ESTIMATION OF MEAN AND VARIANCE RESPONSE

SURFACES IN ROBUST PARAMETER DESIGN

MATTHIAS TAN HWAI YONG

(B.Eng. (Hons.), UTM)

A THESIS SUBMITTED FOR THE DEGREE OF

MASTER OF ENGINEERING

DEPARTMENT OF INDUSTRIAL AND SYSTEMS ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2008

ACKNOWLEDGEMENT

First, I would like to express my deepest gratitude to my parents. I am deeply indebted to them for supporting me financially as I relied almost exclusively on their hard-earned money to pursue my studies as a graduate student in NUS and also as an undergraduate student in UTM. Without them, I would not be able to achieve what I have achieved. In addition, I am also grateful to my entire family for their moral support.

Next, I would like to thank my supervisor Dr. Ng Szu Hui for her guidance and support. Her advice helped me improve this thesis significantly. I also thank NUS for admitting me to the M.Eng program.

Finally, I thank all who have influenced, stimulated, and supported my work in various ways.

TABLE OF CONTENTS

SUMMARY

In robust parameter design, mean and variance models are estimated with data from a combined array experiment, and are subsequently used for process and product optimization. The design of the combined array experiment and estimation of the mean and variance models depend on the means and covariances of the noise variables, which are quantities assumed known with certainty in the literature. However, this is rarely the case in practice, as the parameters are often estimated with field data. Therefore, standard experimentation and optimization conducted with estimated parameters can lead to results that are far from optimal due to variability in the data. To ensure that the best results are obtained with the available resource, field data collection and experiment must be planned in an integrated way.

In this thesis, a methodology that integrates planning of the combined array experiment with planning of the estimation of the means and variances of the noise variables is proposed. It is assumed that random samples from the process are used to estimate those parameters. Novel ideas introduced with the methodology are expounded in this thesis. A method for specifying the levels of the noise variables is presented. The effect of errors in estimating the means and variances of the noise variables on the estimated mean and variance models is investigated. In addition, the variances of the estimators for the mean and variance models are derived. It is demonstrated that the variances can be inflated considerably by sampling variation.

Because sampling error is as significant as experiment error as a source of variability, simultaneous planning of the sampling effort and experiment is proposed so that total resource is optimally allocated for estimation of the mean and variance models. A mathematical program is formulated to find the sample sizes and mixed

resolution design that minimizes the average variance of the estimator for mean model. A similar mathematical program is formulated for the minimization of the average variance of the unbiased estimator for the variance model minus the residual mean square. It is proven that the continuous relaxations of these programs have convex and differentiable objective functions. A third mathematical program is offered for finding solutions that compromise between the minimization of the two objectives. In addition, a greedy algorithm for finding schemes that have low values of the average variances given a candidate set of design points is proposed.

The variances of the estimators for the mean and variance models depend on parameters of the response model. A similar problem, which is the dependence of optimal designs on model parameters, occurs in nonlinear experimental design. A review of methods proposed to address this problem is made. Application of these methods to the problem of specifying unknown parameters in the variance formulas for the estimators of the mean and variance models is discussed. Expected variance criteria are introduced to allow the use of prior distributions instead of point estimates for the parameters in determining the optimal sample sizes and mixed resolution designs. Additionally, a discussion of how ideas from the robust optimization literature can be employed to handle uncertainty in the model parameters is given. Finally, graphical plots are introduced to allow comparison of the performances of alternative combinations of sample sizes and designs.

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

$\boldsymbol{\mu} = \{\mu_j\} = n \times 1$ vector of the mean of the noise variables in un-coded metric, where n is the number of noise variables.

$\hat{\boldsymbol{\mu}} = \{\hat{\mu}_j\} =$ an estimator of $\boldsymbol{\mu}$.

$\boldsymbol{\Sigma} =$ covariance matrix of noise variables in un-coded metric.

$\hat{\boldsymbol{\Sigma}} =$ an estimator of $\boldsymbol{\Sigma}$.

$\sigma_j^2 =$ the jth diagonal element of $\boldsymbol{\Sigma}$, i.e. the variance of the jth noise variable.

$\hat{\sigma}_j^2 =$ the jth diagonal element of $\hat{\boldsymbol{\Sigma}}$.

$\mathbf{x} = \{x_i\} = k \times 1$ vector of control variables in coded units, where k is the number of control variables.

$\boldsymbol{\xi} = \{\xi_j\} = n \times 1$ vector of noise variables in un-coded metric.

$c_j =$ scaling factor for the jth noise variable.

$\mathbf{q} = \{q_j\} = \{(\xi_j - \mu_j)/(c_j\sigma_j)\} = n \times 1$ vector of noise variables in coded units.

$y(\mathbf{x},\mathbf{q}) = \beta_0 + \mathbf{x}'\boldsymbol{\beta} + \mathbf{x}'\mathbf{B}\mathbf{x} + \boldsymbol{\gamma}'\mathbf{q} + \mathbf{x}'\boldsymbol{\Delta}\mathbf{q} + \varepsilon =$ the response model written as a function of $\mathbf{x}$ and $\mathbf{q}$.

$\beta_0 =$ intercept of the response model $y(\mathbf{x},\mathbf{q})$.

$\boldsymbol{\beta} = \{\beta_j\} = k \times 1$ vector of constants, where β_j is the coefficient of x_i in the response model $y(\mathbf{x},\mathbf{q})$.

$\mathbf{B} = \{B_{ij}\} = k \times k$ matrix of constants, where $B_{ii} = \beta_{ii}$ is the coefficient of x_i^2 in the response model and $B_{ij} = \beta_{ij}/2, i \neq j$ is half the coefficient of $x_i x_j$ in the response model $y(\mathbf{x},\mathbf{q})$.

$\boldsymbol{\gamma} = \{\gamma_j\} = n \times 1$ vector of constants, where γ_j is the coefficient of q_j in the response model $y(\mathbf{x},\mathbf{q})$.

$\boldsymbol{\Delta} = \{\delta_{ij}\} = k \times n$ matrix of constants, where δ_{ij} is the coefficient of $x_i q_j$ in the response model $y(\mathbf{x},\mathbf{q})$.

ε = a random variable representing residual variation in the response after accounting for the systematic component, which is the mean of the response given $\mathbf{x}$ and $\boldsymbol{\xi}$.

σ^2 = variance of ε.

$\hat{y}(\mathbf{x},\mathbf{q}) = \hat{\beta}_0 + \mathbf{x}'\hat{\boldsymbol{\beta}} + \mathbf{x}'\hat{\mathbf{B}}\mathbf{x} + \hat{\boldsymbol{\gamma}}'\mathbf{q} + \mathbf{x}'\hat{\boldsymbol{\Delta}}\mathbf{q}$ = least squares estimator of $y(\mathbf{x},\mathbf{q})$.

$\hat{\beta}_0$ = least squares estimator of β_0.

$\hat{\boldsymbol{\beta}}$ = least squares estimator of $\boldsymbol{\beta}$.

$\hat{\mathbf{B}}$ = least squares estimator of $\mathbf{B}$.

$\hat{\boldsymbol{\gamma}}$ = least squares estimator of $\boldsymbol{\gamma}$.

$\hat{\boldsymbol{\Delta}}$ = least squares estimator of $\boldsymbol{\Delta}$.

$\hat{\sigma}^2$ = residual mean square.

$\mu_Y = \beta_0 + \mathbf{x}'\boldsymbol{\beta} + \mathbf{x}'\mathbf{B}\mathbf{x}$ = mean of the response/ the mean model.

$\hat{\mu}_Y = \hat{\beta}_0 + \mathbf{x}'\hat{\boldsymbol{\beta}} + \mathbf{x}'\hat{\mathbf{B}}\mathbf{x}$ = estimator of the mean model obtained using the coefficients of $\hat{y}(\mathbf{x},\mathbf{q})$.

$\mathrm{var}(\mathbf{Q})$ = covariance matrix of $\mathbf{Q}$, the random vector of noise variables in coded units $\mathbf{q}$.

$\sigma_Y^2 = (\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x})'\mathrm{var}(\mathbf{Q})(\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x}) + \sigma^2$ = variance of the response/ the variance model.

$\hat{\sigma}_{YB}^2 = (\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\Delta}}'\mathbf{x})'\mathrm{var}(\mathbf{Q})(\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\Delta}}'\mathbf{x}) + \hat{\sigma}^2$ = biased estimator of the variance model obtained using the coefficients of $\hat{y}(\mathbf{x},\mathbf{q})$.

$\mathbf{C} = \begin{cases} \mathrm{var}(\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\Delta}}'\mathbf{x})/\sigma^2, \text{if the noise variables are coded by } \mathbf{q}. \\ \mathrm{var}[(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\mathbf{x})|\mathbf{z}]/\sigma^2, \text{if the noise variables are coded by } \mathbf{z}. \end{cases}$

$\hat{\sigma}_Y^2 = (\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\Delta}}'\mathbf{x})'\mathrm{var}(\mathbf{Q})(\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\Delta}}'\mathbf{x}) + \hat{\sigma}^2\{1 - trace[\mathrm{var}(\mathbf{Q})\mathbf{C}]\}$ = unbiased estimator of the variance model obtained using the coefficients of $\hat{y}(\mathbf{x},\mathbf{q})$.

N = total number of experiment runs.

$\mathbf{x}_l$ = coded levels of the control variables for the lth experiment run, $l = 1,\ldots,N$.

R = design region for the control variables (contains all permissible values of $\mathbf{x}_l$).

$\mathbf{z} = \{(\xi_j - \hat{\mu}_j)/(c_j \hat{\sigma}_j)\}$ = $n \times 1$ vector representing the coding for the noise variables when $\boldsymbol{\mu}$ is estimated by $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{\Sigma}$ is estimated by $\hat{\boldsymbol{\Sigma}}$.

$\mathbf{z}_l$ = coded levels of the noise variables in coded units $\mathbf{z}$ for the lth experiment run, $l = 1, \ldots, N$.

S = design region for the noise variables (contains all permissible values of $\mathbf{z}_l$).

ξ_l = un-coded levels of the noise variables for the lth experiment run.

S_ξ = experiment region of the noise variables in un-coded units, i.e. the set of ξ corresponding to the set of $\mathbf{z}$ in S.

m_j = sample size for the jth noise variable, $j = 1, \ldots, N$.

$\mathbf{e} = \{e_l\}$ = the vector of experiment errors.

$y(\mathbf{x}, \xi) = \beta_{0\xi} + \mathbf{x}'\boldsymbol{\beta}_\xi + \mathbf{x}'\mathbf{B}_\xi\mathbf{x} + \boldsymbol{\gamma}'_\xi\xi + \mathbf{x}'\boldsymbol{\Delta}_\xi\xi + \varepsilon$ = response model written as a function of $\mathbf{x}$ and ξ, where $\beta_{0\xi}$, $\boldsymbol{\beta}_\xi$, $\mathbf{B}_\xi$, $\boldsymbol{\gamma}_\xi$, and $\boldsymbol{\Delta}_\xi$ are the model coefficients.

τ_{II} = expected proportion of the joint distribution of the noise variables contained by S_ξ.

$y(\mathbf{x}, \mathbf{z}) = \beta_{0z} + \mathbf{x}'\boldsymbol{\beta}_z + \mathbf{x}'\mathbf{B}_z\mathbf{x} + \boldsymbol{\gamma}'_z\mathbf{z} + \mathbf{x}'\boldsymbol{\Delta}_z\mathbf{z} + \varepsilon$ = the response model written as a function of $\mathbf{x}$ and $\mathbf{z}$.

β_{0z} = intercept of the response model $y(\mathbf{x}, \mathbf{z})$.

$\boldsymbol{\beta}_z = \{\beta_{jz}\}$ = $k \times 1$ vector, where β_{jz} is the coefficient of x_i in the response model $y(\mathbf{x}, \mathbf{z})$.

$\mathbf{B}_z = \{B_{ijz}\}$ = $k \times k$ matrix, where $B_{iiz} = \beta_{iiz}$ is the coefficient of x_i^2 in the response model $y(\mathbf{x}, \mathbf{z})$ and $B_{ijz} = \beta_{ijz}/2, i \neq j$ is half the coefficient of $x_i x_j$ in the response model $y(\mathbf{x}, \mathbf{z})$.

$\boldsymbol{\gamma}_z = \{\gamma_{jz}\}$ = $n \times 1$ vector, where γ_{jz} is the coefficient of $z_j = (\xi_j - \hat{\mu}_j)/(c_j \hat{\sigma}_j)$ in the response model $y(\mathbf{x}, \mathbf{z})$.

$\boldsymbol{\Delta}_z = \{\delta_{ijz}\}$ = $k \times n$ matrix, where δ_{ijz} is the coefficient of $x_i(\xi_j - \hat{\mu}_j)/(c_j \hat{\sigma}_j) = x_i z_j$ in the response model $y(\mathbf{x}, \mathbf{z})$.

$\hat{y}(\mathbf{x},\mathbf{z}) = \hat{\beta}_{0\mathbf{z}} + \mathbf{x}'\hat{\boldsymbol{\beta}}_{\mathbf{z}} + \mathbf{x}'\hat{\mathbf{B}}_{\mathbf{z}}\mathbf{x} + \hat{\boldsymbol{\gamma}}'_{\mathbf{z}}\,\mathbf{z} + \mathbf{x}'\hat{\boldsymbol{\Delta}}_{\mathbf{z}}\mathbf{z}$ = least squares estimator of $y(\mathbf{x},\mathbf{z})$.

$\hat{\beta}_{0\mathbf{z}}$ = least squares estimator of $\beta_{0\mathbf{z}}$.

$\hat{\boldsymbol{\beta}}_{\mathbf{z}}$ = least squares estimator of $\boldsymbol{\beta}_{\mathbf{z}}$.

$\hat{\mathbf{B}}_{\mathbf{z}}$ = least squares estimator of $\mathbf{B}_{\mathbf{z}}$.

$\hat{\boldsymbol{\gamma}}_{\mathbf{z}}$ = least squares estimator of $\boldsymbol{\gamma}_{\mathbf{z}}$.

$\hat{\boldsymbol{\Delta}}_{\mathbf{z}}$ = least squares estimator of $\boldsymbol{\Delta}_{\mathbf{z}}$.

$\hat{\mu}_{Y|\mathbf{z}} = \hat{\beta}_{0\mathbf{z}} + \mathbf{x}'\hat{\boldsymbol{\beta}}_{\mathbf{z}} + \mathbf{x}'\hat{\mathbf{B}}_{\mathbf{z}}\mathbf{x}$ = estimator of mean model obtained using the coefficients
$$\text{of } \hat{y}(\mathbf{x},\mathbf{z}).$$

$\mathbf{V}$ = diagonal matrix with j^{th} diagonal element $1/c_j^2$.

$\hat{\sigma}_{YB|\mathbf{z}}^2 = (\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x}) + \hat{\sigma}^2$ = biased estimator of the variance model
$$\text{obtained using the coefficients of } \hat{y}(\mathbf{x},\mathbf{z}).$$

$\hat{\sigma}_{Y|\mathbf{z}}^2 = (\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x}) + \hat{\sigma}^2[1 - trace(\mathbf{V}\mathbf{C})]$ = unbiased estimator of the
$$\text{variance model obtained using the coefficients of } \hat{y}(\mathbf{x},\mathbf{z}).$$

$\underset{\mathbf{s}}{E}(\cdots)$ = the expectation of the quantity in the brackets with respect to $\mathbf{s}$, the vector of
sample observations.

$\underset{\mathbf{s}}{\text{var}}(\cdots)$ = the variance of the quantity in the brackets with respect to $\mathbf{s}$.

$\underset{\mathbf{e}}{E}(\cdots)$ = the expectation of the quantity in the brackets with respect to $\mathbf{e}$, the vector of
experiment error.

$\underset{\mathbf{e}}{\text{var}}(\cdots)$ = the variance of the quantity in the brackets with respect to $\mathbf{e}$.

$E(\cdots)$ = the unconditional expectation of the quantity in the brackets.

$\text{var}(\cdots)$ = the unconditional variance of the quantity in the brackets.

$\mathbf{x}_C = (1, x_1, \ldots, x_k, x_1^2, \ldots, x_k^2, x_1 x_2, \ldots, x_{k-1} x_k)'$

$\mathbf{X}$ = design matrix expanded to the form of the response model with columns arranged in the order

$$(1, x_1, \ldots, x_k, x_1^2, \ldots, x_k^2, x_1 x_2, \ldots, x_{k-1} x_k, z_1, x_1 z_1, \ldots, x_k z_1, \ldots, z_n, x_1 z_n, \ldots, x_k z_n).$$

$\mathbf{M}_C$ = the square matrix obtained by deleting the last $n + nk$ columns and rows of $\mathbf{X'X}$.

$\mathbf{V}_C$ = the square matrix obtained by deleting the last $n + nk$ columns and rows of $(\mathbf{X'X})^{-1}$. In the case of an MRD design, $\mathbf{V}_C = \mathbf{M}_C^{-1}$.

$\mathbf{V}_D$ = the square matrix obtained from the elements indexed by the last $n + nk$ rows and columns of $(\mathbf{X'X})^{-1}$.

ϕ_{2j} = the excess kurtosis of the distribution of the jth noise variable.

$dfSSE$ = number of residual degrees of freedom.

$$\mathbf{w} = \left[\frac{(\mu_1 - \hat{\mu}_1)}{c_1 \sigma_1}, \frac{(\mu_2 - \hat{\mu}_2)}{c_2 \sigma_2}, \ldots, \frac{(\mu_n - \hat{\mu}_n)}{c_n \sigma_n} \right]'.$$

$$M_S = (\gamma + \Delta' \mathbf{x})' \operatorname{var}(\mathbf{w})(\gamma + \Delta' \mathbf{x}).$$

$$M_S = \sum_{j=1}^{n} \left[\frac{1}{c_j^2 m_j} \left(\gamma_j + \sum_{i=1}^{k} \delta_{ij} x_i \right)^2 \right] \text{ when each } \hat{\mu}_j \text{ is the sample mean.}$$

$$M_E = \mathbf{x'}_C \mathbf{V}_C \mathbf{x}_C \sigma^2.$$

$$V_S = \sum_{j=1}^{n} \frac{1}{c_j^4} \left(\gamma_j + \sum_{i=1}^{k} \delta_{ij} x_i \right)^4 \operatorname{var}\left(\frac{\hat{\sigma}_j^2}{\sigma_j^2} \right).$$

$$V_S = \sum_{j=1}^{n} \frac{1}{c_j^4} \left(\gamma_j + \sum_{i=1}^{k} \delta_{ij} x_i \right)^4 \left(\frac{2}{m_j - 1} + \frac{\phi_{2j}}{m_j} \right) \text{ when each } \hat{\sigma}_j^2 \text{ is the sample variance.}$$

$$V_E = 2\sigma^4 \left[\sum_{j=1}^{n} \sum_{l=1}^{n} C_{jl}^2 / (c_j^2 c_l^2) + \frac{1}{dfSSE} \left(1 - \sum_{j=1}^{n} C_{jj} / c_j^2 \right)^2 \right] + 4\sigma^2 \sum_{j=1}^{n} \left[\frac{C_{jj}}{c_j^4} \left(\gamma_j + \sum_{i=1}^{k} \delta_{ij} x_i \right)^2 \right]$$

$$+ 8\sigma^2 \sum_{l=2}^{n} \sum_{j=1}^{l-1} \frac{1}{c_j^2 c_l^2} E\left(\frac{\hat{\sigma}_l}{\sigma_l} \right) E\left(\frac{\hat{\sigma}_j}{\sigma_j} \right) \left(\gamma_j + \sum_{i=1}^{k} \delta_{ij} x_i \right) \left(\gamma_l + \sum_{i=1}^{k} \delta_{il} x_i \right) C_{jl}.$$

$$\overline{V}_E = 2\sigma^4 \left[\sum_{j=1}^{n} \sum_{l=1}^{n} C_{jl}^2 / (c_j^2 c_l^2) + \frac{1}{dfSSE} \left(\sum_{j=1}^{n} C_{jj} / c_j^2 \right)^2 \right] + 4\sigma^2 \sum_{j=1}^{n} \left[\frac{C_{jj}}{c_j^4} (\gamma_j + \sum_{i=1}^{k} \delta_{ij} x_i)^2 \right]$$

$$+ 8\sigma^2 \sum_{l=2}^{n} \sum_{j=1}^{l-1} \frac{1}{c_j^2 c_l^2} E\left(\frac{\hat{\sigma}_l}{\sigma_l} \right) E\left(\frac{\hat{\sigma}_j}{\sigma_j} \right) (\gamma_j + \sum_{i=1}^{k} \delta_{ij} x_i)(\gamma_l + \sum_{i=1}^{k} \delta_{il} x_i) C_{jl}.$$

h_{1j} = the cost of making one observation on the jth noise variable.

h_2 = the cost of performing one experiment run.

K = the available budget/ time for the particular experiment under consideration.

r_f = the number of factorial replicates in an MRD design.

r_a = the number of axial point replicates in an MRD design.

r_c = the number of center points in an MRD design.

φ = objective function in resource allocation.

$$IVV = \int_R \text{var}(\hat{\sigma}_{Y|\mathbf{z}}^2 - \hat{\sigma}^2) d\mathbf{x} / \int_R d\mathbf{x}.$$

p = the number of model coefficients in the response model.

$$F_j = \int_R (\gamma_j + \sum_{i=1}^{k} \delta_{ij} x_i)^4 d\mathbf{x} / \int_R d\mathbf{x}.$$

$$G = \int_R (1 + \sum_{i=1}^{k} x_i^2)^2 d\mathbf{x} / \int_R d\mathbf{x}.$$

$$H_j = \int_R (\gamma_j + \sum_{i=1}^{k} \delta_{ij} x_i)^2 (1 + \sum_{i=1}^{k} x_i^2) d\mathbf{x} / \int_R d\mathbf{x}.$$

$$IVM = \int_R \text{var}(\hat{\mu}_{Y|\mathbf{z}}) d\mathbf{x} / \int_R d\mathbf{x}.$$

$$IM_E / \sigma^2 = \int_R \mathbf{x'}_C \mathbf{V}_C \mathbf{x}_C d\mathbf{x} / \int_R d\mathbf{x}.$$

$$\mathbf{\mu}_R = \int_R \mathbf{x}_C \mathbf{x'}_C d\mathbf{x} / \int_R d\mathbf{x}.$$

$$\mathbf{x}_1 = (1, x_1, x_2, ..., x_k)'.$$

$$\mathbf{x}_2 = (x_1^2, x_2^2, \ldots, x_k^2, x_1 x_2, x_1 x_3, \ldots, x_{k-1} x_k)'.$$

$$\boldsymbol{\mu}_{11} = \int_R \mathbf{x}_1 \mathbf{x}_1' / \int_R d\mathbf{x}.$$

$$\boldsymbol{\mu}_{22} = \int_R \mathbf{x}_2 \mathbf{x}_2' / \int_R d\mathbf{x}.$$

$$\boldsymbol{\mu}_{12} = \int_R \mathbf{x}_1 \mathbf{x}_2' / \int_R d\mathbf{x}.$$

$$R_1 = \{(x_1, \ldots, x_k); x_1^2 + \cdots + x_k^2 \leq \rho^2\}.$$

$$R_2 = \{(x_1, \ldots, x_k); -1 \leq x_i \leq 1, i = 1, \ldots, k\}.$$

$$E_j = \int_R \left(\gamma_j + \sum_{i=1}^k \delta_{ij} x_i \right)^2 d\mathbf{x} / \int_R d\mathbf{x}.$$

α = axial point distance for MRD design.

$\boldsymbol{\Lambda}$ = vector representing $\boldsymbol{\gamma}$, $\boldsymbol{\Delta}$, and σ^2.

$\underset{\Lambda}{E}(\cdots)$ = the expectation of the quantity in the brackets with respect to $\boldsymbol{\Lambda}$.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1   Introduction

The means and covariances of the noise variables are important information in the design and analysis of experiments for robust parameter design. These parameters are the basis with which the levels of the noise variables are set in the experiment. In addition, they are also used in the estimation of the mean and variance models. In practice, the means and covariances of noise variables are often not known with certainty. In some cases, they can be estimated with field data whereas in others, the engineer has to guess the values of the parameters.

However, in the robust parameter design literature, the means and covariances of the noise variables are typically assumed known. This ignores the possibility that standard experimentation and estimation of the mean and variance models can produce results that are seriously in error if the means and covariances of the noise variables are badly estimated. For existing processes, data can be collected to estimate the means and covariances of the noise variables. In this case, the effect of variability in the process data on the estimation of the mean and variance models must be explicitly taken into account in the development of a statistical estimation procedure. In addition, to ensure that the best results are obtained with the available resource, the data collection effort and experiment must be planned in an integrated way. Very little has

been done in these directions. In this thesis, we attempt to fill this gap. We propose a procedure for estimating the mean and variance models that integrates planning of the combined array experiment with planning of the estimation of the means and covariances of the noise variables. Within the framework of the procedure, we treat the problems of estimation of the mean and variance models, and the design of the data collection and experiment plans to optimize the estimation of the models.

The remaining parts of this chapter are organized as follows. The next section introduces robust parameter design. In Section 1.3, we review the literature on experimental designs for robust parameter design; in Section 1.4, we review the literature on the statistical analysis of experiments for robust parameter design. Section 1.5 presents the widely accepted theoretical framework for the estimation of the mean and variance models with a combined array experiment, which assumes that the means and covariances of the noise variables are known. Lastly, Section 1.6 highlights the extensions made by this research to the framework given in Section 1.5 and outlines the structure of this thesis.

1.2   Robust Parameter Design

Robust parameter design (RPD), as it was originally introduced by Taguchi, is a quality improvement methodology based on design of experiments for designing products and processes that are insensitive to variation in a set of variables, called noise variables. Noise variables can usually be controlled during experimentation but not during process operation or product use. Examples include deviations from the nominal values of process variables, variation in raw material properties, variation in tooling geometry, in-plant environmental factors such as humidity and variables

representing customer use conditions (Abraham and MacKay, 1993). On the other hand, control variables are variables whose values are under the control of the process or product designer. The objective of robust parameter design is to find settings of the control variables to neutralize the variability in one or more responses caused by the noise variables and to optimize the responses. This objective relates to Taguchi's quality philosophy, which advocates the minimization of "loss to society" due to deviations of a quality characteristic from its target value (Taguchi et al., 1993). Although the use of statistical design of experiments has been the focus in robust parameter design, awareness of the need to reduce variation by creating insensitivity to noise variables has led to various other methods to achieve this objective (Arvidsson and Gremyr, 2007).

Taguchi not only introduced the concept of robust parameter design, but also experimental designs and analysis methods to achieve the desired objectives (see for example, Taguchi et al. (1993)). However, as pointed out by many authors (for example, Bisgaard, 1996; Myers et al., 1992; Box, 1988), his designs and analysis methods are generally not statistically sound. This led to much research into alternative designs and accompanying analysis approaches that are theoretically better than those proposed by Taguchi. As can be seen in the recent review of the robust parameter design literature by Robinson et al. (2004), modeling of the variance of the response, optimization methods for finding robust solutions, and designs that accommodate both control and noise variables have received the bulk of attention from researchers.

1.3   Experimental Designs for Robust Parameter Design

The designs introduced by Taguchi for RPD experiments are called crossed array designs. A crossed array design consists of a chosen orthogonal array for the control variables, called the inner array, crossed with a chosen orthogonal array for the noise variables, called the outer array. Many degrees of freedom are used to estimate unimportant higher order interactions between the control and noise variables in these designs (Shoemaker et al., 1991). Although heavily fractionated orthogonal arrays in which control x control interactions are confounded with the main effects of the control variables are often used, many of the designs are still uneconomically large (Myers and Montgomery, 2002). This leads to two criticisms of Taguchi's crossed array designs: uneconomical design size and inability to estimate control x control interactions (Myers et al., 1992). However, Shoemaker et al. (1991) point out that the crossed arrays provide some protection against modeling difficulties since they allow direct estimation of a performance measure such as the sample variance at each combination of control variable settings in the inner array. The recent comparison of crossed and combined arrays in a physical experiment by Kunert et al. (2007) illustrates the importance of this built-in robustness to modeling problems.

An alternative to Taguchi's crossed arrays is the combined array designs, which are designs that accommodate both control and noise variables (Shoemaker et al., 1991). Combined arrays are response surface designs such as the central composite designs or computer generated alphabetic optimal designs that allow estimation of all terms in a regression model that contains both control and noise variables (Myers and Montgomery, 2002). Frequently, a model that contains up to second order terms in the control variables, linear terms in the noise variables, and terms representing control x

noise interactions is assumed. The mixed resolution (MRD) designs are a class of combined array designs specifically introduced to estimate models of this form (Borror and Montgomery, 2000; Borkowski and Lucas, 1997). Advantages of the MRD over Taguchi's crossed arrays include control x control interactions that are estimated clear of main effects and control x noise interactions, and a design size that is usually smaller (Borror and Montgomery, 2000; Borkowski and Lucas, 1997). The MRD design also has superior variance properties to most other combined array designs (Borror et al., 2002). However, MRD designs may not be optimal with respect to a specific alphabetic criterion. Alphabetic optimal designs would be desirable if the aim of the experiment is to achieve a specific inference objective such as estimation of a subset of model parameters (Silvey, 1980). Ginsburg and Ben-Gal (2006) show how designs that minimize the variance of the estimated minimum-loss control variable settings can be constructed.

Split-plot designs are another class of designs that are useful for RPD experiments (Box et al., 2005; Box and Jones, 1992). In split-plot designs, a set of factors is placed in the whole-plot and another set is placed in the subplot. Whole-plot treatments are randomly assigned to experiment units and corresponding to each whole-plot treatment, subplot treatments are randomly assigned.

Depending on the manner in which a crossed array design is run, it can be a combined array design or a split-plot design. If a crossed array is fully randomized, it is a combined array design. The structure of crossed arrays, however, suggests that they are often run as split plot designs.

1.4   Statistical Analysis of Experiment Data

Data from a crossed array can be analyzed based on summary measures computed at each combination of control variable levels in the inner array. Taguchi advocates the use of quantities called signal-to-noise ratios as summary measures. Different signal-to-noise ratios are defined for problems in which the objective is to keep the response on target, as large as possible or as small as possible (Myers and Montgomery, 2002). Use of the signal-to-noise ratios for the latter two cases can be very inefficient (Box, 1988). Furthermore, use of the signal-to-noise ratios for the objective of achieving a target value can only be justified with the assumption of specific types of underlying models (Leon et al., 1987). As alternatives to Taguchi's signal-to-noise ratios, Box (1988) proposes the use of transformations based on the observed data. Leon et al. (1987) propose the use of criteria derived from an assumed model for the response that they call performance-measures-independent-of-adjustment.

A better method of analyzing fully randomized crossed array designs is to fit a single model relating the response to both control and noise variables. The resulting model is called a response model (Shoemaker et al., 1991). For combined array designs that are not crossed arrays, analysis with summary measures is not possible and fitting a response model is the appropriate analysis method (Wu and Hamada, 2000). When the residual variance is constant, the response model should be fitted with least squares. However, when the residual variance is not constant, generalized linear modeling methods should be used (Robinson et al., 2004). Myers (1991) and Myers et al. (1992) show how mean and variance models can be derived and estimated. The problem of simultaneous optimization of the mean and variance models has received

considerable attention in the literature (for example, see Koksoy and Doganaksoy (2003) and Lawson and Madrigal (1994)). Various formulations of the problem and solution methods have been proposed to find a solution that achieves a desirable tradeoff between the objective for the mean and the objective for the variance.

Steinberg and Bursztyn (1998) demonstrate that explicit modeling of the noise variables in a response model can lead to significant increases in power of detecting dispersion effects over the summary measure modeling approach. Another advantage of response model fitting over the use of summary measures is that it provides the experimenter an opportunity to better understand the system through examination of control x noise interaction plots (Wu and Hamada, 2000; Shoemaker et al., 1991).

Appropriate analysis methods for split-plot designs are discussed by Box et al. (2005), and Myers and Montgomery (2002). These take into account the error structure of a split plot experiment, which consists of a whole plot error and a subplot error.

1.5 Estimation of the Mean and Variance Models with a Combined Array Experiment: The Dual Response Surface Approach

The objectives of robust parameter design can be achieved by estimating the mean and variance models and then optimizing the process or product based on the estimated models. To estimate the mean and variance models with a combined array experiment in the case where the mean μ and covariance matrix Σ of the noise variables are known, the experimenter follows the standard procedure given in Figure 1.1. This procedure is based on the procedures given by Montgomery (2005b), Khuri and Cornell (1996), and Leon et al. (1993).

Step 1: Selection of the response, control variables, and noise variables.

Step 2: Choice of levels of the control variables that are allowable for the experiment.

Step 3: Choice of levels of the noise variables that are allowable for the experiment.

Step 4: Selection of the design matrix.

Step 5: Execution of the experiment.

Step 6: Estimation of the mean and variance models.

Figure 1.1: Standard Procedure for Estimating the Mean and Variance Models with a Combined Array Experiment: Known μ and Σ

Step 1 is assumed the responsibility of the experimenter, who should use her engineering or process knowledge to make the decisions. In Step 2, the experimenter determines the region of the control variables within which experiment runs may be made. In Step 3, the experimenter determines the region of the noise variables within which experiment runs may be made. Common practice in the literature is to specify the region for the noise variables based on the means and variances of those variables (see Equation (1.2) below). Assuming that the regions for the control and noise variables can be specified independently, the Cartesian product of the regions will give the design space (Silvey, 1980). After the design space is specified, a design is obtained by choosing design points from the design space. Many papers in the literature, such as Borror et al. (2002), discuss designs for Step 4. At this point in our discussion, there are two things to note. Firstly, there is really no precedence relationship between Steps 2 and 3. Secondly, the procedure for choosing a design, specifically Steps 2 to 4 discussed above, is based on the formulation of the design problem in optimal design theory. An alternative formulation of the design problem is presented by Box and Draper (1987). In their formulation, there are two distinct types

of regions: the region of operability and the region of interest. The experimenter is not expected to explicitly specify her region of interest. Rather, the experimenter is supposed to choose a design and the corresponding levels of the factors at the design points based on various considerations, one of which is her interest in predicting at various points. This formulation, however, shall not be adopted in this thesis.

In Step 6, the response is assumed a function of the control and noise variables plus a term representing the contribution of unknown causes of variation. This model, called the response model, is assumed to hold under conditions of process operation or product use in addition to the conditions of the experiment. The commonly assumed form of the response model is given by (Myers et al., 2004; Robinson et al., 2004)

$$y(\mathbf{x},\mathbf{q}) = \beta_0 + \mathbf{x}'\boldsymbol{\beta} + \mathbf{x}'\mathbf{B}\mathbf{x} + \boldsymbol{\gamma}'\mathbf{q} + \mathbf{x}'\boldsymbol{\Delta}\mathbf{q} + \varepsilon, \tag{1.1}$$

where $\mathbf{x}$ is the $k \times 1$ vector of control variables in coded units; $\mathbf{q}$ is the $n \times 1$ vector of noise variables in coded units; β_0, $\boldsymbol{\beta}$, $\mathbf{B}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\Delta}$ are the coefficients of the model and ε is a random variable representing residual variation, which is assumed to have mean zero and constant variance σ^2.

Let $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_n)'$ denote the levels of the noise variables in un-coded units. Common practice in the literature (Miro-Quesada and Del Castillo, 2004; Myers and Montgomery, 2002; Myers et al., 1997) is to assume that the vector $\mathbf{q}$ in Equation (1.1) is given by

$$\mathbf{q} = \left[\frac{(\xi_1 - \mu_1)}{c_1\sigma_1}, \frac{(\xi_2 - \mu_2)}{c_2\sigma_2}, ..., \frac{(\xi_n - \mu_n)}{c_n\sigma_n} \right]', \tag{1.2}$$

where $c_j, j = 1, ..., n$ are the scaling factors, and μ_j and σ_j are the mean and standard deviation of the jth noise variable respectively. This assumes that all noise variables are continuous.

9

Although the noise variables are held fixed in each experiment run, they are random in actual process operation or product use. Let $\mathbf{Q}$ denote the random vector of the noise variables in the coded units $\mathbf{q}$. Substituting $\mathbf{Q}$ for $\mathbf{q}$ in (1.1) and taking expectation with respect to $\mathbf{Q}$ and the residual error ε, we obtain the mean model

$$\mu_Y = \beta_0 + \mathbf{x'\beta} + \mathbf{x'Bx}. \tag{1.3}$$

Similarly, substituting $\mathbf{Q}$ for $\mathbf{q}$ in (1.1) and applying the variance operator with respect to $\mathbf{Q}$ and ε, we obtain the variance model

$$\sigma_Y^2 = (\mathbf{\gamma} + \mathbf{\Delta'x})' \, \mathrm{var}(\mathbf{Q})(\mathbf{\gamma} + \mathbf{\Delta'x}) + \sigma^2, \tag{1.4}$$

where it is assumed that ε is independent of $\mathbf{Q}$ and $\mathrm{var}(\mathbf{Q})$ is the covariance matrix of $\mathbf{Q}$, which is assumed known.

The validity of (1.4) as a model for the variance of the response rests on the assumption that the only sources of heterogeneity of variance (dependence of the variance of the response on $\mathbf{x}$) are the noise variables represented by $\mathbf{Q}$ (Myers and Montgomery, 2002). This assumption is implicit in the assumption that ε has constant variance.

Having performed the experiment, the response model can be fitted with ordinary least squares to give the fitted response model

$$\hat{y}(\mathbf{x},\mathbf{q}) = \hat{\beta}_0 + \mathbf{x'}\hat{\mathbf{\beta}} + \mathbf{x'}\hat{\mathbf{B}}\mathbf{x} + \hat{\mathbf{\gamma}}'\mathbf{q} + \mathbf{x'}\hat{\mathbf{\Delta}}\mathbf{q}. \tag{1.5}$$

An estimator for the mean model $\hat{\mu}_Y$ is obtained by replacing the unknown coefficients in (1.3) with the corresponding least squares estimates in (1.5), giving

$$\hat{\mu}_Y = \hat{\beta}_0 + \mathbf{x'}\hat{\mathbf{\beta}} + \mathbf{x'}\hat{\mathbf{B}}\mathbf{x}. \tag{1.6}$$

Similarly, an estimator of the variance model $\hat{\sigma}_{YB}^2$ is obtained by replacing the unknown coefficients in (1.4) with the corresponding least squares estimates in (1.5) and σ^2 with the residual mean square $\hat{\sigma}^2$, giving

$$\hat{\sigma}_{YB}^2 = (\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\Delta}}'\mathbf{x})' \text{var}(\mathbf{Q})(\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\Delta}}'\mathbf{x}) + \hat{\sigma}^2. \tag{1.7}$$

The estimator $\hat{\mu}_Y$ is an unbiased estimator of μ_Y. However, $\hat{\sigma}_{YB}^2$ is a biased estimator of σ_Y^2 (hence, the subscript). To obtain an unbiased estimator of σ_Y^2, a biased correction term (Myers and Montgomery, 2002) is subtracted from (1.7) to give

$$\hat{\sigma}_Y^2 = (\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\Delta}}'\mathbf{x})' \text{var}(\mathbf{Q})(\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\Delta}}'\mathbf{x}) + \hat{\sigma}^2 \{1 - trace[\text{var}(\mathbf{Q})\mathbf{C}]\}, \tag{1.8}$$

where $\mathbf{C} = \text{var}(\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\Delta}}'\mathbf{x})/\sigma^2$.

The idea of estimating the mean and variance models with the above equations seems to have been first discussed by Myers (1991) and Myers et al. (1992). O'Donnell and Vining (1997) derive the bias and variance of the biased estimator of the variance model. The unbiased estimator of the variance model is recommended by Myers and Montgomery (2002) and Miro-Quesada and Del Castillo (2004).

The approach introduced above for estimating the mean and variance models is called the dual response surface approach (Myers et al., 1992). Several other papers address specific issues in this approach. Myers et al. (1997) discuss the construction of a confidence region for the minimum variance point, a prediction interval for a future response, and one-sided tolerance intervals. Brenneman and Myers (2003) introduce the use of the multinomial distribution as a model for categorical noise variables. An extension to the case of multiple responses is presented by Romano et al. (2004). Miro-Quesada and Del Castillo (2004) discuss a method for specifying the scaling factors. They also introduce a new objective function for finding robust settings, which is said to be robust to errors in estimating the model coefficients. Although the above papers consider various aspects of the dual response surface approach, they assume that the means and covariance matrix of the noise variables are known.

1.6   Outline of Research and Organization of Thesis

In the discussion of the dual response surface approach in Section 1.5, the mean $\mathbf{\mu}$ and covariance matrix $\mathbf{\Sigma}$ (in un-coded units) of the noise variables are assumed known. However, in practice, $\mathbf{\mu}$ and $\mathbf{\Sigma}$ are frequently not known. Variations in the settings of process variables such as fluctuations in the conveyor speed of a wave soldering process may never be recorded. In some cases, measurement of certain quality characteristics can also be costly so that measurements are seldom made. For instance, measuring the various dimensions of a geometrically complicated component may require the use of a Coordinate Measuring Machine and therefore, measurements may be made only when a quality problem is suspected.

The unknown parameters $\mathbf{\mu}$ and $\mathbf{\Sigma}$ are often estimated with process data. Sampling from the process to obtain information about the distributions of the noise variables is well suited for robust design of existing products and processes. For example, in the case studies presented by Radson and Herrin (1995), O'Neill et al. (2000), Shore and Arad (2003), and Dasgupta (2007), information on the distribution of the noise variables was obtained by taking samples of observations on those variables.

When the means and covariances of the noise variables are estimated with data sampled from the process, the levels of the noise variables and estimated mean and variance models are affected by sampling error. Many issues associated with the estimation of the mean and variance models in this situation have not been addressed. In particular, the statistical properties of the estimators for the mean and variance models have not been generalized to take into account sampling variation. Furthermore, the need for simultaneous planning of the sampling effort and experiment so that total

resource is allocated to achieve efficient estimation of the mean and variance models has not been recognized. In this thesis, we examine these problems. We propose a procedure for estimating the mean and variance models that incorporates estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with sampled data. The procedure integrates planning of sample data collection with planning of the combined array experiment to achieve the best possible estimation of the mean and variance models. Novel ideas introduced with the procedure are developed in this thesis. In particular, we address the issues of specification of the levels of the noise variables, estimation of the mean and variance models, repeated sampling properties of the estimators, and optimal allocation of resource to sampling and experimenting. This research is motivated by the suggestions of Miro-Quesada and Del Castillo (2004) and Myers et al. (1997) for further research into the problem where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are replaced with estimates.

The remainder of this thesis is organized as follows. Chapter 2 presents the proposed procedure for estimating the mean and variance models. A method for specifying the levels of the noise variables based on estimates for the means and variances of those variables is proposed. The true means and variances of the noise variables are replaced with estimates in deriving estimators for the mean and variance models. The effect of sampling error, the bias and variances of the estimators, and the increase in the variances due to sampling error are investigated.

Chapter 3 examines the problem of optimal allocation of resource to sampling and experimenting for the case where the specified design is an MRD. We call a combination of sample sizes and a design a scheme, and mathematical programs are formulated to find optimal schemes. Two different objective functions are considered. One is the average variance of the unbiased estimator for the variance model minus the residual mean square, which is a measure of the performance of a scheme at estimating

the variance model. The other is the average variance of the estimator for the mean model, which is a measure of the performance of a scheme at estimating the mean model. The sample sizes, and number of factorial, axial, and center point replicates of the MRD are taken as decision variables. A method for finding schemes that compromise between the optimization of the two objective functions is also discussed. In the last part of the chapter, an algorithm for finding schemes that perform well with respect to the two objectives given a candidate set of design points is introduced.

Chapter 4 suggests solutions to two problems in the optimal allocation of resource. Values of some of the parameters in the response model must be known or estimated if the mathematical programs given in Chapter 3 are to be used. Methods proposed in the literature of nonlinear experimental design to solve the problem of dependence of optimal designs on model parameters are reviewed and their application to the problem of specifying the unknown parameters in the response model is discussed. The mathematical programs given in Chapter 3 are modified to allow the use of prior distributions for the unknown parameters. In addition, a discussion of how uncertainty in model parameters may be handled using ideas from the robust optimization literature is given. The second problem examined in this chapter is the comparison of schemes with designs other than the MRD. For this problem, plots called cumulative distribution plots, which are based on the FDS plots introduced by Zahran et al. (2003), are proposed for comparing schemes.

CHAPTER 2

ESTIMATION OF THE MEAN AND VARIANCE MODELS WHEN MEANS AND VARIANCES OF THE NOISE VARIABLES ARE UNKNOWN

2.1   Introduction

This chapter presents the procedure developed in this research for estimating the mean and variance models. We describe the proposed procedure, which is a modification of the standard procedure presented in Figure 1.1. In order to develop various aspects of the proposed procedure, we make a number of assumptions, which we state explicitly. Two aspects of the proposed procedure that differ from the standard procedure are discussed in this chapter. Firstly, the problem of specifying the levels of the noise variables based on estimates of the means and variances of those variables is addressed. Secondly, estimation of the mean and variance models is examined. The effect of errors in estimating the means and variances of the noise variables on the estimated mean and variance models is investigated. Formulas for the mean squared error of the estimators for the mean and variance models are derived. It is demonstrated that a large part of the variability of the estimators can be due to variability in data sampled from the process.

2.2 Proposed Procedure for Estimating the Mean and Variance Models

We propose the procedure given in Figure 2.1 for estimating the mean and variance models. The main advantage of using this procedure is that it allows for an integrated planning of the experiment and process data collection.

Step 1: Selection of the response, control variables, and noise variables.

Step 2: Specification of the set of coded levels of the control variables from which design points are to be chosen and the corresponding set of un-coded levels.

Step 3: Specification of the scaling factors and the set of coded levels of the noise variables from which design points are to be chosen.

Step 4: Specification of design type/points and optimization of proposed criteria to determine sample sizes and design matrix.

Step 5: Estimation of the means and variances of the noise variables with process data.

Step 6: Computation of the un-coded levels of the noise variables for each experiment run.

Step 7: Execution of the experiment.

Step 8: Estimation of the mean and variance models.

Figure 2.1: Proposed Procedure for Combined Array Experiment

Step 1 in this procedure is identical to Step 1 in the standard procedure in Figure 1.1. The purpose of Steps 2 and 3 is to specify the design space. Denote the coded levels of the control variables by $\mathbf{x}$, and the coded levels of the control variables in the lth design run by $\mathbf{x}_l, l = 1, \ldots, N$. Define R, the design region for the control variables as the set of vectors $\mathbf{x}$ such that $\mathbf{x}_l \in R, l = 1, \ldots, N$ for all permissible design matrices. In Step 2, $\mathbf{x}$ and R are specified. In contrast to the control variables, we fix the coded levels of the noise variables in the design matrix and allow the process data

to determine the corresponding un-coded levels through the coding. In particular, we fix the coding for the noise variables as

$$\mathbf{z} = \left[\frac{(\xi_1 - \hat{\mu}_1)}{c_1 \hat{\sigma}_1}, \frac{(\xi_2 - \hat{\mu}_2)}{c_2 \hat{\sigma}_2}, \ldots, \frac{(\xi_n - \hat{\mu}_n)}{c_n \hat{\sigma}_n} \right]' , \tag{2.1}$$

where $\hat{\mu}_j$ is the jth element of $\hat{\mu}$, an estimator for μ and $\hat{\sigma}_j^2$ is the jth diagonal element of $\hat{\Sigma}$, an estimator for Σ. Denote the coded levels of the noise variables in the lth run by $\mathbf{z}_l$, where $l = 1, \ldots, N$ and define S, the design region for the noise variables, as the set of vectors $\mathbf{z}$ such that $\mathbf{z}_l \in S, l = 1, \ldots, N$, for all permissible design matrices. In Step 3, the design region S, and the scaling factors $c_j, j = 1, \ldots, n$ in Equation (2.1) are specified. Note that although specification of $\mathbf{x}$ and R is labeled as Step 2 while specification of S and $c_j, j = 1, \ldots, n$ is labeled as Step 3, there is really no precedence relationship between the two steps.

Step 4 calls for the design matrix to be specified together with the sample size for each noise variable $m_j, j = 1, \ldots, n$. The design matrix is to be assembled from design points chosen from the design space, which is the Cartesian product of R and S. Observe that the proposed procedure calls for simultaneous consideration of the process data collection and experiment effort. This is desirable because it would then be possible to plan the allocation of effort between the two activities in an optimal way. We shall introduce tools to aid the specification of the design and sample sizes such that estimation of the mean and variance models is optimized. In Step 5, process data collection, which we also call sampling, is carried out. This involves making m_j observations on the jth noise variable.

Steps 3- 5 imply that the design matrix is to be specified before any observations on the noise variables are taken. Therefore, at the point after the design

matrix is specified and before any observations on the noise variables are taken, the un-coded levels of the noise variables for the lth experiment run $\boldsymbol{\xi}_l$ is a random vector given by

$$\boldsymbol{\xi}_l = (\xi_{l1},\ldots,\xi_{ln})' = T(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}, \mathbf{z}_l) = (\hat{\mu}_1 + z_{l1}c_1\hat{\sigma}_1,\ldots,\hat{\mu}_n + z_{ln}c_n\hat{\sigma}_n)', \tag{2.2}$$

where z_{lj} is the jth element of $\mathbf{z}_l$. In addition, observe that S_{ξ}, the region obtained by mapping all points $\mathbf{z}$ in S via the transformation $T(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}, \mathbf{z})$, is random. In Step 6, the un-coded levels of the noise variables for each experiment run are determined through Equation (2.2). This is followed by the execution of the experiment, which is Step 7. In Step 8, the mean and variance models are estimated with data from the experiment.

The proposed procedure is a modification of the standard procedure. Steps 3-4 in the standard procedure are replaced with Steps 3-6 in the proposed procedure. In Step 3 of the standard procedure, both the sets of coded and un-coded levels of the noise variables are specified based on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. This is followed by the construction of the design matrix. Thus, the un-coded levels of the noise variables for the experiment runs do not depend on process data. Another difference between the standard procedure and the proposed procedure is that Step 8 of the proposed procedure involves the use of a theoretically different set of estimators than that used in the standard procedure.

Step 3 and Step 8 of the proposed procedure are discussed in this chapter. Step 4, which is the design step, is treated at length in the next two chapters.

2.2.1 Assumptions

In this section, assumptions that are made throughout this research are stated.

Unless stated otherwise, these assumptions apply wherever they are relevant.

Assumption 2.1 All noise variables are continuous.

Remark: The method of specifying the levels of the noise variables described in the preceding section necessarily requires that this assumption be made. If the noise variables are not continuous, the experimenter may not be able to fix the levels of the noise variables according to (2.2).

Assumption 2.2 Let ω be the union of all possible realizations of S_ξ and let Ω be the set of ξ over which the joint density of the noise variables is non-zero. We assume that for $\xi \in \omega \cup \Omega$ and $\mathbf{x} \in R$, the response model is given by

$$y(\mathbf{x}, \xi) = \beta_{0\xi} + \mathbf{x}'\boldsymbol{\beta}_\xi + \mathbf{x}'\mathbf{B}_\xi\mathbf{x} + \boldsymbol{\gamma}'_\xi\xi + \mathbf{x}'\boldsymbol{\Delta}_\xi\xi + \varepsilon, \tag{2.3}$$

where ε has mean zero and variance σ^2, and $\beta_{0\xi}$, $\boldsymbol{\beta}_\xi$, $\mathbf{B}_\xi$, $\boldsymbol{\gamma}_\xi$, and $\boldsymbol{\Delta}_\xi$ are the model coefficients.

Remark: Note that the response model is written as a function of the coded form of the control variables $\mathbf{x}$ and the un-coded form of the noise variables ξ. The response model given in (2.3) is equivalent to that given by (1.1) since (2.3), when rewritten in the variables $\mathbf{x}$ and $\mathbf{q}$, is of the form given in (1.1). Observe that if the response model given in (2.3) holds for each $\xi \in \Omega$ and $\mathbf{x} \in R$, the true mean and variance models are given in (1.3) and (1.4) respectively. On the other hand, if the response model given in (2.3) holds for each $\xi \in \omega$ and $\mathbf{x} \in R$, the same response model will fit the experiment data without any bias due to model inadequacy. Thus, this assumption implies that the response for the lth experiment run is given by

$$y(\mathbf{x}_l, \xi_l) = \beta_{0\xi} + \mathbf{x}'_l\boldsymbol{\beta}_\xi + \mathbf{x}'_l\mathbf{B}_\xi\mathbf{x}_l + \boldsymbol{\gamma}'_\xi\xi_l + \mathbf{x}'_l\boldsymbol{\Delta}_\xi\xi_l + e_l,$$

where e_l is the experiment error in the lth run. The response is a function of the random variables $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Sigma}}$, and e_l. For illustration, when $k = 1$ and $n = 1$, the response for the lth experiment run, where $(x_1, \xi_1) = (x_{l1}, \xi_{l1})$, is

$$y(x_{l1}, \xi_{l1}) = \beta_{0\xi} + \beta_{1\xi} x_{l1} + \beta_{11\xi} x_{l1}^2 + \gamma_{1\xi}(\hat{\mu}_1 + z_{l1} c_1 \hat{\sigma}_1) + \delta_{11\xi} x_{l1}(\hat{\mu}_1 + z_{l1} c_1 \hat{\sigma}_1) + e_l.$$

Assumption 2.2 appears to be too restrictive because it requires that the response model holds for each $\xi \in \omega \cup \Omega$, which may be a very large set. However, the mean model in (1.3) and the variance model in (1.4) are derived based on the assumption that the response model holds for each $\xi \in \Omega$. Furthermore, the unbiasedness of the estimators in Equations (1.6) and (1.8) are established assuming that the response model holds in $\xi \in \omega_0$ and $\mathbf{x} \in R$, where ω_0 represents the fixed experiment region for the noise variables. Therefore, Assumption 2.2 is, in fact, merely an extension of the assumption implicitly made in the dual response surface approach.

As long as $\omega \subset \Omega$, Assumption 2.2 is not more restrictive than the assumption implicit in the derivation of (1.3) and (1.4), which are the mean and variance models given in the literature (see Section 1.5). To have $\omega \subset \Omega$, the region S_ξ should be within the region of values of the noise variables that are possible to occur. This implies that for the case of independently distributed noise variables (see Assumption 2.4), the range over which each noise variable is varied in the experiment should be within the range of variation of the variable. Reasonable RPD experiments should have $\omega \subset \Omega$ so that the experiment does not study the response across values of the noise variables that never occur in practice. The case of known means and covariances of the noise variables is similar since the RPD experiment should be designed so that $\omega_0 \subset \Omega$.

In the literature, it is commonly assumed that the noise variables are normally distributed (see Assumption 2.5). Theoretically, the normal distribution has an unbounded sample space. Therefore, ω and Ω are the n-dimensional real space if it is assumed that the noise variables are normally distributed. As such, for normally distributed noise variables, we require that Equation (2.3) hold over the n-dimensional real space. However, in any particular practical setting, we cannot really expect Equation (2.3) to hold over the n-dimensional real space nor can we expect the noise variables to be perfectly normally distributed. Thus, despite Assumption 2.2, it would be inappropriate to conduct experiments over wide ranges of values of the noise variables. In the next section, we introduce a method to specify S and $c_j, j = 1,\ldots,n$ that would enable us to control the size of S_ξ.

Assumption 2.3 Each noise variable is distributed independently of the levels of the control variables and each has finite mean and variance.

Remark: This implies that the mean and variance of each noise variable exist, and they are not functions of the levels of any of the control variables.

Assumption 2.4 The noise variables are known to be independently distributed.

Remark: The assumption of independently distributed noise variables is commonly made in the literature (Myers et al., 2004). The fact that the noise variables are independent may be known by physical considerations. For example, when the noise variables are difficult-to-control process variables or raw material properties, it is reasonable to assume that they are independent (Myers et al., 2004; Borror et al., 2002). It follows logically from this assumption that $\hat{\Sigma}$ should also be diagonal.

Assumption 2.5 The noise variables are normally distributed.

Remark: The assumption of normally distributed noise variables is made in many statistical papers and case studies in the literature (for example, see Miro-Quesada et al. (2004), Jeang et al. (2007) and Li et al. (2007)). Therefore, this assumption appears to be reasonable in most cases.

Assumption 2.6 For each $j = 1, \ldots, n$, the estimators $\hat{\mu}_j$ and $\hat{\sigma}_j^2$ are defined on a random sample of size m_j. In other words, the sample observations are independent.

Remark: The assumption of random sampling may not always be valid since in some cases, the values of a noise variable over time may be auto-correlated (Jin and Ding, 2004). However, if data collection were done such that the intervals between successive observations on a noise variable are sufficiently long, the observations for the noise variable would be approximately independent (Montgomery, 2005a).

Assumption 2.7 The estimators $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are independent of the vector of experiment error $\mathbf{e}$.

Remark: Physical considerations suggest that this should be the case. Sampling and experimenting are different activities at two distinct points in time.

Assumption 2.8 The expectation of $\mathbf{e}$, the vector of experiment error, is a zero vector. The elements of $\mathbf{e}$ are independent and identically distributed, each with variance σ^2.

2.3   Specification of Levels of the Noise Variables

Step 3 of the proposed procedure calls for the design region S and the scaling factors $c_j, j = 1, \ldots, n$ to be chosen prior to sampling. This is necessary in order to have the advantage of being able to plan both the experiment and sampling simultaneously. In this section, we address the question of choosing S and $c_j, j = 1, \ldots, n$.

Consider the design of a factorial experiment with a single noise variable that is normally distributed in process operation with known mean μ_1 and known variance σ_1^2. Following common practice, the high and low levels of the noise variable may be set at $\mu_1 + c_1 \sigma_1$ and $\mu_1 - c_1 \sigma_1$ respectively for some c_1. The value chosen for c_1 should be such that the noise variable is varied over a range that is representative of its variation during actual process operation or product use. For example, it does not seem appropriate to choose $\mu_1 + 6\sigma_1$ for the high level and $\mu_1 - 6\sigma_1$ for the low level since the levels are too extreme. It also does not seem appropriate to choose $\mu_1 + 0.1\sigma_1$ for the high level and $\mu_1 - 0.1\sigma_1$ for the low level since the change in the response would be easily masked by experiment error. However, there is no rigid rule for choosing c_1. It appears that any value within the interval [1,2] are reasonable choices for c_1. Now, if μ_1 and σ_1^2 are replaced with $\hat{\mu}_1$ and $\hat{\sigma}_1^2$ respectively, selecting c_1 is not as clear. We propose considering the problem as one of constructing a tolerance region for the distribution of the noise variable with the interval $[\hat{\mu}_1 - c_1\hat{\sigma}_1, \hat{\mu}_1 + c_1\hat{\sigma}_1]$. Let τ_{II} be the proportion of the probability density of the noise variable contained by the interval on the average. Choosing τ_{II} to be moderately large is a logical way to express the rule that "a noise variable should be varied over a range that is representative of its

variation during actual process operation or product use." For instance, given the

sample size, we may choose c_1 so that τ_{II} takes the value of 0.8. This would lead to a

factorial experiment that varies the noise variable over a range that, on the average,

contains 80 percent of the distribution of the noise variable. The idea just introduced

for specifying c_1 is generalized below.

Given the sample sizes and the estimators $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$, we propose that S and

$c_j, j = 1, \ldots, n$ be specified such that S_ξ, the set of un-coded levels corresponding to S,

is a tolerance region of a reasonable size for the joint distribution of the noise variables.

Specifically, we propose that the experimenter choose S and $c_j, j = 1, \ldots, n$ so that the

expected proportion of the joint distribution contained within S_ξ is some suitable value

τ_{II}. This is called a type II tolerance region (Chew, 1966). In addition to the degree

with which S_ξ represents conditions of process operation, specification of τ_{II} for the

type II tolerance region also requires a consideration of the tradeoff between bias due

to model inadequacy and variance of the fitted response model. Hence, a value such as

0.999 for τ_{II} may not be considered appropriate for most cases, as bias due to model

inadequacy may be large.

Assuming that the noise variables are normally and independently distributed

(Assumptions 2.4 and 2.5), a type II tolerance region may be obtained by constructing

type II tolerance intervals with expected coverage of $\tau_{\mathrm{II}}^{1/n}$ for each noise variable. The

Cartesian product of the intervals gives the desired tolerance region. By a result given

by Chew (1966), a $\tau_{\mathrm{II}}^{1/n}$ type II tolerance interval for the jth noise variable is given by

the set of values ξ_j that satisfy the inequality

$$(\xi_j - \hat{\mu}_j)^2 / \hat{\sigma}_j^2 \le (m_j + 1) F(1 - \tau_{\mathrm{II}}^{1/n}, 1, m_j - 1) / m_j, \qquad (2.4)$$

where m_j is the sample size, $\hat{\mu}_j$ and $\hat{\sigma}_j^2$ are the sample mean and sample variance respectively, and $F(1-\tau_{\text{II}}^{1/n},1,m_j-1)$ is the upper $100(1-\tau_{\text{II}}^{1/n})$ percent point of the F distribution with 1 and m_j-1 degrees of freedom. Although there is no hard and fast rule for the value of τ_{II}, reasonable choices are such as 0.7, 0.8, and 0.9. Suppose that $m=m_1=\cdots=m_n$ and $S=\{(z_1,\ldots,z_n);-1\leq z_j\leq 1,j=1,\ldots,n\}$. Then, the scaling factor $c=c_1=\cdots=c_n$ that give a value of $\tau_{\text{II}}^{1/n}$ for each noise variable is

$\sqrt{(m+1)F(1-\tau_{\text{II}}^{1/n},1,m-1)/m}$. The values of c for $\tau_{\text{II}}=0.7,0.8,0.9$ and several different values for n and m are given in Table 2.1. It is seen that for given τ_{II} and m, c increases as n increases. The increase in c when n increases ensures that the tolerance region contains the same proportion of the joint distribution on the average. Table 2.1 also suggests that tolerance regions for $m\geq 30$ are close to the asymptotic ($m=\infty$) tolerance regions. It follows that in the specification of S and $c_j,j=1,\ldots,n$, $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ may be treated as if they were the true values if the sample sizes are sufficiently large. This means that instead of using Equation (2.4) and referring to the F distribution, the experimenter can use the standard normal distribution as a rough guide.

According to Myers et al. (1992), in many of Taguchi's applications, the high and low levels of a noise variable are set at $\sqrt{3/2}$ standard deviations from its mean. They also state that it is common in applications for the high and low levels of a noise variable to be set at 1 or 2 standard deviations from its mean. However, as we shall see in examples in this thesis, arbitrarily using commonly employed values for the scaling factors can lead to experiments that are not representative of process conditions.

Table 2.1: Values of c to Achieve Given τ_{II} for Various Values of n and m

m	$\tau_{\text{II}} = 0.7$			$\tau_{\text{II}} = 0.8$			$\tau_{\text{II}} = 0.9$		
	$n=1$	$n=2$	$n=3$	$n=1$	$n=2$	$n=3$	$n=1$	$n=2$	$n=3$
10	1.15	1.59	1.85	1.45	1.89	2.14	1.92	2.36	2.61
20	1.09	1.49	1.71	1.36	1.74	1.95	1.77	2.13	2.33
30	1.07	1.45	1.67	1.33	1.70	1.90	1.73	2.07	2.26
40	1.06	1.44	1.65	1.32	1.68	1.87	1.71	2.04	2.22
50	1.06	1.43	1.63	1.31	1.67	1.86	1.69	2.02	2.20
60	1.05	1.42	1.63	1.31	1.66	1.85	1.68	2.01	2.18
70	1.05	1.42	1.62	1.30	1.65	1.84	1.68	2.00	2.17
80	1.05	1.42	1.62	1.30	1.65	1.84	1.67	1.99	2.16
90	1.05	1.41	1.61	1.30	1.64	1.83	1.67	1.99	2.16
100	1.05	1.41	1.61	1.30	1.64	1.83	1.67	1.98	2.15
∞	1.04	1.39	1.59	1.28	1.62	1.80	1.64	1.95	2.11

There are two points that should be noted. Firstly, the recommendation that τ_{II} be between 0.7 and 0.9 is based on the assumption that the design points will be selected such that the convex hull of the points $\mathbf{z}_l, l = 1, \ldots, N$ is nearly the size of S. Otherwise, S can be replaced by a smaller design region, which has a smaller τ_{II}. Secondly, because τ_{II} depends on the sample sizes, we need to iterate between Step 3 and Step 4 of the proposed procedure to achieve a tolerance region of the desired size.

2.4   Estimation of the Mean and Variance Models and Propagation of Sampling Error

Consider the case where there is a single noise variable and a single control variable. Suppose that estimates for the mean and variance of the noise variable are $\tilde{\mu}_1 = 10.5$ and $\tilde{\sigma}_1^2 = 1.5^2$ respectively. Suppose that the fitted response model is

$\tilde{y} = 21 + 5x_1 - 3x_1^2 - 4(\xi_1 - 10.5)/1.5 + 3x_1(\xi_1 - 10.5)/1.5$,

and an estimate of the experiment error is $\tilde{\sigma}^2 = 1$.

Given the information above, how may the mean and variance models be estimated? In process operation or product use, ξ_1 will vary randomly with mean μ_1 and variance σ_1^2, which are unknown. The experimenter's best guess of μ_1 and σ_1^2 are 10.5 and 1.5^2 respectively. Therefore, it seems reasonable to estimate the mean model by substituting 10.5 for ξ_1 in the expression for $\tilde{y}$. This gives the estimate $21 + 5x_1 - 3x_1^2$, which can be obtained from (1.6) if $\tilde{\mu}_1$ is treated as if it were μ_1. Similarly, the experimenter's best guess of $\text{var}[(\xi_1 - 10.5)/1.5]$ is 1. Therefore, an apparently reasonable estimate for the variance model is $(-4 + 3x_1)^2 + 1$, which can be obtained from (1.7) by treating $\tilde{\sigma}_1^2$ as if it were σ_1^2. Certainly, an estimate for the variance model can also be obtained from (1.8) by treating $\tilde{\sigma}_1^2$ as if it were σ_1^2.

In the following, we formalize the preceding idea of estimating the mean and variance models. In a subsequent section, we shall examine how errors in estimating $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ affect estimates of the mean and variance models obtained through this method.

The assumed response model in (2.3) when written in terms of the variables $\mathbf{x}$ and $\mathbf{z}$ is given by

$$y(\mathbf{x}, \mathbf{z}) = \beta_{0\mathbf{z}} + \mathbf{x}'\boldsymbol{\beta}_\mathbf{z} + \mathbf{x}'\mathbf{B}_\mathbf{z}\mathbf{x} + \boldsymbol{\gamma}'_\mathbf{z}\mathbf{z} + \mathbf{x}'\boldsymbol{\Delta}_\mathbf{z}\mathbf{z} + \varepsilon , \tag{2.5}$$

where $\beta_{0\mathbf{z}}$, $\boldsymbol{\beta}_\mathbf{z}$, $\mathbf{B}_\mathbf{z}$, $\boldsymbol{\gamma}_\mathbf{z}$, and $\boldsymbol{\Delta}_\mathbf{z}$ are the model coefficients, and as before, ε has mean zero and variance σ^2.

Let the corresponding model fitted with least squares be given by

$$\hat{y}(\mathbf{x}, \mathbf{z}) = \hat{\beta}_{0\mathbf{z}} + \mathbf{x}'\hat{\boldsymbol{\beta}}_\mathbf{z} + \mathbf{x}'\hat{\mathbf{B}}_\mathbf{z}\mathbf{x} + \hat{\boldsymbol{\gamma}}'_\mathbf{z}\mathbf{z} + \mathbf{x}'\hat{\boldsymbol{\Delta}}_\mathbf{z}\mathbf{z} . \tag{2.6}$$

If the experimenter treats $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ as if they were $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ respectively and uses Equation (1.6), the estimator for the mean model actually used is given by

$$\hat{\mu}_{Y|\mathbf{z}} = \hat{\beta}_{0\mathbf{z}} + \mathbf{x}'\hat{\boldsymbol{\beta}}_\mathbf{z} + \mathbf{x}'\hat{\mathbf{B}}_\mathbf{z}\mathbf{x} . \tag{2.7}$$

Similarly, referring to (1.7) and (1.8), and assuming independently distributed noise variables (Assumption 2.4), apparently reasonable estimators for the variance model are given by either

$$\hat{\sigma}^2_{YB|\mathbf{z}} = (\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x}) + \hat{\sigma}^2 \tag{2.8}$$

or

$$\hat{\sigma}^2_{Y|\mathbf{z}} = (\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x}) + \hat{\sigma}^2[1 - trace(\mathbf{V}\mathbf{C})], \tag{2.9}$$

where $\mathbf{V} = \begin{pmatrix} 1/c_1^2 & 0 & \cdots & 0 \\ 0 & 1/c_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/c_n^2 \end{pmatrix}$, and $\mathbf{C} = \operatorname{var}[(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x})|\mathbf{z}]/\sigma^2$.

Equations (1.6)-(1.8) are derived assuming that the means and variances of the noise variables are known. When these parameters are substituted with estimates, Equations (2.7)-(2.9) are obtained. The following example demonstrates that errors in estimating the means and variances of the noise variables can be significant components of errors in the estimation of the mean and variance models.

2.4.1 Example 2.1

Consider the case where there is one control variable and one noise variable. Let the coded level of the control variable be represented by x_1 and let the un-coded level of the noise variable be represented by ξ_1. Suppose that unknown to the experimenter, the mean and variance of the noise variable are $\mu_1 = 3$ and $\sigma_1^2 = 2^2$ respectively and the true response model is

$$y(x_1, q_1) = 5 + 6x_1 + 7x_1^2 + 5q_1 + 8x_1q_1, \text{ where } q_1 = \frac{(\xi_1 - 3)}{2}.$$

Imagine the following scenario. The experimenter specifies $R = \{x_1 : -1 \le x_1 \le 1\}$, $S = \{z_1 : -1 \le z_1 \le 1\}$, and $c_1 = c_2 = 1$. She chooses the MRD design shown in Table 2.2 and specifies a sample size of 10. After sampling from the process, she obtains the estimates $\tilde{\mu}_1 = 3.5$ and $\tilde{\sigma}_1^2 = 3^2$ for the mean and variance of the noise variable. Based on those estimates and the design matrix, she sets the high level of the noise variable at 6.5 un-coded units and the low level at 0.5 un-coded units. Because experiment error is negligible, she observes the response values given by the deterministic model $y(x_1, q_1)$ in the experiment, which are given in the column labeled y in Table 2.2.

Table 2.2: Experiment Design, Un-coded Levels of Noise Variable
and Experiment Data for Example 2.1

x_1	z_1	ξ_1	y
-1	-1	0.5	9.75
1	1	6.5	40.75
-1	-1	0.5	9.75
1	1	6.5	40.75
-1	0	3.5	5.25
1	0	3.5	21.25
0	0	3.5	6.25

Consider estimating the mean and variance models with the data in Table 2.2. The fitted response model is $\tilde{y}(x_1, z_1) = 6.25 + 8x_1 + 7x_1^2 + 7.5z_1 + 12x_1 z_1$, where $z_1 = (\xi_1 - 3.5)/3$. Using (2.7), we estimate the mean model as $\tilde{\mu}_{Y|z} = 6.25 + 8x_1 + 7x_1^2$ and using (2.8) or (2.9), we estimate the variance model as $\tilde{\sigma}_{Y|z}^2 = (7.5 + 12x_1)^2 + 0$.

Note that the true mean model is $\mu_Y = 5 + 6x_1 + 7x_1^2$ whereas the true variance model is $\sigma_Y^2 = (5 + 8x_1)^2$. In Figure 2.2, $\tilde{\mu}_{Y|z}$ and μ_Y are plotted while in Figure 2.3, $\tilde{\sigma}_{Y|z}^2$ and

σ_Y^2 are plotted. These figures show that the estimates are in error. This can only be due to the errors in estimating μ_1 and σ_1^2 as there is no experiment error.
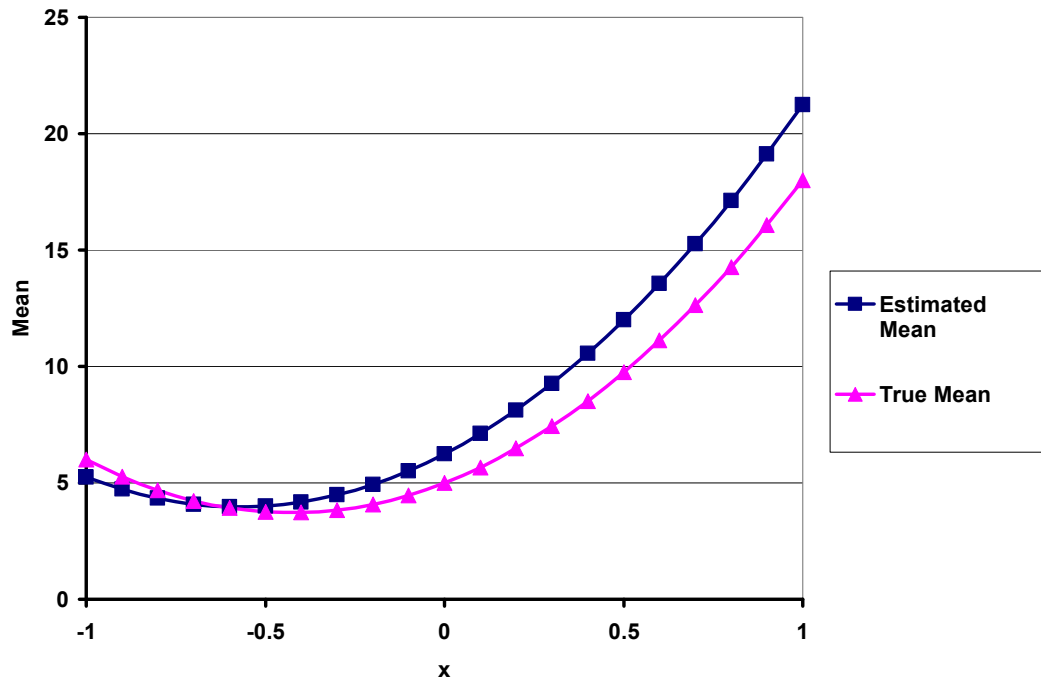
Figure 2.2: Graphs of $\widetilde{\mu}_{Y|\mathbf{z}}$ and μ_Y
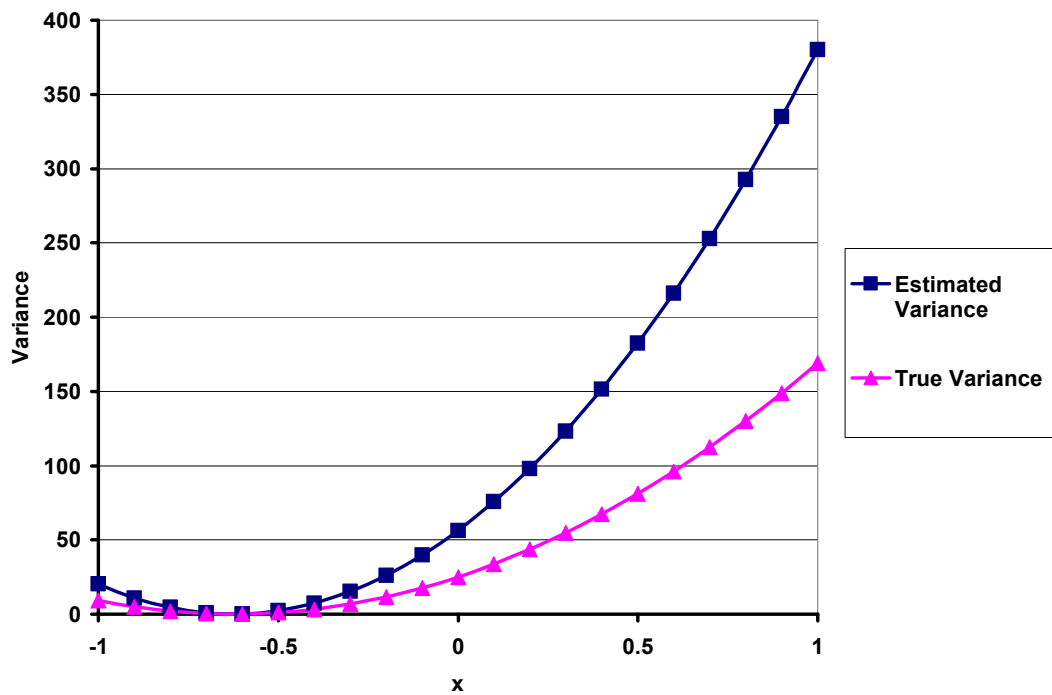
Figure 2.3: Graphs of $\widetilde{\sigma}_{Y|\mathbf{z}}^2$ and σ_Y^2

2.4.2 Relationships Between Coefficients of Response Models

Example 2.1 indicates that the coefficients of the response model in (1.1) and the coefficients of the response model in (2.5) are, in general, different. This occurs because the coding scheme $\mathbf{z}$ in (2.1) is in general different from the coding scheme $\mathbf{q}$ in (1.2). The relationship between the model coefficients $\beta_{0\mathbf{z}}$, $\boldsymbol{\beta}_{\mathbf{z}}$, $\mathbf{B}_{\mathbf{z}}$, $\boldsymbol{\gamma}_{\mathbf{z}}$, and $\boldsymbol{\Delta}_{\mathbf{z}}$, and the model coefficients β_0, $\boldsymbol{\beta}$, $\mathbf{B}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\Delta}$ can be established by using the fact that given a particular $\mathbf{x}$ and $\boldsymbol{\xi}$, Equations (2.5) and (1.1) must yield exactly the same values when the error term ε is set to zero. This gives

$$
\begin{aligned}
\beta_0 &+ \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k}\sum_{j=i}^{k} \beta_{ij} x_i x_j + \sum_{j=1}^{n} \gamma_j \left(\frac{\xi_j - \mu_j}{c_j \sigma_j} \right) + \sum_{i=1}^{k}\sum_{j=1}^{n} \delta_{ij} x_i \left(\frac{\xi_j - \mu_j}{c_j \sigma_j} \right) \\
= \beta_{0\mathbf{z}} &+ \sum_{i=1}^{k} \beta_{i\mathbf{z}} x_i + \sum_{i=1}^{k}\sum_{j=i}^{k} \beta_{ij\mathbf{z}} x_i x_j + \sum_{j=1}^{n} \gamma_{j\mathbf{z}} \left(\frac{\xi_j - \hat{\mu}_j}{c_j \hat{\sigma}_j} \right) + \sum_{i=1}^{k}\sum_{j=1}^{n} \delta_{ij\mathbf{z}} x_i \left(\frac{\xi_j - \hat{\mu}_j}{c_j \hat{\sigma}_j} \right).
\end{aligned}
\tag{2.10}
$$

Since both sides of (2.10) define exactly the same function in the variables $\mathbf{x}$ and $\boldsymbol{\xi}$, we obtain the following relationships by equating the coefficients of each of the variable terms ξ_j, $x_i \xi_j$, x_i, $x_i x_j$ and the "constant" on both sides of (2.10).

$$
\gamma_{j\mathbf{z}} = \frac{\hat{\sigma}_j}{\sigma_j} \gamma_j, \qquad\qquad j = 1,...,n.
\tag{2.11}
$$

$$
\delta_{ij\mathbf{z}} = \frac{\hat{\sigma}_j}{\sigma_j} \delta_{ij}, \qquad\qquad i = 1,...,k; \quad j = 1,...,n.
\tag{2.12}
$$

$$
\beta_{i\mathbf{z}} = \beta_i + \sum_{j=1}^{n} \delta_{ij} \left(\frac{\hat{\mu}_j - \mu_j}{c_j \sigma_j} \right), \qquad\qquad i = 1,...,k.
\tag{2.13}
$$

$$
\beta_{ij\mathbf{z}} = \beta_{ij}, \qquad\qquad i = 1,...,k; \quad j = i,...,k.
\tag{2.14}
$$

$$
\beta_{0\mathbf{z}} = \beta_0 + \sum_{j=1}^{n} \gamma_j \left(\frac{\hat{\mu}_j - \mu_j}{c_j \sigma_j} \right).
\tag{2.15}
$$

From (2.11)-(2.15), it can be seen that the coefficients β_{0z}, $\boldsymbol{\beta}_z$, $\boldsymbol{\gamma}_z$, and $\boldsymbol{\Delta}_z$ are not in general equal to β_0, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\Delta}$ which are used in the definition of the true mean and variance models given in (1.3) and (1.4). This causes estimates computed from $\hat{\mu}_{Y|\mathbf{z}}$, $\hat{\sigma}^2_{YB|\mathbf{z}}$, and $\hat{\sigma}^2_{Y|\mathbf{z}}$ to be in error even if there were no experiment error because given $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$, the expectation of $\hat{\beta}_{0z}$, $\hat{\boldsymbol{\beta}}_z$, $\hat{\boldsymbol{\gamma}}_z$, and $\hat{\boldsymbol{\Delta}}_z$ equal β_{0z}, $\boldsymbol{\beta}_z$, $\boldsymbol{\gamma}_z$, and $\boldsymbol{\Delta}_z$ respectively.

If the activities of sampling from the process and experimenting are repeated, β_{0z}, $\boldsymbol{\beta}_z$, $\boldsymbol{\gamma}_z$, and $\boldsymbol{\Delta}_z$ also vary randomly. Hence, there is a component of variation in the estimators $\hat{\mu}_{Y|\mathbf{z}}$, $\hat{\sigma}^2_{YB|\mathbf{z}}$, and $\hat{\sigma}^2_{Y|\mathbf{z}}$ due to sampling variation in addition to the component due to experiment error. Thus, if either the sampling or experiment plan is poorly specified, optimization or any decisions based on the estimated mean and variance models may produce highly variable results.

2.4.3 Example 2.2

Consider again example 2.1. Due to the fact that $\mathbf{e} = \mathbf{0}$, where $\mathbf{0}$ is a vector of zeros, $y(x_1, z_1) = \tilde{y}(x_1, z_1) = 6.25 + 8x_1 + 7x_1^2 + 7.5z_1 + 12x_1z_1$.

One may verify that Equations (2.11)-(2.15) give the relationships between the coefficients of $y(x_1, z_1)$ and the coefficients of $y(x_1, q_1) = 5 + 6x_1 + 7x_1^2 + 5q_1 + 8x_1q_1$.

It can be seen that because the coefficients of $y(x_1, z_1)$ are different from that of $y(x_1, q_1)$, the estimates $\tilde{\mu}_{Y|\mathbf{z}}$ and $\tilde{\sigma}^2_{Y|\mathbf{z}}$ are in error.

2.5   Sampling Properties of the Estimators for the Mean and Variance Models

The bias and variances of $\hat{\mu}_{Y|\mathbf{z}}$, $\hat{\sigma}^2_{Y|\mathbf{z}}$, and $\hat{\sigma}^2_{YB|\mathbf{z}}$ are important performance measures of those estimators. In addition, a good allocation of experiment effort to sampling and experimenting is one that takes into account the effect of sample sizes and design on the mean squared errors of the estimators. In this section, we establish some results concerning the bias and variance of each of the estimators $\hat{\mu}_{Y|\mathbf{z}}$ and $\hat{\sigma}^2_{Y|\mathbf{z}}$. A reason for preferring $\hat{\sigma}^2_{Y|\mathbf{z}}$ to $\hat{\sigma}^2_{YB|\mathbf{z}}$ is given. In the next section, the variances of $\hat{\mu}_{Y|\mathbf{z}}$ and $\hat{\sigma}^2_{Y|\mathbf{z}}$ are compared with the variances of $\hat{\mu}_Y$ and $\hat{\sigma}^2_Y$ respectively.

In this section, $\mathbf{s}$ is used to represent the vector of sample observations. The notation $\underset{s}{E}(\cdots)$ denotes the expectation of the quantity in the brackets with respect to $\mathbf{s}$. Since an estimator that is a function of $\hat{\boldsymbol{\mu}}$ and/or $\hat{\boldsymbol{\Sigma}}$ can be rewritten as a function of $\mathbf{s}$, expectation with respect to $\mathbf{s}$ implies expectation with respect to $\hat{\boldsymbol{\mu}}$ and/or $\hat{\boldsymbol{\Sigma}}$. The notation $\underset{e}{E}(\cdots)$ denotes the expectation of the quantity in the brackets with respect to $\mathbf{e}$, which is defined as the vector of experiment error. The variance operators $\underset{s}{\mathrm{var}}(\cdots)$ and $\underset{e}{\mathrm{var}}(\cdots)$ are similarly defined and interpreted.

2.5.1 Bias and Variance of the Estimator for the Mean Model

In this section, we give our main results concerning the bias and variance of $\hat{\mu}_{Y|\mathbf{z}}$. Except for Assumption 2.5, all assumptions in Section 2.2.1 are assumed to hold.

Proposition 2.1 If $\hat{\boldsymbol{\mu}}$ is an unbiased estimator for $\boldsymbol{\mu}$, $\hat{\mu}_{Y|\mathbf{z}}$ is an unbiased estimator of the mean model.

Proof:

Equations (2.13)-(2.15) can be rewritten as

$$\beta_{0\mathbf{z}} = \beta_0 - \boldsymbol{\gamma}'\mathbf{w}, \ \boldsymbol{\beta}_{\mathbf{z}} = \boldsymbol{\beta} - \boldsymbol{\Delta}\mathbf{w}, \text{ and } \mathbf{B}_{\mathbf{z}} = \mathbf{B},$$

where $\mathbf{w} = \left[\dfrac{(\mu_1 - \hat{\mu}_1)}{c_1\sigma_1}, \dfrac{(\mu_2 - \hat{\mu}_2)}{c_2\sigma_2}, \ldots, \dfrac{(\mu_n - \hat{\mu}_n)}{c_n\sigma_n} \right]'$.

Since $\hat{\boldsymbol{\mu}}$ is unbiased, $E(\mathbf{w}) = 0$. Thus, $E(\beta_{0\mathbf{z}}) = \beta_0$ and $E(\boldsymbol{\beta}_{\mathbf{z}}) = \boldsymbol{\beta}$. It follows that

$$E(\hat{\mu}_{Y|\mathbf{z}})$$
$$= E_{\mathbf{s}}\left[E_{\mathbf{e}}(\hat{\beta}_{0\mathbf{z}} + \mathbf{x}'\hat{\boldsymbol{\beta}}_{\mathbf{z}} + \mathbf{x}'\hat{\mathbf{B}}_{\mathbf{z}}\mathbf{x})\big|\mathbf{s} \right]$$
$$= E_{\mathbf{s}}(\beta_{0\mathbf{z}} + \mathbf{x}'\boldsymbol{\beta}_{\mathbf{z}} + \mathbf{x}'\mathbf{B}\mathbf{x})$$
$$= \beta_0 + \mathbf{x}'\boldsymbol{\beta} + \mathbf{x}'\mathbf{B}\mathbf{x}.$$

Remark: The result given in this proposition does not require Assumptions 2.4 and 2.5.

Proposition 2.2 The variance of $\hat{\mu}_{Y|\mathbf{z}}$ is given by the formula

$$\mathrm{var}(\hat{\mu}_{Y|\mathbf{z}}) = (\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x})'\mathrm{var}(\mathbf{w})(\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x}) + \mathbf{x}'_C \mathbf{V}_C \mathbf{x}_C \sigma^2, \tag{2.16}$$

where $\mathbf{x}_C = (1, x_1, \ldots, x_k, x_1^2, \ldots, x_k^2, x_1 x_2, \ldots, x_{k-1} x_k)'$ and $\mathbf{V}_C$ is obtained as follows. Let $\mathbf{X}$ be the design matrix expanded to the form of the response model. Let the columns of $\mathbf{X}$ be arranged in the order

$$(1, x_1, \ldots, x_k, x_1^2, \ldots, x_k^2, x_1 x_2, \ldots, x_{k-1} x_k, z_1, x_1 z_1, \ldots, x_k z_1, \ldots, z_n, x_1 z_n, \ldots, x_k z_n).$$

The matrix $\mathbf{V}_C$ is the square matrix obtained by deleting the last $n + nk$ columns and rows of $(\mathbf{X}'\mathbf{X})^{-1}$.

Proof:

Using the conditional variance formula, $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ is given by

$$\begin{aligned}
\text{var}(\hat{\mu}_{Y|\mathbf{z}}) &= \underset{\text{s,e}}{\text{var}}(\hat{\beta}_{0\mathbf{z}} + \mathbf{x}'\hat{\boldsymbol{\beta}}_{\mathbf{z}} + \mathbf{x}'\hat{\mathbf{B}}_{\mathbf{z}}\mathbf{x}) \\
&= \underset{\text{s}}{\text{var}}\left[\underset{\text{e}}{E}(\hat{\beta}_{0\mathbf{z}} + \mathbf{x}'\hat{\boldsymbol{\beta}}_{\mathbf{z}} + \mathbf{x}'\hat{\mathbf{B}}_{\mathbf{z}}\mathbf{x})\Big|\mathbf{s}\right] + \underset{\text{s}}{E}\left[\underset{\text{e}}{\text{var}}(\hat{\beta}_{0\mathbf{z}} + \mathbf{x}'\hat{\boldsymbol{\beta}}_{\mathbf{z}} + \mathbf{x}'\hat{\mathbf{B}}_{\mathbf{z}}\mathbf{x})\Big|\mathbf{s}\right].
\end{aligned}$$
(2.17)

This expresses $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ as the sum of two terms. The first term is reduced as follows.

$$\begin{aligned}
&\underset{\text{s}}{\text{var}}\left[\underset{\text{e}}{E}(\hat{\beta}_{0\mathbf{z}} + \mathbf{x}'\hat{\boldsymbol{\beta}}_{\mathbf{z}} + \mathbf{x}'\hat{\mathbf{B}}_{\mathbf{z}}\mathbf{x})\Big|\mathbf{s}\right] \\
&= \underset{\text{s}}{\text{var}}(\beta_{0\mathbf{z}} + \mathbf{x}'\boldsymbol{\beta}_{\mathbf{z}} + \mathbf{x}'\mathbf{B}\mathbf{x}) \\
&= \underset{\text{s}}{\text{var}}[\beta_0 + \mathbf{x}'\boldsymbol{\beta} + \mathbf{x}'\mathbf{B}\mathbf{x} - (\boldsymbol{\gamma}'\mathbf{w} + \mathbf{x}'\boldsymbol{\Delta}\mathbf{w})] \\
&= (\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x})'\,\text{var}(\mathbf{w})(\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x}).
\end{aligned}$$
(2.18)

Now, note that the design matrix is specified before sampling. Therefore, $\mathbf{X}$ is considered fixed and we have

$$\begin{aligned}
&\underset{\text{s}}{E}\left[\underset{\text{e}}{\text{var}}(\hat{\beta}_{0\mathbf{z}} + \mathbf{x}'\hat{\boldsymbol{\beta}}_{\mathbf{z}} + \mathbf{x}'\hat{\mathbf{B}}_{\mathbf{z}}\mathbf{x})\Big|\mathbf{s}\right] \\
&= \underset{\text{s}}{E}\{\underset{\text{e}}{\text{var}}[\hat{y}(\mathbf{x},\mathbf{z}=\mathbf{0})]\big|\mathbf{s}\} \\
&= \underset{\text{s}}{E}\left[(\mathbf{x}'_C \quad \mathbf{0}')(\mathbf{X}'\mathbf{X})^{-1}\binom{\mathbf{x}_C}{\mathbf{0}}\sigma^2\right] \\
&= \underset{\text{s}}{E}\left[\mathbf{x}'_C\,\mathbf{V}_C\mathbf{x}_C\sigma^2\right] \\
&= \mathbf{x}'_C\,\mathbf{V}_C\mathbf{x}_C\sigma^2,
\end{aligned}$$
(2.19)

where $\mathbf{0}$ is an $(n+nk)\times 1$ vector of zeros.

Putting together (2.17)-(2.19) gives (2.16).

Remark: The result given by Proposition 2.2 does not require Assumption 2.5. (In fact, it also does not require Assumption 2.4). Define $M_S = (\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x})'\,\text{var}(\mathbf{w})(\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x})$ and

$M_E = \mathbf{x}'_C\,\mathbf{V}_C\mathbf{x}_C\sigma^2$. Hence, $\text{var}(\hat{\mu}_{Y|\mathbf{z}}) = M_S + M_E$. Now, if $\hat{\boldsymbol{\mu}}$ is consistent for $\boldsymbol{\mu}$,

$\underset{m_1,\dots,m_n\to\infty}{\lim}\text{var}(\hat{\mu}_{Y|\mathbf{z}}) = M_E = \text{var}(\hat{\mu}_Y)$, where $\hat{\mu}_Y$ is as given in (1.6). This suggests that

M_S may be viewed as the contribution from sampling error whereas M_E may be

viewed as the contribution from experiment error. It can be seen that if $\hat{\boldsymbol{\mu}}$ is restricted to unbiased estimators, choosing each $\hat{\mu}_j$ as the minimum variance unbiased estimator minimizes $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$.

Corollary 2.1 If for each $j = 1,\ldots,n$, $\hat{\mu}_j$ is the sample mean of a random sample of size m_j, the variance of $\hat{\mu}_{Y|\mathbf{z}}$ is given by

$$
\begin{aligned}
&\text{var}(\hat{\mu}_{Y|\mathbf{z}}) \\
&= \sum_{j=1}^{n}\left[\frac{1}{c_j^2 m_j}\left(\gamma_j + \sum_{i=1}^{k}\delta_{ij}x_i\right)^2\right] + \mathbf{x}'_C\,\mathbf{V}_C\mathbf{x}_C\sigma^2 .
\end{aligned}
\tag{2.20}
$$

Proof:

  This follows from Proposition 2.2 and the fact that $\text{var}(\mathbf{w})$ is a diagonal matrix with diagonal elements $1/(c_1^2 m_1), 1/(c_2^2 m_2),\ldots,1/(c_n^2 m_n)$.

Remark: Equation (2.20) also holds when Assumption 2.5 does not hold.

  In order to interpret the variance or standard deviation of $\hat{\mu}_{Y|\mathbf{z}}$, knowledge of the distribution of $\hat{\mu}_{Y|\mathbf{z}}$ would be helpful. The following proposition gives the distribution of $\hat{\mu}_{Y|\mathbf{z}}$.

Proposition 2.3 If in addition to the assumptions in Section 2.2.1, $\mathbf{e}$ has a spherical normal distribution (see Arnold (1981)) and each $\hat{\mu}_j$ is the sample mean of a random sample of size m_j, $\hat{\mu}_{Y|\mathbf{z}}$ at a given $\mathbf{x}$ is normally distributed.

Proof:

Conditioned upon a given $\mathbf{s}$, we know from the theory of linear models (Arnold, 1981) that $\hat{\mu}_{Y|\mathbf{z}}$ is normally distributed with mean

$$\beta_{0\mathbf{z}} + \mathbf{x'}\boldsymbol{\beta}_{\mathbf{z}} + \mathbf{x'Bx} = \beta_0 + \mathbf{x'}\boldsymbol{\beta} + \mathbf{x'Bx} - (\boldsymbol{\gamma'}\mathbf{w} + \mathbf{x'}\boldsymbol{\Delta}\mathbf{w}),$$

and variance

$$\mathbf{x'}_C \mathbf{V}_C \mathbf{x}_C \sigma^2.$$

Since each $\hat{\mu}_j$ is normally distributed, $\beta_0 + \mathbf{x'}\boldsymbol{\beta} + \mathbf{x'Bx} - (\boldsymbol{\gamma'}\mathbf{w} + \mathbf{x'}\boldsymbol{\Delta}\mathbf{w})$ is normally distributed with mean $\beta_0 + \mathbf{x'}\boldsymbol{\beta} + \mathbf{x'Bx} = \mu_Y$ and variance

$$(\boldsymbol{\gamma} + \boldsymbol{\Delta'x})' \operatorname{var}(\mathbf{w})(\boldsymbol{\gamma} + \boldsymbol{\Delta'x}) = \sum_{j=1}^{n} \left[\frac{1}{c_j^2 m_j} \left(\gamma_j + \sum_{i=1}^{k} \delta_{ij} x_i \right)^2 \right].$$

Therefore, the unconditional distribution of $\hat{\mu}_{Y|\mathbf{z}}$ is normal with mean μ_Y and variance given by (2.20).

2.5.2 Bias and Variances of the Estimators for the Variance Model

In this section, we give our main results concerning the bias and variance of $\hat{\sigma}_{Y|\mathbf{z}}^2$. We also compare the mean squared errors of $\hat{\sigma}_{Y|\mathbf{z}}^2$ and $\hat{\sigma}_{YB|\mathbf{z}}^2$. Except for Assumption 2.5, all assumptions in Section 2.2.1 are assumed to hold.

Proposition 2.4 If $\hat{\boldsymbol{\Sigma}}$ is an unbiased estimator of $\boldsymbol{\Sigma}$, i.e. each $\hat{\sigma}_j^2$ is an unbiased estimator of σ_j^2, $\hat{\sigma}_{Y|\mathbf{z}}^2$ is an unbiased estimator of the variance model.

Proof:

$$E(\hat{\sigma}^2_{Y|\mathbf{z}})$$

$$= E_{\mathbf{s}}\left[E_{\mathbf{e}}(\hat{\sigma}^2_{Y|\mathbf{z}})\Big|\mathbf{s}\right]$$

$$= E_{\mathbf{s}}\left[\sum_{j=1}^{n}\frac{1}{c_j^2}(\gamma_{j\mathbf{z}}+\sum_{i=1}^{k}\delta_{ij\mathbf{z}}x_i)^2+\sigma^2\right]$$

$$= E_{\mathbf{s}}\left[\sum_{j=1}^{n}\frac{\hat{\sigma}_j^2}{c_j^2\sigma_j^2}(\gamma_j+\sum_{i=1}^{k}\delta_{ij}x_i)^2+\sigma^2\right]$$

$$= \sum_{j=1}^{n}\frac{1}{c_j^2}(\gamma_j+\sum_{i=1}^{k}\delta_{ij}x_i)^2+\sigma^2$$

$$= \sigma_Y^2 .$$

Remark: The result given in this proposition does not require Assumption 2.5.

Proposition 2.5 Suppose that $\mathbf{e}$ has a spherical normal distribution (see Arnold (1981)). If $\hat{\boldsymbol{\Sigma}}$ is an unbiased estimator of $\boldsymbol{\Sigma}$, then the variance of $\hat{\sigma}^2_{Y|\mathbf{z}}$ is given by

$$\mathrm{var}(\hat{\sigma}^2_{Y|\mathbf{z}}) = V_S + V_E , \tag{2.21}$$

where

$$V_S = \sum_{j=1}^{n}\frac{1}{c_j^4}(\gamma_j+\sum_{i=1}^{k}\delta_{ij}x_i)^4\,\mathrm{var}\left(\frac{\hat{\sigma}_j^2}{\sigma_j^2}\right),$$

$$V_E = 2\sigma^4\left[\sum_{j=1}^{n}\sum_{l=1}^{n}C_{jl}^2/(c_j^2c_l^2)+\frac{1}{dfSSE}\left(1-\sum_{j=1}^{n}C_{jj}/c_j^2\right)^2\right]+4\sigma^2\sum_{j=1}^{n}\left[\frac{C_{jj}}{c_j^4}(\gamma_j+\sum_{i=1}^{k}\delta_{ij}x_i)^2\right]$$

$$+8\sigma^2\sum_{l=2}^{n}\sum_{j=1}^{l-1}\frac{1}{c_j^2c_l^2}E\left(\frac{\hat{\sigma}_l}{\sigma_l}\right)E\left(\frac{\hat{\sigma}_j}{\sigma_j}\right)(\gamma_j+\sum_{i=1}^{k}\delta_{ij}x_i)(\gamma_l+\sum_{i=1}^{k}\delta_{il}x_i)C_{jl} ,$$

C_{jl} is the element in the jth row and l^{th} column of $\mathbf{C}$, which is the covariance matrix defined after Equation (2.9) and *dfSSE* is the residual degrees of freedom. Note that when $\mathbf{C}$ is 1×1, the term $8\sigma^2\sum_{l=2}^{n}\sum_{j=1}^{l-1}\frac{1}{c_j^2c_l^2}E\left(\frac{\hat{\sigma}_l}{\sigma_l}\right)E\left(\frac{\hat{\sigma}_j}{\sigma_j}\right)(\gamma_j+\sum_{i=1}^{k}\delta_{ij}x_i)(\gamma_l+\sum_{i=1}^{k}\delta_{il}x_i)C_{jl}$

should be removed from the expression for V_E.

Proof:

Using the conditional variance formula, we have

$$
\begin{aligned}
\mathrm{var}(\hat{\sigma}^2_{Y|\mathbf{z}}) &= \underset{\mathbf{s},\mathbf{e}}{\mathrm{var}}\{(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x}) + \hat{\sigma}^2[1 - trace(\mathbf{VC})]\} \\
&= \underset{\mathbf{s}}{\mathrm{var}}\Big(\underset{\mathbf{e}}{E}\{(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x}) + \hat{\sigma}^2[1 - trace(\mathbf{VC})]\}\big|\mathbf{s}\Big) \\
&\quad + \underset{\mathbf{s}}{E}\Big(\underset{\mathbf{e}}{\mathrm{var}}\{(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x}) + \hat{\sigma}^2[1 - trace(\mathbf{VC})]\}\big|\mathbf{s}\Big).
\end{aligned}
\tag{2.22}
$$

This expresses $\mathrm{var}(\hat{\sigma}^2_{Y|\mathbf{z}})$ as the sum of two terms. The first term is reduced as follows.

$$
\begin{aligned}
& \underset{\mathbf{s}}{\mathrm{var}}\Big(\underset{\mathbf{e}}{E}\{(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x}) + \hat{\sigma}^2[1 - trace(\mathbf{VC})]\}\big|\mathbf{s}\Big) \\
&= \underset{\mathbf{s}}{\mathrm{var}}[(\boldsymbol{\gamma}_{\mathbf{z}} + \boldsymbol{\Delta}'_{\mathbf{z}}\,\mathbf{x})'\mathbf{V}(\boldsymbol{\gamma}_{\mathbf{z}} + \boldsymbol{\Delta}'_{\mathbf{z}}\,\mathbf{x}) + \sigma^2] \\
&= \underset{\mathbf{s}}{\mathrm{var}}\Bigg[\sum_{j=1}^{n} \frac{\hat{\sigma}^2_j}{c_j^2 \sigma_j^2}(\gamma_j + \sum_{i=1}^{k}\delta_{ij} x_i)^2 \Bigg] \\
&= \sum_{j=1}^{n} \frac{1}{c_j^4}(\gamma_j + \sum_{i=1}^{k}\delta_{ij} x_i)^4 \, \mathrm{var}\!\left(\frac{\hat{\sigma}^2_j}{\sigma^2_j} \right).
\end{aligned}
\tag{2.23}
$$

The derivation of a formula for the second term in (2.22) is simplified by making use of the general expression for $\mathrm{var}(\hat{\sigma}^2_{YB})$ derived by O'Donnell and Vining (1997). Furthermore, note that $\mathbf{C}$ is fixed because it does not depend on any sample or experiment observations. It follows directly from the expression given by O'Donnell and Vining (1997) and the fact that $\mathbf{C}$ is fixed that

$$
\begin{aligned}
& \underset{\mathbf{s}}{E}\Big(\underset{\mathbf{e}}{\mathrm{var}}\{(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\,\mathbf{x}) + \hat{\sigma}^2[1 - trace(\mathbf{VC})]\}\big|\mathbf{s}\Big) \\
&= \underset{\mathbf{s}}{E}\bigg\{ 2\sigma^4\Big[trace(\mathbf{VC})^2 + \frac{1}{dfSSE}(1 - trace(\mathbf{VC}))^2 \Big] + 4\sigma^2(\boldsymbol{\gamma}_{\mathbf{z}} + \boldsymbol{\Delta}'_{\mathbf{z}}\,\mathbf{x})'\mathbf{VCV}(\boldsymbol{\gamma}_{\mathbf{z}} + \boldsymbol{\Delta}'_{\mathbf{z}}\,\mathbf{x}) \bigg\} \\
&= 2\sigma^4\bigg[\sum_{j=1}^{n}\sum_{l=1}^{n} C_{jl}^2/(c_j^2 c_l^2) + \frac{1}{dfSSE}\Big(1 - \sum_{j=1}^{n} C_{jj}/c_j^2 \Big)^2 \bigg] + 4\sigma^2\sum_{j=1}^{n}\bigg[\frac{C_{jj}}{c_j^4}(\gamma_j + \sum_{i=1}^{k}\delta_{ij} x_i)^2 \bigg] \\
&\quad + 8\sigma^2\sum_{l=2}^{n}\sum_{j=1}^{l-1} \frac{1}{c_j^2 c_l^2} E\!\left(\frac{\hat{\sigma}_j}{\sigma_j} \right) E\!\left(\frac{\hat{\sigma}_l}{\sigma_l} \right)(\gamma_j + \sum_{i=1}^{k}\delta_{ij} x_i)(\gamma_l + \sum_{i=1}^{k}\delta_{il} x_i)C_{jl}.
\end{aligned}
$$

$$
\tag{2.24}
$$

Putting (2.22)-(2.24) together yields (2.21).

Remark: This proposition holds whether or not Assumption 2.5 holds. Now, if $\hat{\Sigma}$ is consistent for Σ, $\lim_{m_j \to \infty} E(\hat{\sigma}_j / \sigma_j) = 1$ for each $j = 1, \ldots, n$, (see Theorem B.2 in Appendix B) and

$$\lim_{m_1, \ldots, m_n \to \infty} \text{var}(\hat{\sigma}^2_{Y|\mathbf{z}}) = \lim_{m_1, \ldots, m_n \to \infty} V_S + \lim_{m_1, \ldots, m_n \to \infty} V_E = 0 + \lim_{m_1, \ldots, m_n \to \infty} V_E = \text{var}(\hat{\sigma}^2_Y),$$

where $\hat{\sigma}^2_Y$ is given in (1.8). This suggests that V_S may be thought of as the component of $\text{var}(\hat{\sigma}^2_{Y|\mathbf{z}})$ due to sampling error and V_E as the component due to experiment error.

If $\mathbf{C}$ is diagonal and if $\hat{\Sigma}$ is restricted to unbiased estimators, choosing each $\hat{\sigma}^2_j$ as the minimum variance unbiased estimator minimizes $\text{var}(\hat{\sigma}^2_{Y|\mathbf{z}})$.

For the purposes of computation, expressions for $E(\hat{\sigma}_j / \sigma_j)$, $\text{var}(\hat{\sigma}^2_j / \sigma^2_j)$, and $\mathbf{C}$ are needed. We discuss how to obtain these expressions below.

1. Expression for $E(\hat{\sigma}_j / \sigma_j)$: If the jth noise variable is normally distributed and $\hat{\sigma}^2_j$ is the sample variance,

$$E\left(\frac{\hat{\sigma}_j}{\sigma_j}\right) = \sqrt{\frac{2}{m_j - 1}} \frac{\Gamma(m_j / 2)}{\Gamma[(m_j - 1)/ 2]} \tag{2.25}$$

(Voinov and Nikulin, 1993; Fisher, 1925), where $\Gamma(\cdot)$ denotes the gamma function. However, if the jth noise variable is not normally distributed, the approximation $E(\hat{\sigma}_j / \sigma_j) \approx 1$ may be used. This is justified by the fact that if $\hat{\sigma}^2_j$ is consistent for σ^2_j, $E(\hat{\sigma}_j / \sigma_j) \to 1$ as $m_j \to \infty$. This result follows from probability theory (see Theorem B.2 in Appendix B).

2. Expression for $\text{var}(\hat{\sigma}_j^2 / \sigma_j^2)$: If $\hat{\sigma}_j^2$ is the sample variance of a random sample of

size m_j and the distribution of the jth noise variable has finite moments of order up to

four,

$$\text{var}\left(\frac{\hat{\sigma}_j^2}{\sigma_j^2}\right) = \frac{2}{m_j - 1} + \frac{\phi_{2j}}{m_j}, \tag{2.26}$$

where $\phi_{2j} \in [-2, \infty)$ is the excess kurtosis of the distribution of the noise variable (Box

et al., 1978; Box, 1953).

3. Expression for $\mathbf{C}$: Define $\mathbf{x}_1 = (1, x_1, \ldots, x_k)'$, and denote the $n \times n$ identity matrix

by $\mathbf{I}_n$. Define $\mathbf{V}_D$ as the square matrix obtained from the elements indexed by the last

$n + nk$ rows and columns of $(\mathbf{X'X})^{-1}$, where $(\mathbf{X'X})^{-1}$ is as defined in Proposition 2.2.

The matrix $\mathbf{C}$ is given by

$$\mathbf{C} = (\mathbf{I}_n \otimes \mathbf{x}_1')\mathbf{V}_D(\mathbf{I}_n \otimes \mathbf{x}_1')', \tag{2.27}$$

where $\otimes$ is the Kronecker product (see Harville (1997) for a definition). This

expression is derived by O'Donnell and Vining (1997).

Up to this point, we have only investigated the bias and variance of $\hat{\sigma}_{Y|\mathbf{z}}^2$. A

competitor to the unbiased estimator $\hat{\sigma}_{Y|\mathbf{z}}^2$ is the biased estimator $\hat{\sigma}_{YB|\mathbf{z}}^2$, which is

simpler to compute and use. Hence, it is natural to ask whether the unbiased estimator

is really better than the biased estimator. Note that an unbiased estimator is not

necessarily a good one in the sense that the estimator may not give estimates as close

to the true value as compared to the estimates given by a biased estimator (Kiefer,

1987). A better criterion for comparing the two estimators is the mean squared error. A

comparison based on this criterion yields the following proposition, which is proven in

Appendix A.

Proposition 2.6 If $\hat{\Sigma}$ is unbiased for Σ, $\hat{\sigma}^2_{Y|\mathbf{z}}$ has a smaller mean square error than $\hat{\sigma}^2_{YB|\mathbf{z}}$ for every $\mathbf{x}$ when $dfSSE \geq 2$.

Remark: The result holds whether or not Assumption 2.5 holds.

Proposition 2.6 suggests that $\hat{\sigma}^2_{Y|\mathbf{z}}$ should be used instead of $\hat{\sigma}^2_{YB|\mathbf{z}}$ whenever the design size is two or more than the number of model parameters. Because this is frequently the case, we consider only the estimator $\hat{\sigma}^2_{Y|\mathbf{z}}$ in the rest of this thesis.

2.5.3 Discussion

We do not justify $\hat{\mu}_{Y|\mathbf{z}}$ and $\hat{\sigma}^2_{Y|\mathbf{z}}$ by proving any optimality property of these estimators. However, in Appendix B we show that if $\hat{\mu}$ and $\hat{\Sigma}$ are consistent estimators (so that $\hat{\mu}$ and $\hat{\Sigma}$ converge to μ and Σ respectively as $m_1, \ldots, m_n \to \infty$), $\hat{\mu}_{Y|\mathbf{z}}$ and $\hat{\sigma}^2_{Y|\mathbf{z}}$ converge in distribution to $\hat{\mu}_Y$ and $\hat{\sigma}^2_Y$ respectively as $m_1, \ldots, m_n \to \infty$. This result, which does not require Assumptions 2.5-2.8, justifies the use of $\hat{\mu}_{Y|\mathbf{z}}$ and $\hat{\sigma}^2_{Y|\mathbf{z}}$ because by increasing the sample sizes, the sampling variation transmitted to the estimators decreases and converges to zero. Alternatively, we can justify the estimators by the fact that as the sample sizes and the number of replications of a design increase, $\hat{\mu}_{Y|\mathbf{z}}$ and $\hat{\sigma}^2_{Y|\mathbf{z}}$ converge to μ_Y and σ^2_Y respectively if $\hat{\mu}$ and $\hat{\Sigma}$ are consistent (This result is shown in Appendix B).

There are two other points about the derivations in Sections 2.5.1 and 2.5.2 that deserve some attention. Firstly, it should be noted that if the noise variables are not normally distributed, the sample mean and sample variance might not be efficient

estimators. For example, it is not efficient to estimate the mean and variance of a uniform distribution with the sample mean and sample variance. However, in the case where the noise variables are not normally distributed, a coding different from that of (2.1) may be more appropriate for constructing tolerance regions. Moreover, the response will not be normally distributed and optimizing the response based on the mean and variance models appears to be questionable.

Secondly, the assumption that the noise variables are known to be independently distributed may be relaxed at the expense of a more complicated investigation of the estimator for the variance model. In this case, the (i, j) element in the matrix $\mathbf{V}$ in Equation (2.9) should be $\hat{\rho}_{ij} / (c_i c_j)$, where $\hat{\rho}_{ij}$ is an estimator of the correlation coefficient for the ith and j^{th} noise variables ($\hat{\rho}_{ii} = 1$). However, neglecting the correlations when they in fact exist may cause errors in estimating the variance model.

2.6   Inflation of Variances Due to Sampling Error

In the literature, the fact that $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are often estimated with process data is ignored, giving rise to the use of the estimators $\hat{\mu}_Y$ and $\hat{\sigma}_Y^2$ for the purposes of theoretical development. However, it seems that $\hat{\mu}_{Y|\mathbf{z}}$ and $\hat{\sigma}_{Y|\mathbf{z}}^2$ more closely resemble reality. A comparison of the variances of both sets of estimators is made in this section. We consider only the case where each $\hat{\mu}_j, j = 1, \ldots, n$ is the sample mean and each $\hat{\sigma}_j^2, j = 1, \ldots, n$ is the sample variance.

Using Equation (2.20) and the fact that $\text{var}(\hat{\mu}_Y) = M_E$, we have

$$\text{var}(\hat{\mu}_{Y|\mathbf{z}}) - \text{var}(\hat{\mu}_Y) = M_S = \sum_{j=1}^{n}\left[\frac{1}{c_j^2 m_j}\left(\gamma_j + \sum_{i=1}^{k}\delta_{ij}x_i\right)^2\right].$$

Thus, $\text{var}(\hat{\mu}_{Y|\mathbf{z}}) - \text{var}(\hat{\mu}_Y) = 0$ if and only if $\mathbf{x}$ is such that $(\gamma + \Delta'\mathbf{x})'\mathbf{V}(\gamma + \Delta'\mathbf{x}) = 0$.

In other words, the variance of the estimator for the mean model at points where σ_Y^2 is minimized is unaffected by sampling variation. This however, does not imply that $\text{var}(\mathbf{x}^*)$, where $\mathbf{x}^*$ is such that $(\hat{\gamma}_{\mathbf{z}} + \hat{\Delta}'_{\mathbf{z}}\mathbf{x}^*)'\mathbf{V}(\hat{\gamma}_{\mathbf{z}} + \hat{\Delta}'_{\mathbf{z}}\mathbf{x}^*) = 0$, is not inflated by sampling variation. Note that the difference $\text{var}(\hat{\mu}_{Y|\mathbf{z}}) - \text{var}(\hat{\mu}_Y)$ tends to increase as σ_Y^2 increases. For the case where $m_1 = \cdots = m_n = m$,

$$\text{var}(\hat{\mu}_{Y|\mathbf{z}}) - \text{var}(\hat{\mu}_Y) = (\sigma_Y^2 - \sigma^2)/m.$$

Now, consider the variance of $\hat{\sigma}_{Y|\mathbf{z}}^2$ compared to the variance of $\hat{\sigma}_Y^2$. Assuming normally distributed noise variables and experiment error and that $\mathbf{C}$ is a diagonal matrix,

$$\text{var}(\hat{\sigma}_{Y|\mathbf{z}}^2) - \text{var}(\hat{\sigma}_Y^2) = V_S = \sum_{j=1}^{n}\frac{1}{c_j^4}(\gamma_j + \sum_{i=1}^{k}\delta_{ij}x_i)^4\left(\frac{2}{m_j - 1}\right).$$

The above equation also holds approximately when $\mathbf{C}$ is not a diagonal matrix because $V_E \approx \text{var}(\hat{\sigma}_Y^2)$ when $m_1, \ldots, m_n$ are sufficiently large (see remark after Proposition 2.5). Similar to the case of the mean model, $\text{var}(\hat{\sigma}_{Y|\mathbf{z}}^2) - \text{var}(\hat{\sigma}_Y^2) = 0$ if and only if $\mathbf{x}$ is such that $(\gamma + \Delta'\mathbf{x})'\mathbf{V}(\gamma + \Delta'\mathbf{x}) = 0$. In addition, $\text{var}(\hat{\sigma}_{Y|\mathbf{z}}^2) - \text{var}(\hat{\sigma}_Y^2)$ also tends to increase as σ_Y^2 increases.

It follows from our discussion that in experiments where the noise variables have large effects, the variances of $\hat{\mu}_{Y|\mathbf{z}}$ and $\hat{\sigma}_{Y|\mathbf{z}}^2$ at most points $\mathbf{x}$ in the design region R are inflated considerably by sampling variation. In many RPD experiments, interest is in studying those noise variables that appear to cause a great amount of variation in

the response. Therefore, it is likely that in most cases, M_S will at least be comparable to M_E and V_S will at least be comparable to V_E at many points $\mathbf{x}$ in R.

2.6.1 Example 2.3

Consider the case where $k = 2$, $n = 2$, $R = \{(x_1, x_2); -1 \le x_i \le 1, i = 1, 2\}$, $S = \{(z_1, z_2); -1 \le z_j \le 1, j = 1, 2\}$, $c_1 = c_2 = 1.5$, and $m_1 = m_2 = 60$. This gives $\tau_{\mathrm{II}} = 0.73$. Suppose that the experimenter chooses the MRD design that comprises:

1. The 2^4 factorial in which the coded levels of each factor are at ± 1.

2. One replicate of the axial points for the control variables with axial distance $\alpha = 1$.

3. Four center points.

Suppose that the parameters $\boldsymbol{\gamma}$, $\boldsymbol{\Delta}$, and σ^2 are given by

$$\boldsymbol{\gamma} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \; \boldsymbol{\Delta} = \begin{pmatrix} 1.5 & 1 \\ 1.5 & 1 \end{pmatrix}, \text{ and } \sigma^2 = 1.$$

The sizes of the elements of $\boldsymbol{\gamma}$ and $\boldsymbol{\Delta}$ relative to σ appear to be reasonable based on an inspection of some real and hypothetical examples in the literature. Note that 93 percent of the size of σ_Y^2 at $\mathbf{x} = 0$ is attributed to the noise variables. At $\mathbf{x} = (-1, -1)$, $\sigma_Y^2 - \sigma^2 = 0$ whereas at $\mathbf{x} = (1, 1)$, σ_Y^2 is a maximum. Figure 2.4 plots $\mathrm{var}(\hat{\mu}_{Y|\mathbf{z}})$ and $\mathrm{var}(\hat{\mu}_Y)$ versus $\mathbf{x}$ while Figure 2.5 plots $\mathrm{var}(\hat{\sigma}_{Y|\mathbf{z}}^2)$ and $\mathrm{var}(\hat{\sigma}_Y^2)$ versus $\mathbf{x}$. These figures demonstrate that even with a moderately large sample size for each noise variable, sampling variation can significantly inflate the variances of $\hat{\mu}_{Y|\mathbf{z}}$ and $\hat{\sigma}_{Y|\mathbf{z}}^2$.
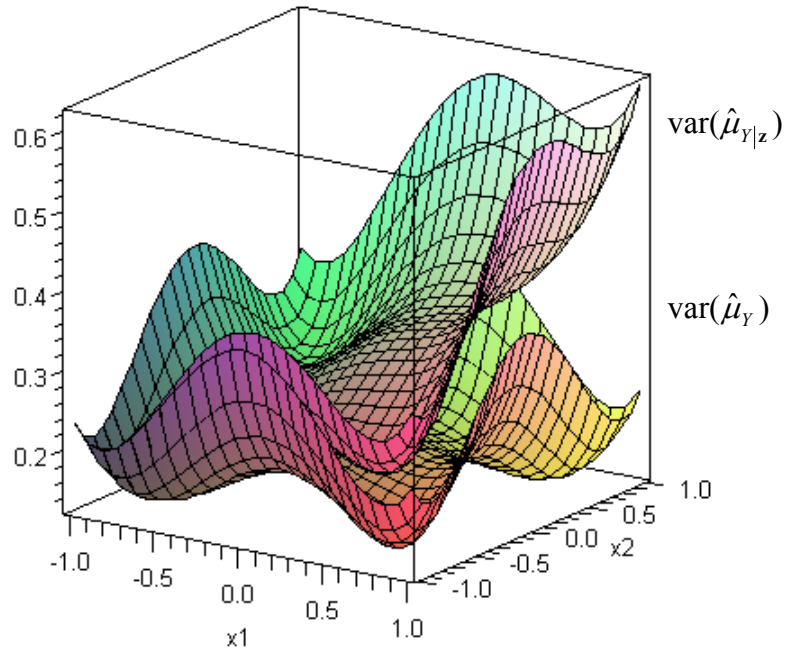
Figure 2.4: Plots of $\mathrm{var}(\hat{\mu}_{Y|\mathbf{z}})$ and $\mathrm{var}(\hat{\mu}_Y)$ versus $\mathbf{x}$
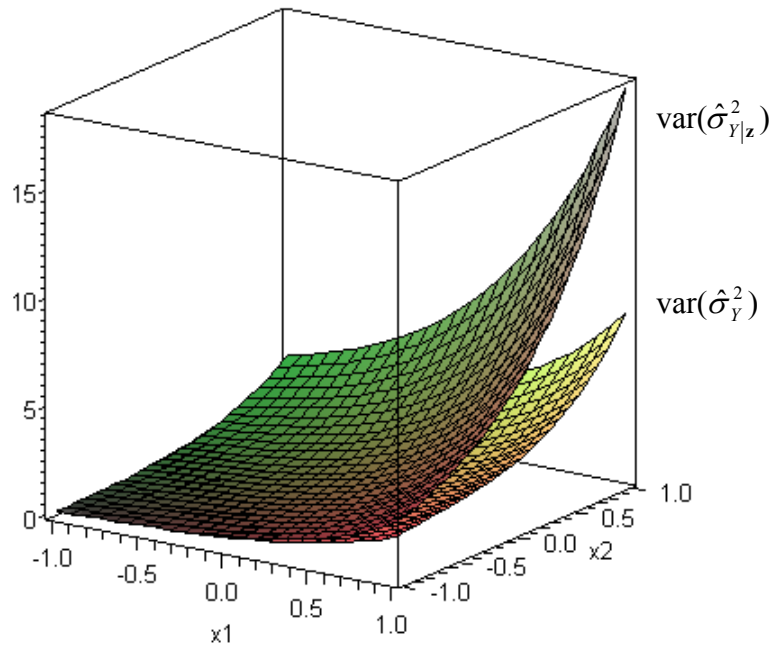
Figure 2.5: Plots of $\mathrm{var}(\hat{\sigma}^2_{Y|\mathbf{z}})$ and $\mathrm{var}(\hat{\sigma}^2_Y)$ versus $\mathbf{x}$

Remark: Example 2.3 suggests that when the means and variances of the noise variables are unknown and estimated with sample data, it makes little sense to focus on choosing the most efficient experimental design only. Efficient experiment designs have received much attention in the literature while the problem of planning process data collection seems to be considered an insignificant problem. Although M_S and V_S tend to be small around the point where $\sigma_Y^2 - \sigma^2 = 0$, they can be very large at other points in R. Frequently, interest is in predicting the mean and variance of the response over the region R rather than at only the point where $\sigma_Y^2 - \sigma^2 = 0$, which in any case is usually unknown. Furthermore, tradeoffs between the objectives of minimizing the variance, minimizing operating or product costs, and optimizing the mean of the response must be made by the decision maker in many cases and accurate estimation of the mean and variance models is required for this purpose.

2.7   Summary

This chapter gives the proposed procedure that combines planning of the sampling effort and planning of the combined array experiment in a single step. Key assumptions that are and shall be made in further developing the procedure into a complete approach are given in Section 2.2.1. The problem of choosing the design region for the noise variables S, and the scaling factors $c_j, j = 1, \ldots, n$ is treated. Equation (2.4) is used to determine the values of $c_j, j = 1, \ldots, n$ that would give a desired τ_{II} given an S that is the Cartesian product of intervals for each noise variable. Estimators for the mean and variance models, i.e. $\hat{\mu}_{Y|\mathbf{z}}$, $\hat{\sigma}_{Y|\mathbf{z}}^2$, and $\hat{\sigma}_{YB|\mathbf{z}}^2$ are given in Equations (2.7)-(2.9). The question of how errors in estimates of the means and

variances of the noise variables are transmitted to estimates of the mean and variance models is resolved with the derivation of Equations (2.11)-(2.15). Based on these equations, the bias and variance of each of the estimators $\hat{\mu}_{Y|\mathbf{z}}$ and $\hat{\sigma}^2_{Y|\mathbf{z}}$ are investigated. In Proposition 2.1, we show that $\hat{\mu}_{Y|\mathbf{z}}$ is unbiased if $\hat{\boldsymbol{\mu}}$ is unbiased and in Proposition 2.2, we derive the variance of $\hat{\mu}_{Y|\mathbf{z}}$. Formulas for the variance of $\hat{\mu}_{Y|\mathbf{z}}$ are given in Equations (2.16) and (2.20), the latter for the case where $\hat{\boldsymbol{\mu}}$ is a vector of sample means. In Proposition 2.4, we show that if each $\hat{\sigma}^2_j$ is unbiased, $\hat{\sigma}^2_{Y|\mathbf{z}}$ is unbiased. The variance of $\hat{\sigma}^2_{Y|\mathbf{z}}$ is derived in Proposition 2.5 and is given in Equation (2.21). Proposition 2.6 gives a reason for preferring $\hat{\sigma}^2_{Y|\mathbf{z}}$ to $\hat{\sigma}^2_{YB|\mathbf{z}}$. In addition, asymptotic properties of the estimators $\hat{\mu}_{Y|\mathbf{z}}$ and $\hat{\sigma}^2_{Y|\mathbf{z}}$ that provide justifications for the use of those estimators are mentioned in Section 2.5.3. Finally, we compare the variance of $\hat{\mu}_{Y|\mathbf{z}}$ with the variance of $\hat{\mu}_Y$ and also the variance of $\hat{\sigma}^2_{Y|\mathbf{z}}$ with the variance of $\hat{\sigma}^2_Y$. The comparisons show that sampling variation can significantly inflate the variance of the estimators for the mean and variance models.

CHAPTER 3

OPTIMAL ALLOCATION OF EXPERIMENT EFFORT

TO SAMPLING AND EXPERIMENTING

3.1   Introduction

Cost can be an important consideration in the practice of design of experiments. A discussion of cost considerations in the selection of the appropriate split plot arrangement for robust design is given by Box et al. (2005). Wu and Hamada (2000) discuss cost considerations in selecting between crossed arrays and combined arrays. Park et al. (2005) present G-optimal designs generated with a genetic algorithm that satisfy certain cost constraints. In practice, any experiment program is allocated a finite budget and must be completed within a specific length of time. Therefore, in the setting of the proposed procedure given in Figure 2.1, it is of practical interest to determine the sample sizes and design that best estimates the mean and variance models given constraints on time and budget. In the remainder of this thesis, a specification of $m_j, j = 1, \ldots n$ and a design shall be called a scheme. Hence, our problem is to find a scheme that best estimates the mean and variance models given the available resource. In considering the problem, we shall always assume that each $\hat{\mu}_j$

and $\hat{\sigma}_j^2$ are the sample mean and sample variance respectively of a random sample of size m_j.

Alternative schemes can be evaluated based on the values of $\mathrm{var}(\hat{\mu}_{Y|\mathbf{z}})$ and $\mathrm{var}(\hat{\sigma}_{Y|\mathbf{z}}^2)$ at various $\mathbf{x} \in R$. However, instead of $\mathrm{var}(\hat{\sigma}_{Y|\mathbf{z}}^2)$, we use $\mathrm{var}(\hat{\sigma}_{Y|\mathbf{z}}^2 - \hat{\sigma}^2)$ as a basis for evaluating alternative schemes in this research. One reason is the following. The variance model σ_Y^2 comprises two components: $(\boldsymbol{\gamma} + \boldsymbol{\Delta'}\mathbf{x})'(\boldsymbol{\gamma} + \boldsymbol{\Delta'}\mathbf{x})$ representing the component of σ_Y^2 due to the noise variables to be studied in the combined array experiment, and σ^2 representing the component of σ_Y^2 due to unidentified noise variables. There is, however, usually more interest in estimating the quantity $(\boldsymbol{\gamma} + \boldsymbol{\Delta'}\mathbf{x})'(\boldsymbol{\gamma} + \boldsymbol{\Delta'}\mathbf{x})$ than the constant σ^2. This can be seen by surveying criteria proposed in the literature for evaluating a combined array design. For instance, Borror et al. (2002) propose evaluating designs based on $\mathrm{var}\left(\hat{\gamma}_j + \sum_{i=1}^{k} \hat{\delta}_{ij} x_i \right) = C_{jj}, j = 1, \ldots, n$, which are called the slope variances. In another paper, Castillo et al. (2007) propose the criterion $\underset{\mathbf{Q}}{E}\left\{ \underset{\hat{\gamma}, \hat{\Delta}}{\mathrm{var}}[(\hat{\boldsymbol{\gamma}}' + \mathbf{x}'\hat{\boldsymbol{\Delta}})\mathbf{Q}] \right\}$ for evaluating and generating designs for RPD experiments. These two criteria represent attempts to quantify the performance of a design at estimating the sensitivity of the response to changes in the noise variables. They do not reflect interest in σ^2. Now, an estimator for $(\boldsymbol{\gamma} + \boldsymbol{\Delta'}\mathbf{x})'(\boldsymbol{\gamma} + \boldsymbol{\Delta'}\mathbf{x})$ is $\hat{\sigma}_{Y|\mathbf{z}}^2 - \hat{\sigma}^2$. Evidently, $\hat{\sigma}_{Y|\mathbf{z}}^2 - \hat{\sigma}^2$ is unbiased for $(\boldsymbol{\gamma} + \boldsymbol{\Delta'}\mathbf{x})'(\boldsymbol{\gamma} + \boldsymbol{\Delta'}\mathbf{x})$. In addition, $\hat{\sigma}_{Y|\mathbf{z}}^2 - \hat{\sigma}^2$ has a smaller mean squared error than $\hat{\sigma}_{YB|\mathbf{z}}^2 - \hat{\sigma}^2$ when $dfSSE \geq 3$. Therefore, when there is more interest in estimating $(\boldsymbol{\gamma} + \boldsymbol{\Delta'}\mathbf{x})'(\boldsymbol{\gamma} + \boldsymbol{\Delta'}\mathbf{x})$ than σ^2, a scheme should be evaluated based on $\mathrm{var}(\hat{\sigma}_{Y|\mathbf{z}}^2 - \hat{\sigma}^2)$. It can be shown that

$$\text{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2) = V_S + \overline{V}_E ,$$ (3.1)

where V_S is as defined in Equation (2.21) and $\overline{V}_E$ is given by

$$\overline{V}_E = 2\sigma^4 \left[\sum_{j=1}^{n}\sum_{l=1}^{n} C_{jl}^2 /(c_j^2 c_l^2) + \frac{1}{dfSSE} \left(\sum_{j=1}^{n} C_{jj}/c_j^2 \right)^2 \right] + 4\sigma^2 \sum_{j=1}^{n} \left[\frac{C_{jj}}{c_j^4} \left(\gamma_j + \sum_{i=1}^{k} \delta_{ij} x_i \right)^2 \right]$$

$$+ 8\sigma^2 \sum_{l=2}^{n}\sum_{j=1}^{l-1} \frac{1}{c_j^2 c_l^2} E\left(\frac{\hat{\sigma}_l}{\sigma_l} \right) E\left(\frac{\hat{\sigma}_j}{\sigma_j} \right) \left(\gamma_j + \sum_{i=1}^{k} \delta_{ij} x_i \right) \left(\gamma_l + \sum_{i=1}^{k} \delta_{il} x_i \right) C_{jl} .$$

Note that $\overline{V}_E$ is obtained from V_E by replacing the term $(2\sigma^4/dfSSE)[1 - trace(\mathbf{VC})]^2$

with $(2\sigma^4/dfSSE)[trace(\mathbf{VC})]^2$.

In this research, we consider only the sample sizes and the design as decision variables in the allocation of resource with the objective of improving the estimation of the mean and variance models. The scaling factors $c_j, j = 1,\ldots,n$ appear in the expressions for $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$, $\text{var}(\hat{\sigma}^2_{Y|\mathbf{z}})$, and $\text{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)$. However, it should be noted that although wider levels of the noise variables reduce $\text{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)$ by reducing $\overline{V}_E$ and reduce $\text{var}(\hat{\sigma}^2_{Y|\mathbf{z}})$ by reducing V_E, they have no effect in reducing V_S and $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$. This can be seen by noting that if the scaling factor for the jth noise variable c_j is replaced by $\overline{c}_j = k_j c_j, k_j > 0$, then the coefficient γ_j of the response model given in (1.1) should be replaced by $\overline{\gamma}_j = k_j \gamma_j$ and the coefficients $\delta_{ij}, i = 1,\ldots,k$ should be replaced by $\overline{\delta}_{ij} = k_j \delta_{ij}, i = 1,\ldots,k$. Therefore, as each c_j increases, $\overline{V}_E$ tends to zero and V_E tends to $2\sigma^4/dfSSE$ but $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ and V_S remain constant. However, as discussed in Section 2.3, larger scaling factors give a tolerance region S_ξ that is expected to contain a larger proportion τ_{II} of the joint distribution of the noise variables and this raises concern about model inadequacy. For this reason, we

do not consider the scaling factors as decision variables to be chosen to improve estimation of the variance model.

3.1.1 General Formulation of Resource Allocation Problem

This chapter considers special cases of the general resource allocation problem, which is formulated below.

General Formulation of Resource Allocation Problem

$$\min \varphi \left[\operatorname{var}(\hat{\mu}_{Y|\mathbf{z}}), \ \operatorname{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2) \right]$$

Explanation of Constraint

subject to: $\mathbf{x}_l \in R, l = 1, \dots, N,$

Coded levels for control variables must be in R.

$\mathbf{z}_l \in S, l = 1, \dots, N,$

Coded levels for noise variables must be in S.

$N \geq p + 1,$

Number of runs must be at least $p + 1$, where p is the number of coefficients in the response model.

$m_j \geq 2, j = 1, \dots, n,$

Sample size for each noise variable must be at least two so that the variances can be estimated.

$Cost(N, m_1, \dots, m_n) \leq K,$

The maximum cost of the scheme is K.

$\pi \left[\operatorname{var}(\hat{\mu}_{Y|\mathbf{z}}), \ \operatorname{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2) \right] \leq U,$

The maximum value of some function of the variances is U.

$N, m_1, \dots, m_n$ are integers.

In the above formulation, the coded levels of the control and noise variables for each experiment run, the number of experiment runs, as well as the sample sizes for each noise variable are decision variables. The objective is to minimize some function of the variances $\varphi\left[\operatorname{var}(\hat{\mu}_{Y|\mathbf{z}}),\ \operatorname{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)\right]$. Explanations of the constraints are provided in the space to the right of the constraint. There is a constraint on the total cost and an upper bound is placed on the value of some function of the variances $\pi\left[\operatorname{var}(\hat{\mu}_{Y|\mathbf{z}}),\ \operatorname{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)\right]$. This latter constraint will be important in cases where the objective is a function of only one of the variances (for instance, $\varphi\left[\operatorname{var}(\hat{\mu}_{Y|\mathbf{z}}),\ \operatorname{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)\right] = \varphi\left[\operatorname{var}(\hat{\mu}_{Y|\mathbf{z}})\right]$) since it would then be possible to place some restriction on the values of the other variance.

The general resource allocation problem is extremely difficult to solve. There appears to be no result in the literature that may readily be used to solve the simpler problem of finding an (exact) design that optimizes $\varphi\left[\operatorname{var}(\hat{\mu}_Y)\right]$ or $\varphi\left[\operatorname{var}(\hat{\sigma}^2_Y - \hat{\sigma}^2)\right]$, where $\hat{\mu}_Y$ and $\hat{\sigma}^2_Y$ are given in Equations (1.6) and (1.8). In fact, cases of exact optimum design problems are frequently simplified by assuming that there is a finite set of candidate points with which to construct the optimal design and researchers seem to have focus only on the D-optimality criterion (Donev and Atkinson, 1988; Welch, 1982). In view of these facts, we do not attempt to solve the general resource allocation problem. Instead, we simplify it by:

1. Assuming that the design to be used is an MRD.

2. Assuming that there is a finite set of candidate points from which the design is to be constructed.

3.1.2 Optimization of Resource Allocation for Schemes with the MRD Design

In this chapter, we focus our attention on the case where the specified design is an MRD. Two optimization problems shall be formulated and solved: the objective function φ of the first is the average of $\mathrm{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)$ over $\mathbf{x} \in R$ whereas that of the second is the average of $\mathrm{var}(\hat{\mu}_{Y|\mathbf{z}})$ over $\mathbf{x} \in R$. The problem of finding schemes that perform well with respect to the two conflicting objectives shall also be considered.

The MRD is the most widely studied and recommended combined array design for RPD experiments. It has three distinct set of points: the factorial points, the axial points, and the center points. The factorial portion of the design is a fractional factorial that is chosen such that all main effects and two-factor interactions corresponding to the response model (2.5) can be estimated. It is a convention to code the high and low levels of each factor in the fractional factorial by $+1$ and -1 respectively. With axial points for the control variables, the pure quadratic coefficients for the control variables can be estimated. In the special case where the axial points are at a distance $\sqrt{k}$ from the origin, at least one center point is also needed. Because there are no axial points for the noise variables and the coded levels of the noise variables in an MRD are either -1, 0, or $+1$, the MRD will be a suitable design for the case where

$$S = \{(z_1, \ldots, z_n); -1 \le z_j \le 1, j = 1, \ldots, n\}. \tag{3.2}$$

We shall assume that S is as given in (3.2) in the remainder of the thesis.

Along with the sample sizes used to estimate the means and variances of the noise variables, the number of replicates of each of the three sets of points in an MRD design determines the values of $\mathrm{var}(\hat{\mu}_{Y|\mathbf{z}})$ and $\mathrm{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)$ at a given $\mathbf{x}$. Therefore,

it is naturally of interest to determine the sample sizes m_j, $j = 1, \ldots n$, the number of

factorial replicates r_f, the number of axial point replicates r_a, and the total number of

runs N (or equivalently, the number of center points r_c) such that the objective

function φ is optimized subject to some constraint on the available resources. The

need for judiciously choosing r_f, r_a, r_c and m_j, $j = 1, \ldots n$ is demonstrated in the

following example.

3.1.3 Motivating Example

Consider the case in which there are two control variables and two noise

variables. Suppose that we set $c_1 = c_2 = 1$ whatever the sample sizes and the true

variance model is

$$\sigma_Y^2 = \sum_{j=1}^{2} \left(\gamma_j + \sum_{i=1}^{2} \delta_{ij} x_i \right)^2 + \sigma^2 = (5 + 6x_1 - 7x_2)^2 + (8 - 4x_1 + 4x_2)^2 + 16.$$

Now, let h_{1j} denote the cost of making one observation on the jth noise

variable, let h_2 denote the cost of performing one experiment run, and let K denote

the available budget/ time for the particular experiment under consideration.

Let R be given by $R = \{(x_1, x_2); -1 \le x_1 \le 1, -1 \le x_2 \le 1\}$, and let the axial

points for the control variables be set at one unit from the origin. Suppose that

$h_{11} = h_{12} = 0.2$, $h_2 = 1$, and $K = 40$. To simplify matters, add the constraint

$m = m_1 = m_2$ to this problem. With an MRD design in which the 2^4 factorial

constitutes one factorial replicate, the experimenter must decide on the values of r_f, r_a,

r_c, and m. We present two possible schemes that costs 40 units each:

A: $m = 10, r_f = 1, r_a = 4, r_c = 4$

B: $m = 40, r_f = 1, r_a = 1, r_c = 4$

In terms of design properties, Scheme A appears to be more attractive since the design size is larger. The larger number of axial points enables each pure quadratic coefficient of the control variables to be estimated with a much smaller variance ($0.125\sigma^2$ for Scheme A versus $0.321\sigma^2$ for scheme B). Considering only experiment error, Scheme A is clearly better than Scheme B.

However, taking into consideration the effect of sampling variation in addition to experiment error, Scheme B turns out to be superior to Scheme A. In fact, the values of $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ and $\text{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)$ for Scheme B are smaller than that for Scheme A everywhere in the region R, as shown in Figures 3.1 and 3.2.

Thus, the experimenter should not pick a scheme arbitrarily or without consideration of variation due to sampling errors because a seemingly reasonable choice may lead to significantly inflated variance.

Remark: Consider the choice of scaling factors $c_1 = c_2 = 1$ in this example. This choice leads to $\tau_{\text{II}} = 0.4$ for Scheme A and $\tau_{\text{II}} = 0.45$ for Scheme B. Consequently, the noise variables are varied over ranges that may be too small for the experiment to effectively capture the range of variation experienced by the response during process operation. Therefore, the scaling factors should be increased and we see that it is not appropriate to choose scaling factors without considering their effect on τ_{II}.
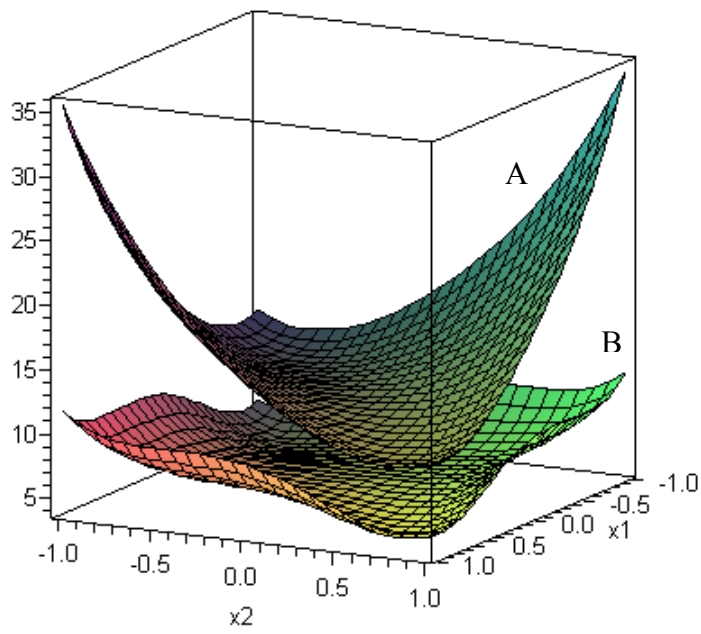
Figure 3.1: Variance of $\hat{\mu}_{Y|\mathbf{z}}$ for Scheme A and Scheme B
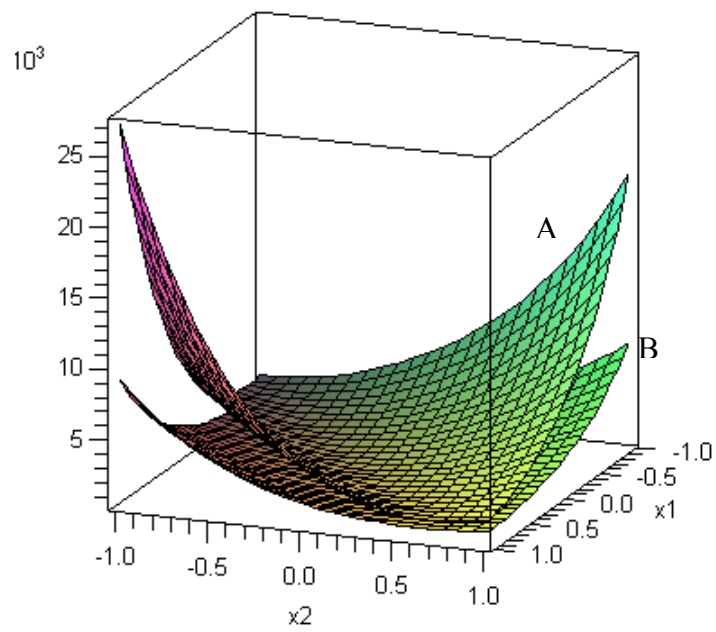
Figure 3.2: Variance of $\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2$ for Scheme A and Scheme B

3.2   Choice of Objective Function

In Section 3.1.3, it is seen that the performance of different schemes in estimating μ_Y and $\sigma_Y^2 - \sigma^2$ may be evaluated by plotting $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ and $\text{var}(\hat{\sigma}_{Y|\mathbf{z}}^2 - \hat{\sigma}^2)$ versus $\mathbf{x}$. However, when $\mathbf{x}$ has three or more elements, it is difficult to compare the performance of different schemes in this manner. Furthermore, when there are many possible schemes, comparison by plotting $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ and $\text{var}(\hat{\sigma}_{Y|\mathbf{z}}^2 - \hat{\sigma}^2)$ versus $\mathbf{x}$ may be awkward. In such cases, it is natural to cast the problem as a mathematical optimization problem with an objective function φ to be optimized. In this section, we discuss briefly, what seems to us some reasonable choices of φ.

Due to research in optimal design theory, many single valued criteria are used for summarizing different aspects of the performance of a design. G-optimality and IV-optimality are two main criteria that quantify a design's performance in prediction. A G-optimal design minimizes the maximum of the variances of the predicted values over the design region while an IV-optimal design minimizes the average of the variances of the predicted values over the design region. For our problem, we consider using summary measures of the behavior of $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ and $\text{var}(\hat{\sigma}_{Y|\mathbf{z}}^2 - \hat{\sigma}^2)$ over R as objective functions. By drawing analogy with optimal design theory, some apparently reasonable alternatives for the objective function φ are the average or maximum of $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ over R and the average or maximum of $\text{var}(\hat{\sigma}_{Y|\mathbf{z}}^2 - \hat{\sigma}^2)$ over R. However, the maximum of $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ and $\text{var}(\hat{\sigma}_{Y|\mathbf{z}}^2 - \hat{\sigma}^2)$ over R tend to occur at points $\mathbf{x}$ where the variance of the response σ_Y^2 is a maximum. Since such points will rarely be

of interest to the researcher, judging the desirability of a scheme by the value of the maximum of $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ or $\text{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)$ can hardly be considered appropriate.

Therefore, it appears that the average of $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ and the average of $\text{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)$ are more reasonable criteria. Note that we consider it more convenient to consider φ as a function of $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ and φ as a function of $\text{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)$ separately. Rather than consider a composite criterion, schemes that perform well when evaluated with respect to both $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ and $\text{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)$ will be found by searching the set of Pareto optimal solutions.

In a particular problem setting, the criterion φ should ideally be chosen to reflect the experimenter's objectives. The average of $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ is an appropriate criterion when the experimenter is interested in estimating μ_Y and the average of $\text{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)$ is an appropriate criterion when the experimenter is interested in estimating $\sigma^2_Y - \sigma^2$ or σ^2_Y. Estimating μ_Y and σ^2_Y is essential when the experimenter faces one of the following situations:

1. The control variables cannot be divided into those that affect the variance of the response and those that affect the mean of the response only. In this situation, tradeoffs between achieving the objective for the mean and achieving the objective for the variance must be considered.

2. The experimenter may want to take into consideration other factors such as cost before deciding on the control variable settings to use. Hence, control variable settings that give a predicted variance slightly higher than the minimum variance may be selected because of lower operating costs.

3. Constraints in design of the product may also exist so that the use of levels of the control variables that give the minimum predicted variance may not be possible. For example, in the case where a component is made of sheet metal, constraints on the supplier's process and standardization of process tooling may necessitate the use of metal sheets of standard thicknesses.

A criterion that is based on $\mathrm{var}(\hat{\mu}_{Y|\mathbf{z}})$ or $\mathrm{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)$ is a natural one in the dual response surface approach to robust parameter design. The distinctive characteristic of this approach to robust parameter design is the construction of response surfaces for the mean and variance models. This is claimed an advantage over Taguchi's approach: "it leads to a better understanding of the system-not just a computation of an optimum condition" (Myers et al., 1992). It is also said that construction of the mean and variance response surfaces allows understanding of the variance-mean tradeoff over the entire design region and gives the decision maker flexibility in selecting alternative product designs or process operating conditions (Myers and Montgomery, 2002; Montgomery, 1999; Myers et al., 1992).

3.3   Design of Scheme for Optimal Estimation of Variance Model

As was discussed, when estimation of the variance model is the primary interest of the experimenter, one reasonable choice for the objective function φ is

$IVV = \int_R \mathrm{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)d\mathbf{x} / \int_R d\mathbf{x}$. In this section, we discuss how values of

$m_j, j = 1,\ldots n$, r_f, r_a, and r_c that minimize IVV may be found.

For an MRD, each $C_{jj}, j = 1,\ldots,n$ is equal to $(fr_f)^{-1}\left(1 + \sum_{i=1}^{k} x_i^2\right)$, where f is

the number of factorial points that constitute one factorial replicate, and $C_{jl} = 0$ for all

$j \neq l$. We consider f a parameter that is specified by the experimenter. Let p denote

the number of model coefficients in the response model and let N denote the total

number of runs. We have $dfSSE = N - p$, where $p = (k + 2 + 2n)(k + 1)/2$. Therefore,

for an MRD design, Equation (3.1) gives

$$\text{var}(\hat{\sigma}_{Y|\mathbf{z}}^2 - \hat{\sigma}^2)$$

$$= \sum_{j=1}^{n} \frac{1}{c_j^4}(\gamma_j + \sum_{i=1}^{k} \delta_{ij} x_i)^4 \left(\frac{2}{m_j - 1} + \frac{\phi_{2j}}{m_j}\right) + 2\sigma^4 \frac{\left(1 + \sum_{i=1}^{k} x_i^2\right)^2}{(fr_f)^2}\left[\sum_{j=1}^{n} \frac{1}{c_j^4} + \frac{1}{dfSSE}\left(\sum_{j=1}^{n} \frac{1}{c_j^2}\right)^2\right]$$

$$+ 4\sigma^2 \frac{1 + \sum_{i=1}^{k} x_i^2}{fr_f} \sum_{j=1}^{n}\left[(\gamma_j + \sum_{i=1}^{k} \delta_{ij} x_i)^2 \frac{1}{c_j^4}\right],$$

(3.3)

where it is assumed that each $\hat{\sigma}_j^2$ is the sample variance, and ϕ_{2j} is the excess kurtosis

of the distribution of the jth noise variable.

Integrating Equation (3.3) over $\mathbf{x} \in R$ and dividing by the volume of R, we get

the following expression for IVV.

$$IVV = \sum_{j=1}^{n}\left(\frac{2}{m_j - 1} + \frac{\phi_{2j}}{m_j}\right)\frac{F_j}{c_j^4} + 2\left(\frac{\sigma^2}{fr_f}\right)^2 G\left[\sum_{j=1}^{n} \frac{1}{c_j^4} + \frac{1}{N - p}\left(\sum_{j=1}^{n} \frac{1}{c_j^2}\right)^2\right] + 4\frac{\sigma^2}{fr_f}\sum_{j=1}^{n} \frac{H_j}{c_j^4},$$

(3.4)

where $F_j = \int_R (\gamma_j + \sum_{i=1}^{k} \delta_{ij} x_i)^4 d\mathbf{x} / \int_R d\mathbf{x}$, $G = \int_R (1 + \sum_{i=1}^{k} x_i^2)^2 d\mathbf{x} / \int_R d\mathbf{x}$, and

$$H_j = \int_R (\gamma_j + \sum_{i=1}^{k} \delta_{ij} x_i)^2 (1 + \sum_{i=1}^{k} x_i^2) d\mathbf{x} / \int_R d\mathbf{x}.$$

Below, we formulate the problem of minimizing IVV as a nonlinear integer program, which we call Program V (explanations of the constraints in the program are also provided).

Program V:

$$\min_{m_1,\ldots,m_n,r_f,N} IVV$$

subject to:

Explanation of Constraint

$$\sum_{j=1}^{n} h_{1j} m_j + h_2 N \leq K$$

Cost constraint.

$$N \geq fr_f + 2k$$

Total number of runs must be greater than or equal to the number of factorial runs and one replicate of axial points.

$$r_f \geq 1$$

Number of factorial replicates must be greater than or equal to one.

$$m_j \geq 2, \, j = 1,\ldots,n$$

Sample size for each noise variable must be at least two so that the variances can be estimated.

$$m_1,\ldots,m_n,r_f,N \text{ are integers}$$

There are several points to note about Program V. Firstly, note that Assumption 2.5 implies that for all $j = 1,\ldots,n$, $\phi_{2j} = 0$. We include the excess kurtosis in (3.3) and (3.4) as a reminder that IVV and consequently Program V can be sensitive to departures from normality. Different values of $\phi_{2j}, j = 1,\ldots,n$ can be tried to assess the sensitivity of the optimal solution to violations of Assumption 2.5.

Secondly, the constraint $N \geq fr_f + 2k$ must be changed to $N \geq fr_f + 2k + 1$ when the axial points are at a distance $\sqrt{k}$ from the origin so that at least a single

center point can be assigned to the design to ensure that $\mathbf{X'X}$ is nonsingular. Thirdly, because one replicate of the fractional factorial allows estimation of all except the pure quadratic terms in the response model, it is always the case that $f + 2k \geq p + 1$. Hence, the constraints $N \geq fr_f + 2k$ and $r_f \geq 1$ ensures that $N \geq p + 1$. Lastly, it can be seen that r_a and r_c are not decision variables in Program V. However, given N and r_f, the possible values of r_a and r_c are limited by the equation $2kr_a + r_c = N - fr_f$.

Define the continuous relaxation of Program V as the nonlinear program that is obtained from Program V by dropping the last constraint. All constraints in the continuous relaxation of Program V are linear functions of the decision variables $m_1, \ldots, m_n$, N, and r_f. In addition, if the integrality requirements on the decision variables are dropped, IVV is a convex differentiable function of those variables on the open set $O_V = \{(m_1, \ldots, m_n, r_f, N); m_j > 1, j = 1, \ldots, n, r_f > 0, N > p\}$. This fact is proven in Appendix C. Let the set of feasible solutions to the continuous relaxation of Program V be denoted by P_V. It can be seen that $P_V \subset O_V$. Therefore, we have the following facts about the continuous relaxation of Program V: its constraints are linear in the decision variables and the objective function of this program is convex and differentiable on an open set that has as its subset the set of feasible solutions. These facts imply that a solution to the continuous relaxation of Program V is a global minimum if and only if the first order Karush-Kuhn-Tucker (KKT) condition is satisfied (Rockafellar, 2007; Bazaraa et al., 1993). This is an important observation because typical nonlinear programming solvers utilize algorithms that converge to the first order KKT condition.

Due to the characteristics of the continuous relaxation of Program V, the global optimal solution of Program V can be obtained by using the branch-and-bound

algorithm (Li and Sun, 2006). In the branch-and-bound algorithm, successive bounds on the decision variables are added as constraints to Program V giving rise to new nodes. At each node, a lower bound for the optimal objective function value is required for deciding whether to prune the node or continue branching from it. A valid lower bound for each node can be obtained by solving the continuous relaxation of the program at the node. Owing to the characteristics of the continuous relaxation of Program V and the fact that bounds on decision variables are linear constraints, the first order KKT condition is necessary and sufficient for a global optimal solution for the continuous relaxation of the program at each node.

There are published studies in the literature that discuss the problem of designing efficient branch-and-bound algorithms for solving nonlinear integer programs. In particular, Gupta and Ravindran (1985) and Sherali and Myers (1985) give detailed descriptions of the branch-and-bound algorithm for solving convex nonlinear integer programs. They investigated the effects of various rules for selecting the branching variables and branching nodes and give recommendations for designing efficient branch-and-bound algorithms. The above studies do not examine the issue of solving the contionos relaxations of the programs generated at the nodes of the branch-and-bound algorithm. However, these programs can be solved by one of many algorithms proposed for solving nonlinear programs and most of these are designed to converge to points that satisfy the first order KKT condition (Bazaraa et al., 1993). Given the developments pointed out above, it is clear that the problem of solving Program V can be achieved by modern mathematical programming methods. In this thesis, Program V and all mathematical programs proposed in later sections are solved by using a software package for solving mathematical programs called Lingo.

3.4   Design of Scheme for Optimal Estimation of Mean Model

In this section, the problem of minimizing $IVM = \int_R \text{var}(\hat{\mu}_{Y|\mathbf{z}}) d\mathbf{x} / \int_R d\mathbf{x}$ is

considered. Recall that $\text{var}(\hat{\mu}_{Y|\mathbf{z}}) = M_S + M_E$, where $M_E = \mathbf{x}'_C \mathbf{V}_C \mathbf{x}_C \sigma^2$. Thus, to

formulate the problem as a nonlinear integer program, the quantity

$IM_E / \sigma^2 = \int_R \mathbf{x}'_C \mathbf{V}_C \mathbf{x}_C d\mathbf{x} / \int_R d\mathbf{x}$ must be expressed explicitly in terms of the decision

variables r_f, r_a, and r_c. Following Khuri and Cornell (1996), we have

$$IM_E / \sigma^2 = \int_R \mathbf{x}'_C \mathbf{V}_C \mathbf{x}_C d\mathbf{x} / \int_R d\mathbf{x} = trace[\mathbf{V}_C \int_R \mathbf{x}_C \mathbf{x}'_C d\mathbf{x} / \int_R d\mathbf{x}]. \tag{3.5}$$

General formulas for $\boldsymbol{\mu}_R = \int_R \mathbf{x}_C \mathbf{x}'_C d\mathbf{x} / \int_R d\mathbf{x}$ can be obtained for two common

cases of R in response surface methodology:

1. The hyper-sphere centered at the origin with radius ρ, which we denote by R_1.

   Mathematically, $R_1 = \{(x_1,\ldots,x_k); x_1^2 + \cdots + x_k^2 \leq \rho^2\}$.

2. The hypercube centered at the origin with sides of length two, which we denote

   by R_2. Mathematically, $R_2 = \{(x_1,\ldots,x_k); -1 \leq x_i \leq 1, i = 1,\ldots,k\}$.

   First, let $\mathbf{x}_1 = (1, x_1, x_2,\ldots,x_k)'$ and $\mathbf{x}_2 = (x_1^2, x_2^2,\ldots,x_k^2, x_1 x_2, x_1 x_3,\ldots, x_{k-1}x_k)'$, and

write $\mathbf{x}'_C = (\mathbf{x}'_1, \mathbf{x}'_2)$. Hence,

$$\boldsymbol{\mu}_R = \int_R \mathbf{x}_C \mathbf{x}'_C d\mathbf{x} / \int_R d\mathbf{x}$$

$$= \int_R \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} (\mathbf{x}'_1 \quad \mathbf{x}'_2) d\mathbf{x} / \int_R d\mathbf{x}$$

$$= \int_R \begin{pmatrix} \mathbf{x}_1 \mathbf{x}'_1 / \int_R d\mathbf{x} & \mathbf{x}_1 \mathbf{x}'_2 / \int_R d\mathbf{x} \\ \mathbf{x}_2 \mathbf{x}'_1 / \int_R d\mathbf{x} & \mathbf{x}_2 \mathbf{x}'_2 / \int_R d\mathbf{x} \end{pmatrix} d\mathbf{x} \tag{3.6}$$

$$= \begin{pmatrix} \boldsymbol{\mu}_{11} & \boldsymbol{\mu}_{12} \\ \boldsymbol{\mu}'_{12} & \boldsymbol{\mu}_{22} \end{pmatrix}.$$

Let $\mathbf{1}_k$ denote a $k \times 1$ vector of 1s and $\mathbf{I}_t$ denote a $t \times t$ identity matrix, where t is a positive integer. Let $\mathbf{0}$ represent a matrix of 0s, with dimensions that shall be clear from the context. Khuri and Cornell (1996) give the following expressions for $\boldsymbol{\mu}_{11}$, $\boldsymbol{\mu}_{12}$, and $\boldsymbol{\mu}_{22}$ for the case where $R = R_1$.

$$\boldsymbol{\mu}_{11} = \begin{pmatrix} 1 & \mathbf{0'} \\ \mathbf{0} & \dfrac{\rho^2}{k+2}\mathbf{I}_k \end{pmatrix}.$$ (3.7)

$$\boldsymbol{\mu}_{12} = \frac{\rho^2}{k+2}\begin{pmatrix} \mathbf{1'}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$ (3.8)

$$\boldsymbol{\mu}_{22} = \frac{\rho^4}{(k+2)(k+4)}\begin{pmatrix} 2\mathbf{I}_k + \mathbf{1}_k\mathbf{1'}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\binom{k}{2}} \end{pmatrix}.$$ (3.9)

For the case where $R = R_2$, it can shown that

$$\boldsymbol{\mu}_{11} = \begin{pmatrix} 1 & \mathbf{0'} \\ \mathbf{0} & \dfrac{1}{3}\mathbf{I}_k \end{pmatrix}.$$ (3.10)

$$\boldsymbol{\mu}_{12} = \frac{1}{3}\begin{pmatrix} \mathbf{1'}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$ (3.11)

$$\boldsymbol{\mu}_{22} = \frac{1}{9}\begin{pmatrix} 0.8\mathbf{I}_k + \mathbf{1}_k\mathbf{1'}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\binom{k}{2}} \end{pmatrix}.$$ (3.12)

The matrices $\boldsymbol{\mu}_{11}$, $\boldsymbol{\mu}_{12}$, and $\boldsymbol{\mu}_{22}$ for other types of region R can be obtained by computing the integrals in Equation (3.6).

An expression for $\mathbf{V}_C$ in terms of r_f, r_a, and r_c is found as follows. For an MRD design, the $\mathbf{X'X}$ matrix has the following form.

$$\mathbf{X'X} = \begin{pmatrix} \mathbf{M}_C & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_D \end{pmatrix},$$

where $\mathbf{M}_C$ corresponds to the columns $\mathbf{X}$ that represent the terms in the mean model whereas $\mathbf{M}_D$ corresponds to the columns of $\mathbf{X}$ that represent the other terms in the response model (the noise main effects and control x noise interactions).

It follows that

$$(\mathbf{X'X})^{-1} = \begin{pmatrix} \mathbf{M}_C^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_D^{-1} \end{pmatrix},$$

This gives us

$$\mathbf{V}_C = \mathbf{M}_C^{-1}.$$

Let α be the distance of the axial points from the origin. It can be shown that $\mathbf{M}_C$ is given by

$$\mathbf{M}_C = \begin{pmatrix} fr_f + 2kr_a + r_c & \mathbf{0} & (fr_f + 2\alpha^2 r_a)\mathbf{1'}_k & \mathbf{0} \\ \mathbf{0} & (fr_f + 2\alpha^2 r_a)\mathbf{I}_k & \mathbf{0} & \mathbf{0} \\ (fr_f + 2\alpha^2 r_a)\mathbf{1}_k & \mathbf{0} & (fr_f)(\mathbf{1}_k\mathbf{1'}_k) + 2\alpha^4 r_a \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & fr_f \mathbf{I}_{\binom{k}{2}} \end{pmatrix}. \quad (3.13)$$

Using the fact that for a square matrix $\mathbf{A}$, $\mathbf{AB} = \mathbf{I}$ when $\mathbf{B} = \mathbf{A}^{-1}$ and $\mathbf{B}$ is unique (Hoffman and Kunze, 2002; Harville, 1997), one may verify that $\mathbf{V}_C$ is as given below.

$$\mathbf{V}_C = \begin{pmatrix} A & \mathbf{0} & B\mathbf{1'}_k & \mathbf{0} \\ \mathbf{0} & \dfrac{1}{(fr_f + 2\alpha^2 r_a)}\mathbf{I}_k & \mathbf{0} & \mathbf{0} \\ B\mathbf{1}_k & \mathbf{0} & D(\mathbf{1}_k\mathbf{1'}_k) + C\mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dfrac{1}{fr_f}\mathbf{I}_{\binom{k}{2}} \end{pmatrix}, \quad (3.14)$$

where

$$A = \frac{kfr_f + 2r_a\alpha^4}{(kfr_f + 2r_a\alpha^4)N - k(fr_f + 2r_a\alpha^2)^2},$$

$$B = \frac{-(fr_f + 2r_a\alpha^2)}{(kfr_f + 2r_a\alpha^4)N - k(fr_f + 2r_a\alpha^2)^2},$$

$$C = \frac{1}{2r_a\alpha^4},$$

$$D = \frac{(fr_f + 2r_a\alpha^2)^2}{(kfr_f + 2r_a\alpha^4)[(kfr_f + 2r_a\alpha^4)N - k(fr_f + 2r_a\alpha^2)^2]} - \frac{fr_f}{(2r_a\alpha^4)(kfr_f + 2r_a\alpha^4)},$$

and $N = fr_f + 2kr_a + r_c$. Observe that when $k = 1$, there is no control x control

interaction. Hence, for $k = 1$, the last columns and rows of $\boldsymbol{\mu}_{22}$ in (3.9) and (3.12),

$\mathbf{M}_C$ in (3.13), and $\mathbf{V}_C$ in (3.14) corresponding to $\mathbf{I}_{\binom{k}{2}}$ should be removed.

Putting together (3.5)-(3.9) and (3.14), the following expression for IM_E / σ^2

for the case where $R = R_1$ is obtained.

$$IM_E / \sigma^2 = A + \frac{2k\rho^2}{k+2}B + \frac{k\rho^2}{(fr_f + 2r_a\alpha^2)(k+2)}$$
$$+ \frac{3k\rho^4}{(k+2)(k+4)}(C+D) + \frac{k(k-1)\rho^4}{(k+2)(k+4)}D + \frac{k(k-1)\rho^4}{2(k+2)(k+4)fr_f}. \tag{3.15}$$

Using Equations (3.5), (3.6), (3.10)-(3.12), and (3.14), the following expression

for IM_E / σ^2 for the case where $R = R_2$ is obtained.

$$IM_E / \sigma^2 = A + \frac{2k}{3}B + \frac{k}{3(fr_f + 2r_a)} + \frac{k}{5}(C+D) + \frac{k(k-1)}{9}D + \frac{k(k-1)}{18fr_f}, \tag{3.16}$$

where A, B, C, and D are obtained from (3.14) by setting $\alpha = 1$.

The average of M_S over R is given by

$$\int_R M_S d\mathbf{x} / \int_R d\mathbf{x} = \sum_{j=1}^{n} \frac{E_j}{m_j c_j^2},$$

where $E_j = \int_R \left(\gamma_j + \sum_{i=1}^{k} \delta_{ij} x_i \right)^2 d\mathbf{x} / \int_R d\mathbf{x}, j = 1, \ldots, n$ and it is assumed that each $\hat{\mu}_j$ is the

sample average. Thus, an expression for $IVM = \int_R \text{var}(\hat{\mu}_{Y|\mathbf{z}}) d\mathbf{x} / \int_R d\mathbf{x}$ is given by

$$IVM = \sum_{j=1}^{n} \frac{E_j}{m_j c_j^2} + IM_E, \tag{3.17}$$

where IM_E is given by (3.15) when $R = R_1$ and is given by (3.16) when $R = R_2$.

   With expressions (3.15)-(3.17), IVM can be written explicitly in terms of the decision variables. Using these results, the minimization problem can be formulated as the following nonlinear integer program, which we call Program M (explanations of the constraints in the program are also provided).

$$\min_{m_1,\ldots,m_n,r_f,r_a,r_c} IVM$$

subject to:

Explanation of Constraint

$$\sum_{j=1}^{n} h_{1j}m_j + h_2(fr_f + 2kr_a + r_c) \le K$$ *Cost constraint.*

$$r_f \ge 1$$ *Number of factorial replicates must be greater than or equal to one.*

Program M:

$$r_a \ge 1$$ *Number of axial point replicates must be greater than or equal to one.*

$$m_j \ge 2, \ j = 1,\ldots,n$$ *Sample size for each noise variable must be at least two so that the variances can be estimated.*

$$r_c \ge 0$$ *Number of center points must be at least zero.*

$m_1,\ldots,m_n,r_a,r_f,r_c$ are integers

   There are several points to note about Program M. Firstly, as long as each $\hat{\mu}_j$ is the sample average, IVM and therefore, Program M is not affected by whether the noise variables are normally distributed. This is in contrast to Program V, which

depends on the excess kurtosis of the distribution of the noise variables. Secondly, the constraint $r_c \geq 0$ must be changed to $r_c \geq 1$ when the axial points are at a distance $\sqrt{k}$ from the origin to ensure that $\mathbf{X'X}$ is nonsingular. Thirdly, although a single observation is sufficient for computing the sample mean, at least two observations are needed for computing the sample variance. Therefore, assuming that both mean and variance models are to be estimated, we must have the constraints $m_j \geq 2$ $j = 1, \ldots, n$.

Now, let us drop the integrality requirements on r_f, r_a, and r_c. For $k \geq 1$, $f > 0$, and $\alpha > 0$, which is always the case, the inverse of the matrix $\mathbf{M}_C$ (3.13) exists, and is given by $\mathbf{V}_C$ (3.14) (where A, B, C, and D are as defined after the equation) for all values of r_f, r_a, and r_c in the set

$$\chi = \{(r_f, r_a, r_c); r_f > 0, r_a > 0, r_c > -(2fr_f r_a)(\alpha^2 - k)^2 / (kfr_f + 2r_a\alpha^4)\}.$$

This result can be obtained by directly verifying that $\mathbf{V}_C$ as given in (3.14) is the inverse of $\mathbf{M}_C$ for all points (r_f, r_a, r_c) in χ. In addition, observe that A, B, C, and D are differentiable with respect to the triple r_f, r_a, and r_c at all points in χ. Therefore, $IM_E / \sigma^2 = trace(\mathbf{V}_C \mathbf{\mu}_R)$, which is a linear function of A, B, C, D, $1/(fr_f + 2\alpha^2 r_a)$, and $1/(fr_f)$, is differentiable at all points in χ. In Appendix D, it is also shown that for any bounded R, which is always the case in practice, IM_E / σ^2 is convex at all points (r_f, r_a, r_c) in the convex set χ. We point out that the differentiability and convexity of IM_E / σ^2 for any bounded region R also follows from the fact that it is a special case of the linear criterion function in optimal design theory (Silvey, 1980), and that the elements of $\mathbf{M}_C$ are linear functions of r_f, r_a, and

r_c. However, optimal design theory does not explicitly provide the set of values of

(r_f, r_a, r_c) over which IM_E / σ^2 is convex and differentiable.

Temporarily forgetting about the integrality requirements on all decision

variables in Program M, it can be seen that IVM is convex and differentiable in the

decision variables $m_1, \ldots, m_n$, r_f, r_a, and r_c when $m_1, \ldots, m_n > 0$, and $(r_f, r_a, r_c) \in \chi$.

Denote by O_M the set $\{(m_1, \ldots, m_n, r_f, r_a, r_c); m_1, \ldots, m_n > 0, (r_f, r_a, r_c) \in \chi\}$ and denote

by P_M the set of feasible solutions to the continuous relaxation of Program M. We see

that $P_M \subset O_M$. Because IVM is convex and differentiable on O_M, and the constraints

in the continuous relaxation of Program M are all linear, a solution to the relaxed

program is a global optimal solution if and only if the first order KKT condition is

satisfied (Rockafellar, 2007; Bazaraa et al., 1993). Consequently, Program M, which

has the requirement of integer-valued decision variables, can be solved for a global

optimal solution with the branch-and-bound algorithm (see the discussion in Section

3.3). A valid lower bound for each node created in the execution of the branch-and-

bound algorithm can be obtained by relaxing the integrality requirements on the

decision variables and solving the resulting mathematical program.

3.5   Pareto Optimal Solutions

In most cases, the optimal solutions for Program M and Program V are

conflicting. This occurs when the optimal values of $m_1, \ldots, m_n$ or r_f for both programs

differ. It can be seen that each decision variable, i.e. $m_j, j = 1, \ldots, n$, r_f, r_a, and r_c

carries about equal weight in the minimization of IVM since the sample observations,

factorial points, axial points, and center points all contribute to the estimation of the

mean model. On the other hand, only $m_j, j = 1, \ldots, n$ and r_f are influential in the minimization of IVV. Given a fixed number of experiment runs N, one often finds that choosing an allocation of r_f, r_a, and r_c such that r_f takes on the maximum possible value minimizes IVV. Therefore, we expect that the optimal solution for Program V can be far from optimal for Program M and vice versa. Since the experimenter is often equally interested in estimating the mean and variance models, some method of finding a compromise solution is needed. In this research, we consider generating a string of Pareto optimal solutions. We assume the generated alternative solutions are presented to the decision maker, who is supposed to choose one from among those solutions for implementation.

To generate a set of Pareto optimal solutions, first solve Program M and Program V. Then, add the constraints $N = fr_f + 2kr_a + r_c$ and $IVV \leq U$ to Program M, where U is greater than or equal to the optimal value of Program V. Let us call the resulting mathematical program, Program P_U. Starting with a value of U near the minimum of IVV, a string of Pareto optimal solutions is obtained by incrementally increasing U and solving Program P_U until the optimal objective value for Program P_U is the same as the optimal objective value for Program M.

The continuous relaxation of Program P_U has a nonlinear constraint $IVV \leq U$. Nevertheless, IVV is convex in the decision variables. This implies that the continuous relaxation of Program P_U is a convex program (Rockafellar, 2007; Bazaraa et al., 1993). Thus, Program P_U can be solved successfully with the branch-and-bound algorithm because the continuous relaxation of the program at each node is a convex program. The first-order KKT condition is sufficient for optimality for the relaxed

program at each node. However, it is not a necessary condition (Rockafellar, 2007; Bazaraa et al., 1993).

3.6   Discussion

In this section, we discuss several issues that concern Program V, Program M, and Program P_U.

Firstly, we point out that the number of decision variables for each of the three programs increases linearly with the number of noise variables. In a practical scenario, the number of decision variables will likely be less than about ten because it is rare for dozens of noise variables to be studied in any one experiment. This can be seen by noting the number of noise variables considered in papers in the RPD literature. For instance, Borkowski and Lucas (1997) provide a catalogue of fractional factorials for MRD designs that covers cases of up to 10 noise variables. We have found that on a Toshiba Portege M6 notebook with two Intel Processors of 2.53GHz, Lingo could solve Program V, Program M, and Program P_U for problems of up to three noise variables in a few seconds. Thus, we believe that the computation effort and solution time required to solve these programs will not be an issue in most cases.

In Section 2.3, it was proposed that the design region S and the scaling factors $c_j, j = 1, \ldots, n$ be specified in such a way that S_ξ is a tolerance region of reasonable size. We assumed that the sample sizes are given in discussing the problem. On the other hand, in considering the problem of optimal allocation in this chapter, we assume that S and $c_j, j = 1, \ldots, n$ are given, and that the sample sizes are decision variables.

Nevertheless, because the noise variables are assumed independently and normally distributed, we may fix S as in (3.2) and use the scaling factors to control τ_{II}.

For the case of minimizing the average of $\mathrm{var}(\hat{\mu}_{Y|\mathbf{z}})$ (Program M), the scaling factors can be adjusted without changing $\mathrm{var}(\hat{\mu}_{Y|\mathbf{z}})$. This implies that the optimal sample sizes are independent of the scaling factors. Thus, we can first choose any values for the scaling factors, solve Program M, and then readjust the scaling factors to achieve a given τ_{II} based on the optimal sample sizes. For the case of minimizing the average of $\mathrm{var}(\hat{\sigma}_{Y|\mathbf{z}}^2 - \hat{\sigma}^2)$ (Program V) and Program $\mathrm{P_U}$, the optimal sample sizes are dependent on the scaling factors because $\mathrm{var}(\hat{\sigma}_{Y|\mathbf{z}}^2 - \hat{\sigma}^2)$ is dependent on the scaling factors. A trial and error approach of specifying the scaling factors can be used to achieve the desired τ_{II} with the optimal sample sizes. However, as was pointed out in Section 2.3, the choice of τ_{II} is generally flexible. Furthermore, Table 2.1 suggests that the values of the scaling factors that give the desired τ_{II} for a large range of sample sizes can be well approximated by the values of the scaling factors that give the desired τ_{II} when all sample sizes become infinitely large. In view of this, the use of asymptotic results for specifying the values of the scaling factors is sufficient in most cases. As a check, the exact value of τ_{II} can be computed after solving Program V or Program $\mathrm{P_U}$. If τ_{II} is within an acceptable range of values, no changes to the scaling factors are required. In the examples in subsequent sections, we shall adopt this approach for all three programs, i.e. Program M, Program V, and Program $\mathrm{P_U}$.

Computation of the quantities G, F_j, and H_j, $j = 1, \ldots, n$ in the objective function of Program V and the quantities E_j, $j = 1, \ldots, n$ in the objective function of Program M requires integration over the region R. In the following, we briefly discuss

how the required integrations can be done for the case where $R = R_1$ and $R = R_2$, where R_1 and R_2 are as defined in Section 3.4. Integration over R_2 is straightforward. One simply integrates over the interval $[-1, 1]$ for each variable in $\mathbf{x}$. For example,

$$E_1 = \int_R \left(\gamma_1 + \sum_{i=1}^{k} \delta_{i1} x_i \right)^2 d\mathbf{x} \Big/ \int_R d\mathbf{x} = \int_{-1}^{1} \cdots \int_{-1}^{1} \left(\gamma_1 + \sum_{i=1}^{k} \delta_{i1} x_i \right)^2 dx_1 \ldots dx_k / 2^k .$$

When $R = R_1$, integration is more complicated. For the case where $k = 2$, integration may be carried out with a transformation to polar coordinates and when $k = 3$, integration may be carried out with a transformation to spherical coordinates. For higher dimensions, an appropriate set of transformations is given by (Edmonson, 1930)

$$
\begin{aligned}
x_1 &= \phi \cos(\theta_1), \\
x_2 &= \phi \sin(\theta_1) \cos(\theta_2), \\
x_3 &= \phi \sin(\theta_1) \sin(\theta_2) \cos(\theta_3), \\
&\vdots \\
x_{k-1} &= \phi \sin(\theta_1) \sin(\theta_2) \cdots \sin(\theta_{k-2}) \cos(\theta_{k-1}), \\
x_k &= \phi \sin(\theta_1) \sin(\theta_2) \cdots \sin(\theta_{k-2}) \sin(\theta_{k-1}); \\
0 &\le \phi \le \rho, \\
0 &\le \theta_1 \le \pi, \\
0 &\le \theta_2 \le \pi, \\
&\vdots \\
0 &\le \theta_{k-2} \le \pi, \\
0 &\le \theta_{k-1} \le 2\pi .
\end{aligned}
\tag{3.18}
$$

In each case, it should be observed that the change of variable formula for multiple integrals should be used (Khuri, 2002). Integration over hyperspheres of high dimension tends to be complicated. However, as noted by Lucas (1974), composite designs for hyperspheres of dimension $k > 4$ with radius $\sqrt{k}$ are seldom used in practice because such designs have axial point distance $\sqrt{k}$ even though the factorial points are at ± 1.

Finally, in choosing the fractional factorial design to use for the MRD, the catalogue provided by Borkowski and Lucas (1997) might be useful. The number of runs of the smallest fractional factorial that allows estimation of all except the pure quadratic terms in the response model is a suitable choice for the value of f (the number of factorial points that constitute a replicate). However, if $r_f > 1$ in the optimal scheme, the experimenter may run a replicate of a larger fractional factorial with fr_f runs if such a fractional factorial exists or two replicates of a fractional factorial with $fr_f / 2$ runs if such a fractional factorial exists, and so on. In other words, a larger fraction replicated so that the total number of runs is the same as the total number of factorial runs in the optimal scheme may be used. The advantage of a larger fractional factorial is that it allows more effects to be estimated. For example, if f represents the number of runs in a quarter fraction and the optimal number of factorial replicates for Program V turns out to be $r_f = 2$, the actual design implemented can be a half fraction so that many more effects are estimable.

3.7   Examples

In the following, we present three examples to illustrate the material we present in this Chapter.

3.7.1 Example 3.1

Consider the motivating example in Section 3.1.3 where the data are the following.

$n = 2$, $k = 2$, $R = R_2$, $c_1 = c_2 = 1$;

$\gamma_1 = 5$, $\gamma_2 = 8$, $\delta_{11} = 6$, $\delta_{21} = -7$, $\delta_{12} = -4$, $\delta_{22} = 4$, $\sigma^2 = 16$;

$h_{11} = 0.2$, $h_{12} = 0.2$, $h_2 = 1$, $K = 40$, $f = 16$.

Numerical integration gives

$$\begin{array}{llll} E_1 = 53.333, & F_1 = 6790.4, & G = \dfrac{133}{45}; & H_1 = 96.444, \\ E_2 = 74.667; & F_2 = 8465.1; & & H_2 = 127.29. \end{array}$$

Adding the constraint $m_1 = m_2 = m$ to Program V and Program M and solving

the programs, the optimal solutions shown in Table 3.1 are obtained. Because $r_f = 1$

and $N = 20$ in the optimal solution for Program V, we must have $r_a = 1$ and $r_c = 0$.

Scheme B in the motivating example is the optimal solution for Program M. In this

case, it is seen that the optimal solution for Program M also performs quite well when

evaluated with respect to IVV.

Table 3.1: Optimal Solutions for Program V and Program M:
$c_1 = c_2 = 1$ (Example 3.1)

	Optimal for Program	
	V	M
IVV	1532.4	1691.1
IVM	12.086	6.2783
m_1	50	40
m_2	50	40
r_f	1	1
r_a	1	1
r_c	0	4

It was pointed out in Section 3.1.3 that the scaling factors are too narrow.

Consider using a new set of scaling factors $\bar{c}_1$ and $\bar{c}_2$. Set $\bar{c}_1 = \bar{c}_2 = 2$ so that

asymptotically, $\tau_{II} = 0.91$. Because the scaling factors change, the coefficients of the

77

response model change. The new set of coefficients are given by $\bar{\gamma}_1 = 10$, $\bar{\gamma}_2 = 16$,

$\bar{\delta}_{11} = 12$, $\bar{\delta}_{21} = -14$, $\bar{\delta}_{12} = -8$, and $\bar{\delta}_{22} = 8$. Let the new set of values for E_j, F_j, and

H_j be represented by $\bar{E}_j$, $\bar{F}_j$, and $\bar{H}_j$. Because $\bar{E}_j = (\bar{c}_j / c_j)^2 E_j$, $\bar{F}_j = (\bar{c}_j / c_j)^4 F_j$,

and $\bar{H}_j = (\bar{c}_j / c_j)^2 H_j$, we have

$$\bar{E}_1 = 213.33, \quad \bar{F}_1 = 108646, \quad \bar{H}_1 = 385.78,$$
$$\bar{E}_2 = 298.67; \quad \bar{F}_2 = 135442; \quad \bar{H}_2 = 509.16.$$

The optimal solutions for Program M and Program V are given in Table 3.2.

They are the same as those given in Table 3.1. However, because of the use of larger

scaling factors, the values of IVV for the optimal solutions are reduced considerably.

The exact values of τ_{II} for both solutions are about 0.9.

Table 3.2: Optimal Solutions for Program V and Program M:
$\bar{c}_1 = \bar{c}_2 = 2$ (Example 3.1)

	Optimal for Program	
	V	M
IVV	847.33	1006.9
IVM	12.086	6.278
m_1	50	40
m_2	50	40
r_f	1	1
r_a	1	1
r_c	0	4

3.7.2 Example 3.2

Consider the following design problem.

$n = 2$, $k = 2$, $R = R_2$, $c_1 = c_2 = 1.5$;

$\gamma_1 = 5$, $\gamma_2 = 8$, $\delta_{11} = 6$, $\delta_{21} = -7$, $\delta_{12} = -4$, $\delta_{22} = 4$, $\sigma^2 = 16$;

$h_{11} = 0.25$, $h_{12} = 0.25$, $h_{12} = 1$, $K = 100$, $f = 16$ (implying a full factorial).

The coefficients γ_j, $j = 1,2$ and δ_{ij}, $i = 1,2$, $j = 1,2$ are the same as those given in Example 3.1 except that they are for larger scaling factors. This implies that for this example, changing either of the noise variables by one standard deviation leads to a smaller absolute change in the response.

A set of five Pareto optimal solutions obtained by solving Program P_U is given in Table 3.3. For each solution in Table 3.3, $\tau_{II} \approx 0.7$. Therefore, the specified scaling factors are acceptable. The optimal solution for Program V is the solution labeled S1 whereas the optimal solution for Program M is the solution labeled S5. It is seen that the optimal solution for Program V performs poorly when evaluated with respect to IVM whereas the optimal solution for Program M performs poorly when evaluated with respect to IVV. Therefore, when estimation of both μ_Y and σ_Y^2 is important, it seems that the solutions labeled S2, S3, and S4 are much better choices.

Table 3.3: Pareto Optimal Solutions: $R = [-1,1]^2$ (Example 3.2)

	S1	S2	S3	S4	S5
IVV	122.45	139.36	159.74	171.37	248.33
IVM	9.4690	2.3502	1.7683	1.7217	1.6827
m_1	91	70	79	68	81
m_2	101	82	93	80	95
r_f	3	3	2	2	1
r_a	1	2	3	4	6
r_c	0	6	13	15	16

Now, suppose that R is a circle of radius $\rho = \sqrt{2}$ instead of the square assumed above, and let $\alpha = \sqrt{2}$. Note that in this case, the design must have at least one center point so that all terms in the response model are estimable. Therefore, the

constraint $r_c \geq 0$ in Program M is changed to $r_c \geq 1$ and the constraint $N \geq fr_f + 2k$

in Program V is changed to $N \geq fr_f + 2k + 1$. Integration with a change of variables to

polar coordinates gives

$$E_1 = 67.5, \quad F_1 = 10612.5, \quad G = 4.3333; \quad H_1 = 149.17,$$
$$E_2 = 80; \quad F_2 = 10752.0; \quad \quad \quad H_2 = 165.33.$$

The optimal solutions for Program V and Program M are given in Table 3.4.

The optimal values of r_f and N for Program V dictate the values of r_a and r_c given

in the table. The solutions in Table 3.4 are very similar to the solutions obtained when

the R is a square, i.e. solutions S1 and S5 in Table 3.3. This suggests that the solutions

to Program M and Program V are not very sensitive to the choice of the region R.

Table 3.4: Optimal Solutions for Program V and Program M:
$R = \{(x_1, x_2); x_1^2 + x_2^2 \leq 2\}$ (Example 3.2)

	Optimal for Program	
	V	M
IVV	173.99	351.56
IVM	7.1632	1.8858
m_1	94	82
m_2	94	90
r_f	3	1
r_a	1	6
r_c	1	17

3.7.3 Example 3.3

Consider an example given by Montgomery (1999) where $n = 3$, $k = 2$,

$c_1 = c_2 = c_3 = 1$, $\tilde{\sigma}^2 = 0.95$, and the fitted response model is

$$\tilde{y} = 30.37 - 2.92x_1 - 4.13x_2 + 2.60x_1^2 + 2.18x_2^2 + 2.87x_1x_2 + 2.73q_1 - 2.33q_2 + 2.33q_3$$
$$- 0.27x_1q_1 + 0.89x_1q_2 + 2.58x_1q_3 + 2.01x_2q_1 - 1.43x_2q_2 + 1.56x_2q_3.$$

In the example, the design used is an MRD design, which consists of a 2_V^{5-1} factorial, one replicate of the axial points for the control variables with axial point distance $\alpha = 2$, and three center points. Asymptotically, $\tau_{II} = 0.32$. Thus, the scaling factors appear to be too small and the results obtained from the experiment may be unrepresentative of actual process conditions.

Suppose we choose to perform another experiment with larger scaling factors $\bar{c}_1$, $\bar{c}_2$, and $\bar{c}_3$. Choose $\bar{c}_1 = \bar{c}_2 = \bar{c}_3 = 2$ so that asymptotically, $\tau_{II} = 0.87$. Rewriting the fitted response model in terms of $\bar{q}_1 = q_1/2$, $\bar{q}_2 = q_2/2$, and $\bar{q}_3 = q_3/2$, we have

$$\tilde{y} = 30.37 - 2.92x_1 - 4.13x_2 + 2.60x_1^2 + 2.18x_2^2 + 2.87x_1x_2 + 5.46\bar{q}_1 - 4.66\bar{q}_2 + 4.66\bar{q}_3$$
$$- 0.54x_1\bar{q}_1 + 1.78x_1\bar{q}_2 + 5.16x_1\bar{q}_3 + 4.02x_2\bar{q}_1 - 2.86x_2\bar{q}_2 + 3.12x_2\bar{q}_3.$$

Let R be the circle centered at the origin with radius 2. Set $\alpha = 2$ and let the 2_V^{5-1} fractional factorial constitute one factorial replicate. Suppose that the cost estimates h_{11}, h_{12}, h_{13}, and budget available K are given by $h_{11} = h_{12} = h_{13} = h_1$, $h_2 = 1$, and $K = 70$. Integration with a change of variables to polar coordinates gives

$$\bar{E}_1 = 46.264, \quad \bar{F}_1 = 4372.8, \qquad\qquad \bar{H}_1 = 149.76,$$
$$\bar{E}_2 = 33.064, \quad \bar{F}_2 = 2207.7, \quad G = \frac{31}{3}; \quad \bar{H}_2 = 106.76,$$
$$\bar{E}_3 = 58.076; \quad \bar{F}_3 = 7853.1; \qquad\qquad \bar{H}_3 = 198.47.$$

The left part of Table 3.5 gives four Pareto optimal solutions for the case where $h_1 = 0.1$ and the right part of Table 3.5 gives four Pareto optimal solutions for the case where $h_1 = 0.2$. The optimal solutions for Program V and Program M are labeled S1 and S4 respectively. All solutions in Table 3.5 seem to perform quite well when evaluated with respect to IVV and IVM.

Table 3.5: Pareto Optimal Solutions: $R = \{(x_1, x_2); x_1^2 + x_2^2 \le 4\}$ (Example 3.3)

	$h_{11} = h_{12} = h_{13} = h_1 = 0.1$				$h_{11} = h_{12} = h_{13} = h_1 = 0.2$			
	S1	S2	S3	S4	S1	S2	S3	S4
IVV	17.011	18.000	18.926	20.175	27.371	27.947	29.999	30.681
IVM	0.52414	0.45916	0.42520	0.42381	0.73102	0.69695	0.67854	0.66035
m_1	164	152	145	132	82	81	79	74
m_2	117	124	123	111	59	62	65	63
m_3	219	184	162	147	109	101	91	83
r_f	1	1	1	1	1	1	1	1
r_a	1	2	2	3	1	1	1	2
r_c	0	0	3	3	0	1	3	2

Suppose that R and α are changed to $R = R_2$ and $\alpha = 1$ respectively.

Numerical integration gives

$$\bar{E}_1 = 35.296, \quad \bar{F}_1 = 1925.0, \qquad \bar{H}_1 = 60.288,$$
$$\bar{E}_2 = 25.498, \quad \bar{F}_2 = 997.09, \quad G = \frac{133}{45}; \quad \bar{H}_2 = 43.506,$$
$$\bar{E}_3 = 33.836; \quad \bar{F}_3 = 2384.3; \qquad \bar{H}_3 = 59.625.$$

Table 3.6 presents a set of four Pareto optimal solutions for the case where

$h_1 = 0.1$. The optimal solution for Program V is labeled S1 whereas the optimal

solution for Program M is labeled S4. Tables 3.5 and 3.6 indicate that the optimal

solutions for Program V and Program M are somewhat insensitive to the choice of R.

However, the optimal solution for Program V performs poorly with respect to IVM

when $R = R_2$, in contrast to the case where $R = R_1$. In fact, in Table 3.6, the optimal

solution for Program V has a value of IVM that is about 2.5 times the minimum

whereas in the left part of Table 3.5, the optimal solution for Program V has a value of

IVM that is about 1.25 times the minimum. The reason for this marked difference is

that estimation of the pure quadratic terms for the control variables is improved with

larger values of α.

Table 3.6: Pareto Optimal Solutions: $R = [-1,1]^2$ (Example 3.3)

	S1	S2	S3	S4
IVV	6.3168	6.6977	7.1991	7.6160
IVM	0.70881	0.33633	0.30342	0.29728
m_1	177	162	145	134
m_2	127	136	116	114
m_3	196	162	149	132
r_f	1	1	1	1
r_a	1	1	2	2
r_c	0	4	5	8

3.8 Greedy Algorithm for Finding Optimal Schemes

In this section, we propose a greedy algorithm for finding schemes that perform well in estimating either the mean model, the variance model, or both models given a candidate set of design points. The algorithm is represented by the following steps.

1. Specify a finite candidate set of design points $\{(\mathbf{x}_l', \mathbf{z}_l'), l = 1, \ldots, L\}$.

2. Set the objective function φ to either IVM, IVV, or $w_1 \left[\frac{IVM}{\min (IVM)} \right] + w_2 \left[\frac{IVV}{\min(IVV)} \right]$, where min (IVM) and min (IVV) are the minimum values of IVM and IVV found so far. In the latter objective, w_1 and w_2, which we may call weights, are positive real numbers such that $w_1 + w_2 = 1$. Specify the cost estimates $K, h_{11}, \ldots, h_{1n}, h_2$.

3. Start with a design $\mathbf{D}_0$ with $N = p + 1$ runs (that allows estimation of the response model). Set $i = 0$, $\mathbf{D}_i^* = \mathbf{D}_0$, and $K_i = K - h_2 N$.

4. Allocate the K_i units of resource to give $m_j = m_{j,i}^*$, $j = 1, \ldots, n$. This is done by

minimizing $\sum_{j=1}^{n} [E_j / (m_j c_j^2)]$ if $\varphi = IVM$, $\sum_{j=1}^{n} \left(\frac{2}{m_j - 1} + \frac{\phi_{2j}}{m_j} \right) \frac{F_j}{c_j^4}$ if $\varphi = IVV$, and

$$\frac{w_1}{\min(IVM)}\sum_{j=1}^{n}[E_j/(m_j c_j^2)] + \frac{w_2}{\min(IVV)}\sum_{j=1}^{n}\left(\frac{2}{m_j-1}+\frac{\phi_{2j}}{m_j}\right)\frac{F_j}{c_j^4}\ \text{if}$$

$\varphi = w_1\left[\frac{IVM}{\min (IVM)}\right] + w_2\left[\frac{IVV}{\min(IVV)}\right]$. (We propose minimizing the latter two

quantities because $E(\hat{\sigma}_j/\sigma_j)\approx 1$ when m_j is large so that $\overline{V}_E$ is approximately

independent of the sample sizes.)

5. If $i = 0$, set φ_0^* equal to the value of φ evaluated at $\mathbf{D}_i^* = \mathbf{D}_0$ and

   $m_j = m_{j,0}^*, j = 1,\ldots,n$. Set φ_{-1}^* to any value greater than φ_0^*, and go to Step 12.

   Otherwise, go to Step 6.

6. Set $l = 1$ and $\varphi_i^* = \varphi_{i-1}^*$.

7. Add the point $(\mathbf{x}_l', \mathbf{z}_l')$ to the design $\mathbf{D}_{i-1}^*$.

8. Evaluate φ for the scheme comprising the design obtained in Step 7 and

   $m_j = m_{j,i}^*, j = 1,\ldots,n$.

9. If $\varphi < \varphi_i^*$, then $\varphi_i^* = \varphi$.

10. If $l = L$, go to Step 11. Otherwise, set $l = l + 1$ and go to Step 7.

11. If $\varphi_i^* \geq \varphi_{i-1}^*$, stop. Return the design $\mathbf{D}_{i-1}^*$ and the sample sizes

    $m_j = m_{j,i-1}^*, j = 1,\ldots,n$. Otherwise, set $\mathbf{D}_i^*$ to be the design corresponding to φ_i^*.

12. Set $i = i + 1$ and $K_i = K_{i-1} - h_2$.

13. If $K_i < 2\sum_{j=1}^{n}h_{1j}$, stop and return the design $\mathbf{D}_{i-1}^*$ and the sample sizes $m_j =$

    $m_{j,i-1}^*, j = 1,\ldots,n$. Otherwise, return to Step 4.

Remark: *IVM* and *IVV* should be computed by integrating Equations (2.20) and (3.1)

respectively, where the expressions in (2.25) and (2.26) are to be substituted for

$E(\hat{\sigma}_j/\sigma_j)$ and $\text{var}(\hat{\sigma}_j^2/\sigma_j^2)$ in Equation (3.1). The equations for *IVM* and *IVV* given

in Sections 3.3 and 3.4 are only valid for the MRD design.

3.8.1 Example 3.4

Consider the case where $k = 1$, $n = 1$, $c_1 = 1$, $\gamma_1 = 1$, $\delta_{11} = 1$, and $\sigma^2 = 1$. Suppose that the cost estimates are $K = 20$, $h_{11} = 0.5$, $h_2 = 1$ and the candidate set of design points (x_1, z_1) is $\{(-1, -1), (0, -1), (1, -1), (-1, 1), (0, 1), (1, 1)\}$.

Let the initial design be specified by $r_1 = \cdots = r_6 = 1$, where $r_1, r_2, \ldots, r_6$ are the number of replicates of each of the six candidate points in the design.

Table 3.7 presents the result of an implementation of the greedy algorithm given in the preceding section with $\varphi = IVM$. The algorithm converges after seven iterations. The final scheme that is obtained is given in the last column of Table 3.7 and for this scheme, $IVM = 0.2658$. An implementation of the greedy algorithm for the case where $\varphi = IVV$ is presented in Table 3.8. We see that the minimum of IVV that is found is 1.0515. Finally, we implement the algorithm for the case where $\varphi = 0.5 \left(\frac{IVM}{0.2658} \right) + 0.5 \left(\frac{IVV}{1.0515} \right)$. The result is given in Table 3.9. Note that we give the values of 100φ in the table, which are percentages, and the ideal percentage is 100.

A comparison of Tables 3.8-3.9 reveals that the sample sizes and design sizes in the optimal schemes for the two cases are the same. The optimal scheme for the case of $\varphi = IVM$ has a slightly larger design. In addition, we see that the optimal scheme for $\varphi = IVM$ has most replications at $(0, -1)$ and $(0, 1)$. Likewise, the scheme that optimizes $\varphi = 0.5 \left(\frac{IVM}{0.2658} \right) + 0.5 \left(\frac{IVV}{1.0515} \right)$ has most replications at these two points. In contrast, the optimal scheme for $\varphi = IVV$ has most replications at $(1, -1)$ and $(1, 1)$.

It should be pointed out that in a few of the iterations shown in Tables 3.7-3.9, there are more than one candidate design point that give the maximum reduction in the value of φ. However, due to Step 9, the lowest indexed candidate point is selected.

Table 3.7: Implementation of Greedy Algorithm with $\varphi = IVM$

i	0	1	2	3	4	5	6	7
IVM	0.4476	0.3713	0.3222	0.3030	0.2889	0.2783	0.2685	0.2658
IVV	1.9315	1.7693	1.6088	1.5550	1.4975	1.4894	1.3563	1.2078
r_1	1	1	1	1	1	2	2	2
r_2	1	2	2	3	3	3	3	3
r_3	1	1	1	1	1	1	1	2
r_4	1	1	1	1	1	1	1	1
r_5	1	1	2	2	3	3	3	3
r_6	1	1	1	1	1	1	2	2
m_1	28	26	24	22	20	18	16	14

Table 3.8: Implementation of Greedy Algorithm with $\varphi = IVV$

i	0	1	2	3	4	5	6
IVV	1.9315	1.5813	1.2696	1.1797	1.0965	1.0675	1.0515
IVM	0.4476	0.4339	0.4222	0.4216	0.4222	0.4121	0.4056
r_1	1	1	1	1	1	2	2
r_2	1	1	1	1	1	1	1
r_3	1	2	2	3	3	3	3
r_4	1	1	1	1	1	1	2
r_5	1	1	1	1	1	1	1
r_6	1	1	2	2	3	3	3
m_1	28	26	24	22	20	18	16

Table 3.9: Implementation of Greedy Algorithm with $\varphi = 0.5\left(\frac{IVM}{0.2658}\right) + 0.5\left(\frac{IVV}{1.0515}\right)$

i	0	1	2	3	4	5	6
100φ	176.0273	153.96	137.1005	124.8193	112.3552	109.0376	107.0813
IVM	0.4476	0.3713	0.3222	0.3095	0.3000	0.2828	0.2722
IVV	1.9315	1.7693	1.6088	1.4009	1.1763	1.1744	1.1753
r_1	1	1	1	1	1	1	1
r_2	1	2	2	2	2	3	3
r_3	1	1	1	2	2	2	2
r_4	1	1	1	1	1	1	1
r_5	1	1	2	2	2	2	3
r_6	1	1	1	1	2	2	2
m_1	28	26	24	22	20	18	16

CHAPTER 4

TWO ISSUES OF PRACTICAL INTEREST IN DESIGN

4.1   Introduction

In this chapter, we address two issues of practical interest. Firstly, observe that before Program V and Program M can be solved, the values of the parameters γ, Δ, and σ^2 must be specified. These are unknown quantities and therefore, it is not obvious as to how Program V and Program M can be utilized in practice. In the first part of this chapter, we discuss how this problem may be overcome. We show how Program V and Program M may be modified when prior knowledge is captured in the form of a prior distribution for the unknown parameters. In addition, we discuss the application of robust optimization ideas to handle uncertainty in estimates of the parameters γ, Δ, and σ^2.

Another problem of practical interest is the comparison of schemes comprising different types of designs. Designs other than the MRD can be used for an RPD experiment. Borror et al. (2002), Robinson et al. (2004), and Castillo et al. (2007) discuss these possibilities. However, Program V and Program M are limited only to finding optimal schemes when the design is constrained to be an MRD. Even though the MRD designs possess many attractive properties, there may be other more desirable designs for a particular problem. For example, when the experimenter's

secondary objective is to estimate the model coefficients as precisely as possible, a D-optimal design is appealing because it minimizes the volume of the confidence ellipsoid for the coefficients of the response model in $\mathbf{x}$ and $\mathbf{z}$. In evaluating alternative schemes, the average of $\mathrm{var}(\hat{\mu}_{Y|\mathbf{z}})$ and $\mathrm{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)$ may not give a good idea of the performance of the schemes over the entire region R. Considering only experiment error, it is known that designs can have a small average for $\mathrm{var}(\hat{\mu}_Y)$ but very large values for $\mathrm{var}(\hat{\mu}_Y)$ at certain points in R. Hence, a graphical tool that gives a more comprehensive picture of the values of $\mathrm{var}(\hat{\mu}_{Y|\mathbf{z}})$ and $\mathrm{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)$ over R can be helpful for evaluating alternative schemes. In the last part of this chapter, we show how schemes with different types of designs can be compared with graphical plots called cumulative distribution plots, which are modifications of the fraction of design space (FDS) plots introduced by Zahran et al. (2003).

4.2 Problem of Unknown Parameters

To allocate the resources of an experiment using Program M and Program V, the unknown parameters $\boldsymbol{\gamma}$, $\boldsymbol{\Delta}$, and σ^2 must be specified or estimated. A similar problem occurs in nonlinear experimental design (Ford et al., 1989) in which the design that is best with respect to a design criterion usually depends on the parameters of the model. Referring to this problem, Steinberg and Hunter (1984) comment "investigators are thus in the rather paradoxical position of having to know at the design stage the very quantities that they are conducting the experiment to estimate!" Similarly, Cochran (1973) remarks that this problem places the statistician in a difficult position, which is literally like telling the experimenter "you tell me the value of θ and

I promise to design the best experiment for estimating θ." To date, there still seems to be no completely satisfactory method of dealing with this problem. However, there are some methods proposed in the literature on nonlinear experimental design for solving the problem. These are reviewed in Sections 4.2.1-4.2.3. In Section 4.2.4, we discuss how the methods reviewed in Sections 4.2.1-4.2.3 can be applied to solve the problem of specifying γ, Δ, and σ^2.

4.2.1 Point Estimates and Prior Distributions

In nonlinear experimental design, either point estimates or prior distributions are specified for the unknown parameters. Technically, a prior distribution is simply a distribution from which it is assumed that an unknown parameter of another distribution is drawn. The use of prior distributions, however, does not necessarily imply that the design criteria used must be motivated by Bayesian considerations (Atkinson, 1996; Atkinson et al., 1995; Chaloner and Verdinelli, 1995). When a point estimate is available, a design that is optimal with respect to the point estimate may be derived. When a prior distribution for the unknown parameters is available, a design that optimizes the expected value of the design criterion taken with respect to the prior distribution can be obtained (Atkinson, 1996; Atkinson et al., 1995; Atkinson and Donev, 1992; Pronzato and Walter, 1985; Atkinson, 1982). Pronzato and Walter (1985) discuss ED-optimal designs for nonlinear models and algorithms for constructing such designs. The ED-criterion is defined as the expectation of the determinant of the Fisher information matrix taken with respect to the prior distribution for the unknown parameters.

There are two ways to obtain point estimates and prior distributions for unknown parameters: prior knowledge and sequential experimentation.

4.2.2 The Use of Prior Knowledge

In problems of nonlinear experimental design in which all runs are to be performed before any analysis of the experiment takes place and the optimal design depends on unknown parameters to be estimated, it is necessary to rely on prior knowledge of the values of the parameters in choosing a design. A prior distribution is used to capture prior knowledge of the values of the unknown parameters. Prior distributions are usually elicited from the experimenter by asking him or her large numbers of simple questions (Press, 2003; Kiefer, 1987). Specific elicitation methods are discussed by Press (2003), Chaloner et al. (1993), and Garthwaite and Dickey (1988). However, Kiefer (1987) remarks that since eliciting prior distributions are often difficult and time-consuming, many Bayesians do not pretend to go through a formal process for eliciting the required prior distributions. He claims that in many practical settings, the prior distributions used are simply rough summaries of the statistician's feelings about the chances of the various states of nature. It is important to point out that prior distributions that quantify the opinion of a person do not have a physical meaning and they are referred to as "subjective" probability laws (Kiefer, 1987). Nevertheless, in some special cases, prior distributions can be specified based on past experiments and past data (Press, 2003; Chaloner and Verdinelli, 1995) so that the element of subjectivity is reduced. In some cases, the experimenter may be willing to provide a guess of the values of the parameters. A point estimate obtained in this way is considered a special type of prior distribution, called a degenerate prior.

4.2.3 Sequential Experimentation

Another way of dealing with the problem of unknown parameters in nonlinear experimental design is to perform experiment runs sequentially. Sequential designs are constructed by adding one run at a time or a number of runs at a time (Ford et al., 1989). After each run or each batch of runs, estimates of the unknown parameters are updated and the next run or next batch of runs is chosen to optimize some design criterion evaluated at the updated estimates. Repeated sampling inference is difficult in the case of sequential designs. Ford et al. (1985) point out the dependence of a design point on the preceding set of design points and observations, and argue that this dependence should not be ignored in the construction of valid confidence intervals. This implies that inference made as if the achieved design were fixed at the start of the experiment is not strictly correct. Theoretical research has focused on providing asymptotic justifications to validate certain inference procedures (for example, see Chaudhuri and Mykland (1993)). However, the conditions required for the asymptotic results to hold are often difficult to verify (Atkinson and Bailey, 2001; Ford et al., 1989).

Note that it is sometimes assumed that point estimates are obtained from some preliminary experiment (Sitter and Wu, 1999; Herzberg and Cox, 1969). However, unless the preliminary experiment is performed on another system of similar characteristics and not on the system on which the planned experiment is to be carried out, it should rightly be regarded as the first phase of a sequence of experiments (Sitter and Wu, 1999).

4.2.4 Specification of γ, $\boldsymbol{\Delta}$, and σ^2

Based on the discussions in the preceding sections, it is seen that either point estimates or a prior distribution can be specified for the unknown parameters γ, $\boldsymbol{\Delta}$, and σ^2. If no experiment precedes the planned experiment, the experimenter can either guess the values of γ, $\boldsymbol{\Delta}$, and σ^2 or use a prior distribution that roughly summarizes his or her belief about the parameters. If each parameter can be assumed independent a priori, percentiles of the prior distribution for each parameter can be assessed using a method given by Press (2003) (page 86). For the case where there is more than one person involved in the RPD experiment and it is desired to use a prior distribution that reflects the belief of all the experimenters, the method of assessing a subjective prior distribution for a group discussed by Press (2003) (pages 94-97) can be utilized. However, there remains the problem of developing a method to assess a joint prior distribution for the parameters for the case where we cannot assume that the parameters are independent. Optimization of resource allocation based on a prior distribution for γ, $\boldsymbol{\Delta}$, and σ^2 is discussed in subsequent sections.

A sequential procedure for our problem would conceivably involve alternating between sampling and performing experiment runs. Such a procedure seems to present serious inference problems as in sequential design for nonlinear models. In addition, when the data is obtained in a sequential manner, the repeated sampling properties of the estimators for the mean and variance models are likely to be very different from those derived in Chapter 2.

Nevertheless, sequential experimentation is a highly recommended practice (Box et al. 2005; Box, 1993, Myers et al., 1992). We suggest the following simple but possibly sub-optimal two-stage procedure. First, collect some process data to estimate

the means and variances of the noise variables, and perform a screening experiment. Use the data to determine the active factors and to estimate the unknown parameters $\boldsymbol{\gamma}$, $\boldsymbol{\Delta}$, and σ^2 by $\hat{\boldsymbol{\gamma}}_z$, $\hat{\boldsymbol{\Delta}}_z$, and $\hat{\sigma}^2$ respectively. These activities constitute the first stage. Thus, the purpose of the first stage is to obtain the necessary information to optimize the design of the second stage, which has the objective of estimating the mean and variance models. A Bayesian analysis of the screening experiment may also be performed to obtain a posterior distribution for $\boldsymbol{\gamma}$, $\boldsymbol{\Delta}$, and σ^2, which will be a prior distribution for the second stage. Next, carry out the second stage according to the proposed procedure given in Figure 2.1. In particular, optimize the allocation of resource for the second stage using the point estimates or prior distribution for $\boldsymbol{\gamma}$, $\boldsymbol{\Delta}$, and σ^2 obtained in the first stage. Then, collect process data and perform the experiment as planned to estimate the mean and variance models. Note that available resource can probably be better utilized if more resource is allocated to the second stage. This is because estimation of the mean and variance models is the main objective, and resource allocation can be optimally planned in the second stage.

4.3   Expected Variance Criteria

When point estimates for the unknown parameters $\boldsymbol{\gamma}$, $\boldsymbol{\Delta}$, and σ^2 are available, they can be used in place of the parameters in solving Program V and Program M. In other words, the point estimates may be treated as if they were the true values. Examples 3.1-3.3 can be viewed as examples where schemes that are optimal with respect to point estimates for the unknown parameters are found. When a prior distribution for the unknown parameters is specified, the criteria *IVV* and *IVM* must

be modified to incorporate uncertainty in the parameters. In this section, we propose

modifications to Program M, Program V, and Program P_U to allow for the use of a

prior distribution for γ, Δ, and σ^2.

Let the elements of γ, Δ, and σ^2 be concatenated in a vector Λ and let

$\underset{\Lambda}{E}(\cdots)$ denote the expectation of the quantity in the brackets with respect to Λ. The

expectation is obtained by multiplying the quantity in the brackets by the prior $P(\Lambda)$

of Λ and integrating over the sample space of Λ. Schemes that minimize $\underset{\Lambda}{E}(IVV)$

appear to be good candidates for estimating σ_Y^2 since they minimize an average of

IVV values, weighted by their plausibility of occurrence. In a similar sense, schemes

that minimize $\underset{\Lambda}{E}(IVM)$ appear to be good candidates for estimating μ_Y. Thus, we

consider replacing IVM and IVV with $\underset{\Lambda}{E}(IVM)$ and $\underset{\Lambda}{E}(IVV)$ respectively. Observe

that these criteria are analogous to the ED-criterion mentioned in Section 4.2.1.

The quantities $\underset{\Lambda}{E}(IVV)$ and $\underset{\Lambda}{E}(IVM)$ are given by

$$\underset{\Lambda}{E}(IVV)$$

$$= \sum_{j=1}^{n}\left(\frac{2}{m_j-1}\right)\frac{\underset{\Lambda}{E}(F_j)}{c_j^4} + 2\frac{\underset{\Lambda}{E}(\sigma^4)}{(fr_f)^2}G\left[\sum_{j=1}^{n}\frac{1}{c_j^4} + \frac{1}{N-p}\left(\sum_{j=1}^{n}\frac{1}{c_j^2}\right)^2\right] + \frac{4}{fr_f}\sum_{j=1}^{n}\frac{\underset{\Lambda}{E}(H_j\sigma^2)}{c_j^4},$$

and $\underset{\Lambda}{E}(IVM) = \sum_{j=1}^{n}\frac{\underset{\Lambda}{E}(E_j)}{m_j c_j^2} + \frac{IM_E}{\sigma^2}\underset{\Lambda}{E}(\sigma^2)$,

where we set $\phi_{21} = \phi_{22} = \cdots = \phi_{2n} = 0$ in the expression for IVV. Note that by

definition, $IM_E/\sigma^2 = \int_R \mathbf{x'}_C \mathbf{V}_C \mathbf{x}_C d\mathbf{x}/\int_R d\mathbf{x}$; hence, it does not depend on Λ.

Computation of the quantities $\underset{\Lambda}{E}(E_j)$, $\underset{\Lambda}{E}(F_j)$, and $\underset{\Lambda}{E}(H_j\sigma^2)$ can be done in the

following way. First, express E_j, F_j, and H_j explicitly in terms of the elements of γ

and $\mathbf{\Lambda}$, i.e. obtain an explicit expression for the integrals defining those terms. It is straightforward to perform the integrations required for E_j, F_j, and H_j by hand when $R = R_1$ or when $R = R_2$. In addition, mathematical software such as MATLAB and MAPLE can be used to perform the integrations. Next, multiply each E_j, F_j, and $H_j \sigma^2$ by the prior $P(\mathbf{\Lambda})$ and integrate over the sample space of $\mathbf{\Lambda}$. This gives the expectation with respect to $\mathbf{\Lambda}$. For some priors, the expectation can be convenient to compute using standard formulas. Alternatively, one can use numerical integration or Monte Carlo simulation to do the computation. To illustrate, consider the case where $R = R_2$. By expanding the integrands and carrying out the integrations in the definitions of E_j, F_j, and H_j, we obtain for $j = 1,\ldots,n$,

$$E_j = \gamma_j^2 + \frac{1}{3}\sum_{i=1}^{k}\delta_{ij}^2, \tag{4.1}$$

$$F_j = \begin{cases} \gamma_j^4 + \frac{1}{5}\sum_{i=1}^{k}\delta_{ij}^4 + \frac{2}{3}\sum_{l=2}^{k}\sum_{i=1}^{l-1}\delta_{ij}^2\delta_{lj}^2 + 2\gamma_j^2\sum_{i=1}^{k}\delta_{ij}^2, & k \geq 2, \\ \gamma_j^4 + \frac{1}{5}\delta_{1j}^4 + 2\gamma_j^2\delta_{1j}^2, & k = 1, \end{cases} \tag{4.2}$$

and $H_j = \left(1 + \frac{k}{3}\right)\gamma_j^2 + \left(\frac{8}{15} + \frac{k-1}{9}\right)\sum_{i=1}^{k}\delta_{ij}^2$. \hfill (4.3)

Now, if we set $\sigma^2 = 1$ and each parameter $\gamma_j, \delta_{1j},\ldots,\delta_{kj}, j = 1,\ldots,n$ is assigned a normally and independently distributed prior with mean 0 and variance σ_P^2, then

$$\underset{\Lambda}{E}(E_j) = \underset{\Lambda}{E}\left(\gamma_j^2 + \frac{1}{3}\sum_{i=1}^{k}\delta_{ij}^2\right) = \sigma_P^2\left(1 + \frac{k}{3}\right), \tag{4.4}$$

$$\underset{\Lambda}{E}(F_j) = \underset{\Lambda}{E}\left(\begin{cases} \gamma_j^4 + \frac{1}{5}\sum_{i=1}^{k}\delta_{ij}^4 + \frac{2}{3}\sum_{l=2}^{k}\sum_{i=1}^{l-1}\delta_{ij}^2\delta_{lj}^2 + 2\gamma_j^2\sum_{i=1}^{k}\delta_{ij}^2, & k \geq 2 \\ \gamma_j^4 + \frac{1}{5}\delta_{1j}^4 + 2\gamma_j^2\delta_{1j}^2, & k = 1 \end{cases}\right)$$
$$= \frac{(k + 9/5)(k + 5)}{3}\sigma_P^4, \tag{4.5}$$

and $E(H_j\sigma^2) = E\left[\left(1+\dfrac{k}{3}\right)\gamma_j^2 + \left(\dfrac{8}{15}+\dfrac{k-1}{9}\right)\sum_{i=1}^{k}\delta_{ij}^2\right] = \dfrac{(k+9/5)(k+5)}{9}\sigma_P^2.$ (4.6)

It can be seen that IVV and $\hat{E}(IVV)$ are of the same form when written as functions of the decision variables. Likewise, IVM and $\hat{E}(IVM)$ are of the same form when written as functions of the decision variables. Thus, replacing IVV with $\hat{E}(IVV)$ and IVM with $\hat{E}(IVM)$ in Program M, Program V, and Program P_U does not change any characteristics of the mathematical programs. In particular, no change in solution method is required. In the following, two numerical examples are given. In the examples, it is assumed that the experimenter specifies a degenerate prior for σ^2, $\sigma^2 = 1$.

4.3.1 Example 4.1

Suppose $n = 2$, $k = 3$, $R = \{(x_1, x_2, x_3); -1 \le x_i \le 1, i = 1,2,3\}$, and $c_1 = c_2 = 1.5$ (so that asymptotically, $\tau_{II} = 0.75$). Assign to each $\gamma_j, \delta_{ij}; i = 1,2,3, j = 1,2$ a uniform prior density over the interval $[-5, 5]$. Assume that $h_{11} = 0.2$, $h_{12} = 0.2$, $h_2 = 1$, $K = 51$, and choose $f = 16$. The 16 distinct factorial points correspond to those of a resolution V fractional factorial. Integration gives $G = 64/15$. Using Equations (4.1)-(4.3) and Monte-Carlo simulation with 30,000 runs, we obtain

$\hat{E}(F_1) = \hat{E}(F_2) = 753.11$, $\hat{E}(H_1) = \hat{E}(H_2) = 35.45$, and $\hat{E}(E_1) = \hat{E}(E_2) = 16.62$.

Minimization of $\hat{E}(IVV)$ gives $m_1 = 72$, $m_2 = 73$, $r_f = 1$, and $N = 22$. For this optimal scheme, $\hat{E}(IVV) = 11.84$, and $\hat{E}(IVM) = 0.5510$. Note that $r_f = 1$ and $N = 22$ implies $r_a = 1$ and $r_c = 0$.

Minimization of $\hat{E}(IVM)$ gives $m_1 = 55$, $m_2 = 55$, $r_f = 1$, $r_a = 2$, and $r_c = 1$.

For this optimal scheme, $\hat{E}(IVV) = 14.54$, and $\hat{E}(IVM) = 0.4626$.

Both solutions perform almost equally well with respect to both objectives. Therefore, it does not really matter which scheme is implemented.

4.3.2 Example 4.2

Consider the problem given in Example 4.1. Suppose now that we assign to each γ_j, δ_{ij}; $i = 1,2,3, j = 1,2$ a normally and independently distributed prior density with mean 0 and variance $\sigma_P^2 = 9$. Suppose $K = 100$ and that all other parameters are the same as in the previous example. Using Equations (4.4)-(4.6), we obtain

$$\hat{E}(F_1) = \hat{E}(F_2) = 1036.8, \ \hat{E}(H_1) = \hat{E}(H_2) = 38.4, \text{ and } \hat{E}(E_1) = \hat{E}(E_2) = 18.$$

Minimization of $\hat{E}(IVV)$ gives $m_1 = 155$, $m_2 = 155$, $r_f = 2$, and $N = 38$. For this scheme, $\hat{E}(IVV) = 7.2194$ and $\hat{E}(IVM) = 0.41180$. Note that $r_f = 2$ and $N = 38$ implies $r_a = 1$ and $r_c = 0$.

Minimization of $\hat{E}(IVM)$ gives $m_1 = 133$, $m_2 = 132$, $r_f = 1$, $r_a = 5$, and $r_c = 1$. For this scheme, $\hat{E}(IVV) = 10.036$ and $\hat{E}(IVM) = 0.23486$.

Again, it appears that both solutions perform almost equally well when judged by the criteria $\hat{E}(IVV)$ and $\hat{E}(IVM)$.

4.4   Robust Optimization

In the case where there is considerable uncertainty in the estimates of $\boldsymbol{\gamma}$, $\boldsymbol{\Delta}$, and σ^2, it may be desirable to utilize the robust optimization approaches of Ben-Tal and Nemirovski (1998) and Xu and Albin (2003). Ben-Tal and Nemirovski (1998) propose using a minimax objective to deal with uncertainty in the parameters of a mathematical program whereas Xu and Albin (2003) propose the use of a minimax deviation objective for response surface optimization. Program M and Program V can be converted into programs with a minimax or a minimax deviation objective. Assume that a confidence interval is available for each element of $\boldsymbol{\Lambda}$ so that the Cartesian product of the intervals form a hypercube Θ (recall that $\boldsymbol{\Lambda}$ represents $\boldsymbol{\gamma}$, $\boldsymbol{\Delta}$, and σ^2). Consider the cases where $R = R_1$ or $R = R_2$. It can be shown that E_j, F_j, and $H_j, j = 1,\ldots,n$ are functions of the squares of each element in $\boldsymbol{\gamma}$ and $\boldsymbol{\Delta}$. This is evident from Equations (4.1)-(4.3) for the case where $R = R_2$. For the case where $R = R_1$, we can see that E_j, F_j, and $H_j, j = 1,\ldots,n$ are functions of the squares of each element in $\boldsymbol{\gamma}$ and $\boldsymbol{\Delta}$ by expanding the integrands in the definition of those terms and noting that $\int_{R_1} x_1^{a_1} x_2^{a_2} \ldots x_k^{a_k} dx_1 dx_2 \ldots dx_k = 0$ whenever one of $a_i, i = 1,\ldots,k$ is an odd integer. It follows that the minimax objectives for Program V and Program M are $\min(IVV|\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^{\max})$ and $\min(IVM|\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^{\max})$, where $\boldsymbol{\Lambda}^{\max}$ is any vector of maximum norm in Θ. These objectives have the same functional forms as IVV and IVM, and so the resulting programs may be solved in the same way as Programs V and M. Theorem 1 in Xu and Albin (2003) can be used to formulate the minimax deviation objective for IVM as a tractable mathematical program. Define $\boldsymbol{\theta}_M = (E_1,\ldots,E_n,\sigma^2)$.

Since the set $\Omega_M = \{\boldsymbol{\theta}_M; \boldsymbol{\theta}_M(\boldsymbol{\Lambda}), \boldsymbol{\Lambda} \in \Theta\}$ is a hypercube, we can convert the semi-infinite program that results from employing the minimax deviation objective to a finite optimization problem (see Theorem 1 in Xu and Albin (2003)). Let $\boldsymbol{\theta}_M^i, i = 1, 2, \ldots, 2^{n+1}$ be the extreme points of Ω_M. The finite optimization problem has constraints $IVM(\boldsymbol{\theta}_M^i) - \min_{m_1, \ldots, m_n, r_f, r_a, r_c} IVM(\boldsymbol{\theta}_M^i) \leq M, i = 1, \ldots, 2^{n+1}$, in addition to the constraints in Program M, and has the objective min M. This problem is a convex nonlinear integer program. Unlike the case with IVM, Xu and Albin's (2003) result does not apply to IVV due to the functional relationship between F_j and H_j.

4.5   Cumulative Distribution Plots for Comparing Alternative Schemes

In this section, we introduce the cumulative distribution plots for comparing alternative schemes. The plots can be constructed with either a point estimate or a prior distribution for $\boldsymbol{\Lambda}$. When a point estimate is used, the construction of a cumulative distribution plot is the same as that of an FDS plot (Zahran et al., 2003) and it can be interpreted in the same manner as an FDS plot. However, the cumulative distribution plots can also be constructed with a prior distribution for the unknown parameters. Because of the different interpretation of the plots in this case, we call the plots cumulative distribution (CD) plots instead of FDS plots. In this section, we discuss the construction and interpretation of CD plots for comparing schemes based on $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ and $\text{var}(\hat{\sigma}_{Y|\mathbf{z}}^2 - \hat{\sigma}^2)$. We call a CD plot constructed with the former criterion a CD plot for the mean model and a CD plot constructed with the latter criterion a CD plot for the variance model.

To construct a CD plot for the mean model with a prior distribution for $\mathbf{\Lambda}$, sample a value for $\mathbf{\Lambda}$ from the prior distribution and a value for $\mathbf{x}$ from the uniform probability density over R. Using the sampled values, compute $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ for each scheme. Repeat the procedure r times for some large number r, order the r values of $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ for each scheme, and plot them versus the quantiles $1/r, 2/r, \ldots, 1$. This is similar to the procedure described by Ozol-Godfrey et al. (2005) for constructing FDS plots but with the added step of sampling from a prior distribution for $\mathbf{\Lambda}$. CD plots for the variance model are constructed in the same way as a CD plot for the mean model except that the values of $\text{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)$ are computed and plotted. Note that $\mathbf{x}$ can be sampled from probability densities other than the uniform density. These place unequal weights over R. The probability density for $\mathbf{x}$ can be viewed as a prior density that summarizes the decision maker's belief about the chances that prediction would be made at various points in R. However, in this thesis, we consider only drawing values of $\mathbf{x}$ from a uniform density.

CD plots constructed with a point estimate for $\mathbf{\Lambda}$ are essentially FDS plots. In this case, we sample $\mathbf{x}$ from a uniform distribution, compute the values of $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ or $\text{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)$ for each scheme, and order and plot the values versus the quantiles. Thus, at a point on the graph for a scheme, the x-coordinate gives the fraction of the volume of R with a variance value at or below the value of the y-coordinate (Zahran et al., 2003). We may also interpret the x-coordinate as the probability that a point $\mathbf{x}$ chosen randomly from R will give a variance value at or below the value of the y-coordinate. On the other hand, when a CD plot is constructed with a prior density for $\mathbf{\Lambda}$, the x-coordinate of a point on the graph for a scheme should be interpreted as the

probability that an **x** chosen randomly from R and a value of $\mathbf{\Lambda}$ sampled from the prior density will give a variance value at or below the value of the y-coordinate.

A cumulative distribution plot is shown in Figure 4.1. We can obtain from a CD plot various performance measures for each scheme that is being evaluated. For instance, we may compare the schemes based on the median variance, the interquartile range of the variance, and the average/expected variance (which is the arithmetic mean of the variance values used to construct the CD plot). The decision maker has the flexibility to compare schemes based on any performance measure that can be derived from the CD plot. The performance measure used for any particular experiment should depend on the preference of the decision maker and the goals of the experiment. For instance, a risk-averse decision maker might prefer a scheme that minimizes the 90th percentile variance value. If this criterion is used, Scheme 2 in Figure 4.1 is superior to Scheme 1. On the other hand, when one goal of the experiment is to achieve a certain precision in prediction of μ_Y or σ_Y^2 over R, a scheme that has maximum probability of achieving that precision, as measured by the variance, might be chosen. Explicitly defining a criterion for comparing schemes will be important to avoid ambiguous comparisons especially when the graphs for the schemes being compared crosses, as is the case with the graphs in Figure 4.1.

It may sometimes be preferred to make a pairwise comparison of schemes. A reasonable way to do this is to employ a CD plot for the difference in variance for each pair of schemes. Suppose that we intend to compare the performance of Scheme 1 and Scheme 2 in Figure 4.1. We may do so by plotting the CD plot for the difference in variance $[\text{var}(\hat{\mu}_{Y|\mathbf{z}})]_1 - [\text{var}(\hat{\mu}_{Y|\mathbf{z}})]_2$, as shown in Figure 4.2. We see clearly from the figure that there is a 60% chance that Scheme 1 will give a lower variance value than Scheme 2 (since there is a 60% chance that the difference is negative). We also see that

despite the higher chance of a lower variance value, the difference in variance tends to be greater when Scheme 1 has a higher variance value. Thus, the CD plot for the difference in variance allows us to determine which of two schemes are better based on the probability of getting a lower variance, and the magnitude of the difference in variance between the two schemes.

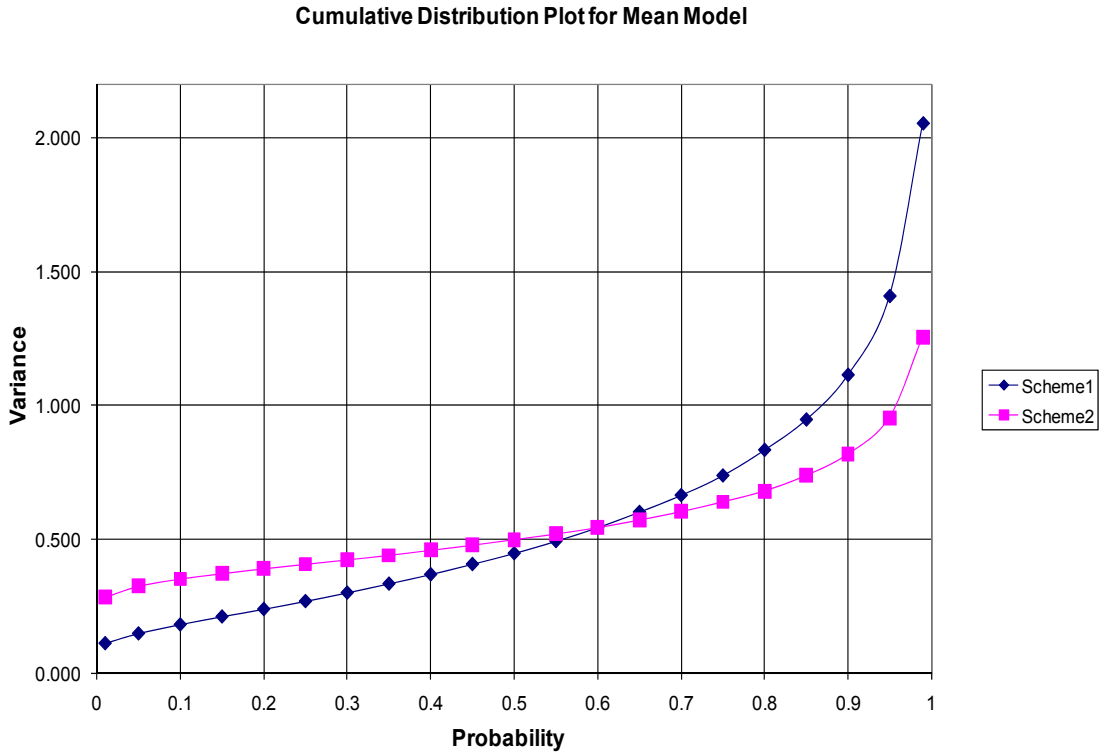**Cumulative Distribution Plot for Mean Model**

Figure 4.1: Example of a Cumulative Distribution Plot

In constructing the CD plots, it would be computationally easier to use explicit expressions for $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ and $\text{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)$. To obtain these expressions, $\mathbf{x}'_C \mathbf{V}_C \mathbf{x}_C$ and $\mathbf{C}$ must be expressed explicitly in terms of the elements of $\mathbf{x}$. This can be done by using software that performs symbolic manipulation.

In the following, we give three examples in which CD plots for the mean and variance models are employed to compare several schemes. Each plot is constructed with $r = 30000$ sampled values. The first example uses data from Example 4.1 and

includes a comparison based on the CD plot for the difference in $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$. The

second example uses data from Example 4.2. In these examples, a point estimate for

the residual variance $\sigma^2 = 1$ is utilized. In the last example, data from Example 3.2 is
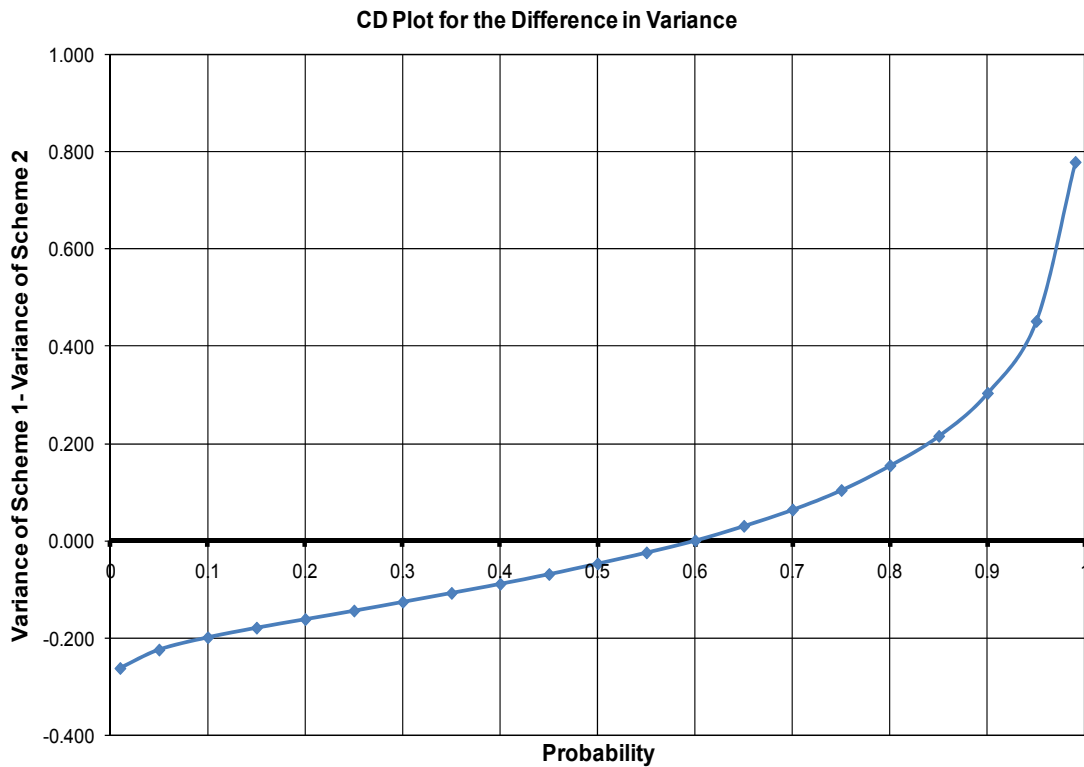
used.

CD Plot for the Difference in Variance

Figure 4.2: CD Plot for the Difference in Variance Values Between Two Schemes

4.5.1 Example 4.3

In this example, we extend Example 4.1 by comparing four different schemes

with CD plots for the mean and variance models. The first scheme consists of sample

sizes $m_1 = m_2 = 35$, and an MRD design determined by $r_f = 1$, $r_a = 3$, and $r_c = 3$.

The second scheme is the solution of Program V and the third scheme is the solution of

Program M. These were given in Example 4.1. The fourth scheme consists of the 24-

run NFS-optimal design given by Castillo et al. (2007) for $k = 3$ and $n = 2$. This design is given in Appendix E. For the fourth scheme, the remaining resource $K - N = 51 - 24 = 27$ is distributed approximately evenly to give $m_1 = 68$ and $m_2 = 67$. Note that for this problem, R is the cube defined in Example 4.1, S is as given in (3.2), and $c_1 = c_2 = 1.5$. The four schemes are summarized in Table 4.1.

Table 4.1: Summary of the Four Schemes for Example 4.3

Scheme	Design	Design Size	m_1	m_2
1	MRD	$r_f = 1, r_a = 3, r_c = 3$	35	35
2	MRD	$r_f = 1, r_a = 1, r_c = 0$	72	73
3	MRD	$r_f = 1, r_a = 2, r_c = 1$	55	55
4	NFS	$N = 24$	68	67

The CD plots for the four schemes given in Table 4.1 are displayed in Figures 4.3 and 4.4. Values for each of the elements of γ and Δ are sampled from a uniform distribution over the interval $[-5,5]$ and values for $\mathbf{x}$ are sampled from a uniform distribution over R.

In Figure 4.3, for any value, say b, of $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$, the corresponding value given by the abscissa axis is the probability that a point $\mathbf{x}$ selected at random from R and each element of γ and Δ drawn from their prior distributions, will yield a value for $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ less than or equal to b. This probability, although of a subjective nature, is a measure of the goodness of a scheme. The CD plot for the variance model can be similarly interpreted.

Figure 4.3: CD Plot for the Mean Model (Example 4.3)
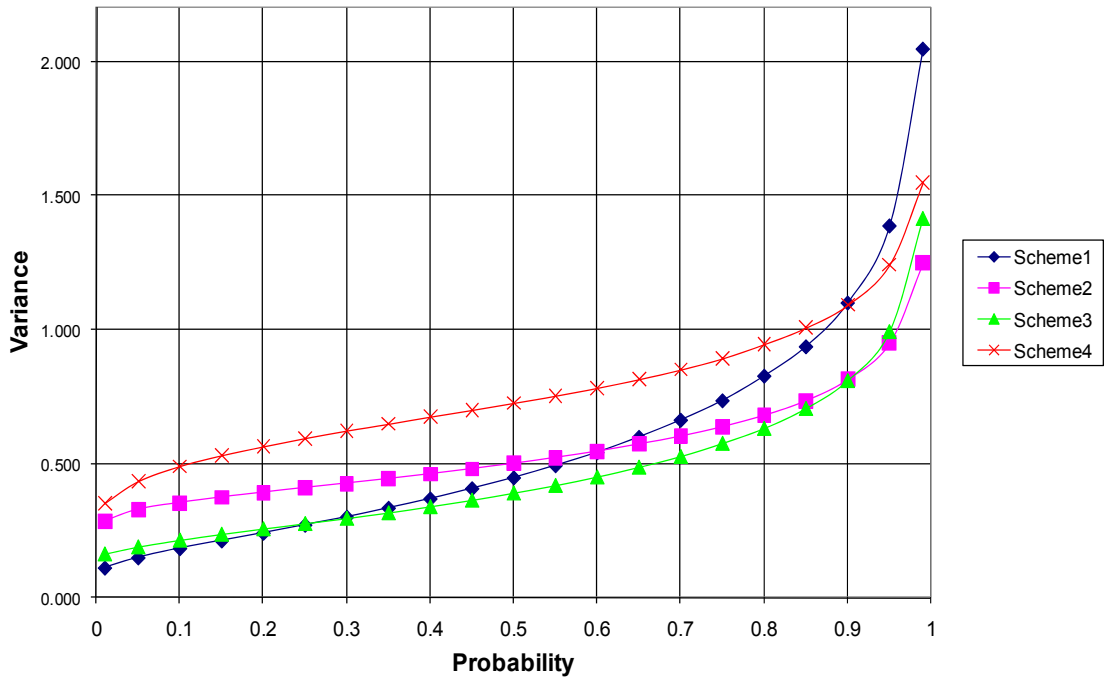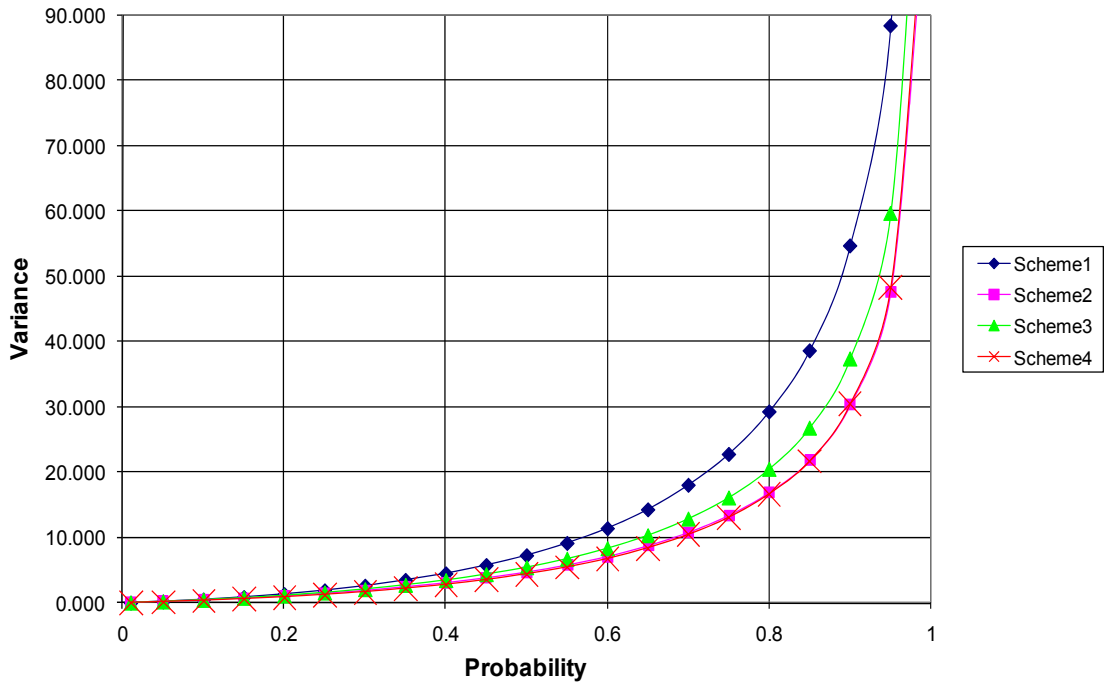

Figure 4.4: CD Plot for the Variance Model (Example 4.3)

Examination of Figure 4.3 reveals that Scheme 4 is a poor candidate for estimating the mean model because the curve for Scheme 4 is higher than the curves for the other three schemes almost everywhere. Although Scheme 1 started with the lowest values of $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$, it rises more steeply than Schemes 2 and 3, eventually rising higher than the graphs for the latter two schemes. If Scheme 1 is used, we have a 90% chance that $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ has a value less than or equal to 1.1. For Schemes 2 and 3, there is a 90% chance that the value is less than or equal to 0.8. Therefore, based on the 90th percentile, Schemes 2 and 3, which are the optimal solutions of Programs V and M respectively, are better candidates for estimating the mean model. Examination of Figure 4.4 reveals that Schemes 2 and 4 perform almost equally well in estimating the variance model, Scheme 3 performs slightly worse than Schemes 2 and 4, whereas Scheme 1 performs badly in estimating the variance model. It appears that all percentiles other than the zero percentile of the probability density of $\text{var}(\hat{\sigma}^2_{Y|\mathbf{z}} - \hat{\sigma}^2)$ for Scheme 1 are larger than the corresponding percentiles for Scheme 3, and the percentiles for Scheme 3 are in turn, larger than the percentiles for Schemes 2 and 4.

A marked feature of the CD plot in Figure 4.4 is that the graphs for each scheme rises sharply to the maximum at the right end. This implies that the maximum variances can be very large. However, based on the discussion in Section 2.6, it is known that large variances tend to occur at points where the variance of the response is a maximum. As such, we should not be too worried about the sharp rise near the right end of each graph.

In Figure 4.5, the CD plot for the difference in $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ for each pair of schemes is plotted. We can see for example, that there is more than a 95% chance that Sheme 2 has a lower variance value than Scheme 4. Table 4.2 summarizes these

probabilities. Each entry in Table 4.2 is the probability that the scheme indicated by the row heading has a lower variance than the scheme indicated by the column heading. The probabilities allow us to rank the schemes in the order 3, 1, 2, 4 in terms of their performance at estimating the mean model.
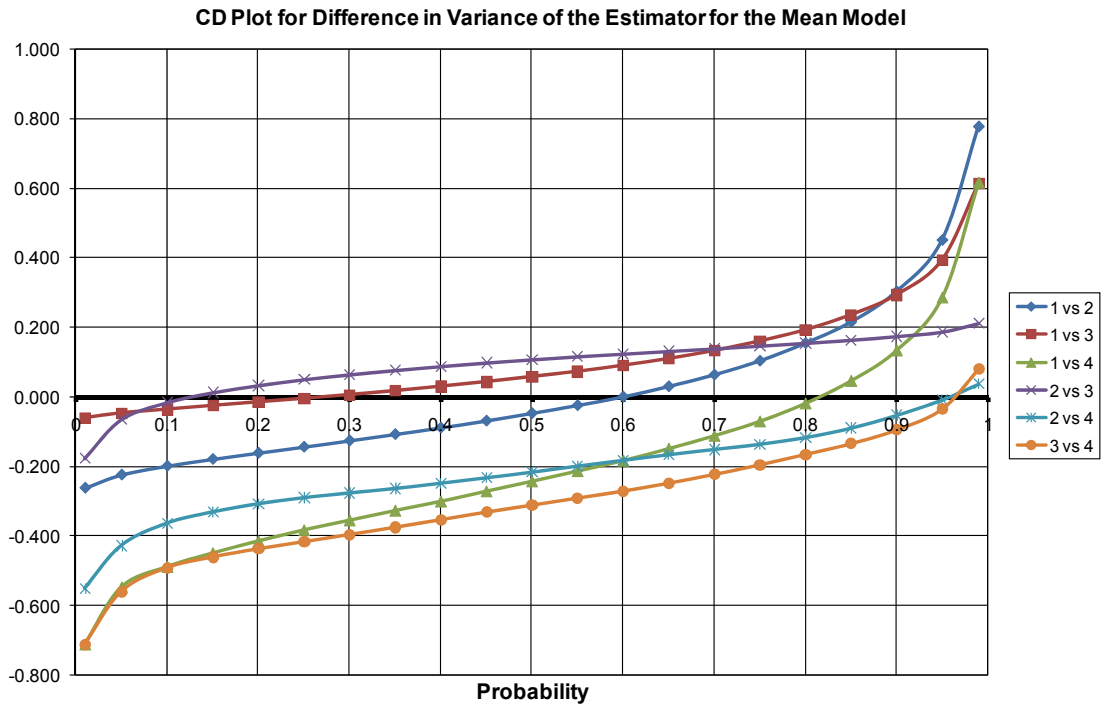
Figure 4.5: CD Plot for Difference in $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ for Each Pair of Schemes

Table 4.2: Probability that Scheme Corresponding to Row has a
Smaller $\text{var}(\hat{\mu}_{Y|\mathbf{z}})$ Than Scheme Corresponding to Column

	Scheme 1	Scheme 2	Scheme 3	Scheme 4
Scheme 1	-	0.6	0.27	0.82
Scheme 2	0.4	-	0.13	0.96
Scheme 3	0.73	0.87	-	0.97
Scheme 4	0.18	0.04	0.03	-

In this example, the main reason for the poor performance of Scheme 1 appears to be the smaller sample sizes for that scheme while the main reason for the poor performance of Scheme 4 in estimating the mean model is the inherent weaknesses of

107

the design. For the design in Scheme 4, the value of IM_E / σ^2, which is the average of

$\mathbf{x}'_C \mathbf{V}_C \mathbf{x}_C$ (see Equation (3.5)), is 0.5475. In contrast, the value of IM_E / σ^2 for the

design in Scheme 2, which has 2 runs less than the design in Scheme 4, is 0.3472. In

fact, it is estimated by simulation that the set of points $\mathbf{x}$ at which Scheme 4 has a

smaller value of $\mathbf{x}'_C \mathbf{V}_C \mathbf{x}_C$ than Scheme 2 occupies a volume of only about 5.8% of

the volume of R. This is not surprising as the NFS criterion is closely linked to the

estimation of the variance model but is not linked to the estimation of the mean model

(see Castillo et al. (2007)). In summary, this example demonstrates that the

performance of a scheme depends as much on the proper choice of sample sizes as on

the design.

4.5.2 Example 4.4

In this example, we extend Example 4.2 by comparing four different schemes

with CD plots for the mean and variance models. The first scheme consists of a D-

Optimal design with 45 runs, constructed by MINITAB using the 3^5 factorial as the

candidate set of points. The sequential optimization option for constructing the initial

design and Fedorov's method for improving the initial design are the chosen options

for constructing the design. Given that the total cost of the scheme must be 100, the

remaining resource of 55 is divided approximately equally to give $m_1 = 138$ and

$m_2 = 137$. The second scheme is the solution of Program V whereas the third scheme

is the solution of Program M. Both schemes were given in Example 4.2. The design in

the fourth scheme is a 25-run D-Optimal design generated by the same method as with

the D-Optimal design in the first scheme. For the fourth scheme, the remaining 75

units of resource are divided approximately equally to give $m_1 = 188$ and $m_2 = 187$.

The designs for the first and fourth schemes are presented in Appendix E. For this

problem, R is the cube defined in Example 4.1, S is as given in (3.2), and

$c_1 = c_2 = 1.5$. A summary of the four schemes is given in Table 4.3.

Table 4.3: Summary of the Four Schemes for Example 4.4

Scheme	Design	Design Size	m_1	m_2
1	D-Optimal	$N = 45$	138	137
2	MRD	$r_f = 2, r_a = 1, r_c = 0$	155	155
3	MRD	$r_f = 1, r_a = 5, r_c = 1$	133	132
4	D-Optimal	$N = 25$	188	187

The CD plots for the four schemes given in Table 4.3 are presented in Figures

4.6 to 4.9. Values for each element of γ and Δ are sampled from a normal prior

density with mean 0 and variance $\sigma_P^2 = 9$. We present 3 CD plots for the variance

model because the graphs for Schemes 1, 2 and 4 are nearly identical so that they

would be difficult to distinguish in a single figure. The figures show that Scheme 3 is

excellent for estimating the mean model, but is poor for estimating the variance model.

It seems that all percentiles other than the zero percentile of the probability density of

$\text{var}(\hat{\sigma}_{Y|z}^2 - \hat{\sigma}^2)$ for Scheme 3 are larger than the corresponding percentiles for the other

three schemes. Although Schemes 1, 2, and 4 perform almost equally well in

estimating the variance model, Scheme 1 performs better than Schemes 2 and 4 in

estimating the mean model. Therefore, if interest lies in estimating both mean and

variance models, Scheme 1, which comprises the 45-run D-optimal design, is a good

candidate. This example demonstrates that D-optimal designs can be better than MRD

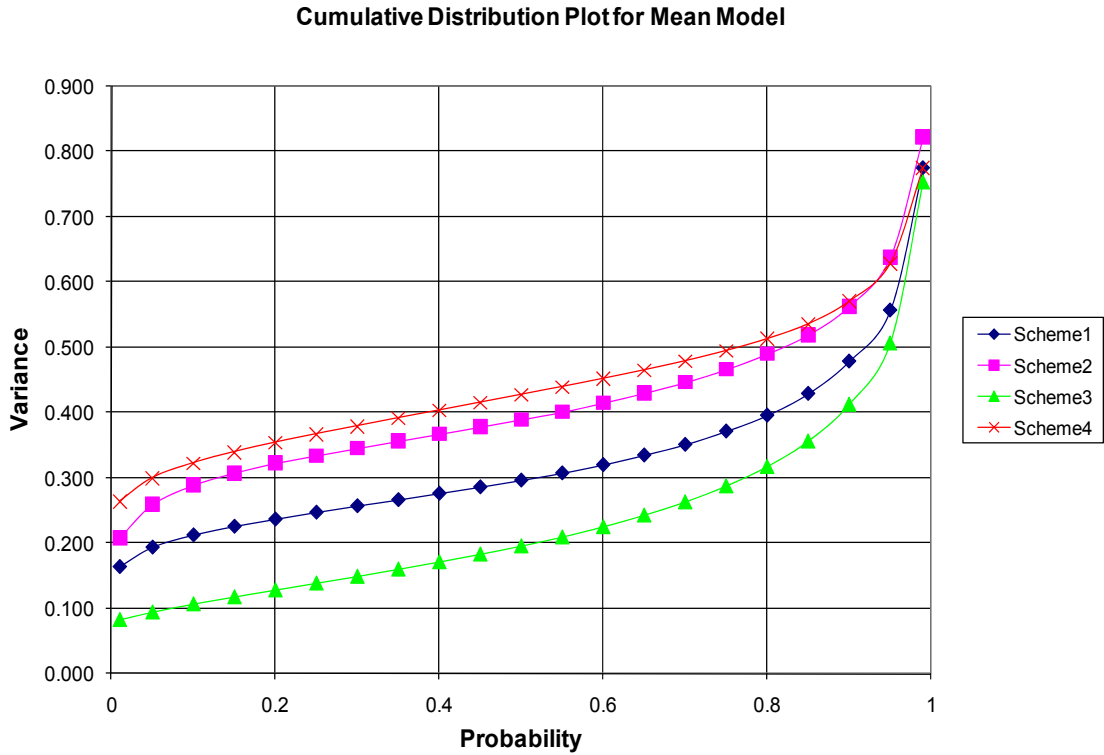designs and so, should be seriously considered for any given problem.

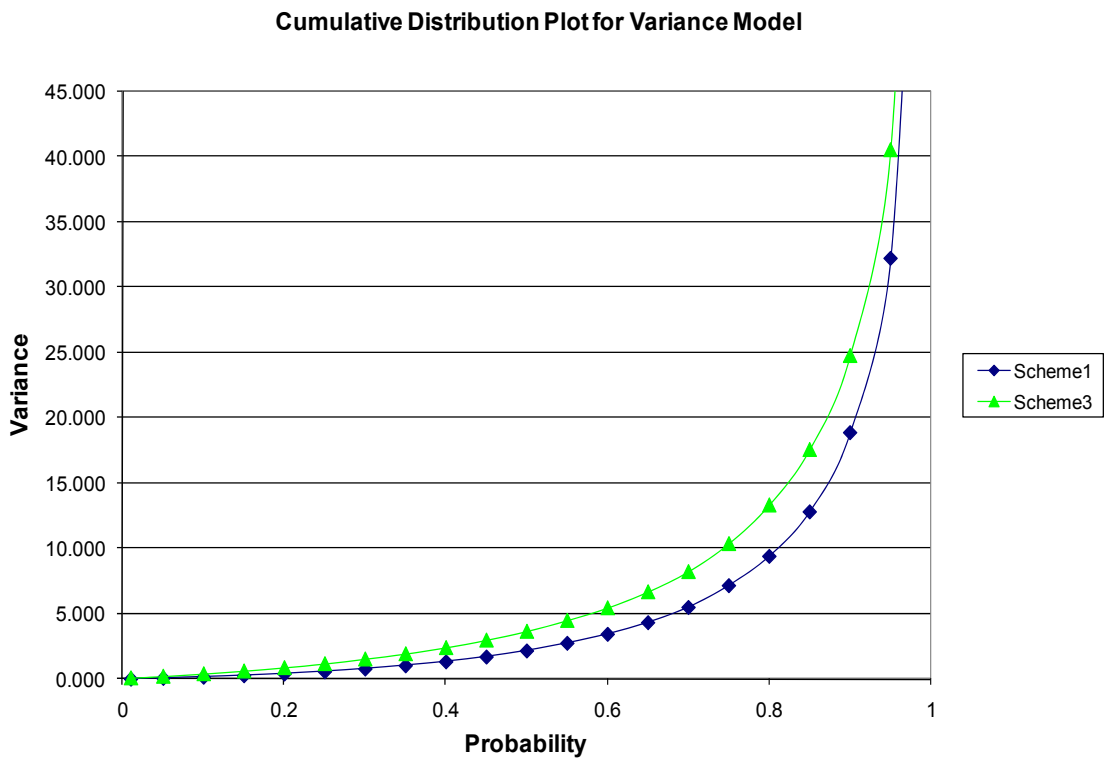Figure 4.6: CD Plot for the Mean Model (Example 4.4)

Figure 4.7: CD Plot for the Variance Model: Schemes 1 and 3 (Example 4.4)
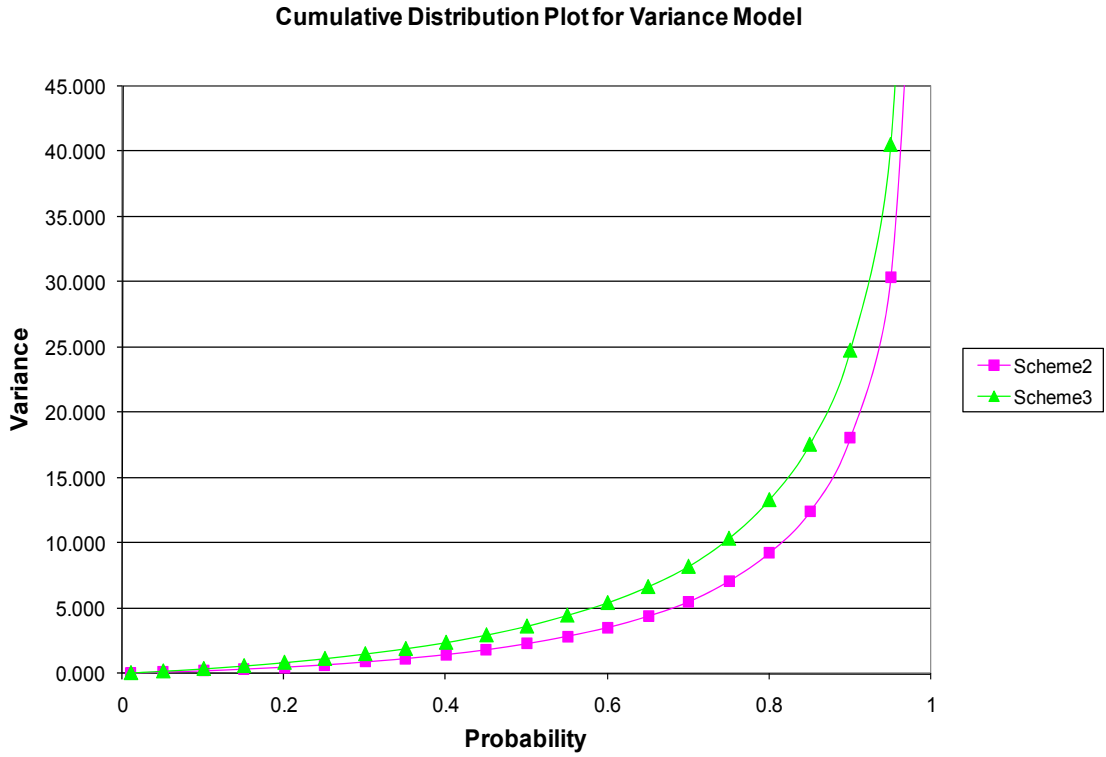
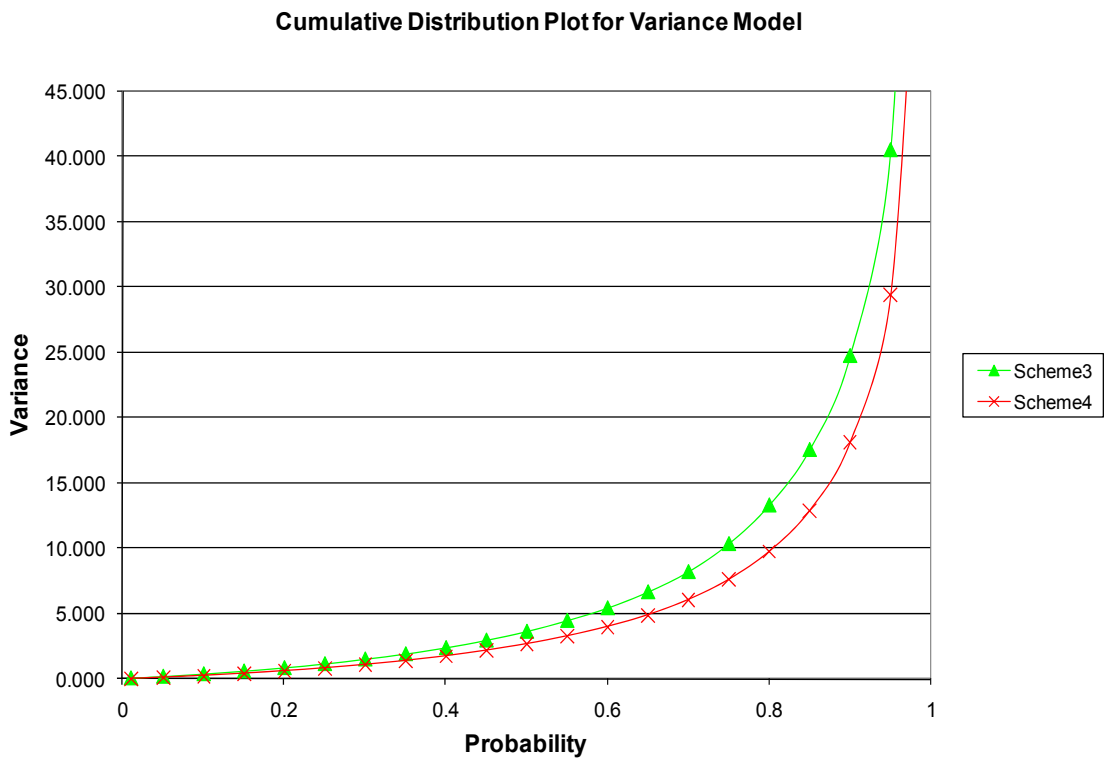Figure 4.8: CD Plot for the Variance Model: Schemes 2 and 3 (Example 4.4)

Figure 4.9: CD Plot for the Variance Model: Schemes 3 and 4 (Example 4.4)

111

4.5.3 Example 4.5

In this example, we extend Example 3.2 by comparing three different schemes chosen from the Pareto optimal solutions in Table 3.3 with the CD plots. In this case, point estimates for γ, Δ, and σ^2 are used in constructing the CD plots. Thus, an interpretation of the CD plots is that given a point on the graph for a scheme with a value b on the ordinate axis, the corresponding value on the abscissa gives the fraction of volume of the design space with a variance at or below b (Zahran et al., 2003). The first scheme to be studied in this example is the solution labeled S3 in Table 3.3. The second scheme is the optimal solution of Program V whereas the third scheme is the optimal solution of Program M. The three schemes are summarized in Table 4.4.

Table 4.4: Summary of the Three Schemes for Example 4.5

Scheme	Design	Design Size	m_1	m_2
1	MRD	$r_f = 2, r_a = 3, r_c = 13$	79	93
2	MRD	$r_f = 3, r_a = 1, r_c = 0$	91	101
3	MRD	$r_f = 1, r_a = 6, r_c = 16$	81	95

The CD plots for the three schemes given in Table 4.4 are displayed in Figures 4.10 and 4.11. They show that Scheme 2, despite being the best scheme for estimating the variance model, performs very badly at estimating the mean model. Scheme 3, which is optimal for estimating the mean model, is undesirable for estimating the variance model. Lastly, Scheme 1 is almost as good as Scheme 3 for estimating the mean model while it is second best for estimating the variance model. If interest is in estimating both the mean and variance models, Scheme 1 is a good choice. This example demonstrates the potential usefulness of Pareto optimal solutions.

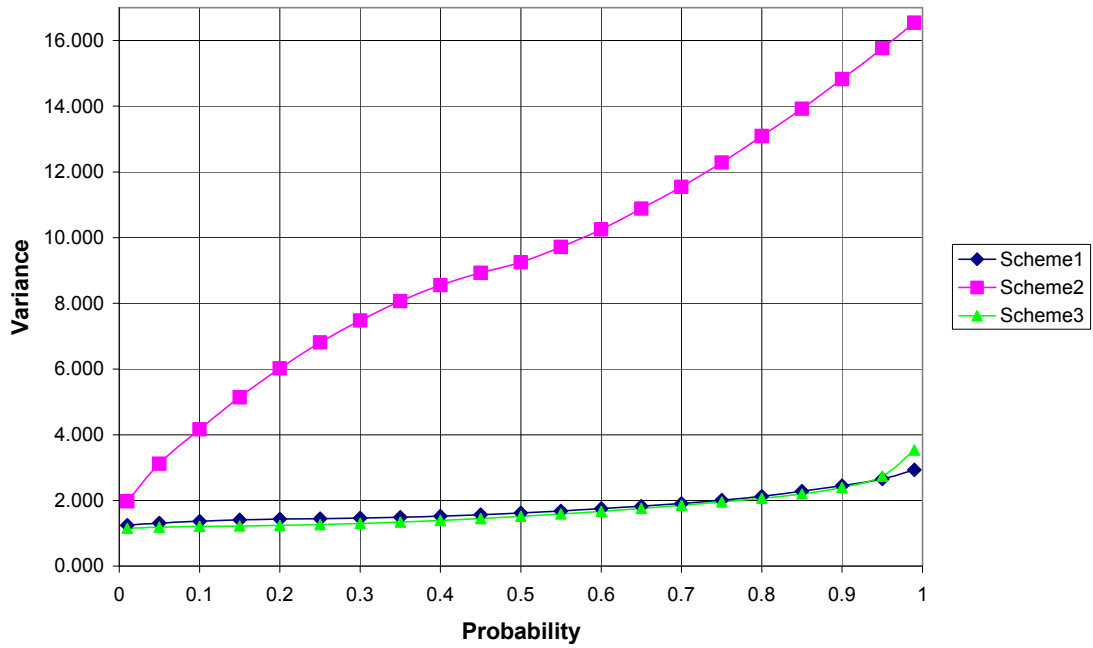**Cumulative Distribution Plot for Mean Model**

Figure 4.10: CD Plot for the Mean Model (Example 4.5)

Cumulative Distribution Plot for Variance Model
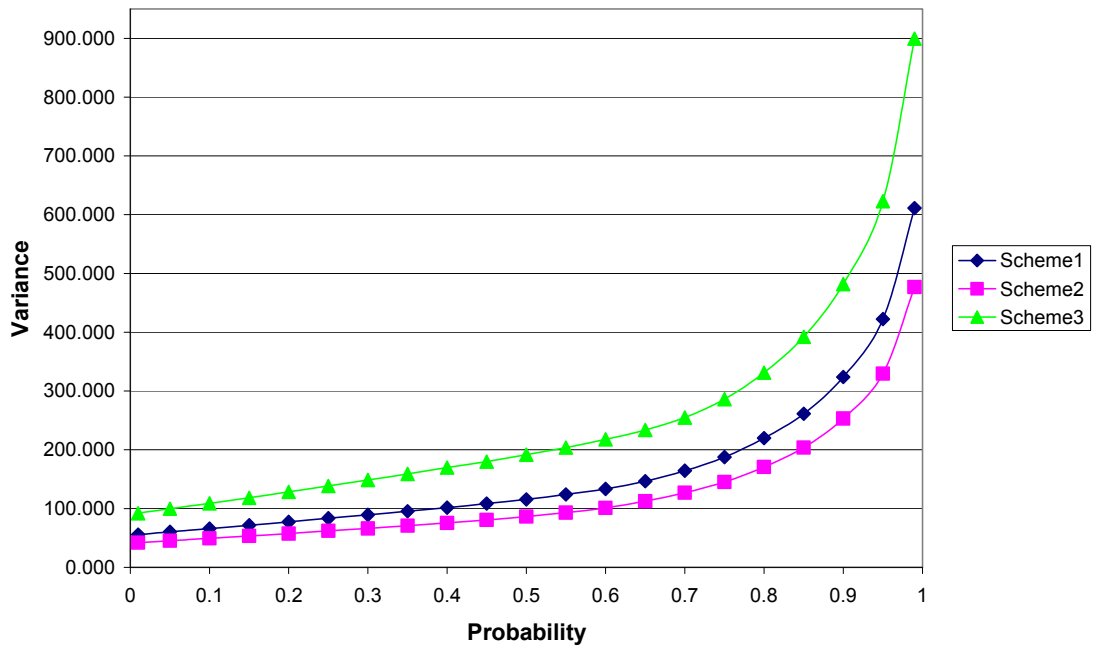
Figure 4.11: CD Plot for the Variance Model (Example 4.5)

CHAPTER 5

CONCLUSIONS AND FURTHER RESEARCH

The main contribution of this work is to propose an approach for estimating the mean and variance models with a combined array experiment for the case where the means and variances of the noise variables are unknown. In the approach, planning of estimation of the means and variances of the noise variables with data sampled from the process is integrated with planning of the combined array experiment. This takes into consideration the fact that in practice, the means and covariances of the noise variables are estimated with process data because they are unknown. Thus, the proposed approach extends the dual response surface approach presented by Myers et al. (1992), and Myers and Montgomery (2002), which assumes the means and covariances of the noise variables are known.

Novel ideas introduced with the proposed procedure are expounded in this thesis. These include specification of the levels of the noise variables, estimation of the mean and variance models, and optimal allocation of resource to sampling and experimenting. We propose a method to determine the appropriate scaling factors and design region so that the noise variables are varied over ranges that are representative of their variation during actual process operation or product use but are not varied over unnecessarily wide ranges.

The consequences of errors in estimating the means and variances of the noise variables on the estimation of the mean and variance models have previously been a

subject that is ignored in the literature. We examine the estimators for the mean and variance models given in the literature in light of sampling and experiment error. Expressions for the bias and variance of the estimators are derived.

Within the framework of our proposed procedure, the problem of allocating experiment effort between sampling and experimenting is of practical interest. This thesis shows how mathematical programs can be used to find sample sizes and MRD designs that optimize estimation of the mean model or that optimize estimation of the variance model. We also show how sample sizes and MRD designs that compromise between the estimation of both models can be found. A greedy algorithm is proposed to find schemes that perform well in estimating either the mean model, the variance model, or both models for the case where the design is to be constructed from a candidate set of points. In addition, cumulative distribution plots are proposed for evaluating schemes that may consist of designs other than the MRD.

The optimal allocation of effort depends on unknown parameters of the response model. Although prior knowledge can be captured in the form of point estimates or a prior distribution, this approach may yield estimates that are far from the true values or a prior density that places little weight on the true values. In addition, the two-stage procedure discussed in Section 4.2.4 may be suboptimal with respect to allocation of total resource because the stages are planned separately. A sequential procedure in which its various stages are considered in an integrated way so that allocation of the total resource is optimized is an interesting extension.

Relaxing the assumption of random sampling, normally and independently distributed noise variables, and generalizing the results in this thesis to cases in which the response model is of a form different from that given in (2.3) will be useful. The robustness of the variance formulas derived in this thesis and the performance of

schemes that minimize the average variances to violations of assumptions are also subjects for further research. Of special interest is robustness to model misspecification because the validity of the mean and variance models and the variance formulas for the estimators of those models depends on the assumption that the response model holds exactly.

Finally, application of the methodology developed in this thesis to real problems may lead to modifications that improve the applicability of the methodology.

REFERENCES

1. Abraham, B. and J. MacKay. (1993). "Variation Reduction and Designed Experiments," International Statistical Review, Vol.61, No.1, Special Issue on Statistics in Industry, pp.121-129.

2. Arnold, S.F. (1981). The Theory of Linear Models and Multivariate Analysis. John Wiley & Sons, New York.

3. Arvidsson, M. and I. Gremyr. (2007). "Principles of Robust Design Methodology," Quality and Reliability Engineering International, (in press).

4. Atkinson, A.C. (1982). "Developments in the Design of Experiments," International Statistical Review, Vol.50, No.2, pp.161-177.

5. Atkinson, A.C. (1996). "The Usefulness of Optimum Experimental Designs," Journal of the Royal Statistical Society. Series B (Methodological), Vol.58, No.1, pp.59-76.

6. Atkinson, A.C. and R.A. Bailey. (2001). "One Hundred Years of the Design of Experiments on and off the Pages of Biometrika," Biometrika, Vol.88, No.1, pp.53-97.

7. Atkinson, A.C., C.G.B. Demetrio, and S.S. Zocchi. (1995). "Optimum Dose Levels when Males and Females Differ in Response," Applied Statistics, Vol.44, No.2, pp.213-226.

8. Atkinson, A.C. and A.N. Donev. (1992). Optimum Experimental Designs. Oxford University Press, Oxford.

9. Bazaraa, M.S., H.D. Sherali, and C.M. Shetty. (1993). Nonlinear Programming: Theory and Algorithms. 2nd Edition. John Wiley & Sons, New York.

10. Ben-Tal, A. and A. Nemirovski. (1998). "Robust Convex Optimization," Mathematics of Operations Research, Vol.23, No.4, pp.76-805.

11. Bisgaard, S. (1996). "A Comparative Analysis of the Performance of Taguchi's Linear Graphs for the Design of Two-Level Fractional Factorials." Applied Statistics, Vol.45, No.3, pp.311-322.

12. Borkowski, J.J. and J.M. Lucas. (1997). "Designs of Mixed Resolution for Process Robustness Studies," Technometrics, Vol.39, No.1, pp.63-70.

13. Borror, C.M. and D.C. Montgomery. (2000). "Mixed Resolution Designs as Alternatives to Taguchi Inner/Outer Array Designs for Robust Design Problems," Quality and Reliability Engineering International, Vol.36, pp.117-127.

14. Borror, C.M., D.C. Montgomery, and R.H. Myers. (2002). "Evaluation of Statistical Design for Experiments Involving Noise Variables," Journal of Quality Technology, Vol.34, No.1, pp. 54-70.

15. Box, G.E.P. (1953). "Non-Normality and Tests on Variances," Biometrika, Vol.40, No.3/4, pp.318-335.

16. Box, G.E.P. (1988). "Signal to Noise Ratios, Performance Criteria and Transformations," Technometrics, Vol.30, No.1, pp.1-17.

17. Box, G.E.P. (1993). "Sequential Experimentation and Sequential Assembly of Designs," Quality Engineering, Vol.5, No.2, pp.321-330.

18. Box, G.E.P. and N.R. Draper. (1987). Empirical Model-Building and Response Surfaces. John Wiley & Sons, New York.

19. Box, G.E.P., J.S. Hunter, and W.G. Hunter. (1978). Statistics for Experimenters. John Wiley & Sons, New York.

20. Box, G.E.P., J.S. Hunter, and W.G. Hunter. (2005). Statistics for Experimenters. 2nd Edition. John Wiley & Sons, New York.

21. Box, G.E.P. and S. Jones. (1992). "Split-plot Designs for Robust Product Experimentation," Journal of Applied Statistics, Vol.19, No.1, pp.3-26.

22. Brenneman, W.A. and W.R. Myers. (2003). "Robust Parameter Design with Categorical Noise Variables," Journal of Quality Technology, Vol.35, No.4, pp.335-341.

23. Castillo, E.D., M.J. Alvarez, L. Ilzarbe, and E. Viles. (2007). "A New Design Criterion for Robust Parameter Experiments," Journal of Quality Technology, Vol.39, No.3, pp.279-295.

24. Chaloner, K., T. Church, T.A. Louis, and J.P. Matts (1993). "Graphical Elicitation of a Prior Distribution for a Clinical Trial," The Statistician, Vol.42, No.4, pp.341-353.

25. Chaloner, K. and Verdinelli, I. (1995). "Bayesian Experimental Design: A Review," Statistical Science, Vol.10, No.3, pp.273-304.

26. Chaudhuri, P. and P.A. Mykland. (1993). "Nonlinear Experiments: Optimal Design and Inference Based on Likelihood," Journal of the American Statistical Association, Vol.88, No.422, pp.538-546.

27. Chew, V. (1966). "Confidence, Prediction, and Tolerance Regions for the Multivariate Normal Distribution," Journal of the American Statistical Association, Vol.61, No.315, pp.605-617.

28. Cochran, W.G. (1973). "Experiments for Nonlinear Functions," Journal of the American Statistical Association, Vol.68, No.344, pp.771-781.

29. Dasgupta, T. (2007). Robust Parameter Design for Automatically Controlled Systems and Nanostructure Synthesis. Ph.D Dissertation. School of Industrial and Systems Engineering, Georgia Institute of Technology.

30. Donev, A.N., and A.C. Atkinson. (1988). "An Adjustment Algorithm for the Construction of Exact D-Optimum Experimental Designs," Technometrics, Vol.30, No.4, pp.429-433.

31. Edmonson, N. (1930). "Poisson's Integral and Plurisegments on the Hypersphere," The Annals of Mathematics, Vol.31, No.1, pp.13-31.

32. Fisher, R.A. (1925). "Theory of Statistical Estimation," Proceedings of the Cambridge Philosophical Society, Vol.22, pp.700-725.

33. Ford, I., D.M. Titterington, and C.P. Kitsos. (1989). "Recent Advances in Nonlinear Experimental Design," Technometrics, Vol.31, No.1, pp.49-60.

34. Ford, I., D.M. Titterington, and C.F.J. Wu. (1985). "Inference and Sequential Design," Biometrika, Vol.72, No.3, pp.545-551.

35. Garthwaite, P.H. and J.M. Dickey. (1988). "Quantifying Expert Opinion in Linear Regression Problems," Journal of the Royal Statistical Society. Series B (Methodological), Vol.50, No.3, pp.462-474.

36. Ginsburg, H. and I. Ben-Gal (2006). "Designing Experiments for Robust-Optimization Problems: the V_S -Optimality Criterion," IIE Transactions, Vol.38, pp.445-461.

37. Groves, T. and T. Rothenberg. (1969). "A Note on the Expected Value of an Inverse Matrix," Biometrika, Vol.56, No.3, pp.690-691.

38. Gupta, O.K., and A. Ravindran. (1985). "Branch and Bound Experiments in Convex Nonlinear Integer Programming," Management Science, Vol.31, No.12, pp.1533-1546.

39. Harville, D.A. (1997). Matrix Algebra from a Statistician's Perspective. Springer-Varlag, New-York.

40. Haven, K., A. Majda, and R. Abramov. (2005). "Quantifying Predictability Through Information Theory: Small Sample Estimation in a Non-Gaussian Framework," Journal of Computational Physics, Vol. 206, pp. 334-362.

41. Herzberg, A.M. and D.R. Cox. (1969). "Recent Work on the Design of Experiments: A Bibliography and a Review," Journal of the Royal Statistical Society, Series A (General), Vol.132, No.1, pp.29-67.

42. Hoffman, K. and R. Kunze. (2002). Linear Algebra. 2nd Edition. Prentice Hall of India, New Delhi.

43. Jeang, A., F. Liang, and C.P. Chung. (2007). "Robust Product Development for Multiple Quality Characteristics Using Computer Experiments and an Optimization Technique," International Journal of Production Research, pp.1-25.

44. Jin, J. and Y. Ding. (2004). "Online Automatic Process Control Using Observable Noise Factors for Discrete-Part Manufacturing," IIE Transactions, Vol. 36, pp.899-911.

45. Khuri, A.I. (2002). Advanced Calculus with Applications in Statistics. 2nd Edition. John Wiley & Sons, New York.

46. Khuri, A.I. and J.A. Cornell. (1996). Response Surfaces: Design and Analyses. 2nd Edition. Marcel Dekker, New York.

47. Kiefer, J.C. (1987). Introduction to Statistical Inference. Edited by G. Lorden. Springer-Verlag, New York.

48. Koksoy, O. and N. Doganaksoy. (2003). "Joint Optimization of Mean and Standard Deviation Using Response Surface Methods," Journal of Quality Technology, Vol.35, No.3, pp. 239-252.

49. Kunert, J., C. Auer, M. Erdbrugge, and R. Ewers. (2007). "An Experiment to Compare Taguchi's Product Array and the Combined Array," Journal of Quality Technology, Vol.39, No.1, pp.17-34.

50. Lawson, J.S. and J.L. Madrigal. (1994). "Robust Design Through Optimization Techniques," Quality Engineering, Vol.6, No.4, pp.593-608.

51. Leon, R.V., A.C. Shoemaker, and R.N. Kacker. (1987). "Performance Measures Independent of Adjustment: An Explanation and Extension of Taguchi's Signal-to-Noise Ratios," Technometrics, Vol.29, No.3, pp.253-265.

52. Leon, R.V., A.C. Shoemaker, and K.L. Tsui. (1993). "A Systematic Approach to Planning for a Designed Industrial Experiment: Discussion," Technometrics, Vol.35, No.1, pp.21-24.

53. Li, D. and X. Sun. (2006). Nonlinear Integer Programming. Springer-Verlag, New York.

54. Li, J., C. Zhang, R. Liang, and B. Wang. (2007). "Robust Design of Composite Manufacturing Processes with Process Simulation and Optimisation Methods," International Journal of Production Research, pp.1-18.

55. Lucas, J.M. (1974). "Optimum Composite Designs," Technometrics, Vol.16, No.4, pp.561-567.

56. Miro-Quesada, G. and E.D. Castillo. (2004). "Two Approaches for Improving the Dual Response Method in Robust Parameter Design," Journal of Quality Technology, Vol.36, No.2, pp.154-168.

57. Miro-Quesada, G., E.D. Castillo, and J.J. Peterson. (2004). "A Bayesian Approach for Multiple Response Surface Optimization in the Presence of Noise Variables," Journal of Applied Statistics, Vol.31, No.3, pp.251-270.

58. Montgomery, D.C. (1999). "Experimental Design for Product and Process Design and Development," The Statistician, Vol.48, Part 2, pp.159-177.

59. Montgomery, D.C. (2005a). Introduction to Statistical Quality Control. 5th Edition. John Wiley & Sons, New York.

60. Montgomery, D.C. (2005b). Design and Analysis of Experiments. 6th Edition. John Wiley & Sons, New York.

61. Myers, R.H. (1991). "Response Surface Methodology in Quality Improvement," Communications in Statistics-Theory and Methods, Vol.20, No.2, pp.457-476.

62. Myers, R.H, A.I Khuri, and G. Vining (1992). "Response Surface Alternatives to the Taguchi Robust Parameter Design Approach," The American Statistician, Vol.46, No.2, pp.131-139.

63. Myers, R.H., Y. Kim, and K.L. Griffiths. (1997). "Response Surface Methods and the Use of Noise Variables," Journal of Quality Technology, Vol.29, No.4, pp.429-440.

64. Myers, R.H. and D.C. Montgomery. (2002). Response Surface Methodology: Process and Product Optimization Using Design of Experiments. John Wiley & Sons, New York.

65. Myers, R.H., D.C. Montgomery, G.G. Vining, C.M. Borror, and S.M. Kowalski. (2004). "Response Surface Methodology: A Retrospective and Literature Survey," Journal of Quality Technology, Vol.36, No.1, pp.53-77.

66. O'Donnell, E.M. and G.G. Vining. (1997). "Mean Squared Error of Prediction Approach to the Analysis of a Combined Array," Journal of Applied Statistics, Vol.24, No.6, pp.733-746.

67. O'Neill, J.C., C.M. Borror, P.Y. Eastman, D.G. Fradkin, M.P. James, A.P. Marks, and D.C. Montgomery. (2000). "Optimal Assignment of Samples to Treatments for Robust Design," Quality and Reliability Engineering International, Vol.16, pp.417-421.

68. Ozol-Godfrey, A., C.M. Anderson-Cook, and D.C. Montgomery. (2005). "Fraction of Design Space Plots for Examining Model Robustness," Journal of Quality Technology, Vol.37, No.3, pp.223-235.

69. Park, Y., D.C. Montgomery, J.W. Fowler, and C.M. Borror. (2005). "Cost-Constrained G-efficient Response Surface Designs for Cuboidal Regions," Quality and Reliability Engineering International, Vol.22, pp.121-139.

70. Press, S.J. (2003). Subjective and Objective Bayesian Statistics. 2nd Edition. John Wiley & Sons, New York.

71. Pronzato, L. and E. Walter (1985). "Robust Experiment Design via Stochastic Approximation," Mathematical Biosciences, Vol.75, pp.103-120.

72. Radson, D. and G.D. Herrin. (1995). "Augmenting a Factorial Experiment When One Factor is an Uncontrollable Random Variable: A Case Study," Technometrics, Vol.37, No.1, pp.70-81.

73. Rahman, M. and M. Ahsanullah. (1973). "A Note on the Expected Value of Powers of a Matrix," The Canadian Journal of Statistics, Vol.1, No.1, pp.123-125.

74. Robinson, T.J., C.M. Borror, and R.H. Myers. (2004). "Robust Parameter Design: A Review," Quality and Reliability Engineering International, Vol.20, pp.81-101.

75. Rockafellar, R.T. (2007). Fundamentals of Optimization. Lecture Notes. Department of Mathematics, University of Washington, Seattle.

76. Romano, D., M.Varetto, and G. Vicario. (2004). "Multiresponse Robust Design: A General Framework Based on Combined Array," Journal of Quality Technology, Vol.36, No.1, pp.27-37.

77. Searle, S.R. (1982). Matrix Algebra Useful for Statistics. John Wiley & Sons.

78. Sherali, H.D. and D.C. Myers. (1985). "The Design of Branch and Bound Algorithms for a Class of Nonlinear Integer Programs," Annals of Operations Research, Vol.5, pp.463-484.

79. Shoemaker, A.C., K.L. Tsui, and C.F.J. Wu (1991). "Economical Experimentation Methods for Robust Design," Technometrics, Vol.33, No.4, pp.415-427.

80. Shore, H. and R. Arad. (2003). "Product Robust Design and Process Robust Design: Are They the Same? (No.)," Quality Engineering, Vol.16, No.2, pp.193-207.

81. Silvey, S.D. (1980). Optimal Design: An Introduction to the Theory for Parameter Estimation. Chapman and Hall, London.

82. Sitter, R.R. and C.F.J. Wu. (1999). "Two-Stage Design of Quantal Response Studies," Biometrics, Vol.55, No.2, pp.396-402.

83. Steinberg, D.M. and D. Bursztyn. (1998). "Noise variables, Dispersion Effects, and Robust Design," Statistica Sinica, Vol.8, pp. 67-85.

84. Steinberg, D.M. and W.G. Hunter. (1984). "Experimental Design: Review and Comment," Technometrics, Vol.26, No.2, pp.71-97.

85. Taguchi, G., Y. Yokoyama, and Y. Wu. (1993). Taguchi Methods: Design of Experiments. Japanese Standards Association, Tokyo.

86. Voinov, V.G. and M.S. Nikulin. (1993). Unbiased Estimators and Their Applications: Volume 1: Univariate Case. Kluwer Academic Publishers, Dordrecht.

87. Welch, W.J. (1982). "Branch-and-Bound Search for Experimental Designs Based on D Optimality and Other Criteria," Technometrics, Vol.24, No.1, pp.41-48.

88. Wu, C.F.J. and M. Hamada. (2000). Experiments: Planning, Analysis, and Parameter Design Optimization. John Wiley & Sons, New York.

89. Xu, D. and S.L. Albin. (2003). "Robust Optimization of Experimentally Derived Objective Functions," IIE Transactions, Vol.35, pp.793-802.

90. Zahran, A., C.M. Anderson-Cook, and R.H. Myers. (2003). "Fraction of Design Space to Assess Prediction Capability of Response Surface Designs," Journal of Quality Technology, Vol.35, No.4, pp.377-386.

APPENDIX A

Proof of Proposition 2.6

Proposition 2.6 If $\hat{\boldsymbol{\Sigma}}$ is unbiased for $\boldsymbol{\Sigma}$, $\hat{\sigma}^2_{Y|\mathbf{z}}$ has a smaller mean square error than

$\hat{\sigma}^2_{YB|\mathbf{z}}$ for every $\mathbf{x}$ when $dfSSE \geq 2$.

Proof:

The expectation of $\hat{\sigma}^2_{YB|\mathbf{z}}$ with respect to $\mathbf{s}$ and $\mathbf{e}$ is given by

$$
\begin{aligned}
\underset{\mathbf{s,e}}{E}(\hat{\sigma}^2_{YB|\mathbf{z}}) &= \underset{\mathbf{s}}{E}\left[\underset{\mathbf{e}}{E}(\hat{\boldsymbol{\gamma}}_\mathbf{z} + \hat{\boldsymbol{\Delta}}'_\mathbf{z}\,\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_\mathbf{z} + \hat{\boldsymbol{\Delta}}'_\mathbf{z}\,\mathbf{x}) + \hat{\sigma}^2 \Big| \mathbf{s} \right] \\
&= \underset{\mathbf{s}}{E}\{(\boldsymbol{\gamma}_\mathbf{z} + \boldsymbol{\Delta}'_\mathbf{z}\,\mathbf{x})'\mathbf{V}(\boldsymbol{\gamma}_\mathbf{z} + \boldsymbol{\Delta}'_\mathbf{z}\,\mathbf{x}) + \sigma^2[1 + trace(\mathbf{VC})]\} \\
&= (\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x})'\mathbf{V}(\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x}) + \sigma^2[1 + trace(\mathbf{VC})].
\end{aligned}
$$

Using the law of conditional variance and the fact that the residual mean square $\hat{\sigma}^2$ is independent of the least squares estimators $\hat{\boldsymbol{\gamma}}_\mathbf{z}$ and $\hat{\boldsymbol{\Delta}}_\mathbf{z}$ when $\mathbf{s}$ is held fixed, the variance of $\hat{\sigma}^2_{YB|\mathbf{z}}$ with respect to $\mathbf{s}$ and $\mathbf{e}$ is given by

$$
\begin{aligned}
var(\hat{\sigma}^2_{YB|\mathbf{z}}) &= \underset{\mathbf{s,e}}{var}\{(\hat{\boldsymbol{\gamma}}_\mathbf{z} + \hat{\boldsymbol{\Delta}}'_\mathbf{z}\,\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_\mathbf{z} + \hat{\boldsymbol{\Delta}}'_\mathbf{z}\,\mathbf{x}) + \hat{\sigma}^2]\} \\
&= \underset{\mathbf{s}}{var}\{\underset{\mathbf{e}}{E}[(\hat{\boldsymbol{\gamma}}_\mathbf{z} + \hat{\boldsymbol{\Delta}}'_\mathbf{z}\,\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_\mathbf{z} + \hat{\boldsymbol{\Delta}}'_\mathbf{z}\,\mathbf{x}) + \hat{\sigma}^2]\Big|\mathbf{s}\} + \underset{\mathbf{s}}{E}\{\underset{\mathbf{e}}{var}[(\hat{\boldsymbol{\gamma}}_\mathbf{z} + \hat{\boldsymbol{\Delta}}'_\mathbf{z}\,\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_\mathbf{z} + \hat{\boldsymbol{\Delta}}'_\mathbf{z}\,\mathbf{x}) + \hat{\sigma}^2]\Big|\mathbf{s}\} \\
&= \underset{\mathbf{s}}{var}\{(\boldsymbol{\gamma}_\mathbf{z} + \boldsymbol{\Delta}'_\mathbf{z}\,\mathbf{x})'\mathbf{V}(\boldsymbol{\gamma}_\mathbf{z} + \boldsymbol{\Delta}'_\mathbf{z}\,\mathbf{x}) + \sigma^2[1 + trace(\mathbf{VC})]\} \\
&\quad + \underset{\mathbf{s}}{E}\{\underset{\mathbf{e}}{var}[(\hat{\boldsymbol{\gamma}}_\mathbf{z} + \hat{\boldsymbol{\Delta}}'_\mathbf{z}\,\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_\mathbf{z} + \hat{\boldsymbol{\Delta}}'_\mathbf{z}\,\mathbf{x})]\Big|\mathbf{s}\} + \underset{\mathbf{s}}{E}[\underset{\mathbf{e}}{var}(\hat{\sigma}^2)\Big|\mathbf{s}] \\
&= \underset{\mathbf{s}}{var}[(\boldsymbol{\gamma}_\mathbf{z} + \boldsymbol{\Delta}'_\mathbf{z}\,\mathbf{x})'\mathbf{V}(\boldsymbol{\gamma}_\mathbf{z} + \boldsymbol{\Delta}'_\mathbf{z}\,\mathbf{x})] + \underset{\mathbf{s}}{E}\{\underset{\mathbf{e}}{var}[(\hat{\boldsymbol{\gamma}}_\mathbf{z} + \hat{\boldsymbol{\Delta}}'_\mathbf{z}\,\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_\mathbf{z} + \hat{\boldsymbol{\Delta}}'_\mathbf{z}\,\mathbf{x})]\Big|\mathbf{s}\} + \frac{2}{dfSSE}\sigma^4.
\end{aligned}
$$

Therefore, the mean squared error of $\hat{\sigma}^2_{YB|\mathbf{z}}$ is

$$
\begin{aligned}
MSE(\hat{\sigma}^2_{YB|\mathbf{z}}) &= [\sigma^2 trace(\mathbf{VC})]^2 + \underset{\mathbf{s}}{var}[(\boldsymbol{\gamma}_\mathbf{z} + \boldsymbol{\Delta}'_\mathbf{z}\,\mathbf{x})'\mathbf{V}(\boldsymbol{\gamma}_\mathbf{z} + \boldsymbol{\Delta}'_\mathbf{z}\,\mathbf{x})] \\
&\quad + \underset{\mathbf{s}}{E}\{\underset{\mathbf{e}}{var}[(\hat{\boldsymbol{\gamma}}_\mathbf{z} + \hat{\boldsymbol{\Delta}}'_\mathbf{z}\,\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_\mathbf{z} + \hat{\boldsymbol{\Delta}}'_\mathbf{z}\,\mathbf{x})]\Big|\mathbf{s}\} + \frac{2}{dfSSE}\sigma^4.
\end{aligned}
\qquad (A1)
$$

The mean squared error of $\hat{\sigma}^2_{Y|\mathbf{z}}$ is equal to its variance. Hence, from (2.22) and (2.23), we have

$$MSE(\hat{\sigma}^2_{Y|\mathbf{z}}) = \operatorname*{var}_{\mathbf{s}}[(\boldsymbol{\gamma}_{\mathbf{z}} + \boldsymbol{\Delta}'_{\mathbf{z}}\mathbf{x})'\mathbf{V}(\boldsymbol{\gamma}_{\mathbf{z}} + \boldsymbol{\Delta}'_{\mathbf{z}}\mathbf{x}) + \sigma^2]$$

$$+ \operatorname*{E}_{\mathbf{s}}(\operatorname*{var}_{\mathbf{e}}\{(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\mathbf{x}) + \hat{\sigma}^2[1 - trace(\mathbf{VC})]\})$$

$$= \operatorname*{var}_{\mathbf{s}}[(\boldsymbol{\gamma}_{\mathbf{z}} + \boldsymbol{\Delta}'_{\mathbf{z}}\mathbf{x})'\mathbf{V}(\boldsymbol{\gamma}_{\mathbf{z}} + \boldsymbol{\Delta}'_{\mathbf{z}}\mathbf{x})] + \operatorname*{E}_{\mathbf{s}}\{\operatorname*{var}_{\mathbf{e}}[(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\mathbf{x})]|\mathbf{s}\}$$

$$+ \operatorname*{E}_{\mathbf{s}}(\operatorname*{var}_{\mathbf{e}}\{\hat{\sigma}^2[1 - trace(\mathbf{VC})]\}|\mathbf{s}) \tag{A2}$$

$$= \operatorname*{var}_{\mathbf{s}}[(\boldsymbol{\gamma}_{\mathbf{z}} + \boldsymbol{\Delta}'_{\mathbf{z}}\mathbf{x})'\mathbf{V}(\boldsymbol{\gamma}_{\mathbf{z}} + \boldsymbol{\Delta}'_{\mathbf{z}}\mathbf{x})] + \operatorname*{E}_{\mathbf{s}}\{\operatorname*{var}_{\mathbf{e}}[(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_{\mathbf{z}} + \hat{\boldsymbol{\Delta}}'_{\mathbf{z}}\mathbf{x})]|\mathbf{s}\}$$

$$+ [1 - trace(\mathbf{VC})]^2 \frac{2}{dfSSE}\sigma^4 .$$

Comparing expressions (A1) and (A2), we see that $\hat{\sigma}^2_{Y|\mathbf{z}}$ is better than $\hat{\sigma}^2_{YB|\mathbf{z}}$ when

$$[1 - trace(\mathbf{VC})]^2 \frac{2}{dfSSE}\sigma^4 \leq [trace(\mathbf{VC})]^2\sigma^4 + \frac{2}{dfSSE}\sigma^4 .$$

The inequality must hold when $dfSSE \geq 2$ since $trace(\mathbf{VC}) > 0$.

APPENDIX B

Asymptotic Properties of the Estimators for the

Mean and Variance Models

We prove two results concerning the asymptotic properties of the estimators $\hat{\mu}_{Y|\mathbf{z}}$ and $\hat{\sigma}^2_{Y|\mathbf{z}}$. In order to prove the results, we need two other results from probability theory, which are stated without proof in Theorem B.1 and Theorem B.2. In the following, we denote by $\mathbf{A}_t \overset{D}{\to} \mathbf{A}$ the statement that $\mathbf{A}_1, \mathbf{A}_2, \ldots$ is a sequence of random variables that converges in distribution to $\mathbf{A}$ and we denote by $\mathbf{B}_t \overset{p}{\to} \boldsymbol{\eta}$ the statement that $\mathbf{B}_1, \mathbf{B}_2, \ldots$ is a sequence of random variables that converges in probability to $\boldsymbol{\eta}$. In addition, we write $\mathbf{b} \to \boldsymbol{\eta}$ to mean that $\mathbf{b}$ approaches $\boldsymbol{\eta}$ in the usual calculus sense.

Theorem B.1 If $g(\mathbf{a}, \mathbf{b})$ is a function jointly continuous at every point of the form $(\mathbf{a}, \boldsymbol{\eta})$ for some fixed $\boldsymbol{\eta}$, and if $\mathbf{A}_t \overset{D}{\to} \mathbf{A}$ and $\mathbf{B}_t \overset{p}{\to} \boldsymbol{\eta}$, then $g(\mathbf{A}_t, \mathbf{B}_t) \overset{D}{\to} g(\mathbf{A}, \boldsymbol{\eta})$.

Remark: This result is given in Haven et al. (2005).

Theorem B.2 If g is a function continuous at the point $\boldsymbol{\eta}$ and $\mathbf{B}_t \overset{p}{\to} \boldsymbol{\eta}$, then

$g(\mathbf{B}_t) \overset{p}{\to} g(\boldsymbol{\eta})$.

Remark: This result is given in Arnold (1981).

Theorem B.3 Assume that Assumptions 2.1-2.4 stated in Section 2.2.1 hold. If $\hat{\boldsymbol{\mu}}$ and

$\hat{\boldsymbol{\Sigma}}$ are consistent estimators, $\hat{\mu}_{Y|\mathbf{z}} \xrightarrow{D} \hat{\mu}_Y$ and $\hat{\sigma}^2_{Y|\mathbf{z}} \xrightarrow{D} \hat{\sigma}^2_Y$ as $m_1, \ldots, m_n \to \infty$.

Proof:

Firstly, we reason that $\hat{\mu}_{Y|\mathbf{z}}$ and $\hat{\sigma}^2_{Y|\mathbf{z}}$ are continuous functions of $\mathbf{e}$, $\hat{\boldsymbol{\mu}}$, and $\hat{\boldsymbol{\Sigma}}$.

This follows from the following observations.

The response for the lth experiment run is

$$y(\mathbf{x}_l, \boldsymbol{\xi}_l) = \beta_{0\xi} + \mathbf{x}'_l \boldsymbol{\beta}_\xi + \mathbf{x}'_l \mathbf{B}_\xi \mathbf{x}_l + \boldsymbol{\gamma}'_\xi \boldsymbol{\xi}_l + \mathbf{x}'_l \boldsymbol{\Delta}_\xi \boldsymbol{\xi}_l + e_l,$$

where $\boldsymbol{\xi}_l = (\xi_{l1}, \ldots, \xi_{ln}) = (\hat{\mu}_1 + z_{l1} c_1 \hat{\sigma}_1, \ldots, \hat{\mu}_n + z_{ln} c_n \hat{\sigma}_n)$. Thus, $y(\mathbf{x}_l, \boldsymbol{\xi}_l)$ is linear in

e_l, $\hat{\boldsymbol{\mu}}$, and the square root of each diagonal element of $\hat{\boldsymbol{\Sigma}}$.

Now, let $\hat{\beta}_{0\mathbf{z}}$, $\hat{\boldsymbol{\beta}}_\mathbf{z}$, $\hat{\mathbf{B}}_\mathbf{z}$, $\hat{\boldsymbol{\gamma}}_\mathbf{z}$, and $\hat{\boldsymbol{\Delta}}_\mathbf{z}$ be represented by $\hat{\boldsymbol{\theta}}_\mathbf{z}$. Note that

$$\hat{\boldsymbol{\theta}}_\mathbf{z} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},$$

where $\mathbf{Y}$ is the column vector of observations on the response, which has elements

$y(\mathbf{x}_l, \boldsymbol{\xi}_l), l = 1, \ldots, N$. Therefore, it is seen that each element of $\hat{\boldsymbol{\theta}}_\mathbf{z}$ is linear in the N

observations on the response $y(\mathbf{x}_l, \boldsymbol{\xi}_l), l = 1, \ldots, N$. Because of this, each element of $\hat{\boldsymbol{\theta}}_\mathbf{z}$

is a linear function of $\mathbf{e}$, $\hat{\boldsymbol{\mu}}$, and the square root of each diagonal element of $\hat{\boldsymbol{\Sigma}}$.

In addition, it can be shown that

$$\hat{\sigma}^2 = \{\mathbf{e}'[\mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{e}\}/(N-p),$$

where N is the number of experiment runs and $\mathbf{I}_N$ is an $N \times N$ identity matrix.

Therefore, it is clear that

$$\hat{\mu}_{Y|\mathbf{z}} = \hat{\beta}_{0\mathbf{z}} + \mathbf{x}'\hat{\boldsymbol{\beta}}_\mathbf{z} + \mathbf{x}'\hat{\mathbf{B}}_\mathbf{z}\mathbf{x}$$

and $\hat{\sigma}^2_{Y|\mathbf{z}} = (\hat{\boldsymbol{\gamma}}_\mathbf{z} + \hat{\boldsymbol{\Delta}}'_\mathbf{z} \mathbf{x})'\mathbf{V}(\hat{\boldsymbol{\gamma}}_\mathbf{z} + \hat{\boldsymbol{\Delta}}'_\mathbf{z} \mathbf{x}) + \hat{\sigma}^2[1 - trace(\mathbf{VC})]$

are continuous functions of $\mathbf{e}$, $\hat{\boldsymbol{\mu}}$, and $\hat{\boldsymbol{\Sigma}}$ (observe that because the noise variables are independent, $\hat{\boldsymbol{\Sigma}}$ is diagonal).

Now, let us write

$$\hat{\mu}_{Y|\mathbf{z}} = g_1[\mathbf{e},(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}})],$$

and $\hat{\sigma}^2_{Y|\mathbf{z}} = g_2[\mathbf{e},(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}})]$.

By Theorem B.1., if $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are consistent estimators so that $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ converge in probability to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ respectively as $m_1,\ldots,m_n \to \infty$, we have

$$\hat{\mu}_{Y|\mathbf{z}} = g_1[\mathbf{e},(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}})] \xrightarrow{D} g_1[\mathbf{e},(\boldsymbol{\mu},\boldsymbol{\Sigma})] = \hat{\mu}_Y$$

and $\hat{\sigma}^2_{Y|\mathbf{z}} = g_2[\mathbf{e},(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}})] \xrightarrow{D} g_2[\mathbf{e},(\boldsymbol{\mu},\boldsymbol{\Sigma})] = \hat{\sigma}^2_Y$

as $m_1,\ldots,m_n \to \infty$.

Theorem B.4 Suppose that Assumptions 2.1-2.4 and Assumption 2.8 stated in Section 2.2.1 hold. Assume that $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are consistent estimators and the design matrix expanded to model form $\mathbf{X}$ has full column rank. Let the number of replicates of the design be denoted by r. Then, $\hat{\mu}_{Y|\mathbf{z}} \xrightarrow{p} \mu_Y$ and $\hat{\sigma}^2_{Y|\mathbf{z}} \xrightarrow{p} \sigma^2_Y$ as $r,m_1,\ldots,m_n \to \infty$.

Proof:

If $\mathbf{X}$ is replicated r times,

$$\hat{\boldsymbol{\theta}}_\mathbf{z} = \frac{1}{r}(\mathbf{X'X})^{-1}\mathbf{X'}\sum_{j=1}^{r}\mathbf{Y}_j, \tag{B1}$$

where $\mathbf{Y}_j$ is the vector of observations on the response in the jth replicate.

Now,

$$\mathbf{Y}_j = \mathbf{X}\boldsymbol{\theta}_\mathbf{z} + \mathbf{e}_j, \tag{B2}$$

where $\boldsymbol{\theta_z}$ represents β_{0z}, $\boldsymbol{\beta_z}$, $\mathbf{B_z}$, $\boldsymbol{\gamma_z}$, and $\boldsymbol{\Delta_z}$, and $\mathbf{e}_j$ is the vector of experiment error for the jth replicate.

Putting together (B1) and (B2), we have

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_\mathbf{z} &= \frac{1}{r}(\mathbf{X'X})^{-1}\mathbf{X'}\sum_{j=1}^{r}(\mathbf{X}\boldsymbol{\theta}_\mathbf{z} + \mathbf{e}_j) \\
&= \boldsymbol{\theta}_\mathbf{z} + \frac{1}{r}(\mathbf{X'X})^{-1}\mathbf{X'}\sum_{j=1}^{r}\mathbf{e}_j
\end{aligned} \tag{B3}$$

Because each of the elements of each of the vectors $\mathbf{e}_1, \mathbf{e}_2,\ldots$ are independently and identically distributed with mean zero and constant variance,

$$\frac{1}{r}(\mathbf{X'X})^{-1}\mathbf{X'}\sum_{j=1}^{r}\mathbf{e}_j \xrightarrow{p} 0 \text{ as } r \to \infty$$

by the Weak Law of Large Numbers.

Furthermore, it can be seen from (2.11)-(2.15) that $\beta_{0z} \xrightarrow{p} \beta_0$, $\boldsymbol{\beta_z} \xrightarrow{p} \boldsymbol{\beta}$, $\mathbf{B_z} \xrightarrow{p} \mathbf{B}$, $\boldsymbol{\gamma_z} \xrightarrow{p} \boldsymbol{\gamma}$, and $\boldsymbol{\Delta_z} \xrightarrow{p} \boldsymbol{\Delta}$ as $m_1,\ldots,m_n \to \infty$ since $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are consistent estimators. Thus, if we write $\boldsymbol{\theta}$ for β_0, $\boldsymbol{\beta}$, $\mathbf{B}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\Delta}$, we have

$$\boldsymbol{\theta}_\mathbf{z} \xrightarrow{p} \boldsymbol{\theta} \text{ as } m_1,\ldots,m_n \to \infty.$$

Hence, by (B3), $\hat{\boldsymbol{\theta}}_\mathbf{z} \xrightarrow{p} \boldsymbol{\theta}$ as $r, m_1,\ldots,m_n \to \infty$. Let N be the total number of experiment runs. Arnold (1981) shows that $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ as $N \to \infty$. Now, since N is a linear function of r, $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ as $r \to \infty$. Thus, we have $(\hat{\boldsymbol{\theta}}_\mathbf{z}, \hat{\sigma}^2) \xrightarrow{p} (\boldsymbol{\theta}, \sigma^2)$ as $r, m_1,\ldots,m_n \to \infty$. Because $\hat{\mu}_{Y|\mathbf{z}}$ and $\hat{\sigma}^2_{Y|\mathbf{z}}$ are continuous functions of $(\hat{\boldsymbol{\theta}}_\mathbf{z}, \hat{\sigma}^2)$, it follows by Theorem B.2 that $\hat{\mu}_{Y|\mathbf{z}} \xrightarrow{p} \mu_Y$ and $\hat{\sigma}^2_{Y|\mathbf{z}} \xrightarrow{p} \sigma^2_Y$.

APPENDIX C

Convexity of the Objective Function of Program V

To proof that IVV is convex on the open convex set

$$O_V = \{(m_1,\ldots,m_n,r_f,N); m_j > 1, j = 1,\ldots,n, r_f > 0, N > p\},$$

we use the fact that a sum of convex functions is convex and a twice-differentiable function is convex if its Hessian is positive semidefinite (Bazaraa et al., 1993). First, observe that IVV is a sum of the functions

$$\left(\frac{2}{m_j-1}+\frac{\phi_{2j}}{m_j}\right)\frac{F_j}{c_j^4}, \quad j=1,\ldots.n, \tag{C1}$$

$$2\left(\frac{\sigma^2}{fr_f}\right)^2 G\left[\sum_{j=1}^{n}\frac{1}{c_j^4}+\frac{1}{N-p}\left(\sum_{j=1}^{n}\frac{1}{c_j^2}\right)^2\right], \tag{C2}$$

$$4\frac{\sigma^2}{fr_f}\sum_{j=1}^{n}\frac{H_j}{c_j^4}. \tag{C3}$$

It is shown that each of the functions given by (C1)-(C3) is convex on O_V.

Since for each $j = 1,\ldots,n$, $\phi_{2j} \geq -2$ and $F_j > 0$,

$$\frac{d^2}{dm_j^2}\left(\frac{2}{m_j-1}+\frac{\phi_{2j}}{m_j}\right)\frac{F_j}{c_j^4}=\left[\frac{4}{(m_j-1)^3}+\frac{2\phi_{2j}}{m_j^3}\right]\frac{F_j}{c_j^4}>0 \quad \forall m_j >1,$$

Therefore, each function in (C1) must be convex on O_V.

The Hessian of the function in (C2) with respect to r_f and N is

$$\begin{pmatrix} 12\dfrac{\sigma^4}{f^2 r_f^4}G\left[\displaystyle\sum_{j=1}^{n}\dfrac{1}{c_j^4}+\dfrac{1}{N-p}\left(\displaystyle\sum_{j=1}^{n}\dfrac{1}{c_j^2}\right)^2\right] & -4\dfrac{\sigma^4}{f^2 r_f^3}G\left[-\dfrac{1}{(N-p)^2}\left(\displaystyle\sum_{j=1}^{n}\dfrac{1}{c_j^2}\right)^2\right] \\[4mm] -4\dfrac{\sigma^4}{f^2 r_f^3}G\left[-\dfrac{1}{(N-p)^2}\left(\displaystyle\sum_{j=1}^{n}\dfrac{1}{c_j^2}\right)^2\right] & 2\dfrac{\sigma^4}{f^2 r_f^2}G\left[\dfrac{2}{(N-p)^3}\left(\displaystyle\sum_{j=1}^{n}\dfrac{1}{c_j^2}\right)^2\right] \end{pmatrix}. \qquad \text{(C4)}$$

Now, a 2×2 matrix is positive semidefinite if and only if its diagonal elements and determinant are non-negative (Bazaraa et al., 1993). It can be seen that the matrix in (C4) satisfies this requirement when $r_f > 0$ and $N > p$ (note that $G > 0$). Therefore, the function in (C2) is convex on O_V.

Finally, since each $H_j > 0$, we have

$$\dfrac{d^2}{dr_f^2}\left(4\dfrac{\sigma^2}{fr_f}\sum_{j=1}^{n}\dfrac{H_j}{c_j^4}\right)=8\dfrac{\sigma^2}{fr_f^3}\sum_{j=1}^{n}\dfrac{H_j}{c_j^4}>0 \quad \forall r_f > 0.$$

This implies that the function in (C3) is convex on O_V.

Since IVV is a sum of functions that are convex on O_V, it is convex on O_V.

APPENDIX D

Convexity of $IM_E\,/\sigma^2$

The convexity of $IM_E\,/\sigma^2$ is proven through the following series of results. Note that it is always assumed that k is a positive integer, and f and α are positive real numbers.

Lemma D.1 A symmetric matrix $\mathbf{H}$ can be expressed as $\mathbf{JJ'}$ for some matrix $\mathbf{J}$ if and only if it is positive semidefinite.

Proof:

By the principle axis theorem, $\mathbf{H} = \mathbf{\Gamma D \Gamma'}$, where $\mathbf{\Gamma}$ is an orthogonal matrix and $\mathbf{D}$ is the matrix of eigenvalues (Arnold, 1981). Suppose $\mathbf{\Gamma D \Gamma'} = \mathbf{JJ'}$, then

$$\mathbf{D} = \mathbf{\Gamma' J J' \Gamma} = (\mathbf{J'\Gamma})'\mathbf{J'\Gamma}\,.$$

Therefore, the eigenvalues of $\mathbf{H}$ cannot be negative so that $\mathbf{H}$ is positive semidefinite.

Conversely, suppose that $\mathbf{H}$ is positive semidefinite. If we let $\mathbf{J} = \mathbf{\Gamma D}^{1/2}$, we have $\mathbf{H} = \mathbf{JJ'}$.

Remark: A slightly different proof of this result is given in Harville (1997).

Lemma D.2 If R is bounded, the matrix of region moments $\mathbf{\mu}_R = \int_R \mathbf{x}_C \mathbf{x'}_C d\mathbf{x} / \int_R d\mathbf{x}$ is a positive semidefinite matrix.

Proof:

Let u denote the number of parameters in the mean model, which is $1 + 2k + k(k-1)/2$.

Note that

$$\int_R \mathbf{x'}_C \mathbf{H}\mathbf{x}_C d\mathbf{x} / \int_R d\mathbf{x} = trace[\mathbf{H}\boldsymbol{\mu}_R] \geq 0$$ for any arbitrary positive semidefinite matrix $\mathbf{H}$

of dimension $u \times u$ that is not a function of $\mathbf{x}$.

Since $\boldsymbol{\mu}_R$ is symmetric, $\boldsymbol{\mu}_R = \boldsymbol{\Gamma}\mathbf{D}\boldsymbol{\Gamma'}$, where $\boldsymbol{\Gamma}$ is an orthogonal matrix and $\mathbf{D}$

is the matrix of eigenvalues.

Now, since $\mathbf{H}$ can be any arbitrary positive semidefinite matrix, choose

$\mathbf{H} = \boldsymbol{\Gamma}\mathbf{W}\mathbf{W}\boldsymbol{\Gamma'}$, where $\mathbf{W}$ is a diagonal matrix with real diagonal elements

$w_j, j = 1, \ldots, u$. Thus, we have

$$\int_R \mathbf{x'}_C \mathbf{H}\mathbf{x}_C d\mathbf{x} / \int_R d\mathbf{x}$$
$$= trace(\boldsymbol{\Gamma}\mathbf{W}\mathbf{W}\boldsymbol{\Gamma'}\boldsymbol{\Gamma}\mathbf{D}\boldsymbol{\Gamma'})$$
$$= trace(\mathbf{W}\mathbf{D}\mathbf{W})$$
$$= trace\left[\begin{pmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_u \end{pmatrix}\begin{pmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_u \end{pmatrix}\begin{pmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_u \end{pmatrix}\right]$$
$$= \sum_{j=1}^{u} w_j^2 d_j \geq 0.$$

Observe that we may choose $w_j > 0$ and $w_i = 0, i \neq j$. Therefore, we see that

each $d_j \geq 0$. This means that $\boldsymbol{\mu}_R$ is positive semidefinite.

Theorem D.1 Suppose the elements of $\mathbf{M}_C$ are linear functions of $\mathbf{t}$ over a convex set

$\mathbf{T}$ such that $\mathbf{M}_C$ is positive definite for all $\mathbf{t} \in \mathbf{T}$. In addition, suppose that R is

bounded. Then, $IM_E / \sigma^2 = \int_R \mathbf{x'}_C \mathbf{V}_C \mathbf{x}_C d\mathbf{x} / \int_R d\mathbf{x}$, where $\mathbf{V}_C = \mathbf{M}_C^{-1}$, is a convex

function of $\mathbf{t}$ for all $\mathbf{t} \in \mathbf{T}$.

Proof:

Let $\boldsymbol{\mu}_R = \mathbf{U}\mathbf{U'}$ for some square matrix $\mathbf{U}$. We have

$$IM_E / \sigma^2 = trace(\mathbf{V}_C \mathbf{U} \mathbf{U}') = trace(\mathbf{U}' \mathbf{V}_C \mathbf{U}).$$

If we let $\mathbf{U}_j$ denote the jth column of $\mathbf{U}$, we can write

$$IM_E / \sigma^2 = \sum_{j=1}^{u} \mathbf{U}'_j \mathbf{V}_C \mathbf{U}_j.$$

Groves and Rothenberg (1969) and Rahman and Ahsanullah (1973) showed that for any two positive definite matrices $\mathbf{A}$ and $\mathbf{B}$, and any vector $\mathbf{d}$, $f(\lambda) = \mathbf{d}'[(1-\lambda)\mathbf{A} + \lambda\mathbf{B}]^{-1}\mathbf{d}$ is a convex function of λ so that $\mathbf{d}'\mathbf{A}^{-1}\mathbf{d}$ is convex in $\mathbf{A}$. Rearrange the elements of $\mathbf{A}$ into a column vector $\mathbf{a}$. Therefore, $\mathbf{d}'\mathbf{A}^{-1}\mathbf{d}$ is a convex function of the elements of $\mathbf{a}$ over Ξ, where Ξ is the set of values of $\mathbf{a}$ such that $\mathbf{A}$ is positive definite. Note that Ξ is a convex set. We may write $g(\mathbf{a}) = \mathbf{d}'\mathbf{A}^{-1}\mathbf{d}$ so that $g(\mathbf{a})$ is a convex function of $\mathbf{a}$ for $\mathbf{a} \in \Xi$. In addition, if $\mathbf{a}$ is a linear function of $\mathbf{t}$ so that $\mathbf{a} = \mathbf{Pt} + \mathbf{b}$ for some matrix $\mathbf{P}$ and column vector $\mathbf{b}$, $g(\mathbf{Pt} + \mathbf{b})$ is a convex function of $\mathbf{t}$ on the convex set $\mathbf{T}_0$, where $\mathbf{T}_0$ is the set of all $\mathbf{t}$ such that $\mathbf{a} = \mathbf{Pt} + \mathbf{b} \in \Xi$ whenever $\mathbf{t} \in \mathbf{T}_0$. Therefore, $g(\mathbf{Pt} + \mathbf{b})$ is also a convex function of $\mathbf{t}$ on a convex set $\mathbf{T} \subseteq \mathbf{T}_0$.

If we set $\mathbf{d} = \mathbf{U}_j$ and $\mathbf{A} = \mathbf{M}_C$ in the arguments in the preceding paragraph, we see that $\mathbf{U}'_j \mathbf{V}_C \mathbf{U}_j$ is a convex function of the elements of $\mathbf{M}_C$ over the set of values where $\mathbf{M}_C$ is positive definite. If these elements are linear functions of a set of variables represented by $\mathbf{t}$, $\mathbf{U}'_j \mathbf{V}_C \mathbf{U}_j$ is a convex function of $\mathbf{t}$ on any convex set $\mathbf{T}$ such that $\mathbf{M}_C$ is positive definite for all $\mathbf{t} \in \mathbf{T}$. Finally, since the sum of convex functions is convex, IM_E / σ^2 is a convex function of $\mathbf{t}$ on $\mathbf{T}$.

Theorem D.2 $\mathbf{M}_C$ is positive definite over

$$\chi = \{(r_f, r_a, r_c); r_f > 0, r_a > 0, r_c > -(2fr_f r_a)(\alpha^2 - k)^2 / (kfr_f + 2r_a\alpha^4)\},$$ which is a

convex set.

Proof:

Let $\mathbf{d}' = (d_1, d_2, \ldots, d_u)$ be an arbitrary vector. First, consider the case where

$k \geq 2$. We have

$\mathbf{d}'\mathbf{M}_C\mathbf{d}$

$$= \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_u \end{pmatrix}' \cdot \begin{pmatrix} fr_f + 2kr_a + r_c & \mathbf{0} & (fr_f + 2\alpha^2 r_a)\mathbf{1}'_k & \mathbf{0} \\ \mathbf{0} & (fr_f + 2\alpha^2 r_a)\mathbf{I}_k & \mathbf{0} & \mathbf{0} \\ (fr_f + 2\alpha^2 r_a)\mathbf{1}_k & \mathbf{0} & (fr_f)(\mathbf{1}_k\mathbf{1}'_k) + 2\alpha^4 r_a\mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & fr_f\mathbf{I}_{\binom{k}{2}} \end{pmatrix} \cdot \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_u \end{pmatrix}$$

$$= (fr_f + 2kr_a + r_c)d_1^2 + \sum_{l=k+2}^{2k+1}(fr_f + 2\alpha^2 r_a)d_l d_1 + \sum_{l=2}^{k+1}(fr_f + 2\alpha^2 r_a)d_l^2$$

$$+ \sum_{l=k+2}^{2k+1}(fr_f + 2\alpha^2 r_a)d_1 d_l + \sum_{l=k+2}^{2k+1}(fr_f + 2\alpha^4 r_a)d_l^2 + 2\sum_{m=k+2}^{2k}\sum_{l=m+1}^{2k+1}fr_f(d_l d_m) + \sum_{l=2k+2}^{u}fr_f d_l^2$$

$$= (fr_f)d_1^2 + 2fr_f d_1\sum_{l=k+2}^{2k+1}d_l + fr_f\sum_{l=k+2}^{2k+1}d_l^2 + 2fr_f\sum_{m=k+2}^{2k}\sum_{l=m+1}^{2k+1}d_l d_m$$

$$+ (2kr_a + r_c)d_1^2 + 4\alpha^2 r_a d_1\sum_{l=k+2}^{2k+1}d_l + 2\alpha^4 r_a\sum_{l=k+2}^{2k+1}d_l^2$$

$$+ \sum_{l=2}^{k+1}(fr_f + 2\alpha^2 r_a)d_l^2 + \sum_{l=2k+2}^{u}fr_f d_l^2$$

$$= fr_f\left(d_1 + \sum_{l=k+2}^{2k+1}d_l\right)^2 + 2r_a\left[\sum_{l=k+2}^{2k+1}(d_1 + \alpha^2 d_l)^2\right] + r_c d_1^2$$

$$+ (fr_f + 2\alpha^2 r_a)\sum_{l=2}^{k+1}d_l^2 + fr_f\sum_{l=2k+2}^{u}d_l^2 .$$

Note that $f > 0$, and $\alpha > 0$. Therefore, $\mathbf{d}'\mathbf{M}_C\mathbf{d} > 0$ for all $\mathbf{d} \neq 0$ only if

$r_f > 0$ and $r_a > 0$. To see this, observe the following.

1. If $r_f \leq 0$, choose nonzero values for $d_l, l = 2k + 2, \ldots, u$ and zero values for all

other elements of $\mathbf{d}$, and we have $\mathbf{d}'\mathbf{M}_C\mathbf{d} \leq 0$ for some $\mathbf{d} \neq 0$.

2. If $r_a \leq 0$, choose $d_1 = 0$; $d_l, l = k+2, \ldots, 2k+1$, not all zero, such that $\sum_{l=k+2}^{2k+1} d_l = 0$;

and zero values for all other elements of $\mathbf{d}$. This leads to $\mathbf{d'M}_C\mathbf{d} \leq 0$ for some $\mathbf{d} \neq 0$.

Given that $r_f > 0$ and $r_a > 0$, we want to find the minimum of $\mathbf{d'M}_C\mathbf{d}$ so that

we can determine the values of r_c such that $\mathbf{d'M}_C\mathbf{d} > 0$ for all $\mathbf{d} \neq 0$. Note the

following facts.

1. $(fr_f + 2\alpha^2 r_a)\sum_{l=2}^{k+1} d_l^2 + fr_f \sum_{l=2k+2}^{u} d_l^2 = 0$ is minimized when each $d_l = 0$, where

$l \in (2, 3, \ldots, k+1, 2k+2, 2k+3, \ldots, u)$.

2. If $d_1 = 0$, $fr_f\left(d_1 + \sum_{l=k+2}^{2k+1} d_l\right)^2 + 2r_a\left[\sum_{l=k+2}^{2k+1}(d_1 + \alpha^2 d_l)^2\right] + r_c d_1^2 = 0$ is minimized when

each $d_l = 0$, where $l \in (k+2, k+3, \ldots, 2k+1)$.

Thus, if the constraint $d_1 = 0$ is imposed on $\mathbf{d}$, the minimum of $\mathbf{d'M}_C\mathbf{d}$ is

zero, and is achieved when $\mathbf{d} = 0$ regardless of the value of r_c.

If $d_1 \neq 0$, we may write

$$fr_f\left(d_1 + \sum_{l=k+2}^{2k+1} d_l\right)^2 + 2r_a\left[\sum_{l=k+2}^{2k+1}(d_1 + \alpha^2 d_l)^2\right] + r_c d_1^2$$

$$= d_1^2\left\{fr_f\left(1 + \sum_{l=k+2}^{2k+1} d_l/d_1\right)^2 + 2r_a\left[\sum_{l=k+2}^{2k+1}(1 + \alpha^2 d_l/d_1)^2\right] + r_c\right\}.$$

Consider the function

$$\psi(\theta_{k+2}, \ldots, \theta_{2k+1}) = fr_f\left(1 + \sum_{l=k+2}^{2k+1} \theta_l\right)^2 + 2r_a\left[\sum_{l=k+2}^{2k+1}(1 + \alpha^2 \theta_l)^2\right] + r_c. \tag{D1}$$

It has the following first and second order derivatives.

$$\frac{\partial \psi}{\partial \theta_m} = 2fr_f \left(1 + \sum_{l=k+2}^{2k+1} \theta_l \right) + 4r_a \alpha^2 (1 + \alpha^2 \theta_m), \quad m = k+2, \ldots, 2k+1.$$

$$\frac{\partial^2 \psi}{\partial \theta_m^2} = 2fr_f + 4r_a \alpha^4, \qquad\qquad m = k+2, \ldots, 2k+1.$$

$$\frac{\partial^2 \psi}{\partial \theta_p \partial \theta_m} = 2fr_f, \quad p \neq m, \, p = k+2, \ldots, 2k+1, \, m = k+2, \ldots, 2k+1.$$

The Hessian matrix of $\psi(\theta_{k+2}, \ldots, \theta_{2k+1})$ is given by

$$\mathbf{H} = 2 \begin{pmatrix} fr_f + 2r_a\alpha^4 & fr_f & \cdots & fr_f \\ fr_f & fr_f + 2r_a\alpha^4 & \cdots & fr_f \\ \vdots & \vdots & \ddots & \vdots \\ fr_f & fr_f & \cdots & fr_f + 2r_a\alpha^4 \end{pmatrix}.$$

Now, let $\mathbf{I}_k$ be the $k \times k$ identity matrix and λ a scalar. Using the diagonal expansion rule of the determinant (Searle, 1982), we find that

$$\det(\mathbf{H} - \lambda \mathbf{I}_k) = (4r_a\alpha^4 - \lambda)^k + (2kfr_f)(4r_a\alpha^4 - \lambda)^{k-1}$$
$$= (4r_a\alpha^4 - \lambda)^{k-1}[(4r_a\alpha^4 - \lambda) + (2kfr_f)].$$

Because the eigenvalues of a symmetric matrix are real (Arnold, 1981), all eigenvalues of $\mathbf{H}$ are positive, and the Hessian is positive definite. This implies that the global minimum of ψ can be found by solving the equations

$$\frac{\partial \psi}{\partial \theta_m} = 2fr_f \left(1 + \sum_{l=k+2}^{2k+1} \theta_l \right) + 4r_a \alpha^2 (1 + \alpha^2 \theta_m) = 0, \quad m = k+2, \ldots, 2k+1. \tag{D2}$$

Now, the following two equations are obtained from (D2).

$$(1 + \alpha^2 \theta_m) = -\frac{fr_f (\alpha^2 - k)}{kfr_f + 2r_a\alpha^4}, \quad m = k+2, \ldots, 2k+1. \tag{D3}$$

$$\sum_{l=k+2}^{2k+1} \theta_l = \frac{-k(2r_a\alpha^2 + fr_f)}{kfr_f + 2r_a\alpha^4}, \quad l = k+2, \ldots, 2k+1. \tag{D4}$$

Substituting (D3) and (D4) into (D1), we have the following expression for the global minimum of ψ, which we denote by $\psi_{\min}$.

$$\psi_{\min} = fr_f \left(1 + \sum_{l=k+2}^{2k+1} \theta_l \right)^2 + 2r_a \left[\sum_{l=k+2}^{2k+1} (1 + \alpha^2 \theta_l)^2 \right] + r_c$$

$$= fr_f \left[\frac{(\alpha^2 - k)(2r_a \alpha^2)}{kfr_f + 2r_a \alpha^4} \right]^2 + 2kr_a \left[-\frac{fr_f (\alpha^2 - k)}{(kfr_f + 2r_a \alpha^4)} \right]^2 + r_c$$

$$= \frac{fr_f (2r_a \alpha^2)^2 (\alpha^2 - k)^2 + 2kr_a (fr_f)^2 (\alpha^2 - k)^2}{(kfr_f + 2r_a \alpha^4)^2} + r_c$$

$$= \frac{(kfr_f + 2r_a \alpha^4)(2fr_f r_a)(\alpha^2 - k)^2}{(kfr_f + 2r_a \alpha^4)^2} + r_c$$

$$= \frac{(2fr_f r_a)(\alpha^2 - k)^2}{kfr_f + 2r_a \alpha^4} + r_c \,.$$

Therefore, $\psi > 0$ for any real values assigned to $\theta_m, m = k + 2, \ldots, 2k + 1$ when

$$r_c > -(2fr_f r_a)(\alpha^2 - k)^2 / (kfr_f + 2r_a \alpha^4) \,.$$

Hence for $k \geq 2$, we conclude that $\mathbf{M}_C$ is positive definite if and only if

$(r_f, r_a, r_c) \in \chi$, where

$$\chi = \{(r_f, r_a, r_c); r_f > 0, r_a > 0, r_c > -(2fr_f r_a)(\alpha^2 - k)^2 / (kfr_f + 2r_a \alpha^4)\} \,.$$

Now, consider the case where $k = 1$. We have

$$\mathbf{d}' \mathbf{M}_C \mathbf{d} = fr_f (d_1 + d_3)^2 + 2r_a (d_1 + \alpha^2 d_3)^2 + r_c d_1^2 + (fr_f + 2\alpha^2 r_a) d_2^2 \,.$$

It can be seen that if $r_f > 0$ and $r_a > 0$, $d_2 = 0$ so that $\mathbf{d}' \mathbf{M}_C \mathbf{d}$ is minimized.

Now, if we set $d_1 = 0$, then $\mathbf{d}' \mathbf{M}_C \mathbf{d}$ is minimized at $\mathbf{d} = 0$ for any r_c. On the other

hand, if $d_1 \neq 0$, the minimum of $d_1^2 [fr_f (1 + d_3 / d_1)^2 + 2r_a (1 + \alpha^2 d_3 / d_1)^2 + r_c]$ is

$d_1^2 [(2fr_f r_a)(\alpha^2 - 1)^2 / (fr_f + 2r_a \alpha^4) + r_c]$. Thus, for the case where $k = 1$, $\mathbf{M}_C$ is

positive definite on

$$\chi = \{(r_f, r_a, r_c); r_f > 0, r_a > 0, r_c > -(2fr_f r_a)(\alpha^2 - 1)^2 / (fr_f + 2r_a \alpha^4)\} \,.$$

Now, since $-(2fr_f r_a)(\alpha^2 - k)^2 /(kfr_f + 2r_a\alpha^4)$ is a convex function of r_f and r_a as can readily be verified by deriving the Hessian of the function, it follows that χ is a convex set.

Corollary D.1 $IM_E / \sigma^2 = trace[\mathbf{V}_C \int_R \mathbf{x}_C \mathbf{x}'_C d\mathbf{x} / \int_R d\mathbf{x}]$ is convex in the variables r_f, r_a, and r_c over $\chi = \{(r_f, r_a, r_c); r_f > 0, r_a > 0, r_c > -(2fr_f r_a)(\alpha^2 - k)^2 /(kfr_f + 2r_a\alpha^4)\}$.

Proof:

This follows directly from Theorems D.1 and D.2, and the fact that the elements $\mathbf{M}_C$ are linear functions of r_f, r_a, and r_c.

APPENDIX E

Experimental Designs for Schemes Compared with CD Plots

D.1   Design for Scheme 4 of Example 4.3

Design given by Castillo et al. (2007) for $k = 3, n = 2$

x_1	x_2	x_3	z_1	z_2
-1	1	-1	-1	-1
1	-1	1	-1	1
-1	-1	-1	1	1
-1	-1	-1	-1	-1
-1	1	-1	1	-1
1	-1	-1	-1	1
-1	1	1	-1	-1
1	1	-1	-1	-1
-1	1	-1	1	-1
1	1	-1	1	1
-1	1	-1	1	1
-1	-1	1	1	-1
1	-1	1	-1	-1
1	1	1	-1	1
-1	1	1	1	1
1	1	1	1	-1
-1	1	-1	-1	1
1	-1	-1	1	-1
1	0	0	1	1
1	-1	1	1	1
0	0	1	0	0
1	0	0	-1	-1
-1	-1	1	-1	1
0	-1	0	0	0

D.2 Design for Scheme 1 of Example 4.4

45 Run D-Optimal Design

x_1	x_2	x_3	z_1	z_2	x_1	x_2	x_3	z_1	z_2
-1	-1	-1	-1	-1	0	1	1	-1	-1
-1	-1	1	1	1	1	0	1	-1	1
-1	1	-1	1	1	1	1	0	-1	1
-1	1	1	-1	-1	0	1	1	1	1
1	-1	-1	-1	1	-1	1	0	1	-1
1	-1	-1	1	-1	-1	-1	-1	1	-1
1	1	1	-1	1	1	1	-1	1	-1
1	1	1	1	-1	-1	-1	-1	-1	1
1	-1	1	-1	-1	-1	1	1	-1	1
1	-1	1	1	1	1	-1	-1	-1	1
1	1	-1	-1	-1	-1	-1	1	-1	-1
1	1	-1	1	1	-1	1	-1	-1	-1
-1	-1	-1	1	1	-1	0	1	1	-1
-1	-1	1	-1	1	0	1	-1	-1	1
-1	1	-1	-1	1	1	0	-1	1	1
0	-1	0	-1	-1	1	1	1	1	1
0	0	-1	-1	-1	1	-1	-1	-1	-1
-1	-1	1	1	-1	1	1	1	-1	-1
-1	1	-1	1	-1	-1	0	0	-1	1
-1	1	1	1	1	1	-1	1	1	-1
0	-1	0	1	1	1	-1	0	1	1
0	0	-1	1	1	0	-1	1	-1	1
1	0	0	-1	-1					

D.3 Design for Scheme 4 of Example 4.4

25 Run D-Optimal Design

x_1	x_2	x_3	z_1	z_2
-1	-1	-1	-1	-1
-1	1	1	-1	-1
1	-1	-1	-1	1
1	-1	-1	1	-1
1	1	1	-1	1
1	1	1	1	-1
1	-1	1	-1	-1
1	-1	1	1	1
1	1	-1	-1	-1
1	1	-1	1	1
-1	-1	-1	1	1
-1	-1	1	-1	1
-1	1	-1	-1	1
-1	-1	1	1	-1
-1	1	-1	1	-1
0	-1	0	1	1
-1	0	1	1	1
-1	1	0	1	1
0	1	1	1	1
1	0	0	-1	1
1	0	0	1	-1
-1	-1	0	-1	-1
0	1	-1	-1	-1
0	0	1	-1	-1
-1	0	-1	-1	1