Bayesian Two Stage Design Under Model Uncertainty

by

Angela R. Neff

Dissertation submitted to the faculty of
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

Raymond H. Myers, Co-chair
Keying Ye, Co-chair
Jesse C. Arnold
Robert V. Foutz
Marvin Lentner

January 16, 1996
Blacksburg, Virginia

Keywords: Bayesian, Design Optimality, Response Surface, Two Stage Design

Bayesian Two Stage Design Under Model Uncertainty

Angela R. Neff

(ABSTRACT)

Traditional single stage design optimality procedures can be used to efficiently generate data for an assumed model $\mathbf{y} = f(\boldsymbol{x}^{(m)},\beta) + \varepsilon$. The model assumptions include the form of *f,* the set of regressors, $\boldsymbol{x}^{(m)}$ **,** and the distribution of ε. The nature of the response, y, often provides information about the model form (f) and the error distribution. It is more difficult to know, apriori, the specific set of regressors which will best explain the relationship between the response and a set of design (control) variables $\boldsymbol{x}$. Misspecification of $\boldsymbol{x}^{(m)}$ will result in a design which is efficient, but for the wrong model.

A Bayesian two stage design approach makes it possible to efficiently design experiments when initial knowledge of $\boldsymbol{x}^{(m)}$ is poor. This is accomplished by using a Bayesian optimality criterion in the first stage which is robust to model uncertainty. Bayesian analysis of first stage data reduces uncertainty associated with $\mathbf{x}^{(m)}$, enabling the remaining design points (second stage design) to be chosen with greater efficiency. The second stage design is then generated from an optimality procedure which incorporates the improved model knowledge. Using this approach, numerous two stage design procedures have been developed for the normal linear model. Extending this concept, a Bayesian design augmentation procedure has been developed for the purpose of efficiently obtaining data for variance modeling, when initial knowledge of the variance model is poor.

Dedication

To my parents, Roy and Pauline Neff, who have always supported and encouraged me in my endeavors. The lessons which they have taught me will last a lifetime.

Acknowledgements

I would like to thank my family and friends, especially my parents, Roy and Pauline Neff, and my best friend, Jeff Patterson, for their patience, love, support, and encouragement. The support of my family and friends has enabled me to make this dream a reality.

I would especially like to thank Dr. Raymond H. Myers, who has been there for me every step of the way... teaching, encouraging and coaching. I will treasure our friendship always.

I would also like to thank Dr. Keying Ye, whose interest in Bayesian methods was transferred to me through his excellent teaching. His patience and wisdom have been invaluable.

Above all, I give thanks to God, for it is through Him that all things are possible.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction and Literature Review

1.1 The Sequential Nature of RSM

Response surface methodology (RSM) is a collection of techniques for experimental design or data analysis useful in the study of new or existing products and processes (see Myers and Montgomery (1995)). Many applications of RSM can be found in industry, where an engineer or scientist (referred to as the *experimenter* throughout this paper) needs to understand how the performance of a product/process *responds* to changes in product/process variables under his or her control. The idea behind RSM is that the dependency of a product/process performance measure (**response**) on a set of controllable variables is defined by an unknown *true response model*

$$y = g(x_1, x_2, ..., x_k) + \varepsilon \qquad (1.1.1)$$

where y denotes the response, $x_1, x_2, ..., x_k$ denote the k **control variables** in coded (standardized) units and ε is an unknown noise component. The response function, g, may be of complex form.

RSM techniques are used to generate and analyze data to arrive at a close approximation to (1.1.1). For a set of n observations of the response, the model is of the form

$$y_i = f(\boldsymbol{x}_i'^{(m)} \beta) + \varepsilon_i, \qquad i = 1,...,n . \qquad (1.1.2)$$

In (1.1.2), $\boldsymbol{x}_i'^{(m)}$ is a set of regressor variables $\{x_1, x_2, ..., x_p\}$ at the i^{th} condition of the control variables. Usually the model contains linear, quadratic or interaction terms. The set of unknown regressor coefficients are denoted by β. The response function, f, may be linear or nonlinear in the β`s.

In most RSM applications, however, knowledge of which regressors are needed to closely approximate (1.1.1) must be obtained through a series of experiments and data analyses. Initially, an experimenter may have an extensive list of control variables which could possibly influence the response. At least one *screening experiment* is needed to eliminate any variables which do not appear to affect the response. Additional experiments are conducted with the remaining variables to identify the actual functional form of the model. Inferences made with the resulting model will be adversely affected if the model contains too few (underspecification) or too many (overspecification) regressors. For this reason, many resources are often expended throughout the model building process.

An extremely important aspect of RSM is the strategy of designing a series of experiments to efficiently build a model which closely approximates (1.1.1). The experimenter (or statistician) must be concerned with the efficiency of the combined experiments, as well as each of the individual experiments. By efficiency, we mean the degree to which the experimental goals are achieved with the number of observations allotted.

1.2 Model Assumptions for Design Optimality Criteria

Based on **optimal design theory**, dating back to Kiefer (1959) and Kiefer and Wolfowitz (1961), numerous *criteria* have been developed to measure the performance (i.e., efficiency) of an experimental design. In addition to providing a mechanism for comparing existing designs, these **design optimality criteria** form a basis for selecting a set of experimental conditions with optimum properties. Since the availability of computer-generated designs in the 1980's, various algorithms have been written to make 'optimal' designs very accessible. The true optimality of these designs is actually conditional on model assumptions.

Traditional design optimality criteria work under the assumption that a **specific model** will be used to make inferences about the response behavior. Suppose that an experimenter specifies that he/she wants to design an experiment for the model, which in matrix form is

$$\mathbf{y} = f(\mathbf{X}\beta) + \varepsilon . \qquad (1.2.1)$$

The n rows of the matrix **X** will be defined by the n conditions of the experimental design. The p columns of **X**, however, are defined by the set of regressor variables $\boldsymbol{x}'^{(m)} = \{x_1, x_2, ..., x_p\}$, which must be specified by the experimenter. The error distribution and response function (f) , which together define the likelihood function $L(\mathbf{y},\beta|\mathbf{X})$, must also be specified. Under these model assumptions, a typical design optimality criterion selects the rows of **X** to maximize information about some or all of the unknown β`s, resulting in improved inferences about the response behavior. This is done by choosing the rows of **X** which optimize some function of the Fisher information matrix,

$$I(\beta|\mathbf{X}) = - E_y \left[\frac{\partial}{\partial \beta'} \left(\frac{\partial \ln L(\beta, \mathbf{y}|\mathbf{X})}{\partial \beta} \right) \right].$$

Obviously $I(\beta|\mathbf{X})$ depends on **X** and in the case of nonlinear models, it also depends on β.

Unfortunately, due to the dependence on model assumptions, successful implementation of design optimality criteria is often difficult in practice. The assumption of $\boldsymbol{x}'^{(m)}$ is often a weak one, since little regressor knowledge may exist prior to conducting the experiment and performing the data analysis. This makes it often difficult to utilize design criteria effectively since misspecification of $\boldsymbol{x}'^{(m)}$ could result in a design which is possibly much less than optimal. Overspecifying $\boldsymbol{x}'^{(m)}$ may result in a design in which some observations will be wasted in trying to estimate unimportant parameters (effects). Even worse, underspecifying $\boldsymbol{x}'^{(m)}$ may result in a design which doesn't allow all necessary effects to be quantified, resulting in an inadequate model.

1.3 Developments in Design Optimality

In recent years, research in the area of optimal design theory has been focused on producing design optimality criteria with less dependence on model assumptions. The Bayesian approach to design optimality is one way to address this. (An extensive review of developments in Bayesian design is given by Chaloner and Verdinelli, (1995) .)

Sometimes, prior to observing data from an experiment, the experimenter has beliefs about β which can be summarized with a probability distribution. If so, a Bayesian approach to design optimality can be more advantageous than the non-Bayesian approach. The basic idea of Bayesian design optimality is to choose the design so as to maximize *posterior* knowledge about some or all of the β`s, *conditional* on the *prior* information available. In 1984, Chaloner derived Bayesian design optimality criteria for the linear model with i.i.d. normal error structure, when apriori β is believed to follow a normal distribution. In 1989, Chaloner and Larntz developed Bayesian design criteria for nonlinear models - specifically, the logistic model. The Bayesian approach results in a design which is optimal over a distribution of β values, as opposed to a single β (the non-Bayesian approach). All of these criteria, however, just like their non-Bayesian counterparts, find optimal designs for fitting a model with a specific set of regressors. DuMouchel and Jones (1994) proposed a Bayesian approach to obtaining D-optimal designs with reduced dependence on regressor specification.

Another approach to obtaining optimal designs when initial model knowledge is poor is to design the experiment in *stages*. In a two stage design procedure, a proportion of observations, allocated to the first stage experiment, produce model information needed to design the balance of the experiment. Abdelbasit and Plackett (1983) and Minkin (1987) found that non-Bayesian two stage D-optimal designs for the logistic model were more robust to poor parameter guesses than the single stage equivalent. Myers, Myers, Carter and White (1996), also working with the

logistic model, combined the D and Q criteria in two stages. The resulting D-Q procedure produces a combined Q-optimal design, conditional on the first stage. Letsinger (1995) combined Bayesian design with a two stage approach to generate efficient D-optimal designs for the logistic model. Using his method, Bayesian D-optimality in the first stage reduces the possibility of initial parameter misspecification. Conditional on Bayesian parameter estimates from the first stage, the balance of the design is selected according to non-Bayesian D-optimality. This two stage approach results in good efficiency and robustness properties.

Although numerous design procedures have been developed to address the problem of parameter misspecification, very little work has been done to address the issue of *regressor misspecification*. The purpose of this research is to develop Bayesian two stage design procedures for the purpose of generating efficient designs with reduced dependence on regressor specification. In these procedures, Bayesian methods are used to construct the first and second stage designs, and to perform the first stage data analysis. Current development has been restricted to applications for the normal linear model.

This paper is organized as follows. Chapters 2 and 3 introduce non-Bayesian and Bayesian design optimality criteria, respectively. These criteria are presented in the context of the normal linear model. The two stage design approach is reviewed in chapter 4, from both the non-Bayesian and Bayesian perspectives. In Chapter 5, Bayesian two stage design procedures with reduced regressor dependence are developed for the normal linear model. Chapter 6 summarizes an evaluation of the performance of these procedures relative to their single stage competitors. The focus shifts to variance modeling in chapter 7, where a Bayesian procedure is developed for the selection of a D-optimal design augmentation for the log linear variance model.

Chapter 2

Non-Bayesian Design Optimality

In this chapter, selected design optimality criteria are presented. Although applications of these criteria are not limited to any specific linear or nonlinear model form, focus is placed on the linear model, as it will be the primary model for applications presented in Chapter 5.

2.1 The Linear Model

In the case of the linear model we have

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon$$

where $\mathbf{y}$ is an n dimensional response vector, $\mathbf{X}$ is an $n \times p$ model matrix of p-1 regressors augmented by a column of ones, β is a p dimensional vector of unknown parameters and ε is an n dimensional vector of unknown errors.

Under the usual assumption that ε is distributed $N(\mathbf{0},\sigma^2\mathbf{I})$, the maximum likelihood estimator of β can be found by the method of ordinary least squares, and is given by

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} .$$

The variance of $\mathbf{b}$ and the variance of prediction are both functions of the Fisher information matrix,

$$\mathbf{I}(\beta) = -E\left[\frac{\partial}{\partial \beta'}\left(\frac{\partial \ln L(\beta, \mathbf{y})}{\partial \beta}\right)\right] = \mathbf{X}'\mathbf{X}(1/\sigma^2) .$$

The variance of $\mathbf{b} = [\mathbf{I}(\beta)]^{-1}$, i.e.,

$$\mathrm{Var}(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} . \qquad (2.1.1)$$

The variance associated with predicting the mean response at a given point in model space $\mathbf{x}_j^{(m)}$, is often called the *variance of prediction,* given by

$$\mathrm{Var}(\hat{y}(\mathbf{x}_j)) = \mathbf{x}_j'^{(m)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_j^{(m)}\sigma^2 \qquad (2.1.2)$$

Note that both (2.1.1) and (2.1.2) are independent of β. The importance of this will become evident in the following sections, when design optimality is discussed.

2.2 Overview of Design Optimality

Optimal design theory dates back to Kiefer [1959, 1961] and Kiefer and Wolfowitz [1959]. One criterion which they formed is D-optimality, which will be discussed in detail later in this chapter. Since the work of Kiefer et al., numerous other design criteria have been developed. Like D-optimality, the various design optimality criteria are characterized by letters of the alphabet, thus the term "alphabetic optimality." See Silvey [1980] and Atkinson [1982] for surveys of the various criteria.

In much of the literature, an experimental design is viewed as a probability measure defined on the set of all conditions available for the experiment. The design measure assigns a proportion (n_i/n) of observations to be taken at each of the conditions, where n_i is not necessarily integer valued. In practical applications, however, it is more reasonable to only consider those measures which produce realistic designs. For the remainder of this paper, let *D* be defined as the set of all designs (or design measures) δ for which all n_i are integer valued.

The basic idea underlying design optimality theory is that statistical inference about quantities of interest can be improved by "optimally" selecting levels of the control variables. Based on this theory, various criteria, or measures, have been developed in order to quantify, and therefore maximize, design performance. In general, a design optimality criterion can be characterized as an "estimation criterion" or a "prediction criterion" depending on the primary purpose for conducting the experiment. An experimental design which is optimal with respect to an estimation criterion is one which maximizes parameter information by minimizing estimator variability. A design which is optimal with respect to a prediction criterion maximizes information about a response surface by focusing on the prediction qualities of the fitted model. In general, an experimental design may not be optimal with respect to both estimation and prediction.

In the remainder of this chapter, three very common criteria will be discussed in detail. They are D and A optimality, which are estimation criteria and Q-optimality, a prediction criterion.

2.3 Selected Estimation Criteria

The most well known and widely used design criterion is **D-optimality**, dating to Kiefer and Wolfowitz (1959). This criterion suggests that the choice of design should maximize the information on β by minimizing the generalized variance of its estimator. In other words, variances as well as covariances of parameter estimates must be controlled. For the linear model with $N(\mathbf{0},\sigma^2\mathbf{I})$ error structure, the D-optimal design does not depend on knowledge of β. The D-optimality criterion is given by

$$\underset{D}{\text{Min}} \; | N(\mathbf{X}'\mathbf{X})^{-1} | \; . \qquad (2.3.1)$$

The choice of the determinant as a norm on Var(**b**) is very intuitive, since $|(\mathbf{X}'\mathbf{X})^{-1}|$ is proportional to the squared volume of the joint confidence region on β. A small determinant reflects a tight confidence region and consequently good estimation of β. Computer algorithms,

such as DETMAX (see Mitchell (1974)), make it quite easy to obtain D-optimal designs for a user's choice of regressors, design space and sample size. Additionally, it is interesting to note that even though | Var(**b**)| is not scale invariant, the design which satisfies (2.3.1) is D-optimal regardless of how the design space is scaled.

Another criterion which places a premium on parameter estimation is **A-optimality**. This criterion uses a trace as the norm on Var(**b**). This results in $\sum_i \text{Var}(b_i)$ as a scalar measure of design performance. Unlike D-optimality, covariances are ignored by this criterion and the criterion is not always invariant to scale. The criterion simplifies to the form in (2.3.2), which is independent of β.

$$\underset{D}{\text{Min}} \ \text{tr}\left(N(\mathbf{X}'\mathbf{X})^{-1} \right) \qquad (2.3.2)$$

2.4 Q-optimality, a Prediction Criterion

When the experimental objective is to gain knowledge about the response surface, a design which allows accurate and precise prediction of the response is desirable. Most prediction criteria place importance on minimizing some function of the scaled prediction variance, which, for any point **x** in the design space, is given by

$$v(\mathbf{x}) = (N/\sigma^2) \text{Var}(\hat{y}(\mathbf{x})) = N \mathbf{x}'^{(m)} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}^{(m)}$$

Note that for the linear model, $v(\mathbf{x})$ is a scale free quantity, and therefore unaffected by any rescaling of the design space.

An optimality criterion which addresses prediction variance is **Q-optimality**. A single number which quantifies predictive quality is generated by averaging the scaled prediction variance over some region of interest, *R*. The averaging is accomplished by integrating $v(\mathbf{x})$ over *R*, then dividing through by k, the volume of *R*. For the linear model, the general form of the Q criterion is given by

$$\underset{D}{\text{Min}} \quad (N / k) \int_R \mathbf{x}'^{(m)} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}^{(m)} w(\mathbf{x}) \, dx$$

where $w(\mathbf{x})$ is a weight function used to weight certain parts of R more heavily. In many applications , $w(\mathbf{x}) = 1.0$, representing uniform interest throughout R . The Q-optimal design is implemented by choosing the rows of **X**, in model form, which minimize the Q criterion.

Chapter 3

Bayesian Design

As with many areas of Bayesian statistics, Bayesian experimental design has experienced rapid growth in recent years. A thorough review of Bayesian design is given by Chaloner and Verdinelli (1995). In this chapter, an introduction to Bayesian design optimality for the linear model is discussed. Also presented are two applications of Bayesian design.

3.1 Fundamentals of Bayesian Design Optimality Theory

In his 1972 review of Bayesian statistics (pages 19-20), Lindley presented a decision theory approach to experimental design. Lindley's argument suggests that a good method for designing experiments is to specify a utility function which reflects the purpose of the experiment, then choose the design that maximizes the expected utility. A parallel approach is to specify a loss function which, when minimized, reflects the purpose of the experiment. The choice of design can then be viewed as a decision theory problem for which the Bayes solution is to select the design (or design measure) δ^*, that minimizes the Bayes risk, i.e.,

$$R(\delta^*) = \underset{D}{\text{Min}} \int L(\theta,\delta,\mathbf{y})\, p(\theta,\mathbf{y}|\delta)\, d\theta\, d\mathbf{y} \qquad (3.1.1)$$

where $p(\cdot)$ denotes a probability density function with respect to an appropriate measure, θ is the set of unknown parameters, $\mathbf{y}$ is a set of data to be observed in the experiment and D represents the collection of all possible designs.

In developing design criteria from (3.1.1), $p(\theta,\mathbf{y}|\delta)$ can be written as $p(\theta|\mathbf{y})m(\mathbf{y})$, the product of the posterior density of θ and the marginal of $\mathbf{y}$. Integrating (3.1.1) first with respect to θ then results in the posterior expected loss. For example, if $L(\theta,\delta,\mathbf{y}) = (\theta-\hat{\theta})'(\theta-\hat{\theta})$ (squared error loss)

and $\hat{\theta}$ is the Bayes estimator of θ (i.e., $\hat{\theta} = E[\theta|\mathbf{y}]$), then the posterior expected loss is the sum of the posterior variances of the individual parameters. Integration once more with respect to $\mathbf{y}$ results in the Bayes risk.

When developing design criteria from (3.1.1) for nonlinear models, it is often necessary to write $p(\theta,\mathbf{y}|\delta)$ as $f(\mathbf{y}|\theta)\pi(\theta)$, the product of the likelihood of $\mathbf{y}$ and the prior distribution of θ. This is often necessary because it is difficult (if not impossible) to find the exact posterior of θ. Integrating (3.1.1) first with respect to $\mathbf{y}$ then results in the *frequentist risk*. For example, assuming again that $L(\theta,\delta,\mathbf{y}) = (\theta-\hat{\theta})'(\theta-\hat{\theta})$, where $\hat{\theta}$ is the Bayes estimator and a function of $\mathbf{y}$, the frequentist risk is the sum of the posterior variances of the individual parameters. Taking expectation again with respect to the prior distribution of θ results in the Bayes risk.

3.2 Design Optimality for the Linear Model with Normal Error Structure

Optimal experimental designs for estimation and prediction in normal linear models were derived by Chaloner (1984) with a more extensive review given by Chaloner and Verdinelli (1995). Consider the linear model $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ where $\mathbf{y}$ is a response vector containing n observations and $\mathbf{y}|\beta,\sigma^2 \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I})$. The experimental design in model format is contained in $\mathbf{X}$ of dimension $n \times p$. Suppose that the prior information on β is that $\beta|\sigma^2,\tau^2 \sim N(\beta_0, \sigma^2\,\mathbf{T}^{-1})$ where $\mathbf{T} = (1/\tau^2)\mathbf{I}_p$ for a known τ. Under these distributional assumptions, the posterior distribution of β is also normal with mean $\mathbf{b} = (\mathbf{X}'\mathbf{X} + \mathbf{T})^{-1}(\mathbf{X}'\mathbf{y} + \mathbf{T}\beta_0)$ and covariance matrix $\sigma^2\mathbf{V} = \sigma^2(\mathbf{X}'\mathbf{X} + \mathbf{T})^{-1}$ (Box and Tiao, 1973). Note that $\mathbf{V}$ is a function of the design (through $\mathbf{X}$) and the prior variance of β. This posterior variance will play an integral role in several design criteria, which will now be discussed.

Chaloner and Verdinelli show that Bayes D-optimality has been derived from several different loss functions. One such loss function is the negative Kullback-Leibler distance between the posterior and prior distributions;

$$L(\beta,\delta,\mathbf{y}) = - \log \frac{p(\beta|\mathbf{y},\delta)}{p(\beta)}$$

where $-L(\beta,\delta,\mathbf{y})$ is also known as the gain in Shannon information (Shannon, 1948). Following Lindley's approach, the corresponding Bayes risk is

$$R(\delta) = \int - \log \frac{p(\beta|\mathbf{y},\delta)}{p(\beta)} \; p(\mathbf{y},\beta|\delta)\, d\beta d\mathbf{y} \; .$$

The design which minimizes $R(\delta)$ is that which maximizes posterior knowledge of β over and above that reflected in the prior. For the normal linear model with p parameters,

$$R(\delta) = \; - c - \tfrac{1}{2} \log \det\{\sigma^{-2}(\mathbf{X}'\mathbf{X} + \mathbf{T})\}$$

where c is a constant. Choosing a design δ to minimize R is therefore equivalent to selecting the Bayes D-optimal design, δ^* for which

$$R(\delta^*) = \underset{D}{\text{Min}} \;\; \log | \; (\mathbf{X}'\mathbf{X} + \mathbf{T})^{-1} \; | \; .$$

A set of Bayesian design criteria, collectively called Bayes A-optimality criteria, is useful when inference about functions of β are of interest. An A-optimal design criterion can be derived from a quadratic loss function,

$$L(\delta,\beta,\mathbf{y}) = (\beta-\mathbf{b})'\mathbf{A}(\beta-\mathbf{b}) \qquad (3.2.1)$$

for any symmetric non negative matrix **A**, of dimension p×p. The Bayes risk corresponding to (3.2.1) is

$$\begin{aligned} R(\delta) \;\; &= \int (\beta-\mathbf{b})'\mathbf{A}(\beta-\mathbf{b}) \; p(\mathbf{y},\beta|\delta)\, d\beta \; d\mathbf{y} \\ &= \int \mathrm{tr}[\mathbf{A}(\beta-\mathbf{b})(\beta-\mathbf{b})' \;] \; p(\mathbf{y},\beta|\delta)\, d\beta \; d\mathbf{y} \qquad (3.2.2) \end{aligned}$$

For the normal linear model, (3.2.2) simplifies to

$$R(\delta) = \text{tr}\,(\mathbf{A}\,\sigma^2(\mathbf{X}'\mathbf{X} + \mathbf{T})^{-1}\,),$$

thus resulting in the Bayes A-optimality criterion

$$\underset{D}{\text{Min}}\;\; R(\delta) = \text{tr}\,(\mathbf{A}\,\sigma^2(\mathbf{X}'\mathbf{X} + \mathbf{T})^{-1}\,)$$

A special case of Bayesian A-optimality will be referred to as Bayesian Q-optimality, since it parallels the non-Bayesian Q criterion. The integrated prediction variance, which is the non-Bayesian Q criterion, serves as the loss function , i.e.,

$$L(\delta,\beta,\mathbf{y}) = (N/k)\int_R \mathbf{x}'^{(m)}\,(\beta-\mathbf{b})(\beta-\mathbf{b})'\,\mathbf{x}^{(m)}\;\; w(\mathbf{x})\,d\mathbf{x}\,. \qquad (3.2.3)$$

Equivalently, (3.2.3) can be written in the form of (3.2.1), namely

$$L(\delta,\beta,\mathbf{y}) = (\beta-\mathbf{b})'\,\big[(N/k)\int_R \mathbf{x}^{(m)}\,\mathbf{x}'^{(m)}\,w(\mathbf{x})\,d\mathbf{x}\,\big]\,(\beta-\mathbf{b})\,.$$

The corresponding Bayes risk is

$$\begin{aligned} R(\delta) \;&=\; \int (\beta-\mathbf{b})'\,\big[(N/k)\int_R \mathbf{x}^{(m)}\,\mathbf{x}'^{(m)}\,w(\mathbf{x})\,d\mathbf{x}\,\big](\beta-\mathbf{b})\;p(\mathbf{y},\beta|\delta)\,d\beta\,d\mathbf{y} \\ &=\; \text{tr}\,\Big(\,\big[(N/k)\int_R \mathbf{x}^{(m)}\,\mathbf{x}'^{(m)}\,w(\mathbf{x})\,d\mathbf{x}\big]\int (\beta-\mathbf{b})\,(\beta-\mathbf{b})'\;p(\mathbf{y},\beta|\delta)\,d\beta\,d\mathbf{y}\Big) \qquad (3.2.4) \end{aligned}$$

For the normal linear model, (3.2.4) simplifies to

$$R(\delta) = \text{tr}\,\Big(\,\big[(N/k)\int_R \mathbf{x}^{(m)}\mathbf{x}'^{(m)}\,w(\mathbf{x})\,d\mathbf{x}\,\big]\;\sigma^2(\mathbf{X}'\mathbf{X} + \mathbf{T})^{-1}\,\Big)$$

and thus the Bayes Q-optimal design is that which minimizes $R(\delta)$. This criterion should be used when the reason for conducting the experiment is to control the precision with which the mean response can be predicted. This becomes very important when knowledge about the response surface is needed in order to gain control of the process or system under study.

The last criterion to be presented in this chapter is useful when the experimenter is interested in both estimation of β and prediction of future observations. The expected utility function, as suggested by Verdinelli (1992) is a linear combination of the expected gain in

Shannon information for both $p(\beta|\mathbf{y},\delta)$ and a future observation at a condition $\mathbf{x}$. Interestingly enough, for the normal linear model, the design which maximizes Verdinelli's utility function also minimizes

$$R_v(\delta,\mathbf{x}) = \sigma^2 [\mathbf{x}'^{(m)}(\mathbf{X}'\mathbf{X} + \mathbf{T})^{-1}\mathbf{x}^{(m)} + 1] \cdot \det(\sigma^2(\mathbf{X}'\mathbf{X} + \mathbf{T})^{-1}) \ . \qquad (3.2.5)$$

Note that $R_v(\delta)$ is a function of the prediction variance at a point $\mathbf{x}$, i.e.,

$$\text{Var}\,(\hat{y}\,(\mathbf{x})) = \sigma^2\,\mathbf{x}'^{(m)}(\mathbf{X}'\mathbf{X} + \mathbf{T})^{-1}\mathbf{x}^{(m)},$$

as well as the posterior $\text{Var}(\beta)$, which is $\sigma^2(\mathbf{X}'\mathbf{X} + \mathbf{T})^{-1}$. Also note that this simplified criterion is independent of any priorities that the experimenter may have regarding estimation or prediction.

3.3 Applications of Bayesian Design

In certain applications, Bayesian designs have been shown to outperform their non-Bayesian counterparts. Two of these applications will be discussed.

In 1989, Chaloner and Larntz developed Bayesian design criteria for the logistic regression model. The logistic model is an important special case of the generalized linear model. Its primary use is in modeling binary response data which can result from bioassay experiments or reliability studies. When working with a logistic model, non-Bayesian design criteria are not very appealing because they depend on knowledge of unknown model parameters. Designs which are 'locally optimum' are constructed using the experimenter's best guess of the parameter values. This approach leaves a lot of room for misspecification of the model as well as the design. Bayesian methods, however, permit the experimenter to design optimally over a distribution of parameter values. It is not surprising therefore, that Bayesian designs for the logistic model have been shown to be more robust to model misspecification. For more details, see Letsinger (1995).

In 1994, DuMouchel and Jones proposed a Bayesian modification to D-optimal designs for the normal linear model. The purpose of their work was to develop a criterion with less

dependence on model specification. Essentially, this was accomplished with Bayesian D-optimality by structuring the prior distribution on β to reflect priorities in estimation. For example, high priority is placed on estimating linear terms with low priority on higher order terms. The high priority terms are called 'primary terms' while the low priority terms are called 'potential terms'. Through a case study the performance of the Bayesian design was compared to that of more traditional designs. The Bayesian design was more powerful for detecting the presence of potential terms, when only the primary terms were modeled. Additionally, under the assumption that the true model contained potential terms, the Bayesian design resulted in a lower mean squared error of prediction.

Chapter 4

Two Stage Design

The two stage design procedure is best defined by contrasting it with that of a single stage design. A single stage design of size N is one for which all N design points are optimally selected and executed as a group. The resulting N observations are analyzed together upon completion of the experiment. In contrast, a two stage design is one in which the combined experiment, of size N, consists of two experiments run sequentially, but not independently of one another. The first stage experiment is designed optimally, then conditional on the information provided by the first stage, the second stage design is chosen to create certain optimal conditions in the combined design. Statistical inferences are then made based on all N observations, as if the experiment had been completed in a single stage.

In this chapter, the two stage design procedure will be presented from both the non-Bayesian and Bayesian perspectives. Related issues such as design performance evaluation and optimal sample allocation will also be addressed.

4.1 Non-Bayesian Two Stage Design

Surprisingly, there has been very limited work in the development of two stage design procedures. In fact, the only non-Bayesian two stage procedures appearing in the literature are those developed for nonlinear models (such as the logistic regression model). Similar to two stage designs are the well known fold-over and central composite designs (CCD) , developed for use with normal linear models. Both the fold-over and CCD can be executed in two blocks or stages,

the foldover being constructed from two fractional factorials and the CCD consisting of a factorial (or fractional factorial) design and a set of additional runs called axials (see Myers and Montgomery, 1995). The difference between these designs and two stage procedures, however, being that the second block or stage is only executed when necessary, as indicated from the first stage data analysis.

Critical to the development of any non-Bayesian two stage design procedure is the ability to express the Fisher information matrix in a form which reflects the two stage manner in which information is gained about the unknown parameters. This is easily accomplished, as shown in the context of the linear model,

$$y_i = \mathbf{x}_i'^{(m)} \beta + \varepsilon_i, \qquad i = 1, 2, ..., N.$$

If the assumed probability distribution of the response is $p(y|\beta)$, then the likelihood function for all N observations generated from a design defined by $\mathbf{X}$ (in model space), is

$$L(\mathbf{y},\beta|\mathbf{X}) = \prod_{i=1}^{N} p(y_i|\beta). \tag{4.1.1}$$

For this likelihood , the Fisher information matrix for the unknown parameter vector β of dimension $p\times1$ is

$$I(\beta) = (1/\sigma^2)\mathbf{X}'\mathbf{X}$$

where $\mathbf{X}$ is the $N\times p$ model matrix.

If the experiment is run in two stages, with n_1 observations generated from the first stage design $\mathbf{X}_1$ and the remaining $n_2=(N-n_1)$ observations generated by the second stage design $\mathbf{X}_2$, an equivalent form of (4.1.1) is

$$L_{1,2}(\mathbf{y}_1,\mathbf{y}_2,\beta|\mathbf{X}_1,\mathbf{X}_2) = L_1(\mathbf{y}_1,\beta|\mathbf{X}_1)\, L_{2|1}(\mathbf{y}_2,\beta|\mathbf{y}_1,\mathbf{X}_1,\mathbf{X}_2). \tag{4.1.2}$$

In other words, the joint likelihood of the two stages is equal to the product of the first stage likelihood and the second stage conditional likelihood. Taking the log of both sides of 4.1.2 results in

$$\log L_{1,2} = \log L_1 + \log L_{2|1} \tag{4.1.3}$$

Critical to the application of optimality criteria in the two stage procedure is the Fisher information matrix which can easily be derived from (4.1.3). Simply speaking, the combined information is the sum of the information from both stages. That is

$$I_{1,2}(\beta) = I_1(\beta) + I_{2|1}(\beta) \tag{4.1.4}$$

$$= (1/\sigma^2)[\mathbf{X}_1'\mathbf{X}_1 + \mathbf{X}_2'\mathbf{X}_2]$$

where $\mathbf{X}_1$, of dimension $n_1 \times p$ and $\mathbf{X}_2$, of dimension $n_2 \times p$ are the first and second stage designs, respectively, expressed in model form. The strategy behind designing in two stages, however, is to generate information about β (through parameter estimates, for example) from the first stage in order to maximize the combined information through the choice of $\mathbf{X}_2$ (conditional on $\mathbf{X}_1$, $\mathbf{I}_1$, $\mathbf{y}_1$, but independent of β).

Design optimality criteria are extremely useful in generating two stage designs. A two stage procedure may implement any pair of design criteria which meet the first stage objective as well as the objective of the combined design. For example, Abdelbasit and Plackett (1983) and Minkin (1987) studied the efficiency of two stage D-optimal designs for binary responses, thus applying D-optimality to both stages. Myers, Myers, Carter and White (1996) developed a two stage procedure for the logistic regression model which utilizes D-optimality in the first stage followed by Q-optimality in the second. Following the convention set by Myers, et al., these two procedures will be called D-D optimality and D-Q optimality, respectively. To illustrate the two stage method, the D-D and D-Q procedures will now be presented.

The first step in both the D-D and D-Q procedures is the selection of a first stage D-optimal design. Recall the D-optimality criterion introduced in Chapter 2,

$$\underset{D}{\text{Min}} \; | \; N\,[I(\beta)]^{-1} \; |$$

where N is the design size and D is the set of all possible designs (no size restriction). As in the case of nonlinear models such as the logistic (addressed by Myers et al.), $I(\beta)$ may depend on β. In order to implement D-optimality in the first stage, the experimenter must estimate the unknown β with a best guess, $\mathbf{b_0}$. The first stage D-optimality criterion can then be written

$$\underset{D}{\text{Min}} \; | \; n_1[\mathbf{I}_1(\beta)]^{-1} \; |_{\beta=\mathbf{b}0}$$

with β replaced by $\mathbf{b}_0$, and D representing all possible designs of size n_1.

After design and execution of the first stage experiment, n_1 observations are available to estimate β. The 'best guess' of β is updated by replacing $\mathbf{b}_0$ with the MLE of β. The second stage of the two stage process will utilize $\mathbf{b}$, thus making it *conditional* on the results of the first stage.

To complete the D-D procedure, it is necessary to choose a set of n_2 second stage design points which will create a combined design which is conditionally D-optimal. The n_2 points are chosen by

$$\underset{D}{\text{Min}} \; | \; N(\mathbf{I}_1(\beta) + \mathbf{I}_{2|1}(\beta))^{-1} \; |_{\beta=\mathbf{b}} \qquad (4.1.5)$$

where D is now the set of all possible designs of size n_2 and $\mathbf{I}_1(\beta)$ is fixed after the first stage.

To complete the D-Q procedure, a set of n_2 second stage design points is chosen to create a combined design which is conditionally Q-optimal. When there is uniform interest over the region, *R*, the criterion for selecting the n_2 points is given by

$$\underset{D}{\text{Min}} \; N\,k^{-1} \int_R \mathbf{x}'^{(m)} (\mathbf{I}_1(\beta) + \mathbf{I}_{2|1}(\beta))^{-1} \mathbf{x}^{(m)} \; \text{var}(y(\mathbf{x})) \; d\mathbf{x} \; |_{\beta=\mathbf{b}}$$

where *D* is the set of all designs of size n_2, $k=\int d\mathbf{x}$ is the volume of *R*, N = (n1 + n2) and $\mathbf{I}_1(\beta)$ is fixed.

In addition to the choice of optimal design points, other important issues related to the two stage design procedure are sample size allocation (ie, choice of n_1/N and n_2/N) and methods for evaluating the efficiency of the procedure. These issues will be addressed collectively for both non-Bayesian and Bayesian procedures in section 4.3.

4.2 Bayesian Two Stage Design

Development of Bayesian two stage design procedures has been even more limited than that of the non-Bayesian procedures. In 1966, Draper and Hunter developed a method for selecting a Bayesian D-optimal second stage design when the first stage data was already available. Although the method was developed for a multiresponse situation, their use of a non-informative prior makes the procedure identical to a non-Bayesian procedure, which is given in (4.1.5) for the single response case. In 1995, Letsinger developed a two stage procedure for designing D-optimal designs for the logistic regression model. Bayesian D-optimality was used to select the first stage design but non-Bayesian D-optimality was applied to the second stage. Neither procedure is considered to be a Bayesian two stage design procedure, since Bayesian design criteria are not used to select both the first and second stage designs. To illustrate the use of

Bayesian methods in both stages, development of a two stage procedure for the normal linear model will now be demonstrated.

Consider the linear model $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, where $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$ is the N dimensional vector of responses from a two stage design and $\mathbf{X}$ is the N×p model matrix for the first and second stage combined. It is assumed that $\mathbf{y}_i|\beta,\sigma^2 \sim N(\mathbf{X}_i\beta, \sigma^2\mathbf{I})$ for each stage i, with n_1 observations to be collected in stage 1 and the remaining n_2 observations coming from stage 2. Suppose that the information on β prior to the first stage experiment is that $\beta|\sigma^2,\tau^2 \sim N(\beta_0, \sigma^2\mathbf{T}^{-1})$ where $\mathbf{T} = (1/\tau^2)\mathbf{I}_p$ for a known τ. Under these distribution assumptions, the first stage posterior distribution of β is normal,

$$\beta|\sigma^2,\tau^2,\mathbf{y_1} \sim N(\mathbf{b_1}, \sigma^2(\mathbf{X_1}'\mathbf{X_1} + \mathbf{T})^{-1})$$

with mean $\mathbf{b_1} = (\mathbf{X_1}'\mathbf{X_1} + \mathbf{T})^{-1}(\mathbf{X_1}'\mathbf{y} + \mathbf{T}\beta_0)$. The second stage posterior of β is also normal,

$$\beta|\sigma^2,\tau^2,\mathbf{y_1}, \mathbf{y_2} \sim N(\mathbf{b_2}, \sigma^2(\mathbf{X_1}'\mathbf{X_1} + \mathbf{X_2}'\mathbf{X_2} + \mathbf{T})^{-1}),$$

with $\mathbf{b_2} = (\mathbf{X}'\mathbf{X} + \mathbf{T})^{-1}(\mathbf{X}'\mathbf{y} + \mathbf{T}\beta_0)$.

Let $L(\beta,\delta_1,\mathbf{y_1})$ be any loss function measuring the performance of a first stage design δ_1. Following Lindley's method, the best first stage design is the one that minimizes the Bayes risk,

$$R(\delta_1) = \int L(\beta,\delta_1,\mathbf{y_1})\, p(\beta,\mathbf{y_1}|\delta_1)\, d\beta\, d\mathbf{y_1}, \qquad (4.2.1)$$

where $p(\beta,\mathbf{y_1}|\delta_1) = p(\beta|\sigma^2,\tau^2, \mathbf{y_1}, \delta_1)\, m(\mathbf{y}_1)$, the product of the first stage posterior of β and the marginal distribution of $\mathbf{y}_1$. As an example of applying (4.2.1), let $L(\beta,\delta_1,\mathbf{y_1})$ be the negative Kullback-Leibler distance between the first stage posterior and prior distributions . The corresponding Bayes risk is

$$R^*(\delta_1) = \int -\log \frac{p(\beta|\sigma^2,\tau^2,\mathbf{y}_1,\delta_1)}{p(\beta)}\, p(\beta|\sigma^2,\tau^2, \mathbf{y_1}, \delta_1)\, m(\mathbf{y}_1)\, d\beta\, d\mathbf{y_1}.$$

As shown in section 3.3, minimizing R* for the normal linear model is equivalent to minimizing the log determinant of the posterior covariance matrix of β.

$$R^*(\delta_1) \qquad \log | [\ \sigma^2(\mathbf{X_1}'\mathbf{X_1} + \mathbf{T})^{-1}\ ] |.$$

Therefore, minimizing $R^*(\delta_1)$ results in finding the Bayes D-optimal first stage design.

To find the optimal second stage design, let $L(\beta,\delta_2,\mathbf{y}_1,\mathbf{y_2})$ be any loss function measuring the performance of a second stage design δ_2 in the presence of δ_1. For the optimal second stage design δ_2^*,

$$R(\delta_2^{*}) \;=\; \int L(\beta,\delta_2,\mathbf{y}_1,\mathbf{y_2})\ p(\beta,\mathbf{y_2}|\delta_2)\ d\beta\ d\mathbf{y_2}\ , \qquad (4.2.2)$$

where $p(\beta,\mathbf{y_2}|\delta_2) = p(\beta|\sigma^2,\tau^2,\ \mathbf{y_1},\ \mathbf{y}_2,\ \delta_1)\ m(\mathbf{y})$, the product of the second stage posterior of β and the marginal distribution of $\mathbf{y}$. Again, as an example of applying (4.2.2), let $L(\beta,\delta_1,\mathbf{y_1},\mathbf{y_2})$ be the negative Kullback-Leibler distance between the second stage posterior and prior distributions. Since the most current information about β just prior to observing y_2 is reflected in the first stage posterior distribution, $p(\beta|\sigma^2,\tau^2,\ \mathbf{y_1},\ \delta_1)$ can be used as the second stage prior. The Bayes risk associated with the second stage design then becomes

$$R^*(\delta_2) \;=\; \int -\log\ \frac{p(\beta|\sigma^2,\tau^2,\mathbf{y}_1,\mathbf{y}_2,\delta_1)}{p\left(\beta\,|\,\sigma^2,\tau,\mathbf{y}_1,\delta_1\right)}\ p(\beta|\sigma^2,\tau^2,\ \mathbf{y_1},\ \mathbf{y}_2,\ \delta_2)\ m(\mathbf{y})\ d\beta\ d\mathbf{y}$$

$$\log \det[\mathrm{Var}(\beta|\sigma^2,\tau^2,\ \mathbf{y_1},\ \mathbf{y}_2,\ \delta_2)]$$

$$=\ \log | \sigma^2(\mathbf{X_1}'\mathbf{X_1} + \mathbf{X_2}'\mathbf{X_2} + \mathbf{T})^{-1} | .$$

Minimizing $R^*(\delta_2)$ results in finding the Bayes two stage D-optimal design.

4.3 Evaluation of Two Stage Designs

Extending the evaluation method used by Myers, et al. (1996), the evaluation of non-Bayesian two stage designs is as follows. In the two stage procedure, $\mathbf{X}_2$ is a random variable dependent on first stage parameter estimates. As a result, the **average** performance of the two stage procedure must be determined, where the performance measure would typically be the second stage design criterion.

From the laws of probability it is known that

$$\mathrm{Var}(\mathbf{b}) = E_{\mathbf{X}_2}\mathrm{Var}(\mathbf{b}|\,\mathbf{X}_2) + \mathrm{Var}_{\mathbf{X}_2}E(\mathbf{b}|\,\mathbf{X}_2)\,. \qquad (4.3.1)$$

If **b** is the MLE of β, then (4.3.1) can be used in conjunction with a simulation procedure to find the asymptotic variance-covariance properties of **b**. Since the second term on the right hand side of (4.3.1) is zero, the Var(**b**) for the two stage procedure can be found by averaging Var($\mathbf{b}|\mathbf{X}_2$) over numerous simulated two stage designs. Since this quantity may depend on β, a fixed value of β is assumed throughout the evaluation. Knowledge of Var(**b**) results ultimately in knowing the average performance of the two stage procedure. For example, for the D-D procedure discussed in section 4.1, the performance measure is the generalized variance of **b**. The generalized variance of the two stage procedure is found by taking the determinant of Var(**b**).

It is not entirely clear what the 'best' method is for evaluating Bayesian two stage procedures. One approach is to repeat the above method for various values of β, as done by Letsinger (1995). In this manner, the two stage designs are evaluated over a range of β values extending beyond the first stage prior.

Using the methods just described, the performance of two stage procedures can be evaluated relative to single stage designs or other two stage procedures. For example, one method of determining how sample allocation affects a given two stage procedure is to evaluate the procedure over a range of (n_1/N, n_2/N) combinations. For example, Myers, et al. (1996) used sample allocation pairs (.3,.7), (.5,.5) and (.7,.3) when evaluating the average performance of their D-Q procedure. In most cases, the best average performance resulted when 70% of the observations were reserved for the second stage. Letsinger actually modeled the average performance of his procedure as a function of sample size and allocation, obtaining results similar to those of Myers, et al. (1996).

Chapter 5

Bayesian Two Stage Design Approach Under Model (Regressor) Uncertainty

Design procedures like those discussed in previous chapters allow one to efficiently generate data for an assumed model $\mathbf{y} = f(\boldsymbol{x}^{(m)}, \beta) + \varepsilon$. The model assumptions include the form of f, the set of regressors, $\boldsymbol{x}^{(m)}$, and the distribution of ε. The nature of the response, y, often provides information about the model form (f) and the error distribution. It is more difficult to know, apriori, the specific set of regressors which will best explain the relationship between the response and a set of design (control) variables $\boldsymbol{x}$. Misspecification of $\boldsymbol{x}^{(m)}$ will result in a design which is efficient, but for the wrong model.

A Bayesian two stage design approach makes it possible to efficiently design experiments when initial knowledge of $\boldsymbol{x}^{(m)}$ is poor. This is accomplished by using a Bayesian *D*-optimality criterion in the first stage which works well under model uncertainty. Bayesian analysis of first stage data reduces uncertainty associated with $\mathbf{x}^{(m)}$, enabling the remaining design points (second stage design) to be chosen with greater efficiency. The second stage design is then generated from any optimality procedure which incorporates the improved model knowledge. Using this approach, numerous two stage design procedures have been developed under the assumption of a linear model form with normal error distribution. Six of these procedures will be presented in detail in this chapter.

5.1 First Stage Design and Analysis

Suppose that an experimenter identifies a set of regressors, $\boldsymbol{x}^{(f)}$, containing all regressors he/she believes might be needed in modeling the behavior of a response, y. This set of regressors defines the *full model* , $\mathbf{y} = f(\mathbf{x}^{(f)}, \beta) + \varepsilon$, for which the first stage experiment will be designed.

Assuming a linear model form, the full model is written as $\mathbf{y}_1 = \mathbf{X}_1\beta + \varepsilon_1$, with $\mathbf{y}_1$ denoting the n_1 observations to be collected via the first stage experiment. It is assumed that $\mathbf{y}_1|\beta,\sigma^2 \sim N(\mathbf{X}_1\beta, \sigma^2\mathbf{I})$. The model matrix, $\mathbf{X}_1$, has dimension $n_1 \times (p+q)$ with the p+q columns defined by the full set of regressors $\boldsymbol{x}^{(f)} = \{ x_1, x_2, ..., x_p, x_{(p+1)}, ..., x_{(p+q)} \}$, with $x_1=1$ denoting the intercept term. It is assumed that prior to observing $\mathbf{y}_1$, the experimenter has knowledge of the process or system that allows him or her to identify p of the regressors as **primary terms**. These are terms that the experimenter strongly believes are needed in modeling the response. The remaining q regressors are the **potential terms**, i.e., those terms about which the experimenter has uncertainty. For example, the experimenter may know from past experience that certain process variables must be included in the model as main effects (i.e. linear terms) but is uncertain if higher order terms (such as quadratics) are needed. Let β_{pri} (of dimension $p \times 1$) and β_{pot} (of dimension $q \times 1$) be the model parameters attached to primary and potential terms, respectively.

5.1.1 Bayesian *D*-optimality as a First Stage Criterion

The challenge of the first stage is to design an experiment supporting estimation of β_{pri} and β_{pot}, while saving as many resources as possible for the second stage experiment. This is accomplished by using a Bayesian D-optimality criterion proposed by DuMouchel and Jones (1994) for the efficient study of both primary and potential terms. Following their approach, the following distributional assumptions are made. A diffuse prior distribution (i.e., arbitrary prior mean with infinite prior variance) is assumed for β_{pri}. This is reasonable since these parameters are expected to be significantly different from zero, but no assumption of direction is made. The

potential terms, however, are perceived to have smaller coefficients. For this reason, β_{pot} is assigned a $N(\mathbf{0}, \sigma^2\tau^2\mathbf{I})$ prior distribution, with τ^2 and σ^2 known. Fortunately, the first stage design can be constructed independently of σ^2. The value of τ^2, however, affects the choice of design, since it reflects the degree of uncertainty associated with potential terms, relative to σ^2.

Before defining the joint prior distribution of β_{pri} and β_{pot}, it is necessary to discuss a scaling convention recommended by DuMouchel and Jones. Traditionally, each linear regressor x_j is scaled such that $\max[x_{jk}] = +1$ and $\min[x_{jk}] = -1$, for $k = 1,...,n$. The resulting scale of any quadratic term x_j^2 would then be $(x_j^2)_k \in [0,1]$ for $k = 1,...,n$. Although the traditional scaling attempts to reduce correlation between regressors, often the intercept and quadratic terms are still slightly correlated. DuMouchel and Jones recommend a modification to this scaling convention for the purpose of eliminating (as much as possible) any correlation between primary and potential terms. In practical applications, an 'optimal' design is constructed from a set of candidate points (defining a grid over the continuous design space), therefore the scaling can be accomplished as follows. Let $\boldsymbol{X} = [\boldsymbol{X}_{pri}|\boldsymbol{X}_{pot}]$ represent the set of candidate points in model space, partitioned by the two sets of terms. Regressing the potential terms on the primary terms results in $\alpha = (\boldsymbol{X}_{pri}' \boldsymbol{X}_{pri})^{-1} \boldsymbol{X}_{pri}' \boldsymbol{X}_{pot}$, the alias matrix measuring how confounded the primary and potential terms are over the candidate set. Replacing $\boldsymbol{X}_{pot}$ with the residual matrix, $\mathbf{R} = \boldsymbol{X}_{pot} - \boldsymbol{X}_{pri}\alpha$, or a scaled version of $\mathbf{R}$, essentially rescales the model space so as to make the primary and potential terms nearly uncorrelated.

Under the assumption that primary and potential terms are uncorrelated, the joint prior distribution assigned to β_{pri} and β_{pot} is the $N(\mathbf{0},\sigma^2\tau^2\mathbf{K}^{-1})$ distribution, where $\mathbf{K}$ is a $(p+q)\times(p+q)$ diagonal matrix whose first p diagonal elements equal 0 and remaining q diagonal elements equal 1. This results in infinite prior variance for the p primary β`s, corresponding to the uninformative prior assigned to β_{pri} (i.e., Uniform$(-\infty,+\infty)$ prior). Since it is assumed $\mathbf{y}_1|\beta,\sigma^2 \sim N(\mathbf{X}_1\beta, \sigma^2\mathbf{I})$, it can

be shown that the first stage posterior distribution of β is also normal, with mean $\mathbf{b_1} = (\mathbf{X_1}'\mathbf{X_1} + \mathbf{K}/\tau^2)^{-1}\,\mathbf{X_1}'\mathbf{y}$ and posterior variance $\mathbf{V_1} = \sigma^2(\mathbf{X_1}'\mathbf{X_1} + \mathbf{K}/\tau^2)^{-1}$ (Box and Tiao, 1973).

Following the method given in section 4.2, the first stage Bayes D-optimal design is found by choosing the design that minimizes the Bayes risk, proportional to

$$\log|\mathbf{V_1}| = \log|\sigma^2(\mathbf{X_1}'\mathbf{X_1} + \mathbf{K}/\tau^2)^{-1}|\,.$$

In practice, this is done by selecting the rows of $\mathbf{X}_1$ from $\boldsymbol{X}$ so that $|\mathbf{V_1}|$ is minimized. Notice that the diagonals of $\mathbf{V_1}$ associated with β_{pot} are somewhat stabilized through prior information (given through τ). The other diagonals of $\mathbf{V_1}$ (those associated with β_{pri}) are more dependent on design. The resulting first stage Bayes D-optimal design is one which will support estimation of both β_{pri} and β_{pot}, but with higher priority given to β_{pri}.

As τ approaches infinity, the Bayes D-optimal design becomes equivalent to the non-Bayesian D-optimal design with p+q terms, treating all p+q terms equally. During preliminary investigation, setting τ to be as much as 10 still resulted in a design which differs slightly from the non-Bayesian design produced by DETMAX. Further study is needed to determine the affect of τ on the effectiveness of the combined (two-stage) design.

5.1.2 Bayesian Q-optimality as a First Stage Criterion

When predictive properties of the response surface model are given a higher priority than parameter estimation, Bayesian Q-optimality can be used to generate the first stage design. This is done by implementing the approach of DuMouchel and Jones, as discussed in (5.1.1), replacing the Bayesian D-optimality criterion with the Bayesian Q criterion.

Recall from the previous section that the joint prior distribution assigned to β_{pri} and β_{pot} is the $N(\mathbf{0},\sigma^2\tau^2\mathbf{K}^{-1})$ distribution, where $\mathbf{K}$ is a (p+q)×(p+q) diagonal matrix whose first p diagonal elements equal 0 and remaining q diagonal elements equal 1. Under the assumption $\mathbf{y}_1|\beta,\sigma^2$ $N(\mathbf{X}_1\beta, \sigma^2\mathbf{I})$, the first stage posterior distribution of β is also normal, with mean $\mathbf{b_1} = (\mathbf{X_1}'\mathbf{X_1} + \mathbf{K}/\tau^2)^{-1}\ \mathbf{X_1}'\mathbf{y}$ and posterior variance $\mathbf{V_1} = \sigma^2(\mathbf{X_1}'\mathbf{X_1} + \mathbf{K}/\tau^2)^{-1}$.

Assuming that there is uniform interest in prediction over the entire design region R, recall that the Bayesian Q criterion, developed in section 3.2, is given by

$$\underset{D}{\text{Min}} \quad \text{tr}\ \left[\text{Var}(\beta|\mathbf{y},\sigma^2,\tau^2)\ (N/k) \int_R \mathbf{x}^{(m)}\mathbf{x}'^{(m)}\ d\mathbf{x}\ \right].$$

Therefore, the first stage Bayes Q-optimal design is found by choosing the N rows of $\mathbf{X}_1$ in order to minimize

$$Q = \text{tr}\left[\sigma^2(\mathbf{X_1}'\mathbf{X_1} + \mathbf{K}/\tau^2)^{-1}\ (N/k) \int_R \mathbf{x}^{(m)}\mathbf{x}'^{(m)}\ d\mathbf{x}\ \right].$$

5.1.3 First Stage Analysis

Prior to observing the first stage data, the experimenter has specified a set of p+q regressors defining the full model. The 'best' model (that which best describes the response behavior) is believed to contain the p primary terms and some subset of the potential terms. Uniform uncertainty about the q potential terms results in a set of m+1 candidate models, each being a candidate for the 'best' model, with $m = \sum_{i=1}^{q} \binom{q}{i}$.

The purpose of the first stage analysis is to use the information in the data in conjunction with prior information to reduce model uncertainty. This can be done by ranking/scoring the p+q model terms based on their importance in explaining the response. Equivalently, model knowledge can be gained by scoring each of the candidate models according to its likelihood of

being the 'best' model. Either way, the resulting scores can be used as inputs to a second stage criterion, ensuring that remaining data points will provide information on important terms and/or models. In sections 5.2 - 5.7 which follow, the details of the first stage analysis will be given for each of the six Bayesian two stage design procedures presented.

5.2 Bayesian D-D Optimality Procedure

(Based on First Stage Parameter Estimates)

In this section, the second stage of a Bayesian two stage D-optimality procedure is developed. The first stage design is chosen according to the method of DuMouchel and Jones, presented in section 5.1.1. The second stage design is again chosen according to the method in section 5.1.1, however, this time based on a more informative prior distribution. The resulting Bayesian D-D procedure provides a method of constructing D-optimal designs when initial model knowledge is poor.

Initially, the p primary terms were specified to have equal importance. Similarly, the q potential terms were treated with uniform uncertainty and as a group were believed to be less important than the primary terms. This was reflected through the first stage prior distribution, $\beta|\sigma^2,\tau^2 \quad N(\mathbf{0},\sigma^2\tau^2\mathbf{K}^{-1})$. After analyzing the first stage data, there is additional information about the relative importance of the p+q regressors. One approach to sorting and ranking the p+q model terms is to compare the magnitude of their estimated effects. Since the estimated effect for any regressor x_j is proportional to its standardized estimated coefficient,

$$\hat{\beta}_j^* = \frac{\hat{\beta}_j}{\sqrt{\text{var}(\hat{\beta}_j)}}$$

the relative importance of the various model terms can be estimated by the relative sizes of the $\hat{\beta}_j^*$'s (in absolute value).

Basing inferences on the first stage posterior distribution of β, the Bayes estimator of β_j is its posterior mean, $E(\beta_j|\mathbf{y}_1)$, equal to the j^{th} element of $\mathbf{b_1} = (\mathbf{X_1}'\mathbf{X_1} + \mathbf{K}/\tau)^{-1}\ \mathbf{X_1}'\mathbf{y}$. The standard error of $\hat{\beta}_j$ is $\sigma\sqrt{c_{jj}}$, where c_{jj} is the j^{th} diagonal element of $(1/\sigma^2)\mathbf{V_1} = (\mathbf{X_1}'\mathbf{X_1} + \mathbf{K}/\tau^2)^{-1}$. Thus the p+q standardized estimates of model parameters (coefficients) are

$$\hat{\beta}_j^* = \frac{E\left(\beta_j \middle| \mathbf{y}_1\right)}{\sigma\sqrt{c_{jj}}}, \quad j = 1, 2, ..., p+q .$$

Normalizing these estimates (in absolute value) results in a set of discrete scores or 'weights of evidence' which quantify the relative importance of each model term,

$$p(x_j) = \frac{\left|\hat{\beta}_j^*\right|}{\sum_{j=1}^{p+q}\left|\hat{\beta}_j^*\right|}, \quad j = 1, 2, ..., p+q$$

These scores are built into a second stage prior distribution for β after some initial rescaling. First the regressor scores are scaled so that $\max[p(x_j)]= 1$ (i.e., removing any requirement that the scores sum to 1). The scores are then scaled once more to put them back in the same scale as the first stage prior standard deviation. The resulting score for a regressor x_j is τ_j, where

$$\tau_j = \tau\ [p(x_j) / \max p(x_j)\ ], \quad j = 1, 2, ..., p+q$$

and τ is the known value from the first stage prior. For example, assuming τ=10 in the first stage, the resulting τ_j's range from 0 to 10, so that the most important term's prior variance is very large (approaching infinity).

Prior to observing the second stage data $\mathbf{y}_2$, beliefs about the relative importance of the p+q model terms are expressed as

$$\beta|\sigma^2, \tau^2, \mathbf{y}_1 \quad N(\mathbf{0}, \sigma^2\mathbf{T}),$$

where T is a (p+q)×(p+q) diagonal matrix with $\tau_1, \tau_2, \ldots, \tau_{p+q}$ appearing on the diagonals. Setting the prior mean to **0** is arbitrary at this point as it will have no impact on the second stage design criterion. It is also assumed that $\mathbf{y}_2|\beta,\sigma^2 \quad N(\mathbf{X}_2\beta,\sigma^2 I)$, where $\mathbf{X}_2$ represents a second stage design with columns expanded to contain all p+q regressors. From these distribution assumptions, the resulting second stage posterior distribution of β is also normal, with posterior covariance matrix $\mathbf{V}_2 = \sigma^2(\mathbf{X_1}'\mathbf{X_1} + \mathbf{X_2}'\mathbf{X_2} + \mathbf{T}^{-1})^{-1}$.

Thus, the second stage conditionally D-optimal design is found by selecting the n_2 rows of $\mathbf{X}_2$ from a candidate list $\boldsymbol{X}$ such that $|\mathbf{V}_2|$ is minimized, i.e.,

$$\underset{\mathbf{X}_2 \quad \boldsymbol{X}}{\text{Min}} \quad |(\mathbf{X_1}'\mathbf{X_1} + \mathbf{X_2}'\mathbf{X_2} + \mathbf{T}^{-1})^{-1}| \ .$$

Due to the structure of $\mathbf{T}^{-1}$, the diagonals of $\mathbf{V}_2$ corresponding to less important regressors are already somewhat stabilized. The design points in $\boldsymbol{X}$ which provide information about the more important regressors (and thus stabilize the corresponding diagonals) will be those chosen for the second stage design.

5.3 Bayesian D-D Optimality Procedure

(A Linear Combination of D-optimality Criteria)

Another approach for generating a Bayesian two stage design is given in this section. As with the previous procedure, the first stage design is chosen from the Bayesian D-optimality criterion presented in section 5.1.1. Letting the first stage posterior become the second stage prior distribution for β, the Bayesian D-optimality criterion can be used again in the second stage. The second stage design is chosen to minimize a weighted average of m+1 D-optimality

criteria - averaging over the weighted set of candidate models. Model weights are posterior probabilities, calculated according to the method of Box and Meyer (1993).

The Box and Meyer method of calculating posterior probabilities for candidate models is as follows (in the context of the first stage analysis). Analogous to the approach of all-subsets regression, consider the set of m+1 candidate models, described previously. Label these models as M_0, M_1, ..., M_m, with M_0 denoting the model containing primary terms only. Each model M_i contains the parameters $\beta_i = \begin{bmatrix} \beta_{PRI} \\ \beta_{POT(i)} \end{bmatrix}$, where $\beta_{POT(0)}$ is the null vector. The predictive density of $\mathbf{y}_1$, given model M_i, is written $f(\mathbf{y}_1|M_i)$, given by the expression

$$f(\mathbf{y}_1|M_i) = \int_{\beta_i} f(\mathbf{y}_1|M_i, \beta_i)\, f(\beta_i|\, M_i)\, d\beta_i \quad , \tag{5.3.1}$$

where β_i is the set of all possible values of β_i, $f(\mathbf{y}_1|M_i, \beta_i)$ is the sampling distribution of $\mathbf{y}_1$, given the model M_i, and $f(\beta_i|\, M_i)$ is the prior probability density of β_i. Given the first stage data, $\mathbf{y}_1$, the posterior probability of the model M_i becomes

$$p(M_i|\mathbf{y}_1) = \frac{p(M_i)\, f(\mathbf{y}_1|M_i)}{\sum_{j=0}^{m} p(M_j)\, f(\mathbf{y}_1|M_j)} \quad , \tag{5.3.2}$$

when the prior probability of M_i is $p(M_i)$. The $p(M_i|\mathbf{y}_1)$ quantifies the likelihood that M_i is the 'best' model, upon observation of the first stage data. Thus, calculating $p(M_i|\mathbf{y}_1)$ for i=0, 1, ..., m, results in a discrete probability density defined over the set of candidate models.

Each candidate model M_i contains q_i potential terms, $(0 \le q_i \le q)$ and all p primary terms. Since there is uniform uncertainty about all potential terms, prior to observing $\mathbf{y}_1$, the prior probability of model M_i being the 'best' model is given by

$$p(M_i) = \pi^{q_i} (1 - \pi)^{q - q_i} \quad , i = 0, 1, \ldots , m ,$$

for a specified probability, π. The value of π should be chosen to represent the proportion of potential terms believed to be active, or equivalently, the likelihood of a potential term being active, relative to a primary term. Box and Meyer found that sensible results can be achieved with a nominal value of π=0.25. A more conservative approach would be to evaluate the posterior model probabilities for each of several values of π, such as 0.20, 0.30, and 0.50.

Since we have assumed a normal linear model,

$$f(\mathbf{y}_1|M_i,\sigma,\beta_i) \qquad \sigma^{-n_1} \exp[-(\mathbf{y}_1-\mathbf{X}_i\beta_i)'(\mathbf{y}_1-\mathbf{X}_i\beta_i)/2\sigma^2] .$$

Assuming that $\beta|\sigma^2,\tau^2 \quad N(\mathbf{0}, \sigma^2\tau^2 K^{-1})$ and integrating as in (5.3.1) (see Box and Meyer for details) results in

$$f(\mathbf{y}_1|M_i) \quad \tau^{-q_i} \, |\mathbf{T}_i + \mathbf{X}_i'\mathbf{X}_i|^{-1/2} \, (S(\hat{\beta}_i) + \hat{\beta}_i' \, \mathbf{T}_i \, \hat{\beta}_i)^{-(n_1-1)/2}$$

where

$\mathbf{X}_i$ = first stage design in model M_i space

$$\mathbf{T}_i = \begin{bmatrix} \frac{1}{\tau^2}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_i \end{bmatrix}$$

$\hat{\beta}_i = E[\beta_i|\mathbf{y}_1]$, assuming Model M_i

and

$S(\hat{\beta}_i)$ = SS(error) for Model M_i .

The resulting posterior probability for model M_i becomes

$$p(M_i|\mathbf{y}_1) = C \, \pi^{q_i} (1-\pi)^{q-q_i} \, \tau^{-q_i} \, |\mathbf{T}_i + \mathbf{X}_i'\mathbf{X}_i|^{-1/2} \, (S(\hat{\beta}_i) + \hat{\beta}_i' \, \mathbf{T}_i \, \hat{\beta}_i)^{-(n_1-1)/2} \qquad (5.3.3)$$

where C is the normalization constant (i.e., the sum of the posterior model probabilities).

The second stage criterion will now be developed. Recall that in the first stage it is assumed that $\beta|\sigma^2,\tau^2 \quad N(\mathbf{0}, \sigma^2\tau^2\mathbf{K^{-1}})$ and $\mathbf{y}_1|\beta,\sigma^2 \quad N(\mathbf{X}_1\beta, \sigma^2\mathbf{I})$, resulting in the first stage posterior which is $N(\mathbf{b_1}, \mathbf{V_1})$, with $\mathbf{b_1} = (\mathbf{X_1}'\mathbf{X_1} + \mathbf{K}/\tau^2)^{-1} \, \mathbf{X_1}'\mathbf{y}$ and $\mathbf{V_1} = \sigma^2(\mathbf{X_1}'\mathbf{X_1} + \mathbf{K}/\tau^2)^{-1}$. Let the second stage

prior distribution of β be the first stage posterior. Also assume that $\mathbf{y}_2$, the n_2 second stage observations, comes from a normal sampling distribution, namely, $\mathbf{y}_2|\beta,\sigma^2 \quad N(\mathbf{X}_2\beta,\sigma^2 I)$, where $\mathbf{X}_2$ is a second stage design expanded to full model form. It can then be shown that the second stage posterior distribution of β is also normal, with posterior covariance matrix $\mathbf{V}_2 = \sigma^2(\mathbf{X_1}'\mathbf{X_1} + \mathbf{X_2}'\mathbf{X_2} + \mathbf{K}/\tau^2)^{-1}$. It is important to note that β contains all p+q parameters of the full model, and therefore the Bayes D-D optimal design for the full model can be found by choosing $\mathbf{X}_2$ such that

$$\underset{\mathbf{X}_2 \quad \boldsymbol{X}}{\text{Min}} \quad |\, N(\mathbf{X_1}'\mathbf{X_1} + \mathbf{X_2}'\mathbf{X_2} + \mathbf{K}/\tau^2)^{-1} \,|$$

The full model, however, is only one of the candidate models and in most cases is not the 'best' model. Consider any of the subset models M_i (i=0, 1, ... , m) as defined previously, with each model M_i identified by its parameters β_i. The posterior variance of β_i is

$$\mathbf{V}_{2(i)} = \sigma^2(\mathbf{X_1}^{(i)}{}'\mathbf{X_1}^{(i)} + \mathbf{X_2}^{(i)}{}'\mathbf{X_2}^{(i)} + \mathbf{K}^{(i)}/\tau^2)^{-1} \qquad (5.3.4)$$

where $\mathbf{X_1}^{(i)}$ and $\mathbf{X_2}^{(i)}$ are the first and second stage design matrices, respectively, expanded to model M_i space (i.e., containing only the $p+q_i$ regressors in M_i) and $\mathbf{K}^{(i)} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_i \end{bmatrix}$ is of dimension $(p+q_i) \times (p+q_i)$. A Bayes D-D optimal design for model M_i is the set of design points $\mathbf{X_2}$ which minimizes $D_i = |\mathbf{V}_{2(i)}|$.

Utilizing the Box and Meyer posterior probabilities for models $M_0, M_1, \ldots, M_m$ (5.3.3), a weighted average D criterion can be used to select the second stage design. The objective is to choose the second stage design points so as to minimize D_i for each model M_i having a high probability of being the 'best' model. This is done by choosing the second stage design so as to minimize

$$\underset{\mathbf{X}_2 \quad \boldsymbol{X}}{\text{Min}} \quad \sum_{Mi} D_i \, p(M_i|\mathbf{y}_1)$$

where $\mathbf{X}_2$ and $\boldsymbol{X}$ are still in full model space. It is important to point out that the second stage design, just like the first, contains any design (control) variables included in the full model.

5.4 Bayesian D-Q Optimality Procedure

(A Linear Combination of Q-optimality Criteria)

A Bayesian D-Q optimality procedure is developed in this section. Once again, the first stage design is chosen to be D-optimal, according to the procedure discussed in section 5.1.1. The second stage design is chosen to minimize the integrated prediction variance, which has been averaged over a set of candidate models. Box and Meyer posterior probabilities are used to weight the candidate models, just as in the second stage of the D-D procedure presented in section 5.3.

Recall that in the first stage it is assumed that $\beta|\sigma^2,\tau^2 \quad N(\mathbf{0}, \sigma^2\tau^2\mathbf{K}^{-1})$ and $\mathbf{y}_1|\beta,\sigma^2 \quad N(\mathbf{X}_1\beta, \sigma^2\mathbf{I})$, resulting in the first stage posterior which is also normal with posterior covariance matrix $\mathbf{V_1} = \sigma^2(\mathbf{X_1}'\mathbf{X_1} + \mathbf{K}/\tau^2)^{-1}$. Using this first stage posterior as the prior for the second stage and assuming $\mathbf{y}_2|\beta,\sigma^2 \quad N(\mathbf{X}_2\beta,\sigma^2\mathbf{I})$, it can then be shown that the second stage posterior distribution of β is also normal, with posterior covariance matrix $\mathbf{V}_2 = \sigma^2(\mathbf{X_1}'\mathbf{X_1} + \mathbf{X_2}'\mathbf{X_2} + \mathbf{K}/\tau^2)^{-1}$. Note that β contains all p+q parameters of the full model and therefore $\mathbf{X}_1$ and $\mathbf{X}_2$ are the first and second stage design matrices, respectively, expanded to full model form.

Recall the Bayesian Q-optimality criterion, presented for the normal linear model in section 3.3. The Bayesian Q-criterion is derived from the Bayes risk,

$$R^* = \int \{ (N/k) \int_R (\beta-\mathbf{b})' \, \mathbf{x}^{(m)} \mathbf{x}'^{(m)} \, (\beta-\mathbf{b}) \} \, p(\mathbf{y},\beta) \, d\mathbf{x} \, d\beta \, d\mathbf{y}$$

when there is uniform interest in prediction of the mean response over a region R with volume k. In the second stage,

$$p(\mathbf{y},\beta) = p(\beta|\sigma^2,\tau^2,\mathbf{y}_1, \mathbf{y}_2) \, p(\mathbf{y_1},\mathbf{y_2})$$

so, $$R^* = \text{tr}\left[\ \text{Var}(\beta|\sigma^2,\tau^2,\mathbf{y}_1,\mathbf{y}_2)\ \ (N/k)\int_R \mathbf{x}^{(m)}\,\mathbf{x}'^{(m)}\,d\mathbf{x}\ \right]$$

$$\text{tr}\left[(\ \mathbf{X_1}'\mathbf{X_1} + \mathbf{X_2}'\mathbf{X_2} + \mathbf{K}/\tau^2\ )^{-1}\ \ (N/k)\int_R \mathbf{x}^{(m)}\,\mathbf{x}'^{(m)}\,d\mathbf{x}\right].$$

Consider any candidate model M_i, identified by its parameters $\beta_i = \begin{bmatrix}\beta_{PRI}\\ \beta_{POT(i)}\end{bmatrix}$, where $\beta_{POT(i)}$ is a subset of q_i of the q potential parameters. Since the second stage posterior variance of β_i is $\mathbf{V}_{2(i)}$ (given in 5.3.4), the Bayes risk for model M_i, is

$$Q_i = \text{tr}\left[\mathbf{V}_{2(i)}\ (N/k)\int_R \mathbf{x}^{(m(i))}\,\mathbf{x}'^{(m(i))}\,d\mathbf{x}\right]$$

where $\mathbf{x}^{(m(i))}$ denotes a design point in model M_i space .

Paralleling the approach taken in section 5.2.2, Box and Meyer posterior probabilities are used to construct a weighted average Q criterion. This criterion will be used to select the second stage design. The objective is to choose the n_2 second stage design points so as to minimize Q_i for each model M_i having a high probability of being the 'best' model. This is done by choosing the second stage design so as to minimize

$$\underset{\mathbf{X}_2\ \ \boldsymbol{X}}{\text{Min}}\quad \sum_{Mi}\ Q_i\ p(M_i|\mathbf{y}_1)$$

where $\mathbf{X}_2$ and $\boldsymbol{X}$ are still in full model space. Once again, it is important to point out that the second stage design, just like the first, contains any design (control) variables included in the full model.

5.5 Bayesian Q-Q Optimality Procedure

(A Linear Combination of Q-optimality Criteria)

The Bayesian Q-Q optimality procedure is almost identical to that of the D-Q procedure, developed in the previous section. The key difference is that the first stage design of the Q-Q

procedure focuses on prediction more than estimation. The first stage design is chosen according to the method described in section 5.1.2. The first stage analysis and choice of second stage design are performed exactly as discusse d in section 5.4.

5.6 Bayesian D-DQ Optimality Procedure

(A Linear Combination of DQ-optimality Criteria)

A Bayesian D-DQ optimality procedure is developed in this section. The first stage design is chosen to be D-optimal, according to the procedure discussed in section 5.1.1. The second stage design simultaneously addresses the objectives of parameter *estimation* (i.e., D-optimality) and the *prediction* of future observations (similar to Q-optimality). This procedure is similar to those presented in sections 5.3 - 5.5, in that Box and Meyer posterior probabilities are used to weight candidate models, producing a single DQ criterion which is a weighted average of 2^q DQ criteria.

In the first stage it is assumed that $\beta|\sigma^2,\tau^2 \quad N(\mathbf{0}, \sigma^2\tau^2\mathbf{K^{-1}})$ and $\mathbf{y}_1|\beta,\sigma^2 \quad N(\mathbf{X}_1\beta, \sigma^2\mathbf{I})$, resulting in the first stage posterior which is also normal with posterior covariance matrix $\mathbf{V_1} = \sigma^2(\mathbf{X_1}'\mathbf{X_1} + \mathbf{K}/\tau^2)^{-1}$. Using this first stage posterior as the prior for the second stage and assuming $\mathbf{y}_2|\beta,\sigma^2 \quad N(\mathbf{X}_2\beta,\sigma^2\mathbf{I})$, the second stage posterior distribution of β is also normal, with posterior covariance matrix $\mathbf{V}_2 = \sigma^2(\mathbf{X_1}'\mathbf{X_1} + \mathbf{X_2}'\mathbf{X_2} + \mathbf{K}/\tau^2)^{-1}$. Note that β contains all p+q parameters of the full model and therefore $\mathbf{X}_1$ and $\mathbf{X}_2$ are the first and second stage design matrices, respectively, expanded to full model form.

Both estimation and prediction can be addressed in the second stage through the implementation of a utility function suggested by Verdinelli (1992). Recall from section 3.2 that this utility function is a linear combination of the expected gain in Shannon information for both $p(\beta|\mathbf{y},\delta)$ and a future observation at a condition $\mathbf{x}$. As pointed out by Chaloner and Verdinelli

(1995), the design which maximizes Verdinelli's utility function for the normal linear model also minimizes the risk function

$$R_v(\delta,\mathbf{x}) = [\mathbf{x}'^{(m)}\, \mathrm{Var}(\beta|\mathbf{y})\, \mathbf{x}^{(m)} + \sigma^2] \cdot \det(\mathrm{Var}(\beta|\mathbf{y}))\ .$$

The quantity in brackets is the *predictive variance*, i.e., the variance associated with predicting a single observation at a location $\mathbf{x}^{(m)}$ in model space. Integrating $R_v(\delta,\mathbf{x})$ over a region of interest, *R*, and dividing through by the volume of *R* produces a design criterion which is the product of the *average predictive variance* (similar to the Bayes Q criterion) and the Bayes D-optimality criterion. In other words,

$$\begin{aligned}(1/k)\int_R R_v(\delta,\mathbf{x}) &= (1/k)\int_R (\mathbf{x}'^{(m)}\, \mathrm{Var}(\beta|\mathbf{y})\, \mathbf{x}^{(m)} + \sigma^2) \cdot \det(\mathrm{Var}(\beta|\mathbf{y}))\, d\mathbf{x} \\ &= \det(\mathrm{Var}(\beta|\mathbf{y}))\, [(1/k)\int_R (\mathbf{x}'^{(m)}\, \mathrm{Var}(\beta|\mathbf{y})\, \mathbf{x}^{(m)} + \sigma^2) d\mathbf{x}\,]\end{aligned}$$

when there is uniform interest over R which has volume $k = \int_R d\mathbf{x}$.

Under the assumption that the full model is correct, the second stage posterior variance of β is $\mathbf{V}_2 = \sigma^2(\,\mathbf{X_1}'\mathbf{X_1} + \mathbf{X_2}'\mathbf{X_2} + \mathbf{K}/\tau^2\,)^{-1}$, and thus the second stage design can be chosen by selecting the n_2 rows of X_2 in order to minimize

$$\det(\mathbf{V}_2)\, [(1/k)\int_R (\mathbf{x}'^{(m)}\, \mathbf{V}_2\, \mathbf{x}^{(m)} + \sigma^2) d\mathbf{x}\,].$$

For any given subset model M_i, the second stage design would be chosen to minimize

$$(DQ)_i = \det(\mathbf{V}_{2(i)})\, [(1/k)\int_R (\mathbf{x}'^{(m(i))}\, \mathbf{V}_{2(i)}\, \mathbf{x}^{(m(i))} + \sigma^2) d\mathbf{x}\,],$$

where the subscript (i) denotes the appropriate model space.

As with the criteria presented in sections 5.3-5.5, Box and Meyer posterior model probabilities are used to construct a weighted average DQ criterion. This criterion is used to select the second stage design. The objective is to choose the n_2 second stage design points so as to minimize $(DQ)_i$ for each model having a high probablity of being the 'best' model.

This is done by choosing the second stage design so as to minimize

$$\sum_{Mi} (DQ)_i \, p(M_i|\mathbf{y}_1) ,$$

where $p(M_i|\mathbf{y}_1)$ is the posterior probability that model M_i is the 'best' model.

5.7 Bayesian D-Q(w) Optimality Procedure

The utility of this procedure is in choosing a second stage design which minimizes prediction variance, especially in the portion of the design region which produces high response values. Like the D-Q procedure developed in section 5.4, a first stage Bayesian D-optimal design is followed by a design which minimizes a weighted average of Q criteria (weighting over the set of candidate models with Box and Meyer posterior probabilities). Additionally, however, each Q criterion results from integrating prediction variance over a weighted region of interest, *R*. The density, $w(\mathbf{x})$, used to weight *R*, is chosen so that high weights are assigned to the portion of *R* expected to produce high response values. It is believed that $w(\mathbf{x})$ should be a function of $\hat{y}(\mathbf{x})$, the estimated response after the first stage. One such function is

$$w(\mathbf{x}) = \frac{\mathbf{x}'\mathbf{b}}{\int \mathbf{x}'\mathbf{b} \, d\mathbf{x}}$$

where $\mathbf{b}$ is the Bayes estimator of β after observing the first stage data. Thus the second stage design is found by minimizing

$$\sum_{Mi} Q(w)_i \, p(M_i|\mathbf{y}_1) , \qquad (5.7.1)$$

where

$$Q(w)_i = \text{tr} \left[\mathbf{V}_{2(i)} \, (N/k) \int_R \mathbf{x}^{(m(i))} \, \mathbf{x}'^{(m(i))} \, w(\mathbf{x}^{(m(i))}) \, d\mathbf{x} \right] , \qquad (5.7.2)$$

and $p(M_i|\mathbf{y}_1)$ is the posterior probability that model M_i is the 'best' model. Recall that $\mathbf{V}_{2(i)}$, given in (5.3.4), is the posterior variance matrix of the parameters in M_i.

In practical applications of the D-Q(w) procedure, the design space is represented by a finite grid of points. An approximation of (5.7.2) is obtained by calculating the average prediction variance over the grid of points, weighting each point by the predicted response at that

location. When the list of candidate models is extensive, the computer time required to find the second stage design which minimizes (5.7.1) will be excessive. It is therefore recommended that the first stage data be analyzed in order to reduce the candidate set to contain no more than four or five models. This initial model screening must be performed cautiously, however, to avoid underspecification of the true model. For example, if p-values are used to identify any of the q potential terms as being inactive, one rule of thumb is to eliminate only those terms with a p-value greater than .20.

Three examples will now be given in order to illustrate the usefulness of the D-Q(w) procedure. For these examples, it is assumed that the full model under consideration by the experimenter is defined by the regressors

$$\boldsymbol{x}^{(f)} = \{\ 1, x_1, x_2, x_1x_2, x_1^2, x_2^2\ \}$$

including p=4 primary terms, $\{1, x_1, x_2, x_1x_2\}$ and q=2 potential terms, $\{x_1^2, x_2^2\}$. The *true model* (unknown to the experimenter), is the full quadratic model, defined by $\boldsymbol{x}^{(f)}$. Response data from the first stage experiment is simulated from the true model,

$$y_i = 53.0 + 10.0\, x_{1i} + 8.0\, x_{2i} + 2.5\, x_{1i}x_{2i} - 2.5\, x_{1i}^2 - 6.0\, x_{2i}^2 + \varepsilon_i \qquad (5.7.3)$$

under the assumption that all ε_i are i.i.d. N(0,1). A contour plot of the true response surface is shown in figure 5.7.1. Notice that the response increases as both x_1 and x_2 increase, resulting in maximum response values in the upper right quadrant of the design region.

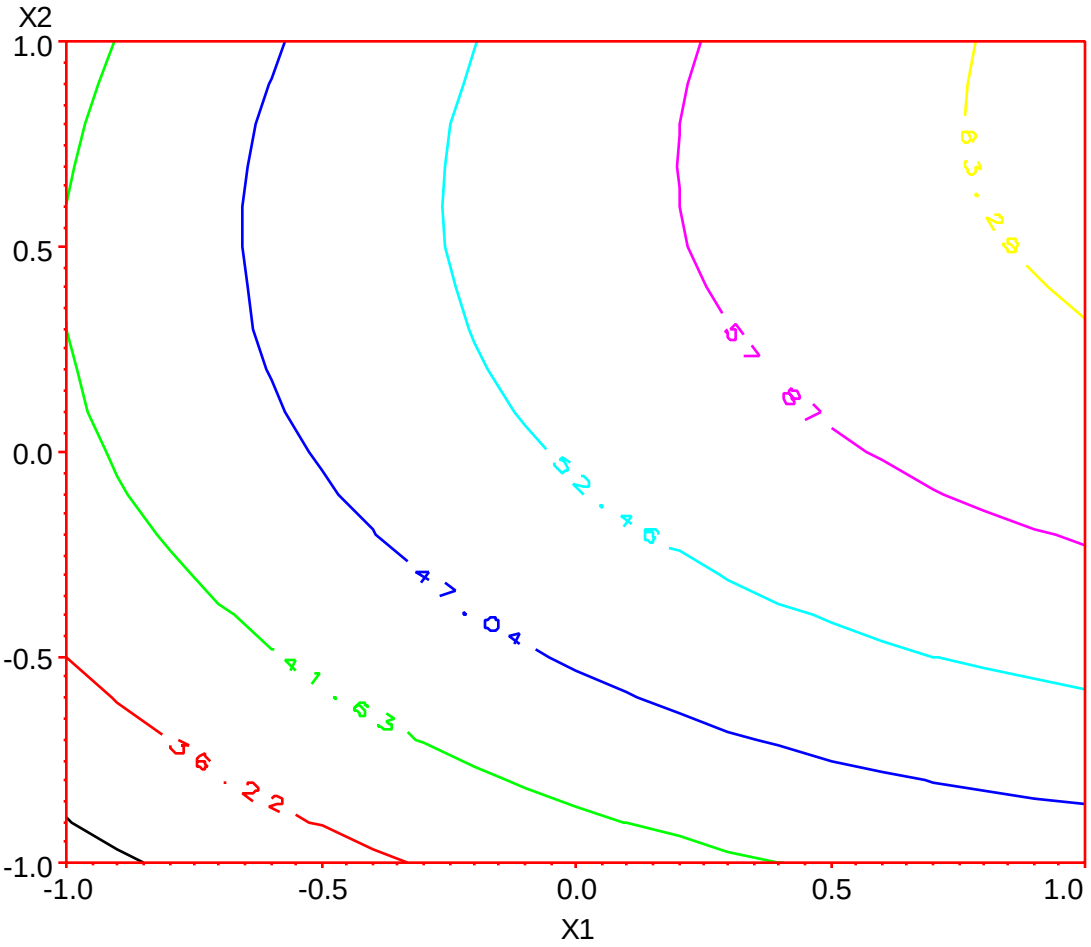

Figure 5.7.1 Contour Plot of the True Response Surface Model

5.7.1 An Example of the D-Q(w) Procedure Under Mild Model Uncertainty

Based on the full model defined by $\boldsymbol{x}^{(f)}$, Bayesian D-optimality is used to select the first stage design consisting of seven design points. In selecting this design, the prior standard deviation assumed for any potential term is $\tau=5$ (recommendations for sample size allocation and the size of τ will be given in section 6.1). The resulting design is shown in figure 5.7.1(a). Notice that this design contains the four corner points which are used in estimating β_1, β_2 and β_{12}. Two face center points and one overall center point are also included. The overall center point provides some information about both quadratic terms. The point at the center of the x_1 axis,

however, provides independent information about β_{11}. Similarly, the point at the center of the x_2 axis provides independent information about β_{22}, thus allowing these two quadratic terms to be estimated separately.

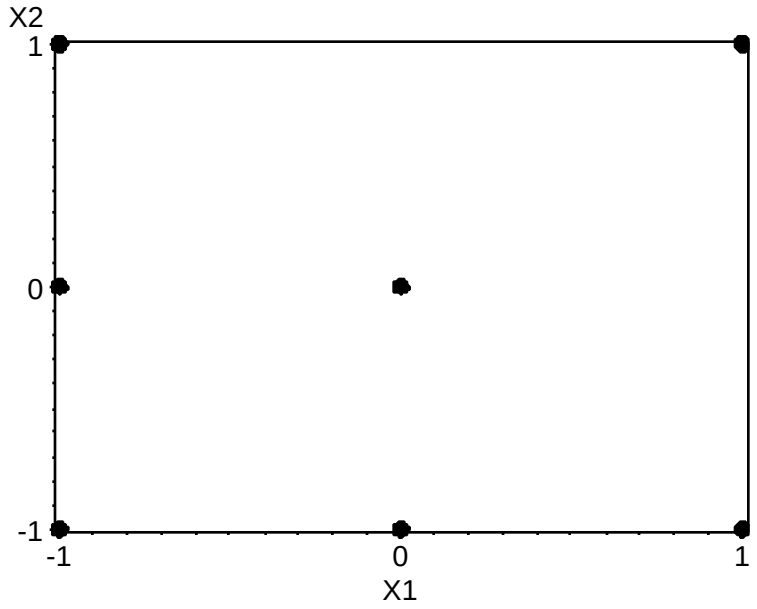

Figure 5.7.1(a) First stage Bayes D-optimal design

Data from the first stage experiment serves a dual role. As with many of the two stage procedures discussed previously in this chapter, the first stage data is used to calculate posterior probabilities to weight the importance of each candidate model. For this example, the resulting Box and Meyer posterior probabilities are included in table 5.7.1. From these probabilities it is easy to see that in selecting a second stage design, the majority of the weight will be placed on the full quadratic model. It should be noted that there is still a little uncertainty about the importance of the x_1^2 term, since the candidate model which excludes x_1^2 has been given the second largest probability of approximately 0.15.

Table 5.7.1 Box and Meyer posterior probabilities for candidate models

Posterior Probability	**Terms Included in the Candidate Model**
0.002	Intercept, x_1, x_2, x_1x_2
0.001	Intercept, x_1, x_2, x_1x_2, x_1^2
0.149	Intercept, x_1, x_2, x_1x_2, x_2^2
0.848	Intercept, x_1, x_2, x_1x_2, x_1^2, x_2^2

The first stage data must also be used to produce parameter estimates for each of the candidate models. Each of the resulting models is used to predict the response at each point of a finite grid placed over the design region. Those areas of the design region with the highest predicted values are given more weight by the Q(w) procedure used to choose the second stage design points.

The second stage design is then chosen to minimize the weighted average prediction variance for the candidate model(s) of most importance. In this example, the second stage design was restricted to be of size n_2=6 design points. The resulting design is shown in figure 5.7.1(b), where the design points have been plotted over contours of the true response surface. The majority of the second stage design points are placed around the upper right hand quadrant of the design space, i.e., the quadrant containing the highest response values. Three of the points are face center points, two of them supporting estimation of β_{22}, the remaining point supporting estimation of β_{11} (recall that there was some question as to the significance of β_{11}). As will be shown in section 5.7.2, certainty about the importance of the x_1^2 term would have resulted in the (0,1) point being replicated, instead of placing a point at the (+1,-1) location.

The combined two stage D-Q(w) design is shown in figure 5.7.1(c), plotted over contours of the resulting prediction variance over the design region. The prediction variance contours show that the mean response can be predicted more precisely in the upper and lower right corners of the design region. Since the upper right corner is the area of highest response, this two stage design has done a good job of minimizing the weighted prediction variance under model uncertainty.

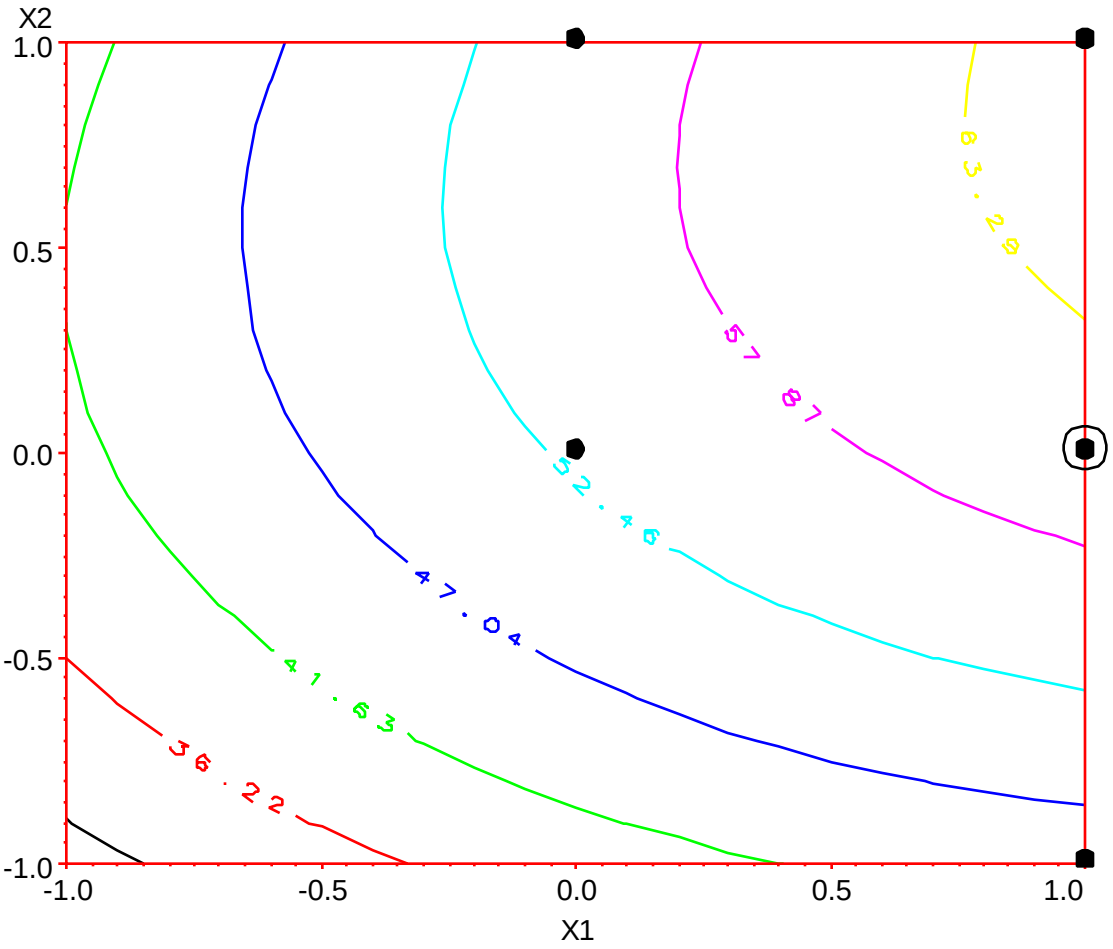

Figure 5.7.1(b) Second stage design plotted over contours of true response surface

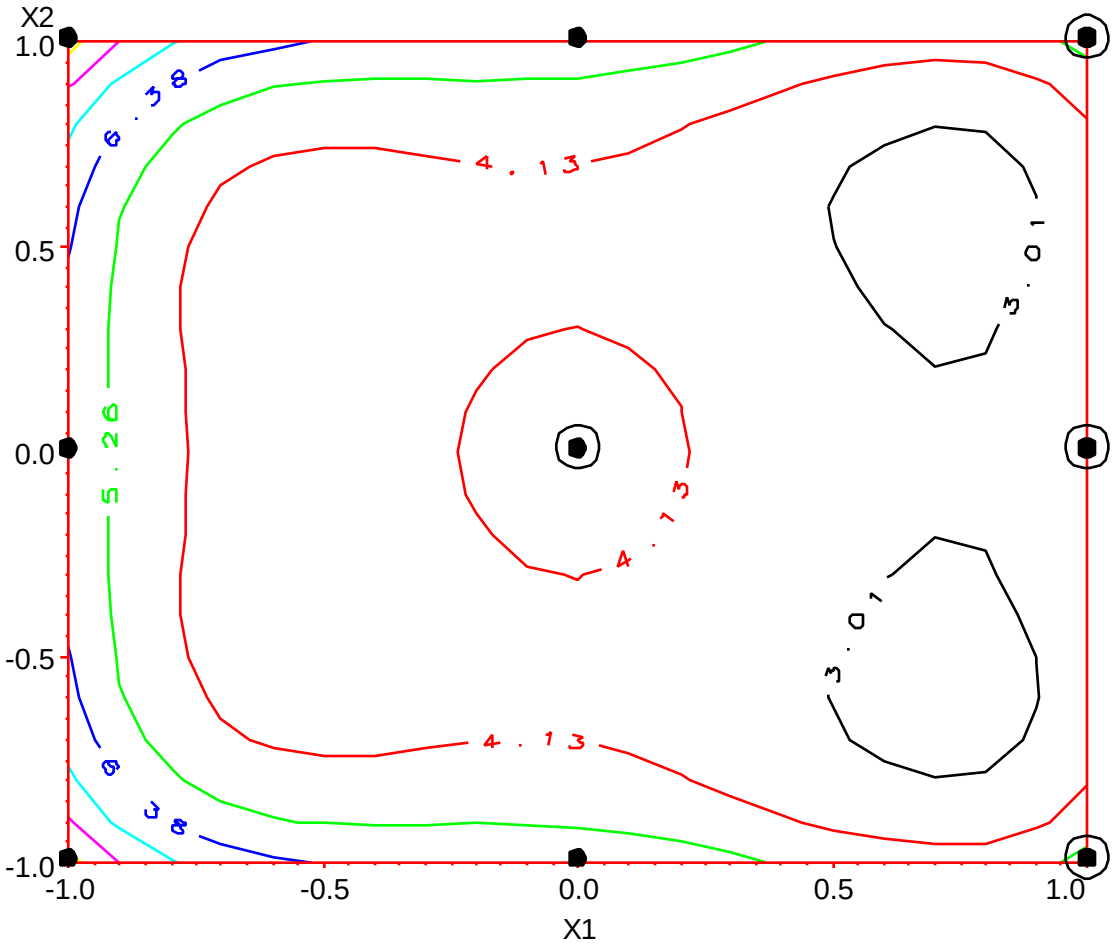

Figure 5.7.1(c) Two stage D-Q(w) design plotted over contours of its prediction variance throughout the design region

5.7.2 An Example of the D-Q(w) Procedure Under No Model Uncertainty

Once again assume that the first stage design is the seven run Bayes D-optimal design which is represented graphically in figure 5.7.1(a). Upon completion of the first stage experiment, the seven data points (simulated from the true response model (5.7.3)) are analyzed using a classical

approach. The results of the Student's *t* tests for the significance of parameter estimates are included in table 5.7.2(a). Using a conservative p-value of .20, it is concluded that the most appropriate model is the full quadratic model. The second stage design will therefore be selected with no model uncertainty.

Table 5.7.2(a) Analysis of data from the first stage Bayes D-optimal design

Variable	**Parameter Estimate**	**Standard Error**	**P-value of *t* test**
Intercept	53.47	0.402	0.0048
X_1	10.38	0.227	0.0139
X_2	7.46	0.227	0.0194
X_1X_2	3.10	0.235	0.0482
X_1X_1	-2.47	0.446	0.1137
X_2X_2	-7.09	0.446	0.0400

Just as in the previous example, the second stage is restricted to a size of n_2=6 design points. The second stage design which results when the model is known to be the full quadratic model is shown in figure 5.7.2(a). It is plotted over contours of the true response surface. Notice that all six design points are located at the corners of the upper right quadrant of the design space...the area of highest response. The lack of model uncertainty has allowed the Q(w) procedure to put all of the second stage resources in the area of high response.

The combined two stage design is shown in figure 5.7.2(b), plotted over contours of the resulting prediction variance over the design region. It is a Face Center Cube (FCC) design with the four corners of the upper right quadrant replicated. The prediction variance contours show that the mean response can be predicted more precisely in the upper right corner of the design

region. Since the upper right corner is the area of highest response, this two stage design has done an excellent job of meeting the objective of the experiment.

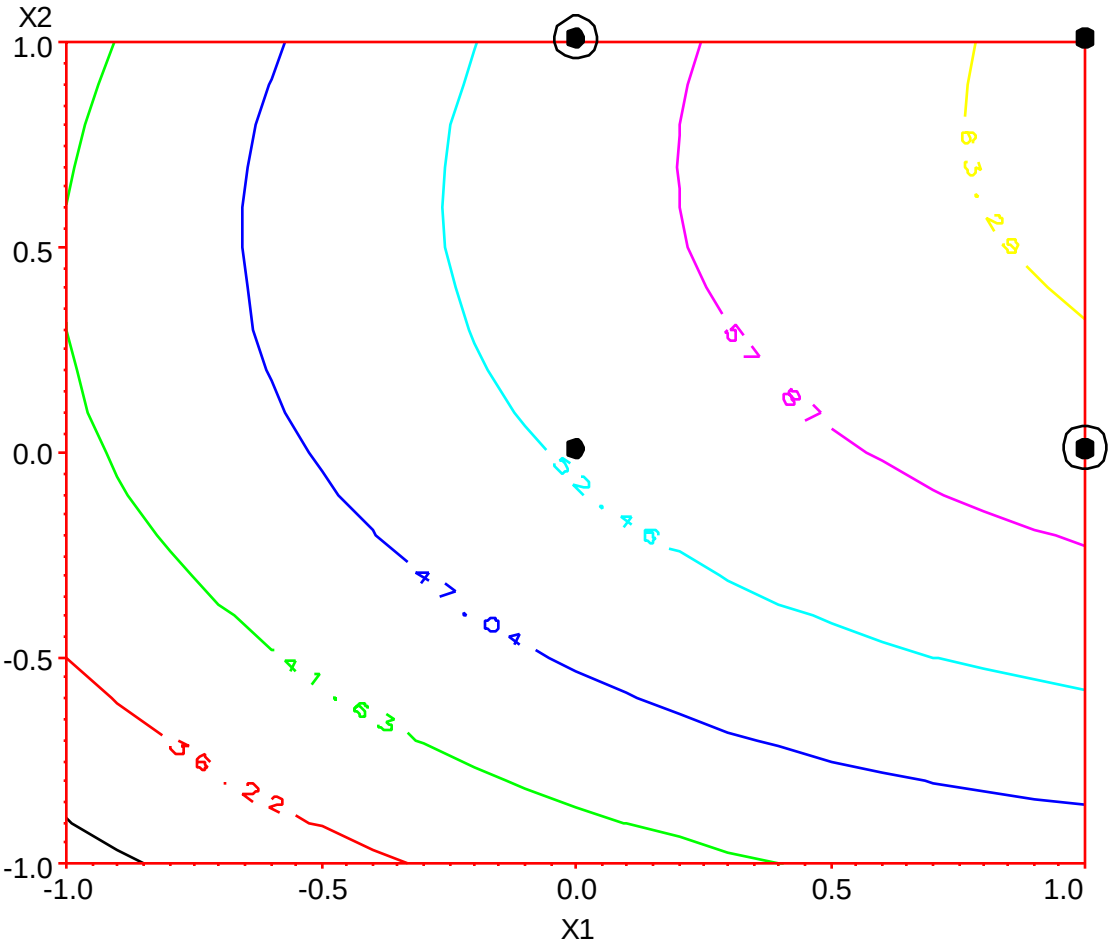

Figure 5.7.2(a) Second stage design under no model uncertainty
(plotted over contours of the true response surface)

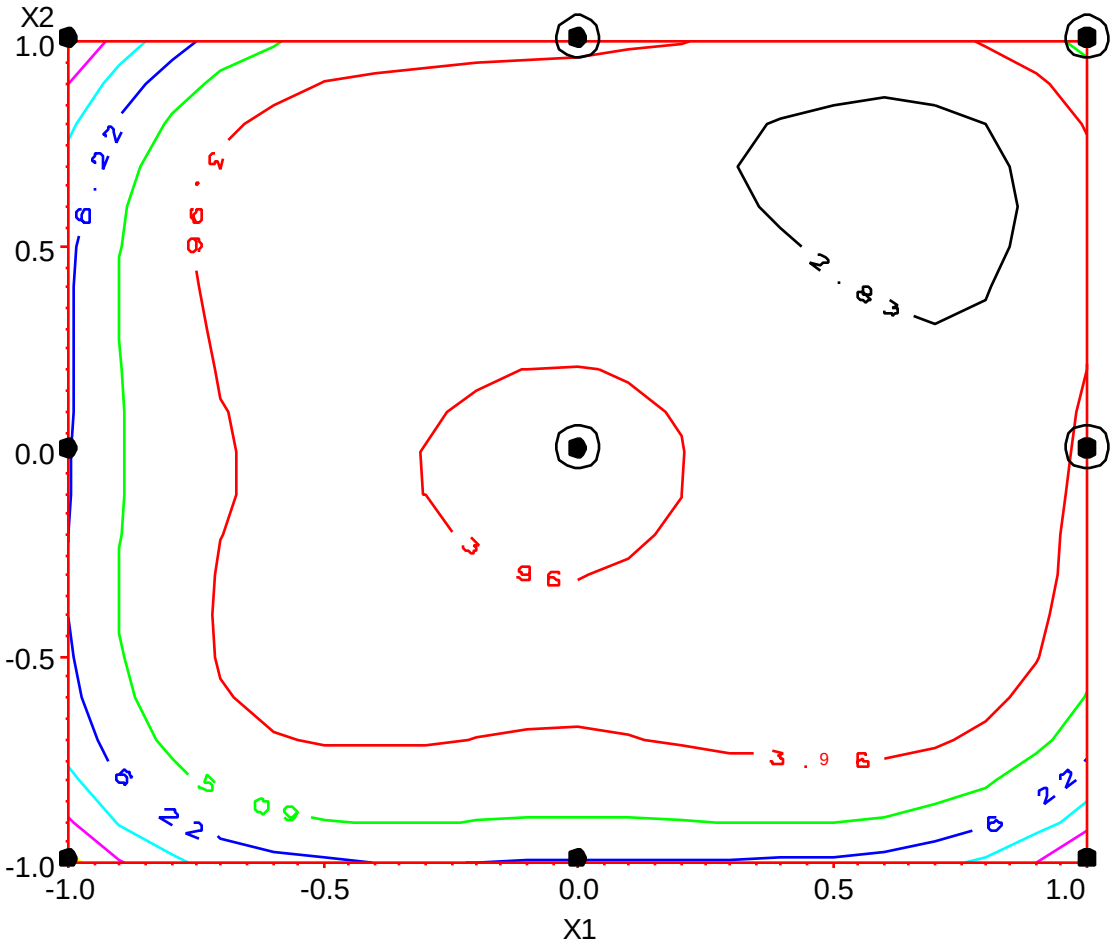

Figure 5.7.2(b) Two stage D-Q(w) design (under no model uncertainty) plotted over contours of its prediction variance throughout the design region

5.7.3 An Example of the Q(w) Criterion as an Augmentation Procedure

The purpose of this example is to show how the Q(w) procedure might be used to augment a more classical design in order to reduce prediction variance in areas of high response. Assume that the first stage design is an FCC design with two center runs. This design is shown in figure 5.7.3(a).

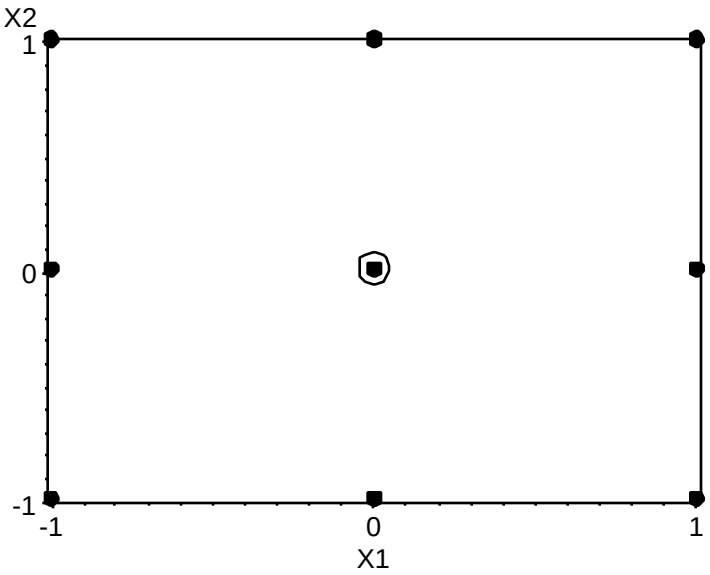

Figure 5.7.3(a) Face Center Cube design

Upon completion of the first stage experiment, the ten data points (simulated from the true response model (5.7.3)) are analyzed using a classical approach. The results of the Student's *t* tests for the significance of parameter estimates are included in table 5.7.3(a). Using a conservative p-value of .20, it is concluded that the most appropriate model is the full quadratic model.

Table 5.7.3(a) Analysis of data from the FCC design

Variable	**Parameter Estimate**	**Standard Error**	**P-value of T-test**
Intercept	53.03	0.696	0.0001
X_1	10.00	0.475	0.0001
X_2	8.15	0.475	0.0001
X_1X_2	2.36	0.582	0.0155
X_1X_1	-2.64	0.762	0.0256
X_2X_2	-6.03	0.762	0.0014

A contour plot of the fitted response model is given in figure 5.7.3(b). Additionally, a contour plot of the prediction variance of the FCC design is given in figure 5.7.3(c). Together, these two plots show that the prediction variance needs to be reduced in the upper right quadrant of the design space (i.e., in the area of high response).

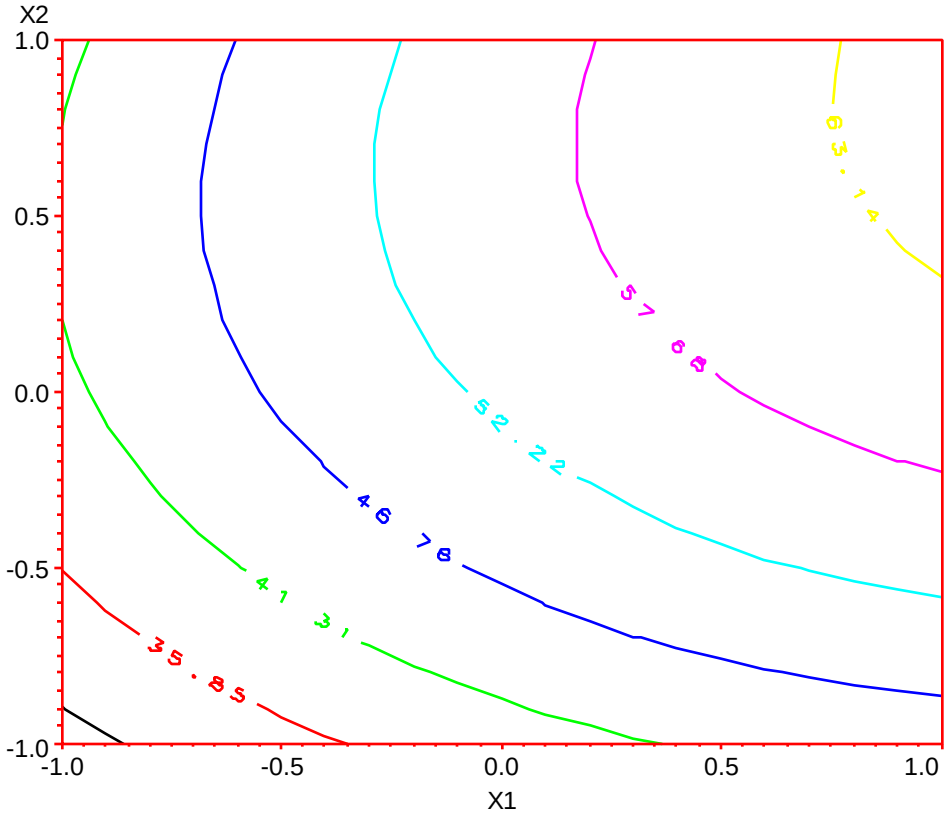

Figure 5.7.3(b) Contour plot of the fitted response surface

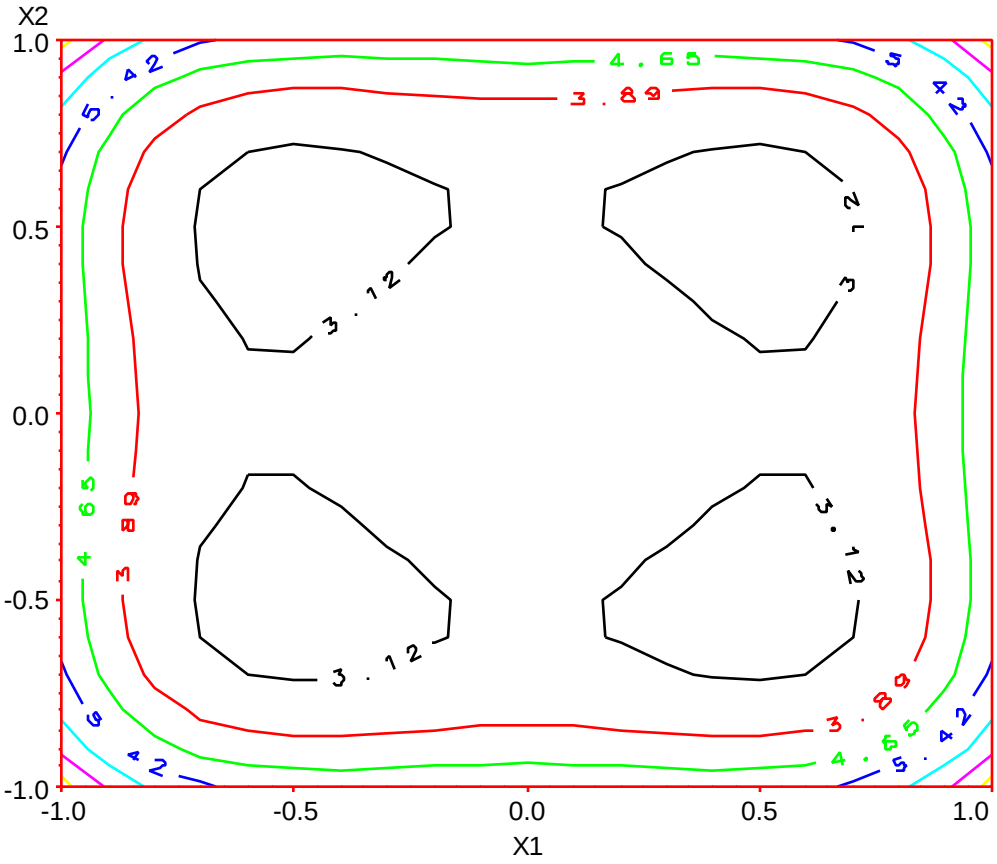

Figure 5.7.3(c) Contour plot of the prediction variance of the FCC design

A second stage design will be used to improve the prediction variance in the area of high response. The Q(w) procedure is used to select an additional three design points with which to augment the FCC design. These points, plotted in figure 5.7.3(d), are selected under the assumption that the model is known to be the full quadratic model. Notice that the three points are located at the upper right corner of the design space, since this is the area requiring more resources in order to improve prediction precision.

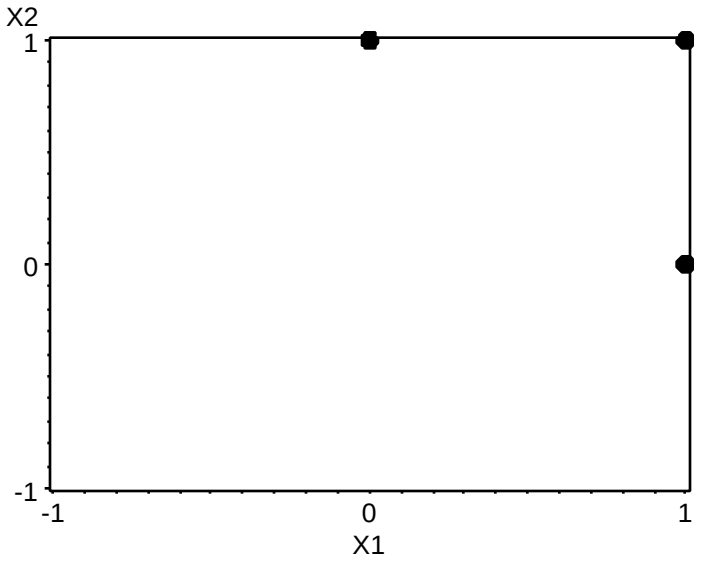

Figure 5.7.3(d) Second stage design points chosen to augment the FCC design

The augmented FCC design is shown in figure 5.7.3(e) with a contour plot of its prediction variance over the design region. There has been a noticeable improvement in the ability to predict the mean response in the upper right quadrant of the design space (i.e., the quadrant of high response).

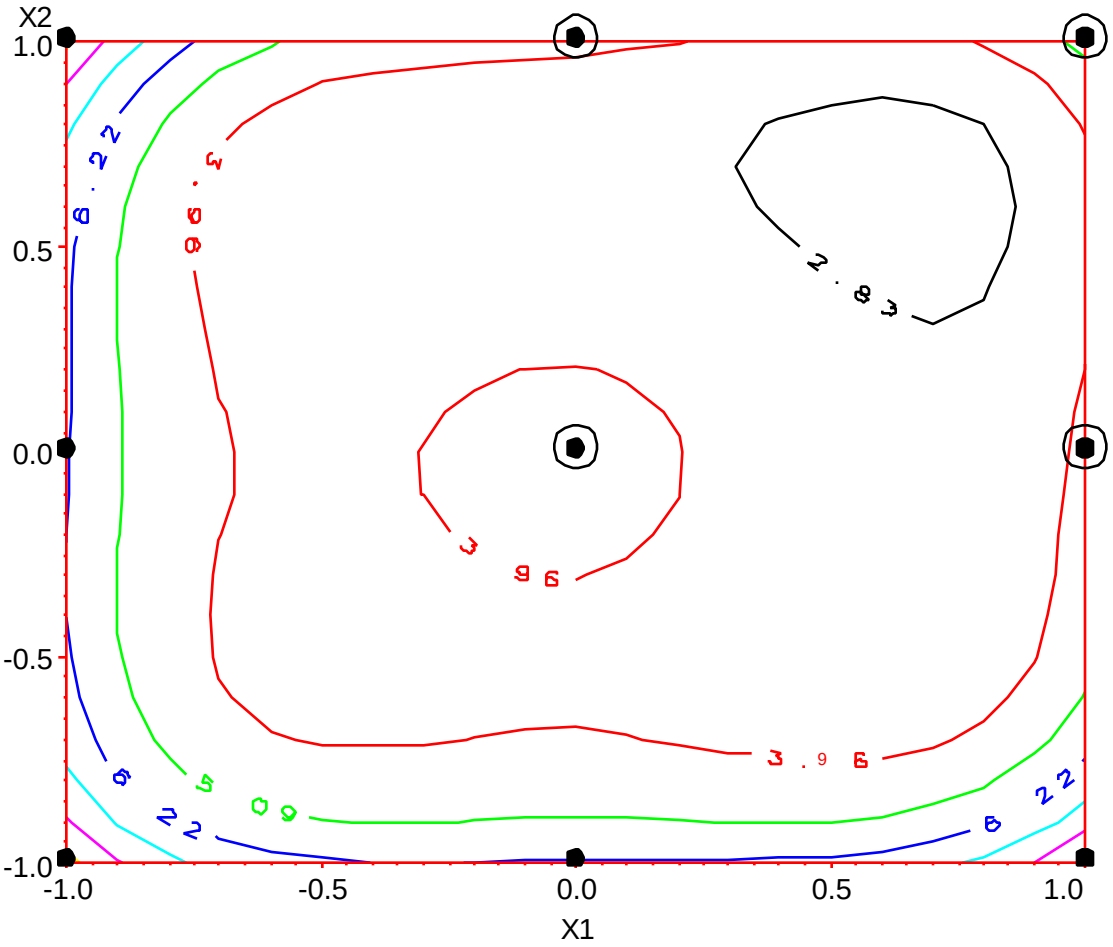

Figure 5.7.3(e) Augmented FCC and contours of its prediction variance

5.7.4 A Class of D-Q(w) Optimality Procedures

Thus far in section 5.7, the D-Q(w) procedure has been utilized in a specific application - namely, one in which the experimenter has a high priority on predicting high response values. There are other situations in which one may not have uniform interest in prediction over the experimental region. For example, there may be a high priority on predicting response values close to a given target. As another example, there may be a high priority on predicting low response values. A more general D-Q(w) procedure could be adapted for use in either of these two cases, as well as that previously demonstrated.

The second stage of the D-Q(w) procedure is the selection of a design which minimizes the Q(w) optimality criterion, given in general form by

$$Q(w) = \sum_{Mi} Q(w)_i \, p(M_i|\mathbf{y}_1) \qquad (5.7.4a)$$

where,

$$Q(w)_i = \text{tr}\left[\mathbf{V}_{2(i)} (N/k) \int_R \mathbf{x}^{(m(i))} \mathbf{x}'^{(m(i))} w(\mathbf{x}^{(m(i))}) \, d\mathbf{x}\right] . \qquad (5.7.4b)$$

In general, w(**x**) can be any density function on **x** which incorporates information from the first stage data. For example, if high priority is placed on predicting response values close to a target of ***t***, one possible choice of the density function is given by

$$w(\mathbf{x}) = \frac{\left(|\mathbf{x}'\mathbf{b} - t|\right)^{-1}}{\int \left(|\mathbf{x}'\mathbf{b} - t|\right)^{-1} d\mathbf{x}} .$$

On the other hand, if high priority is placed on predicting low response values, the density function may be defined as

$$w(\mathbf{x}) = \frac{\left(\mathbf{x}'\mathbf{b}\right)^{-1}}{\int \left(\mathbf{x}'\mathbf{b}\right)^{-1} d\mathbf{x}} .$$

Utilizing this flexibility in the definition of w(**x**) results in an entire class of D-Q(w) optimality procedures, all of which are special cases of the general form given in (5.7.4a) and (5.7.4b).

5.8 Examples of Bayesian Two Stage Designs

Given in this section are example designs produced by the Bayesian D-D (section 5.3) and Bayesian Q-Q (section 5.5) optimality procedures. For both examples, it is assumed that the full model under consideration by the experimenter is defined by the regressors

$$\boldsymbol{x}^{(f)} = \{\, 1, x_1, x_2, x_1x_2, x_3, x_1x_3, x_2x_3, x_1^2, x_2^2 \,\}$$

including p=4 primary terms, **{**1, x_1, x_2, x_1x_2 **}** and q=5 potential terms, **{**x_3, x_1x_3, x_2x_3, x_1^2, x_2^2 **}**. The *true model* (unknown to the experimenter), contains the four primary terms and only three of the five potential terms. The true model is defined by the set of regressors

$$\boldsymbol{x}^* = \{\ 1, x_1, x_2, x_1x_2, x_1x_3, x_2x_3, x_1^2\},$$

with coefficients chosen so that x_1x_3, x_2x_3 are only marginally significant. For the examples in this section, response data from the first stage experiment is simulated from the true model. Twelve runs are allocated to each stage of the experiment and the prior standard deviation assumed for any potential term is τ=5 (recommendations for sample size allocation and the size of τ will be given in section 6.1).

5.8.1 A Bayesian D-D Optimal Design

The Bayesian D-optimality criterion presented in section 5.1.1 is used to select the first stage design, based on the full model defined by $\boldsymbol{x}^{(f)}$. The resulting design is shown in figure 5.8.1(a). It is not surprising that the eight corner points make up the majority of this design, since it is well known that the D-optimality criterion pushes points to extremes, except when the model dictates otherwise. Notice that the three edge points and the face center point are also included in the design, providing three levels of both x_1 and x_2 for the estimation of β_{11} and β_{22}. Only two levels of x_3 are necessary, since there is no interest in estimating β_{33} (recall that x_3^2 was not included in the full model).

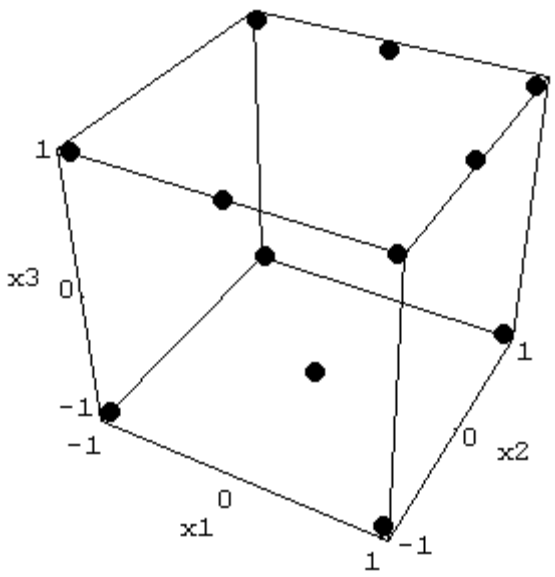

Figure 5.8.1(a) First stage of Bayes D-D optimal design

Recall from section 5.8.2 that the response data resulting from the first stage experiment is used in the calculation of probabilities/weights associated with each of the candidate models. For this example, 12 data points were simulated from the true model defined by $\boldsymbol{x}^* = \{ 1, x_1, x_2, x_1x_2, x_1x_3, x_2x_3, x_1^2\}$, under the assumption that σ=1.0. The resulting set of posterior probabilities/weights was such that the highest weight was placed on the model defined by $\boldsymbol{x}^*$ (i.e., correctly choosing the true model). The second stage design, shown in figure 5.3.1(b), was then chosen to produce a two stage design which is D-optimal with high priority given to the model defined by $\boldsymbol{x}^*$.

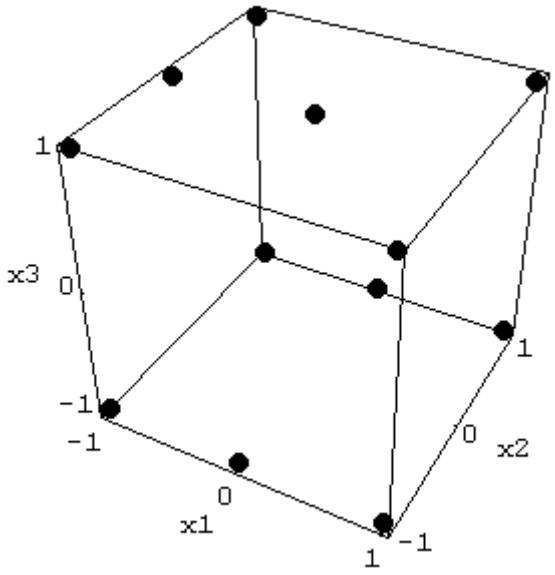

Figure 5.8.1(b) Second stage of Bayes D-D optimal design

The combined two stage design, shown in figure 5.8.1(c) , contains two replicates at each of the eight corners, six edge points supporting estimation of β_{11} and β_{22} and a center point in each of the x_1,x_2 planes. Notice that little emphasis is placed on x_3 since it is only involved in two marginal interactions in the true model. In fact, if this design were collapsed onto the x_1,x_2 plane, the resulting design would be a fully replicated FCC design in x1 and x2, except for replication of the x_2 edge points (since x_2^2 is absent from the true model).

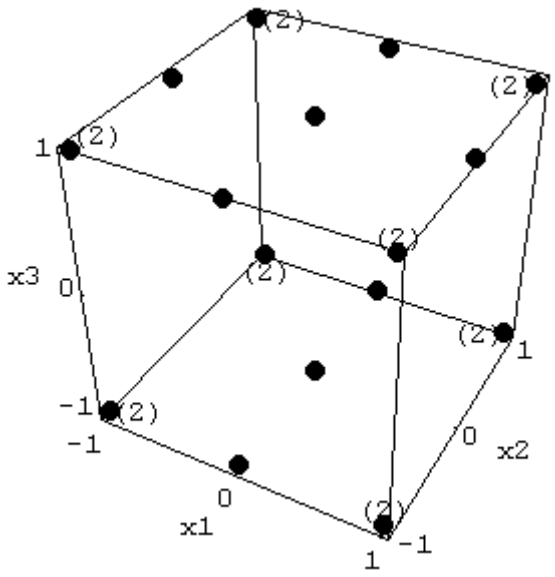

Figure 5.8.1(c) The combined two stage Bayes D-optimal design

5.8.2 A Bayesian Q-Q Optimal Design

The first stage design is chosen according to the Bayesian Q-optimality criterion presented in section 5.8.2. The resulting design, shown in figure 5.8.2(a), contains five levels of x_1 and three levels of x_2, to support estimation of β_{11} and β_{22}. Only two levels of x_3 are included, since there is no interest in estimating β_{33}.

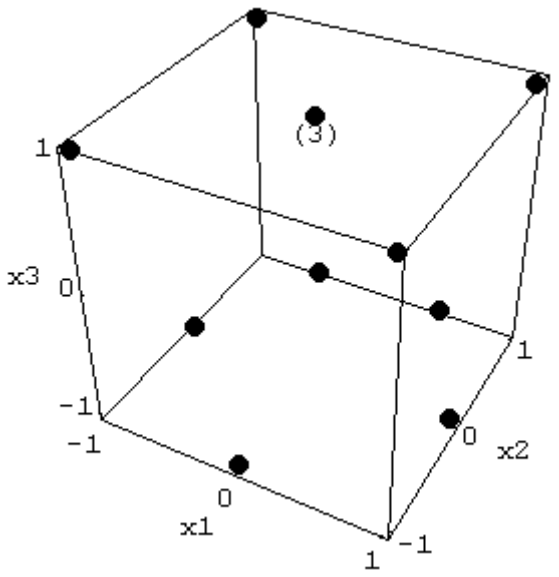

Figure 5.8.2(a) First stage of Bayes Q-Q optimal design

For this example, like that of section 5.8.1, 12 data points were simulated from the true model defined by $\boldsymbol{x}^* = \{ 1, x_1, x_2, x_1x_2, x_1x_3, x_2x_3, x_1^2\}$, under the assumption that $\sigma=1.0$. The resulting set of posterior probabilities/weights was such that the highest weight was placed on the model defined by $\boldsymbol{x}^*$ (i.e., correctly choosing the true model). The second stage design was then selected to produce a two stage design which is Q-optimal with high priority given to the model defined by $\boldsymbol{x}^*$. The resulting second stage design is shown in figure 5.8.2(b). Notice that this design contains numerous edge points which contribute to the estimation of β_{11}.

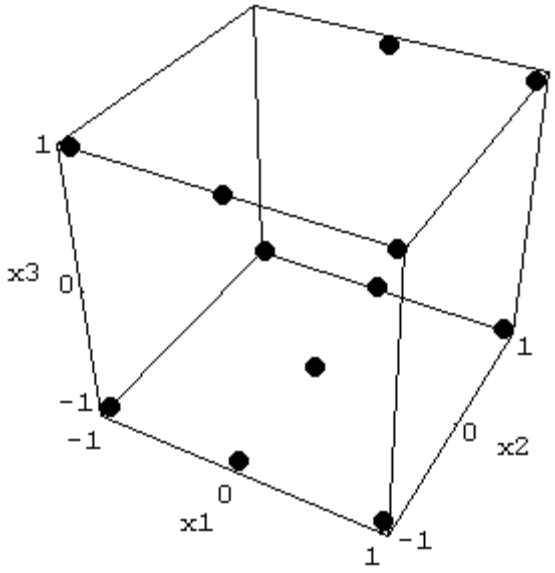

Figure 5.8.2(b) Second stage of Bayes Q-Q optimal design

The combined two stage design, shown in figure 5.8.2(c), clearly reflects the importance of x_1 and x_2 in the true model.

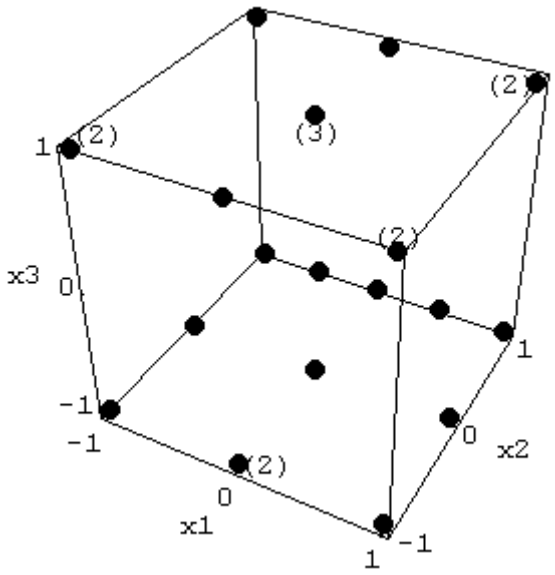

Figure 5.8.2(c) The combined two stage Bayes Q-optimal design

Chapter 6

Evaluation of the Bayesian Two Stage Procedures

6.1 Preliminary Evaluation for Choice of Sample Size Allocation and τ

Before making recommendations for sample size allocation (n_1/N, n_2/N) and choice of τ, it was necessary to determine how these parameters affect the performance of the Bayesian two stage procedures. A simulation approach was taken to evaluate the efficiency of designs which result when various values of (n_1/N, n_2/N) and τ were used in the two stage procedures. Based on this evaluation, it is recommended that 50% of the total design size N be used in each of the two stages, setting $n_1 = n_2 \geq (p+q)$, and if possible $n_1 = n_2 = (p+q+3)$. (Recall that p+q is the total number of terms under consideration for the full model.) Additionally, it is recommended that τ=5 in the first stage (and the second stage when appropriate). These values of n_1, n_2 and τ are recommended for their ability to produce designs which are most robust to model (regressor) misspecification.

6.2 Method of Evaluating Bayesian Two Stage Designs vs. Single Stage Designs

The performance of each of the two stage procedures presented in sections 5.2-5.6 was to be evaluated relative to more traditional single stage designs. As discussed in section 4.3, the second stage design is actually a random variable dependent on first stage data. For this reason, a simulation approach is taken to evaluate the average performance of each two stage procedure,

where averaging occurs over N_s first stage data sets. (In this case N_s=50.) The *full model* under consideration throughout the evaluation is a linear model defined by the full set of regressors

$$\boldsymbol{x}^{(f)} = \{\ 1, x_1, x_2, x_1x_2, x_3, x_1x_3, x_2x_3, x_1^2, x_2^2\ \}$$

including p=4 primary terms, $\{1, x_1, x_2, x_1x_2\ \}$ and q=5 potential terms, $\{x_3, x_1x_3, x_2x_3, x_1^2, x_2^2\ \}$. For simulation purposes, four separate cases are considered, each defined by the *true model* from which response data is simulated. For each of the following models it is assumed that $\varepsilon_i \sim N(0,1)$.

CASE 1: Primary Terms Only

$$y_i = 70.0 + 11.5x_{1(i)} + 7.3x_{2(i)} + 8.0x_{1(i)}x_{2(i)} + \varepsilon_i$$

CASE 2: Primary Terms + interactions

$$y_i = 70.0 + 11.5x_{1(i)} - 7.3x_{2(i)} + 8.0x_{1(i)}x_{2(i)} + 1.1x_{1(i)}x_{3(i)} - 1.3x_{2(i)}x_{3(i)} + \varepsilon_i$$

CASE 3: Primary Terms + interactions and a quadratic

$$y_i = 70.0 - 7.3x_{1(i)} + 10.0x_{2(i)} + 8.0x_{1(i)}x_{2(i)} + 1.1x_{1(i)}x_{3(i)} - 1.3x_{2(i)}x_{3(i)} - 5.8x_{1(i)}^2 + \varepsilon_i$$

CASE 4: Primary Terms + all potential terms

$$y_i = 70.0 - 7.3x_{1(i)} + 10.0x_{2(i)} + 8.0x_{1(i)}x_{2(i)} - 3.0\ x_{3(i)} + 4.1x_{1(i)}x_{3(i)} - 5.3x_{2(i)}x_{3(i)} - 5.8x_{1(i)}^2 + 6.0x_{2(i)}^2 + \varepsilon_i$$

Note that the four *true models* defined above represent a full range of subsets of the *full model.* The smallest and largest possible subsets are given by *CASE 1* and *CASE 4*, respectively, while the other two cases are intermediate subsets. Together, these four cases represent a full range of potential regressor misspecification.

Since the true response model is unknown to the experimenter, the logical alternatives (competitors) to the two stage designs are single stage experiments designed 'optimally' for the *full model*. These competitors to the two stage procedure are as follows:

- 1-stage non-Bayesian D-optimal design for the full model, N=24
  (generated by SAS/QC from the 125 point candidate list , $\boldsymbol{X} = (-1, -.5, 0, .5, 1)^3$)
- 1-stage Bayes D-optimal design for the full model, N=24 (generated from $\boldsymbol{X}$)
- Central Composite Design for the 3 control variables with axial points set at ± 1 (Face Center Cube (FCC) design)
  - FCC with 2 center runs for a total of N=16 design points
  - FCC with 3 center runs for a total of N=17 design points

It can be argued that the single stage Q-optimal design should also be included in the preceding list. However, in order to find a Q-optimal design a very complex computer algorithm would be required. In fact, there are no commercial algorithms available for selecting Q-optimal designs. For this reason, the single stage Q-optimal design was not included as a competitor in this study.

Using the two stage procedures presented in sections 5.2-5.6, Bayesian two stage designs were generated in the following manner. Since p+q = 9 for the assumed full model, the recommended size for the first stage design is n_1=12 (as discussed in section 6.1). The first stage design was generated by selecting n_1=12 design points from $\boldsymbol{X}$ according to one of the methods presented in section 5.1. It was assumed throughout that τ=5.

From each of the four *true models* identified previously, fifty data sets were then simulated, with each data set containing 12 observations. Each of the fifty data sets produced a first stage data analysis and corresponding second stage design. For all analyses using the Box and Meyer model probabilities, a moderate value of π=.33 was assumed. (Recall that π is the prior probability that any potential term is important or 'active'.) Each second stage design was

generated by selecting n_2=12 design points from $\boldsymbol{X}$ according to one of the methods presented in sections 5.2 - 5.6.

In this evaluation, the performance of each design was measured by its design efficiency, relative to the *true models*. The two design measures used for comparison are the D and Q criteria, or more specifically,

$$D^* = |\, N(X^{*\prime}X^*)^{-1}| \ ,$$

$$Q^* = \mathrm{tr}\left[\, (X^{*\prime}X^*)^{-1}\, (N/k) \int_R \mathbf{x}^{(m)}\, \mathbf{x}'^{(m)}\, d\mathbf{x}\right] .$$

In the definitions of D* and Q* above, X* contains the N point design matrix with traditional ±1 scaling, expanded to contain only those regressors of the true model. Therefore, for evaluation purposes, the Bayesian two stage designs were rescaled when necessary to match the traditional designs.

The performance of each Bayesian two stage design procedure is measured by its average D* or Q* for each of the four true models. In other words, the reported D* and Q* values are actually

$$\overline{D}^* = \frac{\sum_{i=1}^{50} (D^*)_i}{50} \ , \quad \overline{Q}^* = \frac{\sum_{i=1}^{50} (Q^*)_i}{50},$$

which give the average performance of the 50 designs resulting from the simulation. The standard errors of $\overline{D}^*$ and $\overline{Q}^*$ are also recorded. Unlike the two stage procedures, single stage design procedures are not data dependent, and thus can be evaluated by a single D* or Q* value for each of the four true models. The results of the evaluation of the Bayesian two stage design procedures and their single stage competitors are given in the sections which follow.

6.3 Evaluation of Bayesian D-D Optimality Procedures

Table 6.3.1 contains the evaluation results for the Bayesian D-D optimality procedures presented in sections 5.2 and 5.3. The performance of these procedures is measured by $\overline{D}^*$ which is given in the first two columns of the table, along with the standard error of $\overline{D}^*$ (in parentheses). The remaining entries in table 6.3.1 are values of $\overline{D}^*$ for competing designs. The results show that the Bayesian D-D optimal designs are more efficient than their competitors in each case for which the true model contains fewer regressors than the full model. For CASE 4, in which the true model is the full model, the two stage D-optimal designs are just as efficient as the single stage D-optimal designs (which were constructed under this model assumption).

Table 6.3.1 Values of determinant for evaluation of Bayesian D-D optimal designs and single stage competitors

True Model 1: $y = 70.0 + 11.5x_1 + 7.3x_2 + 8.0x_1x_2 + \varepsilon$

D-D procedure (section 5.2)	**D-D procedure (section 5.3)**	**1-stage non-Bayes *D*-opt**	**1-stage Bayes *D*-opt**	**FCC 2 center runs**	**FCC 3 center runs**
1.97 *(0.04)*	2.03 *(0.04)*	2.28	2.28	5.12	6.14

True Model 2: $y = 70.0 + 11.5x_1 - 7.3x_2 + 8.0x_1x_2 + 1.1x_1x_3 - 1.3x_2x_3 + \varepsilon$

D-D procedure (section 5.2)	**D-D procedure (section 5.3)**	**1-stage non-Bayes *D*-opt**	**1-stage Bayes *D*-opt**	**FCC 2 center runs**	**FCC 3 center runs**
2.52 *(0.07)*	2.88 *(0.08)*	3.47	3.47	20.48	27.73

Table 6.3.1 continued

True Model 3: $y = 70.0 - 7.3x_1 + 10.0x_2 + 8.0x_1x_2 + 1.1x_1x_3 - 1.3x_2x_3 - 5.8x_1^2 + \varepsilon$

D-D procedure (section 5.2)	D-D procedure (section 5.3)	1-stage non-Bayes *D*-opt	1-stage Bayes *D*-opt	FCC 2 center runs	FCC 3 center runs
19.77 *(0.18)*	20.20 *(0.19)*	21.08	21.08	87.38	114.49

True Model 4: $y = 70.0 - 7.3x_1 + 10.0x_2 + 8.0x_1x_2 - 3.0x_3 + 1.1x_1x_3 - 1.3x_2x_3 - 5.8x_1^2 + 6.0x_2^2 + \varepsilon$

D-D procedure (section 5.2)	D-D procedure (section 5.3)	1-stage non-Bayes *D*-opt	1-stage Bayes *D*-opt	FCC 2 center runs	FCC 3 center runs
158.31 *(0.00)*	158.31 *(0.00)*	158.31	158.31	762.60	1092.5

Note: Tabled value is D*, the non-Bayesian D-criterion value of the true model.

6.4 Evaluation of the Bayesian D-Q Optimality Procedure (A Linear Combination of Q-optimality Criteria)

Performance of the Bayesian D-Q procedure, relative to its competitors, is measured by both D* and Q* , since the design objective differs between the two stages. The results are included in table 6.4.1. Note that the standard errors of $\overline{D}$* and $\overline{Q}$*for the two stage procedure are shown in parentheses in the first column of the table. The results show that the D-Q designs are more D-efficient than all competitors other than the single stage D-optimal designs. Interestingly, the single stage D-optimal designs also outperformed the D-Q designs in terms of Q-efficiency when the model was greatly misspecified (i.e., cases 1 and 2). The roles reversed, however, for the remaining cases.

Table 6.4.1 Values of determinant and Q for evaluation of Bayesian D-Q optimal designs and single stage competitors

True Model 1: $y = 70.0 + 11.5x_1 + 7.3x_2 + 8.0x_1x_2 + \varepsilon$

	D-Q procedure	**1-stage non-Bayes *D*-opt**	**1-stage Bayes *D*-opt**	**FCC 2 ctr runs**	**FCC 3 ctr runs**
D*	3.87 *(0.29)*	2.28	2.28	5.12	6.14
Q*	2.16 *(0.02)*	1.99	1.99	2.29	2.37

True Model 2: $y = 70.0 + 11.5x_1 - 7.3x_2 + 8.0x_1x_2 + 1.1x_1x_3 - 1.3x_2x_3 + \varepsilon$

	***D*-Q procedure**	**1-stage non-Bayes *D*-opt**	**1-stage Bayes *D*-opt**	**FCC 2 ctr runs**	**FCC 3 ctr runs**
D*	6.88 *(0.43)*	3.47	3.47	20.48	27.73
Q*	2.44 *(0.02)*	2.27	2.27	2.73	2.84

True Model 3: $y = 70.0 - 7.3x_1 + 10.0x_2 + 8.0x_1x_2 + 1.1x_1x_3 - 1.3x_2x_3 - 5.8x_1^2 + \varepsilon$

	D-Q procedure	**1-stage non-Bayes *D*-opt**	**1-stage Bayes *D*-opt**	**FCC 2 ctr runs**	**FCC 3 ctr runs**
D*	56.08 *(4.01)*	21.08	21.08	87.38	114.49
Q*	3.26 *(0.01)*	4.08	4.08	3.48	3.48

Table 6.4.1 continued

True Model 4: $y = 70.0 - 7.3x_1 + 10.0x_2 + 8.0x_1x_2 - 3.0\,x_3 + 1.1x_1x_3 - 1.3x_2x_3 - 5.8x_1^2 + 6.0\,x_2^2 + \varepsilon$

	D-Q procedure	**1-stage non-Bayes *D*-opt**	**1-stage Bayes *D*-opt**	**FCC 2 ctr runs**	**FCC 3 ctr runs**
D*	726.01 *(5.30)*	158.31	158.31	762.60	1092.53
Q*	4.40 *(0.00)*	6.83	6.83	4.73	4.76

6.5 Evaluation of the Bayesian Q-Q Optimality Procedure (A Linear Combination of Q-optimality Criteria)

The evaluation results for the Bayesian Q-Q optimality procedure can be seen in table 6.5.1. This table includes both D* and Q* values so that comparisons can be made between the performance of the Bayesian Q-Q and D-Q optimality procedures. Like the D-Q designs, the Bayesian Q-Q optimal designs are more D-efficient than all competitors other than the single stage D-optimal designs. The single stage D-optimal designs again outperformed the Q-Q designs in terms of Q* when the model was greatly misspecified (i.e., cases 1 and 2). In fact, by comparing the results of tables 6.4.1 and 6.5.1, it seems that there is very little difference in the Bayesian Q-Q and D-Q designs, as measured by Q*.

Table 6.5.1 Values of determinant and Q for evaluation of Bayesian Q-Q optimal designs and single stage competitors

True Model 1: $y = 70.0 + 11.5x_1 + 7.3x_2 + 8.0x_1x_2 + \varepsilon$

	Q-Q procedure	**1-stage non-Bayes D-opt**	**1-stage Bayes D-opt**	**FCC 2 ctr runs**	**FCC 3 ctr runs**
D*	3.57 *(0.22)*	2.28	2.28	5.12	6.14
Q*	2.13 *(0.02)*	1.99	1.99	2.29	2.37

True Model 2: $y = 70.0 + 11.5x_1 - 7.3x_2 + 8.0x_1x_2 + 1.1x_1x_3 - 1.3x_2x_3 + \varepsilon$

	Q-Q procedure	**1-stage non-Bayes D-opt**	**1-stage Bayes D-opt**	**FCC 2 ctr runs**	**FCC 3 ctr runs**
D*	7.51 *(0.89)*	3.47	3.47	20.48	27.73
Q*	2.45 *(0.02)*	2.27	2.27	2.73	2.84

True Model 3: $y = 70.0 - 7.3x_1 + 10.0x_2 + 8.0x_1x_2 + 1.1x_1x_3 - 1.3x_2x_3 - 5.8x_1^2 + \varepsilon$

	Q-Q procedure	**1-stage non-Bayes D-opt**	**1-stage Bayes D-opt**	**FCC 2 ctr runs**	**FCC 3 ctr runs**
D*	60.93 *(3.30)*	21.08	21.08	87.38	114.49
Q*	3.31 *(0.01)*	4.08	4.08	3.48	3.48

Table 6.5.1 continued

True Model 4: $y = 70.0 - 7.3x_1 + 10.0x_2 + 8.0x_1x_2 - 3.0\,x_3 + 1.1x_1x_3 - 1.3x_2x_3 - 5.8x_1^2 + 6.0\,x_2^2 + \varepsilon$

	Q-Q procedure	1-stage non-Bayes D-opt	1-stage Bayes D-opt	FCC 2 ctr runs	FCC 3 ctr runs
D*	693.02 *(2.63)*	158.31	158.31	762.60	1092.5
Q*	4.41 *(0.00)*	6.83	6.83	4.73	4.76

6.6 Evaluation of the Bayesian D-DQ Optimality Procedure (A Linear Combination of DQ-optimality Criteria)

Table 6.6.1 contains the evaluation results for the Bayesian D-DQ optimality procedure, as measured by both D* and Q*. For all models except *True Model 4* (the full model), the Bayesian D-DQ designs are more D and Q efficient than their competitors. In fact, comparing the results in table 6.6.1 to those in tables 6.4.1 and 6.5.1, the Bayesian D-DQ optimality procedure seems to be a very strong alternative to both the Bayesian D-Q and Q-Q procedures.

There are two possible reasons why the D-DQ designs do not compete well with the single stage designs for *True Model 4*. First, *Model 4* is the full model, for which the single stage designs were constructed. In other words, this is the model for which the single stage designs perform the best. The D-DQ designs, however, were constructed under some model uncertainty. The second reason is that the selection of the D-DQ designs is an attempt to simultaneously satisfy both the D and Q criteria, resulting in a design which is somewhat of a compromise between the Bayes D-optimal and Q-optimal designs.

Table 6.6.1 Values of determinant and Q for evaluation of Bayesian D-DQ optimal designs and single stage competitors

True Model 1: $y = 70.0 + 11.5x_1 + 7.3x_2 + 8.0x_1x_2 + \varepsilon$

	D-DQ procedure	**1-stage non-Bayes D-opt**	**1-stage Bayes D-opt**	**FCC 2 ctr runs**	**FCC 3 ctr runs**
D*	1.60 *(0.03)*	2.28	2.28	5.12	6.14
Q*	1.90 *(0.01)*	1.99	1.99	2.29	2.37

True Model 2: $y = 70.0 + 11.5x_1 - 7.3x_2 + 8.0x_1x_2 + 1.1x_1x_3 - 1.3x_2x_3 + \varepsilon$

	D-DQ procedure	**1-stage non-Bayes D-opt**	**1-stage Bayes D-opt**	**FCC 2 ctr runs**	**FCC 3 ctr runs**
D*	2.14 *(0.07)*	3.47	3.47	20.48	27.73
Q*	2.17 *(0.01)*	2.27	2.27	2.73	2.84

True Model 3: $y = 70.0 - 7.3x_1 + 10.0x_2 + 8.0x_1x_2 + 1.1x_1x_3 - 1.3x_2x_3 - 5.8x_1^2 + \varepsilon$

	D-DQ procedure	**1-stage non-Bayes D-opt**	**1-stage Bayes D-opt**	**FCC 2 ctr runs**	**FCC 3 ctr runs**
D*	18.08 *(0.15)*	21.08	21.08	87.38	114.49
Q*	3.46 *(0.02)*	4.08	4.08	3.48	3.48

Table 6.6.1 continued

True Model 4: $y = 70.0 - 7.3x_1 + 10.0x_2 + 8.0x_1x_2 - 3.0\,x_3 + 1.1x_1x_3 - 1.3x_2x_3 - 5.8x_1^2 + 6.0\,x_2^2 + \varepsilon$

	D-DQ procedure	**1-stage non-Bayes D-opt**	**1-stage Bayes D-opt**	**FCC 2 ctr runs**	**FCC 3 ctr runs**
D*	241.43 *(6.60)*	158.31	158.31	762.60	1092.5
Q*	9.26 *(0.19)*	6.83	6.83	4.73	4.76

6.7 Evaluation Summary

When the true process model is a subset of an assumed full model, designs which are more efficient than traditional designs can be obtained by using a Bayesian two stage procedure. Both of the Bayesian D-D optimality procedures performed particularly well relative to their single stage competitors. If prediction properties are important to the experimenter, then the Bayesian D-DQ procedure appears to be the best choice. It not only minimizes prediction variance, but minimizes variability associated with estimation as well.

Chapter 7

Design Augmentation for Variance Modeling

Thus far, the focus in this paper has been on finding efficient experimental designs for building a process mean model. It has been assumed that the process variance is constant across all levels of the control variables. However, there are many applications in RSM for which some control variables act as *dispersion effects*, resulting in a heteroscedastic variance structure across the design region. When this is the case, it is of interest to obtain additional data in order to model the process variance as a function of control variables. This chapter presents a Bayesian approach to design augmentation for efficiently collecting data for variance modeling.

7.1 Variance Modeling

In most RSM applications, the objective is to build a statistical model which describes the relationship between the mean of a response and a set of control variables. This process mean model,

$$\mathbf{y} = f(\mathbf{X}\beta) + \varepsilon$$

is typically built under the assumption that

$$\varepsilon \quad N(\mathbf{0}, \sigma^2\mathbf{I}).$$

In other words, a common assumption is that the variance of the response, y, is constant across all levels of the control variables (i.e., the *x*'s). The residuals of the fitted model are then used to gain diagnostic information about the validity of this assumption. For example, suppose that a plot of the residuals against the level of a factor of a two-level design takes the form in figure 7.1.1. This plot appears to give some evidence that the model error variance increases as the level of x increases. In this case, factor x would be considered as a possible *dispersion effect*.

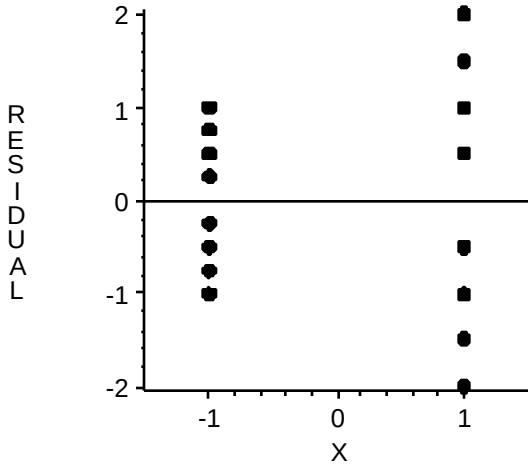

Figure 7.1.1 Plot of residuals vs. levels of factor x

When one or more control variables are suspected to be dispersion effects, additional process knowledge can be gained by modeling the error variance as a function of these variables. This is known as *variance modeling.* (See chapter 10 of Myers and Montgomery (1995) for a thorough discussion on variance modeling.)

Although it is possible to model the variance using residuals of the fitted mean model, it is not necessarily a good practice since the residuals may be impacted by model misspecification (i.e. lack-of-fit). Instead, it is recommended that replicate observations are collected at various points in the design space. The sample variances at these d design points are then commonly modeled using a log linear model of the type given in (7.1.1).

$$\ln s_i^2 = \mathbf{z}_i'\gamma + \varepsilon_i^* , \quad i = 1, 2,..., d \qquad (7.1.1)$$

In (7.1.1), the $\mathbf{z}_i'\gamma$ portion of the model refers to a linear model in a set of control variables (i.e., the *x*'s). The notation $\mathbf{z}_i$ is used to distinguish the set of p regressors used in the variance model from that of the mean model. Although $\mathbf{z}_i$ could include cross-product and/or quadratic terms, in most practical applications $\mathbf{z}_i'\gamma$ is a first order model in a subset of the control variables under study.

Under the assumption that the errors are normal around the mean model, it can be shown that the model in (7.1.1) is a constant variance model. In other words, if

$$y_{ij} = \mathbf{x}_i'\beta + \varepsilon_{ij}\sigma_i, \quad i = 1, 2,..., d; \; j = 1, 2,..., r$$

where $\varepsilon_{ij} \sim N(0,1)$, $\ln \sigma_i^2 = \mathbf{z}_i'\gamma$ and r is the number of observations collected at each of the d replicated design points, then ε_i^* follows a log χ^2_{r-1} distribution with variance 2/(r-1). Details of this are given in Appendix A. With constant error variance, it is appropriate to estimate the unknown γ's using ordinary least squares, i.e.,

$$\hat{\gamma} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{ln\ s^2}) \qquad (7.1.2)$$

where $(\mathbf{ln\ s^2})' = [\ln s_1^2, \ln s_2^2, \ldots, \ln s_d^2]$ and $\mathbf{Z}$ is the d×p model matrix. It follows from (7.1.2) that

$$\mathrm{Var}(\hat{\gamma}) = (\mathbf{Z}'\mathbf{Z})^{-1}\, 2/(r-1)\,. \qquad (7.1.3)$$

Based on (7.1.2) and (7.1.3), approximate z-tests can be formed and utilized in model selection.

From the preceding paragraphs, hopefully it is clear that variance modeling is very straightforward after replicated observations have been obtained at each of the d design points. The real difficulty in most practical applications is in finding the resources needed to collect this additional data. With a limited total sample size available for the experiment, there comes a tradeoff between the number of design points which can be replicated and the degrees of freedom available for each sample variance. It is obvious that *design efficiency* becomes an extremely important issue. For this reason, a Bayesian method has been developed for the purpose of efficiently augmenting a previous experiment in order to acquire the replication needed for variance modeling. This Bayesian design augmentation method is presented in the next section.

7.2 A Bayesian D-optimal Design Augmentation Procedure

Very little work has been published in Bayesian design procedures for the variance model. Vining and Schaub (1996) developed what they refer to as a 'semi-Bayesian' approach to finding a D-optimal design for simultaneously modeling the mean and variance of a response. Using their approach, the resulting single stage design is D-optimal for a specific set of regressors in the mean and variance models, leaving no room for regressor uncertainty.

As illustrated in section 7.1, mean and variance modeling often occur as a two stage process. In the first stage, the mean model is determined after executing an appropriately designed experiment. This 'first stage' design will be referred to as the *base design*. In the second stage, the base design is augmented to supply the replicates needed for variance modeling. Although the mean model is assumed to be known after the first stage, there is still uncertainty as to which control variables should be included in the variance model. A Bayesian D-optimal design augmentation procedure makes it possible to add the needed replication even though initial knowledge of the variance model is poor.

The Bayesian procedure actually consists of two phases; model discrimination and design augmentation. First, initial variance information is used in the calculation of Box and Meyer posterior probabilities for each of the candidate variance models. These probabilities are used as inputs to a Bayesian D-optimality criterion which controls the design augmentation. The idea is to ensure that the replicates are chosen to provide information about the candidate model(s) with the highest likelihood of being the 'best' variance model.

7.2.1 Initial Variance Information

It is assumed that a log linear model of the type given in (7.1.1) will be used to model the variance. Recall that this model is of the form

$$\ln s_i^2 = \mathbf{z}_i'\gamma + \varepsilon_i^* , \quad i = 1, 2,..., d$$

where d is the number of design points at which there has been some replication. Additionally, it is assumed that the linear portion, $\mathbf{z}_i`\gamma$, is a first order model in at most k control variables. In other words, the full variance model under consideration is defined by

$$\mathbf{z}_i` = [1, x_{1i}, x_{2i}, ..., x_{ki}] , \quad i = 1, 2,..., d.$$

This implies that there are $m = \sum_{i=1}^{k} \binom{k}{i}$ candidate variance models in addition to the intercept model, with each model defined by a subset of the k control variables. Let the intercept model be denoted as M_0, with the remaining candidate models denoted as M_1, M_2, ..., M_m.

Before the process of model discrimination can begin, there must be enough data available to provide information about all candidate variance models, including the full model. As mentioned

previously, this information is best obtained through the calculation of sample variances. Since each sample variance must be based on at least two observations, design replication is required.

It is quite possible that intuition about potential dispersion effects led the experimenter to include adequate replication in the base design. If this is not the case, then he/she would need to add enough points to the base design to obtain r=2 observations at each point of a *resolution III fractional factorial design* in $x_1, x_2, ..., x_k$. The resolution III fraction is chosen for four reasons. First, by definition the resolution III is the smallest fraction which supports estimation of main effects independently of one another. Secondly, the resolution III fraction is contained in many classical designs, such as the full factorial and central composite designs, which are often used in the development of the process mean model. This means that often only a single replicate of the resolution III fraction would need to be added in order to obtain variance information. Thirdly, Vining and Schaub (1996) found that a replicated resolution III fraction is a very D-efficient single stage design for a first order log linear variance model. The fourth and final reason for choosing to augment the base design in this manner is simplicity. No computer algorithm is needed to choose the design augmentation, since the experimenter can simply examine the base design and determine which points should be added to result in a replicated resolution III fraction.

7.2.2 Model Discrimination

The two observations at each point of the resolution III fraction (of size d) are summarized by their sample variance,

$$s_i^2 = \sum_{j=1}^{2} \left(y_{ij} - \bar{y}_{i.} \right)^2, \; i = 1, 2, ..., d.$$

Since each s_i^2 (i = 1, 2, ..., d) has one degree of freedom, $s_i^2 \quad \sigma_i^2 \chi_1^2$, and thus its probability density function is of the form

$$f(s_i^2) = c\left(\sigma_i^2 s_i^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\frac{s_i^2}{\sigma_i^2}\right\}$$

where c is a constant. See Appendix B for details. Under the assumption that $\sigma_i^2=\exp\{\mathbf{z}_i'\gamma\}$, then the density can be rewritten as

$$f(s_i^2) = c\left(\exp\{\mathbf{z}_i'\gamma\} s_i^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\frac{s_i^2}{\exp\{\mathbf{z}_i'\gamma\}}\right\}.$$

Since $s_1^2, s_2^2, ..., s_d^2$ are independent of one another, their joint density can be written as

$$f(s_1^2, s_2^2, ..., s_d^2 | \gamma) = c^d \left(\prod_{i=1}^{d} \exp\{\mathbf{z}_i'\gamma\} s_i^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{d}\frac{s_i^2}{\exp\{\mathbf{z}_i'\gamma\}}\right\}. \qquad (7.2.2a)$$

The above density function for $\mathbf{s^2} = (s_1^2, s_2^2, ..., s_d^2)$ is used in the calculation of Box and Meyer posterior probabilities for the m+1 candidate variance models. These probabilities will weight the importance of each candidate model when determining the D-optimal design augmentation. Recall that in using the Box and Meyer approach, the posterior probability that candidate model M_j is the most appropriate variance model is given by

$$p(M_j|\mathbf{s^2}) = \frac{p(M_j) f(\mathbf{s^2}|M_j)}{\sum_{i=0}^{m} p(M_i) f(\mathbf{s^2}|M_i)} \qquad (7.2.2b)$$

when the prior probability of M_j is $p(M_j)$.

Each candidate model M_j contains k_j control variables, ($0 \le k_j \le k$). It is assumed that there is uniform uncertainty about all k variables prior to observing $\mathbf{s^2}$, so the prior probability of model M_j being the 'best' model is given by

$$p(M_j) = \pi^{k_j}(1-\pi)^{k-k_j}, \quad j = 0, 2, ..., m$$

for a specified probability, π. The value of π should be chosen to represent the proportion of control variables which are believed to be dispersion effects. It should be noted that setting π=.5 would result in a uniform prior for the set of candidate models.

In (7.2.2b), the predictive density of $\mathbf{s^2}$, given model M_j, is given by the expression

$$f(\mathbf{s^2} \mid M_j) = \int_{\Gamma^{(j)}} f(\mathbf{s^2} \mid M_j, \gamma^{(j)})\, f(\gamma^{(j)} \mid M_j)\, d\gamma^{(j)}$$

where $\gamma^{(j)}$ is the parameter vector of M_j. Substituting (7.2.2a) into the predictive density function results in

$$f(\mathbf{s^2} \mid M_j) = \int_{\Gamma_j} c^d \left[\prod_{i=1}^{d} \exp\left\{\mathbf{z}_i'\gamma^{(j)}\right\} s_i^2\right]^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{d} \frac{s_i^2}{\exp\left\{\mathbf{z}_i'\gamma^{(j)}\right\}}\right] f(\gamma^{(j)} \mid M_j)\, d\gamma^{(j)} \qquad (7.2.2c)$$

where $f(\gamma^{(j)} \mid M_i)$ is the prior density of $\gamma^{(j)}$ and $\Gamma^{(j)}$ is the set of all possible values of $\gamma^{(j)}$.

Prior to observing $\mathbf{s^2}$, it is reasonable to assume that $\gamma^{(j)} \sim N(\gamma_0^{(j)}, \tau^2 I)$. The prior mean, , $\gamma_0^{(j)}$ is a (k_j+1) dimensional vector of the form

$$\gamma_0^{(j)} = \begin{bmatrix} \ln MSE_1 \\ \mathbf{0}_{k_j \times 1} \end{bmatrix}$$

where MSE_1 is the mean squared error resulting from the ANOVA of the mean model (using only the data from the base design). Since there is usually a great deal of uncertainty about γ prior to observing $\mathbf{s}^2$, it is reasonable that τ would be assigned a large value, such as τ=10.

Having defined the prior distribution of γ, the next step is to carry out the integration as defined in (7.2.2c). Since this is an intractable task, Monte Carlo methods are utilized in order to solve for the predictive density, $f(\mathbf{s}^2 \mid M_j)$, j = 0, 1, ..., m.

7.2.3 Design Augmentation with Reduced Model Uncertainty

The Box and Meyer posterior model probabilities described in section 7.2.2 are incorporated into a Bayesian D-optimality procedure for the variance model. The objective of this augmentation procedure is to determine the most D-efficient plan for replicating the resolution III fraction, given the variance model information provided by the Box and Meyer probabilities. The development of the procedure follows.

Recall that the model form to be used in modeling the response variance is

$$\ln s_i^2 = \mathbf{z}_i'\gamma + \varepsilon_i^* , \quad i = 1, 2,..., d$$

with ν_i +1 observations going into each s_i^2 after the final design augmentation is complete.
Since there are m+1 candidate variance models under consideration, let a given candidate model M_j be denoted as

$$\ln s_i^2 = (\mathbf{z}_i'\gamma)^{(j)} + \varepsilon_i^* , \quad i = 1, 2,..., d ,$$

using the superscript (j) to reflect the appropriate model space for M_j. The D-optimal design for model M_j is that which maximizes the determinant of the Fisher information matrix (or

equivalently minimizes the generalized variance of $\gamma^{(j)}$). In other words, the D-optimality criterion is

$$\underset{D}{\text{Max}} \; | \; \mathbf{I}(\gamma^{(j)}) \; | \tag{7.2.3a}$$

where

$$\mathbf{I}(\gamma^{(j)}) = \sum_{i=1}^{d} \left[\frac{\nu_i \left(\mathbf{z}_i' \omega^{(j)} \right)}{\exp\left(\mathbf{z}_i' \gamma^{(j)} \right)} - \frac{1}{2} \nu_i \left(\mathbf{z}_i' \omega^{(j)} \right) \right]. \tag{7.2.3b}$$

See Appendix C for the derivation of the Fisher information matrix. Notice, however, that (7.2.3b) depends on the unknown parameter vector, $\gamma^{(j)}$, and therefore a criterion other than that in (7.2.3a) is needed. Following the approach of Vining and Schaub (1996), $\gamma^{(j)}$ can be integrated out of the information matrix, based on the most recent information available about $\gamma^{(j)}$.

Prior to performing the design augmentation, it is reasonable to assume that

$$\gamma^{(j)} \quad N(\gamma_0^{(j)}, \tau^2 I),$$

where

$$\gamma_0^{(j)} = \begin{bmatrix} \ln \text{MSE}_2 \\ \mathbf{0}_{k_j \times 1} \end{bmatrix}.$$

The term MSE_2 is the mean squared error for the mean model using all data which exists at this point. Since there is still a great deal of uncertainty about $\gamma^{(j)}$ prior to the final augmentation, it is reasonable that τ would be assigned a large value, such as τ=10.

Taking the expectation of the information matrix in (7.2.3b) over $\pi(\gamma^{(j)})$ (defined above), results in

$$E^{\gamma}[\mathbf{I}(\gamma^{(j)})] = \sum_{i=1}^{k} \left[\nu_i (\mathbf{z}_i \mathbf{z}_i')^{(j)} \exp\left\{ -(\mathbf{z}_i' \gamma_0)^{(j)} + \frac{1}{2}\tau^2 \mathbf{z}_i' \mathbf{z}_i \right\} - \frac{1}{2} \nu_i (\mathbf{z}_i \mathbf{z}_i')^{(j)} \right]. \qquad (7.2.3c)$$

Expressing (7.2.3c) in matrix form gives

$$E^{\gamma}[\mathbf{I}(\gamma^{(j)})] = c_j \mathbf{Z}^{(j)'} \mathbf{V} \mathbf{Z}^{(j)}$$

where c_j is a constant for model M_j, $Z^{(j)}$ is the model matrix for M_j and V is a d×d diagonal matrix with diagonal elements $\nu_1, \nu_2, ..., \nu_d$. Thus the Bayes D-optimal design for model M_j is that which maximizes

$$D_j = | c_j \mathbf{Z}^{(j)'} \mathbf{V} \mathbf{Z}^{(j)} |.$$

Model M_j is just one of the m candidate variance models under consideration. Utilizing the Box and Meyer posterior probabilities for models M_0, M_1, ..., M_m, a weighted average D criterion can be used to select the design augmentation. The objective is to choose the remaining design points so as to minimize D_j for each model M_j having a high probability of being the 'best' model. This is done by choosing the design augmentation so as to maximize

$$\sum_{M_j} D_j p(M_j | \mathbf{s}^2). \qquad (7.2.3d)$$

7.3 An Example Using the Bayesian Design Augmentation Procedure

Suppose that a full factorial experiment has been run in order to model the mean of a response, y, as a function of three variables (x_1, x_2, x_3). The full factorial design, which shall be refered to as the 'base design', is shown in figure 7.3.1.

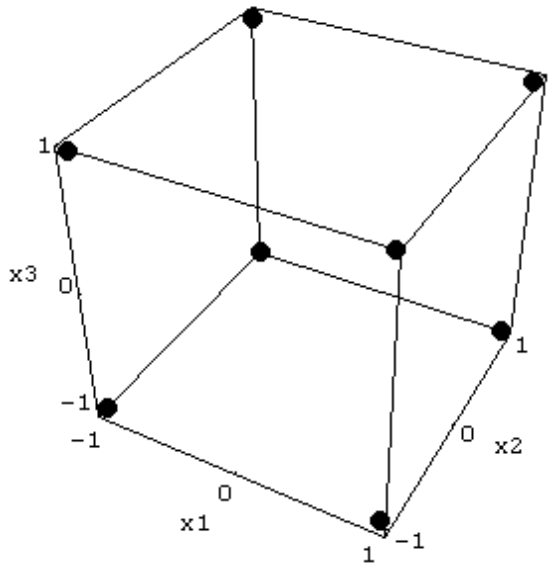

Figure 7.3.1 Full factorial design in variables x_1, x_2, x_3

Table 7.3.1 contains the data collected from this experiment. It should be pointed out that this data was actually simulated, under the assumption that the true variance model is given by

$$\ln \sigma_i^2 = 3.92 + 3.09x_{1i}, \; i = 1, 2, ..., 8 .$$

In other words, there is a heteroscedastic error variance structure, due to the fact that variable x_1 is a dispersion effect.

Table 7.3.1 Full factorial design and resulting data

x_1	x_2	x_3	y
-1	-1	-1	-112.47
-1	+1	+1	397.80
+1	-1	+1	562.99
+1	+1	-1	385.69
+1	+1	+1	684.66
+1	-1	-1	226.09
-1	-1	+1	156.64
-1	+1	-1	122.64

The fitted model,

$$y_i = 303.00 + 161.85x_1 + 94.69x_2 + 147.52x_3 - 24.37x_1x_2$$

has a mean squared error of 472.62. The residuals from this model were plotted against each of the three variables in order to check the assumption of homoscedasticity. These plots are included in figures 7.3.2, 7.3.3 and 7.3.4. From these plots, it appears that both x_1 and x_2 are possible dispersion effects and therefore variance modeling is warranted.

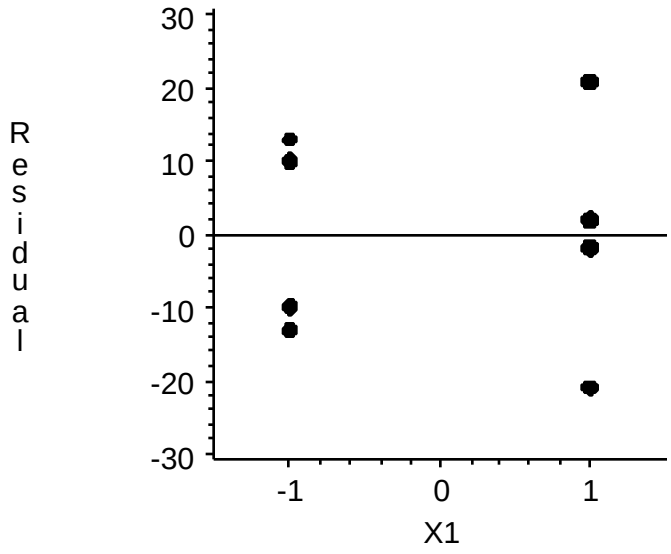

Figure 7.3.2 Plot of residuals against levels of x_1

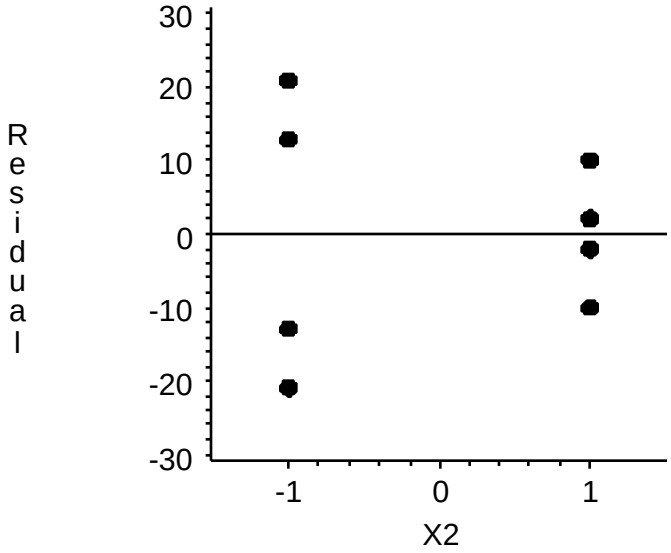

Figure 7.3.3 Plot of residuals against levels of x_2

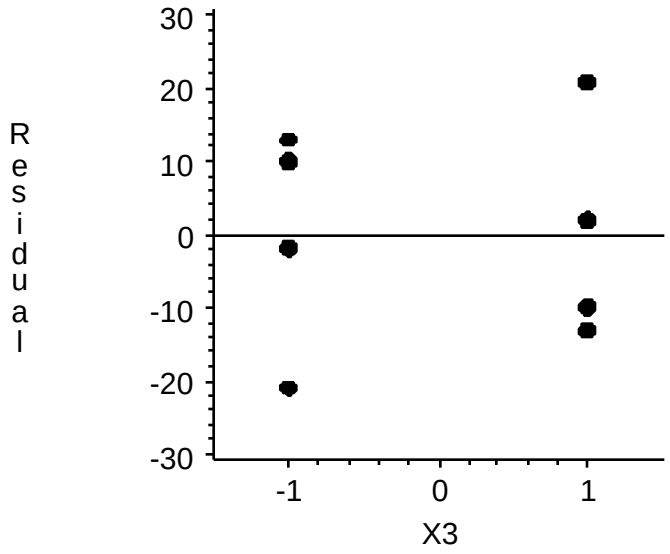

Figure 7.3.4 Plot of residuals against levels of x_3

The base design must be augmented to include a single replication of the resolution III fraction. The resulting design is shown in figure 7.3.5, with a (2) denoting the points at which there are now two observations. Note, however, that the base design would likely have included this replication had there been any prior intuition about the need for variance modeling.

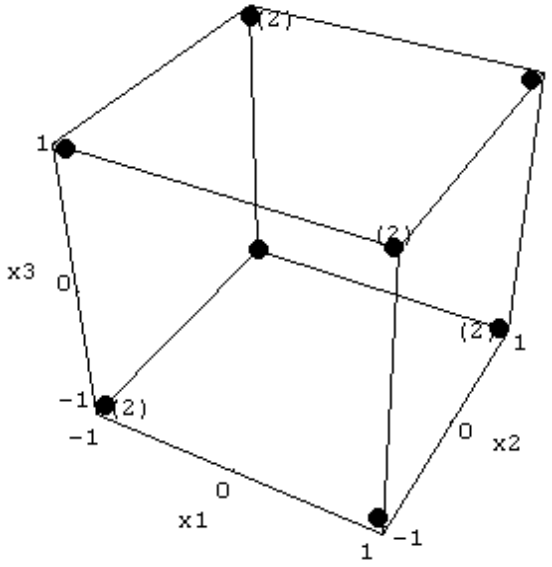

Figure 7.3.5 The base design with replication

Sample variances are calculated for each point of the resolution III fraction. The results are included in Table 7.3.2.

Table 7.3.2 Data from the replicated resolution III fraction

x_1	x_2	x_3	y_1	y_2	s^2
-1	-1	-1	-112.47	-110.67	1.62
-1	+1	+1	397.80	395.80	2.00
+1	-1	+1	562.99	488.65	2763.22
+1	+1	-1	385.69	365.43	205.23

The next step in the augmentation procedure is to use the four sample variances as inputs to the model discrimination process. The full variance model is of the form

$$\ln s_i^2 = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 x_{3i} + \varepsilon_i^* .$$

Since there are k=3 variables in the full model, there are eight candidate variance models. Prior to observing the data from the first augmentation, it is believed that $\gamma = [\gamma_0, \gamma_1, \gamma_2, \gamma_3]'$ is normally distributed with prior mean, γ_0 of the form

$$\gamma_0 = \begin{bmatrix} \ln MSE \\ \mathbf{0}_{3\times 1} \end{bmatrix} = \begin{bmatrix} \ln(472.62) \\ \mathbf{0}_{3\times 1} \end{bmatrix},$$

where 472.62 is the MSE of the mean model prior to the initial augmentation. Since there is still great deal of uncertainty about γ, the prior variance of γ is specified to be $\tau^2 I$, where τ=10. The Box and Meyer probabilities are computed under this distribution assumption using Monte Carlo methods (n=2000). The resulting probabilities are included in table 7.3.3.

Table 7.3.3 Box and Meyer posterior probabilities for candidate variance models

Posterior Probability	**Terms Included in the Candidate Model**
.03	Intercept
.65	Intercept, x_1
.01	Intercept, x_2
.01	Intercept, x_3
.12	Intercept, x_1, x_2
.15	Intercept, x_1, x_3
.01	Intercept, x_2, x_3
.02	Intercept, x_1, x_2, x_3

Recall that the true variance model from which the data was simulated contained only an intercept and the x_1 term. From table 7.3.3 it is clear that the Box and Meyer approach has done a good job in discriminating among the candidate models, since the highest probability is given to this model.

Suppose that the experimenter can afford to collect 10 additional data points for the purpose of variance modeling. Based on the probabilities/weights given in table 7.3.3, the Bayes D-optimal augmentation is selected, using the criterion defined in (7.2.3d). The design criterion determines the most efficient method of replicating the resolution III fraction, based on a specified number of runs (10, in this case). The complete design, including the base design, first augmentation and Bayes D-optimal augmentation, is shown in figure 7.3.6. Notice that of the 10 available runs for the final augmentation, 8 are used to obtain two more replicates of the entire resolution III fraction. The remaining two runs are used to get additional information about variable x_1, since it was the only variable included in the highest weighted candidate model. It is interesting to see that if the design were collapsed down to include only x_1, the resulting design would have 11 replicates at each level of x_1. This would clearly be a strong design for modeling the variance as a function of x_1.

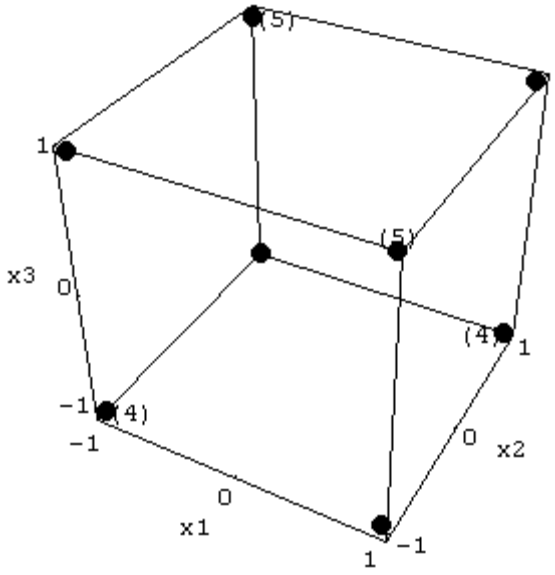

Figure 7.3.6 Base design with first augmentation and Bayes D-optimal augmentation

7.4 Special Case: Uniform Replication of a Resolution III Fraction

Suppose that after modeling the mean of a response it is determined that variance modeling is necessary or suppose it is known at the outset that variance modeling is needed. Assume that the base design contains at least a resolution III fraction for all control variables in the mean model (and hence defining the full variance model). Let N denote the number of additional data points that the experimenter can afford to collect to accomplish this task. In what follows we will show that if N is a multiple of d, where d is the number of points in the appropriate resolution III fraction, then the optimal design augmentation is to replicate the entire fraction N/d times.

Theorem 1

Let D_z denote the base design. Assume that the log linear variance model is first order. Assume that D_z contains at least a resolution III fraction for the variables in the full variance model. Then the D-optimal design augmentation is to replicate the resolution III fraction r times (r=1, 2, ...).

Proof

In general, if **A** is any square matrix partitioned as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

then $|\mathbf{A}| = |\mathbf{A}_{22}|\,|\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}|$, assuming $|\mathbf{A}_{22}| \neq 0$. Let X be a first order model matrix, such that

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix}.$$

Let $A = X'X$ be partitioned as

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \left[\begin{array}{c|ccc} n & \sum_{i=1}^{n} x_{1i} & \cdots & \sum_{i=1}^{n} x_{ki} \\ \hline \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}^2 & \cdots & \sum_{i=1}^{n} x_{1i}x_{ki} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{n} x_{ki} & \sum_{i=1}^{n} x_{1i}x_{ki} & \cdots & \sum_{i=1}^{n} x_{ki}^2 \end{array}\right],$$

then $|\mathbf{X}'\mathbf{X}| = |\mathbf{A}| = |\mathbf{A}_{22}|\,|n - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}|$.

Note that $|X'X|$ is maximized when

(a) $|\mathbf{A}_{22}|$ is maximized

and (b) $\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} = 0 \rightarrow \mathbf{A}_{12} = \mathbf{0} \rightarrow \sum_{i=1}^{n} x_{ij} = 0\ ,\ j = 1, 2, ..., k,$

since $\mathbf{A}_{12}$ is the vector of a positive definite quadratic form.

In order to determine the conditions which maximize $|\mathbf{A}_{22}|$, now let $\mathbf{A}=\mathbf{A}_{22}$, partitioned as

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \left[\begin{array}{c|ccc} \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i}x_{2i} & \cdots & \sum_{i=1}^{n} x_{1i}x_{ki} \\ \hline \sum_{i=1}^{n} x_{1i}x_{2i} & \sum_{i=1}^{n} x_{2i}^2 & \cdots & \sum_{i=1}^{n} x_{2i}x_{ki} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{n} x_{1i}x_{ki} & \sum_{i=1}^{n} x_{2i}x_{ki} & \cdots & \sum_{i=1}^{n} x_{ki}^2 \end{array}\right].$$

Thus, $|\mathbf{A}| = |\mathbf{A}_{22}|\ |\sum x_{1i}^2 - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}|$ is maximized when the following conditions hold:

(a) $|\mathbf{A}_{22}|$ is maximized

(b) $\sum_{i=1}^{n} x_{1i}^2$ is maximized

(c) $\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} = 0 \rightarrow \sum_{i=1}^{n} x_{1i}x_{ji} = 0\ ,\ j = 2, ..., k$

The same technique used to find the above conditions is repeated for each submatrix A_{22}, until finally

$$\mathbf{A}_{22} = \begin{bmatrix} \sum_{i=1}^{n} x_{(k-1)i}^2 & \sum_{i=1}^{n} x_{(k-1)i}x_{ki} \\ \sum_{i=1}^{n} x_{(k-1)i}x_{ki} & \sum_{i=1}^{n} x_{ki}^2 \end{bmatrix}.$$

The determinant of this last submatrix $\mathbf{A}_{22}$ is maximized when

(a) $\sum_{i=1}^{n} x_{(k-1)i}^2$ and $\sum_{i=1}^{n} x_{ki}^2$ are maximized

and (b) $\sum_{i=1}^{n} x_{(k-1)i} x_{ki} = 0$.

Summarizing all conditions, then, $|\mathbf{X}'\mathbf{X}|$ is maximized when

(a) all diagonal elements ($\sum_{i=1}^{n} x_{ji}^2$, j=1,2, ..., k) are maximized

and (b) all columns of X are orthogonal to each other such that off-diagonal elements are zero.

Now consider the Bayes D-optimal augmentation procedure for the variance model. Recall that the objective is to maximize

$$D_j = | c_j \mathbf{Z}^{(j)'} \mathbf{V} \mathbf{Z}^{(j)} |$$

or equivalently

$$D_j^* = | \mathbf{Z}^{(j)'} \mathbf{V} \mathbf{Z}^{(j)} |$$

for each candidate model, M_j, with some models weighted more heavily than others. Also recall that $\mathbf{Z}^{(j)}$ is the first order model matrix for model M_j and V is a d×d diagonal matrix with diagonal elements $\nu_1, \nu_2, ..., \nu_d$. Notice that $\mathbf{Z}^{(j)'} \mathbf{V} \mathbf{Z}^{(j)}$ can be rewritten as

$$P^{(j)'} P^{(j)} = (\mathbf{Z}^{(j)'} \mathbf{V}^{1/2}) (\mathbf{V}^{1/2} \mathbf{Z}^{(j)}).$$

Therefore, $|P^{(j)'}P^{(j)}|$ is maximized by an experimental design which causes

(a) the diagonal elements of $P^{(j)'}P^{(j)}$ to be maximized

and (b) the columns of $P^{(j)}$ to be orthogonal to each other.

The structure of $P^{(j)}$ is given in (7.4.1), with x_{ij} representing the i^{th} control variable at the j^{th} condition of the resolution III fraction (of size d). (It is assumed that there are k variables included in model M_j.)

$$\begin{bmatrix} \sqrt{\nu_1} & x_{11}\sqrt{\nu_1} & x_{21}\sqrt{\nu_1} & \cdots & x_{k1}\sqrt{\nu_1} \\ \sqrt{\nu_2} & x_{12}\sqrt{\nu_2} & x_{22}\sqrt{\nu_2} & \cdots & x_{k2}\sqrt{\nu_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sqrt{\nu_d} & x_{1d}\sqrt{\nu_d} & x_{2d}\sqrt{\nu_d} & \cdots & x_{kd}\sqrt{\nu_d} \end{bmatrix} \qquad (7.4.1)$$

From (7.4.1), it is obvious that the product of any two columns s and t ($s \neq t$) will be equal to $\sum_{i=1}^{d} x_{si}x_{ti}\nu_i$. But $\sum_{i=1}^{d} x_{si}x_{ti}\nu_i = 0$ (and thus condition (b) above is satisfied) only if $\nu_i = \nu$, i=1,2, ...d, since $\sum_{i=1}^{d} x_{si}x_{ti} = 0$, i=1,2, ...d. This implies that all d points in the resolution III fraction should be replicated uniformly. Again observing the structure of $P^{(j)}$, it is easy to see that the jth diagonal of $P^{(j)'}P^{(j)}$ is given by

$$\sum_{i=1}^{d} x_{ji}^2 \nu_i = \sum_{i=1}^{d} \nu_i = N\text{-}d$$

where N is the overall design size. Therefore, once the experimenter has determined the design size that he/she can afford, the diagonals of $P^{(j)'}P^{(j)}$ are fixed and thus condition (a) is automatically satisfied.

Therefore, it has been proven that for an arbitrary candidate model M_j, the D-optimal design augmentation is to uniformly replicate the resolution III fraction N/d times. This implies that this same augmentation strategy is D-optimal for all (m+1) candidate models, making any weighting of the models irrelevant.

Chapter 8

Summary

The primary goal of this research was to develop and study Bayesian two stage design optimality procedures for the normal linear model which work well under model (regressor) uncertainty. The performance of these procedures was evaluated relative to more traditional single stage designs. Additional research has shown that the concept of Bayesian two stage design can be extended to applications in variance modeling.

In chapter 5, six Bayesian two stage design optimality procedures were developed. The objective of each procedure is to find optimal designs with reduced dependence on regressor specification. This is achieved by selecting a first stage design which is robust to regressor misspecification. Data from the first stage design provides model information, enabling the second stage design to be chosen efficiently with reduced model uncertainty.

Chapter 6 contains the results of an evaluation of the performance of these procedures relative to more traditional single stage designs. These results show that under model misspecification, more efficient designs can be obtained by using the Bayesian two stage design procedures.

Chapter 7 presents a Bayesian approach to design augmentation for efficiently collecting data for variance modeling. Data from a base design (typically used for initially modeling the mean of a response) is used to gain information about the variance model. Under reduced variance model uncertainty, the Bayesian design procedure selects the most efficient design augmentation.

Bibliography

Bibliography

Abdelbasit, K. M. and Plackett, R. L. (1983), "Experimental Design for Binary Data," *Journal of the American Statistical Association,* **78**, 90-98.

Atkinson, A. C. (1982), "Developments in the Design of Experiments," *International Statistical Review*, **50**, 161-177.

Bartlett, M. S., and Kendall, D. G. (1946), "The Statistical Analysis of Variance Heterogeneity and the Logarithmic Transformation," *Journal of the Royal Statistical Society, Series B,* **8**, 128-150.

Box, G. E. P. and Meyer, R. D. (1993), "Finding the Active Factors in Fractionated Screening Experiments," *Journal of Quality Technology*, **25**, 94-105.

Box, G. E. P. and Tiao, G. C. (1973), *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, Massachusetts.

Chaloner, K. (1984), "Optimal Bayesian Experimental Design for Linear Models," Annals of Statistics, **12**, 283-300.

Chaloner, K. and Larntz, K. (1989), "Optimal Bayesian Design Applied to Logistic Regression Experiments," *Journal of Statistical Planning and Inference,* **21**, 191-208.

Chaloner, K. and Verdinelli, I. (1995), "Bayesian Experimental Design: A Review," University of Minnesota technical report.

Draper, N. R. and Hunter, W. G. (1966), "Design of Experiments for Parameter Estimation in Multiresponse Situations", *Biometrika,* **53**, 525-533.

Dumouchel, W and Jones, B. (1994), "A Simple Bayesian Modification of D-Optimal Designs to Reduce Dependence on an Assumed Model," *Technometrics,* **36**, 37-47.

Kiefer, J. (1959), "Optimum Experimental Designs," *Journal of the Royal Statistical Society, Series B,* **21**, 272-304.

Kiefer, J. (1961), "Optimum Designs in Regression Problems," *Annals of Mathematical Statistics*, **32**, 298-325.

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

Letsinger, W. C. (1995) "Optimal One and Two-stage Designs for the Logistic Regression Model, *VPI & SU Dissertation.*

Lindley, D. V. (1972), "Bayesian Statistics - A Review," *SIAM*, Philadelphia.

Minkin, S. (1987), "Optimal Designs for Binary Data," *Journal of the American Statistical Association*, **82**, 1098-1103.

Myers, R. H. and Montgomery, D. C. (1995), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, New York: John Wiley & Sons, Inc.

Myers, W. R., Myers, R. H., Carter, W. H., Jr. and White, K. L, Jr. (1996), "Two Stage Designs for the Logistic Regression Model in Single Agent Bioassays," *Journal of Biopharmaceutical Statistics*, April issue.

Shannon, C. E. (1948), "A Mathematical Theory of Communication," *Bell System Tech. J.*, **27**, 379-423 and 623-656.

Silvey, S. D., (1980), *Optimal Designs*, London: Chapman & Hall.

Verdinelli, I. (1992), "Advances in Bayesian Experimental Design," *Bayesian Statistics*, 467-481, Oxford University Press, Oxford.

Vining, G. G., and Schaub, D. (1996), "Experimental Designs for Estimating Both Mean and Variance Functions", Journal of Quality Technology, **28**, 135-147.

Appendix

Appendix A

Under the assumption that the errors are normal around the mean model, i.e.,

$$y_{ij} = \mathbf{x}_i'\beta + \varepsilon_{ij}\sigma_i, \quad i = 1, 2,..., d; \; j = 1, 2,..., r_i$$

where $\varepsilon_{ij} \quad N(0,1)$, $\ln \sigma_i^2 = \mathbf{z}_i'\gamma$ and r_i is the number of observations collected at the i^{th} replicated design point, then

$$s_i^2 = \frac{\sum_{j=1}^{r} (y_{ij} - \bar{y}_{i.})^2}{r_i - 1}$$

is distributed as

$$\left(\frac{\sigma_i^2}{r_i - 1}\right)\chi^2_{r_i - 1} \; . \qquad \text{(A1)}$$

It follows from (A1) that

$$E(s_i^2) = \sigma_i^2$$

and

$$Var(s_i^2) = 2(r_i - 1)\left(\frac{\sigma_i^2}{r_i - 1}\right)^2 = 2\left(\frac{\sigma_i^4}{r_i - 1}\right).$$

Note that $(\ln s_i^2)$ can be written as the following Taylor series expansion

$$\ln s_i^2 = \ln E(s_i^2) + \left.\frac{\partial \ln s_i^2}{\partial s_i^2}\right|_{E(s_i^2)} \left(s_i^2 - E\left(s_i^2\right)\right)$$

$$= \ln \sigma_i^2 + (s_i^2 - \sigma_i^2)/ \sigma_i^2 \qquad \text{(A2)}$$

Rearranging the terms in (A2) and then squaring both sides of the equation results in

$$(\ln s_i^2 - \ln \sigma_i^2)^2 = (s_i^2 - \sigma_i^2)^2 / \sigma_i^4. \qquad \text{(A3)}$$

Taking expectation of both sides of (A3) with respect to s_i^2 gives the approximate variance of (ln s_i^2) as follows;

$$\text{Var } (\ln s_i^2) \quad E(\ln s_i^2 - \ln \sigma_i^2)^2 = \text{Var } (s_i^2)/ \sigma_i^4 ,$$

therefore

$$\text{Var } (\ln s_i^2) \quad 2/(r_i - 1), \quad i = 1, 2 ,..., d.$$

Notice that this variance is constant for all s_i^2 when the same number of replicates is obtained for each design point, in other words, when $r_i = r$ for i =1, 2,..., d.

It should be pointed out that the log χ^2 distribution can be approximated by the Normal distribution, and that this normal approximation is often used in the literature (Bartlett and Kendall (1946)). For r <5, however, the approximation does not appear to be very good. For this reason, the normal approximation is not used in this paper since it is often not feasible for an experimenter to collect five or more observations at each design point.

Appendix B

Let

$$w_i = \frac{\nu_i s_i^2}{\sigma_i^2}\ , i = 1, 2, ..., d$$

where s_i^2 is the sample variance of ν_i+1 observations collected at the i^{th} design point. Recall from Appendix A that $w_i \quad \chi^2_{\nu_i}$. The density function of w_i is given by

$$g(w_i) = \frac{\left(\frac{1}{2}\right)^{\nu_i/2}}{\Gamma\left(\frac{\nu_i}{2}\right)} w_i^{\frac{\nu_i}{2}-1} \exp\left\{-\frac{w_i}{2}\right\}, \quad w_i > 0 .$$

The change of variable formula is used to find the density of $s_i^2 = \frac{\sigma_i^2 w_i}{\nu_i}$ as follows:

$$f(s_i^2) = \frac{\partial w_i}{\partial s_i^2} g(w_i)$$

$$= \frac{\nu_i}{\sigma_i^2} g\left(\frac{\nu_i s_i^2}{\sigma_i^2}\right),$$

therefore,

$$f(s_i^2) = \frac{\left(\frac{1}{2}\right)^{\nu_i/2}}{\Gamma\left(\frac{\nu_i}{2}\right)} \left(\frac{\nu_i}{\sigma_i^2}\right)\left(\frac{\nu_i s_i^2}{\sigma_i^2}\right)^{\frac{\nu_i}{2}-1} \exp\left\{-\frac{\nu_i s_i^2}{2\sigma_i^2}\right\}$$

$$= c_i \left(\frac{\nu_i}{\sigma_i^2}\right)^{\nu_i/2} \left(s_i^2\right)^{\frac{\nu_i}{2}-1} \exp\left\{-\frac{\nu_i s_i^2}{2\sigma_i^2}\right\}, \quad s_i^2 > 0$$

where c_i is a constant which is dependent on ν_i. Since s_1^2, s_2^2, ..., s_d^2 are independent of one another, their joint density, or likelihood is given by

$$L(s_1^2, s_2^2, \ldots, s_d^2 | \gamma) = \prod_{i=1}^{d} c_i \prod_{i=1}^{d} \left(\frac{\nu_i}{\sigma_i^2}\right)^{\nu_i/2} \prod_{i=1}^{d} \left(s_i^2\right)^{\frac{\nu_i}{2}-1} \exp\left\{-\frac{1}{2}\sum_{i=1}^{d} \frac{\nu_i s_i^2}{\sigma_i^2}\right\}$$

$$= \prod_{i=1}^{d} c_i \prod_{i=1}^{d} \left(\frac{\nu_i}{\exp\{\mathbf{z}_i' \gamma\}}\right)^{\nu_i/2} \prod_{i=1}^{d} \left(s_i^2\right)^{\frac{\nu_i}{2}-1} \exp\left\{-\frac{1}{2}\sum_{i=1}^{d} \frac{\nu_i s_i^2}{\exp\{\mathbf{z}_i' \gamma\}}\right\}.$$

Appendix C

The log-likelihood is given by

$$L\left(s_1^2, s_2^2, \ldots, s_d^2 \middle| \gamma\right) \qquad \sum_{i=1}^{d} \frac{\nu_i}{2}\left(\ln \nu_i - \mathbf{z}_i'\gamma\right) + \sum_{i=1}^{2}\left(\frac{\nu_i}{2} - 1\right)\ln s_i^2 - \frac{1}{2}\sum_{i=1}^{d} \frac{\nu_i s_i^2}{\exp\left(\mathbf{z}_i'\gamma\right)}. \qquad \text{(C1)}$$

The Fisher Information matrix is derived from (C1) as

$$\mathbf{I}(\gamma) = -E^{s^2}\left[\frac{\partial}{\partial \gamma'}\left[\frac{\partial L}{\partial \gamma}\right]\right].$$

First, taking the derivative of L with respect to γ gives

$$\frac{\partial L}{\partial \gamma} = -\sum_{i=1}^{d} \frac{\nu_i}{2}\mathbf{z}_i + \frac{1}{2}\sum_{i=1}^{d} \frac{\nu_i s_i^2 \exp\left\{\mathbf{z}_i'\gamma\right\}\mathbf{z}_i}{\left(\exp\left\{\mathbf{z}_i'\gamma\right\}\right)^2}.$$

Then taking the derivative with respect to γ' results in

$$\frac{\partial}{\partial \gamma'}\left[\frac{\partial L}{\partial \gamma}\right] = \frac{1}{2}\sum_{i=1}^{d}\left[\frac{\nu_i s_i^2 \mathbf{z}_i\mathbf{z}_i'}{\exp\left\{\mathbf{z}_i'\gamma\right\}} - \frac{2\nu_i s_i^2 \mathbf{z}_i\mathbf{z}_i'}{\left(\exp\left\{\mathbf{z}_i'\gamma\right\}\right)^2}\right].$$

Since $E(s_i^2) = \exp\{z_i'\gamma\}$,

$$\mathbf{I}(\gamma) = \sum_{i=1}^{d} \left[\frac{\nu_i \mathbf{z}_i \mathbf{z}_i'}{\exp\{\mathbf{z}_i'\gamma\}} - \frac{1}{2}\nu_i \mathbf{z}_i \mathbf{z}_i' \right].$$

Appendix D

This computer algorithm, written in SAS Proc IML, determines the first stage Bayes D-optimal design using the method of DuMouchel and Jones (1994) as described in section 5.1.

This algorithm is written for an application in which the full model contains four primary terms {intercept, x_1, x_2, x_1x_2} and five potential terms {x_3, x_1x_3, x_2x_3, x_1^2, x_2^2}. The code which is unique to this application is given in italics.

The output of this program consists of a set of Box and Meyer posterior probabilities for the candidate models and the resulting second stage design. The design levels are reported with ±1 scaling.

```
/*****************************************************************
 **    Generate the candidate list....                            **
 **     Note:  It is not necessary to use PROC PLAN to generate the candidate    **
 **            list, but the candidate list must be supplied and placed in data set    **
 **          ' CANDID '.                                          **
 *****************************************************************/

 PROC PLAN;
    FACTORS X1=5 ORDERED X2=5 ORDERED X3=5 ORDERED/NOPRINT;
    OUTPUT  OUT=CANDID   X1 NVALS= (-1, -.5, 0, .5, 1)
                         X2 NVALS= (-1, -.5, 0, .5, 1)
                         X3 NVALS= (-1, -.5, 0, .5, 1);
 RUN;
```

```
/*************************************************************
 **                    Convert the Candidate list to model format          **
 *************************************************************/
  DATA CANDID; SET CANDID;
       INTRCPT = 1.0;
       X1X2 = X1*X2;
       X1X3 = X1*X3;
       X2X3 = X2*X3;
       X1X1 = X1*X1;
       X2X2 = X2*X2;


PROC IML;


      TAU = 5;   /*******************************************************
                 **     Note:  TAU=5 is the recommended value, but can be        **
                 **     changed to better reflect the user's beliefs                 **
                 *******************************************************/

      /**  Specify the number of points to be included in the first stage design  **/
       N1 = ????;


  /*************************************************************************
   **     Read the list of candidate design points into IML and scale as                **
   **     recommended by DuMouchel and Jones.                                           **
   **     Note: Begin by reading the p columns of primary terms into the matrix       **
   **     PRI and the q columns of potential terms into the matrix POT.                 **
   **     The rescaled list will be placed in matrix 'CANDID1'.                         **
   *************************************************************************/
    USE CANDID;

    READ ALL VAR{INTRCPT X1 X2 X1X2} INTO PRI;
      READ ALL VAR{X3 X1X3 X2X3 X1X1 X2X2} INTO POT;
      BETA = INV(PRI`*PRI)*PRI`*POT;
      RESID = POT - PRI*BETA;
      COLMIN = RESID[><,];
      COLMAX = RESID[<>,];
       DIFF = COLMAX - COLMIN;
```

```
        DIFFMAT = DIAG(DIFF);
        INVDIFF = INV(DIFFMAT);
        ZMAT = RESID*INVDIFF;
        SCALED = ROW||PRI||ZMAT;
        CANDID1 = PRI||ZMAT;
    P=NCOL(PRI);
   Q=NCOL(POT);
   GAMMA = ( J(P,P,0)||J(P,Q,0) )//( J(Q,P,0)||((TAU**-2.0)*I(Q)) );
   DELTA = (.001/125)*CANDID1`*CANDID1 ;
   NCAND = NROW(CANDID1);
   RIDGE = GAMMA + DELTA;

/*********************    Begin loss module     ***************************/
   START LOSS(DOX) GLOBAL(RIDGE);
     LOSS = 1/DET(DOX`*DOX + RIDGE);
     RETURN(LOSS);
   FINISH LOSS;
 /*********************    End loss module     ***************************/


 DO COUNT = 1 TO 10;    /***  Take best of 10 designs  ***/

 /*******          Randomly select the initial n1 design pts          ******/
 J=N1-1;
 SEED = UNIFORM(0:J);
 DEVIATE = UNIFORM(SEED);
 INDEX = ROUND(NCAND*DEVIATE);
 INDEX2 = INDEX + (INDEX=0);
 BESTD1 = CANDID1[INDEX2,];
 BESTDOLD = LOSS(BESTD1);


  /**** Initializing  values for the very first exchange ****/

  BESTD = BESTDOLD;
  DHOLD = {100000 100000};
  BESTD1N1 = BESTD1;
  LOSSVEC = J(NCAND,1,0);     /*** Initializing a loss vector ***/
  ND1 = NROW(BESTD1);
```

```
/**********                Exchange                **********/

DO WHILE ( (ND1 > 0) & (ABS(N1-NROW(BESTD1) ) < 8 ) );

/*****          What to do when returned to an n-point design          ******/

IF ND1=N1 THEN DO;
   IF BESTDOLD <= BESTD THEN FAILEDD = DHOLD;
   IF BESTDOLD > BESTD THEN DO;
    BESTD1N1 = BESTD1;
    BESTDOLD = BESTD;
    DHOLD = REMOVE(DHOLD,1:NCOL(DHOLD));  /**   set failedd and dhold **
                                          **   to null                **/
    FAILEDD = {100000};
   END;


 DEVIATE = RANUNI(0);

 IF DEVIATE < .50000 THEN DO;     /**** remove worst point   ****/

   BESTD1 = BESTD1N1[2:N1,];       /*** initialize best design and best q  **/
   BESTD =  LOSS(BESTD1) ;

   D1 = BESTD1//BESTD1N1[1,];
   DO I= 2 TO N1;
      NEWD1 = D1[2:N1,];
      DNEW  = LOSS(NEWD1);
      IF DNEW<BESTD THEN DO;
        BESTD = DNEW;
        BESTD1= NEWD1;
      END;
      D1=NEWD1//BESTD1N1[I,];
   END;
 END;

 IF DEVIATE >= .50000 THEN DO;
```

```
      DO C=1 TO NCAND;

        CDIDROW = CANDID1[C,];
        LOSSVEC[C] = LOSS(BESTD1N1//CANDID1[C,] ) ;

      END;
      MINLOC = LOSSVEC[>:<];
      BESTD1 = BESTD1N1//CANDID1[MINLOC,];
      BESTD = LOSSVEC[MINLOC,];

  END;
   DHOLD=DHOLD||BESTD;

END;     /*******  End block for n-point design    ********/



  /*********     What to do if design size < N1    *********/
  ND1 = NROW(BESTD1);
  IF ND1 < N1 THEN DO;
    IF ( ANY(BESTD = FAILEDD) ) | (SUM(BESTD=DHOLD)>1)  THEN DO;

    IF ND1=1 THEN DO;
      ND1=0;
      GOTO FINAL;
    END;

    D1 = BESTD1;
    BESTD1 = BESTD1[2:ND1,];     /*** Initialize best design and best d  **/
    BESTD =  LOSS(BESTD1);


    D1 = BESTD1//D1[1,];

    DO I= 2 TO ND1;

       NEWD1 = D1[2:ND1,];
       DNEW  = LOSS(NEWD1) ;

       IF DNEW<BESTD THEN DO;
```

```
        BESTD = DNEW;
        BESTD1=NEWD1;
      END;
      D1=NEWD1//D1[1,];
  END;
 ND1 = NROW(BESTD1);
 GOTO FINAL1;
 END;
 IF (ALL(BESTD ^= FAILEDD)) & (SUM(BESTD=DHOLD)=1) THEN DO;
     DO C=1 TO NCAND;
      LOSSVEC[C] = LOSS(BESTD1//CANDID1[C,] ) ;
     END;

     MINLOC = LOSSVEC[>:<];
     BESTD1 = BESTD1//CANDID1[MINLOC,];
     BESTD = LOSSVEC[MINLOC,];
 END;
 ND1 = NROW(BESTD1);
 FINAL1: DHOLD = DHOLD||BESTD;

END; /*** End block for design size < N1  ****/


 /******  What to do if design size > N1  *******/
ND1 = NROW(BESTD1);

IF ND1 > N1 THEN DO;

 IF ( ALL(BESTD ^= FAILEDD) ) & (SUM(BESTD=DHOLD)=1) THEN DO;
   D1 = BESTD1;
   BESTD1 = BESTD1[2:ND1,];        /*** initialize best design and best q  **/
   BESTD = LOSS(BESTD1);

   D1 = BESTD1//D1[1,];

   DO I= 2 TO ND1;

      NEWD1 = D1[2:ND1,];
      DNEW  = LOSS(NEWD1);
```

```
      IF DNEW<BESTD THEN DO;
        BESTD = DNEW;
        BESTD1=NEWD1;
      END;
      D1=NEWD1//D1[1,];
   END;
  ND1 = NROW(BESTD1);
  GOTO FINAL2;
  END;

 IF ( ANY(BESTD = FAILEDD)) | (SUM(BESTD=DHOLD)>1) THEN DO;
     DO C=1 TO NCAND;
       LOSSVEC[C] = LOSS(BESTD1//CANDID1[C,] ) ;
     END;
     MINLOC = LOSSVEC[>:<];
     BESTD1 = BESTD1//CANDID1[MINLOC,];
     BESTD = LOSSVEC[MINLOC,];
  END;
  ND1 = NROW(BESTD1);
  FINAL2:  DHOLD = DHOLD||BESTD;

 END;

FINAL:  END;     /**** End of outside do-loop *****/


 DOXMAT = DOXMAT//BESTD1N1;
 DVEC = DVEC//BESTDOLD;

 END;
 PRINT DVEC;
 LOC = DVEC[>:<];
 BEGIN = (LOC-1)*N1 + 1;
 END = LOC*N1;
 BESTD1 = DOXMAT[BEGIN:END,{2 3 5}];  /*****************************
                                       **  Design variables x1, x2 and x3  **
                                        **  are in columns 2, 3 and 5        **
                                         **  respectively                     **
                                          *****************************/
```

```
CREATE DESIGN FROM BESTD1 [COLNAME={"X1'  'X2'  'X3'}];
    APPEND FROM BESTD1;

QUIT;

DATA DESIGN; SET DESIGN;
X3=2*X3;                    /**  Rescale the design to +/-1 format  **/
PROC PRINT DATA=DESIGN;
```

Appendix E

This computer algorithm, written in SAS Proc IML, determines the first stage Bayes Q-optimal design using the method of DuMouchel and Jones (1994) as described in section 5.1.

This algorithm is written for an application in which the full model contains four primary terms {intercept, x_1, x_2, x_1x_2} and five potential terms {x_3, x_1x_3, x_2x_3, x_1^2, x_2^2}. The code which is unique to this application is given in italics. The user must supply the *region moment matrix* of the full model, so that the integrated prediction variance (IPV) can be calculated for each candidate design.

The output of this program consists of a set of Box and Meyer posterior probabilities for the 2^q candidate models and the resulting second stage design. The design levels are reported with ± 1 scaling.

```
/*****************************************************************
 **    Generate the candidate list....                            **
 **     Note:  It is not necessary to use PROC PLAN to generate the candidate    **
 **              list, but the candidate list must be supplied and placed in data set    **
 **           ' CANDID '.                                         **
 *****************************************************************/

  PROC PLAN;
     FACTORS X1=5 ORDERED X2=5 ORDERED X3=5 ORDERED/NOPRINT;
      OUTPUT  OUT=CANDID   X1 NVALS= (-1, -.5, 0, .5, 1)
                                X2 NVALS= (-1, -.5, 0, .5, 1)
                                X3 NVALS= (-1, -.5, 0, .5, 1);
RUN;



/*****************************************************************
 **                  Convert the Candidate list to model format                **
 *****************************************************************/
DATA CANDID; SET CANDID;
     INTRCPT = 1.0;
     X1X2 = X1*X2;
     X1X3 = X1*X3;
     X2X3 = X2*X3;
     X1X1 = X1*X1;
     X2X2 = X2*X2;


PROC IML;


     TAU = 5;   /*****************************************************
                **    Note:  TAU=5 is the recommended value, but can be        **
                **    changed to better reflect the user's beliefs             **
                *****************************************************/

     /**  Specify the number of points to be included in the first stage design  **/
      N1 = ????;
```

```
/*****************************************************************
 **     Read the list of candidate design points into IML and scale as          **
 **     recommended by DuMouchel and Jones.                                     **
 **     Note: Begin by reading the p columns of primary terms into the matrix   **
 **     PRI and the q columns of potential terms into the matrix POT.           **
 **     The rescaled list will be placed in matrix 'CANDID1'.                   **
 *****************************************************************/
 USE CANDID;

 READ ALL VAR{INTRCPT X1 X2 X1X2} INTO PRI;
   READ ALL VAR{X3 X1X3 X2X3 X1X1 X2X2} INTO POT;
   BETA = INV(PRI`*PRI)*PRI`*POT;
   RESID = POT - PRI*BETA;
   COLMIN = RESID[><,];
   COLMAX = RESID[<>,];
    DIFF = COLMAX - COLMIN;
    DIFFMAT = DIAG(DIFF);
   INVDIFF = INV(DIFFMAT);
   ZMAT = RESID*INVDIFF;
   SCALED = ROW||PRI||ZMAT;
   CANDID1 = PRI||ZMAT;

/*****************************************************************
 **     Enter MSTAR, the Region Moment Matrix of the full model....             **
 **      Note:  This matrix must be in the same order as X'X for the full  model,  **
 **             where X = [XPRI | XPOT].                                         **
 *****************************************************************/
 MSTAR = { 1     0     0     0     0     0     0    .333  .333,
           0    .333   0     0     0     0     0     0     0   ,
           0     0    .333   0     0     0     0     0     0   ,
           0     0     0    .111   0     0     0     0     0   ,
           0     0     0     0    .333   0     0     0     0   ,
           0     0     0     0     0    .111   0     0     0   ,
           0     0     0     0     0     0    .111   0     0   ,
          .333   0     0     0     0     0     0    .20   .111,
          .333   0     0     0     0     0     0    .111  .20};


 P=NCOL(PRI);
Q=NCOL(POT);
```

```
   GAMMA = ( J(P,P,0)||J(P,Q,0) )//( J(Q,P,0)||((TAU**-2.0)*I(Q)) );
   DELTA = (.001/125)*CANDID1`*CANDID1 ;
   NCAND = NROW(CANDID1);
   RIDGE = GAMMA + DELTA;

/*********************  Begin loss module  ***************************/
   START LOSS(DOX) GLOBAL(RIDGE);
     LOSS = TRACE(INV(DOX`*DOX + RIDGE)*MSTAR);
     RETURN(LOSS);
   FINISH LOSS;
/*********************  End loss module  ***************************/


 DO COUNT = 1 TO 10;   /***  Take best of 10 designs  ***/

 /*******          Randomly select the initial n1 design pts          ******/
 J=N1-1;
 SEED = UNIFORM(0:J);
 DEVIATE = UNIFORM(SEED);
 INDEX = ROUND(NCAND*DEVIATE);
 INDEX2 = INDEX + (INDEX=0);
 BESTD1 = CANDID1[INDEX2,];
 BESTDOLD = LOSS(BESTD1);


  /**** Initializing  values for the very first exchange ****/

  BESTD = BESTDOLD;
  DHOLD = {100000 100000};
  BESTD1N1 = BESTD1;
  LOSSVEC = J(NCAND,1,0);    /*** Initializing a loss vector ***/
  ND1 = NROW(BESTD1);



 /**********                    Exchange                    **********/
 DO WHILE ( (ND1 > 0) & (ABS(N1-NROW(BESTD1) ) < 8 ) );


/*****              What to do when returned to an n-point design              ******/
```

```
IF ND1=N1 THEN DO;
    IF BESTDOLD <= BESTD THEN FAILEDD = DHOLD;
    IF BESTDOLD > BESTD THEN DO;
      BESTD1N1 = BESTD1;
      BESTDOLD = BESTD;
      DHOLD = REMOVE(DHOLD,1:NCOL(DHOLD));  /**   set failedd and dhold to null **/
      FAILEDD = {100000};
    END;
  DEVIATE = RANUNI(0);

  IF DEVIATE < .50000 THEN DO;      /**** remove worst point   ****/

    BESTD1 = BESTD1N1[2:N1,];       /*** initialize best design and best q  **/
    BESTD =  LOSS(BESTD1) ;

    D1 = BESTD1//BESTD1N1[1,];
    DO I= 2 TO N1;

       NEWD1 = D1[2:N1,];
       DNEW  = LOSS(NEWD1);

       IF DNEW<BESTD THEN DO;
         BESTD = DNEW;
         BESTD1= NEWD1;
       END;
       D1=NEWD1//BESTD1N1[I,];
    END;

  END;

 IF DEVIATE >= .50000 THEN DO;

       DO C=1 TO NCAND;

         CDIDROW = CANDID1[C,];
         LOSSVEC[C] = LOSS(BESTD1N1//CANDID1[C,] ) ;

       END;
```

```
      MINLOC = LOSSVEC[>:<];
      BESTD1 = BESTD1N1//CANDID1[MINLOC,];
      BESTD = LOSSVEC[MINLOC,];

   END;
    DHOLD=DHOLD||BESTD;

END;     /*******  End block for n-point design    ********/
  /*********     What to do if design size < N1    *********/
  ND1 = NROW(BESTD1);
  IF ND1 < N1 THEN DO;
    IF ( ANY(BESTD = FAILEDD) ) | (SUM(BESTD=DHOLD)>1)  THEN DO;

    IF ND1=1 THEN DO;
      ND1=0;
      GOTO FINAL;
    END;

    D1 = BESTD1;
    BESTD1 = BESTD1[2:ND1,];     /*** Initialize best design and best d  **/
    BESTD =  LOSS(BESTD1);
    D1 = BESTD1//D1[1,];

    DO I= 2 TO ND1;

       NEWD1 = D1[2:ND1,];
       DNEW  = LOSS(NEWD1) ;

       IF DNEW<BESTD THEN DO;
         BESTD = DNEW;
         BESTD1=NEWD1;
       END;
       D1=NEWD1//D1[1,];
    END;
   ND1 = NROW(BESTD1);
   GOTO FINAL1;
   END;
```

```
 IF (ALL(BESTD ^= FAILEDD)) & (SUM(BESTD=DHOLD)=1) THEN DO;
    DO C=1 TO NCAND;
     LOSSVEC[C] = LOSS(BESTD1//CANDID1[C,] ) ;
    END;

    MINLOC = LOSSVEC[>:<];
    BESTD1 = BESTD1//CANDID1[MINLOC,];
    BESTD = LOSSVEC[MINLOC,];
 END;
 ND1 = NROW(BESTD1);
 FINAL1: DHOLD = DHOLD||BESTD;

END; /*** End block for design size < N1  ****/
/******  What to do if design size > N1  *******/

ND1 = NROW(BESTD1);
IF ND1 > N1 THEN DO;

 IF ( ALL(BESTD ^= FAILEDD) ) & (SUM(BESTD=DHOLD)=1) THEN DO;
  D1 = BESTD1;
  BESTD1 = BESTD1[2:ND1,];       /*** initialize best design and best q  **/
  BESTD = LOSS(BESTD1);

  D1 = BESTD1//D1[1,];

  DO I= 2 TO ND1;

     NEWD1 = D1[2:ND1,];
     DNEW  = LOSS(NEWD1);

     IF DNEW<BESTD THEN DO;
       BESTD = DNEW;
       BESTD1=NEWD1;
     END;
     D1=NEWD1//D1[1,];
  END;
 ND1 = NROW(BESTD1);
 GOTO FINAL2;
 END;
```

```
   IF ( ANY(BESTD = FAILEDD)) | (SUM(BESTD=DHOLD)>1) THEN DO;
      DO C=1 TO NCAND;
        LOSSVEC[C] = LOSS(BESTD1//CANDID1[C,] ) ;
      END;

      MINLOC = LOSSVEC[>:<];
      BESTD1 = BESTD1//CANDID1[MINLOC,];
      BESTD = LOSSVEC[MINLOC,];
    END;
   ND1 = NROW(BESTD1);
   FINAL2:  DHOLD = DHOLD||BESTD;

  END;

FINAL:  END;     /**** End of outside do-loop *****/
 DOXMAT = DOXMAT//BESTD1N1;
 DVEC = DVEC//BESTDOLD;

 END;
 PRINT DVEC;
 LOC = DVEC[>:<];
 BEGIN = (LOC-1)*N1 + 1;
 END = LOC*N1;
 BESTD1 = DOXMAT[BEGIN:END,{2 3 5}];  /***********************************
                                       **  Design variables x1, x2 and x3      **
                                        **  are in columns 2, 3 and 5           **
                                        **  respectively                        **
                                        ***********************************/


 CREATE DESIGN FROM BESTD1 [COLNAME={"X1'  'X2'  'X3'}];
     APPEND FROM BESTD1;

 QUIT;

 DATA DESIGN; SET DESIGN;
 X3=2*X3;            /**  Rescale the design to +/-1 format  **/
 PROC PRINT DATA=DESIGN;
```

Appendix F

This computer algorithm, written in SAS Proc IML, determines the optimal second stage of a two stage design. Given a first stage design and first stage data, the program finds the Bayes D-optimal design using the criterion described in section 5.2.

The first stage design must be provided by the user in full model form. This algorithm is written for an application in which the full model contains four primary terms {intercept, x_1, x_2, x_1x_2} and five potential terms {x_3, x_1x_3, x_2x_3, x_1^2, x_2^2}. The code which is unique to this application is given in italics.

Additionally, the first stage design should first be read/entered into the program with all terms scaled to ±1. The potential terms must then be rescaled according to the convention of DuMouchel and Jones (1994), which is accomplished in the following manner. Suppose that x_i is a primary term and that { x_j, x_ix_j, x_i^2 } are potential terms, all scaled to ±1. The newly rescaled potential terms are:

$$x_j\_new = .5*x_j$$

$$x_ix_j\_new = x_i*x_j\_new$$

$$x_i^2\_new = x_i^2 - .5$$

The algorithm automatically rescales the list of candidate design points in the same manner.

The output of this program consists of a set of Box and Meyer posterior probabilities for the candidate models and the resulting second stage design. The design levels are reported with ±1 scaling.

```
/*********************************************************************
**    Read (or construct) the first stage data set which contains the following:      **
**         a)  p primary terms (including a column of 1's for the intercept)          **
**         b)  q potential terms                                                      **
**             Note:  All p+q terms should be scaled to +/- 1                         **
**         c)  response (y)                                                           **
*********************************************************************/

*  (example data set)  *
   DATA DESIGN;
      INPUT X1 X2 X3 Y;
      INTRCPT = 1.0;
      X1X3 = X1*X3;
      X2X3 = X2*X3;
      X1X1 = X1*X1;
      X2X2 = X2*X2;
  LINES;
     .
     .
     .
   ;
  RUN;



/*********************************************************************
 **   Rescale the potential terms according to the scaling convention of DuMouchel   **
 **   and Jones  (1994).                                                             **
 *********************************************************************/
  DATA DESIGN; SET DESIGN;
  X3 = X3/2;
  X1X3 = X1*X3;
  X2X3 = X2*X2;
  X1X1 = X1*X1 - .5;
  X2X2 = X2*X2 - .5;
  RUN;
```

```
/*****************************************************************
 **    Generate the candidate list....                            **
 **     Note:  It is not necessary to use PROC PLAN to generate the candidate    **
 **              list, but the candidate list must be supplied and placed in data set    **
 **           ‘ CANDID ’.                                          **
 *****************************************************************/

  PROC PLAN;
     FACTORS X1=5 ORDERED X2=5 ORDERED X3=5 ORDERED/NOPRINT;
      OUTPUT  OUT=CANDID   X1 NVALS= (-1, -.5, 0, .5, 1)
                                  X2 NVALS= (-1, -.5, 0, .5, 1)
                                  X3 NVALS= (-1, -.5, 0, .5, 1);
 RUN;



/*****************************************************************
**                    Convert the Candidate list to model format              **
*****************************************************************/
 DATA CANDID; SET CANDID;
      INTRCPT = 1.0;
      X1X2 = X1*X2;
      X1X3 = X1*X3;
      X2X3 = X2*X3;
      X1X1 = X1*X1;
      X2X2 = X2*X2;


PROC IML;

  USE DESIGN;

     /*********************************************************
      **       Reading from the first stage data set, named ‘DESIGN’,          **
      **       read the p columns of primary terms into the matrix XPRI.       **
      *********************************************************/
      READ ALL VAR{INTRCPT X1 X2 X1X2 }INTO XPRI;
```

```
  /*************************************************************
   **       Reading from the first stage data set, named 'DESIGN',          **
   **       read the q columns of potential terms into the matrix XPOT.    **
   *************************************************************/
   READ ALL VAR{X3 X1X3 X2X3 X1X1 X2X2} INTO XPOT;
 /**       Construct the combined model matrix of the form [XPRI | XPOT]   **/
   D1 = XPRI||XPOT;
   P = NCOL(XPRI);
   Q = NCOL(XPOT);

   TAU = 5;   /*********************************************************
                **     Note:  TAU=5 is the recommended value, but can be          **
                **     changed to better reflect the user's beliefs                **
                *********************************************************/

   /**  Specify the number of points to be included in the second stage design  **/
   N2 = ????;

/*********************************************************************
 **     Read the list of candidate design points into IML and scale as              **
 **     recommended by DuMouchel and Jones.                                         **
 **     Note: Begin by reading the p columns of primary terms into the matrix       **
 **     PRI and the q columns of potential terms into the matrix POT.               **
 **     The rescaled list will be placed in matrix 'CANDID1'.                       **
 *********************************************************************/
  USE CANDID;

    READ ALL VAR{INTRCPT X1 X2 X1X2} INTO PRI;
    READ ALL VAR{X3 X1X3 X2X3 X1X1 X2X2} INTO POT;
    BETA = INV(PRI`*PRI)*PRI`*POT;
    RESID = POT - PRI*BETA;
    COLMIN = RESID[><,];
    COLMAX = RESID[<>,];
     DIFF = COLMAX - COLMIN;
     DIFFMAT = DIAG(DIFF);
     INVDIFF = INV(DIFFMAT);
     ZMAT = RESID*INVDIFF;
    SCALED = ROW||PRI||ZMAT;
     CANDID1 = PRI||ZMAT;
```

```
GAMMA = ( J(P,P,0)||J(P,Q,0) )//( J(Q,P,0)||((TAU**-2.0)*I(Q)) );

/*************************************************************
**       Reading from the first stage data set, named 'DESIGN',          **
**       read the response data into the vector Y .                      **
*************************************************************/
USE DESIGN;   READ ALL VAR {Y} INTO Y;
/*************************************************************
**    Using first stage data, calculate the Bayes estimates and their standard  **
**    errors (apart from sigma**2).                                              **
*************************************************************/
BETAEST1 = INV(D1`*D1 + GAMMA) *D1`*Y;
BETASE1 = SQRT(DIAG(INV(D1`*D1+ GAMMA)));

/*******   Calculate the vector of standardized beta estimates   *******/
STDBETA1 = (INV(BETASE1))*BETAEST1;
BETASE1 = VECDIAG(BETASE1);

/*******    Calculate the updated tau-squared's ,  unscaled       *******/
 NUMPARMS = NROW(STDBETA1);
 SUM = SUM(ABS(STDBETA1));
 TUNSCLD = STDBET1A`*(I(NUMPARMS) /SUM);

/*************************************************************
**       Scale the tau's, by setting the max of unscaled tau-squared's          **
**       equal to 100 and scaling the others relatively.                        **
*************************************************************/
 TMAT = DIAG(TUNSCLD);
 MAXT = MAX(TMAT);
 TMATSCLD = 100*ABS(TMAT)/MAXT;

/*************************************************************
**   Check for singularity of TMATSCLD, adding shrinkage parameter      **
**   if necessary.                                                      **
*************************************************************/
 DO WHILE (DET(TMATSCLD) = 0 );
    MINLOC = (VECDIAG(TMATSCLD))[>:<];
     TMATSCLD[MINLOC,MINLOC] = TMATSCLD[MINLOC,MINLOC] +
    (RANUNI(0))/100000;
 END;
```

```
   TMATINV = INV(TMATSCLD);
   TSCLD = VECDIAG(TMATSCLD);
   TSCLDINV = VECDIAG(TMATINV);

  /**  Sequentially select the optimal N2 design points for the second stage design  **/
 BASE = D1;  N1 = NROW(D1);
 DO I = 1 TO N2;
   BAYESVAR = INV(BASE`*BASE + TMATINV);
   PVARVEC = VECDIAG(CANDID1*BAYESVAR*CANDID1`);
   MAXLOC = PVARVEC[<:>];
   BASE = BASE||CANDID1[MAXLOC,];
 END;

 BEGIN2 = N1+1; END2 = N1+N2;
 STAGE2 = BASE[BEGIN2:END2, {2 3 5}];   /** x1, x2 and x3 are in columns  **
                                        **  2, 3 and 5 respectively       **/

CREATE DOX FROM STAGE2[COLNAME = { 'X1' 'X2' 'X3'}];
APPEND FROM STAGE2;

QUIT;

/**     Convert the optimum design to +/- 1 scaling  and print the design  **/;
 DATA DOX; SET DOX;
  X3=2*X3;

 PROC PRINT DATA=DOX;
```

Appendix G

This computer algorithm which is written in SAS Proc IML determines the optimal second stage of a two-stage design. Given a first stage design and first stage data, the program calculates Box and Meyer posterior probabilities for all candidate models. These probabilities are then used to find the optimal second stage design using the Bayes D-optimality criterion described in section 5.3.

The first stage design must be provided by the user in full model form. This algorithm is written for an application in which the full model contains four primary terms {intercept, x_1, x_2, x_1x_2} and five potential terms {x_3, x_1x_3, x_2x_3, x_1^2, x_2^2}. The code which is unique to this application is given in italics. This algorithm can easily accommodate other applications with five potential terms through modification of the italicized code. (Additional code would be required to expand the list of candidate models to include those containing more than five potential terms).

Additionally, the first stage design should first be read/entered into the program with all terms scaled to ±1. The potential terms must then be rescaled according to the convention of DuMouchel and Jones (1994), which is accomplished in the following manner. Suppose that x_i is a primary term and that { x_j, x_ix_j, x_i^2 } are potential terms, all scaled to ±1. The newly rescaled potential terms are:

$$x_j\_new = .5*x_j$$

$$x_ix_j\_new = x_i*x_j\_new$$

$$x_i^2\_new = x_i^2 - .5$$

The algorithm automatically rescales the list of candidate design points in the same manner.

The output of this program consists of a set of Box and Meyer posterior probabilities for the candidate models and the resulting second stage design. The design levels are reported with ±1 scaling.

```
/*************************************************************************
**    Read (or construct) the first stage data set which contains the following:      **
**         a)  p primary terms (including a column of 1's for the intercept)          **
**         b)  q potential terms                                                      **
**         c)  response (y)                                                           **
*************************************************************************/

 *  (example data set)  *
    DATA DESIGN;
       INPUT X1 X2 X3 Y;
       INTRCPT = 1.0;
       X1X3 = X1*X3;
       X2X3 = X2*X3;
       X1X1 = X1*X1;
       X2X2 = X2*X2;
   LINES;
       .
       .
       .
    ;
   RUN;


/*************************************************************************
 **   Rescale the potential terms according to the scaling convention of DuMouchel   **
 **   and Jones  (1994).                                                             **
 *************************************************************************/
   DATA DESIGN; SET DESIGN;
   X3 = X3/2;
   X1X3 = X1*X3;
   X2X3 = X2*X2;
   X1X1 = X1*X1 - .5;
   X2X2 = X2*X2 - .5;
   RUN;
```

```
/*****************************************************************
 **    Generate the candidate list....                            **
 **     Note:  It is not necessary to use PROC PLAN to generate the candidate   **
 **            list, but the candidate list must be supplied and placed in data set   **
 **          ‘ CANDID ’.                                          **
 *****************************************************************/

 PROC PLAN;
    FACTORS X1=5 ORDERED X2=5 ORDERED X3=5 ORDERED/NOPRINT;
     OUTPUT  OUT=CANDID   X1 NVALS= (-1, -.5, 0, .5, 1)
                          X2 NVALS= (-1, -.5, 0, .5, 1)
                          X3 NVALS= (-1, -.5, 0, .5, 1);
RUN;


 /*****************************************************************
 **                Convert the Candidate list to model format                **
 *****************************************************************/

  DATA CANDID; SET CANDID;
      INTRCPT = 1.0;
      X1X2 = X1*X2;
      X1X3 = X1*X3;
      X2X3 = X2*X3;
      X1X1 = X1*X1;
      X2X2 = X2*X2;


PROC IML;

  USE DESIGN;

    /*********************************************************
     **      Reading from the first stage data set, named ‘DESIGN’,        **
     **      read the p columns of primary terms into the matrix XPRI.     **
     *********************************************************/
     READ ALL VAR{INTRCPT X1 X2 X1X2 }INTO XPRI;
```

```
/*************************************************************
 **       Reading from the first stage data set, named 'DESIGN',          **
 **       read the q columns of potential terms into the matrix XPOT.     **
 *************************************************************/
 READ ALL VAR{X3 X1X3 X2X3 X1X1 X2X2} INTO XPOT;
/**       Construct the combined model matrix of the form [XPRI | XPOT]   **/
 D1 = XPRI||XPOT;
 P = NCOL(XPRI);
 Q = NCOL(XPOT);

 TAU = 5;   /*********************************************************
              **     Note:  TAU=5 is the recommended value, but can be        **
              **     changed to better reflect the user's beliefs             **
              *********************************************************/

 PI = 1/3;    /*********************************************************
              **     Note:  PI is the prior probability that any potential term      **
              **     is active and therefore should reflect the user's beliefs.     **
              **     (  Recommended values are .25 ≤ PI ≤ .50  )                    **
              *********************************************************/



/*****************************************************************
 **     Read the list of candidate design points into IML and scale as          **
 **     recommended by DuMouchel and Jones.                                     **
 **     Note: Begin by reading the p columns of primary terms into the matrix   **
 **     PRI and the q columns of potential terms into the matrix POT.           **
 **     The rescaled list will be placed in matrix 'CANDID1'.                   **
 *****************************************************************/
 USE CANDID;

  READ ALL VAR{INTRCPT X1 X2 X1X2} INTO PRI;
  READ ALL VAR{X3 X1X3 X2X3 X1X1 X2X2} INTO POT;
  BETA = INV(PRI`*PRI)*PRI`*POT;
  RESID = POT - PRI*BETA;
  COLMIN = RESID[><,];
  COLMAX = RESID[<>,];
   DIFF = COLMAX - COLMIN;
   DIFFMAT = DIAG(DIFF);
```

```
      INVDIFF = INV(DIFFMAT);
      ZMAT = RESID*INVDIFF;
      SCALED = ROW||PRI||ZMAT;
      CANDID1 = PRI||ZMAT;

/**  Specify the number of design points to be included in the second stage design  **/
 N2 = ????;

 /**   List the primary terms, in the order they appear in the first stage data set    **/
 PTERMS = {'B0'  'B1'  'B2'  'B12'};

 /**   List the potential terms, in the order they appear in the first stage data set   **/
 QTERMS = {'B3'  'B13'  'B23'  'B11'  'B22'};

  N = NROW(XPOT);
NN = (N-1)/2;
 GAMMA = ( J(P,P,0)||J(P,Q,0) )//( J(Q,P,0)||((TAU**-2.0)*I(Q)) );

/*************************************************************
**       Reading from the first stage data set, named 'DESIGN',          **
**       read the response data into the vector Y .                      **
*************************************************************/
USE DESIGN;   READ ALL VAR {Y} INTO Y;

 /**********      Begin Box-Meyer module  **********/
 START BOXMEYER;
    B = INV(X`*X + GAMMAI)*X`*Y;
   ESS = SSQ(Y - X*B);
    ONE = ( PI/((1-PI)*TAU) )**QQ;
    TWO = DETPRI/SQRT(DET(GAMMAI + X`*X) );
    THREE = ( (ESS + B`*GAMMAI*B)/ESSPRI)**-NN;
    PP = ONE*TWO*THREE;
    PPROB = PPROB//PP;
 FINISH;
/**********    End  Box-Meyer module   **********/
```

```
/**********     Begin  Lmod1 module      **********/
 START   LMOD1(DOX2)
       GLOBAL(P, Q, N1, NROW2, MSTAR, D2D2, D1D1, GAMMA, PPROB);
    NROW2 = NROW(DOX2);
    D2D2 = DOX2`*DOX2;

    INDEX = 1:P;                              /***  0 potential terms ***/
    LOSS = LMOD2(INDEX);
   LVEC = LVEC//LOSS;
    DO J = 1 TO Q;                           /***  1 potential  term  ***/
      INDEX = (1:P)||(P+J);
      LOSS = LMOD2(INDEX);
     LVEC = LVEC//LOSS;
    END;
    DO J = 1 TO Q;                            /***  2 potential terms  ***/
     K = J;
     DO WHILE ( K < Q);
       K = K+1;
       INDEX = (1:P)||(P+J)||(P+K);
       LOSS = LMOD2(INDEX);
      LVEC = LVEC//LOSS;
     END;
    END;

    DO J = 1 TO Q;                            /***  3 potential terms  ***/
     K = J;
     DO WHILE ( K < Q);
       K = K+1; L=K;
       DO WHILE (L < Q);
         L = L+1;
         INDEX = (1:P)||(P+J)||(P+K)||(P+L);
         LOSS= LMOD2(INDEX);
        LVEC = LVEC//LOSS;
       END;
     END;
    END;
```

```
DO J = 1 TO Q;                          /*** 4 potential terms  ***/
      K = J;
      DO WHILE ( K < Q);
        K = K+1; L=K;
        DO WHILE (L < Q);
          L = L+1; M = L;
          DO WHILE (M < Q);
            M = M+1;
            INDEX = (1:P)||(P+J)||(P+K)||(P+L)||(P+M);
            LOSS = LMOD2(INDEX);
           LVEC = LVEC//LOSS;
          END;
        END;
      END;
   END;

  DO J = 1 TO Q;                            /*** 5 potential terms  ***/
      K = J;
      DO WHILE ( K < Q);
        K = K+1; L=K;
        DO WHILE (L < Q);
          L = L+1; M = L;
          DO WHILE (M < Q);
            M = M+1; N = M;
           DO WHILE (N < Q);
             N = N+1;
             INDEX = (1:P)||(P+J)||(P+K)||(P+L)||(P+M)||(P+N);
             LOSS = LMOD2(INDEX) ;
            LVEC = LVEC//LOSS;
           END;
          END;
        END;
      END;
   END;


   LCOMP = PPROB`*LVEC;   RETURN(LCOMP);
 FINISH LMOD1;
/**********   End  Lmod1 module       **********/
```

```
/*********      Begin Loss module for Bayes D-optimality criterion     ********/
  START LMOD2(I)
      GLOBAL(N1,NROW2, MSTAR, D1D1, D2D2, GAMMA);
      DOXLOSS = DET(  INV(D1D1[I,I] + D2D2[I,I] + GAMMA[I,I])  );
      RETURN(DOXLOSS);
  FINISH LMOD2;
/*********      End Loss module for Bayes D-optimality criterion     ********/


/**       Calculate  Box and Meyer posterior probabilities for all candidate models       **/

  /*************  Model with 0 potential terms  *************/
   B=INV(XPRI`*XPRI)*XPRI`*Y;
  ESSPRI = SSQ(Y-XPRI*B);  DETPRI = SQRT(DET(XPRI`*XPRI));
  PP=1;     /**   since all probs are divided by prob of this model   **/
  PPROB = PPROB||PP; MODELS = PTERMS||J(1,Q,' ');


   /**************  Models with 1 potential term  **************/
  GAMMAI = GAMMA[1:(P+1),1:(P+1)];

  DO J = 1 TO Q;
       X = XPRI||XPOT[,J];  QQ=1;
       RUN BOXMEYER;
       IF Q>1 THEN
         MODELI = PTERMS||QTERMS[,J]||J(1,(Q-1),' ');
       ELSE
         MODELI = PTERMS||QTERMS[,J];
       MODELS = MODELS//MODELI;
   END;
  /*************  Models with 2 potential terms  *************/
   GAMMAI = GAMMA[1:(P+2),1:(P+2)];
```

```
DO J = 1 TO Q;
      K = J;
      DO WHILE ( K < Q);
        K = K+1;
        X = XPRI||XPOT[,(J||K)];  QQ=2;
        RUN BOXMEYER;
        IF Q>2 THEN
          MODELI = PTERMS||QTERMS[,(J||K)]||J(1,(Q-2),' ');
        ELSE
          MODELI = PTERMS||QTERMS[,(J||K)];
        MODELS = MODELS//MODELI;
      END;
  END;


   /*************  Models with 3 potential terms  *************/
   GAMMAI = GAMMA[1:(P+3),1:(P+3)];

   DO J = 1 TO Q;
      K = J;
      DO WHILE ( K < Q);
        K = K+1; L=K;
        DO WHILE (L < Q);  L = L+1;
           X = XPRI||XPOT[,(J||K||L)];  QQ=3;
           RUN BOXMEYER;
        IF Q>3 THEN
          MODELI = PTERMS||QTERMS[,(J||K||L)]||J(1,(Q-3),' ');
        ELSE
          MODELI = PTERMS||QTERMS[,(J||K||L)];
        MODELS = MODELS//MODELI;
        END;
      END;
  END;

   /*************  Models with 4 potential terms  *************/
   GAMMAI = GAMMA[1:(P+4),1:(P+4)];

   DO J = 1 TO Q;
      K = J;
      DO WHILE ( K < Q);
```

```
      K = K+1; L=K;
      DO WHILE (L < Q);
        L = L+1; M = L;
        DO WHILE (M < Q);
          M = M+1;
          X = XPRI||XPOT[,(J||K||L||M)];  QQ=4;
          RUN BOXMEYER;
          IF Q>4 THEN
            MODELI = PTERMS||QTERMS[,(J||K||L||M)]||J(1,(Q-4),' ');
          ELSE
            MODELI = PTERMS||QTERMS[,(J||K||L||M)];
          MODELS = MODELS//MODELI;
        END;
      END;
    END;
 END;

/**************  Models with 5 potential terms  **************/
 GAMMAI = GAMMA[1:(P+5),1:(P+5)];

 DO J = 1 TO Q;
    K = J;
    DO WHILE ( K < Q);
      K = K+1; L=K;
      DO WHILE (L < Q);
        L = L+1; M = L;
        DO WHILE (M < Q);
          M = M+1; N = M;
          DO WHILE (N < Q);
            N = N+1;
            X = XPRI||XPOT[,(J||K||L||M||N)];  QQ=5;
            RUN BOXMEYER;
            IF Q>5 THEN
              MODELI = PTERMS||QTERMS[,(J||K||L||M||N)]||J(1,(Q-5),' ');
            ELSE
              MODELI = PTERMS||QTERMS[,(J||K||L||M||N)];
            MODELS = MODELS//MODELI;
          END;
        END;
      END;
```

```
    END;
  END;


   /******** Normalize the posterior probabilities ********/
   PPROB = PPROB/SUM(PPROB);
   PRINT MODELS PPROB;

  NTOT = N1+N2;
  NCAND = NROW(CANDID1);
  N1 = NROW(D1);
  D1D1 = D1`*D1;                  /**** X1`X1 for the full model ****/
  LOSSVEC = J(NCAND,1,0);    /**** Initialize the loss vector ****/

   /** Sequentially select the optimal N2 design points for the second stage design **/
     DO I = 1 TO N2;
        DO C=1 TO NCAND;
           NEWD2 = BESTD2//CANDID1[C,];
           LOSSVEC[C] = QMOD1(NEWD2);
        END;

        MINLOC = LOSSVEC[>:<];
        BESTD2 = BESTD2//CANDID1[MINLOC,];
     END;

/*************************************************************************
 ** Read the second stage design matrix into the data set named 'STAGE2',          **
 ** quit PROC IML and print the second stage design.                                **
 *************************************************************************/
 DOX2 = BESTD2[,{2 3 5}];
 CREATE STAGE2 FROM DOX2[COLNAME={ 'X1' 'X2' 'X3'}];
 APPEND FROM DOX2;
 QUIT;

 DATA STAGE2; SET STAGE2;
   X3 = 2*X3;              /** Rescale the design matrix to +/-1 form **/
 PROC PRINT DATA=STAGE2;   RUN;
```

Appendix H

This computer algorithm which is written in SAS Proc IML determines the optimal second stage of a two-stage design. Given a first stage design and first stage data, the program calculates Box and Meyer posterior probabilities for all candidate models. These probabilities are then used to find the optimal second stage design using a Bayes Q, DQ or Q(w) optimality criterion. (It should be noted that although the weight function of the Q(w) criterion is that which emphasizes areas of high response, the algorithm could easily be modified to accommodate other weight functions.) The user controls the selection of one of these four criteria by implementing the appropriate loss function.

The first stage design must be provided by the user in full model form. This algorithm is written for an application in which the full model contains four primary terms {intercept, x_1, x_2, x_1x_2} and five potential terms {x_3, x_1x_3, x_2x_3, x_1^2, x_2^2}. The code which is unique to this application is given in italics. This algorithm can easily accommodate other applications with five potential terms through modification of the italicized code. (Additional code would be required to expand the list of candidate models to include those containing more than five potential terms).

If either the Q or DQ criterion is selected, the user must also supply the *region moment matrix* of the full model, so that the integrated prediction variance (IPV) can be calculated for each candidate design. (For the Q(w) criterion, the IPV is approximated by averaging the prediction variance over the grid of candidate points.)

The output of this program consists of a set of Box and Meyer posterior probabilities for the 2^q candidate models and the resulting second stage design. The design levels are reported with ±1 scaling.

```
/*************************************************************************
**    Read (or construct) the first stage data set which contains the following:      **
**        a)  p primary terms (including a column of 1's for the intercept)          **
**        b)  q potential                                                            **
**        c)  response (y)                                                           **
*************************************************************************/

 *  (example data set)  *
    DATA DESIGN;
       INPUT X1 X2 X3 Y;
       INTRCPT = 1.0;
       X1X3 = X1*X3;
       X2X3 = X2*X3;
       X1X1 = X1*X1;
       X2X2 = X2*X2;
   LINES;
      .
      .
      .
    ;
   RUN;



    /*************************************************************************
     **    Generate the candidate list....                                     **
     **     Note:  It is not necessary to use PROC PLAN to generate the candidate   **
     **            list, but the candidate list must be supplied and placed in data set  **
     **          ' CANDID '.                                                   **
     *************************************************************************/

     PROC PLAN;
        FACTORS X1=5 ORDERED X2=5 ORDERED X3=5 ORDERED /NOPRINT;
        OUTPUT  OUT=CANDID   X1 NVALS= (-1, -.5, 0, .5, 1)
                             X2 NVALS= (-1, -.5, 0, .5, 1)
                             X3 NVALS= (-1, -.5, 0, .5, 1);
    RUN;
```

```
/*************************************************************
**                  Convert the Candidate list to model format              **
*************************************************************/

 DATA CANDID; SET CANDID;
     INTRCPT = 1.0;
     X1X2 = X1*X2;
     X1X3 = X1*X3;
     X2X3 = X2*X3;
     X1X1 = X1*X1;
     X2X2 = X2*X2;


PROC IML;

  USE DESIGN;

   /*************************************************************
    **      Reading from the first stage data set, named 'DESIGN',          **
    **      read the p columns of primary terms into the matrix XPRI.       **
    *************************************************************/
    READ ALL VAR{INTRCPT X1 X2 X1X2 }INTO XPRI;

   /*************************************************************
    **      Reading from the first stage data set, named 'DESIGN',          **
    **      read the q columns of potential terms into the matrix XPOT.     **
    *************************************************************/
    READ ALL VAR{X3 X1X3 X2X3 X1X1 X2X2} INTO XPOT;

   /**      Construct the combined model matrix of the form [XPRI | XPOT]   **/
    D1 = XPRI||XPOT;
    P = NCOL(XPRI);
    Q = NCOL(XPOT);

    TAU = 5;   /*******************************************************
                **    Note:  TAU=5 is the recommended value, but can be        **
                **    changed to better reflect the user's beliefs             **
                *******************************************************/
```

```
 PI = 1/3;   /*********************************************************
               **   Note:  PI is the prior probability that any potential term      **
               **   is active and therefore should reflect the user's beliefs.      **
               **   (  Recommended values are .25 ≤ PI ≤ .50  )                     **
              *********************************************************/


 /*****************************************************************
  **   Read the list of candidate design points into the matrix 'CANDID1'      **
  **   Note:  This list must be in full model form, with columns in the same    **
  **          order as [XPRI | XPOT].                                           **
  *****************************************************************/
 USE CANDID;
     READ ALL VAR {INTRCPT  X1  X2  X1X2  X3  X1X3  X2X3  X1X1  X2X2}
     INTO CANDID1;

/**  Specify the number of design points to be included in the second stage design  **/
 N2 = ????;

 /**   List the primary terms, in the order they appear in the first stage data set    **/
 PTERMS = {'B0'  'B1'  'B2'  'B12'};

 /**   List the potential terms, in the order they appear in the first stage data set   **/
 QTERMS = {'B3'  'B13'  'B23'  'B11'  'B22'};

 N = NROW(XPOT);
NN = (N-1)/2;
 GAMMA = ( J(P,P,0)||J(P,Q,0) )//( J(Q,P,0)||((TAU**-2.0)*I(Q)) );

/*********************************************************
 **       Reading from the first stage data set, named 'DESIGN',          **
 **       read the response data into the vector Y .                      **
 *********************************************************/
 USE DESIGN;   READ ALL VAR {Y} INTO Y;
```

```
/*****************************************************************
 **      If using the Q or DQ criterion, enter MSTAR, the Region Moment Matrix    **
 **      of the full model....                                                    **
 **      Note:  This matrix must be in the same order as X'X for the full  model,  **
 **             where X = [XPRI | XPOT].                                          **
 *****************************************************************/

 MSTAR = { 1     0     0     0     0     0     0    .333  .333,
           0    .333   0     0     0     0     0     0     0   ,
           0     0    .333   0     0     0     0     0     0   ,
           0     0     0    .111   0     0     0     0     0   ,
           0     0     0     0    .333   0     0     0     0   ,
           0     0     0     0     0    .111   0     0     0   ,
           0     0     0     0     0     0    .111   0     0   ,
          .333   0     0     0     0     0     0    .20   .111,
          .333   0     0     0     0     0     0    .111  .20};

 /**********    Begin Box-Meyer module  **********/
 START BOXMEYER;
    B = INV(X`*X + GAMMAI)*X`*Y;
   ESS = SSQ(Y - X*B);
    ONE = ( PI/((1-PI)*TAU) )**QQ;
    TWO = DETPRI/SQRT(DET(GAMMAI + X`*X) );
    THREE = ( (ESS + B`*GAMMAI*B)/ESSPRI)**-NN;
    PP = ONE*TWO*THREE;
    PPROB = PPROB//PP;
 FINISH;
/**********    End  Box-Meyer module   **********/



/**********    Begin  Lmod1 module     **********/
 START   LMOD1(DOX2)
      GLOBAL(P, Q, N1, NROW2, MSTAR, D2D2, D1D1, GAMMA, PPROB);
   NROW2 = NROW(DOX2);
   D2D2 = DOX2`*DOX2;

   INDEX = 1:P;                          /***  0 potential terms ***/
   LOSS = LMOD2(INDEX);
  LVEC = LVEC//LOSS;
```

```
DO J = 1 TO Q;                          /***  1 potential  term  ***/
   INDEX = (1:P)||(P+J);
   LOSS = LMOD2(INDEX);
  LVEC = LVEC//LOSS;
END;

DO J = 1 TO Q;                          /***  2 potential terms  ***/
  K = J;
  DO WHILE ( K < Q);
     K = K+1;
     INDEX = (1:P)||(P+J)||(P+K);
     LOSS = LMOD2(INDEX);
    LVEC = LVEC//LOSS;
  END;
END;

DO J = 1 TO Q;                          /***  3 potential terms  ***/
  K = J;
  DO WHILE ( K < Q);
     K = K+1; L=K;
     DO WHILE (L < Q);
        L = L+1;
        INDEX = (1:P)||(P+J)||(P+K)||(P+L);
        LOSS= LMOD2(INDEX);
       LVEC = LVEC//LOSS;
     END;
  END;
END;


DO J = 1 TO Q;                          /*** 4 potential terms  ***/
  K = J;
  DO WHILE ( K < Q);
     K = K+1; L=K;
     DO WHILE (L < Q);
        L = L+1; M = L;
        DO WHILE (M < Q);
           M = M+1;
          INDEX = (1:P)||(P+J)||(P+K)||(P+L)||(P+M);
```

```
            LOSS = LMOD2(INDEX);
           LVEC = LVEC//LOSS;
          END;
        END;
      END;
    END;

    DO J = 1 TO Q;                          /*** 5 potential terms  ***/
      K = J;
      DO WHILE ( K < Q);
        K = K+1; L=K;
        DO WHILE (L < Q);
          L = L+1; M = L;
          DO WHILE (M < Q);
            M = M+1; N = M;
            DO WHILE (N < Q);
              N = N+1;
              INDEX = (1:P)||(P+J)||(P+K)||(P+L)||(P+M)||(P+N);
              LOSS = LMOD2(INDEX) ;
             LVEC = LVEC//LOSS;
            END;
          END;
        END;
      END;
    END;


    LCOMP = PPROB`*LVEC;   RETURN(LCOMP);
  FINISH LMOD1;
/**********   End  Lmod1 module       **********/



/*************************************************************************
**     Of the following Loss modules, use only that which is appropriate for the      **
**     criterion of interest.                                                          **
*************************************************************************/
```

```
/**********     Begin Loss module for Bayes Q-optimality criterion      **********/
    START LMOD2(I)
        GLOBAL(N1,NROW2, MSTAR, D1D1, D2D2, GAMMA);
        DOXLOSS = TRACE( ( N1+NROW2)*INV(D1D1[I,I] + D2D2[I,I]
                                + GAMMA[I,I])*MSTAR[I,I] );
       RETURN(DOXLOSS);
    FINISH LMOD2;
   /*********     End of Loss module for Bayes Q-optimality criterion      **********/


  /*********     Begin Loss module for Bayes DQ-optimality criterion     ********/
    START LMOD2(I)
        GLOBAL(N1,NROW2, MSTAR, D1D1, D2D2, GAMMA);
        VAR = INV(D1D1[I,I] + D2D2[I,I] + GAMMA[I,I]);
        DOXLOSS = DET(VAR)*TRACE( ( N1+NROW2)*INV(D1D1[I,I] + D2D2[I,I]
                                + GAMMA[I,I])*MSTAR[I,I] )  + DET(VAR);
       RETURN(DOXLOSS);
    FINISH LMOD2;
  /*********     End Loss module for Bayes DQ-optimality criterion     ********/


  /*********     Begin Loss module for Bayes Q(w)-optimality criterion     ********/
    START LMOD2(I)
        GLOBAL(N1,NROW2, D1, D1D1, D2D2, GAMMA, CANDID1, Y);
        BETAHAT = INV(D1D1[I,I] + GAMMA[I,I])*D1[,I]`*Y;
        VAR = INV(D1D1[I,I] + D2D2[I,I] + GAMMA[I,I]);
        DOXLOSS = VECDIAG( CANDID1[,I]*VAR*CANDID1[,I]`)`
                                                    *CANDID1[,I]*BETAHAT;
       RETURN(DOXLOSS);
    FINISH LMOD2;
   /*********     End of Loss module for Bayes Q(w)-optimality criterion   **********/

   /**     Calculate  Box and Meyer posterior probabilities for all candidate models       **/

    /*************  Model with 0 potential terms  *************/
     B=INV(XPRI`*XPRI)*XPRI`*Y;
    ESSPRI = SSQ(Y-XPRI*B);  DETPRI = SQRT(DET(XPRI`*XPRI));
    PP=1;     /**   since all probs are divided by prob of this model   **/
    PPROB = PPROB||PP; MODELS = PTERMS||J(1,Q,' ');
```

```
 /*************  Models with 1 potential term  *************/
GAMMAI = GAMMA[1:(P+1),1:(P+1)];

  DO J = 1 TO Q;
     X = XPRI||XPOT[,J];  QQ=1;
     RUN BOXMEYER;
     IF Q>1 THEN
       MODELI = PTERMS||QTERMS[,J]||J(1,(Q-1),' ');
     ELSE
       MODELI = PTERMS||QTERMS[,J];
     MODELS = MODELS//MODELI;
 END;
/*************  Models with 2 potential terms  *************/
 GAMMAI = GAMMA[1:(P+2),1:(P+2)];

 DO J = 1 TO Q;
     K = J;
     DO WHILE ( K < Q);
        K = K+1;
        X = XPRI||XPOT[,(J||K)];  QQ=2;
        RUN BOXMEYER;
        IF Q>2 THEN
          MODELI = PTERMS||QTERMS[,(J||K)]||J(1,(Q-2),' ');
        ELSE
          MODELI = PTERMS||QTERMS[,(J||K)];
        MODELS = MODELS//MODELI;
     END;
 END;


 /*************  Models with 3 potential terms  *************/
  GAMMAI = GAMMA[1:(P+3),1:(P+3)];

  DO J = 1 TO Q;
     K = J;
     DO WHILE ( K < Q);
        K = K+1; L=K;
        DO WHILE (L < Q);  L = L+1;
           X = XPRI||XPOT[,(J||K||L)];  QQ=3;
```

```
        RUN BOXMEYER;
       IF Q>3 THEN
        MODELI = PTERMS||QTERMS[,(J||K||L)]||J(1,(Q-3),' ');
       ELSE
        MODELI = PTERMS||QTERMS[,(J||K||L)];
       MODELS = MODELS//MODELI;
       END;
    END;
END;

 /*************  Models with 4 potential terms  *************/
 GAMMAI = GAMMA[1:(P+4),1:(P+4)];

 DO J = 1 TO Q;
    K = J;
    DO WHILE ( K < Q);
      K = K+1; L=K;
      DO WHILE (L < Q);
        L = L+1; M = L;
        DO WHILE (M < Q);
          M = M+1;
          X = XPRI||XPOT[,(J||K||L||M)];  QQ=4;
          RUN BOXMEYER;
          IF Q>4 THEN
            MODELI = PTERMS||QTERMS[,(J||K||L||M)]||J(1,(Q-4),' ');
          ELSE
            MODELI = PTERMS||QTERMS[,(J||K||L||M)];
          MODELS = MODELS//MODELI;
        END;
      END;
    END;
 END;


/*************  Models with 5 potential terms  *************/
GAMMAI = GAMMA[1:(P+5),1:(P+5)];

DO J = 1 TO Q;
    K = J;
    DO WHILE ( K < Q);
```

```
        K = K+1; L=K;
        DO WHILE (L < Q);
          L = L+1; M = L;
          DO WHILE (M < Q);
            M = M+1; N = M;
            DO WHILE (N < Q);
              N = N+1;
              X = XPRI||XPOT[,(J||K||L||M||N)];  QQ=5;
              RUN BOXMEYER;
              IF Q>5 THEN
                MODELI = PTERMS||QTERMS[,(J||K||L||M||N)]||J(1,(Q-5),' ');
              ELSE
                MODELI = PTERMS||QTERMS[,(J||K||L||M||N)];
              MODELS = MODELS//MODELI;
            END;
          END;
        END;
      END;
  END;

   /********  Normalize the posterior probabilities  ********/
   PPROB = PPROB/SUM(PPROB);
   PRINT MODELS PPROB;

   NTOT = N1+N2;
   NCAND = NROW(CANDID1);
   N1 = NROW(D1);
   D1D1 = D1`*D1;                  /****  X1`X1 for the full model  ****/
   LOSSVEC = J(NCAND,1,0);    /****  Initialize the loss vector  ****/

/**   Sequentially select the optimal N2 design points for the second stage design **/
     DO I = 1 TO N2;
          DO C=1 TO NCAND;
               NEWD2 = BESTD2//CANDID1[C,];
               LOSSVEC[C] = QMOD1(NEWD2);
          END;

          MINLOC = LOSSVEC[>:<];
          BESTD2 = BESTD2//CANDID1[MINLOC,];
     END;
```

```
/*****************************************************************
**   Read the second stage design matrix into the data set named 'STAGE2',          **
**   quit  PROC IML  and print the second stage design.                              **
*****************************************************************/
DOX2 = BESTD2[,{2 3 5}];
CREATE STAGE2 FROM DOX2[COLNAME={ 'X1' 'X2' 'X3'}];
APPEND FROM DOX2;
QUIT;
PROC PRINT DATA=STAGE2;    RUN;
```

Appendix I

This computer algorithm, written in SAS Proc IML, selects an optimal second stage design using the Q(w) criterion under the assumption that the model is known after the first stage. Given a first stage design and first stage data, the program determines the Q(w)-optimal design (i.e., that which minimizes the average prediction variance for a specified model, especially in areas of high response). It should be noted that the loss module in this program could be modified to accommodate other Q(w) criteria, such as a criterion emphasizing areas of low response or a criterion emphasizing a target response.

The first stage design and the list of candidate design points must be provided by the user in the known model form. This algorithm is written for an application in which the known model is a full quadratic model in two variables, x_1 and x_2. This algorithm can be used for other applications through modification of the italicized code.

The output of this program contains the optimal second stage design. The design levels are reported with ± 1 scaling.

```
/*****************************************************************
**    Read (or construct) the first stage data set which contains the following:      **
**        a)  all known model terms (including a column of 1's for the intercept)     **
**        b)  response (y)                                                            **
*****************************************************************/

* (example data set) *
DATA DESIGN;
   INPUT X1 X2 Y;
   INTRCPT = 1.0;
   X1X2 = X1*X2;
   X1X1 = X1*X1;
   X2X2 = X2*X2;
   LINES;
         .
         .
         .
     ;
 RUN;

 /*****************************************************************
  **    Generate the candidate list....                                          **
  **     Note:  It is not necessary to use PROC PLAN to generate the candidate   **
  **            list, but the candidate list must be supplied and placed in data set **
  **          ' CANDID '.                                                        **
  *****************************************************************/
    PROC PLAN;
     FACTORS X1=5 ORDERED X2=5 ORDERED /NOPRINT ;
     OUTPUT OUT=CANDID   X1 NVALS= (-1, -.5, 0, .5, 1)
                         X2 NVALS= (-1, -.5, 0, .5, 1);
   RUN;


    /*****************************************************************
     *******          Convert the Candidate list to model format          ******
     *****************************************************************/

      DATA CANDID;   SET CANDID;
         INTRCPT = 1.0;
         X1_X2 = X1*X2;
         X1_X1 = X1*X1;
         X2_X2 = X2*X2;
```

```
PROC IML;

 /************          Begin Loss Module      **********/
  START LOSS(DOX2) GLOBAL( CANDID1,D1D1, BETAHAT,NCAND);
  VAR = INV(D1D1 + DOX2`*DOX2);
  Q=VECDIAG(CANDID1*VAR*CANDID1`)`*CANDID1*BETAHAT/NCAND;
  RETURN(Q);
  FINISH LOSS;
 /************          End Loss Module       **********/


 /*********************************************************************
  **      Reading from the first stage data set named 'DESIGN', read the first      **
  **      stage model matrix into the matrix 'DOX1'.  Read the response data        **
  **      into the vector Y.                                                        **
  *********************************************************************/
  USE DESIGN;
  READ ALL VAR{INTCPT X1 X2 X1X2  X1X1 X2X2} INTO DOX1;
  READ ALL VAR{Y} INTO Y;


 /*********************************************************************
  **      Read the list of candidate design points into the matrix 'CANDID1'        **
  **       Note:  This list must be read in the same order as the first stage model  **
  **              matrix, 'DOX1'.                                                    **
  *********************************************************************/
  USE CANDID;
  READ ALL VAR{INTRCPT X1 X2 X1_X2 X1_X1 X2_X2} INTO CANDID1;


 /**  Specify the number of design points to be included in the second stage design  **/
   N2 = ????;


  NCAND = NROW(CANDID1);
  D1D1 = DOX1`*DOX1;
  BETAHAT = INV(D1D1)*DOX1`*Y;
```

```
    /**  Sequentially select the optimal N2 design points for the second stage design  **/
     LOSSVEC = J(NCAND,1,0);
     DO I=1 TO N2;
        DO C=1 TO NCAND;
           NEWD2 = BESTD2//CANDID1[C,];
          LOSSVEC[C] = LOSS(NEWD2);
        END;
        MINLOC = LOSSVEC[>:<];
        BESTD2 = BESTD2//CANDID1[MINLOC,];
     END;

     PRINT BESTD2;

QUIT;
```

Appendix J

This computer algorithm which is written in SAS Proc IML determines the D-optimal augmentation for variance modeling. Given at least two data points at each point of a resolution III fraction, the program calculates Box and Meyer posterior probabilities for all candidate variance models. These probabilities are then used to find the D-optimal augmentation under variance model uncertainty.

This algorithm is written for an application in which the full variance model contains the following terms, {intercept, x_1, x_2, x_3}. The code which is unique to this application is given in italics. This algorithm can easily be modified to accommodate other first order models containing no more than five main effect terms (additional code is necessary for larger models).

The output of this program is the D-optimal design augmention. The design levels are reported with ± 1 scaling.

```
/*************************************************************************
**      Generate the candidate list which is a resolution III fraction in model format.     **
**      Enter the sample variances already observed at each point of the resolution        **
**      III fraction, along with the corresponding observed sample size, n                 **
**      (i.e., each sample variance has n-1 degrees of freedom).                           **
*************************************************************************/

  DATA CANDID;
   INPUT X1 X2 X3 n smplvar;
   INTCPT = 1.0;
   LINES;
    .
    .
    .
   ;
```

```
/*************************************************************************
**        Enter the first stage design for the mean model and the observed data (Y).      **
**        All regressors in the known mean model should be included in this data set.   **
*************************************************************************/
  DATA FIRST;
  INPUT RUN X1 X2 X3 Y;
  INTCPT = 1.0;
  X1X2 = X1*X2;
  X1X3 = X1*X3;
  X1X1 = X1*X1;
  X3X3 = X3*X3;

  LINES;
    .
    .
    .
   ;



  PROC IML;

    USE FIRST;
          READ ALL VAR{INTCPT X1 X2 X3 X1X2 X1X3 X1X1 X3X3}INTO X;
          READ ALL VAR{Y} INTO Y;
    USE CANDID;
          READ ALL VAR{INTCPT X1 X2 X3} INTO CANDID1;
          READ ALL VAR{SMPLVAR} INTO S;
          READ ALL VAR{N} INTO N;

  /** Specify the number of  design points to be included in the augmentation  **/
   N2 = ????;

  /** Specify the prior standard deviation of any variance model parameter  **/
   GAMMASTD=????;

  /** Specify the prior probability that any variance model term is active  **/
   PI = ????;            /** Note:  PI=.5 results in a noninformative prior **/

   Z=CANID1;  D1=Z;  DF1=N-1;
   P=NCOL(X);
```

```
Q = NCOL(Z); N = NROW(X); M = Q-1;   /** M is # of main effects **/
NCAND = NROW(CANDID1);
N1=NROW(D1);
NTOT = N1+N2;

/***  Calculate log(mse), the estimated prior mean of gamma0  ***/
BETAHAT = INV(X`*X)*X`*Y;
LOGMSE = LOG( (Y-X*BETAHAT)`*(Y-X*BETAHAT)/(N-P) );
GMAMEAN = LOGMSE//J(M,1,0);
PROB=J((2**(Q-1)),1,0);

/****        Module which calculates marginal prob. of data given M_i          *****/
 START MARGPROB(I) GLOBAL(GAMMA, S, Q, Z);
    SIGMA = EXP(Z[,I]*GAMMA[I]);
    PRODUCT=1;
    DO J=1 TO Q;
       PRODUCT=PRODUCT*SQRT(SIGMA[J]*S[J]);
    END;
    BMPROB = (1/PRODUCT)*EXP(-.5*S`*(SIGMA##-1));
    RETURN(BMPROB);
 FINISH MARGPROB;
/*****************************************************************/

/***       Use Monte Carlo method to find the marginal density of  y       ***/
  DO COUNT2 = 1 TO 2000;          /***  gamma loop  ***/
    GAMINIT=J(Q,1,0);
    NMLERR = NORMAL(GAMINIT);
    GAMMA = GMAMEAN + NMLERR#GAMMASTD;

   /*****  Calculate marginal prob. for each candidate model  ******/
    MODEL=1;                    /****  constant var model  ****/
    INDEX=1;
    PROB[MODEL] = PROB[MODEL] + MARGPROB(INDEX);

    DO J=2 TO Q;                /****  intercept and 1 term ****/
       MODEL=MODEL+1;
       INDEX=1||J;
       PROB[MODEL]= PROB[MODEL] + MARGPROB(INDEX);
    END;
```

```
  DO J=2 TO Q;                  /****  intercept and 2 terms ****/
    K=J;
    DO WHILE (K<Q);
      K=K+1;
      MODEL=MODEL+1;
      INDEX=1||J||K;
       PROB[MODEL]= PROB[MODEL] + MARGPROB(INDEX);
    END;
  END;
  DO J=2 TO Q;                  /****  intercept and 3 terms ****/
    K=J;
    DO WHILE (K<Q); K=K+1;L=K;
      DO WHILE (L<Q);
        L=L+1;
        MODEL=MODEL+1;
        INDEX=1||J||K||L;
         PROB[MODEL]= PROB[MODEL] + MARGPROB(INDEX);
      END;
    END;
  END;

DO J=2 TO Q;                    /****  intercept and 4 terms ****/
    K=J;
    DO WHILE (K<Q); K=K+1;L=K;
      DO WHILE (L<Q);
        L=L+1; M=L;
        DO WHILE (M<Q); M=M+1;
          MODEL=MODEL+1;
          INDEX=1||J||K||L|M;
           PROB[MODEL]= PROB[MODEL] + MARGPROB(INDEX);
        END;
      END;
    END;
END;

DO J=2 TO Q;                    /****  intercept and 5 terms ****/
    K=J;
    DO WHILE (K<Q); K=K+1;L=K;
      DO WHILE (L<Q);
        L=L+1; M=L;
```

```
          DO WHILE (M<Q); M=M+1; O=M;
           DO WHILE (O<Q); O=O+1;
             MODEL=MODEL+1;
            INDEX=1||J||K||L|M|O;
             PROB[MODEL]= PROB[MODEL] + MARGPROB(INDEX);
           END;
         END;
       END;
     END;
  END;

END;


 /*************************************************************
  *****    Calculate prior prob. for each candidate variance model          *****
  *****    where the prior prob. that any term is active is pi.              *****
  *************************************************************/
  NUMODELS=2**M;
  PP = J(NUMODELS,1,0);      /** initialize the vector of prior probabilities **/
  PP[1] = (1-PI)**M;          /** prior prob. of constant variance model **/;

  DO K=2 TO Q;                /** candidate models containing only one main effect **/
    PP[K] = PI*(1-PI)**(M-1);
  END;

  LAST=M*(M-1)/2;
  IF LAST^=0 THEN DO;
  K=Q;
  DO J=1 TO LAST;             /** candidate models containing two main effects **/
    K=K+1;
    PP[K] = (PI**2)*(1-PI)**(M-2);
  END;
  END;

  LAST=M*(M-1)*(M-2)/6;
  IF LAST^=0 THEN DO;
  DO J=1 TO LAST;             /** candidate models containing three main effects **/
    K=K+1;
    PP[K] = (PI**3)*(1-PI)**(M-3);
```

```
 END;
 END;

 LAST=M*(M-1)*(M-2)*(M-3)/24;
 IF LAST^=0 THEN DO;
 DO J=1 TO LAST;              /** candidate models containing four main effects **/
   K=K+1;
   PP[K] = (PI**4)*(1-PI)**(M-4);
 END;
 END;

 LAST=M*(M-1)*(M-2)*(M-3)*(M-4)/120;
 IF LAST^=0 THEN DO;
 DO J=1 TO LAST;            /** candidate models containing five main effects **/
   K=K+1;
   PP[K] = (PI**5)*(1-PI)**(M-5);
 END;
 END;

 FINLPROB=PP#PROB/SUM(PP#PROB);

 /*****************************************************************
  **          Wmod:  Module for calculating a collapsed version                  **
  **                    of the W matrix for each subset model                    **
  *****************************************************************/
 START WMOD(I) GLOBAL(D1, GMAMEAN, GAMMASTD, N1);
   WVEC = VECDIAG(EXP(DIAG(D1[,I]*D1[,I]`#(.5*GAMMASTD**2))
                - DIAG(D1[,I]*GMAMEAN[I])) - .5#I(N1) )`;
   RETURN(WVEC);
 FINISH WMOD;


/*****************************************************************
 **   Calculate W matrix (part of the information matrix) for each candidate      **
 **   (ie. subset)  model. Note: The diagonal elements of each W matrix are stored  **
 **    as a single row in the matrix denoted as Wtotal.                           **
 *****************************************************************/
  WTOTAL = J(NUMODELS,Q,0);        /****  initialize W matrix ****/
```

```
MODEL=1;                              /****  constant var model  ****/
 INDEX=1;
 WTOTAL[MODEL,] = WTOTAL[MODEL,] + WMOD(INDEX);

 DO J=2 TO Q;                         /****  intercept and 1 term ****/
    MODEL=MODEL+1;
    INDEX=1||J;
    WTOTAL[MODEL,] = WTOTAL[MODEL,] + WMOD(INDEX);
 END;
 DO J=2 TO Q;                         /****  intercept and 2 terms ****/
    K=J;
    DO WHILE (K<Q);
      K=K+1;
      MODEL=MODEL+1;
      INDEX=1||J||K;
      WTOTAL[MODEL,] = WTOTAL[MODEL,] + WMOD(INDEX);
    END;
 END;

 DO J=2 TO Q;                         /****  intercept and 3 terms ****/
    K=J;
    DO WHILE (K<Q); K=K+1;L=K;
      DO WHILE (L<Q);
        L=L+1;
        MODEL=MODEL+1;
        INDEX=1||J||K||L;
        WTOTAL[MODEL,] = WTOTAL[MODEL,] + WMOD(INDEX);
      END;
    END;
 END;

DO J=2 TO Q;                          /****  intercept and 4 terms ****/
    K=J;
    DO WHILE (K<Q); K=K+1;L=K;
      DO WHILE (L<Q);
        L=L+1; M=L;
        DO WHILE (M<Q); M=M+1;
          MODEL=MODEL+1;
          INDEX=1||J||K||L|M;
          WTOTAL[MODEL,] = WTOTAL[MODEL,] + WMOD(INDEX);
```

```
        END;
      END;
    END;
END;

DO J=2 TO Q;                          /****  intercept and 5 terms ****/
    K=J;
    DO WHILE (K<Q); K=K+1;L=K;
      DO WHILE (L<Q);
        L=L+1; M=L;
        DO WHILE (M<Q); M=M+1; O=M;
          DO WHILE (O<Q); O=O+1;
            MODEL=MODEL+1;
            INDEX=1||J||K||L|M|O;
             WTOTAL[MODEL,] = WTOTAL[MODEL,] + WMOD(INDEX);
          END;
        END;
      END;
    END;
END;


 /********  Begin Loss1 module   **********/
 START LOSS1(DOX2) GLOBAL(P, Q, N1, NROW2, GMAMEAN,
   GAMMASTD, D1, FINLPROB, DF1, DF);
   NROW2 = NROW(DOX2);
   DF=DF1;
   DO I=1 TO N1;
      DO J=1 TO NROW2;
        DF[I] = DF[I] + ALL(D1[I,]=DOX2[J,]);
      END;
   END;

   INDEX=1;                                /****  constant var model  ****/
   DET = QMOD2(INDEX);
   DETVEC = DETVEC//DET;
```

```
DO J=2 TO Q;                          /****  intercept and 1 term ****/
  INDEX=1||J;
   DET = QMOD2(INDEX);
  DETVEC = DETVEC//DET;
END;

DO J=2 TO Q;                          /****  intercept and 2 terms ****/
  K=J;
  DO WHILE (K<Q);
    K=K+1;
    INDEX=1||J||K;
     DET = QMOD2(INDEX);
    DETVEC = DETVEC//DET;
  END;
END;

DO J=2 TO Q;                           /****  intercept and 3 terms ****/
  K=J;
  DO WHILE (K<Q); K=K+1;L=K;
    DO WHILE (L<Q);
      L=L+1;
      INDEX=1||J||K||L;
       DET = QMOD2(INDEX);
      DETVEC = DETVEC//DET;
    END;
  END;
END;

 DO J=2 TO Q;                            /****  intercept and 4 terms ****/
    K=J;
    DO WHILE (K<Q); K=K+1;L=K;
      DO WHILE (L<Q);
        L=L+1; M=L;
         DO WHILE (M<Q); M=M+1;
           INDEX=1||J||K||L|M;
            DET = QMOD2(INDEX);
           DETVEC = DETVEC//DET;
         END;
      END;
   END;
```

```
  END;

  DO J=2 TO Q;                          /****  intercept and 5 terms ****/
    K=J;
    DO WHILE (K<Q); K=K+1;L=K;
      DO WHILE (L<Q);
         L=L+1; M=L;
         DO WHILE (M<Q); M=M+1; O=M;
           DO WHILE (O<Q); O=O+1;
              INDEX=1||J||K||L|M|O;
               DET = QMOD2(INDEX);
              DETVEC = DETVEC//DET;
           END;
        END;
      END;
    END;
  END;

 DETCOMP = FINLPROB`*DETVEC; RETURN(DETCOMP);
FINISH LOSS1;
/********  End of Loss1 module   **********/

/********  Loss2 module  *********/
 START LOSS2(I) GLOBAL(DF, DOX2, WTOTAL, MODEL, D1,N1);
  Q = NCOL(I);
  LOGD = N1*LOG(WTOTAL[MODEL,1]) + LOG(DET(D1[,I]`*
            DIAG(DF`#WTOTAL[MODEL,]/WTOTAL[MODEL,1])*D1[,I]));
  LOSS = -LOGD;
  RETURN(LOSS);
 FINISH LOSS2;
/********  End of Loss2 module   **********/
 D1D1 = D1`*D1;                  /**    X1`X1 for the Full Variance Model  **/
 LOSSVEC = J(NCAND,1,0);         /**   Initialize the loss vector    **/

 /**   Sequentially select the optimal N2 design points for the second stage design **/
    DO I = 1 TO N2;
        DO C=1 TO NCAND;
            NEWD2 = BESTD2//CANDID1[C,];
           LOSSVEC[C] = LOSS1(NEWD2);
        END;
```

```
        MINLOC = LOSSVEC[>:<];
        BESTD2 = BESTD2//CANDID1[MINLOC,];
    END;


/*****************************************************************
 **   Read the second stage design matrix into the data set named 'STAGE2',          **
 **   quit  PROC IML  and print the second stage design.                              **
 *****************************************************************/
 DOX2 = BESTD2[,{2 3 4}];     /**  Design variables X1 - X3 are in columns 2-4  **/
 CREATE STAGE2 FROM DOX2[COLNAME={ 'X1' 'X2' 'X3'}];
 APPEND FROM DOX2;
 QUIT;
 PROC PRINT DATA=STAGE2;    RUN;
```

Vita

Angela R. Neff

Angela Neff was born July 24, 1964 to Roy and Pauline Neff of Roanoke, Virginia. She graduated from Lord Botetourt High School in Daleville, VA in 1982. In May 1986 she graduated Magna Cum Laude from James Madison University in Harrisonburg, VA with a B.S. degree in Mathematics. The author was then employed by Automated Sciences Group of Dahlgren, VA as a Junior Programmer/Analyst. In the fall of 1987, Angela left ASG to pursue an M.S. degree in Statistics at Virginia Tech. She was awarded that degree in December, 1988.

Angela's career in statistics began with the Applied Statistics group of Eastman Chemical Company. Her career continued at Alcatel Telecommunications Cable. After having gained five years experience consulting in industry, the author returned to Virginia Tech in January, 1994, to pursue a Ph.D. in Statistics, which she received three years later. She was awarded the Jesse C. Arnold Award for Teaching Excellence from Virginia Tech in the fall of 1996.

The author is a member of the honor societies Pi Mu Epsilon and Mu Sigma Rho. She is also a member of the American Statistical Association.