

$$m_0(x) = a(x^2 + b)^+, \quad a \geq 0, \quad x \in [-0.5, 0.5],$$

where a and b are chosen to satisfy the conditions that m_0 be a density function and minimize $\mathcal{L}_*$. They depend on the parameters η^2 and σ^2 only through the ratio $\frac{\sigma^2}{\eta^2}$. Table 1 gives some representative values. Note that $m_0(x) \equiv 0$ on $[-\sqrt{-b}, \sqrt{-b}]$ if $b < 0$.

We now give strategies by which design points may be chosen to satisfy (3.3)–(3.5). For this, let $M_0(x)$ be the distribution function corresponding to $m_0(x)$ and define

$$u_i = M_0^{-1}(\frac{i - 0.5}{n}), \quad i = 1, \ldots, n,$$

with $u_{\frac{n+1}{2}} = 0$ if n is odd. We first discuss the case $\rho < 0$, for which (3.3) and (3.4) require

$$\lim_{n\to\infty} \frac{1}{n-1} \sum_{i=1}^{n-1} x_i x_{i+1} = \int_S x^2 m_0(x) dx, \tag{3.6}$$

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n-1} \sum_{i=1}^{n-1} x_i f(x_{i+1}) &= \lim_{n\to\infty} \frac{1}{n-1} \sum_{i=1}^{n-1} x_{i+1} f(x_i) \\ &= \int_S x f(x) m_0(x) dx \end{aligned} \tag{3.7}$$

Table 1: Constants a and b for minimax marginal density $m_0(x)$

Variance Ratio $\frac{\sigma^2}{\eta^2}$	a	b
0	0	∞
.0001	.0006	1666
.001	.0060	166.6
.01	.0595	16.72
.1	.5580	1.709
1	3.810	.1780
10	15.55	-.0240
100	90.23	-.1487
1000	737.0	-.2136
10000	6886	-.2379
∞	∞	-.25

and (3.5) requires

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i\neq j} x_i x_j \rho^{|i-j|} \leq 0. \tag{3.8}$$

Define design points by

$$x_i = u_i, i = 1, \ldots, n.$$

Then M_0 is the marginal design distribution, asymptotically. To verify (3.6) and (3.7), note that

$$x_{i+1} = x_i + O(n^{-1})$$

for all but at most two values of i, corresponding to the interval on which $M_0(x) \equiv .5$ if $b < 0$. Thus

$$\begin{aligned}\frac{1}{n-1}\sum_{i=1}^{n-1} x_i x_{i+1} &= \frac{1}{n-1}\sum_{i=1}^{n-1} x_i^2 + O(n^{-1}) \\ &\to \int_S x^2 m_0(x)dx, \quad as \quad n \to \infty.\end{aligned}$$

This gives (3.6), the verification of (3.7) is similar. For (3.8) we take $n = 2m$ and find that, upon re-arranging terms, the left-hand side becomes $\lim_{n\to\infty}(L'_n - L''_n)$, where

$$L''_n = \frac{2}{m}\sum_{1\le i<j\le m} u_i u_j \rho^{2m+1-i-j} \to 0, \quad as\ n \to \infty,$$

and

$$\begin{aligned}L'_n &= \frac{2}{m}\sum_{1\le i<j\le m} u_i u_j \rho^{j-i} \\ &= \frac{2}{m}\sum_{k=1}^{m-1} \rho^k c_k\end{aligned}$$

with $c_k = \sum_{l=1}^{m-k} u_l u_{k+l}$. The sequence $\{c_k\}$ is positive and decreasing, and (3.8) follows.

If $\frac{\sigma^2}{\eta^2} = 0$ the optimal density is $m_0(x) = 1$, the continuous uniform design. The design points chosen as above are then equally spaced and (weakly) increasing. As $\frac{\sigma^2}{\eta^2} \to \infty$, the optimal design tends to $P(X = 0.5) = P(X = -0.5) = 0.5$ and the above prescription calls for the first half of the design points to be placed at -0.5, the last half at 0.5 if n is even. If n is odd there is as well a design point $x_{\frac{n+1}{2}} = 0$. We remark that both here and below one may change the signs of all of the design points without affecting the optimality.

Now consider the case $\rho > 0$. In a similar manner to that above we find that (3.3)–(3.5) are satisfied asymptotically by the design with

$$x_i = \begin{cases} u_{\frac{i+1}{2}}, \text{if } i \text{ is odd,} \\ u_{n+1-\frac{i}{2}}, \text{if } i \text{ is even.} \end{cases}$$

The design points then alternate in sign and decrease in magnitude. If $\frac{\sigma^2}{\eta^2} \to 0$ the u_i are equally spaced. In the limit as $\frac{\sigma^2}{\eta^2} \to \infty$, the design tends to $(-0.5, 0.5, -0.5, 0.5, \ldots, -0.5, 0.5)$ for even n, with as well $x_n = 0$ when n is odd.

In the case of even n, the limiting designs as $\frac{\sigma^2}{\eta^2} \to \infty$ were also given by Jenkins and Chanmugam (1962). Constantine (1989) gave these designs as well for the BLUE and MA(1) errors. For the OLSE it is easy to see that Theorem 3.1 also applies to, and yields the optimality of, ξ^* for MA(1) errors.

4 Appendix: Proof of Theorem 3.1

First consider the BLUE. By (3.3) and (3.4) we have

$$b_1(f, \xi^*, \rho) = (1-\rho)^2 \int_S f(x) m_0(x) dx,$$

$$b_2(f, \xi^*, \rho) = (1+|\rho|)^2 \int_S x f(x) m_0(x) dx,$$

$$r(\xi^*, \rho) = (1 + |\rho|)^2 r_0(m_0),$$

so that by (2.4) and (3.1)

$$IMSE'_{BLU}(f, \xi^*, \rho) = (\int_S f(x)m_0(x)dx)^2 + \frac{(\int_S xf(x)m_0(x)dx)^2}{12r_0^2(m_0)} + \frac{\sigma^2(1-|\rho|)}{12(1+|\rho|)r_0(m_0)}.$$

Huber (1981, pp. 247–249) showed that

$$max_{f\in\mathcal{F}}\left((\int_S f(x)m_0(x)dx)^2 + \frac{(\int_S f(x)m_0(x)dx)^2}{12r_0^2(m_0)}\right) = \eta^2 \int_S (m_0(x)-1)^2 dx \quad (4.1)$$

so that

$$\max_{f,\rho} IMSE'_{BLU}(f, \xi^*, \rho) = \max_{f} IMSE'_{BLU}(f, \xi^*, 0) = \mathcal{L}_*(m_0). \quad (4.2)$$

For any other $\xi \in \mathcal{C}(\xi)$, put

$$f_\xi(x) = \eta\{\int_S (m(x)-1)^2 dx\}^{-1/2}(m(x)-1).$$

Note that $\int_S f_\xi(x)dx = 0$, $\int_S xf_\xi(x)dx = 0$, and $\int_S f_\xi^2(x)dx = \eta^2$, so that $f_\xi(x) \in \mathcal{F}$.

Then

$$\begin{aligned} max_{f,\rho} IMSE'_{BLU}(f, \xi, \rho) &\geq IMSE'_{BLU}(f_\xi, \xi, 0) \\ &\geq (b_1(f_\xi, \xi, 0))^2 + \frac{\sigma^2}{12r(\xi, 0)} \\ &= \mathcal{L}_*(m), \end{aligned}$$

and the result follows by (4.2) and (3.2).

For the OLSE, note that $Q_{22}(\xi,0) = r_0(m)$ for $\xi \in C(\xi)$, so that

$$\begin{aligned} max_{f,\rho} IMSE'_{OLS}(f,\xi,\rho) &\geq IMSE'_{OLS}(f_\xi,\xi,0) \\ &\geq \eta^2 \int_S (m(x)-1)^2 dx + \frac{\sigma^2 Q_{22}(\xi,0)}{12 r_C^2(m)} \\ &= \mathcal{L}_*(m). \end{aligned} \tag{4.3}$$

By (4.1) and (3.5) we have

$$max_{f,\rho} IMSE'_{OLS}(f,\xi^*,\rho) = \mathcal{L}_*(m_0)$$

and the result follows by (4.3) and (3.2). □.

References

Berenblut, I.I. and G.I. Webb (1974). Experimental design in the presence of autocorrelated errors. *Biometrika* **61**, 427-437.

Bischoff, W. (1992). On exact D-optimal designs for regression models with correlated observations. *Ann. Inst. Statist. Math.* **44**, 229-238.

Bischoff, W. (1993). On D-optimal designs for linear models under correlated observations with an application to a linear model with multiple response. *J. Statist. Plann. Inference* **37**, 69-80.

Box, G.E.P. and N. Draper (1959). A basis for the selection of a response surface design. *J. Amer. Statist. Assoc.* **54**, 622-654.

Constantine, G.M. (1989). Robust designs for serially correlated observations. *Biometrika* **76**, 245-251.

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

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

Jenkins, G.M. and J. Chanmugam. (1962). The estimation of slope when errors are autocorrelated. *J.R. Statist. Soc. B* **24**, 199-214.

Kiefer, J. and H.P. Wynn (1981). Optimum balanced block and Latin squares designs for correlated observations. *Ann. Statist.* **9**, 737-757.

Kiefer, J. and H.P. Wynn (1984). Optimum and minimax exact treatment designs for one-dimentional autoregressive error process. *Ann. Statist.* **12**, 414-450

Li, K.C. and W. Notz (1982). Robust designs for nearly linear regression. *J. Statist. Plann. Inference* **6**, 135-151.

Liu, S.X. and D.P. Wiens (1994). Robust designs for approximately polynomial regression. Preprint.

Pesotchinsky, L. (1982). Optimal robust designs: Linear regression in R^k. *Ann. Statist.* **10**, 511-525.

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

Wiens, D.P. (1991). Designs for approximately linear regression: two optimality properties of uniform designs. *Statist. Probab. Lett.* **12**, 217-221.

Wiens, D.P. (1992). Minimax designs for approximately linear regression. *J. Statist. Plann. Inference* **31**, 353-371.

Wiens, D.P. (1993). Designs for approximately linear regression: maximizing the minimum coverage probability of confidence ellipsoids. *Cdn. J. Statist.* **21**, 59-70.

Wiens, D.P. (1994). Robust designs for approximately linear regression: M-estimated parameters. *J. Statist. Plann. Inference* **40**, 135-160.

Wiens, D.P. and J. Zhou (1994). Minimax regression designs for approximately linear models with autocorrelated errors. Preprint.

Chapter Four

Robust Designs Based on the Infinitesimal Approach

1 Introduction

In this paper we introduce an *infinitesimal* approach to the construction of robust designs for linear models. These designs are robust against small departures from the assumed linear regression response and/or small departures from the usual assumption of uncorrelated errors. Subject to satisfying a robustness constraint, they minimize the mean squared error of the estimator at the ideal model.

The infinitesimal approach and the minimax approach are two basic approaches in robust *estimation* theory. The infinitesimal approach was first introduced by F. Hampel (see Hampel (1974)), and the minimax approach was proposed by P. Huber (see Huber (1964)). In robust *design* theory, the minimax approach has been adopted by, e.g., Huber (1975), Kiefer and Wynn (1984), Li and Notz (1982), Liu (1994), Pesotchinsky (1982), Tang (1993), Wiens (1991, 1992), Wiens and Zhou (1994, 1995).

[0] A version of this chapter has been submitted for publication. Douglas P. Wiens and Julie Zhou, 1995. *Journal of the American Statistical Association.*

It is perhaps surprising that the infinitesimal approach has not been previously investigated with regard to robust design theory.

The infinitesimal robustness of an estimator can be investigated by means of the Influence Function (IF) of the estimator and by the Change-of-Variance Function (CVF) - see Hampel et al (1986) and more recently Hössjer (1991). While the IF provides an intuitive measure of the local robustness of the value of an estimator, the CVF quantifies the local stability of the variance.

The bias and the variance of the resulting estimates are two important measures in the process of choosing an optimal design. If the true nature of the model is unknown, it is important not only that the bias and variance be small at the ideal model, but that they remain small in a neighbourhood of this model. To this end, we introduce the Change-of-Bias Function (CBF) and the Change-of-Variance Function (CVF) as basic components of an infinitesimal approach to robust design theory. The CBF (resp., CVF) is the Gateaux derivative, evaluated at the ideal model, of the bias (resp., variance) functional of the estimator in the direction of a contaminating response function (resp., autocorrelation structure). Based on the CBF and the CVF, various optimality criteria are proposed. Correspondingly optimal designs are obtained.

An outline of the paper is as follows. In Section 2, the CBF and the CVF are defined for approximately linear models with correlated errors:

$$\left.\begin{array}{l} y_i = \mathbf{z}^T(\mathbf{x}_i)\theta + f(\mathbf{x}_i) + \epsilon_i, \quad i = 1, \ldots, n; \\ \mathbf{x}_i \in \mathcal{S} \subset \mathcal{R}^q, \ \mathbf{z}(\mathbf{x}_i) \in \mathcal{R}^p, \ \boldsymbol{\theta} \in \mathcal{R}^p; \\ E[\epsilon] = \mathbf{0}, \quad COV[\epsilon] = \sigma^2\mathbf{P}. \end{array}\right\} \tag{1.1}$$

In (1.1), $f(\mathbf{x})$ is assumed to be constrained in a manner ensuring that the regression parameters are well-defined - see Section 5 below.

Roughly speaking, the CBF measures the rate at which the bias of the regression estimates changes as $f(\mathbf{x})$ departs from the zero function. Similarly the CVF measures the rate of change in the variance as the autocorrelation matrix $\mathbf{P}$ departs from the identity matrix. The Change-of-Bias Sensitivity (CBS) is defined as an upper bound for the CBF when the disturbance function f varies over a neighbourhood. Similarly the Change-of-Variance Sensitivity (CVS) is defined as an upper bound for the CVF when $\mathbf{P}$ varies.

In Section 3, three classes of optimality criteria for robust designs are given. We define a *V-robust* design as one which minimizes a scalar valued function of the mean squared error matrix, subject to the constraint that the CVS not exceed a pre-assigned bound, and a *most V-robust* design as a V-robust design for which the minimum upper bound is attained. A V-robust design is robust against increases in the variance caused by autocorrelated errors. In an analogous manner, based on the CBF we define *B-robust* and *most B-robust* designs. A B-robust design is robust against biases caused by departures from the assumed regression response. Finally, *M-robust* designs, for which bounds on both the CBS and CVS are given, are proposed. These designs are robust against both of the above types of model violations.

In Sections 4 to 6, V-, B- and M- robust designs are obtained. Explicit V-robust designs for two particular error processes are constructed in Section 4. In Section 5 we note that B-robust designs coincide with the *Bounded Bias* designs of Liu (1994) and Liu and Wiens (1994). Examples with respect to two different neighbourhood structures for f are given. The marginals of the M-robust design distributions studied in Section 6 coincide with those for B-robust designs. However, the order in which the design points are allocated determines the robustness against autocorrelated errors.

2 The CVF and CBF for designs

Let ξ be the design measure, i.e. the empirical distribution function of the design points, and define

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}^T(\mathbf{x}_1) \\ \mathbf{z}^T(\mathbf{x}_2) \\ \vdots \\ \mathbf{z}^T(\mathbf{x}_n) \end{bmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} f(\mathbf{x}_1) \\ f(\mathbf{x}_2) \\ \vdots \\ f(\mathbf{x}_n) \end{pmatrix},$$

$$\mathbf{B}_\xi(m) = \frac{1}{n}\sum_{i=1}^{n-m} \mathbf{z}(\mathbf{x}_i)\mathbf{z}^T(\mathbf{x}_{i+m}), \quad for \quad 0 \le m \le n-1,$$

$$\mathbf{B}_\xi(m) = \mathbf{B}_\xi^T(-m), \quad for \quad m < 0,$$

$$\mathbf{b}_{f,\xi} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}(\mathbf{x}_i) f(\mathbf{x}_i).$$

In the hopes that $f \equiv 0$ and $\mathbf{P} = \mathbf{I}$ in (1.1), the experimenter will compute the Ordinary Least Squares Estimator $\hat{\boldsymbol{\theta}}_{LS} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y}$. Should these hopes not be realized, a design is desired that will afford a measure of protection against the consequent increases in bias and/or variance.

As a measure of the accuracy of $\hat{\boldsymbol{\theta}}_{LS}$ we shall use a scalar-valued function of the mean squared error matrix

$$\begin{aligned} MSE(\sqrt{n}\hat{\boldsymbol{\theta}}_{LS}, f, \xi, \mathbf{P}) &= E[(\sqrt{n}\hat{\boldsymbol{\theta}}_{LS} - \sqrt{n}\boldsymbol{\theta})(\sqrt{n}\hat{\boldsymbol{\theta}}_{LS} - \sqrt{n}\boldsymbol{\theta})^T] \\ &= \sigma^2 \mathbf{B}_\xi^{-1}(0)\frac{\mathbf{Z}^T\mathbf{P}\mathbf{Z}}{n}\mathbf{B}_\xi^{-1}(0) + n\mathbf{B}_\xi^{-1}(0)\mathbf{b}_{f,\xi}\mathbf{b}_{f,\xi}^T\mathbf{B}_\xi^{-1}(0). \end{aligned} \quad (2.1)$$

The first term in (2.1) is the variance-covariance matrix, and the second arises from the bias. The MSE depends on the autocorrelation matrix $\mathbf{P}$ only through the variance-covariance matrix, and on the disturbance function f only through the bias.

For a convex class $\mathcal{F}$ of disturbance functions and a convex class $\mathcal{P}$ of autocorrelation matrices, define

$$\begin{aligned} f_s &= (1-s)f_0 + sf_1, \; f_0 \equiv 0, \; f_1 \in \mathcal{F}, \\ \mathbf{P}_t &= (1-t)\mathbf{P}_0 + t\mathbf{P}_1, \; \mathbf{P}_0 = \mathbf{I}, \; \mathbf{P}_1 \in \mathcal{P}, \end{aligned}$$

where $0 \le s, t \le 1$.

2.1 Change-of-Variance Function

Let $\mathcal{L}$ be a non-negative, twice differentiable function of p.s.d. matrices, increasing with respect to the usual ordering by positive definiteness.

Definition 2.1 *The Change-of-Variance Function $CVF(\xi, \mathbf{I}, \mathbf{P}_1)$ for a design ξ at* $\mathbf{I}$ *in the direction* $\mathbf{P}_1$ *is*

$$CVF(\xi, \mathbf{I}, \mathbf{P}_1) = \frac{\partial}{\partial t}\mathcal{L}(MSE(\sqrt{n}\hat{\theta}_{LS}, f_0, \xi, \mathbf{P}_t))_{|_{t=0}}.$$

The CVF measures the rate at which the MSE changes under infinitesimal departures from the ideal model, in the direction of a particular autocorrelation structure.

We take $\mathcal{L} = determinant$ throughout this paper, and find

$$CVF(\xi, \mathbf{I}, \mathbf{P}) = \sigma^{2p} det(\mathbf{B}_\xi^{-1}(0)) \cdot trace(\frac{\mathbf{Z}^T(\mathbf{P}-\mathbf{I})\mathbf{Z}}{n}\mathbf{B}_\xi^{-1}(0)). \tag{2.2}$$

Of particular interest is the supremum of CVF, as $\mathbf{P}$ varies.

Definition 2.2 *Let $\mathcal{P}$ be a convex class of autocorrelation matrices. The Change-of-Variance Sensitivity $CVS(\xi, \mathbf{I})$ of ξ over $\mathcal{P}$ is*

$$CVS(\xi, \mathbf{I}) = \sup_{\mathbf{P}\in\mathcal{P}} \frac{CVF(\xi, \mathbf{I}, \mathbf{P})}{\mathcal{L}(MSE(\sqrt{n}\hat{\theta}_{LS}, f_0, \xi, \mathbf{I}))}.$$

The normalizing matrix in the denominator is

$$MSE(\sqrt{n}\hat{\theta}_{LS}, f_0, \xi, \mathbf{I}) = \sigma^2 \mathbf{B}_\xi^{-1}(0), \tag{2.3}$$

the covariance matrix of $\sqrt{n}\hat{\theta}_{LS}$ under uncorrelated errors. From this and (2.2), we find that for $\mathcal{L}$ = *determinant*,

$$CVS(\xi, \mathbf{I}) = \sup_{\mathbf{P} \in \mathcal{P}} \left\{ trace \left(\frac{\mathbf{Z}^T(\mathbf{P} - \mathbf{I})\mathbf{Z}}{n} \mathbf{B}_\xi^{-1}(0) \right) \right\}. \quad (2.4)$$

2.2 Change-of-Bias Function

Definition 2.3 *The Change-of-Bias Function $CBF(\xi, f_1)$ of ξ in the direction f_1 is*

$$CBF(\xi, f_1) = \frac{1}{2}\frac{\partial^2}{\partial s^2}\mathcal{L}(MSE(\sqrt{n}\hat{\theta}_{LS}, sf_1, \xi, \mathbf{I}))_{|_{s=0}}.$$

The use of the second derivative is motivated by the observation that $MSE(\sqrt{n}\hat{\theta}_{LS}, sf_1, \xi, \mathbf{I})$ is a linear function of s^2. The CBF measures changes in MSE due to increased bias, as one moves away from the ideal model in the direction of a particular disturbance function. For $\mathcal{L}$ = *determinant*,

$$CBF(\xi, f) = \sigma^{2p} det(\mathbf{B}_\xi^{-1}(0)) \cdot \frac{n}{\sigma^2}\mathbf{b}_{f,\xi}^T \mathbf{B}_\xi^{-1}(0)\mathbf{b}_{f,\xi}.$$

Note that the CBF is always positive, whereas the CVF may be negative. This reflects the feeling that one generally wants a design which *decreases* the variance of the estimator.

If f varies over a class $\mathcal{F}$, we define the Change-of-Bias Sensitivity as the supremum of CBF.

Definition 2.4 *Let $\mathcal{F}$ be a convex class of disturbance functions. The Change-of-Bias Sensitivity $CBS(\xi)$ of ξ over $\mathcal{F}$ is*

$$CBS(\xi, f_0) = \sup_{f \in \mathcal{F}} \frac{CBF(\xi, f)}{\mathcal{L}(MSE(\sqrt{n}\hat{\theta}_{LS}, f_0, \xi, \mathbf{I}))}.$$

For $\mathcal{L}$ = *determinant*,

$$CBS(\xi, f_0) = \sup_{f \in \mathcal{F}} \left\{ \frac{n}{\sigma^2} \mathbf{b}_{f,\xi}^T \mathbf{B}_\xi^{-1}(0) \mathbf{b}_{f,\xi} \right\}. \tag{2.5}$$

3 Optimality criteria

3.1 V-robustness

For given α, we say that a design ξ is *V-robust* if it minimizes (2.3) subject to the constraint

$$CVS(\xi, \mathbf{I}) \leq \alpha. \tag{3.1}$$

If α is the infimum of the CVS over a given class of designs, then the V-robust design is called *most V-robust* in this class. In Section 4 we construct V-robust and most V-robust designs for two classes $\mathcal{P}$.

3.2 B-robustness

For given β, we say that a design ξ is *B-robust* if it minimizes (2.3) subject to the constraint

$$CBS(\xi, f_0) \leq \beta. \tag{3.2}$$

If β is the infimum of the CBS over a given class of designs, then the B-robust design is called *most B-robust* in this class. In Section 5 we note that when $\mathcal{L} = determinant$, B-robust designs coincide with the *Bounded Bias* designs of Liu (1994) and Liu and Wiens (1994). Examples are given for two classes $\mathcal{F}$.

3.3 M-robustness

We say that a design is *M-robust* if it minimizes (2.3) subject to both (3.1) and (3.2), and *most M-robust* if it is both most V-robust and most B-robust. We consider M-robust designs in Section 6. It is not known (to us) if most M-robust designs exist.

4 V-robust designs

Let Ξ be a given class of designs and let ρ be the autocorrelation function corresponding to $\mathbf{P} \in \mathcal{P}$. Upon substituting

$$\mathbf{P}(i,j) = \rho(|i-j|), \quad i,j = 1, \ldots, n$$

into (2.4) and using (2.3), the V-robust design problem for $\{\Xi, \mathcal{P}\}$ becomes

$$\max_{\xi \in \Xi} |\mathbf{B}_\xi(0)|, \quad \textit{subject to}$$

$$CVS(\xi, \mathbf{I}) = sup_{\mathbf{P} \in \mathcal{P}} \{ \sum_{0<|s|\leq n-1} \rho(s)\ trace(\mathbf{B}_\xi(s)\mathbf{B}_\xi^{-1}(0)) \} \leq \alpha. \qquad (4.1)$$

In this section we obtain V-robust and most V-robust designs for the classes

$$\mathcal{P}_1 = \{\mathbf{P} |\ \rho(s) = 0\ for\ |s| \geq 2\ and\ c_0 \leq \rho(1) < 1\ with\ c_0 > 0\}$$

$$\mathcal{P}_2 = \{\mathbf{P} |\ \rho(s) = 0\ for\ |s| \geq 2\ and\ -1 < \rho(1) \leq -c_1\ with\ c_1 > 0\}.$$

These classes correspond to MA(1) processes with, respectively, positive and negative lag-1 correlations bounded away from 0. The most V-robust designs presented here do not depend on the values of c_0 and c_1. We consider straight line regression, i.e. $q = 1, p = 2$ and $\mathbf{z}(x) = (1, x)^T$ in (1.1). Denote by $\mathbf{x}$ the vector $(x_1, \ldots, x_n)^T$ of design points. Without loss of generality we take $\mathcal{S} = [-0.5, 0.5]$ as design space. We restrict to the class Ξ of designs with $\sum_{i=1}^n x_i = 0$.

For this model we find that

$$\begin{aligned} \mathbf{B}_\xi(0) &= diag\left(1, \frac{\sum_{i=1}^n x_i^2}{n}\right), \\ diag(\mathbf{B}_\xi(1) + \mathbf{B}_\xi(-1)) &= \left(\frac{2(n-1)}{n}, \frac{2\sum_{i=1}^{n-1} x_i x_{i+1}}{n}\right), \end{aligned}$$

so that

$$CVS(\xi, \mathbf{I}) = \sup_{\rho(1)} 2\rho(1) \left(\frac{n-1}{n} + F(\mathbf{x})\right),$$

where

$$F(\mathbf{x}) := \frac{\sum_{i=1}^{n-1} x_i x_{i+1}}{\sum_{i=1}^{n} x_i^2} = \frac{\mathbf{x}^T\mathbf{M}\mathbf{x}}{\mathbf{x}^T\mathbf{x}},$$

and $\mathbf{M}$ is the $n \times n$ tridiagonal matrix with $(i,j)^{th}$ element $m_{ij} = \frac{1}{2}I(|i-j|=1)$.

For $\mathcal{P}_1$ the CVS is then

$$CVS_1(\xi,\mathbf{I}) = \begin{cases} 2c_0\left(\frac{n-1}{n} + F(\mathbf{x})\right), & if\ F(\mathbf{x}) \le -\frac{n-1}{n}, \\ 2\left(\frac{n-1}{n} + F(\mathbf{x})\right), & if\ F(\mathbf{x}) > -\frac{n-1}{n}. \end{cases}$$

while for $\mathcal{P}_2$ it is

$$CVS_2(\xi,\mathbf{I}) = \begin{cases} -2c_1(\frac{n-1}{n} + F(\mathbf{x})), & if\ F(\mathbf{x}) > -\frac{n-1}{n}, \\ -2(\frac{n-1}{n} + F(\mathbf{x})), & if\ F(\mathbf{x}) \le -\frac{n-1}{n}. \end{cases}$$

Propositions 4.1 and 4.2 below show that if α is sufficiently large, then (4.1) imposes no restriction.

Proposition 4.1 *Let n be even.*

(i) If $\alpha \ge 0$ a V-robust design for $\mathcal{P}_1$ is

$$\mathbf{x} = < 0.5, -0.5, 0.5, -0.5, \ldots, 0.5, -0.5 >$$

with $CVS_1(\xi,\mathbf{I}) = 0$. This design minimizes $CVS_1(\xi,\mathbf{I})$ among those designs in Ξ which maximize $|\mathbf{B}_\xi(0)| = \frac{\mathbf{x}^T\mathbf{x}}{n}$.

(ii) If $\alpha \ge -4c_1\frac{n-2}{n}$ a V-robust design for $\mathcal{P}_2$ is

$$\mathbf{x} = < \underbrace{0.5, \ldots, 0.5}_{n/2}, \underbrace{-0.5, \ldots, -0.5}_{n/2} >$$

with $CVS_2(\xi, \mathbf{I}) = -4c_1\frac{n-2}{n}$. *This design minimizes* $CVS_2(\xi, \mathbf{I})$ *among those designs in* Ξ *which maximize* $|\mathbf{B}_\xi(0)| = \frac{\mathbf{x}^T\mathbf{x}}{n}$.

Proof: Each design maximizes $\mathbf{x}^T\mathbf{x}$ unconditionally, and has the stated value of $CVS(\xi, \mathbf{I})$. That each design minimizes $CVS(\xi, \mathbf{I})$ among designs in Ξ which maximize $\mathbf{x}^T\mathbf{x}$ is obvious. □

Remark: If $(x_1, \ldots, x_n)$ is a V-robust design, then $-(x_1, \ldots, x_n)$ is also V-robust.

The designs in Proposition 4.1 are given in Jenkins and Chanmugam (1962) and Constantine (1989). Jenkins and Chanmugam (1962) minimize $var(\hat{\theta}_{LS}) = (X^TX)^{-1}X^T\mathbf{P}X(X^TX)^{-1}$ over a subset of the set of all possible designs carried out at two levels only. The subset consists of all discrete 'square wave' designs in which a run of m experiments at one level is followed by m experiments at the other and so on. Then the design problem becomes that of choosing the block size m. In Constantine (1989), the goal is to maximize $trace((COV(\hat{\theta}_{BLUE}))^{-1}) = trace(X^T\mathbf{P}^{-1}X)$ over Ξ. A linear approximation to $X^T\mathbf{P}^{-1}X$ is used to derive the optimal design.

Proposition 4.2 *Let n be odd.*

(i) If $\alpha \geq \frac{2}{n(n-1)}$ *a V-robust design for* $\mathcal{P}_1$ *is*

$$\mathbf{x} = < 0.5, -0.5, 0.5, -0.5, \ldots, 0.5, -0.5, 0 >$$

with $CVS_1(\xi, \mathbf{I}) = \frac{2}{n(n-1)}$. *This design minimizes* $CVS_1(\xi, \mathbf{I})$ *among those designs in* Ξ *which maximize* $|\mathbf{B}_\xi(0)| = \frac{\mathbf{x}^T\mathbf{x}}{n}$.

(ii) If $\alpha \geq -2c_1 \frac{2n^2-5n+1}{n(n-1)}$ a V-robust design for $\mathcal{P}_2$ is

$$\mathbf{x} = < \underbrace{0.5, \ldots, 0.5}_{(n-1)/2}, 0, \underbrace{-0.5, \ldots, -0.5}_{(n-1)/2} >$$

with $CVS_2(\xi, \mathbf{I}) = -2c_1 \frac{2n^2-5n+1}{n(n-1)}$. This design minimizes $CVS_2(\xi, \mathbf{I})$ among those designs in Ξ which maximize $|\mathbf{B}_\xi(0)| = \frac{\mathbf{x}^T\mathbf{x}}{n}$.

Proof: First we show that each design maximizes $\mathbf{x}^T\mathbf{x}$ in Ξ. This seemingly obvious result does not appear to be in the literature.

Let $\mathbf{x}^*$ be any one of the V-robust designs and $\mathbf{x}^0$ be any other design in Ξ. By replacing $\mathbf{x}^0$ by $-\mathbf{x}^0$ and permuting the elements if necessary we may assume that $x_i^0 \geq 0$ for $1 \leq i \leq m$, $x_i^0 < 0$ for $m+1 \leq i \leq n$ for some $m \geq \frac{n+1}{2}$. Since $|x_i^0| \leq .5$ and $\sum x_i^0 = 0$ we have

$$\begin{aligned} \sum_{i=1}^{n} (x_i^0)^2 &\leq \sum_{i=1}^{n} \frac{1}{2} |x_i^0| \\ &= \frac{1}{2} \left(\sum_{i=1}^{m} x_i^0 - \sum_{i=m+1}^{n} x_i^0 \right) \\ &= -\sum_{i=m+1}^{n} x_i^0 \leq (n-m) \cdot \frac{1}{2} \\ &\leq \frac{n-1}{4} = \sum_{i=1}^{n} (x_i^*)^2. \end{aligned}$$

It is easy to check that $CVS(\xi, \mathbf{I})$ is as stated for each design, and is a minimum among designs which maximize $\mathbf{x}^T\mathbf{x}$. □

In order to get the most V-robust designs, it is necessary to evaluate the minimum values of $CVS_1(\xi, \mathbf{I})$ and $CVS_2(\xi, \mathbf{I})$ over Ξ. This requires an investigation of the extrema of $F(\mathbf{x})$ over the hyperplane $\mathbf{x}^T\mathbf{1}_n = 0$. These extrema turn out to correspond to characteristic roots of $\mathbf{M}$.

Lemma 4.3 *For even n, the matrix* $\mathbf{M}$ *has characteristic roots*

$$\lambda_1 > \ldots > \lambda_{\frac{n}{2}} > 0 > \lambda_{\frac{n}{2}+1} > \ldots > \lambda_n,$$

and corresponding orthonormal characteristic vectors $\mathbf{x}_j$*, given by*

$$\lambda_j = \cos\frac{j\pi}{n+1}, \quad j = 1, \ldots, n; \tag{4.2}$$

$$(\mathbf{x}_j)_k = \sqrt{\frac{2}{n+1}}\sin\frac{kj\pi}{n+1}, \quad k = 1, \ldots, n. \tag{4.3}$$

We have

$$min_{\mathbf{x}\perp\mathbf{1}}\frac{\mathbf{x}^T\mathbf{M}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \mathbf{x}_n^T\mathbf{M}\mathbf{x}_n = \lambda_n, \tag{4.4}$$

$$max_{\mathbf{x}\perp\mathbf{1}}\frac{\mathbf{x}^T\mathbf{M}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \mathbf{x}_2^T\mathbf{M}\mathbf{x}_2 = \lambda_2. \tag{4.5}$$

Proof: The verification of (4.2) and (4.3) is straightforward. We then calculate

$$\mathbf{x}_j^T\mathbf{1} = \sqrt{\frac{2}{n+1}}\cot\frac{j\pi}{2(n+1)} \cdot I(j\ odd). \tag{4.6}$$

Then (4.4) is immediate, since $\mathbf{x}_n$ is orthogonal to $\mathbf{1}$ and is the unconditional minimizer. For (4.5), let $\mathbf{x} \perp \mathbf{1}$. Define

$$u_j = \mathbf{x}_j^T\mathbf{x}, \quad j = 1, 3, \ldots, n-1; \quad c^2 = \sum_{j\geq 3} u_j^2 \quad (= 0\ if\ n = 2);$$

$$v_j = \mathbf{x}_j^T\mathbf{x}, \quad j = 2, 4, \ldots, n; \quad w_j = \mathbf{x}_j^T\mathbf{1}, \quad j = 1, 3, \ldots, n-1.$$

By (4.6) we have $\sum w_j^2 = n$ and $\sum u_j w_j = 0$ so that, applying the Cauchy-Schwarz inequality,

$$u_1^2 = \frac{(-\sum_{j\geq 3} u_j w_j)^2}{w_1^2} \leq c^2 \left(\frac{n}{w_1^2} - 1\right).$$

Then

$$\begin{aligned} \mathbf{x}^T(\mathbf{M} - \lambda_2 \mathbf{I})\mathbf{x} &= \sum_{j=1}^{n} (\lambda_j - \lambda_2)(\mathbf{x}_j^T \mathbf{x})^2 \\ &= (\lambda_1 - \lambda_2)u_1^2 + \sum_{j\geq 3} (\lambda_j - \lambda_2)u_j^2 + \sum_{j\geq 2} (\lambda_j - \lambda_2)v_j^2. \end{aligned}$$

This is zero if $n = 2$; if $n \geq 4$ it is

$$\leq (\lambda_1 - \lambda_2)c^2 \left(\frac{n}{w_1^2} - 1\right) + c^2(\lambda_3 - \lambda_2),$$

which is ≤ 0 as long as

$$w_1^2 \geq \frac{n(\lambda_1 - \lambda_2)}{\lambda_1 - \lambda_3}.$$

This is easily checked, so that

$$\mathbf{x}^T(\mathbf{M} - \lambda_2 \mathbf{I})\mathbf{x} \leq 0 \quad \textit{for all} \quad \mathbf{x} \perp \mathbf{1}.$$ □

Lemma 4.4 *For odd n, the matrix $\mathbf{M}$ has characteristic roots*

$$\lambda_1 > \ldots > \lambda_{\frac{n-1}{2}} > \lambda_{\frac{n+1}{2}} = 0 > \lambda_{\frac{n+3}{2}} > \ldots > \lambda_n,$$

and corresponding orthonormal characteristic vectors $\mathbf{x}_j$, given by (4.2) and (4.3).

We have

$$\begin{aligned} min_{\mathbf{x}\perp\mathbf{1}} \frac{\mathbf{x}^T\mathbf{M}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} &= \mathbf{x}_{n-1}^T \mathbf{M}\mathbf{x}_{n-1} = \lambda_{n-1}, \\ max_{\mathbf{x}\perp\mathbf{1}} \frac{\mathbf{x}^T\mathbf{M}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} &= \mathbf{x}_2^T \mathbf{M}\mathbf{x}_2 = \lambda_2. \end{aligned}$$

The proof is similar to that of Lemma 4.3 and is omitted.

Theorem 4.5 *With notation as above, we have*

$$\min_{\xi\in\Xi} CVS_1(\xi,\mathbf{I}) = \begin{cases} 2c_0\cdot\left(cos\frac{n\pi}{n+1}+\frac{n-1}{n}\right), & n \ \ even, \\ 2c_0\cdot\left(cos\frac{(n-1)\pi}{n+1}+\frac{n-1}{n}\right), & n \ \ odd. \end{cases}$$

The most V-robust design for $\mathcal{P}_1$ - unique up to sign changes - is given by

$$x_i^* = \begin{cases} (-1)^{i+1}\frac{sin\frac{\pi i}{n+1}}{sin\frac{\pi(\frac{n}{2}+1)}{n+1}}\cdot 0.5, & i=1,\ldots,n, \quad n \ \ even, \\ \\ (-1)^{i+1}\frac{sin\frac{2\pi i}{n+1}}{sin\frac{2[\frac{n+1}{4}]\pi}{n+1}}\cdot 0.5, & i=1,\ldots,n, \quad n \ \ odd. \end{cases}$$

For $\mathcal{P}_2$, we have

$$\min_{\xi\in\Xi} CVS_2(\xi,\mathbf{I}) = -2c_1\cdot\left(cos\frac{2\pi}{n+1}+\frac{n-1}{n}\right).$$

The most V-robust design for $\mathcal{P}_2$ - unique up to sign changes - is given by

$$x_i^{**} = \begin{cases} \frac{sin\frac{2\pi i}{n+1}}{sin\frac{\pi(\frac{n}{2}+1)}{n+1}}\cdot 0.5, & i=1,\ldots,n, \quad n \ \ even, \\ \\ \frac{sin\frac{2\pi i}{n+1}}{sin\frac{2[\frac{n+1}{4}]\pi}{n+1}}\cdot 0.5, & i=1,\ldots,n, \quad n \ \ odd. \end{cases}$$

Proof: For $\mathcal{P}_1$ we are to maximize $\mathbf{x}^T\mathbf{x}$, subject to $F(\mathbf{x})$ being a minimum. But by Lemmas 4.3 and 4.4 any minimizer of $F(\mathbf{x})$ is a scalar multiple $c\mathbf{x}^*$, and then $max_i|cx_i^*| \leq .5$ requires $|c| \leq 1$. For such c, the maximum is attained at $c = \pm 1$. The proof for $\mathcal{P}_2$ is similar. □

Figure 1: M st V-robust designs for $\mathcal{P}_1$ and $\mathcal{P}_2$; $n = 9, 10, 19, 20$. Design points x_i plotted against the index i.

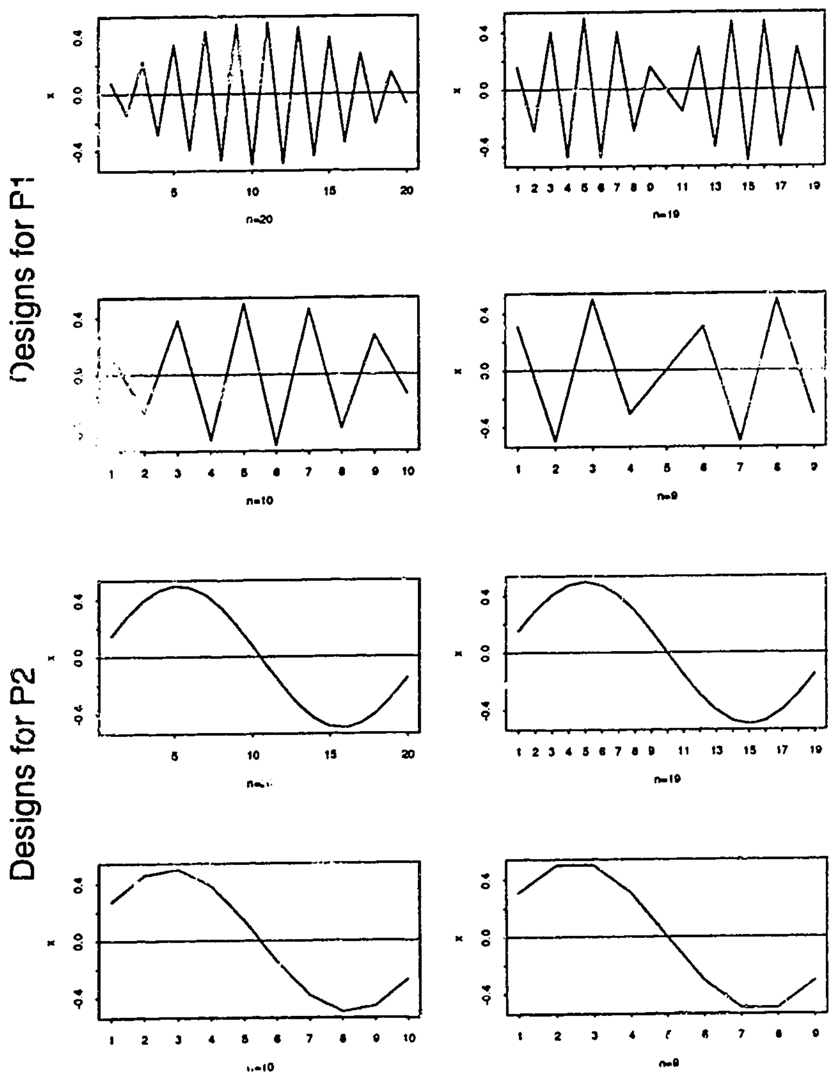


Table 1: Most V-robust designs for $\mathcal{P}_1$ and $\mathcal{P}_2$; $n = 9, 10, 19, 20$.

$\mathcal{P}$	n	Most V-robust designs $x_1, x_2, \cdots, x_n$
$\mathcal{P}_1$	9	0.309, -0.5, 0.5, -0.309, 0, 0.309, -0.5, 0.5, -0.309
	10	0.142, -0.273, 0.382, -0.459, 0.5, -0.5, 0.459, -0.382, 0.273, -0.142
	19	0.155, -0.294, 0.405, -0.476, 0.5, -0.476, 0.405, -0.294, 0.155, 0, -0.155, 0.294, -0.405, 0.476, -0.5, 0.476, -0.405, 0.294, -0.155
	20	0.075, -0.148, 0.218, -0.282, 0.341, -0.392, 0.434, -0.467, 0.489, -0.5, 0.5, -0.489, 0.467, -0.434, 0.392, -0.341, 0.282, -0.218, 0.148, -0.075
$\mathcal{P}_2$	9	0.309, 0.5, 0.5, 0.309, 0, -0.309, -0.5, -0.5, -0.309
	10	0.273, 0.459, 0.5, 0.382, 0.142, -0.142, -0.382, -0.5, -0.459, -0.273
	19	0.155, 0.294, 0.405, 0.476, 0.5, 0.476, 0.405, 0.294, 0.155, 0, -0.155 -0.294, -0.405, -0.476, -0.5, -0.476, -0.405, -0.294, -0.155
	20	0.148, 0.282, 0.392, 0.467, 0.5, 0.489, 0.434, 0.341, 0.218, 0.075, -0.075, -0.218, -0.341, -0.434, -0.489, -0.5, -0.467, -0.392, -0.282, -0.148

Remarks:

1. The designs $\mathbf{x}^*$ and $\mathbf{x}^{**}$ are also most V-robust for regression through the origin with respect to $\mathcal{P}_1$ and $\mathcal{P}_2$ respectively, and then

$$min_\xi CVS_1(\xi, \mathbf{I}) = 2c_0 \left(cos\frac{n\pi}{n+1} \cdot I(n\ even) + cos\frac{(n-1)\pi}{n+1} \cdot I(n\ odd) \right)$$

and $min_\xi CVS_2(\xi, \mathbf{I}) = -2c_1 cos\frac{2\pi}{n+1}$. The proof of this entails only minor modifications to those above, and so is omitted.

2. For each n, the unordered vectors x^* and $\mathbf{x}^{**}$ are equal - the designs differ only with respect to the order in which the points are allocated.

3. The empirical distribution functions of the design points of the most V-robust designs above are easily seen to converge weakly, as $n \to \infty$, to the *arcsine* distribution with distribution function $\frac{1}{2} + \frac{sin^{-1}(2x)}{\pi}$ and density $\frac{2}{\pi\sqrt{1-4x^2}}$, $|x| \leq \frac{1}{2}$. An asymptotic description of the V-robust designs is then given by this limit together with the order in which the design points are to be applied. For $\mathcal{P}_1$ the design points alternate in sign. For n even these points increase in magnitude over the first half, then decrease. For n odd each half (after omitting the middle zero) is identical and oscillates like the design on $[\frac{n}{2}]$ points. For $\mathcal{P}_2$ the first half (for n odd omitting the middle zero) of the points are positive, the second half negative. Within each half the points increase and then decrease in magnitude. Figure 1 shows the most V-robust designs for $\mathcal{P}_1$ and $\mathcal{P}_2$ when $n = 9,\ 10,\ 19,\ 20$, and numerical values are in Table 1.

It is interesting to note that the *arcsine* design also arises in another context - that of optimal polynomial regression design, as the degree of the fitted polynomial tends to ∞. See Pukelsheim (1993) for a discussion.

4. As $n \to +\infty$, $CVS_1(\xi^*, \mathbf{I}) \to 0$ and $CVS_2(\xi^{**}, \mathbf{I}) \to -4c_1$.

5 B-robust designs

By virtue of (2.3) and (2.5), the B-robustness problem is that of constructing a design to maximize $|\mathbf{B}_\xi(0)|$, subject to a bound on $sup_{f\in\mathcal{F}}\mathbf{b}_{f,\xi}^T\mathbf{B}_\xi^{-1}(\mathbf{0})\mathbf{b}_{\mathbf{f},\xi}$. Such designs then coincide with the *Bounded Bias* designs of Liu (1994) and Liu and Wiens (1994).

Example 5.1: Consider the case of approximate multiple linear regression with an intercept and with uncorrelated errors, i.e. (1.1) with $\mathbf{z}(\mathbf{x}) = (1, \mathbf{x})^T$ and $\mathbf{P} = \mathbf{I}$. Let $\mathcal{S}$ be a q-dimensional sphere centred at the origin, with radius r determined by the requirement that $\mathcal{S}$ have unit volume. Take

$$\mathcal{F} = \{\ f \mid \int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})d\mathbf{x} = 0,\ \int_S f^2(\mathbf{x})d\mathbf{x} \leq \eta^2\ \}.$$

Minimax designs for this model are considered in Wiens (1992), and in Huber (1975) when $q = 1$. Bounded Bias designs are constructed in Liu (1994) when $q = 1$.

It is shown in Wiens (1992) that only absolutely continuous designs are admissible for this problem. A convexity argument then gives a further reduction to spherically symmetric measures. The maximization of $\mathbf{b}_{f,\xi}^T\mathbf{B}_\xi^{-1}(\mathbf{0})\mathbf{b}_{\mathbf{f},\xi}$ over $\mathcal{F}$ can be carried out as at Theorem 1 of Wiens (1992); the methods of Section 3 of that paper then show that the B-robust design is the solution to the problem

$$\textit{maximize } E_G[U^2] \textit{ subject to a bound on } E_G[g(U)],$$

where g is the density and G the distribution function, under ξ, of $U := ||\mathbf{x}||$. This is a standard variational problem whose solution is as follows. Let β be given by (3.2), and set $\beta' = \frac{\sigma^2\beta}{n\eta^2}$. Define $\gamma = E[U^2]/q$. For the B-robust design the density of U is of the form

$$g_0(u) = \lambda(u^2 + \delta)^+, \quad \lambda > 0, \quad u \in [0, r],$$

where λ and δ are chosen to make g_0 a proper density, and to attain the required bound on $E[g(U)]$.

1. $0 < \beta' \le \frac{4}{q(q+4)}$:

$$g_0(u) = 1 + (\frac{\gamma}{\gamma_0} - 1)(\frac{q+4}{4})(\frac{u^2}{\gamma_0} - q),$$

   where

$$\gamma_0 = \frac{r^2}{q+2}, \quad \gamma = \gamma_0 + \gamma_0 \cdot \sqrt{\frac{4\beta'}{q(q+4)}}.$$

2. $\beta' \ge \frac{4}{q(q+4)}$:

   Define

$$K_q(b) = (1-b) - \frac{2(1 - b^{\frac{q}{2}+1})}{q+2}.$$

   Determine b and γ from the equations

$$\gamma = \gamma_0 \cdot \frac{K_{q+2}(b)}{K_q(b)},$$
$$\frac{q\gamma - br^2}{r^2 K_q(b)} - 1 = \beta'.$$

Then

$$g_0(u) = [(\frac{u}{r})^2 - b]/K_q(b), \quad r\sqrt{b} \le u \le r.$$

The minimum value of CBS is $\beta = 0$. This is attained only by the Uniform distribution on $\mathcal{S}$, which is then most B-robust.

Example 5.2 For the approximate polynomial regression model - $\mathbf{z}^T(x) = (1, x, x^2, \dots,$ in (1.1) - with uncorrelated errors, Liu and Wiens (1994) construct Bounded Bias designs for the class

$$\mathcal{F} = \{ f \mid |x^{q+1} f(x)| \le \phi(x) \; \forall x \in \mathcal{S} = [-1, 1]\},$$

where ϕ is a given non-negative function. Liu (1994) considers the case $q = 1$. The designs are similar to the classically optimal designs which minimize variance alone, in that they have all mass at $q + 1$ symmetrically placed points. For $q = 2$ and $\phi(x) \equiv 1$ the solution is

$$\xi = \begin{cases} \frac{1}{2}\delta_0 + \frac{1}{2}\delta_{\pm(2\beta')^{\frac{1}{6}}}, & for\ 0 \le \beta' \le \frac{1}{2}; \\ (1 - \beta')\delta_0 + \beta'\delta_{\pm 1}, & for\ \frac{1}{2} \le \beta' \le \frac{2}{3}; \\ \frac{1}{3}\delta_0 + \frac{2}{3}\delta_{\pm 1}, & for\ \beta' \ge \frac{2}{3}; \end{cases}$$

where $\beta' = \frac{\sigma^2 \beta}{n}$ and δ_a is pointmass at a. In this case then, the concept 'most B-robust' leads to the clearly impractical design δ_0.

6 M-robust designs

In this section we outline an approach to the construction of M-robust desig Let ξ^* be a B-robust design for $\{\Xi, \mathcal{F}\}$. The B-robustness is unaffected by a p rmutation of the design points. Suppose then that there is a permutation for wh the corresponding design ξ^{**} has $CVS(\xi^{**}, \mathbf{I}) \leq \alpha$. Then ξ^{**} satisfies both (3.1) a (3.2). It minimizes (2.3) in the class of designs satisfying (3.2), hence *a fortiori* in t smaller class of designs satisfying both (3.1) and (3.2). It is thus M-robust.

Proposition 6.1 *Consider the model* $\{\mathcal{F}, \mathcal{P}_1\}$ *with* $\mathcal{F}$, $\mathcal{S}$ *and* $\mathbf{z}(\mathbf{x})$ *as in Exam 5.1. Let* $\alpha \geq 2\left(\frac{n-1}{n}\right)$ *and suppose that* ξ^* *is B-robust. If there is a permutat* $< \mathbf{x}_1^{**}, \ldots, \mathbf{x}_n^{**} >$ *of* $< \mathbf{x}_1^{*}, \ldots, \mathbf{x}_n^{*} >$ *for which*

$$\tau_j := \frac{\sum_{i=1}^{n-1} x_{i,j}^{**} x_{i+1,j}^{**}}{\sum_{i=1}^{n} x_{i,j}^{**2}} < 0 \quad \forall\, j = 1, \ldots, q$$

then the corresponding design ξ^{**} *is M-robust.*

Proof: As in Section 4 we find

$$CVS(\xi^{**}, \mathbf{I}) = \sup_{c_0 < \rho(1) < 1} 2\rho(1) \left[\frac{n-1}{n} + \sum_{j=1}^{q} \tau_j\right] \leq \alpha.$$

Similarly, we have

Proposition 6.2 *Consider the model* $\{\mathcal{F}, \mathcal{P}_2\}$ *with* $\mathcal{F}$, $\mathcal{S}$ *and* $\mathbf{z}(\mathbf{x})$ *as in Exam 5.1. Let* $\alpha \geq -2c_1\left(\frac{n-1}{n}\right)$ *and suppose that* ξ^* *is B-robust. If there is a permutat*

$< \mathbf{x}_1^{**}, \ldots, \mathbf{x}_n^{**} >$ *of* $< \mathbf{x}_1^{*}, \ldots, \mathbf{x}_n^{*} >$ *for which* $\tau_j > 0$ *for all* $j = 1, \ldots, q$ *then the corresponding design* ξ^{**} *is M-robust.*

In the following example we propose ways of implementing the requirements of Propositions 6.1 and 6.2 asymptotically, for $q = 2$.

Example 6.1: To implement Proposition 6.1 asymptotically, suppose that n is a multiple of 4 and that we are given design points

$$\{\mathbf{x}_i^J \in \textit{ quadrant } J | i = 1, \ldots, \frac{n}{4}; J \in \{I, II, III, IV\}\}.$$

Now apply the design by alternating between quadrants I and III for the first $\frac{n}{2}$ points, and then between quadrants II and IV for the remaining $\frac{n}{2}$ points:

$$\mathbf{x}_1^{**} = < \mathbf{x}_1^I, \mathbf{x}_1^{III}, \mathbf{x}_2^I, \mathbf{x}_2^{III}, \ldots, \mathbf{x}_{\frac{n}{4}}^I, \mathbf{x}_{\frac{n}{4}}^{III}, \mathbf{x}_1^{II}, \mathbf{x}_1^{IV}, \mathbf{x}_2^{II}, \mathbf{x}_2^{IV}, \ldots, \mathbf{x}_{\frac{n}{4}}^{II}, \mathbf{x}_{\frac{n}{4}}^{IV} > .$$

It is readily checked that $\tau_j < 0$ for $j = 1, 2$; the design is then asymptotically M-robust if the empirical distribution of the $\mathbf{x}_i^J$ tends weakly to ξ^*. This can be arranged as follows.

Define a probability distribution by

$$p_0 = 1 - \frac{[\sqrt{n}]^2}{n}, \quad p_i = \frac{[\sqrt{n}]}{n}, \quad i = 1, 2, \cdots, [\sqrt{n}].$$

Define

$$a_{-1} = 0, \quad a_i = G_0^{-1}\left(\sum_{j=0}^{i} p_j\right), \quad i = 0, 1, \cdots, [\sqrt{n}],$$

where G_0 is the distribution function corresponding to the optimal density function g_0 of Example 5.1. Divide the design space $\mathcal{S} = \{ \mathbf{x} \mid ||\mathbf{x}|| \leq r \}$ into $[\sqrt{n}] + 1$ annuli $A_0, A_1, \cdots, A_{[\sqrt{n}]}$ with

$$A_i = \{ \mathbf{x} \mid ||\mathbf{x}|| \in (a_{i-1}, a_i] \}, \quad and \quad P_{G_0}(\mathbf{x} \in A_i) = p_i.$$

In each A_i we select np_i points equally spaced over $||\mathbf{x}|| = \frac{a_{i-1}+a_i}{2}$. We then have $n - [\sqrt{n}]^2$ points on A_0 and $[\sqrt{n}]$ points on each A_i for $i = 1, \cdots, [\sqrt{n}]$. It is easy to verify that the empirical distribution of $||\mathbf{x}||$ for the n points converges weakly to G_0. Also $\frac{\mathbf{x}}{||\mathbf{x}||}$ is, under ξ^*, independent of $||\mathbf{x}||$ and uniformly distributed over the unit sphere. It follows that the empirical distribution of the design points tends weakly to ξ^*.

Similarly, we obtain an asymptotically M-robust design for $\mathcal{P}_2$ by applying the design points quadrant-by-quadrant:

$$\mathbf{x}_2^{**} = < \mathbf{x}_1^I, \ldots, \mathbf{x}_{\frac{n}{4}}^I, \mathbf{x}_1^{II}, \ldots, \mathbf{x}_{\frac{n}{4}}^{II}, \mathbf{x}_1^{III}, \ldots, \mathbf{x}_{\frac{n}{4}}^{III}, \mathbf{x}_1^{IV}, \ldots, \mathbf{x}_{\frac{n}{4}}^{IV} >,$$

so that $\tau_j > 0$ for $j = 1, 2$.

References

Constantine, G.M. (1989). Robust designs for serially correlated observations. *Biometrika* **76**, 245-251.

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

Hampel, F.R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69**, 383-393.

Hampel, F.R., R. Ronchetti, R.J. Rousseeuw and W. Stahel (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

Huber, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Stat.* **35**, 73-101.

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

Hössjer, O. (1991). The change-of-variance function for dependent data. *Probab. Theory Relat. Fields* **90**, 447-467.

Jenkins G.M. and J. Chanmugam (1962). The estimation of slope when the errors are autocorrelated. *J.R. Statist. Soc.* **B 24**, 199-214.

Kiefer, J. and H.P. Wynn (1984). Optimum and minimax exact treatment designs for one-dimensional autoregressive error process. *Ann. Statist.* **12**, 414-450.

Li, K.C. and W. Notz (1982). Robust designs for nearly linear regression. *J. Statist. Plann. Inference* **6**, 135-151.

Liu, S.X. (1994). Some topics in robust estimation and experimental design. Ph.D. Thesis, University of Calgary.

Liu, S.X. and D.P. Wiens (1994). Robust designs for approximately polynomial regression. Preprint.

Pesotchinsky, L. (1982). Optimal robust designs: Linear regression in R^k. *Ann. Statist.* **10**, 511-525.

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

Tang, D. (1993). Minimax regression designs under uniform departure models. *Ann. Statist.* **21**, 434-446.

Wiens, D.P. (1991). Designs for approximately linear regression: Two optimality properties of uniform designs. *Statist. Probab. Lett.* **12**, 217-221.

Wiens, D.P. (1992). Minimax designs for approximately linear regression. *J. Statist. Plann. Inference* **31**, 353-371.

Wiens, D.P. and J. Zhou (1994). Minimax regression designs for approximately linear models with autocorrelated errors. Preprint.

Wiens, D.P. and J. Zhou (1995). Minimax designs for approximately linear models with AR(1) errors. Preprint.