# Model Selection and Analysis Tools in Response Surface Modeling of the Process Mean and Variance

by

Kristi L. Griffiths

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

DOCTOR OF PHILOSOPHY

IN

STATISTICS

APPROVED:

Raymond H. Myers, Chair

Marvin M. Lentner

Jeffrey B. Birch

Marion R. Reynolds, Jr.

Klaus H. Hinkelmann

April, 1995
Blacksburg, Virginia

# MODEL SELECTION AND ANALYSIS TOOLS IN RESPONSE SURFACE MODELING OF THE PROCESS MEAN AND VARIANCE

by

Kristi L. Griffiths

Raymond H. Myers, Chair

Statistics

(ABSTRACT)

Product improvement is a serious issue facing industry today. And while response surface methods have been developed which address the process mean involved in improving the product there has been little research done on the process variability. Lack of quality in a product can be attributed to its inconsistency in performance thereby highlighting the need for a methodology which addresses process variability. The key to working with the process variability comes in the handling of the two types of factors which make up the product design: control and noise factors. Control factors can be fixed in both the lab setting and the real application. However, while the noise factors can be fixed in the lab setting, they are assumed to be random in the real application. A response-model can be created which models the response as a function of both the control and noise factors. This work introduces criteria for selecting an appropriate response-model which can be used to create accurate models for both the process mean and process variability. These two models can then be used to identify settings of the control factors which minimize process variability while maintaining an acceptable process mean.

If the response-model is known, or at least well estimated, response surface methods can be extended to building various confidence regions related to the process variance. Among these are a confidence region on the location of minimum process variance and a confidence region on the ratio of the process variance to the error variance. It is easy to see the importance for research on the process variability and this work offers practical methods for improving the design of a product.

# Acknowledgement

It has been an honor and a pleasure to work with my advisor, Dr. Ray Myers, on this research. I wish to thank him for his invaluable guidance and great insight, both in the research process and in helping me decide my future. Dr. Myers is truly an excellent teacher.

I would also like to thank the members of my committee: Marvin Lentner, Jeff Birch, Klaus Hinkelmann, and Marion Reynolds. Their suggestions and questions helped direct the course of the research.

Throughout my academic career my family has shown me their constant and unending support. Special thanks are due to my parents for encouraging me to continue my education and reach this success. Their love and support have kept me going and kept me grounded.

# Table of Contents


**List of Figures** ..... vii

**List of Tables** ..... viii

**Chapter 1. Overview of Taguchi Methods**

§1.1 Introduction ..... 1
§1.2 Location and Dispersion Effects ..... 2
§1.3 Taguchi's Robust Parameter Design (RPD) ..... 3
§1.4 Inefficiencies Associated with Taguchi's RPD Approach ..... 4

**Chapter 2. Response Surface Methodology Approach**

§2.1 Improvements Over Taguchi's RPD ..... 6
§2.2 The Combined Array ..... 7
Example 2.1 ..... 8
§2.3 Modeling Product Performance Using the Response-Model ..... 10
§2.4 Importance of Accurately Modeling the Process Variance ..... 11

**Chapter 3. Criteria for Process Variance Model Selection**

§3.1 Introduction to the Process Variance ..... 13
§3.2 Estimators of the Process Mean and Variance ..... 15
§3.3 Effect of Response-Model Selection on the Process Variance ..... 21
Example 3.1 (Chemical Data) ..... 21
§3.4 Average MSE of the Estimated Process Variance ..... 25
§3.5 Average MSE of the Estimated Slopes of the Response-Model ..... 26
§3.6 Extending the Response-Model ..... 29

**Chapter 4. Inference on Individual Process Variance Model Terms**

§4.1 Introduction ..... 32
§4.2 Distributions of the Estimated Coefficients ..... 32
§4.3 Procedure for Calculating p-values ..... 35
§4.4 Interpreting p-values ..... 38

**Chapter 5. Analysis of "Residuals" for the Process Variance Model**

§5.1 Introduction ..... 39

§5.2 Relationship Between $PV_B$ and $s^2$ ..... 39

§5.3 Use of Residuals ..... 40

Example 5.1 ..... 45

§5.4 Extending RSM Analysis ..... 47

**Chapter 6. Further RSM Relating to the Process Variance**

§6.1 Confidence Region on the Location of Minimum Process Variance 52

Example 6.1 (Filtration Rate Example) ..... 53

§6.2 Confidence Region on $\frac{\underline{l}'\underline{l} + \sigma_\epsilon^2}{\sigma_\epsilon^2}$ ..... 57

Example 6.2 ..... 58

§6.3 One-Sided Tolerance Intervals on Future Observations ..... 59

§6.4 Prediction Limits on a Future Response ..... 63

Example 6.3 ..... 66

§6.5 Results of Simulation ..... 66

§6.5.1 Results of Simulation on the Confidence Region on $\frac{\underline{l}'\underline{l} + \sigma_\epsilon^2}{\sigma_\epsilon^2}$ ... 67

§6.5.2 Results of Simulation on Prediction Limits ..... 69

§6.6 Effect of the Extended Response-Model on the Various Confidence Regions ..... 71

§6.6.1 Confidence Region on the Location of Minimum Process Variance 71

§6.6.2 Confidence Region on $\frac{\underline{l}'\underline{l} + \sigma_\epsilon^2}{\sigma_\epsilon^2}$ ..... 71

§6.6.3 One-Sided Tolerance Intervals on Future Observations ..... 72

§6.6.4 Prediction Limits on a Future Response ..... 72

**Chapter 7. RSM in the Presence of Categorical Noise Factors**

§7.1 Introduction ..... 74

§7.2 Estimation of the Process Variance ..... 76

§7.3 One-Sided Confidence Region on $\frac{\underline{l}'\mathrm{V}\,\underline{l} + \sigma_\epsilon^2}{\sigma_\epsilon^2}$ ..... 77

§7.4 Lower One-Sided Tolerance Intervals on Future Observations .... 79

§7.5 Prediction Limits on a Future Response ..... 81

**Chapter 8. Summary and Future Research**

§8.1 Summary and Future Research ..... 84

**Bibliography** ..... 85

**Appendix A. Computation of AVGMSE(B) and AVGMSE(U)** ..... 90

**Appendix B. Properties of $s^2(\underline{x})$**

§B.1 Distribution of $s^2(\underline{x})$ ..... 95

§B.2 Equality of $s^2(\underline{x})$ and $PV_B$ when the Response-Model is Saturated ..... 96

**Appendix C. Approximate Distributions Required for the Various Confidence Regions in the Case of the Extended Response-Model**

§C.1 Confidence Region on $\frac{\underline{l}'\underline{l} + \sigma_\epsilon^2}{\sigma_\epsilon^2}$ ..... 98

§C.2 One-Sided Tolerance Intervals on Future Observations ..... 100

§C.3 Prediction Limits on a Future Response ..... 100

**Appendix D. Derivation for the Lower One-Sided Tolerance Interval in the Presence of Categorical Noise Factors** ..... 102

**Vita** ..... 105

# List of Figures


| Figure | | Page |
|---|---|---|
| 1.1 | Taguchi's $2^2$ x $2^2$ Crossed Array | 4 |
| 2.1 | Control by Noise Interactions | 10 |
| 3.1 | Values of the Correction Factor in Case I | 18 |
| 3.2 | Values of the Correction Factor in Case II | 19 |
| 3.3 | Values of the Correction Factor in Case III | 20 |
| 3.4 | Contour plot of the biased estimate of the process variance resulting from Model A | 23 |
| 3.5 | Contour plot of the biased estimate of the process variance resulting from Model B | 24 |
| 5.1 | Standardized residuals vs Ln $PV_B$ corresponding to the "true" response-model. | 48 |
| 5.2 | Standardized residuals vs Ln $PV_B$ corresponding to the response-model without $z_1$ | 49 |
| 5.3 | Standardized residuals vs Ln $PV_B$ corresponding to the response-model without $x_1z_2$ | 50 |
| 5.4 | Standardized residuals vs Ln $PV_B$ corresponding to the response-model without $x_4z_1$ | 51 |
| 6.1 | Contour plot of the mean filtration rate and the line of minimum process variance along with its 95% confidence region | 55 |
| 6.2 | Contour plot of both the mean filtration rate and the process standard deviation | 56 |

# List of Tables


Table — Page

6.1 Coverage Probabilities and Standard Errors in % from the Simulation of the Confidence Region on $\frac{\underline{l}'\underline{l} + \sigma_\epsilon^2}{\sigma_\epsilon^2}$ when $r_x$=3 and dfE=40 ..... 68

6.2 Coverage Probabilities and Standard Errors in % from the Simulation of the Confidence Region on $\frac{\underline{l}'\underline{l} + \sigma_\epsilon^2}{\sigma_\epsilon^2}$ when $r_x$=2 and dfE=20 ..... 68

6.3 Coverage Probabilities and Standard Errors in % from the Simulation of Prediction Limits when $r_x$=3 and dfE=40 ..... 70

6.4 Coverage Probabilities and Standard Errors in % from the Simulation of Prediction Limits when $r_x$=2 and dfE=20 ..... 70

# Chapter 1
# Overview of Taguchi Methods

## §1.1 Introduction

There has been a significant improvement over the past 60 years in the area of product development. The 1920's saw Shewhart developing sampling inspection plans as a means of quality control in the manufacturing stage of a product. Then, the design of the process was improved using process control techniques during the 1940's. But more recently, industry has moved to the idea of better product design. The awareness of improving product design is due in large part to Genichi Taguchi. He emphasized the notion that the lack of quality in a product is due to its inconsistency in performance. Product variability was seen as a major contributor to poor quality. In 1980 Taguchi introduced a methodology by which a product could be designed in such a way as to be insensitive to factors influencing the variability, e.g. tolerances on design factors, variations in environmental variables, and variations in consumer usage (see Taguchi and Wu (1980), Kackar (1985), Myers, Khuri, and Vining (1992), and Myers and Kim (1995)).

The basic principle behind this methodology is to choose the levels of the design factors such that:

1. the product meets or exceeds existing quality specifications and

2. the product is not easily influenced by variations in manufacturing or consumer usage environments.

This type of product design is said to be robust to the environment. Taguchi calls this approach Robust Parameter Design (RPD). While the concept of RPD has

been widely accepted by engineers and scientists, Taguchi's specific methodology behind RPD has come under attack by Box (1985)(1988)(1992), Lucas (1985), Wu (1992), Myers and Vining (1992), and Shoemaker, Tsui, and Wu (1991). The criticisms produced alternative ideas for applying RPD. In this chapter we will discuss how factors may influence the mean and/or variability of a product, how to handle those factors, and how Taguchi's RPD methodology leads to alternative methodologies.

## §1.2 Location and Dispersion Effects

Some design factors affect the mean value of the performance characteristic. These factors are called *location* effects. Other design factors, called *dispersion* effects, affect the variability of the performance characteristic. A design factor may be both a location and dispersion effect. When this occurs, the design factors are said to exhibit "crosstalk" (Box (1988)). Box (1988) suggested modeling the process mean and variance separately to isolate location and dispersion effects.

While the idea of modeling the variability is not due to Taguchi he did realize that by understanding dispersion effects the process variability could be reduced. This would in turn lead to an increase in quality. According to Kackar (1985), "In industrial processes, controlling variability is usually much harder than controlling the mean value, and variability is often a root cause of high manufacturing costs".

Modeling the mean response is standard procedure in statistical methodology. However, much has yet to be learned about modeling the variability. Myers, Khuri, and Vining (1992) discuss the notion that the model for the process variance may be more important than the model for the mean, since consistency in the performance of the product is of vital importance. Fortunately, the concept of variability is quite natural to engineers and there are diagnostic techniques which now include examining dispersion effects. For general discussion, see Box and Meyer (1986), Nair and Pregibon (1988), and Bartlett and Kendall (1946).

## §1.3 Taguchi's Robust Parameter Design (RPD)

When applying RPD it is important to understand the nature of the design factors. In a given process, design factors can be classified into two separate categories: control factors and noise factors. Control factors, denoted by $\underline{x}$, can be controlled in the lab, pilot plant, and the real application, e.g. consumer use or full-scale production. Noise factors, on the other hand, denoted by $\underline{z}$, are assumed to be controllable in the lab and pilot plant setting, but may vary in the real application.

Consider an example of baking a cake. The control factors could be the ingredients and the noise factors could be the amount of milk or flour added, baking time, baking temperature, and position of cake in the oven. The objective of the food scientist creating the cake recipe would be to determine the type and amount of ingredients which would make the response, say taste of the cake, robust to the baking time, baking temperature, etc. Myers, Khuri, and Vining note that, "This robustness to noise variables can only be incorporated at the product design stage, which, of course, is ideally the stage in which experimental design is most useful" (1992).

Taguchi's approach uses performance statistics which summarize the information produced by the variability in the noise factors, called Signal-to-Noise (S-N) ratios. S-N ratios address both the mean and variability. He suggests modeling the S-N ratios against the control factors. The different S-N ratios will not be presented here, but they are outlined in Kackar (1985) and detailed in Box (1988). The experimental designs advocated by Taguchi are based on orthogonal arrays. A common approach to creating an orthogonal array is to combine all possible settings of the various design factors which are orthogonal to each other. Taguchi's designs combine two orthogonal arrays. The "inner array" is made up of settings in the control factors and the "outer array" is made up of settings in the noise factors. The inner and outer arrays are completely cross-classified often resulting in a relatively large experiment. For example, two control factors at two levels each crossed with two noise factors at two levels each would result in a $2^2x2^2$ completely randomized crossed array requiring 16 experimental runs as

depicted in Figure 1.1.

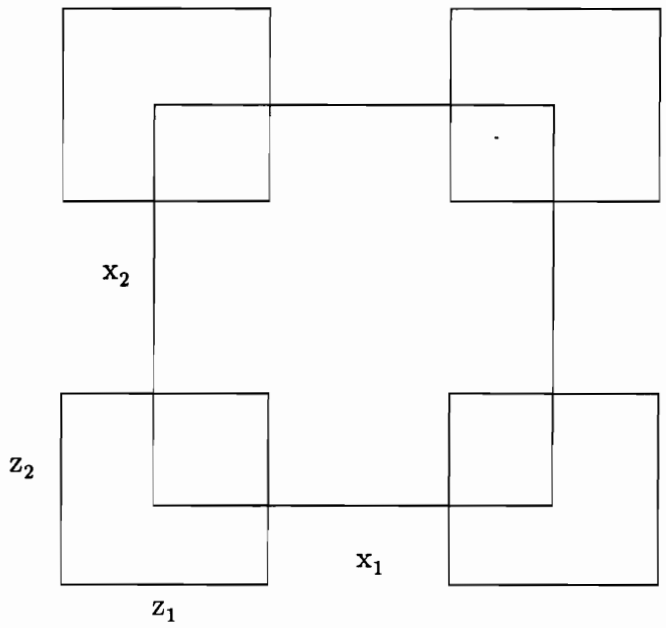



**Figure 1.1** Taguchi's $2^2x2^2$ Crossed Array

## §1.4 Inefficiencies Associated with Taguchi's RPD Approach

Box (1985, 1992) states that Taguchi's engineering ideas are important and should be understood. However, it would be foolish to learn the details of many of his proposed methods. While Taguchi's approach represents a significant advance in the application of experimental designs to the improvement of product quality, it has several shortcomings. Among these shortcomings are the following:

1. the advocated designs are limited and fail to adequately deal with interactions between control factors,

2. the sequential nature of investigations is not exploited,
3. the Taguchi approach involves using techniques which are inefficient and complicated,
4. the universal use of S-N ratios as performance statistics is unconvincing,
5. Taguchi's approach is preoccupied with finding the optimal conditions rather than understanding the underlying process, and
6. crossed arrays often require an inordinate number of experimental runs.

Vining and Myers (1990), Lucas (1990), Welch, Yu, Kang, and Sacks (1990), Box (1985, 1992), Shoemaker, Tsui, and Wu (1991), Shoemaker and Tsui (1992), Myers, Khuri, and Vining (1992), and Myers and Kim (1995) all recognize that these shortcomings can be overcome by using a different approach to RPD. The improved approach uses a more efficient design, called the combined array, intended for handling interactions and sequential investigations. The combined array also utilizes the information about a product's performance allowing for a greater knowledge of the underlying process. This new approach nicely accommodates the use of response surface methodology (RSM) and has the ability to pay particular attention to the process variability.

# Chapter 2
# Response Surface Methodology Approach

## §2.1 Improvements Over Taguchi's RPD

Combining Taguchi's parameter design principles with conventional response surface methodology (RSM) yields an approach to RPD that is an alternative to the Taguchi approach. The advantages are due to the fact that RSM incorporates a modeling strategy with a systematic approach to variable screening, sequential investigation, and exploration of the design region. Unlike Taguchi's approach, RSM places a great deal of emphasis on understanding the underlying process while striving to accurately estimate the optimum conditions.

By combining Taguchi's notion of heterogeneous variability induced by uncontrollable design factors with RSM modeling and design techniques, the process can be improved. The RSM approach accomplishes this by:

1. identifying settings of the control factors which minimize the effect of the noise factors and thus minimize the variability,

2. understanding the underlying process in the region of interest for the purpose of seeing tradeoffs between the process mean and variance, and

3. selecting and using models to maintain the process after improvements have been made.

The techniques used in RSM require that both the control and noise factors be treated as fixed effects for the purpose of experimental design and estimation. Once the experimental design is carried out and the data are collected the noise factors are treated as random effects with certain distribution properties. This

allows both the process mean and process variance to be expressed as functions of only the control factors. Fortunately, it is possible to create a relatively efficient experimental design to accommodate these modeling techniques.

## §2.2 The Combined Array

Recall Taguchi's approach to experimental design from §1.3. He advocates the use of two completely crossed orthogonal arrays which typically results in a relatively large experiment. A more efficient approach "combines" the control and noise factors in the same array. This allows for more flexibility in the type of design. The combined array is more efficient than the crossed array for several reasons. Welch, Yu, Kang, and Sacks (1990), Shoemaker, Tsui, and Wu (1991), and Myers, Khuri, and Vining (1992) discuss many of these advantages.

The clearest advantage is simply the smaller size of the experiment required. Dealing with a single fractional factorial, for example, will cut the number of experimental runs at least in half when compared to the crossing of two fractional designs. Another advantage is how well it accommodates a wide range of models. Typically, depending on the surface of the true region of interest, it is important to use a model which uses first or second order terms in the control factors, a first order model in the noise factors, and control by noise interactions. Other terms, such as noise by noise interactions or quadratic noise effects, could also be included.

Unlike Taguchi's crossed array the combined array is very flexible, particularly when it comes to handling interactions. Taguchi's designs do accommodate control by noise interactions, but they use too many degrees of freedom by estimating terms which are typically unimportant. In the combined array you can specify which interactions you want to estimate. The RSM approach places a great deal of emphasis on the control by noise interactions for good reason. These interactions drive the process variance model, allowing us to manipulate the process variance through the levels of the control factors. As a result, it is important that an experimental design be used which allows for efficient estimation of the control by noise interactions.

The following example, taken from Myers, Khuri, and Vining (1992), illustrates a comparison between the Taguchi and RSM approaches to experimental design.

**Example 2.1**

Consider a situation in which a simple experiment is to be run. There are three control factors ($x_1$, $x_2$, and $x_3$) at three levels each and two noise factors ($z_1$ and $z_2$) at two levels each. The Taguchi approach would be to use the $L_9$ design for the inner array crossed with a $2^2$ factorial design for the outer array. The $L_9$ is a three-level Plackett-Burman design for three factors constructed from a Latin Square. It is given by:

| $x_1$ | $x_2$ | $x_3$ |
|---|---|---|
| -1 | -1 | -1 |
| -1 | 0 | 0 |
| -1 | 1 | 1 |
| 0 | -1 | 0 |
| 0 | 0 | 1 |
| 0 | 1 | -1 |
| 1 | -1 | 1 |
| 1 | 0 | -1 |
| 1 | 1 | 0 |

The coding, -1, 0, and 1, represents the three levels of the control factors. Crossing this design with the outer array, $2^2$ factorial design, yields an experiment requiring 36 runs. The resulting array does not allow estimation of all control by control interactions due to the fact that control by control interactions are aliased to some degree with other important effects.

An obvious design for the RSM approach would be the central composite design. Recall this approach combines the control and noise factors into the same array. It is given by:

| $x_1$ | $x_2$ | $x_3$ | $z_1$ | $z_2$ |
|---|---|---|---|---|
| $\pm 1$ | $\pm 1$ | $\pm 1$ | $\pm 1$ | $\pm 1$ |
| -1 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 |
| 0 | -1 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 0 |
| 0 | 0 | -1 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 |
| $\underline{0}$ | $\underline{0}$ | $\underline{0}$ | $\underline{0}$ | $\underline{0}$ |

The first row of this array represents a resolution V fraction of a $2^5$ factorial design containing 16 experimental runs. The last row represents a set of $n_0 \geq 1$ center runs. This design allows linear and quadratic effects as well as two-factor interactions to be estimated in the control factors and linear effects and two-factor interactions to be estimated in the noise factors. It also allows the important control by noise interactions to be estimated. An appropriate fitted model for this example might be:

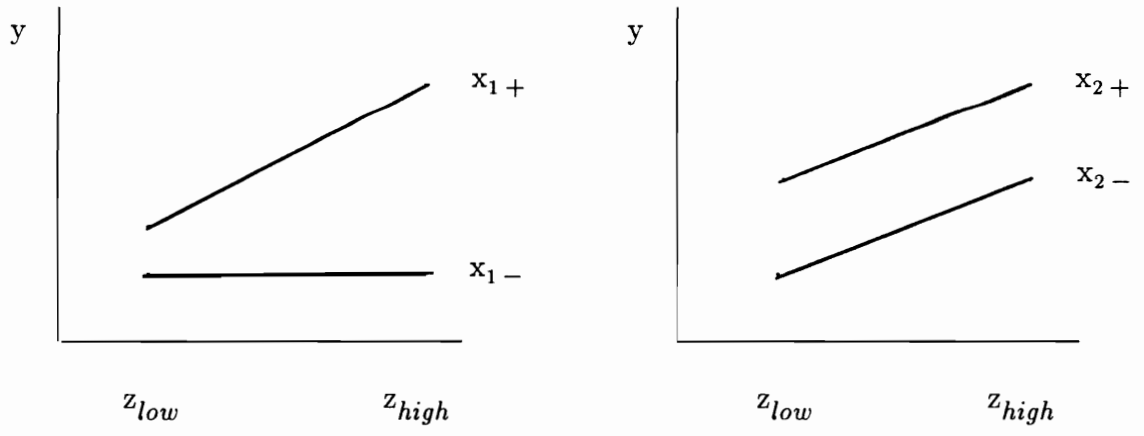
$$\hat{y} = b_0 + \sum_{i=1}^{3} b_i x_i + \sum_{i<k} \sum b_{ik} x_i x_k + \sum_{i=1}^{3} b_{ii} x_i^2 + \sum_{j=1}^{2} \hat{\gamma}_j z_j + \sum_i \sum_j \hat{\delta}_{ij} x_i z_j. \quad (2.2.1)$$

The above design requires 22 + $n_0$ experimental runs as compared to 36 runs using Taguchi's approach. As mentioned earlier, Taguchi's crossed array wastes degrees of freedom on certain higher order control by noise interactions which are not important. For example, Taguchi's design allows the estimation of $x_1^2 z_1$, $x_2^2 z_1$, $x_1^2 z_1 z_2$, $x_1^2 z_2$, $x_2^2 z_2$, etc. Experience shows that these interactions are typically not important. The degrees of freedom wasted on these interactions could be more efficiently used on control by control interactions with the combined array. This is just one of several examples that shows how the RSM designs are more economical than those advocated by Taguchi and can be used to produce information on process variability.

## §2.3 Modeling Product Performance Using the Response-Model

In the environment, e.g. consumer use, noise factors are not set or fixed. They are random according to some distribution. In RSM the noise factors are fixed effects. RSM takes advantage of assumptions about the noise factors for the purpose of estimating the process mean and variance. When at least one control by noise interaction is found to be significant in the model, process variability actually depends on the control factor which is involved in the interaction. Details will be given later. It is this interaction structure which allows us to estimate the process variability.

A simple illustration will show how the control by noise interactions play an important role in understanding the dependence of the process variability on the control factors. Consider an experiment which has two control factors at two levels each and one noise factor. The interaction plot below, Figure 2.1, could have been generated from a fixed-effects model on the continuous factors $x_1$, $x_2$, and z.




(a) $x_1$ has a positive dispersion effect (b) $x_2$ has no dispersion effect

**Figure 2.1** Control by Noise Interactions

If z is truly random in the process, then setting $x_1$ at its low level will certainly minimize variability in the process, i.e. $x_1$ has a positive dispersion effect. However, changes in the settings of $x_2$ will not affect the variability in the response. Thus, if a high response is desired, one should set $x_2$ at the high level. So, the fitted fixed-effects model on $\underline{x}$ and z and the variance of z allow the process variance to be viewed as a function of $\underline{x}$. Since the process variance can be controlled by the settings of the control factors it would be beneficial to understand the relationship between the process variance and the control factors. This would allow us to determine the settings of the control factors which minimize the process variability, assuming the settings exist in the design region.

A generalization of the model of Equation (2.2.1) for the natural response as a function of the control and noise factors is given by

$$y = \beta_0 + \sum_{i=1}^{r_x} \beta_i x_i + \sum\sum_{i<k} \beta_{ik} x_i x_k + \sum_{i=1}^{r_x} \beta_{ii} x_i^2 + \sum_{j=1}^{r_z} \gamma_j z_j + \sum_i \sum_j \delta_{ij} x_i z_j + \epsilon. \qquad (2.3.1)$$

This model is called the ***response-model*** (Shoemaker et al. (1991)). This particular form has been considered by Box and Jones (1990), Lucas (1990), and Myers, Khuri, and Vining (1992). Noise by noise interaction terms as well as quadratic noise factors could also be included in the response-model and the methodology that follows will address this issue.

### §2.4 Importance of Accurately Modeling the Process Variance

In order to determine the optimum operating conditions or to better understand the underlying process, the estimation of the process variance is critical. Shoemaker and Tsui (1992), Shoemaker, Tsui, and Wu (1991), and Myers and Kim (1995) discuss the sensitivity of the estimated process variance to response-model selection. Shoemaker et al. (1991) recommend that both examination of control by noise interaction plots and prediction of loss from the fitted response-model be included in the RSM approach. They go on to say that if a noise factor that has an interaction with a control factor is eliminated from the model, the combined array experiment may lead to settings of the control factors

which actually increase the process variability. In essence, if the analysis of data deals with the use of a model for the process variance, a methodology must be developed to ascertain the quality of the model (Myers and Kim (1995)). Otherwise, an RSM approach to RPD may lead to incorrect representations of the process. In turn, this could lead to a parameter design which is not robust to uncontrollable design factors.

The purpose of this research is to investigate several aspects of the process variance model. In particular, Chapter 3 addresses the effect of response-model selection on estimation of the process variance and considers various criteria for selecting a response-model. Once a response-model has been chosen, Chapters 4 and 5 provide techniques for evaluating its resulting estimated process variance model by calculating p-values of the estimated coefficients and examining residuals to detect nonideal conditions. Chapters 6 and 7 explore various confidence regions and sampling properties of the estimated process variance model through the use of the slopes of the estimated response-model. For example, confidence intervals on the mean response and prediction intervals on a future response are standard techniques in regression modeling and RSM. While these topics are a natural part of standard modeling or RSM, they are obviously more complicated in the case where the distribution and properties of the response depend on changing variance as well as mean.

# Chapter 3
# Criteria for Process Variance Model Selection

## §3.1 Introduction to the Process Variance

The purpose of this chapter is to introduce various criteria for selecting an appropriate response-model which will yield an accurate process variance model. The response-model, as defined in Equation (2.3.1), is the basis for determining the process mean and process variance models. Thus, it is necessary to understand how it is formed. Consider a model in $r_x$ control factors and $r_z$ noise factors for some process with natural response y. The response-model, written in matrix notation, is given by

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'B\,\underline{x} + \underline{z}'\underline{\gamma} + \underline{x}'\Delta\,\underline{z} + \epsilon, \tag{3.1.1}$$

where $\epsilon \sim N(0, \sigma_\epsilon^2)$,

$\underline{x}' = [x_1, x_2, \ldots, x_{r_x}]$,

$\underline{z}' = [z_1, z_2, \ldots, z_{r_z}]$,

with parameters $\underline{\beta}' = [\beta_1, \beta_2, \ldots, \beta_{r_x}]$

and $\underline{\gamma}' = [\gamma_1, \gamma_2, \ldots, \gamma_{r_z}]$.

The matrix B is $r_x$ by $r_x$ and contains $\beta_{ii}$ (i = 1, 2, ..., $r_x$) on the main diagonal and $\frac{1}{2}\beta_{ik}$ (i $\neq$ k) on the (i, k) off-diagonal. The matrix $\Delta$ is $r_x$ by $r_z$ and contains the $\delta_{ij}$, the coefficients of control by noise interactions. It can be quite common for the control factors to contain only two levels in the design. As a result, no quadratic control factors would be included. However, this model is general and

could be extended to include quadratic noise factors and noise by noise interaction terms. The extended response-model will be discussed in §3.6

The basic idea behind RPD is to choose the levels of the control factors which minimize the effect of the noise factors. A response surface of the process variability can be produced as a function of the control factors through the response-model (3.1.1). While the noise factors are considered fixed or controlled in the experimental design, the RSM practitioner assumes certain properties about them for the purpose of estimating the process mean and variance. These assumptions are made in accordance with design level centering and scaling. We assume that

$$\begin{aligned} \mathrm{E}(\underline{z}) &= \underline{0} \\ \mathrm{Var}(\underline{z}) &= \mathrm{V}. \end{aligned} \tag{3.1.2}$$

As a result, the response surfaces for the process mean and variance can be obtained by taking the expectation and variance operators, respectively, through the response-model (3.1.1) making use of the assumptions in (3.1.2) (Myers, Khuri, and Vining (1992)). Thus, the process mean model is given by

$$\mathrm{E}_{\underline{z}}(\mathrm{y}) = \beta_0 + \underline{\mathrm{x}}'\underline{\beta} + \underline{\mathrm{x}}'\mathrm{B}\,\underline{\mathrm{x}}. \tag{3.1.3}$$

And, more important to the topic of product design, carrying the variance operator through the response-model (3.1.1) forms the process variance model given by

$$\mathrm{Var}_{\underline{z}}(\mathrm{y}) = (\underline{\gamma} + \Delta'\underline{\mathrm{x}})'\mathrm{V}(\underline{\gamma} + \Delta'\underline{\mathrm{x}}) + \sigma_\epsilon^2 \tag{3.1.4}$$

where V is the variance-covariance matrix of $\underline{z}$. Now, for the purpose of illustration, if we assume uncorrelated noise factors with common variance

$$\sigma_{z_j}^2 = \sigma_z^2 \text{ , for all } \mathrm{j} = 1, 2, \ldots, \mathrm{r}_z$$

then Equation (3.1.4) can be written

$$\mathrm{Var}_{\underline{z}}(y) = (\underline{\gamma} + \Delta'\underline{x})'(\underline{\gamma} + \Delta'\underline{x})\sigma_z^2 + \sigma_\epsilon^2. \tag{3.1.5}$$

Furthermore, by coding the noise factors to $\pm 1$ corresponding to $\pm\sigma_z$, and by assuming they are independent and identically distributed with $\sigma_z^2 = 1$, then

$$V = I. \tag{3.1.6}$$

And, the process variance can be further simplified as

$$\mathrm{Var}_{\underline{z}}(y) = (\underline{\gamma} + \Delta'\underline{x})'(\underline{\gamma} + \Delta'\underline{x}) + \sigma_\epsilon^2. \tag{3.1.7}$$

Notice that the derivative of the response with respect to the noise factors can be written

$$\frac{\partial y}{\partial \underline{z}} = (\underline{\gamma} + \Delta'\underline{x}).$$

Denote this derivative by $\underline{l}$. Then the process variance can be expressed

$$\mathrm{Var}_{\underline{z}}(y) = \underline{l}'\underline{l} + \sigma_\epsilon^2. \tag{3.1.8}$$

This equation is a quadratic form in the slopes, $l_j$, of the response-model where each $l_j$ is a polynomial function of the control factors, $\underline{x}$. It is the presence of $\underline{x}$ in $\underline{l}$ which allows us to minimize the process variability by adjusting the levels of the control factors. However, before this can be done the process variance needs to be estimated.

### §3.2 Estimators of the Process Mean and Variance

The most obvious estimator of the process mean can be constructed by replacing parameters by their estimates. Thus, the estimated process mean is

$$\widehat{E}_{\underline{z}}(y) = \widehat{\beta}_0 + \underline{x}'\underline{\widehat{\beta}} + \underline{x}'\widehat{B}\,\underline{x}. \tag{3.2.1}$$

This estimator is unbiased for the process mean.

There are two estimators of the process variance which will be considered. An intuitive, but biased, estimator and an unbiased estimator. Recall the process variance of Equation (3.1.8)

$$\mathrm{Var}_{\underline{z}}(\mathrm{y}) = \underline{l}\,'\underline{l} + \sigma_\epsilon^2. \tag{3.2.2}$$

An intuitive estimator for the process variance model would simply be

$$\widehat{\mathrm{Var}}_{\underline{z}}(\mathrm{y}) = \underline{\hat{l}}\,'\underline{\hat{l}} + \hat{\sigma}_\epsilon^2, \tag{3.2.3}$$

where $\underline{\hat{l}} = (\underline{\hat{\gamma}} + \underline{\hat{\Delta}}'\underline{\mathrm{x}})$ and $\hat{\sigma}_\epsilon^2$ = estimated mean square error from the fitted response-model. The parameters in (3.2.2) are replaced by their estimates taken from the fitted response-model. For ease of notation this estimator will be called $\mathrm{PV}_B$ for "biased estimate of the process variance". $\mathrm{PV}_B$ is indeed biased since

$$\mathrm{E}(\mathrm{PV}_B) = \mathrm{E}(\underline{\hat{l}}\,'\underline{\hat{l}} + \hat{\sigma}_\epsilon^2) = \underline{l}\,'\underline{l} + \sigma_\epsilon^2\,\mathrm{tr}(\mathrm{C}) + \sigma_\epsilon^2 \tag{3.2.4}$$

where C is the variance-covariance matrix of $\underline{\hat{l}}$, apart from $\sigma_\epsilon^2$.

Another estimator of the process variance is

$$\widehat{\mathrm{Var}}_{\underline{z}}(\mathrm{y}) = \underline{\hat{l}}\,'\underline{\hat{l}} + [1 - \mathrm{tr}(\mathrm{C})]\,\hat{\sigma}_\epsilon^2. \tag{3.2.5}$$

This estimator, denoted $\mathrm{PV}_U$, is unbiased for the process variance. Notice that both estimators, $\mathrm{PV}_B$ and $\mathrm{PV}_U$, involve the estimated slopes, $\underline{\hat{l}}$. Accurately estimating $\underline{\hat{l}}$ would lead to a more accurate estimate for the process variance. This idea leads to two criteria which will be discussed in §3.5.

$\mathrm{PV}_B$ and $\mathrm{PV}_U$ each have qualities which make them attractive. $\mathrm{PV}_B$ is an intuitive estimator. Parameters are replaced by their estimates taken from the fitted response-model. This makes it very simple to use. On the other hand, $\mathrm{PV}_U$ has the quality of being unbiased for the process variance.

Consideration of the ratio of $\frac{PV_U}{PV_B}$ would depict how much smaller values of $PV_U$ are as compared to $PV_B$ estimates in real life situations. This ratio can be simplified to:

$$1 - \frac{\hat{\sigma}_\epsilon^2 \, \mathrm{tr}(C)}{\underline{\hat{l}}' \underline{\hat{l}} + \hat{\sigma}_\epsilon^2} \quad . \tag{3.2.6}$$

In this simpler form the second term contains the useful information for determining if there are substantial differences between the estimators. This term will be called the *correction factor.* Thus, it would be useful to examine the values of the correction factor throughout the design region for various sizes of $\underline{\hat{l}}$. Values of the correction factor near zero would suggest that the two estimators, $PV_B$ and $PV_U$, yielded similar estimates. One should bear in mind that the process variance is a function of the levels of the control factors.

For the purpose of illustration, 2 controls factors and 2 noise factors will be considered at two levels each. A combined array constructed of 16 experimental runs is assumed for the purpose of calculating the standard errors of the estimated slopes. The standard errors will be used to consider three cases for the size of $\underline{\hat{l}}$.

Case I: $\hat{l}_j = 1 \; \mathrm{SE}(\hat{l}_j)$
Case II: $\hat{l}_j = 2 \; \mathrm{SE}(\hat{l}_j)$
Case III: $\hat{l}_j = 3 \; \mathrm{SE}(\hat{l}_j)$

Notice that the estimation of $\hat{l}_j$ improves from Case I to Case III. This example allows the values of the correction factor to be plotted against settings of $\underline{x}$. Figures 3.1 - 3.3 display graphs of each case of the correction factors throughout the design region, respectively.

In summary, examination of the three cases illustrates that as the estimation of the slopes improve, the value of the correction factor throughout the majority of the design region approaches zero. Thus, the more accurate the slopes are estimated, the less difference there is between the estimates produced by $PV_B$ and $PV_U$. Generally there is not a large discrepancy between the two estimators in practical situations.

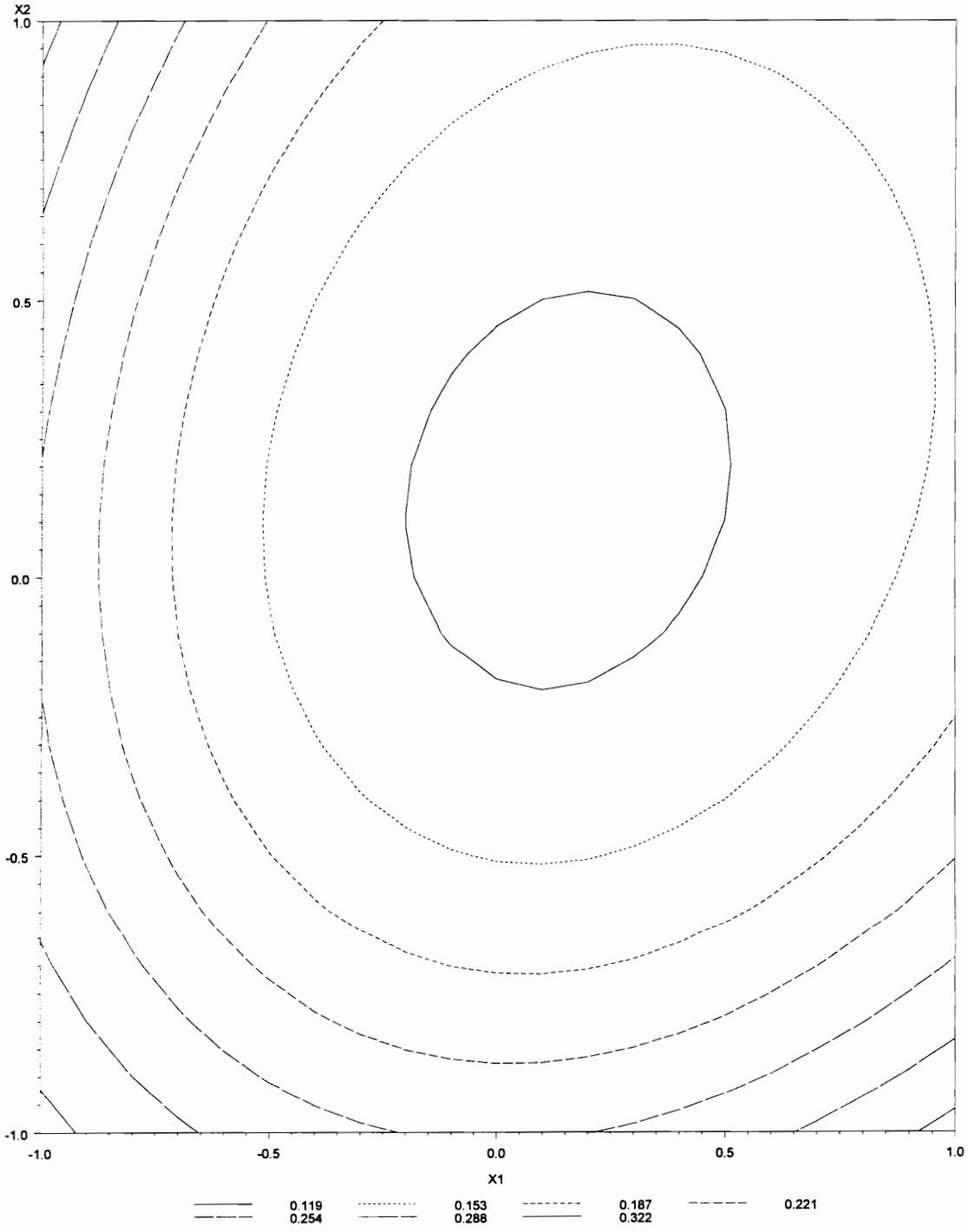



**Figure 3.1** Values of the Correction Factor in Case I.

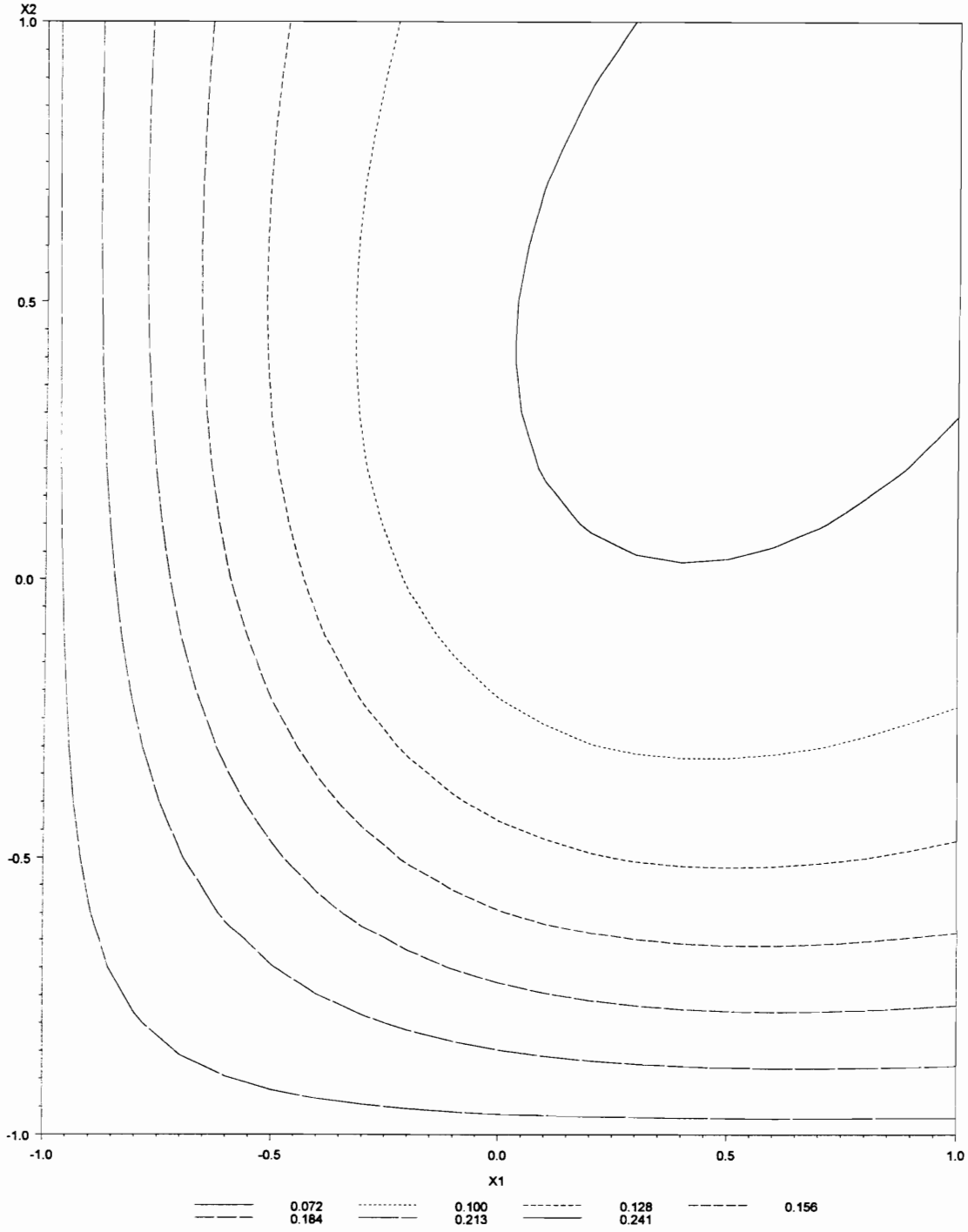



**Figure 3.2** Values of the Correction Factor in Case II.

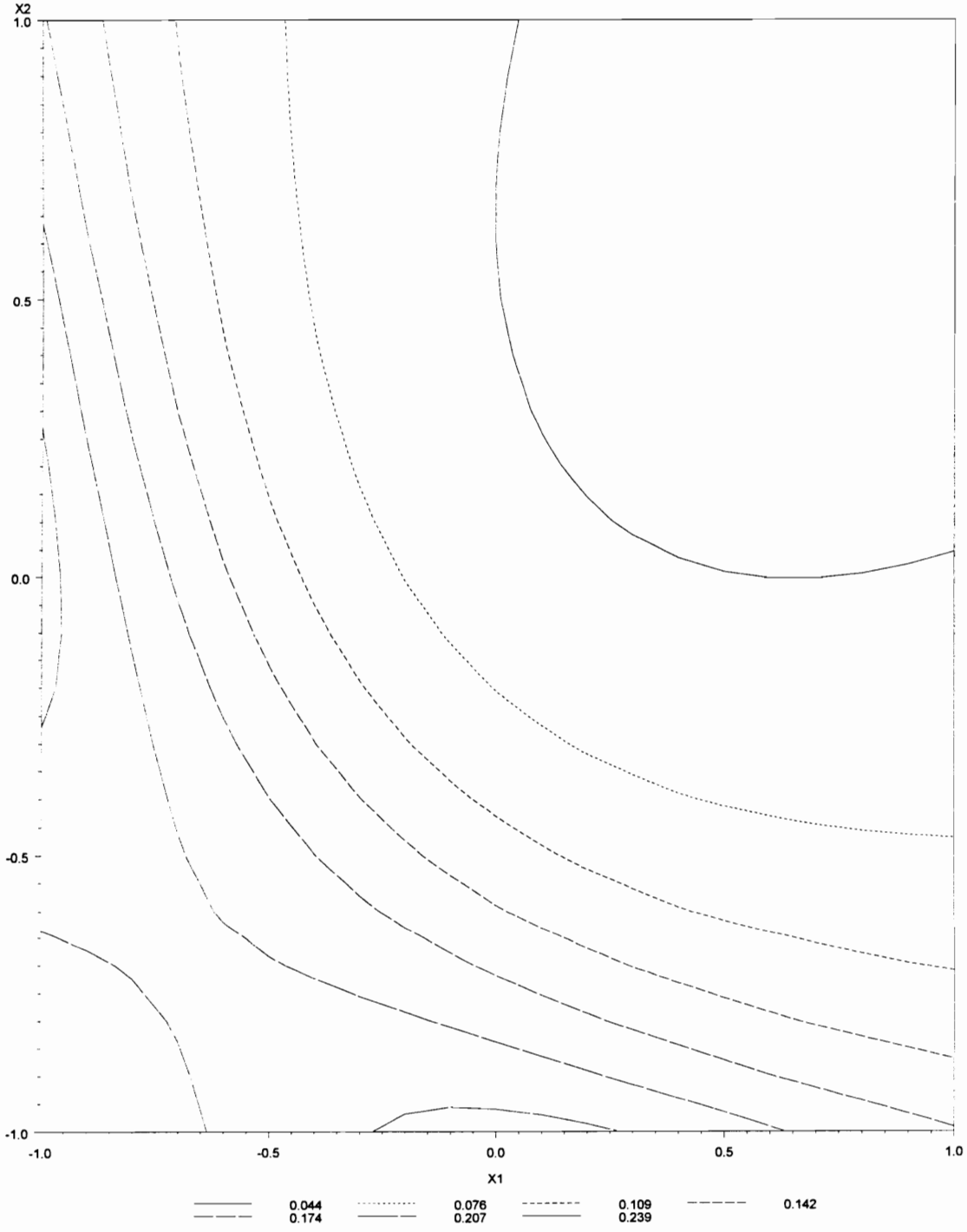



**Figure 3.3** Values of the Correction Factor in Case III.

The following sections introduce various criteria for selecting an appropriate response-model. These criteria will involve each type of estimator: $PV_B$ and $PV_U$. Before these criteria are introduced it is necessary to understand the effect that response-model selection has on the process variance.

### §3.3 Effect of Response-Model Selection on the Process Variance

As discussed in §2.4 the estimated process variance depends heavily on the choice of model terms in the response-model. To illustrate the effect of response-model selection on the process variance consider the following example.

**Example 3.1 (Chemical Data)**

This example involves a chemical data set taken from Myers (1976) (page 85) which contains 2 control factors ($x_1$ and $x_2$) and 2 noise factors ($z_1$ and $z_2$) at two levels each. Based on the data two models for modeling the response, y, against these factors were chosen. Model A was selected on the basis of standard t-tests on the estimated coefficients in the response-model using a liberal significance level of 0.15. Model B is simply a competing model selected using criteria to be discussed in §3.5.

Model A: $\hat{y} = 26.6437 - 9.50625\, z_1 + 8.00625\, x_2 z_1$

Model B: $\hat{y} = 26.6437 - 9.50625\, z_1 + 5.11875\, z_2 + 1.58125\, x_1 z_2 + 8.00625\, x_2 z_1 + 2.80625\, x_2 z_2$

The two response-models are rather different. Model A is much simpler than Model B and, as a result, the process variance models which these response-models yield are also different. To illustrate the differences, the biased estimate of the process variance will be constructed from each response-model. Using (3.2.3) the estimated process variance models are

$PV_B(A) = 249.14865 - 152.21883\, x_2 + 64.10004\, x_2^2$

$$PV_B(B) = 264.46102 + 16.18805\ x_1 - 123.48984\ x_2 + 8.87477\ x_1x_2 + 2.50035\ x_1^2 + 71.97508\ x_2^2$$

A graphical way to investigate the differences is through contour plots. A contour plot was created for each process variance model. See Figures 3.4 and 3.5. According to Model A, in order to minimize the process variance we should operate at the high level of $x_2$ and any level of $x_1$. However, according to Model B, to minimize the process variance we should operate at the high level of $x_2$ and at the low level of $x_1$. There is a definite difference in optimum operating conditions for minimal process variance. Also, notice the differences in the estimated process variances for each model. It is clear that the selection of the response-model greatly affects the optimum operating conditions as well as the estimates of the process variance.

One way of comparing process variance models is to calculate their estimated mean square error (MSE). MSE is a measure of variance and bias which can be expressed as

$$\text{MSE} = \text{variance} + (\text{bias})^2.$$

The bias is due to comparing the process variance models to a more complete model assumed to be "true". This is the approach taken in developments in standard regression. The Cp statistic and other criteria are developed via this approach (see Mallows (1973)). Thus, any competing model would be biased (see Appendix A).

Since the MSE depends on location in $\underline{x}$ space another idea would be to compute the average estimated MSE over the design region. This criterion could be calculated for each process variance model and the selected response-model would be that with the smallest average estimated MSE of the estimated process variance over the design region. This leads to two new criteria for discerning between response-models.

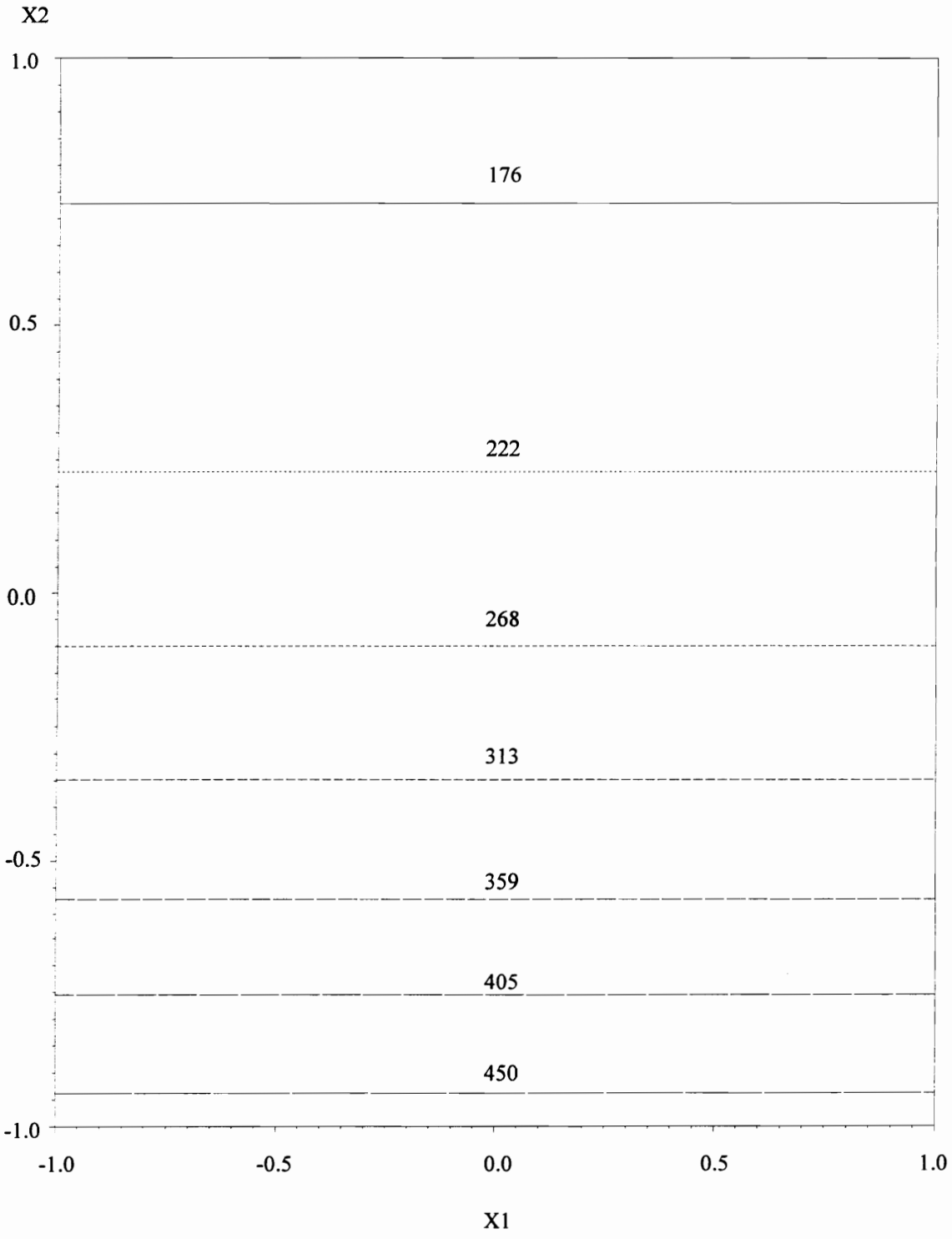



**Figure 3.4** Contour plot of the biased estimate of the process variance resulting from Model A.

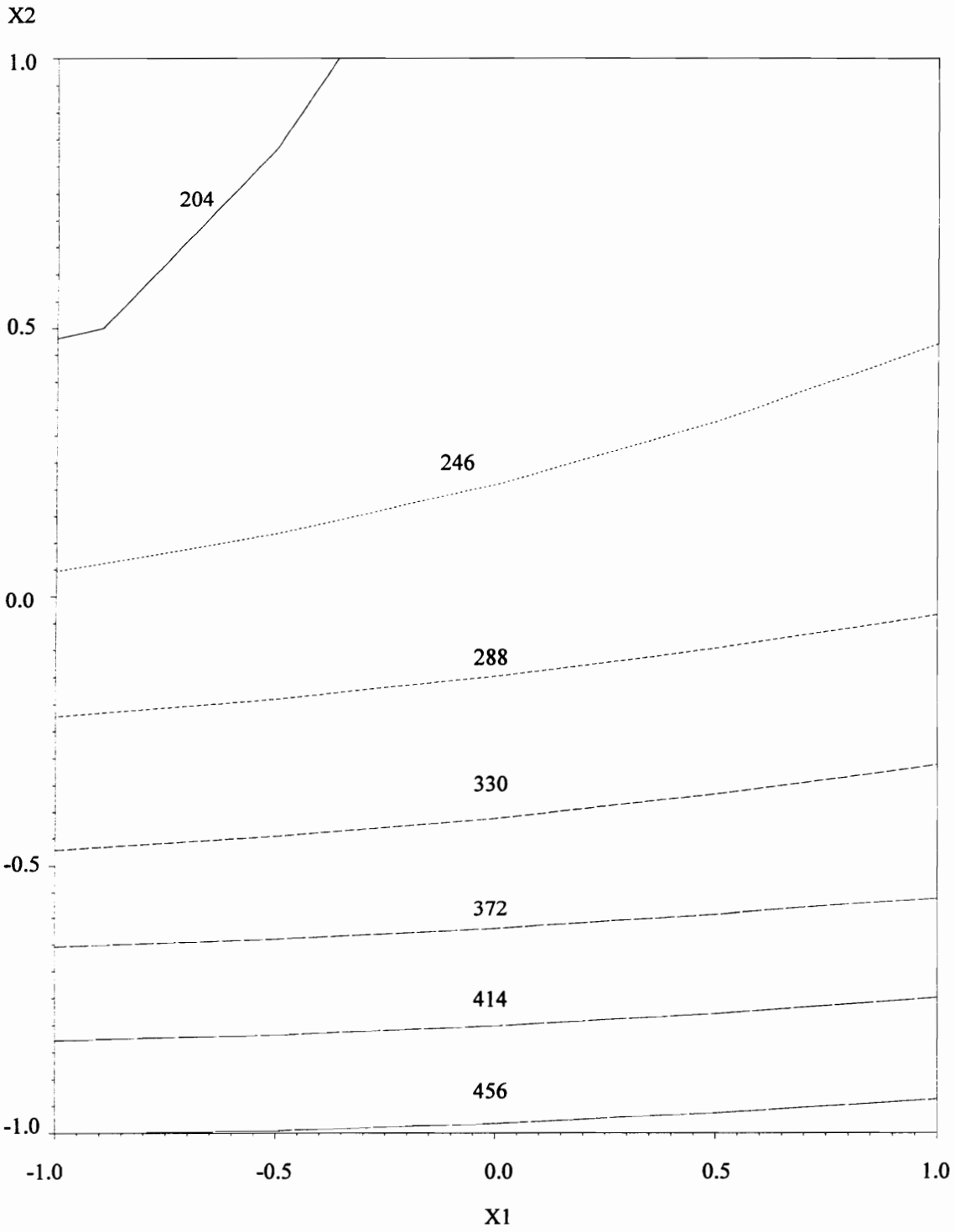



**Figure 3.5** Contour plot of the biased estimate of the process variance resulting from Model B.

## §3.4 Average MSE of the Estimated Process Variance

Two criteria arise when discussing the average MSE for a process variance model:

1. AVGMSE(B) = average MSE of the process variance using the ***biased*** estimator and
2. AVGMSE(U) = average MSE of the process variance using the ***unbiased*** estimator.

Each criterion is calculated in a similar fashion:

$$\begin{aligned} \text{AVGMSE}(\bullet) &= \frac{1}{k\sigma_\epsilon^4} \int_R \text{MSE}(\text{PV}_\bullet)\, d\underline{x} \\ &= \frac{1}{k\sigma_\epsilon^4} \int_R \left\{ \text{Var}(\text{PV}_\bullet) + [\, \text{E}(\text{PV}_\bullet) - \text{Var}_{\underline{z}}(y)\, ]^2 \right\} d\underline{x} \end{aligned}$$

where k = volume of the design region = $\int_R d\underline{x}$ and $\bullet$ refers to the type of estimator, biased (B) or unbiased (U). The final result, with no details, for AVGMSE(B) and AVGMSE(U) can be found in Appendix A. The results for these criteria assume orthogonality among the control by noise interactions and the noise main effects. Although the equations look formidable they are easy to handle using a computer and calculations for several process variance models can be done at one time in only a matter of seconds. Note that these criteria are calculated on the process variance models and not on the response-models.

These criteria have a similarity to the Cp statistic used in regression analysis. The Cp statistic provides a suitable balance between the variance and bias of the model. It is the standardized MSE of prediction summed over all design points. AVGMSE($\bullet$), on the other hand, is the estimated MSE of the estimated process variance model integrated and averaged over the entire design region and then standardized. Both criteria, Cp and AVGMSE($\bullet$), put a high priority on parsimonious models with small bias. As a result, the process variance

model with the smallest AVGMSE( • ) will be chosen. Judging AVGMSE(B) and AVGMSE(U) as variations on the Cp statistic provides an interpretation which is easy to understand.

A traditional way of choosing a model has been to look at the t-tests associated with the estimated coefficients in the model. Insignificant terms would be discarded and the resulting new model would be examined. Contrary to intuition this method is not recommended for selecting a response-model when the goal is to achieve an accurate process variance model. Though experience is limited at this point, it would seem that the use of t-tests, or even Cp statistics, on the response-model is not appropriate for selection of model terms. Experience suggests that the standard regression criteria tend to underspecify, particularly in the area of control by noise interactions.

The two criteria, AVGMSE(B) and AVGMSE(U), provide new methods for selecting response-models which yield accurate process variance models. Another approach, mentioned earlier, suggests minimizing the variability of the estimated slopes of the response-model. Recall Figures 2.1 a) and b). Also, recall Equation (3.1.8). The slopes play a critical role in the process variance model and it is easy to see graphically that the variability in the slopes affects the variability in the response. One way of quantifying this variability is to calculate an average of the estimated mean square error of the estimated slopes. This introduces two more criteria for selecting a response-model.

### §3.5 Average MSE of the Estimated Slopes of the Response-Model

There are two ways of calculating the average estimated mean square error of the slopes. One method involves integrating over the design region and the other method involves summing over the design points. These criteria have been named, respectively, INTMSE(L) and SUMMSE(L) and are calculated using the following equations:

1. $$\text{INTMSE(L)} = \frac{1}{r_z} \sum_{j=1}^{r_z} \frac{\text{N}}{\text{k}\sigma_\epsilon^2} \int_R \text{MSE}(\hat{l}_j)\, \text{d}\underline{\text{x}}$$

2. $\text{SUMMSE(L)} = \frac{1}{\sigma_\epsilon^2 \, r_z} \sum_{j=1}^{r_z} \sum_{\underline{x}} \text{MSE}(\hat{l}_j)$

where N refers to the number of experimental runs.

The latter equation is equivalent to a Cp-like statistic on estimated slopes. Assume an orthogonal experimental design with the "true" response-model being

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}' \, B \, \underline{x} + \underline{z}'\underline{\gamma} + \underline{x}'\Delta \, \underline{z} + \epsilon.$$

Suppose the actual model to be fit is of the form

$$y = \beta_0 + \underline{x}'\underline{\beta}^* + \underline{x}' \, B^* \, \underline{x} + \underline{z}'\underline{\gamma}^* + \underline{x}'\Delta^* \, \underline{z} + \epsilon$$

where $*$ indicates that one or more parameters may be left out of the response-model. It follows then that the true vector of slopes is $\underline{l} = (\underline{\gamma} + \Delta'\underline{x})$ and the estimated vector of slopes is $\hat{\underline{l}} = (\hat{\underline{\gamma}}^* + \hat{\Delta}^{*\prime}\underline{x})$. As a result, each criterion defined above can be rewritten using this notation

$$\text{INTMSE(L)} = \frac{N}{\sigma_\epsilon^2 \, r_z} \left\{ \sum_{j=1}^{r_z} \text{Var}(\hat{\gamma}_j^*) + \frac{1}{3} \sum_{j=1}^{r_z} \sum_{i=1}^{r_x} \text{Var}(\hat{\delta}_{ij}^*) + \sum_{j=1}^{r_z} \left(\gamma_j - \hat{\gamma}_j^*\right)^2 \right.$$

$$\left. + \sum_{j=1}^{r_z} \sum_{i=1}^{r_x} \frac{1}{3} \left(\delta_{ij} - \hat{\delta}_{ij}^*\right)^2 \right\} \qquad (3.5.1)$$

$$\text{SUMMSE(L)} = \frac{N}{\sigma_\epsilon^2 \, r_z} \left\{ \sum_{j=1}^{r_z} \text{Var}(\hat{\gamma}_j^*) + \sum_{j=1}^{r_z} \sum_{i=1}^{r_x} \text{Var}(\hat{\delta}_{ij}^*) + \sum_{j=1}^{r_z} \left(\gamma_j - \hat{\gamma}_j^*\right)^2 \right.$$

$$\left. + \sum_{j=1}^{r_z} \sum_{i=1}^{r_x} \left(\delta_{ij} - \hat{\delta}_{ij}^*\right)^2 \right\} \qquad (3.5.2)$$

where the parameters corresponding to the "true" model would be replaced by their estimates.

These criteria look very similar. Obviously they are, in fact, different. Each criterion is simply a variation on the Cp statistic discussed earlier. Each strives for a balance between the variance and bias of the estimated response-model slopes. Notice that norms, or yardsticks, can be constructed for each criterion. For instance, the norm for the Cp statistic can be calculated using the following form of the statistic

$$\text{Cp} = \text{p} + \frac{(\text{s}^2 - \hat{\sigma}_\epsilon^2)(\text{n} - \text{p})}{\hat{\sigma}_\epsilon^2} .$$

If the model being evaluated is not misspecified, then one expects $s^2$ to be near $\hat{\sigma}_\epsilon^2$ and thus the latter portion of the Cp statistic should be near zero. Thus, a reasonable norm by which to judge a model is Cp = p (Myers (1990)). Similarly, norms, or yardsticks, can be constructed for the criteria presented in (3.5.1) and (3.5.2). For example, the norm for SUMMSE(L) can be derived by noticing the form of the criterion. Recall (3.5.2).

$$\text{SUMMSE(L)} = \frac{\text{N}}{\sigma_\epsilon^2 \, r_z} \left\{ \sum_{j=1}^{r_z} \text{Var}(\hat{\gamma}_j^*) + \sum_{j=1}^{r_z} \sum_{i=1}^{r_x} \text{Var}(\hat{\delta}_{ij}^*) + \sum_{j=1}^{r_z} \left(\gamma_j - \hat{\gamma}_j^*\right)^2 \right.$$

$$\left. + \sum_{j=1}^{r_z} \sum_{i=1}^{r_x} \left(\delta_{ij} - \hat{\delta}_{ij}^*\right)^2 \right\} \qquad (3.5.3)$$

If the model under evaluation is not misspecified, then there will be no bias due to model misspecification. Thus, the last two sums involving estimated coefficients will drop out. All that is left is the sum of variances. Under the conditions of an orthogonal design, each variance equals $\frac{1}{\text{N}}$ (apart from $\sigma_\epsilon^2$). The equation can then be reduced to

$$\begin{aligned} \text{SUMMSE(L)} &= \frac{\text{N}}{\sigma_\epsilon^2 \, r_z} \left\{ \sum_{j=1}^{r_z} \text{Var}(\hat{\gamma}_j^*) + \sum_{j=1}^{r_z} \sum_{i=1}^{r_x} \text{Var}(\hat{\delta}_{ij}^*) \right\} \\ &= \frac{\text{N}}{r_z} \left\{ \text{r}_z \left(\frac{1}{\text{N}}\right) + \text{r}_x \, \text{r}_z \left(\frac{1}{\text{N}}\right) \right\} \\ &= 1 + \text{r}_x \, . \end{aligned} \qquad (3.5.4)$$

Thus, a reasonable yardstick is $(1 + r_x)$. Similarly, the norm for INTMSE(L) can be derived. The result is $(1 + \frac{r_x}{3})$. These norms allow the engineer to gauge a selected response-model with the full response-model just as regression models can be compared using the Cp statistic with its norm of p. However, as in standard regression, the model which is chosen should yield the minimum criterion value. As a result, if one employs $r_x$=4 control variables, then a value for SUMMSE(L) of approximately 5 would suggest a good (though not optimal) model.

These new methods for selecting a response-model based on INTMSE(L) or SUMMSE(L) differ from the traditional method of using t-tests. Recall Example 3.1 which considered two response-models: Model A and Model B. This example actually uses standard t-tests to select Model A and uses INTMSE(L) and SUMMSE(L) to select Model B. Recall there was a difference not only in the estimated process variances but also in the optimum operating conditions. It is therefore critical to accurately estimate the process variability, through the selection of the response-model, when the goal is to accommodate RPD.

### §3.6 Extending the Response-Model

The response-model presented in (3.1.1) is considered a general form which is commonly used. However, this model can be extended to include quadratic noise factors and noise by noise interaction terms. The response-model would then appear as

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'B\,\underline{x} + \underline{z}'\underline{\gamma} + \underline{x}'\Delta\,\underline{z} + \underline{z}'\Psi\,\underline{z} + \epsilon, \qquad (3.6.1)$$

where the matrix $\Psi$ is $r_z$ by $r_z$ and contains $\psi_{jj}$ (j = 1, 2, ..., $r_z$) on the main diagonal and $\frac{1}{2}\psi_{jk}$ (j $\neq$ k) on the (j, k) off-diagonal. These are the coefficients of both the quadratic noise factors and the noise by noise interaction terms, respectively. All other terms are defined as in (3.1.1). If, for the purpose of illustration, we use the assumptions on the noise factors as in §3.1, particularly uncorrelated noise factors with V=I, then the process mean model is given by

$$E_{\underline{z}}(y) = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'B\,\underline{x} + tr(\Psi). \tag{3.6.2}$$

Note that $tr(\Psi)$ is a constant value and not a function of $\underline{x}$. Thus, the addition of quadratic noise factors and noise by noise interaction terms in the response-model does not affect the optimum settings of $\underline{x}$ with respect to the process mean. If we further assume the noise factors follow a normal distribution then the process variance model is given by

$$Var_{\underline{z}}(y) = (\underline{\gamma} + \Delta'\underline{x})'(\underline{\gamma} + \Delta'\underline{x}) + 2\sum_{j=1}^{r_z}\psi_{jj}^2 + \sum\sum_{j<k}\psi_{jk}^2 + \sigma_\epsilon^2. \tag{3.6.3}$$

The assumption of normally distributed noise factors is strictly for the purpose of illustration. A similar expression can be written if we merely assume symmetric distributions of the noise factors. Of course, some distributional assumption must be made. When deriving the process variance the $Var(\underline{z}'\ \Psi\ \underline{z})$ needs to be calculated. This requires knowledge of the third and fourth central moments of each $z_j$ which demands assumptions about the distributions of the $z_j$. Fortunately, the $Var(\underline{z}'\ \Psi\ \underline{z})$ enters the process variance as a constant and not as a function of $\underline{x}$ regardless of the distributional assumption. Thus, the addition of quadratic noise factors and noise by noise interaction terms in the response-model does not affect the optimum settings of $\underline{x}$ with respect to the process variance.

The estimators of the process mean and variance using the extended response-model are quite similar to the estimators discussed in §3.2. The estimator of the process mean is

$$\hat{E}_{\underline{z}}(y) = \hat{\beta}_0 + \underline{x}'\underline{\hat{\beta}} + \underline{x}'\hat{B}\,\underline{x} + tr(\hat{\Psi}). \tag{3.6.4}$$

Again, this estimator is unbiased for the process mean derived from the extended response-model. Also, the biased estimate of the process variance is given by

$$\hat{Var}_{\underline{z}}(y) = (\underline{\hat{\gamma}} + \hat{\Delta}'\underline{x})'(\underline{\hat{\gamma}} + \hat{\Delta}'\underline{x}) + 2\sum_{j=1}^{r_z}\hat{\psi}_{jj}^2 + \sum\sum_{j<k}\hat{\psi}_{jk}^2 + \hat{\sigma}_\epsilon^2. \tag{3.6.5}$$

An unbiased estimator could be formed by subtracting off

$$\mathrm{tr}(\mathrm{C})\hat{\sigma}_{\epsilon}^{2} + 2\sum_{j=1}^{r_z}\widehat{\mathrm{Var}}(\hat{\psi}_{jj}) + \sum_{j<k}\sum\widehat{\mathrm{Var}}(\hat{\psi}_{jk}). \tag{3.6.6}$$

Although the forms of the process mean and variance models are somewhat more complicated, the addition of quadratic noise factors and noise by noise interaction terms in the response-model only affect the process mean and variance as constants. Optimization of the process mean and variance would be unaffected. Note also that the notion of slopes in the $\underline{z}$ - direction is still the dominant concept in the process variance of (3.6.3). An alternative development of (3.6.3) can be derived by calculating the process variance in pieces. The first term in (3.6.3) is identical to the process variance model without the quadratic noise factors or the noise by noise interaction terms. The other pieces of the process variance model are caused by the addition of these extra terms and can be derived using

$$\begin{aligned}\mathrm{Var}_{\underline{z}}(\mathrm{y}^{*}) &= \tfrac{1}{2}\,E_{\underline{z}}\left\{\left(\frac{\partial \mathrm{y}^{*}}{\partial \underline{z}}\right)'\left[\mathrm{Var}(\underline{z})\right]\left(\frac{\partial \mathrm{y}^{*}}{\partial \underline{z}}\right)\right\} \\ &= \tfrac{1}{2}\,E_{\underline{z}}\left\{\left(\frac{\partial \mathrm{y}^{*}}{\partial \underline{z}}\right)'\left(\frac{\partial \mathrm{y}^{*}}{\partial \underline{z}}\right)\right\}\end{aligned}$$

where $\mathrm{y}^{*} = \underline{z}'\ \Psi\ \underline{z}$ and $\mathrm{V}=\mathrm{I}$. Notice the dominant use of slopes. Thus, estimation of the slopes remains of paramount importance if one wishes to estimate the process variance well.

The criteria created for selecting a response-model which would yield an accurate process variance model would only be affected slightly by the extra terms in the extended response-model. AVGMSE(B) and AVGMSE(U), as described in §3.4, would be calculated in a similar fashion, but the form of the computer program would have to be altered to accommodate the extra terms. Otherwise, the basic idea and interpretation of criteria results would remain unchanged. INTMSE(L) and SUMMSE(L) would not be affected by the form of the extended response-model. These two criteria are concerned solely with $\hat{l} = (\hat{\underline{\gamma}} + \hat{\Delta}'\underline{x})$ which does not contain any information from the quadratic noise factors or the noise by noise interaction terms.

# Chapter 4
# Inference on Individual Process Variance Model Terms

## §4.1 Introduction

In multiple linear regression, hypothesis testing can be used to determine if terms found in a regression model are statistically significant. The process variance model can be examined in a similar fashion by assessing the significance of the model terms. However, the concepts involved are much more complicated.

## §4.2 Distributions of the Estimated Coefficients

Recall the process variance of Equation (3.1.8)

$$\mathrm{Var}_{\underline{z}}(\mathrm{y}) = \underline{l}\,'\underline{l} + \sigma_\epsilon^2. \tag{4.2.1}$$

An equivalent and more revealing form of this process variance model is given by

$$\mathrm{Var}_{\underline{z}}(\mathrm{y}) = \sum_{j=1}^{r_z}\gamma_j^2 + \sum_{i=1}^{r_x}\left(\sum_{j=1}^{r_z}2\,\gamma_j\delta_{ij}\right)\mathrm{x}_i + \sum_{i<k}\sum\left(\sum_{j=1}^{r_z}2\,\delta_{ij}\delta_{kj}\right)\mathrm{x}_i\mathrm{x}_k$$

$$+ \sum_{i=1}^{r_x}\left(\sum_{j=1}^{r_z}\delta_{ij}^2\right)\mathrm{x}_i^2 + \sigma_\epsilon^2 \tag{4.2.2}$$

Notice that the coefficients involve either squares or cross-products of coefficients related to noise factors found in the response-model. This process variance model (4.2.2) can be estimated by replacing the parameters with their estimates. In what follows the distributions of the estimated coefficients will be investigated.

The coefficients of (4.2.2) are of two types: sums of squares and sums of

cross-products. The first distribution to be discussed deals with the sum of squares. The coefficients of interest are the intercept and the coefficients of the squared control factors. Recall that the assumption of normal errors in the response-model of (3.1.1) allows us to assume that the distribution of the estimated coefficients in the response-model also follow a normal distribution. If the true mean of each normally distributed variable is constrained to equal zero (clearly the null distribution if variable screening were carried out) then the resulting distribution of the sum of squared, independent and normally distributed random variables is proportional to a central chi-square with degrees of freedom equal to $r_z$. In other words,

$$\sum_{j=1}^{r_z} \hat{\gamma}_j^2 \sim \sigma_\epsilon^2 \, \mathrm{Var}(\hat{\gamma}_j) \, \chi^2_{r_z} \quad \text{and} \quad \left( \sum_{j=1}^{r_z} \hat{\delta}_{ij}^2 \right) \sim \sigma_\epsilon^2 \, \mathrm{Var}(\hat{\delta}_{ij}) \, \chi^2_{r_z} \qquad (4.2.3)$$

under the hypothesis that $\mathrm{E}(\hat{\gamma}_j) = 0$ and $\mathrm{E}(\hat{\delta}_{ij}) = 0$ where $\mathrm{Var}(\hat{\gamma}_j)$ and $\mathrm{Var}(\hat{\delta}_{ij})$ are taken to be apart from $\sigma_\epsilon^2$. Since $\sigma_\epsilon^2$ is unknown it is necessary to use an F distribution such as

$$\frac{\sum_{j=1}^{r_z} \hat{\delta}_{ij}^2}{\mathrm{r}_z \, \mathrm{Var}(\hat{\delta}_{ij}) \, \sigma_\epsilon^2} \Big/ \frac{\hat{\sigma}_\epsilon^2}{\sigma_\epsilon^2} = \frac{\chi^2_{\mathrm{r}_z}/\mathrm{r}_z}{\chi^2_{dfE}/\mathrm{dfE}} = \mathrm{F}_{(\mathrm{r}_z,\, dfE)} \qquad (4.2.4)$$

where dfE refers to the error degrees of freedom. This distribution allows us to make inferences about $\sum_{j=1}^{r_z} \hat{\delta}_{ij}^2$ and, similarly, $\sum_{j=1}^{r_z} \hat{\gamma}_j^2$.

The next type of distribution to be discussed is more complicated. The estimated coefficients of the linear control factors and of the control by control interaction terms are of interest now. Thus, we need the distributions of

$$\left( \sum_{j=1}^{r_z} 2 \, \hat{\gamma}_j \hat{\delta}_{ij} \right) \quad \text{and} \quad \left( \sum_{j=1}^{r_z} 2 \, \hat{\delta}_{ij} \hat{\delta}_{kj} \right). \qquad (4.2.5)$$

The distribution of the sum of cross-products of independent and normally distributed random variables is essential.

For the purpose of illustration let $q_{ij} = 2\ \hat{\gamma}_j \hat{\delta}_{ij}$ and let $Q_i = \sum_{j=1}^{r_z} q_{ij}$ where $\hat{\gamma}_j \sim N(0, \sigma^2)$ and $\hat{\delta}_{ij} \sim N(0, \sigma^2)$ for all i, j. The method taken to derive the distribution of $Q_i$ will make use of the characteristic function (C. F.) of $q_{ij}$. The first step is to derive the density function of $q_{ij}$. Dr. I. J. Good had researched a similar problem years ago and aided in the following development. Using the assumptions of independence and normality along with rules of probability for deriving the distribution function, the density function of $q_{ij}$ is given by

$$f_{q_{ij}}(q_{ij}) = \int_{-\infty}^{\infty} \frac{1}{|2\hat{\gamma}_j|} f(\hat{\gamma}_j, \frac{q_{ij}}{2\hat{\gamma}_j})\ d\hat{\gamma}_j\ ,\ -\infty < q_{ij} < \infty.$$

$$= 2 \int_0^{\infty} \frac{1}{2\hat{\gamma}_j} \left( \frac{1}{\sqrt{2\pi}\ \sigma} \exp\left\{ - \frac{\hat{\gamma}_j^2}{2\sigma^2} \right\} \right) \left( \frac{1}{\sqrt{2\pi}\ \sigma} \exp\left\{ - \frac{(q_{ij}/2\hat{\gamma}_j)^2}{2\sigma^2} \right\} \right) d\hat{\gamma}_j$$

$$= \frac{1}{2\pi\sigma^2} \int_0^{\infty} \frac{1}{\hat{\gamma}_j} \exp\left\{ - \frac{(\hat{\gamma}_j^2 + \frac{q_{ij}^2}{4\hat{\gamma}_j^2})}{2\sigma^2} \right\} d\hat{\gamma}_j \qquad (4.2.6)$$

Since this integration is not trivial the computer package *Mathematica* was used to carry out the integration. The final result for the density function of $q_{ij}$ is

$$f_{q_{ij}}(q_{ij}) = \frac{1}{2\pi\sigma^2} \text{BesselK}[0, \frac{\sqrt{q_{ij}^2}}{2\sigma^2}]\ ,\ -\infty < q_{ij} < \infty \qquad (4.2.7)$$

(Gray and Mathews, 1931, Farrell and Ross, 1963, Johnson and Kotz, 1970b-c). This density function can be used to obtain the C. F. of $q_{ij}$. It is given by

$$\text{C. F. } (q_{ij};\ t) = \int_{-\infty}^{\infty} e^{itq_{ij}} f_{q_{ij}}(q_{ij})\ d\hat{\gamma}_j\ ,\ -\infty < q_{ij} < \infty. \qquad (4.2.8)$$

With the aid of *Mathematica*, the characteristic function can be expressed as

$$\text{C. F. } (q_{ij};\ t) = \frac{1}{\sqrt{1 + 4\sigma^4 t^2}}\ . \qquad (4.2.9)$$

This characteristic function can be used to find the density function of $Q_i$ by making use of

$$\text{C. F. } (Q_i = \sum_{j=1}^{r_z} q_{ij} \; ; \; t) = \prod_{j=1}^{r_z} \text{C. F. } (q_{ij} \; ; \; t)$$

$$= \left( \frac{1}{\sqrt{1 + 4\sigma^4 t^2}} \right)^{r_z}$$

$$= \phi_{q_{ij}}(t) \tag{4.2.10}$$

and $$\text{pdf}\left( \sum_{j=1}^{r_z} q_{ij} \; ; \; Q_i \right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (e^{-itq_{ij}}) \; \phi_{q_{ij}}(t) \; dt. \tag{4.2.11}$$

The distribution of the sum of cross-products of independent and normally distributed random variables can be determined by integrating the above expression appropriately.

In summary, Equation (4.2.11) provides the means for determining the distributions of

$$\left( \sum_{j=1}^{r_z} 2 \; \hat{\gamma}_j \hat{\delta}_{ij} \right) \quad \text{and} \quad \left( \sum_{j=1}^{r_z} 2 \; \hat{\delta}_{ij} \hat{\delta}_{kj} \right) \tag{4.2.12}$$

under the hypothesis that $E(\hat{\gamma}_j) = 0$ and $E(\hat{\delta}_{ij}) = 0$.

The purpose of determining the two types of distributions is to find a method for assessing significance of process variance model terms through the use of significance testing and probability values. The next section discusses how to calculate p-values.

### §4.3 Procedure for Calculating p-values

There are two types of estimated coefficients involved in the process variance model and the previous section discussed methods for determining their distributions. These distributions can be used to calculate the significance of the estimated coefficients through the use of p-values. The first case to be discussed involves the sums of squared independent normally distributed random variables. Namely,

$$\sum_{j=1}^{r_z} \hat{\gamma}_j^2 \sim r_z \, \mathrm{Var}(\hat{\gamma}_j) \, \hat{\sigma}_\epsilon^2 \, F_{(r_z,\, dfE)}$$

and

$$\sum_{j=1}^{r_z} \hat{\delta}_{ij}^2 \sim r_z \, \mathrm{Var}(\hat{\delta}_{ij}) \, \hat{\sigma}_\epsilon^2 \, F_{(r_z,\, dfE)} \,. \tag{4.3.1}$$

The appropriate null hypotheses from which p-values are derived are

$$H_0: \sum_{j=1}^{r_z} \gamma_j^2 = 0 \quad \text{and} \quad H_0: \left( \sum_{j=1}^{r_z} \delta_{ij}^2 \right) = 0 \tag{4.3.2}$$

or equivalently,

$$H_0: \gamma_j = 0 \ (\forall \ j=1, 2, \ldots, r_z) \quad \text{and} \quad H_0: \delta_{ij} = 0 \ (\forall \ j=1, 2, \ldots, r_z) \,. \tag{4.3.3}$$

Under these conditions, the distributions given by (4.3.1) are correct. These distributions can be used to calculate p-values using the following probability statements:

$$P\Big( \sum_{j=1}^{r_z} \hat{\gamma}_j^2 > \text{observed intercept} \Big) = P\Big(F_{(r_z,\, dfE)} > \frac{\text{observed intercept}}{r_z \, \mathrm{Var}(\hat{\gamma}_j) \, \hat{\sigma}_\epsilon^2} \Big)$$

$$P\Big( \sum_{j=1}^{r_z} \hat{\delta}_{ij}^2 > \text{observed coef of } x_i^2 \Big) = P\Big(F_{(r_z,\, dfE)} > \frac{\text{observed coef of } x_i^2}{r_z \, \mathrm{Var}(\hat{\delta}_{ij}) \, \hat{\sigma}_\epsilon^2}\Big) \tag{4.3.4}$$

where $\mathrm{Var}(\hat{\gamma}_j)$ and $\mathrm{Var}(\hat{\delta}_{ij})$ are apart from $\sigma_\epsilon^2$.

The second type of estimated coefficients involves the sum of cross-products of independent and normally distributed random variables.

$$\left( \sum_{j=1}^{r_z} 2 \, \hat{\gamma}_j \hat{\delta}_{ij} \right) \quad \text{and} \quad \left( \sum_{j=1}^{r_z} 2 \, \hat{\delta}_{ij} \hat{\delta}_{kj} \right) \tag{4.3.5}$$

Recall the general density function derived in the previous section

$$\mathrm{pdf}\Big( \sum_{j=1}^{r_z} q_{ij} \, ; \, Q_i \Big) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itq_{ij}} \, \phi_{q_{ij}}(t) \, dt, \tag{4.3.6}$$

where, for example, $Q_i = \sum_{j=1}^{r_z} 2\hat{\gamma}_j\hat{\delta}_{ij}$ and the means of the random variables $\hat{\gamma}_j$ and $\hat{\delta}_{ij}$ were assumed to be zero. Thus, the "null hypothesis" associated with the density function of (4.3.6) can be either

$$\begin{aligned} &H_0\text{: } \gamma_j = 0 \text{ and } \delta_{ij} = 0 \ (\forall \text{ j=1, 2, ... , } r_z) \\ \text{or} \quad &H_0\text{: } \delta_{ij} = 0 \text{ and } \delta_{kj} = 0 \ (\forall \text{ j=1, 2, ... , } r_z) \ . \end{aligned} \tag{4.3.7}$$

However, there are other ways to render

$$\left( \sum_{j=1}^{r_z} 2\ \gamma_j\delta_{ij} \right) = 0 \quad \text{and} \quad \left( \sum_{j=1}^{r_z} 2\ \delta_{ij}\delta_{kj} \right) = 0 \ . \tag{4.3.8}$$

Unfortunately, it would be quite a difficult task to determine the distributions of (4.3.5) under all possible ways of satisfying (4.3.8). Information about the true mean of one estimated coefficient in each product would be necessary to calculate the distributions of (4.3.5). This was attempted for a simple case involving only one product. Using the procedure of (4.2.6) the calculation of the density function of $2\ \hat{\gamma}_j\hat{\delta}_{ij}$, under the assumptions that $E(\hat{\gamma}_j) = 0$ and $E(\hat{\delta}_{ij})$=constant (other than 0), was attempted with the aid of *Mathematica*. However, this package was not powerful enough to solve the integral. Furthermore, knowledge of true means of estimated coefficients is unlikely even if the integral could be calculated. This is the reason why a specific case of zero means was undertaken to illustrate how to calculate p-values. Interpretation of these p-values will be discussed in the next section.

Calculating p-values for the estimated coefficients of (4.3.5) under the conditions of (4.3.7) can be accomplished by

$$P\left(\sum_{j=1}^{r_z} 2\hat{\gamma}_j\hat{\delta}_{ij} \ > \ | \ Q_i = \text{observed coef of } x_i|\right) = 2\int_{Q_i}^{\infty}\left(\frac{1}{2\pi}\int_{\infty}^{\infty} e^{-iut}\phi_u(t)\ dt\right)du$$

and

$$P\left(\sum_{j=1}^{r_z} 2\hat{\delta}_{ij}\hat{\delta}_{kj} \ > \ | \ Q_{ik} = \text{observed coef of } x_ix_k|\right) = 2\int_{Q_i}^{\infty}\left(\frac{1}{2\pi}\int_{\infty}^{\infty} e^{-iut}\phi_u(t)\ dt\right)du. \tag{4.3.9}$$

The absolute value signs within the probability statements are valid since the BesselK function is symmetric about zero.

Equations (4.3.4) and (4.3.9) supply the means by which p-values can be calculated for all estimated coefficients found in the estimated process variance model. The next section discusses how to interpret these p-values.

## §4.4 Interpreting p-values

The p-values calculated for the estimated coefficients having a chi-square distribution can be interpreted in the standard fashion. Interpretation of the p-values calculated for the estimated coefficients having a BesselK distribution is much more involved. Recall (4.3.7). While the conditions under which the distributions were derived were specific, some information can be gained by the associated p-values.

If a p-value is calculated that is insignificant in the eyes of the practitioner, then it is reasonable to assume that the estimated coefficient is not significantly different from zero. For example, consider the case of the estimated coefficient of $x_i$, denoted by $\sum_{j=1}^{r_z} \hat{\gamma}_j \hat{\delta}_{ij}$. If each piece of this sum is not significantly different from zero, it seems reasonable to conclude that the sum is not significantly different from zero. However, if a calculated p-value is significant, then the interpretation is not as clear. It is quite possible for each piece of the sum to be statistically significant while the true sum is zero by cancellation. As a result, the practitioner should use these p-values as diagnostics.

# Chapter 5
# Analysis of "Residuals" for the Process Variance Model

## §5.1 Introduction

In multiple linear regression, model misspecification and outlying observations can be detected through the use of various diagnostics. Many of these diagnostics make use of residuals based on the observed response, y, and the predicted response, $\hat{y}$. The basic idea is that the difference between what is observed and what is predicted should be relatively small when conditions are ideal, in which case $E[y - \hat{y}]=0$. This type of idea for the use of residuals can be extended to the case of detecting nonideal conditions such as model misspecification in the estimated process variance model. However, as in the case of all forms of inference, the situation is more complicated. The difference lies in how the residuals are formulated.

## §5.2 Relationship between $PV_B$ and $s^2(\underline{x})$

As discussed in §3.2, $PV_B$ or $PV_U$ provides the analyst with a methodology to predict the process variance at any setting in the control factors. Another statistic, the sample variance, $s^2(\underline{x})$, at a particular location in $\underline{x}$, provides observed information about the process variance. This sample variance is taken over the levels of the noise factors. A closer look at these two statistics will reveal some similarities. Let us assume at this point that the design is a combined array with each noise factor having 2 levels.

First, the expected value of $s^2(\underline{x})$ is equal to the population process variance apart from a small constant. This can be shown by first examining the distribution of $s^2(\underline{x})$. Details of the derivation of the distribution can be found in Appendix §B.1.

$$E_y\left\{s^2(\underline{x})\right\} = E_y\left\{\frac{\sigma_\epsilon^2}{2^{r_z}}\chi^{2'}_{(2^{r_z}-1,\ \frac{2^{r_z}\,\underline{l}'\underline{l}}{\sigma_\epsilon^2})}\right\}$$

$$= \frac{\sigma_\epsilon^2}{2^{r_z}}\left\{2^{r_z} - 1 + \frac{2^{r_z}\,\underline{l}'\underline{l}}{\sigma_\epsilon^2}\right\}$$

$$= \underline{l}'\underline{l} + \sigma_\epsilon^2(1 - \frac{1}{2^{r_z}}) \,. \qquad (5.2.1)$$

This holds true for any response-model being fit. Thus, an unbiased estimate of the population process variance can be given by

$$s^2(\underline{x}) + \frac{\widehat{\sigma}_\epsilon^2}{2^{r_z}} \,. \qquad (5.2.2)$$

Another similarity is that $s^2(\underline{x})$ is exactly equal to $PV_B$ when the response-model is saturated. An algebraic illustration is seen with a small example that appears in Appendix §B.2. This should certainly be intuitive to the reader.

Another way of viewing $s^2(\underline{x})$ and $PV_B$ is that of y and $\hat{y}$, respectively, in multiple linear regression. Namely, $s^2(\underline{x})$ acts as the "observed" process variance and, for a given response-model, $PV_B$ acts as the "predicted" process variance. These two statistics can be combined in a way that is very much like a residual to detect the presence of outliers or model misspecification.

### §5.3 Use of Residuals

The statistics to be used for creating a residual are $s^2(\underline{x}) + \frac{\widehat{\sigma}_\epsilon^2}{2^{r_z}}$ and $PV_B$. These two statistics are estimating the same process variance. However, before a residual is constructed it would be wise to examine the distributions of these two statistics. It is important that the residual be constructed in such a way that diagnostic yardsticks are available much like those that are used in standard regression.

From Appendix §B.1 and (5.2.1) the true distribution of $s^2(\underline{x}) + \frac{\widehat{\sigma}_\epsilon^2}{2^{r_z}}$ is

$$\left(\frac{\sigma_\epsilon^2}{2^{r_z}}\,\chi^{2'}_{(2^{r_z}\text{ - }1,\ \frac{2^{r_z}\,\underline{l}'\underline{l}}{\sigma_\epsilon^2})}\right) + \left(\frac{\sigma_\epsilon^2}{2^{r_z}\ \mathrm{dfE}}\,\chi^2_{dfE}\right)$$

$$= \frac{\sigma_\epsilon^2}{2^{r_z}}\left(\chi^{2'}_{(2^{r_z}\text{ - }1,\ \frac{2^{r_z}\,\underline{l}'\underline{l}}{\sigma_\epsilon^2})} + \frac{1}{\mathrm{dfE}}\,\chi^2_{dfE}\right). \tag{5.3.1}$$

It is of interest to express the quantity in (5.3.1) as a single central chi-square distribution. One needs a function of $PV_B$ and $s^2(\underline{x})$ for which a reasonable yardstick can be computed for diagnostic purposes. This will involve the use of approximate distributions.

The first approximate distribution needed is that of the distribution of $s^2(\underline{x})$. Namely,

$$\chi^{2'}_{(2^{r_z}\text{ - }1,\ \frac{2^{r_z}\,\underline{l}'\underline{l}}{\sigma_\epsilon^2})} \sim K^*\,\chi^2_{D^*}$$

$$\text{where } K^* = \frac{2^{r_z}\text{ - }1 + \frac{2(2^{r_z})\,\underline{l}'\underline{l}}{\sigma_\epsilon^2}}{2^{r_z}\text{ - }1 + \frac{2^{r_z}\,\underline{l}'\underline{l}}{\sigma_\epsilon^2}} \quad \text{and} \quad D^* = \frac{[2^{r_z}\text{ - }1 + \frac{2^{r_z}\,\underline{l}'\underline{l}}{\sigma_\epsilon^2}]^2}{2^{r_z}\text{ - }1 + \frac{2(2^{r_z})\,\underline{l}'\underline{l}}{\sigma_\epsilon^2}}. \tag{5.3.2}$$

This approximation is due to Patnaik (1949). It allows a non-central chi-square distribution to be approximated by a central chi-square distribution.

Thus, the approximate distribution of $s^2(\underline{x}) + \frac{\widehat{\sigma}_\epsilon^2}{2^{r_z}}$ is given by

$$\frac{\sigma_\epsilon^2}{2^{r_z}}\left(K^*\,\chi^2_{D^*} + \frac{1}{\mathrm{dfE}}\,\chi^2_{dfE}\right). \tag{5.3.3}$$

This sum can be approximated further by a single central chi-square distribution using results due to Satterthwaite (1946) and Welch (1956). Thus, the practical approximate distribution is given by

$$s^2(\underline{x}) + \frac{\hat{\sigma}_\epsilon^2}{2^{r_z}} \sim \frac{\sigma_\epsilon^2}{2^{r_z}} K \chi^2_D$$

$$\text{where } K = \frac{K^{*2}D^* + \left(\frac{1}{dfE}\right)^2 dfE}{K^*D^* + \left(\frac{1}{dfE}\right) dfE} \qquad \text{and} \qquad D = \frac{\left(K^*D^* + \left(\frac{1}{dfE}\right) dfE\right)^2}{K^{*2}D^* + \left(\frac{1}{dfE}\right)^2 dfE}. \quad (5.3.4)$$

To develop the distribution of $PV_B$ the distribution of $\hat{\underline{l}}'\hat{\underline{l}}$ must first be determined. It has been stated earlier that $\hat{\underline{l}} \sim N(\underline{l}, \sigma_\epsilon^2 C)$. Thus, $\hat{l}_j \sim N(l_j, \sigma_\epsilon^2 c_j)$ where $c_j = Var(\hat{l}_j)$, apart from $\sigma_\epsilon^2$. Then

$$\hat{\underline{l}}'\hat{\underline{l}} = \sum_{j=1}^{r_z} \hat{l}_j^2 \sim \sum_{j=1}^{r_z} \chi^{2'}(1, \lambda_j)\, \sigma_\epsilon^2 c_j \quad (5.3.5)$$

where $\lambda_j = \frac{l_j^2}{\sigma_\epsilon^2 c_j}$. The noncentral distribution of Equation (5.3.5) can be approximated by a linear combination of central chi squares:

$$\hat{\underline{l}}'\hat{\underline{l}} \sim \sum_{j=1}^{r_z} \sigma_\epsilon^2 c_j \rho_j \chi^2_{\nu_j} \quad (5.3.6)$$

where $\rho_j = \frac{1 + 2\lambda_j}{1 + \lambda_j}$ and $\nu_j = \frac{(1 + \lambda_j)^2}{1 + 2\lambda_j}$. This approximation is due to Patnaik (1949). We can further approximate this linear combination of central chi squares with

$$\hat{\underline{l}}'\hat{\underline{l}} \sim \sigma_\epsilon^2 d\, \chi^2_\nu \quad (5.3.7)$$

$$\text{where } d = \frac{\sum_{j=1}^{r_z} (c_j\rho_j)^2\nu_j}{\sum_{j=1}^{r_z} c_j\rho_j\nu_j} \quad \text{and} \quad \nu = \frac{\left(\sum_{j=1}^{r_z} \sigma_\epsilon^2 c_j\rho_j\nu_j\right)^2}{\sum_{j=1}^{r_z} (\sigma_\epsilon^2 c_j\rho_j)^2\nu_j}. \text{ This approximation is due}$$

to Satterthwaite (1946) and Welch (1956). Thus, we have an approximate distribution of $\hat{\underline{l}}'\hat{\underline{l}}$. It is also true that $\hat{\sigma}_\epsilon^2 \sim \frac{\sigma_\epsilon^2}{\mathrm{dfE}}\chi^2_{dfE}$. So we can approximate $\hat{\underline{l}}'\hat{\underline{l}} + \hat{\sigma}_\epsilon^2$ by

$$\sigma_\epsilon^2 \left\{ \mathrm{d}\chi^2{}_\nu + \frac{1}{\mathrm{dfE}}\chi^2_{dfE} \right\} \sim \sigma_\epsilon^2 \, \mathrm{e}\chi^2{}_{\nu^*}$$

$$\text{where} \quad \mathrm{e} = \frac{\mathrm{d}^2\nu + \left\{\frac{1}{\mathrm{dfE}}\right\}^2 \mathrm{dfE}}{\mathrm{d}\nu + \left\{\frac{1}{\mathrm{dfE}}\right\}\mathrm{dfE}} \quad \text{and} \quad \nu^* = \frac{\left\{\mathrm{d}\nu + \left\{\frac{1}{\mathrm{dfE}}\,\mathrm{dfE}\right\}\right\}^2}{\mathrm{d}^2\nu + \left\{\frac{1}{\mathrm{dfE}}\right\}^2 \mathrm{dfE}} \,. \tag{5.3.8}$$

In summary, the approximate distribution of the statistics to be used are:

$$\mathrm{s}^2(\underline{\mathrm{x}}) + \frac{\hat{\sigma}_\epsilon^2}{2^{r_z}} \sim \frac{\sigma_\epsilon^2}{2^{r_z}} \mathrm{K}\, \chi^2\mathrm{D}$$

$$\mathrm{PV}_B \sim \sigma_\epsilon^2 \, \mathrm{e}\chi^2{}_{\nu^*} \,. \tag{5.3.9}$$

These distributions will be used to find the appropriate form of the residual.

Bartlett and Kendall (1946) discussed the use of the natural log transformation on chi-square variates. The transformation results in an approximately normal variate. This approximation improves as the degrees of freedom increase. Fortunately, it is reasonably close for degrees of freedom greater than 5. The degrees of freedom in the chi-square statistics of (5.3.9) are dependent upon unknown parameter values. However, since the residual to be formed will only be used as a diagnostic tool and not for formal hypothesis testing exact normality of ln $\chi^2$ is not necessary. Limited experience has shown that the degrees of freedom are generally much larger than 30.

As a result, it is reasonable to use

$$\ln\left(\frac{\mathrm{s}^2(\underline{\mathrm{x}}) + \frac{\hat{\sigma}_\epsilon^2}{2^{r_z}}}{\mathrm{PV}_B}\right) = \ln\left(\mathrm{s}^2(\underline{\mathrm{x}}) + \frac{\hat{\sigma}_\epsilon^2}{2^{r_z}}\right) - \ln(\mathrm{PV}_B) \tag{5.3.10}$$

as a residual. It has been established previously that each statistic has an asymptotically chi-square distribution. This log ratio can be used to detect process variance model misspecification just as the residual (y - $\hat{y}$) of multiple linear regression detects misspecification. Under ideal conditions the above "residual" has expectation approximately zero. However, if the process variance model is misspecified, then the predicted process variance at a given setting in $\underline{x}$ will be different from the observed process variance given by the statistic involving $s^2(\underline{x})$. This dissimilarity will be detected through the log ratio.

One drawback of the residual is that it is not scale-free. This can be countered by standardizing the residual along the lines of studentized residuals in regression analysis. Derivation of the standard deviation of the log ratio requires the variances of the natural log of each of the statistics as well as the covariance between them. Asymptotic variances can be approximated through the use of a Taylor's Series expansion around the mean (see, for example, Crowder and Hamilton (1992)). Again, since the purpose of the residual is purely diagnostic, an approximately standardized residual is acceptable. The approximate standard deviation of the log ratio is given by

$$\sqrt{2\left\{\frac{1}{D}+\frac{1}{\nu^*}+\frac{\ln(\sigma_\epsilon^2\, e\, \nu^*)}{D}+\frac{\ln\left(\frac{\sigma_\epsilon^2\, K\, D}{2^{r_z}}\right)}{\nu^*}\right\}}. \tag{5.3.11}$$

Thus, the form of the standardized residual is given by

$$R = \frac{\ln\left(\frac{s^2(\underline{x})+\frac{\hat{\sigma}_\epsilon^2}{2^{r_z}}}{PV_B}\right)}{\sqrt{2\left\{\frac{1}{D}+\frac{1}{\nu^*}+\frac{\ln(\sigma_\epsilon^2\, e\, \nu^*)}{D}+\frac{\ln\left(\frac{\sigma_\epsilon^2\, K\, D}{2^{r_z}}\right)}{\nu^*}\right\}}}. \tag{5.3.12}$$

The standardized residual, R, is simple to compute and can be used to detect model misspecification. Graphing the standardized residual against the predicted process variance (or its natural log transformation) will depict the extent of model misspecification for various predicted values of the process variance. This technique is used in multiple linear regression where plots of residuals against $\hat{y}$ are examined to detect model misspecification. Since the standardized residual is approximately normal with a mean of zero and a variance of 1 it is reasonable to set upper and lower limits on these graphs. Reasonable "red flag" limits would be $\pm 2$. Values of the standardized residual falling outside of these limits would indicate process variance model misspecification or a variance inflation outlier. The following example will illustrate the use of the standardized residual.

**Example 5.1**

Consider an experiment which involves 5 control factors and 2 noise factors at two levels each. The experimental design used in this situation was a combined array involving all factors and all levels. It required 128 experimental runs. The data was simulated using a response-model with selected parameter values and normal error. The simulated responses were then used to fit several competing response-models. These models ranged from the "true" response-model to varying degrees of misspecified response-models. The fitted "true" response-model is given by

$$\begin{aligned}\hat{y} = \hat{E}_{\underline{z}}(\hat{y}) &+ 0.633\, z_1 + 1.356\, z_2 + 0.641\, x_1z_1 + 1.290\, x_1z_2 + 0.728\, x_2z_1 \\ &+ 1.323\, x_2z_2 + 1.396\, x_3z_1 + 1.223\, x_3z_2 + 1.918\, x_4z_1 + 1.369\, x_4z_2 \\ &+ 2.631\, x_5z_1 + 1.391\, x_5z_2 \qquad (5.3.13)\end{aligned}$$

where $\sigma_\epsilon^2 = (0.75)^2$ and $\hat{E}_{\underline{z}}(\hat{y})$ is the estimated process mean model. The $\hat{E}_{\underline{z}}(\hat{y})$ in (5.3.13) involves only the control factors and thus has no effect on the process variance model. The competing response-models left out one term for the purpose of examining the process variance model misspecification. The fitted response-

models were then used to create their respective process variance models using the form of $PV_B$. The standardized residuals were then calculated and plotted against the natural log of the predicted process variance for each of the response-models.

The following cases represent the various response-models and Figures 5.1 - 5.4 represent plots of the standardized residuals against the natural log of the predicted process variance for each of these cases. Recall that process variance model misspecification is diagnosed when the value of the standardized residual falls outside the limits of $\pm 2$.

**CASE I:** (see Figure 5.1)

"True" response-model.

**CASE II:** (see Figure 5.2)

One noise main effect, $z_1$, is left out of the response-model. The process variance model misspecification caused by the absence of this term is considered to be mild.

**CASE III:** (see Figure 5.3)

One control by noise interaction term, $x_1z_2$, is left out of the response-model. The process variance model misspecification caused by the absence of this term is considered to be medium.

**CASE IV:** (see Figure 5.4)

One control by noise interaction term, $x_4z_1$, is left out of the response-model. The process variance model misspecification caused by the absence of this term is considered to be severe.

In summary, the plots adequately display the effect of the process variance model misspecification for mild to severe cases. The purpose of these plots is to diagnose the problem. Correcting the misspecification amounts to examining competing response-models and selecting the model with the smallest amount of process variance model misspecification.

### §5.4 Extending RSM Analysis

The discussion to this point has involved several new methods for selecting and evaluating an appropriate response-model. If, in fact, the response-model is already known or at least well-estimated based on the engineer's prior knowledge, then sampling properties of various RSM statistics can be calculated which take advantage of this information. For example, confidence regions on the true process variance would be of interest to the engineer. Important items such as prediction intervals and tolerance intervals can also be computed.

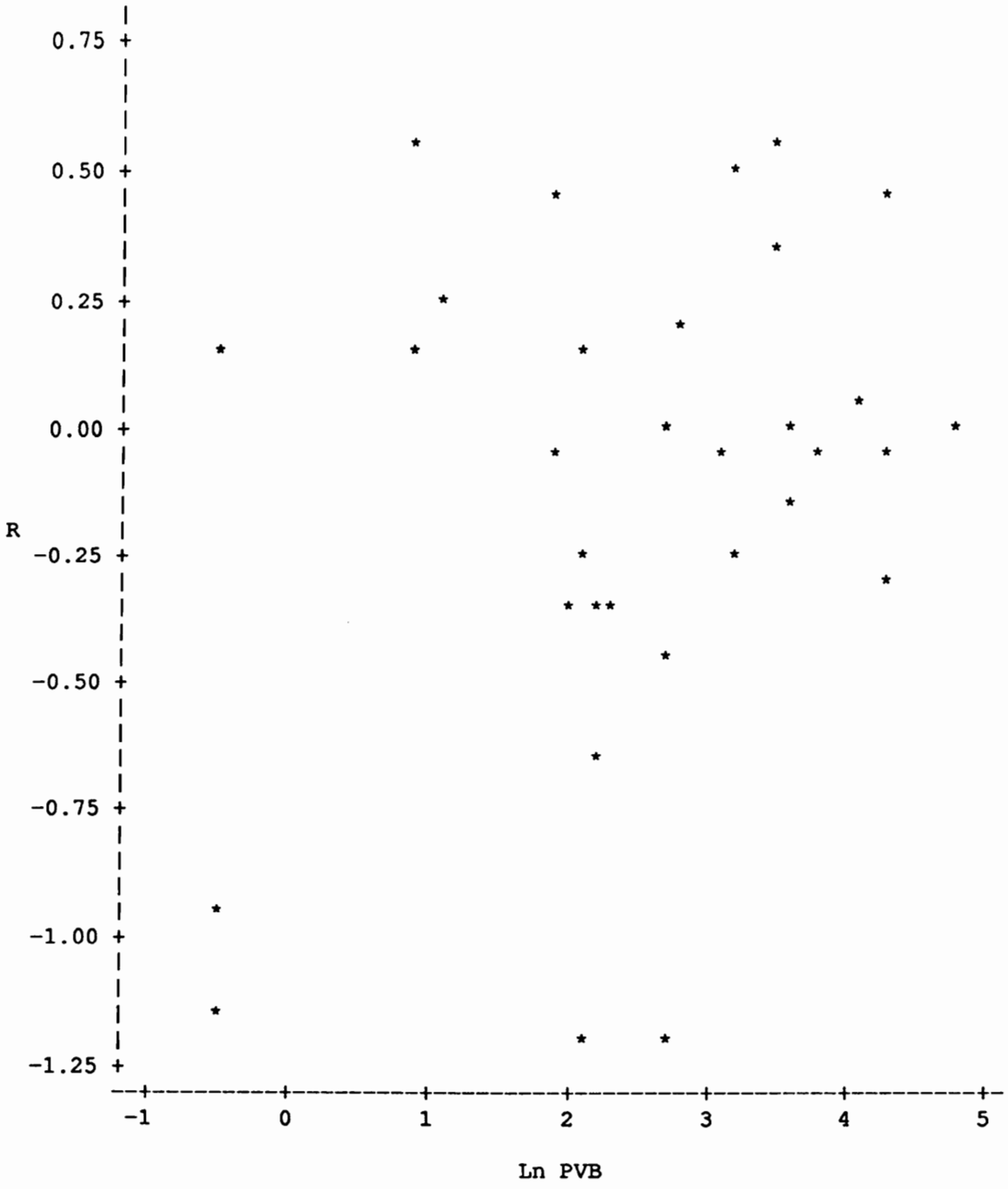



**Figure 5.1** Standardized residuals vs Ln $PV_B$ corresponding to the "true" response-model.

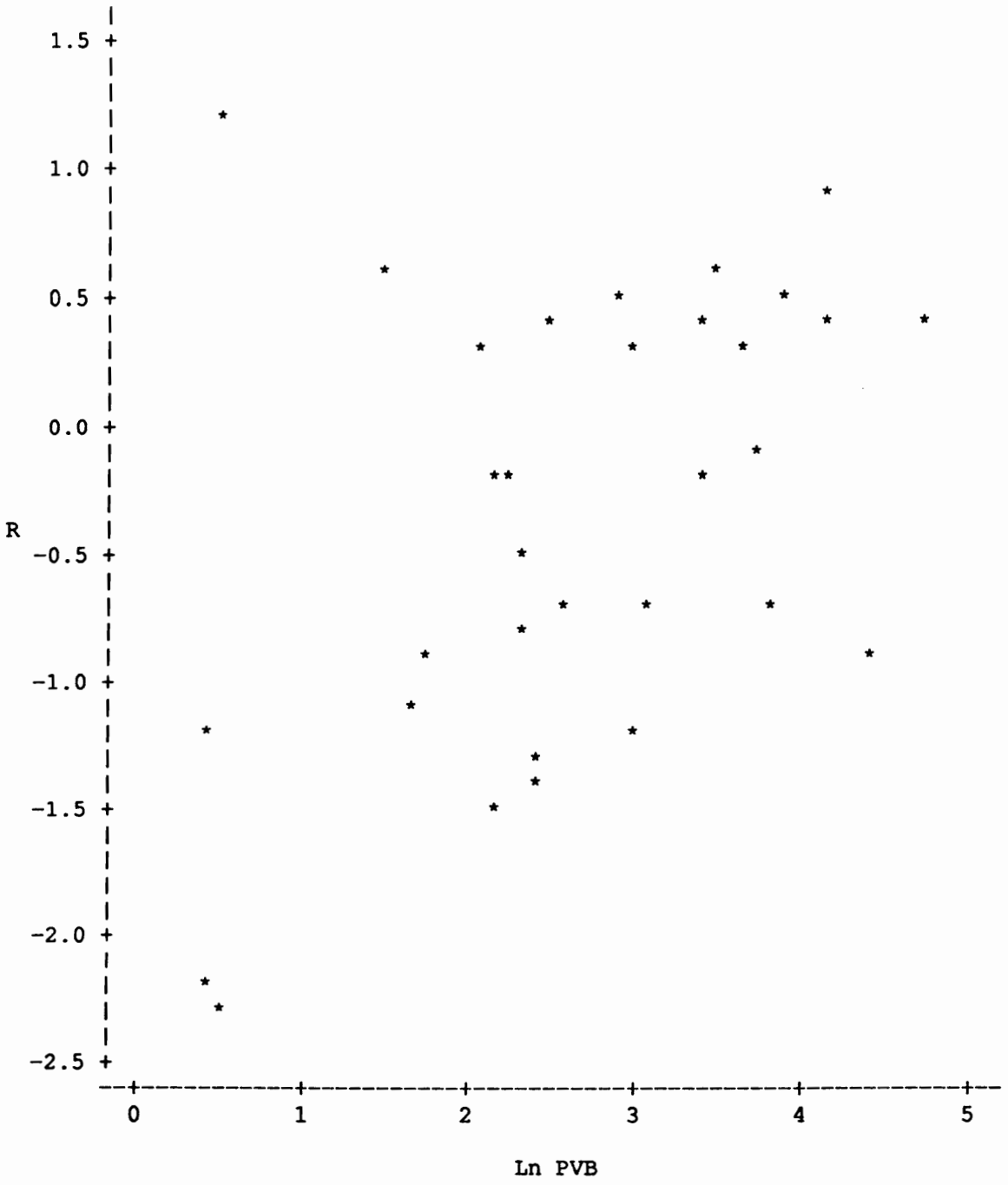



**Figure 5.2** Standardized residuals vs Ln $PV_B$ corresponding to the response-model without $z_1$.

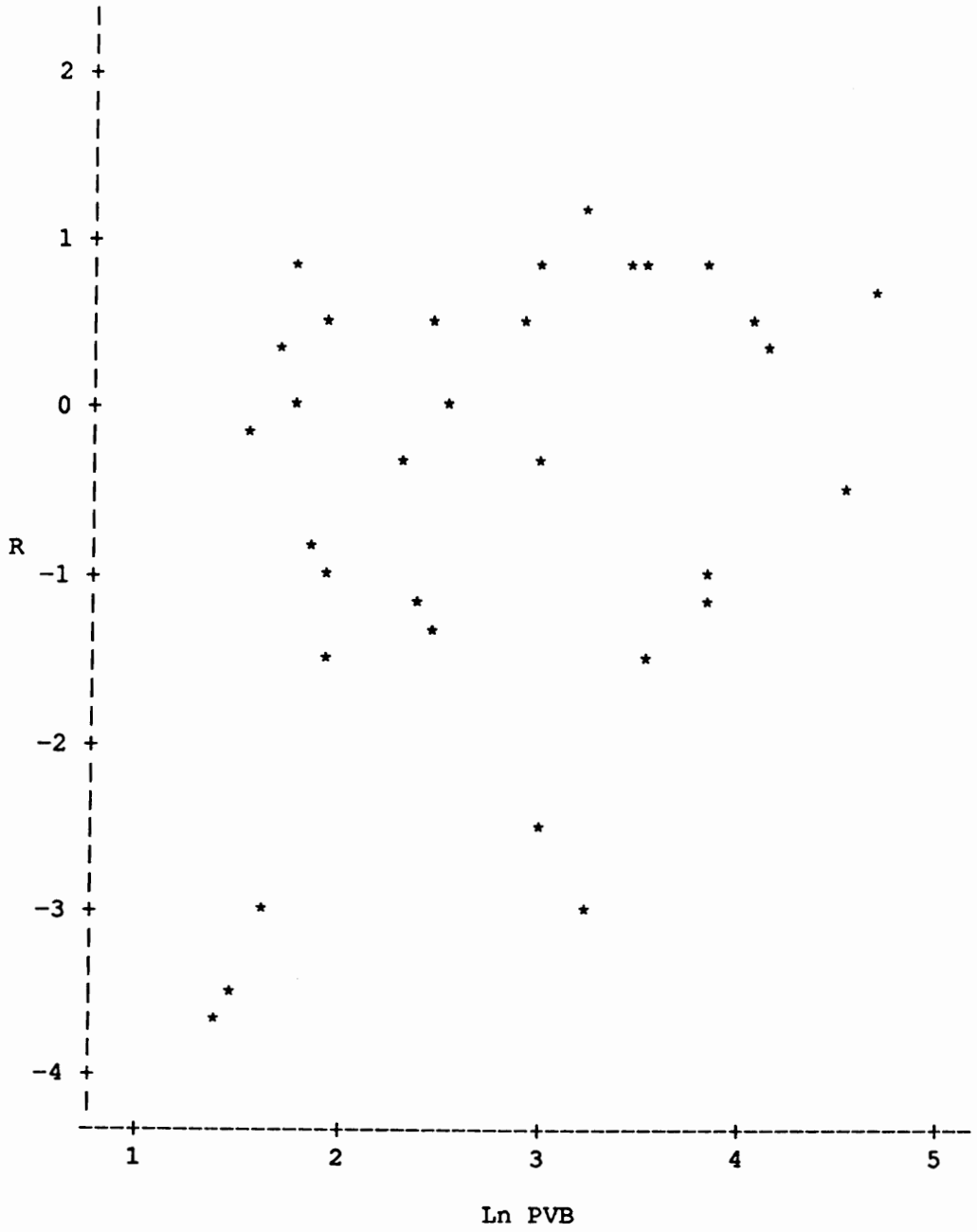



**Figure 5.3** Standardized residuals vs Ln $PV_B$ corresponding to the response-model without $x_1z_2$.

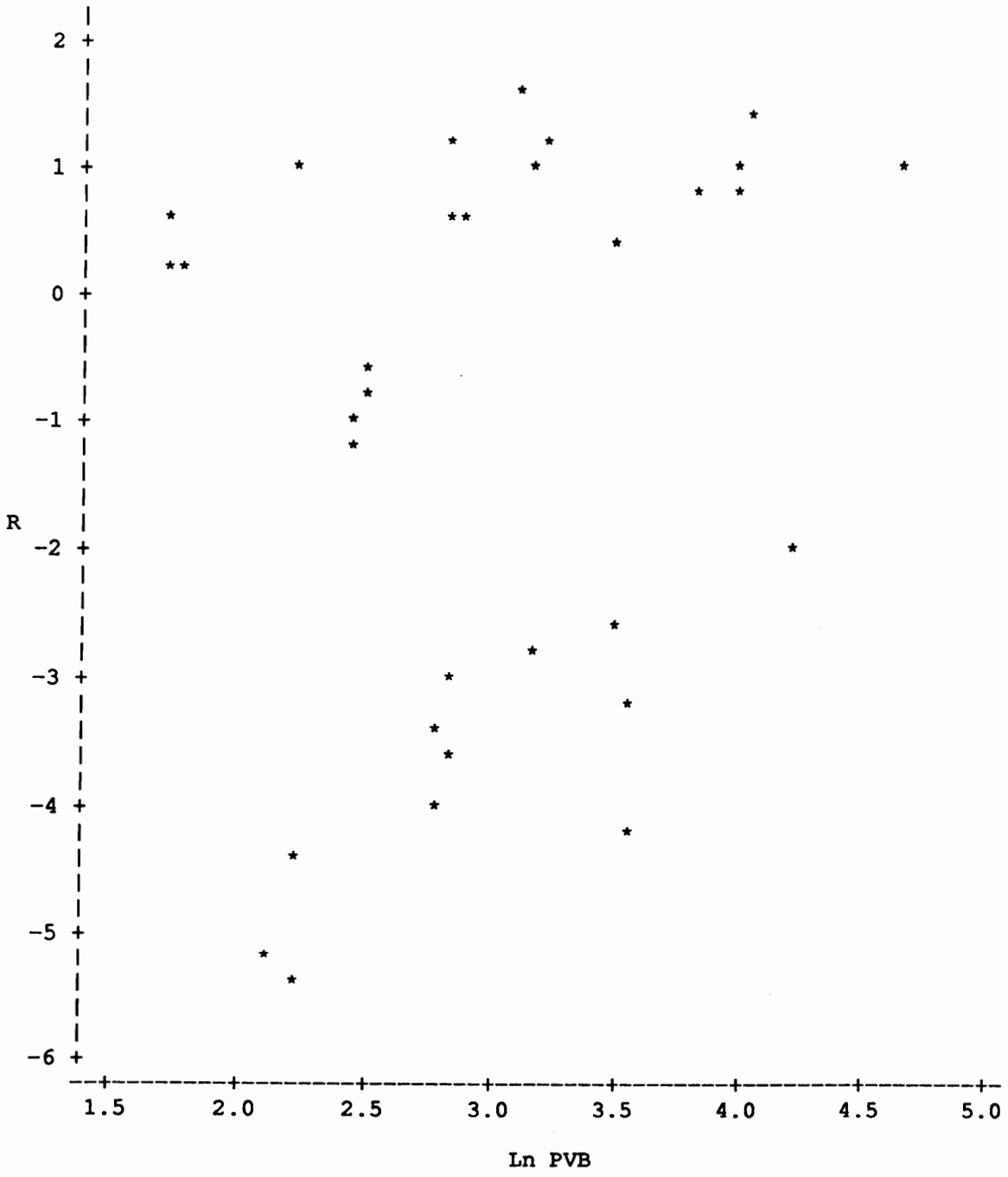



**Figure 5.4** Standardized residuals vs Ln $PV_B$ corresponding to the response-model without $x_4z_1$.

# Chapter 6
# Further RSM Relating to the Process Variance

## §6.1 Confidence Region on the Location of Minimum Process Variance

In the Robust Parameter Design problem the slopes of the response-model (3.1.1) in the direction of the noise factors contain the important information in the process variance. The two control by noise interaction plots in Figure 2.1 clearly illustrate the effect of changing slopes on the variability in the response. Recall that the process variance can be written as a quadratic form in the slopes, namely

$$\mathrm{Var}_{\underline{z}}(\mathrm{y}) = \underline{l}'\underline{l} + \sigma_\epsilon^2. \tag{6.1.1}$$

The slopes, $\underline{l} = (\underline{\gamma} + \Delta'\underline{\mathrm{x}})$, are merely linear functions of the control factors, $\underline{\mathrm{x}}$. Thus, theoretically, minimum estimated process variance can be achieved by setting $\widehat{\underline{l}} = \underline{0}$. Myers and Kim (1995) discuss how a confidence region on the location of zero process variance, apart from $\sigma_\epsilon^2$, can be created. If $\underline{\mathrm{x}}_0$ denotes the conditions where the slopes of the true response-model are simultaneously zero, i.e. $\underline{l}(\underline{\mathrm{x}}_0) = \underline{0}$, then

$$\widehat{\underline{l}}(\underline{\mathrm{x}}_0) \sim \mathrm{N}\left(\underline{0},\ \mathrm{Var}(\widehat{\underline{l}}(\underline{\mathrm{x}}_0))\right)$$

and

$$\frac{\widehat{\underline{l}}'(\underline{\mathrm{x}}_0)\,[\mathrm{Var}(\widehat{\underline{l}}(\underline{\mathrm{x}}_0))]^{-1}\,\widehat{\underline{l}}(\underline{\mathrm{x}}_0)}{r_z} \sim \mathrm{F}_{(\mathrm{r}_z,\ dfE)}. \tag{6.1.2}$$

$\mathrm{Var}(\widehat{\underline{l}}(\underline{\mathrm{x}}_0))$ can be rewritten as $\sigma_\epsilon^2\,\mathrm{C}(\underline{\mathrm{x}}_0)$ where $\mathrm{C}(\underline{\mathrm{x}}_0)$ represents the variance-covariance matrix of $\widehat{\underline{l}}$ apart from $\sigma_\epsilon^2$. When the noise factors and control by

noise interactions are orthogonal $C(\underline{x}_0)$ will be a diagonal matrix which can be written as $\text{diag}\{\underline{x}_0' \; B_{jj} \; \underline{x}_0\}$ where $B_{jj}$ is a diagonal matrix whose diagonal elements are the variances of the coefficients of the noise main effects and control by noise interactions and $\underline{x}_0' = [1, \; x_{1,0} \; , \; x_{2,0} \; , \quad \cdots \quad , \; x_{r_x,0} \; ]$. Also, dfE is the number of error degrees of freedom produced by the fitted response-model.

Using the distribution of $\hat{\underline{l}}\,(\underline{x}_0)$ and Equation (6.1.2), a confidence region can be constructed with the following probability statement:

$$P\left\{\frac{\hat{\underline{l}}\,'(\underline{x}_0)\;[C(\underline{x}_0)]^{-1}\;\hat{\underline{l}}\,(\underline{x}_0)}{r_z\hat{\sigma}_\epsilon^2} \leq F_{(r_z,\; dfE,\alpha)}\right\} \geq 1 - \alpha \qquad (6.1.3)$$

where $F_{(r_z, dfE, \alpha)}$ represents the upper $\alpha^{\text{th}}$ percent point of the F distribution. Values of $\underline{x}_0$ which satisfy the inequality within the above probability statement would form the $(1\text{-}\alpha)100\%$ confidence region on the location of minimum process variance.

**Example 6.1 (Filtration Rate Example)**

The purpose of this example is to illustrate how to produce a confidence region on the location of minimum process variance and to see the tradeoff between the process mean and variance. The example involves a chemical data set taken from Montgomery (1991). A factorial experiment was carried out to study factors thought to influence the filtration rate of a chemical product. The factors consist of two control factors ($x_1$ and $x_2$) and two noise factors ($z_1$ and $z_2$). The process engineer is interested in maximizing the filtration rate. All factors are varied at $\pm 1$ levels in a $2^4$ factorial array. The $\pm 1$ levels associated with $z_1$ and $z_2$ are assumed to be at $\pm\sigma_{z_1}$ and $\pm\sigma_{z_2}$, respectively, representing the variability in the process. The second noise factor was found to have no impact on the filtration rate and was eliminated from the model. The selected response-model is given by

$$\hat{y} = 70.0625 + 4.9375\ x_1 + 7.3125\ x_2 - 0.5625\ x_1x_2 + 10.8125\ z_1$$
$$- 9.0625\ x_1z_1 + 8.3125\ x_2z_1$$

with $R^2 = 0.9668$ and $\hat{\sigma}^2_\epsilon = 21.11806$. The estimated process mean model is given by

$$\hat{E}_{\underline{z}}\,(y) = 70.0625 + 4.9375\ x_1 + 7.3125\ x_2 - 0.5625\ x_1x_2$$

and the slope of the response-model in the direction of $z_1$ is estimated by

$$\hat{l} = 10.8125 - 9.0625\ x_1 + 8.3125\ x_2.$$

Thus, the unbiased estimate of the process variance is given by

$$PV_U = [10.8125 - 9.0625\ x_1 + 8.3125\ x_2]^2 + [1 - 0.0625(1 + x_1^2 + x_2^2)]21.11806.$$

Here we have already simplified $tr(C) = \frac{Var(\hat{l})}{\sigma^2_\epsilon} = \frac{1}{2^4}(1 + x_1^2 + x_2^2)$.

Figure 6.1 shows contours of constant mean filtration rate along with the conditions of $x_1$ and $x_2$ which yield a line of minimum process variance. The shaded region around this line depicts its 95% confidence region. To get an idea of the tradeoff between the process mean and variance, Figure 6.2 superimposes the contours of constant mean filtration rate and process standard deviation. Notice that conditions of $\underline{x}$ which yield higher estimated means also yield larger standard deviations. The process engineer should be careful when selecting operating conditions. For example, the point $x_1 = 1$ and $x_2 = 1$ yields the highest estimated mean filtration rate, but it also has an enormous corresponding process variance. Perhaps a better idea would be to operate at $x_1 = 1$ and $x_2$ close to the region around the minimum process variance, e.g. $x_2 = 0.2$. This example will serve as a basis of illustration for other forms of inference in future sections.

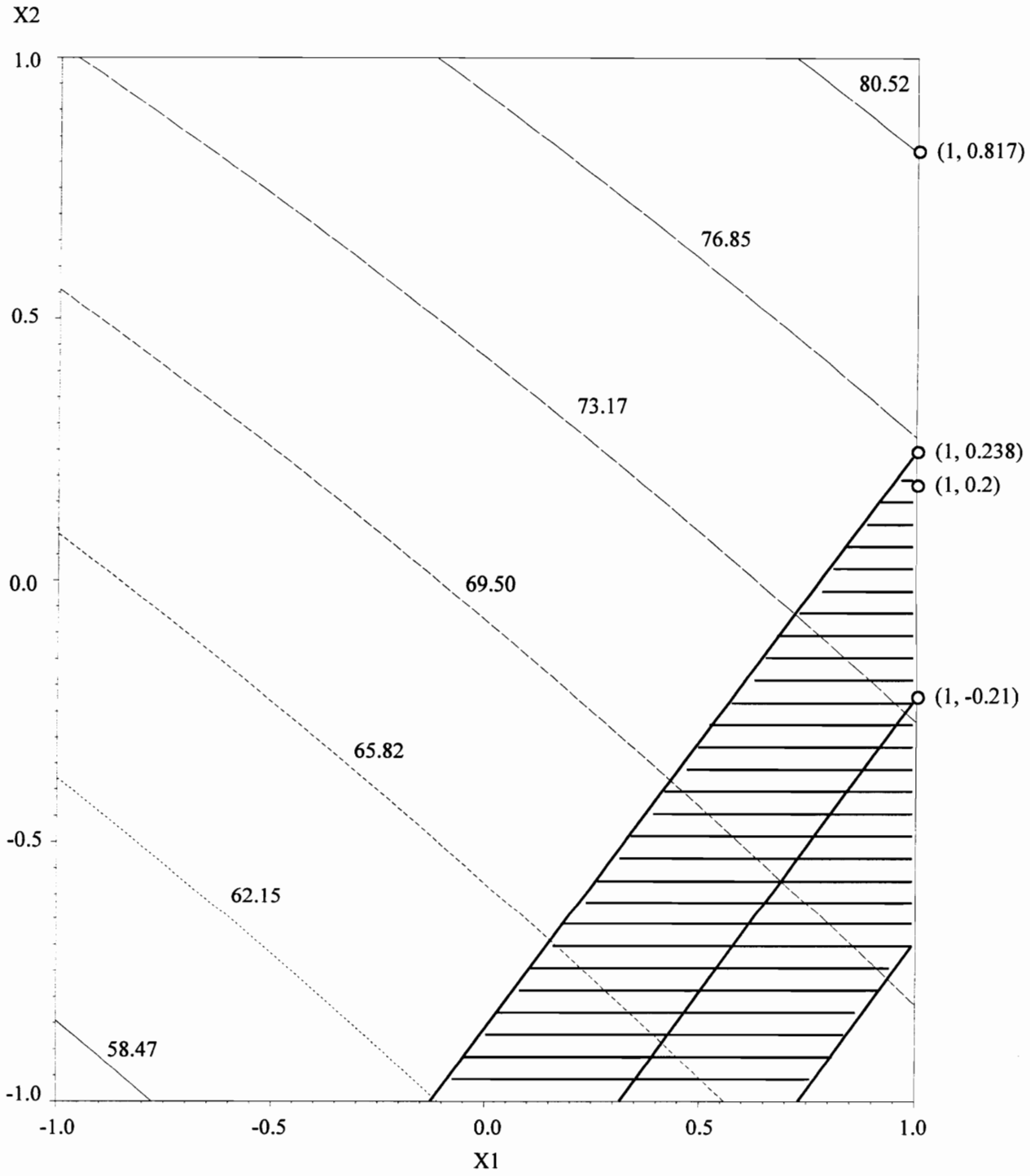



**Figure 6.1** Contour plot of the mean filtration rate and the line of minimum process variance along with its 95% confidence region.

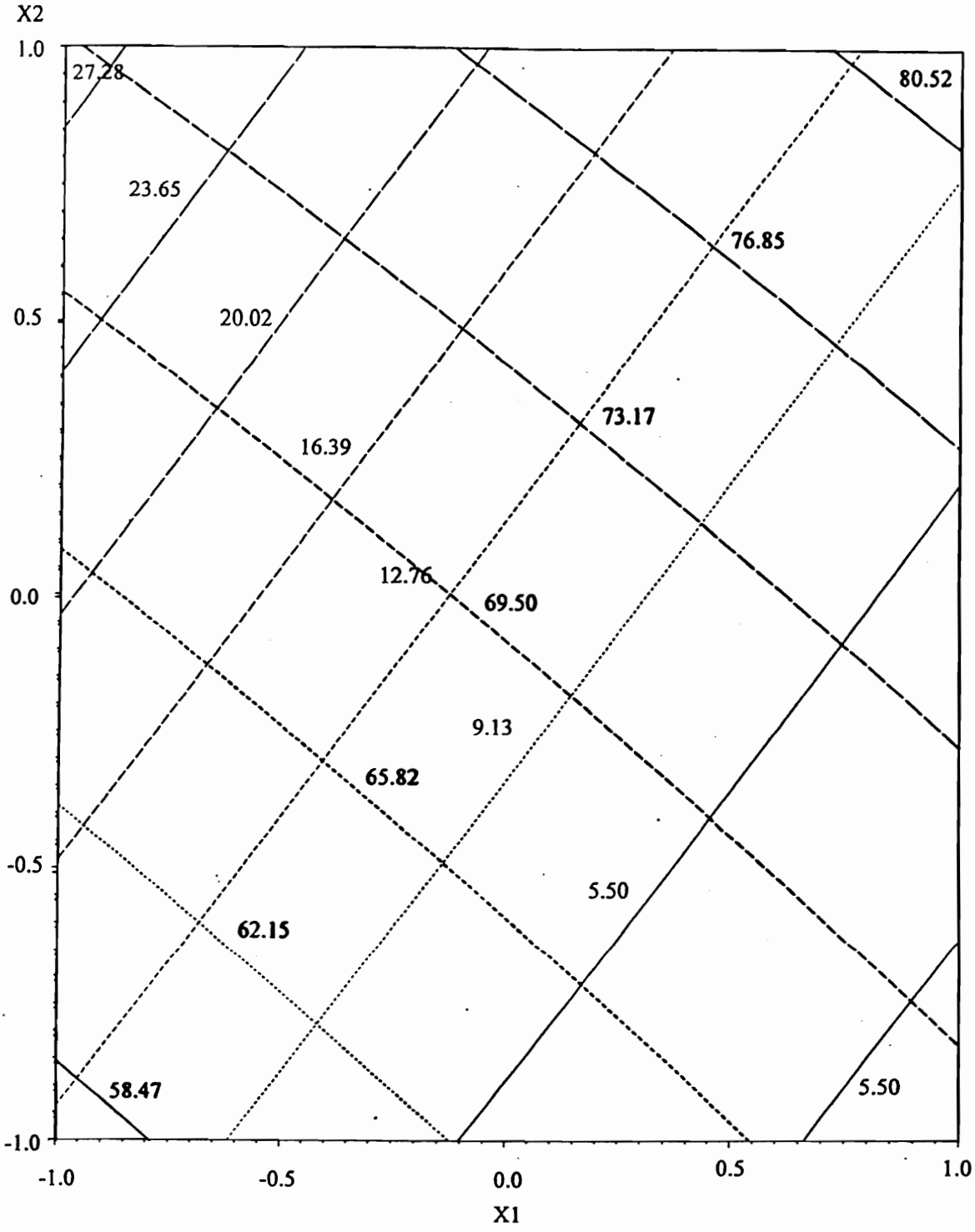



**Figure 6.2** Contour plot of both the mean filtration rate and the process standard deviation.

## §6.2 Confidence Region on $\frac{\underline{l}'\underline{l} + \sigma_\epsilon^2}{\sigma_\epsilon^2}$

In many cases the engineer may not have any intuition for what a large process variance is in the context of product development. Using the ratio of the process variance to $\sigma_\epsilon^2$ will allow for a better understanding. In fact, creating a confidence region around this ratio gives the engineer an idea of how much larger the process variance is when compared to the error variance. For example, if a setting in the control factors yields a confidence interval of [1.25, 1.75], this implies that variability due to the noise factors is 25% to 75% larger than the variance due to model error or the inability to reproduce a response with all control factors held constant. Confidence limits close to 1 would indicate a successful parameter design experiment in terms of the variability. This ratio becomes a measure of efficiency of the parameter design effort. Myers and Kim (1995) develop this confidence region in detail for the specific case where V=I and the variances of all estimated slopes are the same. A general case allowing $V \neq I$ will be considered in Chapter 7 in the discussion regarding discrete noise factors. The following development is for the case of V=I allowing the variances of the estimated slopes to differ.

To develop the more general case the distribution of $\hat{\underline{l}}'\hat{\underline{l}}$ must be determined. The derivation of this distribution was completed in §5.3. Recall that the approximate distribution is given by

$$\hat{\underline{l}}'\hat{\underline{l}} \sim \sigma_\epsilon^2 \, \mathrm{d} \, \chi^2_\nu$$

$$\text{where } \mathrm{d} = \frac{\sum_{j=1}^{r_z} (\mathrm{c}_j \rho_j)^2 \nu_j}{\sum_{j=1}^{r_z} \mathrm{c}_j \rho_j \nu_j} \quad \text{and} \quad \nu = \frac{\left( \sum_{j=1}^{r_z} \sigma_\epsilon^2 \mathrm{c}_j \rho_j \nu_j \right)^2}{\sum_{j=1}^{r_z} (\sigma_\epsilon^2 \mathrm{c}_j \rho_j)^2 \nu_j} \; . \tag{6.2.1}$$

It is also true that $\hat{\sigma}_\epsilon^2 \sim \frac{\sigma_\epsilon^2}{\mathrm{dfE}} \chi^2_{dfE}$. Using these two distributions a ratio can be formed which is distributed approximately as an F. In particular

$$\frac{\frac{\hat{\underline{l}}'\hat{\underline{l}}}{\sigma_\epsilon^2 \, \mathrm{d}} \Big/ \nu}{\frac{\mathrm{dfE}\ \hat{\sigma}_\epsilon^2}{\sigma_\epsilon^2} \Big/ \mathrm{dfE}} \sim \mathrm{F}_{(\nu,\ dfE)} \cdot \tag{6.2.2}$$

This ratio can be simplified and used to form a probability statement using the approximate F distribution:

$$\mathrm{P}\left\{ \mathrm{F}_{(\nu,\ dfE);\ 1 - \frac{\alpha}{2}} \leq \frac{\hat{\underline{l}}'\hat{\underline{l}}}{\hat{\sigma}_\epsilon^2 \, \mathrm{d}\ \nu} \leq \mathrm{F}_{(\nu,\ dfE);\ \frac{\alpha}{2}} \right\} = 1 - \alpha \tag{6.2.3}$$

where $\mathrm{d}\nu = \mathrm{tr}(\mathrm{C}) + \frac{\underline{l}'\underline{l}}{\sigma_\epsilon^2}$ .

To form the confidence bounds for $\frac{\underline{l}'\underline{l} + \sigma_\epsilon^2}{\sigma_\epsilon^2}$ choose a setting of the control factors, $\underline{x}_0$, and calculate the statistic $\frac{\hat{\underline{l}}'\hat{\underline{l}}}{\hat{\sigma}_\epsilon^2}$, then follow the steps below.

Forming the Lower Confidence Bound:

1. Set the statistic equal to $\mathrm{d}\nu \mathrm{F}_{(\nu,\ dfE);\ \frac{\alpha}{2}}$ .
2. Solve for $\mathrm{d}\nu$.
3. Lower Confidence Bound = $\mathrm{d}\nu$ - tr(C) + 1 .

Forming the Upper Confidence Bound:

1. Set the statistic equal to $\mathrm{d}\nu \mathrm{F}_{(\nu,\ dfE);\ 1 - \frac{\alpha}{2}}$ .
2. Solve for $\mathrm{d}\nu$.
3. Upper Confidence Bound = $\mathrm{d}\nu$ - tr(C) + 1 .

**Example 6.2**

Recall Example 6.1 which dealt with the Filtration Rate Data. This example can be extended to illustrate confidence regions on the ratio of the process variance to the error variance. When we examine Figures 6.1-6.2 it is reasonable that we would be interested in several locations. Two of these are $\underline{x}_A$=[1, 1] and $\underline{x}_B$=[1, 0.2]. The first location produces the maximum estimated process mean within the design region while the second location falls within the

95% confidence region on the location of minimum process variance. This location will produce a more consistent product than the first location. Using steps similar to the procedure outlined above, 95% upper one-sided confidence regions on the ratio of the process variance to $\sigma_\epsilon^2$ can be calculated for each of the locations through the following steps.

Step 1: $\underline{x}_A = [1, 1]$

Step 2: $$\frac{\hat{\underline{l}}'\hat{\underline{l}}}{\hat{\sigma}_\epsilon^2} = \frac{(10.8125 - 9.0625\, x_1 + 8.3125\, x_2)^2}{21.11806} = 4.795$$

Step 3: $4.795 = d\nu F_{(14,\, 9);0.95} = d\nu\,(0.37796) \Rightarrow d\nu = 12.6865$

Step 4: Upper Confidence Bound = 13.499

The engineer can be 95% confident that the true ratio of the process variance to the error variance at $\underline{x}_A$=[1, 1] is less than 13.499. Similarly, the upper bound at $\underline{x}_B$=[1, 0.2] can be calculated. The result is 5.732. Thus, the engineer can be 95% confident that the true ratio is less than 5.732. This is a strong argument for operating at the second location.

### §6.3 One-Sided Tolerance Intervals on Future Observations

In this section the prediction of future process observations is discussed. Tolerance intervals involving these observations relate to quality control concepts such as specification limits and process capability considerations. The importance of these concepts evolves from a need for the engineer to know the process capability and whether or not it is going to meet specification limits. One-sided tolerance intervals are particularly important when the engineer is interested in either maximizing or minimizing the response. For example, if a maximum response is desired, one approach is to find a location in the design space which yields a large value, L, for the lower one-sided tolerance limit [L, $\infty$). In terms of industry production, specifications might dictate that any manufactured part which measures less than L be considered defective. Thus, a certain variability is tolerated and the interval [L, $\infty$) may have even arisen from product design considerations, cost break-even points, quality of available manufacturing

materials, etc. Also, the tolerance limits can provide the manufacturer with a better understanding of the performance of the production process through the probability that a future response is greater than L.

**Definition 6.3.1**: A p-content *lower* one-sided tolerance interval at confidence level $(1-\gamma)$ is denoted $[\hat{\mu} - k\ \hat{\sigma}_{\underline{z}}\, y(\underline{x}),\ \infty)$ if

$$P_{\hat{\mu},\hat{\sigma}_{\underline{z}}\, y(\underline{x})}[P_{y_f}\{y_f \geq \hat{\mu} - k\ \hat{\sigma}_{\underline{z}}\, y(\underline{x})\} \geq p] \geq 1-\gamma$$

where $\hat{\sigma}_{\underline{z}}\, y(\underline{x}) = \sqrt{\hat{Var}_{\underline{z}}\, y(\underline{x})}$ is the estimated process standard deviation at a particular setting in the control factors, $\underline{x}$. The tolerance interval contains at least 100p% of the population or process being sampled with a confidence level of $(1-\gamma)$.

The goal is to compute k so that we can be $(1-\gamma)100\%$ certain that at least 100p% of the sampled population will be above $\hat{\mu} - k\ \hat{\sigma}_{\underline{z}}\, y(\underline{x})$. If the engineer is interested in a minimum response, then an upper one-sided tolerance interval should be calculated. Details are given in Myers and Kim (1995) for the specific case of V=I along with equal variances of the estimated slopes. A more general situation allowing $V \neq I$ will be discussed in Chapter 7. The following development allows the variances of the estimated slopes to differ in the case of V=I.

Recall the probability statement from Definition 6.3.1

$$P_{\hat{\mu},\hat{\sigma}_{\underline{z}}\, y(\underline{x})}[P_{y_f}\{y_f \geq \hat{\mu}(\underline{x}) - k\ \hat{\sigma}_{\underline{z}}\, y(\underline{x})\} \geq p] \geq 1-\gamma \qquad (6.3.1)$$

where $\hat{\mu}(\underline{x}) = \hat{\beta}_0 + \underline{x}'\underline{\hat{\beta}} + \underline{x}'\hat{B}\,\underline{x}$

and $\hat{\sigma}_{\underline{z}}\, y(\underline{x}) = \sqrt{\underline{\hat{l}}'(\underline{x})\ \underline{\hat{l}}(\underline{x}) + \hat{\sigma}_{\epsilon}^2}$ is the estimated process standard deviation.

The probability statement can be modified to form a standard normal variate.

$$P_{\hat{\mu},\hat{\sigma}_{\underline{z}}\, y(\underline{x})}[P_{y_f}\{\ \frac{y_f - E_{\underline{z}}(y)}{\sqrt{\underline{l}'\underline{l} + \sigma_\epsilon^2}}\ \geq\ \frac{\hat{\mu} - E_{\underline{z}}(y) - k\ \hat{\sigma}_{\underline{z}}\, y(\underline{x})}{\sqrt{\underline{l}'\underline{l} + \sigma_\epsilon^2}}\ \}\ \geq\ p\ ]\ \geq\ 1\text{-}\gamma. \qquad (6.3.2)$$

Denote the standard normal percentage point as $K_p$. Then (6.3.2) can be rewritten as

$$P_{\hat{\mu},\hat{\sigma}_{\underline{z}}\, y(\underline{x})}[\ \frac{\hat{\mu} - E_{\underline{z}}(y) - k\ \hat{\sigma}_{\underline{z}}\, y(\underline{x})}{\sqrt{\underline{l}'\underline{l} + \sigma_\epsilon^2}}\ \leq\ K_p\ ]\ \geq\ 1\text{-}\gamma. \qquad (6.3.3)$$

Further modification yields

$$P_{\hat{\mu},\hat{\sigma}_{\underline{z}}\, y(\underline{x})}[\ \frac{\hat{\mu} - E_{\underline{z}}(y)}{\sigma_\epsilon \sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}\ \leq\ \frac{K_p \sqrt{\underline{l}'\underline{l} + \sigma_\epsilon^2} + k\ \hat{\sigma}_{\underline{z}}\, y(\underline{x})}{\sigma_\epsilon \sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}\ ]\ \geq\ 1\text{-}\gamma. \qquad (6.3.4)$$

$$P_{\hat{\mu},\hat{\sigma}_{\underline{z}}\, y(\underline{x})}[\ Z\ \leq\ \lambda + \frac{k\ \hat{\sigma}_{\underline{z}}\, y(\underline{x})}{\sigma_\epsilon \sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}\ ]\ \geq\ 1\text{-}\gamma \qquad (6.3.5)$$

where $\lambda = \dfrac{K_p \sqrt{\underline{l}'\underline{l} + \sigma_\epsilon^2}}{\sigma_\epsilon \sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}$

$$P_{\hat{\mu},\hat{\sigma}_{\underline{z}}\, y(\underline{x})}[\ \frac{Z - \lambda}{\sqrt{\dfrac{\hat{\underline{l}}'\hat{\underline{l}} + \hat{\sigma}_\epsilon^2}{\sigma_\epsilon^2}}}\ \leq\ \frac{k}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}\ ]\ \geq\ 1\text{-}\gamma \qquad (6.3.6)$$

We must now determine the distribution of the denominator of the left hand side. This can also be written $\dfrac{PV_B}{\sigma_\epsilon^2}$. Recall that the approximate distribution of $PV_B$ was derived in §5.3. It is given by

$$\hat{\underline{l}}'\hat{\underline{l}} + \hat{\sigma}_\epsilon^2 \sim \sigma_\epsilon^2 \, e \, \chi^2_{\nu^*} .$$

Thus, the approximate distribution of $\frac{\hat{\underline{l}}'\hat{\underline{l}} + \hat{\sigma}_\epsilon^2}{\sigma_\epsilon^2}$ is given by $e\,\chi^2{}_{\nu^*}$

$$\text{where} \quad e = \frac{d^2\nu + \left\{\frac{1}{dfE}\right\}^2 dfE}{d\nu + \left\{\frac{1}{dfE}\right\} dfE} \quad \text{and} \quad \nu^* = \frac{\left\{d\nu + \left\{\frac{1}{dfE}\, dfE\right\}\right\}^2}{d^2\nu + \left\{\frac{1}{dfE}\right\}^2 dfE} \; . \tag{6.3.7}$$

Using this approximation, Equation (6.3.6) can be rewritten as

$$P_{\hat{\mu},\hat{\sigma}_{\underline{z}}\, y(\underline{x})}\left[ \frac{Z - \lambda}{\sqrt{e\,\chi^2{}_{\nu^*}}} \leq \frac{k}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}} \right] \geq 1-\gamma \tag{6.3.8}$$

which becomes

$$P_{\hat{\mu},\hat{\sigma}_{\underline{z}}\, y(\underline{x})}\left[ \frac{Z - \lambda}{\sqrt{\frac{\chi^2{}_{\nu^*}}{\nu^*}}} \leq \frac{k\sqrt{e\,\nu^*}}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}} \right] \geq 1-\gamma \; . \tag{6.3.9}$$

Now, we will turn the left hand side into a noncentral t variate by writing Equation (6.3.9) as

$$P_{\hat{\mu},\hat{\sigma}_{\underline{z}}\, y(\underline{x})}\left[ \frac{Z + \lambda}{\sqrt{\frac{\chi^2{}_{\nu^*}}{\nu^*}}} \geq \frac{-k\sqrt{e\,\nu^*}}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}} \right] \geq 1-\gamma \tag{6.3.10}$$

And, finally,

$$P_{\hat{\mu},\hat{\sigma}_{\underline{z}}\, y(\underline{x})}\left[ t'_{\nu^*,\,\lambda} \geq m^* \right] \geq 1-\gamma \tag{6.3.11}$$

or equivalently,

$$P_{\hat{\mu},\hat{\sigma}_{\underline{z}}\, y(\underline{x})}\left[ t'_{\nu^*,\,\lambda} \leq m^* \right] \leq \gamma \tag{6.3.12}$$

where $m^* = \frac{-k\sqrt{e\,\nu^*}}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}$ .

Equation (6.3.12) allows us to calculate k through the following procedure:

(i) Estimate all model terms in the response-model as well as $\sigma_\epsilon^2$.

(ii) Select probability levels p and $\gamma$.

(iii) Select a setting of the control factors, $\underline{x}_0$, and calculate $\underline{x}'_0(X'X)^{-1}\underline{x}_0$, $\hat{\mu}(\underline{x}_0)$, and $\hat{\sigma}_{\underline{z}}\, y(\underline{x}_0)$.

(iv) Estimate $\lambda$ and calculate e and $\nu^*$.

(v) Using the estimate of $\lambda$ find the percentage point of the noncentral t, e.g. using SAS, found in Equation (6.3.12) and equate it to

$$m^* = \frac{-k\sqrt{e\,\nu^*}}{\sqrt{\underline{x}'_0(X'X)^{-1}\underline{x}_0}} .$$

(vi) Solve for k.

(vii) The lower tolerance limit at $\underline{x}_0$ is then $\hat{\mu}(\underline{x}_0) - k\,\hat{\sigma}_{\underline{z}}\, y(\underline{x}_0)$.

Note that the lower one-sided tolerance interval is used primarily when the engineer is interested in maximizing the response. If a minimum response is of interest an upper one-sided tolerance interval would be more appropriate. This interval can be created in a similar fashion as that derived above. Adjustments can be easily made.

## §6.4 Prediction Limits on a Future Response

Prediction intervals play an important role when an engineer desires probabilistic bounds on a new response at fixed conditions on the control factors. They have a familiar interpretation in a multiple linear regression setting. However, the development in this setting is somewhat more complicated. In order to develop the bounds, the distribution of $\hat{\underline{l}}'\hat{\underline{l}}$ must first be determined.

Recall the derivation of the approximate distribution of $\hat{\underline{l}}'\hat{\underline{l}}$ in §5.3. It was determined that

$$\hat{\underline{l}}'\hat{\underline{l}} \sim \sigma_\epsilon^2 \, d\chi^2_\nu$$

$$\text{where } d = \frac{\sum_{j=1}^{r_z} (c_j \rho_j)^2 \nu_j}{\sum_{j=1}^{r_z} c_j \rho_j \nu_j} \quad \text{and} \quad \nu = \frac{\left( \sum_{j=1}^{r_z} \sigma_\epsilon^2 c_j \rho_j \nu_j \right)^2}{\sum_{j=1}^{r_z} (\sigma_\epsilon^2 c_j \rho_j)^2 \nu_j} \; . \tag{6.4.1}$$

It is also true that $\hat{\sigma}_\epsilon^2 \sim \frac{\sigma_\epsilon^2}{dfE} \chi^2_{dfE}$. By combining the distributions of $\hat{\underline{l}}'\hat{\underline{l}}$ and $\hat{\sigma}_\epsilon^2$ we can also determine the distribution of the estimated variance of $[y_f(\underline{x}_0) - \hat{\mu}(\underline{x}_0)]$. It should be clear that, approximately,

$$\hat{\sigma}^2_{y\text{-}\hat{\mu}} = [\hat{\underline{l}}'\hat{\underline{l}} + \hat{\sigma}_\epsilon^2 (1 + \underline{x}'_0 (X'X)^{-1} \underline{x}_0)] \sim \sigma_\epsilon^2 \, a \, \chi^2_{\nu_4} \tag{6.4.2}$$

$$\text{where } a = \frac{d^2\nu + \left\{ \frac{1}{dfE} (1 + \underline{x}'_0 (X'X)^{-1} \underline{x}_0) \right\}^2 dfE}{d\nu + \left\{ \frac{1}{dfE} (1 + \underline{x}'_0 (X'X)^{-1} \underline{x}_0) \right\} dfE} \quad \text{and}$$

$$\nu_4 = \frac{\left\{ d\nu + \left\{ \frac{1}{dfE} (1 + \underline{x}'_0 (X'X)^{-1} \underline{x}_0) \right\} dfE \right\}^2}{d^2\nu + \left\{ \frac{1}{dfE} (1 + \underline{x}'_0 (X'X)^{-1} \underline{x}_0) \right\}^2 dfE} \; .$$

The goal of the engineer is to find the limits around $\hat{y}(\underline{x}_0)$ on $y_f(\underline{x}_0)$ corresponding to a probability level of $(1\text{-}\alpha)$ at fixed process conditions. The probability statement for determining the prediction limits can be written

$$P \{ \hat{\mu}(\underline{x}_0) - k \, \hat{\sigma}_{y\text{-}\hat{\mu}}(\underline{x}_0) \leq y_f(\underline{x}_0) \leq \hat{\mu}(\underline{x}_0) + k \, \hat{\sigma}_{y\text{-}\hat{\mu}}(\underline{x}_0) \} = 1 - \alpha \tag{6.4.3}$$

$$= P\{ -k \leq \frac{y_f(\underline{x}_0) - \hat{\mu}(\underline{x}_0)}{\hat{\sigma}_\epsilon \sqrt{\frac{\hat{\underline{l}}'\hat{\underline{l}}}{\hat{\sigma}_\epsilon^2} + 1 + \underline{x}'_0 (X'X)^{-1} \underline{x}_0}} \leq k \} = 1 - \alpha. \tag{6.4.4}$$

Denote $Z = \dfrac{y_f(\underline{x}_0) - \hat{\mu}(\underline{x}_0)}{\sigma_\epsilon \sqrt{\dfrac{\underline{l}'\underline{l}}{\sigma_\epsilon^2} + 1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0}} \sim N(0, 1),$

then Equation (6.4.4) can be rewritten as

$$P\left\{ -k \leq \frac{Z\sigma_\epsilon \sqrt{\dfrac{\underline{l}'\underline{l}}{\sigma_\epsilon^2} + 1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0}}{\hat{\sigma}_\epsilon \sqrt{\dfrac{\hat{\underline{l}}'\hat{\underline{l}}}{\hat{\sigma}_\epsilon^2} + 1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0}} \leq k \right\} = 1 - \alpha \qquad (6.4.5)$$

Now, the distribution of $[\hat{\underline{l}}'\hat{\underline{l}} + \hat{\sigma}_\epsilon^2(1 + \underline{x}'_0(X'X)^{-1}\underline{x}_0)]$ in Equation (6.4.2) can be applied, yielding

$$P\left\{ -k \leq \frac{Z \sqrt{\dfrac{\underline{l}'\underline{l}}{\sigma_\epsilon^2} + 1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0}}{\sqrt{a\,\chi^2_{\nu_4}}} \leq k \right\} = 1 - \alpha \qquad (6.4.6)$$

Further manipulations yield

$$P\left\{ -m \leq t_{\nu_4} \leq m \right\} = 1 - \alpha \qquad (6.4.7)$$

where $m = \dfrac{k\sqrt{a\nu_4}}{\sqrt{\dfrac{\underline{l}'\underline{l}}{\sigma_\epsilon^2} + 1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0}}$ and $t_{\nu_4} = \dfrac{Z}{\sqrt{\dfrac{\chi^2_{\nu_4}}{\nu_4}}}$. (6.4.8)

In order to determine the estimated prediction limits, k must be calculated. Thus, the following procedure should be implemented:

(i) Estimate all model terms in the response-model as well as $\sigma_\epsilon^2$.

(ii) Select the probability $\alpha$.

(iii) Select fixed conditions on the control factors, $\underline{x}_0$, and calculate

$\underline{x}_0'(X'X)^{-1}\underline{x}_0$, $\hat{\mu}(\underline{x}_0)$, $\hat{\sigma}_{y\text{-}\hat{\mu}}(\underline{x}_0)$ , and $\hat{\underline{l}}'\hat{\underline{l}}$.

(iv) Estimate a and $\nu_4$ through estimates of d and $\nu$.

(v) Set m equal to the percentage point of $t_{\nu_4}$ and solve for k.

(vi) The prediction limits at $\underline{x}_0$ are then $[\hat{\mu}(\underline{x}_0) \pm k\, \hat{\sigma}_{y\text{-}\hat{\mu}}(\underline{x}_0)]$.

**Example 6.3**

Recall Examples 6.1 and 6.2 which dealt with the Filtration Rate Data. Suppose the engineer is interested in operating at a specific location, but would like to know the lower one-sided prediction limit for a future response. When we examine Figures 6.1-6.2 it is reasonable that we would be interested in several locations. Two of which are $\underline{x}$ =[1, 1] and $\underline{x}$ =[1, 0.2]. The first location produces the maximum estimated process mean within the design region while the second location falls within the 95% confidence region on the location of minimum process variance. This location will produce a more consistent product than the first location. Using steps similar to the procedure outlined above, 95% lower one-sided prediction limits on a future response at these locations are

| $\underline{x}$ | 95% Lower One-Sided Prediction Interval | |
|---|---|---|
| [1, 1] | [62.5279, $\infty$) | |
| [1, 0.2] | [65.9982, $\infty$) | ** |

Notice that the first lower prediction limit contains a lower value than that of the second location even though its estimated process mean, as depicted in Figure 6.1, is quite large. The engineer can be 95% certain that a future response at the specified location will be covered by the above lower one-sided prediction limits.

**§6.5 Results of Simulation**

Whenever approximate distributions are used in place of the true distributions for the purpose of creating confidence regions one should expect the

confidence level to be in error. Issues at hand are how close the approximate distributions are to the true distributions and how precise the confidence levels are. To address these two issues, limited simulation experiments were conducted on two of the previously discussed confidence regions. The purpose of these experiments was to provide an idea of how well the approximations behaved by examining the coverage probabilities and comparing them to the associated confidence levels.

### §6.5.1 Results of Simulation on the Confidence Region on $\frac{\underline{l}'\underline{l}+\sigma_\epsilon^2}{\sigma_\epsilon^2}$

The confidence region developed in §6.2 depended heavily on the approximate distribution of $\hat{\underline{l}}'\hat{\underline{l}}$. This approximation used methods developed by Patnaik (1949), Satterthwaite (1946), and Welch (1956). The true distribution of $\hat{\underline{l}}'\hat{\underline{l}}$ , shown in (6.2.1), is a weighted sum of noncentral chi-squares. This sum was the focus of the simulation.

In detail, this particular simulation experiment was run under various circumstances. For the purpose of illustration three noise factors were considered. Values of the noncentrality parameter associated with the noncentral chi-squares in the true distribution of $\hat{\underline{l}}'\hat{\underline{l}}$ ranged from the case where $\lambda_1$=2, $\lambda_2$=2.5, and $\lambda_3$=3 to a more extreme case where $\lambda_1$=19, $\lambda_2$=20, and $\lambda_3$=21. This range contained reasonable, and possibly unreasonable, values of the noncentrality parameter. Also, the distribution of $\hat{\sigma}_\epsilon^2$ was simulated to mimic the estimated mean square error. The upper and lower confidence bounds were then created and examined to see if they covered the true ratio of $\frac{\underline{l}'\underline{l}+\sigma_\epsilon^2}{\sigma_\epsilon^2}$. Next, two sizes of experiments were considered. One experiment contained 3 control factors and a reasonable number of degrees of freedom for error was chosen to be 40. The other experiment was smaller and contained 2 control factors with 20 degrees of freedom for error. The average coverage probabilities and their standard errors are based on 10,000 repetitions each. The simulations were also run using various confidence levels, CL=(1-$\alpha$)100%, ranging from 90% to 95%. The results of the simulation are outlined in Tables 6.1-6.2.

**Table 6.1** Coverage Probabilities and Standard Errors in % from the Simulation of the Confidence Region on $\frac{\underline{l}'\underline{l} + \sigma_\epsilon^2}{\sigma_\epsilon^2}$ when $r_x$=3 and dfE=40

| $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | CL=90% | CL=95% | CL=99% |
|---|---|---|---|---|---|
| 2 | 2.5 | 3 | 88.00 (0.33) | 93.26 (0.28) | 98.20 (0.14) |
| 3 | 3.5 | 4 | 89.12 (0.46) | 94.42 (0.36) | 98.48 (0.15) |
| 4 | 5 | 6 | 89.14 (0.36) | 94.18 (0.20) | 98.63 (0.08) |
| 8 | 10 | 12 | 90.31 (0.32) | 95.27 (0.25) | 98.79 (0.04) |
| 19 | 20 | 21 | 89.21 (0.28) | 95.35 (0.19) | 98.92 (0.10) |

**Table 6.2** Coverage Probabilities and Standard Errors in % from the Simulation of the Confidence Region on $\frac{\underline{l}'\underline{l} + \sigma_\epsilon^2}{\sigma_\epsilon^2}$ when $r_x$=2 and dfE=20

| $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | CL=90% | CL=95% | CL=99% |
|---|---|---|---|---|---|
| 2 | 2.5 | 3 | 89.72 (0.12) | 94.69 (0.12) | 99.04 (0.14) |
| 3 | 3.5 | 4 | 88.86 (0.28) | 93.31 (0.31) | 98.17 (0.12) |
| 4 | 5 | 6 | 90.06 (0.38) | 95.05 (0.18) | 98.78 (0.11) |
| 8 | 10 | 12 | 89.95 (0.39) | 94.69 (0.19) | 98.91 (0.12) |
| 19 | 20 | 21 | 89.80 (0.31) | 94.56 (0.30) | 98.82 (0.14) |

In summary, Tables 6.1-6.2 show that the coverage probabilities are quite close to the associated confidence levels for the various cases of the lambda values. As the values of lambda increase, the coverage probabilities also increase. And, for all practical purposes, the confidence region does a reasonable job at maintaining a high coverage probability. Based on the simulation, the practitioner should choose the confidence level which yields a coverage probability of interest.

### §6.5.2 Results of Simulation on Prediction Limits

As in §6.2, the development of Prediction Limits depends heavily on the approximate distribution of $\hat{\underline{l}}'\hat{\underline{l}}$. Data used to create these limits would be used in conjunction with an approximate distribution. Again, important issues at hand are how well the limits perform and how well the approximate distribution mimics the true distribution. To address these issues another simulation experiment was conducted.

The true distribution of the variance of $[y_f(\underline{x}_0) - \hat{\mu}(\underline{x}_0)]$ was simulated. This simulated data was used along with the procedure outlined in §6.4 to create prediction limits. Now, other parameter information need be supplied. For example, $\underline{x}'_0(X'X)^{-1}\underline{x}_0$ was set at 0.5. This is a reasonable distance away from the center of the design region and likely to be of interest to a practitioner. Also, various values of the noncentrality parameter are considered along with 2 sizes of experiments involving 3 noise factors. For the purpose of illustration this simulation is quite similar to that conducted in §6.5.1. Average coverage probabilities and their standard errors are based on 10,000 repetitions. Confidence levels, CL=(1-$\alpha$)100%, of 90% and 95% are considered. Tables 6.3-6.4 display the results of the simulation.

In summary, the coverage probabilities are always slightly larger than the confidence level used to create the prediction limits. However, the values do suggest that the prediction regions are quite useful. Again, the practitioner should choose the confidence level which provides the coverage probability of interest. For example, if a coverage probability of 95% is desired, then, based on the

**Table 6.3** Coverage Probabilities and Standard Errors in % from the Simulation of Prediction Limits when $r_x$=3 and dfE=40

| $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | CL=90% | CL=95% |
|---|---|---|---|---|
| 2 | 2.5 | 3 | 91.74 (0.29) | 96.31 (0.11) |
| 3 | 3.5 | 4 | 92.50 (0.26) | 96.59 (0.19) |
| 4 | 5 | 6 | 93.00 (0.35) | 96.90 (0.16) |
| 8 | 10 | 12 | 94.69 (0.19) | 97.83 (0.17) |
| 19 | 20 | 21 | 95.92 (0.14) | 98.32 (0.12) |

**Table 6.4** Coverage Probabilities and Standard Errors in % from the Simulation of Prediction Limits when $r_x$=2 and dfE=20

| $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | CL=90% | CL=95% |
|---|---|---|---|---|
| 2 | 2.5 | 3 | 92.67 (0.21) | 96.40 (0.13) |
| 3 | 3.5 | 4 | 92.64 (0.23) | 96.68 (0.20) |
| 4 | 5 | 6 | 93.67 (0.30) | 97.12 (0.18) |
| 8 | 10 | 12 | 94.97 (0.18) | 98.02 (0.17) |
| 19 | 20 | 21 | 96.15 (0.21) | 98.66 (0.06) |

simulation information, a confidence level of 95% should be chosen. It offers a coverage probability of at least 95% for various values of the noncentrality parameters.

## §6.6 Effect of the Extended Response-Model on the Various Confidence Regions

Recall the extended response-model of (3.6.1)

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'B\,\underline{x} + \underline{z}'\underline{\gamma} + \underline{x}'\Delta\,\underline{z} + \underline{z}'\Psi\,\underline{z} + \epsilon, \tag{6.6.1}$$

The addition of quadratic noise factors and noise by noise interaction terms would affect the construction of the previously discussed confidence regions through the distribution of the estimated process variance. Each type of confidence region would be affected differently. The following sections outline the effect of the extended response-model by explaining how to re-derive the confidence regions. Appendix C contains details on deriving the appropriate approximate distributions involving the "extra" terms.

### §6.6.1 Confidence Region on the Location of Minimum Process Variance

This particular confidence region is unaffected by the addition of quadratic noise factors or noise by noise interaction terms in the extended response-model. It requires only information supplied by $\underline{\hat{l}} = (\hat{\gamma} + \hat{\Delta}'\underline{x})$ and the variance-covariance matrix of $\underline{\hat{l}}$, $\sigma_\epsilon^2 C$.

### §6.6.2 Confidence Region on $\frac{\underline{l}'\underline{l} + \sigma_\epsilon^2}{\sigma_\epsilon^2}$

This region is formed similarly to that of §6.2. The main difference lies in the approximate distribution of

$$\underline{\hat{l}}'\underline{\hat{l}} + 2\sum_{j=1}^{r_z}\hat{\psi}_{jj}^2 + \sum_{j<k}\sum\hat{\psi}_{jk}^2 \tag{6.6.2.1}$$

which is derived in detail in Appendix §C.1. This approximate distribution is found to be $\sigma_\epsilon^2 K\chi_\nu^2$ and is used along with the distribution of $\hat{\sigma}_\epsilon^2$ to form the F distribution used in the probability statement of (6.2.5). All terms are defined in

Appendix §C.1. The procedure used to create the confidence bounds is now similar to that used in §6.2.

### §6.6.3 One-Sided Tolerance Intervals on Future Observations

The derivation for this region is similar to that of §6.3. The only difference lies in the approximate distribution of

$$\hat{\underline{l}}'\hat{\underline{l}} + 2\sum_{j=1}^{r_z}\hat{\psi}_{jj}^2 + \sum_{j<k}\sum\hat{\psi}_{jk}^2 + \hat{\sigma}_\epsilon^2 . \qquad (6.6.3.1)$$

This distribution is derived in §C.2 and is found to be $\sigma_\epsilon^2 K^{**}\chi^2_{\nu^{**}}$ . This is used in place of $\sigma_\epsilon^2 e\chi^2_{\nu^*}$ in the derivation of the tolerance intervals. The procedure used to create the tolerance interval is now similar to that used in §6.3.

### §6.6.4 Prediction Limits on a Future Response

Again, the derivation of these limits is similar to that found in §6.4. The difference can be found in the approximate distribution of the estimated variance of $[y_f(\underline{x}_0) - \hat{\mu}(\underline{x}_0)]$ denoted by

$$\hat{\sigma}^2_{y\text{-}\hat{\mu}} = [\hat{\underline{l}}'\hat{\underline{l}} + 2\sum_{j=1}^{r_z}\hat{\psi}_{jj}^2 + \sum_{j<k}\sum\hat{\psi}_{jk}^2 + \hat{\sigma}_\epsilon^2(1 + \underline{x}'_0(X'X)^{-1}\underline{x}_0)] . \qquad (6.6.4.1)$$

Derivations found in Appendix §C.3 show that the approximate distribution of (6.6.4.1) is $\sigma_\epsilon^2 a^*\chi^2_{\nu_5}$ . This approximation is used in place of $\sigma_\epsilon^2 a\chi^2_{\nu_4}$ found in (6.4.2) when calculating the prediction limits. The procedure found in §6.4 can be followed with the appropriate distributional changes made.

The changes made in the derivations and calculations of the various confidence regions discussed in this chapter are not complicated. The addition of the quadratic noise factors and the noise by noise interaction terms affect the distributions required in the derivation of each type of region, but do not cause a complete disruption in the procedure for calculating the regions.

The techniques presented in this chapter all assumed that the noise factors were continuous. This assumption is not always correct. It is entirely possible for a noise factor to be categorical and consist of different discrete levels. The next chapter discusses how this could occur and how to handle the noise factors in this type of setting.

# Chapter 7
# RSM in the Presence of Categorical Noise Factors

## §7.1 Introduction

It is quite common for a noise factor to be categorical. Examples would include different brands of materials, different types of engines, different types of handling by consumer, etc. In this chapter we combine what has previously been developed with the introduction of noise factors which are discrete rather than continuous. In particular, we will discuss the case of multinomial noise factors using indicator variables.

Consider a response-model in control, $\underline{x}$, and noise, $\underline{z}$, for some process with response y given by

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}' B \underline{x} + \underline{z}'\underline{\gamma} + \underline{x}'\Delta \underline{z} + \epsilon \qquad (7.1.1)$$

where $\epsilon \sim N(0, \sigma_\epsilon^2)$. Now, for the purpose of illustration, suppose there are two control factors and two noise factors. Furthermore, suppose each noise factor, $z_1$ and $z_2$, consists of two and three categories, respectively, with known probabilities of occurrence in the process. For example, $z_1$ may be two different types of engines and $z_2$ may be three different brands of oils. In this case we could define indicator variables for each of the noise factors. Let

$$I_1 = \begin{cases} 1, \text{ for engine type 1} \\ 0, \text{ otherwise,} \end{cases} \qquad \text{and}$$

$$I_2 = \begin{cases} 1, \text{ for oil 1} \\ 0, \text{ otherwise,} \end{cases} \qquad I_3 = \begin{cases} 1, \text{ for oil 2,} \\ 0, \text{ otherwise.} \end{cases}$$

Now model (7.1.1) can be rewritten as

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'\,B\,\underline{x} + \gamma_1 I_1 + \gamma_2 I_2 + \gamma_3 I_3 + \delta_{11} x_1 I_1 + \delta_{12} x_1 I_2 + \delta_{13} x_1 I_3 + \delta_{21} x_2 I_1 + \delta_{22} x_2 I_2 + \delta_{23} x_2 I_3 + \epsilon\ . \tag{7.1.2}$$

Indicator variable $I_1$ partitions $z_1$ and indicator variables $I_2$ and $I_3$ partition $z_2$. Note that the indicator variables $I_1$, $I_2$, and $I_3$ are distributed as Bernoulli random variables with success probabilities $p_1$, $p_2$, and $p_3$, respectively. Thus, response-model (7.1.2) can be considered a special case of the response-model of (3.1.1). In matrix notation, a response-model containing $r_I$ indicator variables can be written in general as

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'\,B\,\underline{x} + \underline{I}'\underline{\gamma} + \underline{x}'\Delta\,\underline{I} + \epsilon \tag{7.1.3}$$

where $\epsilon \sim N(0, \sigma_\epsilon^2)$,

$\underline{x}' = [x_1,\ x_2,\ \ldots\ ,\ x_{r_x}]$,

$\underline{I}' = [I_1,\ I_2,\ \ldots\ ,\ I_{r_I}]$,

with parameters $\underline{\beta}'$, $\underline{\gamma}'$, B, and $\Delta$.

The vector $\underline{\beta}'$ and matrix B are defined in (3.1.1). The vector $\underline{\gamma}'$ has dimensions $r_I$ x 1 and the matrix $\Delta$ has dimensions $r_x$ x $r_I$.

Recall that the indicator variables, $\underline{I}$, are Bernoulli trials and

$$E(\underline{I}) = [p_1, p_2, \cdots, p_{r_I}],$$

$$\mathrm{Var}(\underline{I}) = V = \begin{bmatrix} V_1 & & & \underline{0} \\ & V_2 & & \\ & & \ddots & \\ \underline{0} & & & V_{r_z} \end{bmatrix}, \tag{7.1.4}$$

where $V_j$ , j = 1, 2, $\cdots$ , $r_z$, is a block matrix corresponding to the variance-covariance matrix of the $j^{th}$ noise factor. For example, $V_1$ can be written

$$V_1 = \begin{bmatrix} p_1(1-p_1) & -p_1p_2 & \cdots & -p_1p_{l_1} \\ & p_2(1-p_2) & \cdots & -p_2p_{l_1} \\ & & \ddots & \vdots \\ \text{sym} & & & p_{l_1}(1-p_{l_1}) \end{bmatrix},$$

where the size of the block, $l_1$, is the number of indicator variables required to partition the first noise factor. The diagonal elements of V are the variances of Bernoulli trials, $p_i(1-p_i)$, and the off-diagonal elements of each "block" submatrix are covariances among Bernoulli trials, $-p_ip_j$, where $p_i$ and $p_j$ represent success probabilities of indicator variables $I_i$ and $I_j$ ($i \neq j$), respectively.

From the response-model (7.1.3) the process mean model is given by

$$E_{\underline{I}}(y) = [\beta_0 + E(\underline{I})'\underline{\gamma}] + \underline{x}'[\underline{\beta} + \Delta\, E(\underline{I})] + \underline{x}'B\,\underline{x} \qquad (7.1.5)$$

where $E(\underline{I})$ is given in (7.1.4). Notice that this mean model is quite different from the mean model of (3.1.3) in the case of continuous noise factors. However, it can be estimated in a similar fashion.

## §7.2 Estimation of the Process Variance

The process variance model can be determined by the response-model of (7.1.3) and is given by

$$Var_{\underline{I}}(y) = (\underline{\gamma} + \Delta'\underline{x})'V(\underline{\gamma} + \Delta'\underline{x}) + \sigma_\epsilon^2 \qquad (7.2.1)$$

where V, the variance-covariance matrix of $\underline{I}$, is given in (7.1.4). Equation (7.2.1) can be rewritten as

$$\mathrm{Var}_{\underline{I}}(y) = \underline{l}'V\underline{l} + \sigma_\epsilon^2 \tag{7.2.2}$$

where $\underline{l} = (\underline{\gamma} + \Delta'\underline{x})$ consists of linear or polynomial functions of the control factors just as before.

It should be clear that an intuitive, but biased, estimator of the process variance is

$$\widehat{\mathrm{Var}}_{\underline{I}}(y) = \underline{\hat{l}}'V\underline{\hat{l}} + \hat{\sigma}_\epsilon^2. \tag{7.2.3}$$

An unbiased estimator of the process variance is given by

$$\widehat{\mathrm{Var}}_{\underline{I}}(y) = \underline{\hat{l}}'V\underline{\hat{l}} + [1 - \mathrm{tr}(CV)]\hat{\sigma}_\epsilon^2. \tag{7.2.4}$$

Note that all estimation depends on the fact that V, the variance-covariance matrix of $\underline{I}$, is known. This, in turn, requires that the success probabilities of the indicator variables, namely the $p_i$, be known as well. Often, it is not possible to know the true probabilities. As a result, estimates of the $p_i$ are used. These estimates may have a profound effect on the quality of estimating the process mean and process variance. It is not uncommon for engineers to have accurate estimates of the probabilities based on past experience. Since V is considered a known constant matrix, these probabilities should be estimated as accurately as possible. When this is accomplished there are three types of confidence regions which can be created to better understand the performance of the product.

## §7.3 One-Sided Confidence Region on $\frac{\underline{l}'V\underline{l} + \sigma_\epsilon^2}{\sigma_\epsilon^2}$

The first confidence region to be presented creates an approximate upper bound on $\frac{\underline{l}'V\underline{l} + \sigma_\epsilon^2}{\sigma_\epsilon^2}$. This region can be quite important to the engineer who is concerned about the ratio of the process variance to $\sigma_\epsilon^2$. In order to develop this

bound, the distribution of $\hat{\underline{l}}'V\,\hat{\underline{l}}$ must first be determined.

Consider the matrices $\sigma_\epsilon^2[\mathrm{Var}(\hat{\underline{l}})]^{-1} = C^{-1}$ and V. Since both matrices are symmetric and $C^{-1}$ is positive definite there exists a nonsingular matrix Q such that $Q'C^{-1}Q = I$ and $Q'VQ = D$, where D is a diagonal matrix whose diagonal elements are the eigenvalues of CV. Now, define $\hat{\underline{z}} = Q^{-1}\hat{\underline{l}}$. Then $\hat{\underline{l}}'V\,\hat{\underline{l}} = \hat{\underline{z}}'Q'VQ\,\hat{\underline{z}} = \hat{\underline{z}}'D\,\hat{\underline{z}}$ where $\hat{\underline{z}} \sim N(Q^{-1}\underline{l}\,,\, \sigma_\epsilon^2 I)$. And

$$\hat{\underline{z}}'D\,\hat{\underline{z}} = \sum_{j=1}^{r_z} d_{jj}\,\sigma_\epsilon^2\,\frac{\hat{z}_j^2}{\sigma_\epsilon^2} \tag{7.3.1}$$

where $d_{jj}$ is the $j^{th}$ diagonal element of D. The distribution of (7.3.1) is

$$\hat{\underline{z}}'D\,\hat{\underline{z}} \sim \sum_{j=1}^{r_z} d_{jj}\sigma_\epsilon^2\chi^{2'}_{(1,\lambda_j)} \tag{7.3.2}$$

where $\lambda_j = \frac{z_j^2}{\sigma_\epsilon^2}$. Equation (7.3.2) can be approximated, using a method by Patnaik (1949), as

$$\hat{\underline{z}}'D\,\hat{\underline{z}} \sim \sigma_\epsilon^2 \sum_{j=1}^{r_z} d_{jj}\rho_j\chi^2_{\nu_j} \tag{7.3.3}$$

where $\rho_j = \dfrac{1 + 2\lambda_j}{1 + \lambda_j}$ and $\nu_j = \dfrac{(1 + \lambda_j)^2}{1 + 2\lambda_j}$. A further approximation from Satterthwaite (1946) and Welch (1956) yields

$$\hat{\underline{z}}'D\,\hat{\underline{z}} \sim \sigma_\epsilon^2 d\,\chi^2_{\nu_2}$$

where $d = \dfrac{\sum_{j=1}^{r_z}(d_{jj}\rho_j)^2\nu_j}{\sum_{j=1}^{r_z} d_{jj}\rho_j\nu_j}$ and $\nu_2 = \dfrac{\left\{\sum_{j=1}^{r_z} d_{jj}\rho_j\nu_j\right\}^2}{\sum_{j=1}^{r_z}(d_{jj}\rho_j)^2\nu_j}$. (7.3.4)

Thus, we have the approximate distribution of $\hat{\underline{l}}'V\,\hat{\underline{l}}$ as $\sigma_\epsilon^2\, d\,\chi^2_{\nu_2}$. And, since $\hat{\sigma}_\epsilon^2 \sim \left(\frac{\sigma_\epsilon^2}{dfE}\right)\chi^2_{dfE}$,

$$\frac{\hat{\underline{l}}'V\,\hat{\underline{l}}}{d\nu_2\hat{\sigma}_\epsilon^2} \sim F_{(\nu_2,\ dfE)}\,. \tag{7.3.5}$$

An upper confidence limit can be formed using

$$P\left\{\frac{\hat{\underline{l}}'V\,\hat{\underline{l}}}{d\nu_2\hat{\sigma}_\epsilon^2} \geq F_{(\nu_2,\ dfE);1-\alpha}\right\} = 1 - \alpha \tag{7.3.6}$$

where $d\nu_2 = \sum_{j=1}^{r_z} d_{jj} + \frac{\underline{l}'V\,\underline{l}}{\sigma_\epsilon^2}$.

Notice that, through $d\nu_2$, large values of $\frac{\underline{l}'V\,\underline{l}}{\sigma_\epsilon^2}$ force the statistic to become small and not satisfy the inequality within the brackets of the probability statement in (7.3.6). To form the upper confidence limit choose a setting of the control factors, $\underline{x}_0$, and calculate the statistic $\frac{\hat{\underline{l}}'V\,\hat{\underline{l}}}{\hat{\sigma}_\epsilon^2}$. Set this value equal to $d\nu_2 F_{(\nu_2,\ dfE);1-\alpha}$ and solve for

$$d\nu_2 = \sum_{j=1}^{r_z} d_{jj} + \frac{\underline{l}'V\,\underline{l}}{\sigma_\epsilon^2}\,.$$

Then the approximate upper confidence limit is $(\zeta_u + 1)$ where $\zeta_u = (d\nu_2 - \sum_{j=1}^{r_z} d_{jj})$.

## §7.4 Lower One-Sided Tolerance Intervals on Future Observations

As in §6.3, tolerance intervals can be applied in the case of categorical noise factors to give bounds on future observations. The interpretation is that at least 100p% of the population or process being sampled is above the lower limit of $\hat{\mu}(\underline{x}) - k\,\hat{\sigma}_{\underline{l}}\,y(\underline{x})$ with $(1 - \gamma)100\%$ confidence. Definition 6.3.1 can be modified for the case of categorical noise factors. Consider Definition 7.4.1.

**Definition 7.4.1:** A p-content *lower* one-sided tolerance interval at confidence level $(1-\gamma)$ is denoted $[\hat{\mu} - k\,\hat{\sigma}_{\underline{I}}\,y(\underline{x}),\ \infty)$ if

$$P_{\hat{\mu},\hat{\sigma}_{\underline{I}}\,y(\underline{x})}[P_{y_f}\{y_f \geq \hat{\mu} - k\,\hat{\sigma}_{\underline{I}}\,y(\underline{x})\} \geq p] \geq 1-\gamma \tag{7.4.1}$$

where $\hat{\sigma}_{\underline{I}}\,y(\underline{x}) = \sqrt{\hat{Var}_{\underline{I}}\,y(\underline{x})}$ is the estimated process standard deviation in the presence of categorical noise factors. Probability statement (7.4.1) can be manipulated to form approximate bounds on a future observation in the presence of categorical noise factors. Details are presented in Appendix D, but the final result is

$$P_{\hat{\mu},\hat{\sigma}_{\underline{I}}\,y(\underline{x})}[\,t'_{\nu^*,\,\lambda} \leq m^*\,] \leq \gamma. \tag{7.4.2}$$

Equation (7.4.2) allows us to calculate k through the following procedure:

(i) Estimate all model terms in the response-model as well as $\sigma_\epsilon^2$.

(ii) Select probability levels p and $\gamma$.

(iii) Select a setting of the control factors, $\underline{x}_0$, and calculate $\underline{x}'_0(X'X)^{-1}\underline{x}_0$, $\hat{\mu}(\underline{x}_0)$, and $\hat{\sigma}_{\underline{I}}\,y(\underline{x}_0)$.

(iv) Estimate $\lambda$ and calculate $d^*$ and $\nu^*$.

(v) Using the estimate of $\lambda$ find the percentage point of the noncentral t, e.g. using SAS, found in Equation (7.4.2) and equate it to

$$m^* = \frac{-k\sqrt{d^*\nu^*}}{\sqrt{\underline{x}'_0(X'X)^{-1}\underline{x}_0}}\ .$$

(vi) Solve for k.

(vii) The lower tolerance limit at $\underline{x}_0$ is then $\hat{\mu}(\underline{x}_0) - k\,\hat{\sigma}_{\underline{I}}\,y(\underline{x}_0)$.

Note that the lower one-sided tolerance interval is used primarily when the engineer is interested in maximizing the response. If a minimum response is of interest an upper one-sided tolerance interval would be more appropriate. This interval can be created in a similar fashion as that found in Appendix D. Adjustments can be easily made. The details presented make use of the unbiased estimate of the process variance. If the biased estimate is to be used instead simply replace [1 - tr(CV)] with a 1 in all computations.

### §7.5 Prediction Limits on a Future Response

When an engineer desires probabilistic bounds on a new response at a fixed setting of the control factors prediction intervals play an important role. The categorical noise factors can be accommodated and only mildly affect the manipulations of the probability statements. As in §6.4, before the bounds are developed, we must first determine the distribution of $\underline{\hat{l}}\,'\mathrm{V}\,\underline{\hat{l}}$.

Recall the derivations of §7.3. In particular, by simultaneously diagonalizing the matrices $\mathrm{C}^{-1}$ and V the approximate distribution of $\underline{\hat{l}}\,'\mathrm{V}\,\underline{\hat{l}}$ was found to be

$$\underline{\hat{l}}\,'\mathrm{V}\,\underline{\hat{l}} \;\sim\; \sigma_\epsilon^2\,\mathrm{d}\,\chi^2_{\nu_2} \tag{7.5.1}$$

where d and $\nu_2$ are defined in Equation (7.3.4). Also, the distribution of $\hat{\sigma}_\epsilon^2$ is $\left(\frac{\sigma_\epsilon^2}{\mathrm{dfE}}\right)\chi^2_{dfE}$. And,

$$\hat{\sigma}_\epsilon^2[1 + \underline{\mathrm{x}}\,'_0(\mathrm{X}'\mathrm{X})^{-1}\underline{\mathrm{x}}_0] \;\sim\; \left(\frac{\sigma_\epsilon^2}{\mathrm{dfE}}\right)[1 + \underline{\mathrm{x}}\,'_0(\mathrm{X}'\mathrm{X})^{-1}\underline{\mathrm{x}}_0]\chi^2_{dfE} \tag{7.5.2}$$

By combining the distributions of (7.5.1) and (7.5.2) it can be shown that

$$\hat{\sigma}^2_{y\text{-}\hat{\mu}} = \underline{\hat{l}}\,'\mathrm{V}\,\underline{\hat{l}} + \hat{\sigma}_\epsilon^2[1 + \underline{\mathrm{x}}\,'_0(\mathrm{X}'\mathrm{X})^{-1}\underline{\mathrm{x}}_0] \;\sim \sigma_\epsilon^2\left\{\mathrm{d}\chi^2_{\nu_2} + \frac{1}{\mathrm{dfE}}[1 + \underline{\mathrm{x}}\,'_0(\mathrm{X}'\mathrm{X})^{-1}\underline{\mathrm{x}}_0]\chi^2_{dfE}\right\}$$

$$\sim \sigma_\epsilon^2 \, a \, \chi^2_{\nu_3} \tag{7.5.3}$$

$$\text{where } a = \frac{d^2\nu_2 + [\frac{1}{dfE}(1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0)]^2 dfE}{d\nu_2 + [\frac{1}{dfE}(1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0)]\, dfE} \text{ and} \tag{7.5.4}$$

$$\nu_3 = \frac{\{d\nu_2 + [\frac{1}{dfE}(1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0)]\, dfE\}^2}{d^2\nu_2 + [\frac{1}{dfE}(1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0)]^2 dfE} . \tag{7.5.5}$$

This approximation is due to Satterthwaite (1946) and Welch (1956).

The probability statement for forming a prediction interval is given by

$$P\ \{\ \hat{\mu}(\underline{x}_0) - k\ \hat{\sigma}_{y\text{-}\hat{\mu}}(\underline{x}_0)\ \le\ y_f(\underline{x}_0)\ \le\ \hat{\mu}(\underline{x}_0) + k\ \hat{\sigma}_{y\text{-}\hat{\mu}}(\underline{x}_0)\ \} = 1\text{-}\alpha \tag{7.5.6}$$

By isolating k,

$$P = \{\ -k\ \le\ \frac{y_f(\underline{x}_0) - \hat{\mu}(\underline{x}_0)}{\hat{\sigma}_\epsilon \sqrt{\frac{\hat{\underline{l}}'V\hat{\underline{l}}}{\hat{\sigma}_\epsilon^2} + 1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0}}\ \le\ k\ \} = 1 - \alpha \tag{7.5.7}$$

Note that $(y_f(\underline{x}_0) - \hat{\mu}(\underline{x}_0))$ can be standardized by dividing by $\sigma_\epsilon \sqrt{\frac{\underline{l}'V\underline{l}}{\sigma_\epsilon^2} + 1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0}$. Thus, denote

$$Z = \frac{y_f(\underline{x}_0) - \hat{\mu}(\underline{x}_0)}{\sigma_\epsilon \sqrt{\frac{\underline{l}'V\underline{l}}{\sigma_\epsilon^2} + 1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0}} \sim N(0, 1),$$

Also, the denominator of the statistic inside Equation (7.5.7) is distributed approximately $\sqrt{\sigma_\epsilon^2\ a\ \chi^2_{\nu_3}}$. Thus, Equation (7.5.7) can be rewritten as

$$P\left\{ -k \leq \frac{Z\sqrt{\frac{\underline{l}'V\underline{l}}{\sigma_\epsilon^2} + 1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0}}{\sqrt{a\,\chi^2_{\nu_3}}} \leq k \right\} = 1 - \alpha \qquad (7.5.8)$$

Further manipulations yield

$$P\left\{ -m \leq t_{\nu_3} \leq m \right\} = 1 - \alpha \qquad (7.5.9)$$

$$\text{where } m = \frac{k\sqrt{a\nu_3}}{\sqrt{\frac{\underline{l}'V\underline{l}}{\sigma_\epsilon^2} + 1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0}} \text{ and } t_{\nu_3} = \frac{Z}{\sqrt{\frac{\chi^2_{\nu_3}}{\nu_3}}}. \qquad (7.5.10)$$

In order to determine the estimated prediction limits, k must be calculated. Thus, the following procedure should be implemented:

(i) Estimate all model terms in the response-model as well as $\sigma_\epsilon^2$.

(ii) Select the probability $\alpha$.

(iii) Select fixed conditions on the control factors, $\underline{x}_0$, and calculate

$\underline{x}_0'(X'X)^{-1}\underline{x}_0$, $\hat{\mu}(\underline{x}_0)$, $\hat{\sigma}_{y\text{-}\hat{\mu}}$, and $\hat{\underline{l}}'V\hat{\underline{l}}$.

(iv) Estimate a and $\nu_3$ through estimates of d and $\nu_2$.

(v) Set m equal to the percentage point of $t_{\nu_3}$ and solve for k.

(vi) The prediction limits at $\underline{x}_0$ are then $[\hat{\mu}(\underline{x}_0) \pm k\,\hat{\sigma}_{y\text{-}\hat{\mu}}]$.

The engineer can say with (1 - $\alpha$) probability that a future response is contained by the interval at the specified location, $\underline{x}_0$. Prediction limits are an important approach for finding process operating conditions.

# Chapter 8
# Summary and Future Research

## §5.1 Summary and Future Research

In this dissertation several criteria were presented for aiding in the selection of a response-model which would yield an accurate process variance model. These criteria are quite similar to techniques used in standard regression. Cross-validation procedures, such as the PRESS statistic, could also be developed and used to select an appropriate response-model. Standard regression tools involving influence diagnostics could be adapted and used on the response-model to identify unusual data which may cause the process variance to produce inaccurate estimates. This would allow the practitioner to identify unusually large estimated process variances.

Another avenue for future research involves the various confidence regions discussed in Chapters 6 and 7. In particular, the development of the Tolerance Intervals could be extended to include two-sided limits. This development is more complicated than that for the one-sided limits and would be of interest to the engineer who is trying to attain a target response consistently throughout the process.

Research on optimal experimental designs which minimize the volume of the various confidence regions would be quite useful. These designs would be particularly useful in the case of categorical noise factors since they would likely depend on the values of the estimated probabilities. An interesting issue is how robust experimental designs are to the estimation of the probabilities in the variance-covariance matrix of the discrete noise factors, V. Similarly, another interesting issue is how robust the estimated process variance model is to the estimation of V in the case of discrete noise factors.

# Bibliography


Bartlett, M. S., and Kendall, D. G. (1946), "The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation," *Journal of the Royal Statistical Society*, Series B, Vol. 8, 128-138.

Box, G. E. P. (1985), "Discussion of Off-Line Quality Control, Parameter Design, and the Taguchi Method," *Journal of Quality Technology*, 17, 198-206.

---------- (1988), "Signal-to-Noise Ratios, Performance Criteria, and Transformations (with Discussion)," *Technometrics*, 30, 1-40.

---------- (1992), "Taguchi's Parameter Design: A Panel Discussion," *Technometrics*, 34, 2, 127-161.

Box, G. E. P. and Meyer, R. D. (1986), "Dispersion Effects From Fractional Designs," *Technometrics*, 28, 19-27.

Crowder, S. V., and Hamilton, M. D. (1992), "An EWMA for Monitoring a Process Standard Deviation," *Journal of Quality Technology*, 24, 12-21.

Farrell, O. J., and Ross, B. (1963), "*Solved Problems: Gamma and Beta Functions, Legendre Polynomials, Bessel Functions*," The Macmillan Company, New York.

Freeny, A. E., and Nair, V. N. (1992), "Robust Parameter Design with Uncontrolled Noise Variables," *Statistica Sinica*, Vol. 2, 313-334.

Gray, A., and Mathews, G. B. (1931), *A Treatise on Bessel Functions and Their Applications to Physics*, Macmillan and Co., Ltd, London.

Graybill, F. A. (1976), *An Introduction to Linear Statistical Models*, Boston: Duxbury.

Johnson, N. L., and Kotz, S. (1970a), *Distributions in Statistics: Discrete Distributions*, Boston: Houghton Mifflin Company.

---------- (1970b), *Distributions in Statistics: Continuous Univariate Distributions - 1*, Boston: Houghton Mifflin Company.

---------- (1970c), *Distributions in Statistics: Continuous Univariate Distributions - 2*, Boston: Houghton Mifflin Company.

Hoel, P. G., Port, S. C., and Stone, C. J. (1971), *Introduction to Probability Theory*, Boston: Houghton Mifflin Company.

Hunter, J. S. (1985), "Statistical Design Applied to Product Design," *Journal of Quality Technology*, 17, 4, 210-221.

Kackar, R. N. (1985), "Off-Line Quality Control, Parameter Design, and the Taguchi Method," *Journal of Quality Technology*, 17, 176-209.

Khattree, R. (1993), "Some Issues in Robust Parameter Designs: A Response Surface Approach," Technical Report No. 93-11, Department of Mathematical Sciences, Oakland University of Rochester, MI.

Kim, Y. (1992), "A Response Surface Approach to Data Analysis in Robust Parameter Design," Dissertation, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Lucas, J. M. (1985), Comment on "Off-Line Quality Control, Parameter Design, and the Taguchi Method," by R. N. Kackar, *Journal of Quality Technology*, 17, 195-197.

---------- (1990), "Achieving a Robust Process Using Response Surface Methodology," Technical Report, E. I. DuPont de Nemours Company, Wilmington, DE.

Mallows, C. L. (1973), "Some Comments on Cp," *Technometrics*, 15, 661-675.

Montgomery, D. C. (1984), *Design and Analysis of Experiments*, 2nd Edition, John Wiley and Sons, New York.

---------- (1990), "Using Fractional Factorial Designs for Robust Process Development," *Quality Engineering*, 3, 193-205.

---------- (1991), *Design and Analysis of Experiments*, 3rd Edition, John Wiley and Sons, New York.

Myers, R. H. (1976), *Response Surface Methodology*, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA.

---------- (1990), *Classical and Modern Regression and Applications*, PWS-KENT Publishing Company, Boston, MA.

---------- (1992), "Response Surface Methodology in Quality Improvement," Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA. (To be published in 1995 by Wiley)

Myers, R. H., Khuri, A. I., and Carter, W. H. (1989), "Response Surface Methodology: 1966-1988," *Technometrics*, 31, 2, 137-157.

Myers, R. H., Khuri, A. I., and Vining, G. G. (1992), "Response Surface Alternatives to the Taguchi Robust Parameter Design Approach," *The American Statistician*, 46, 2, 131-139.

Myers, R. H., and Kim, Y. (1995), "Response Surface Methods and the Use of Noise Variables," *Technometrics* (to appear).

Myers, R. H., and Milton, J. S. (1991), "*A First Course in the Theory of Linear Statistical Models*," PWS-KENT Publishing Company, Boston, MA.

Nair, V. N. (1992), "Taguchi's Parameter Design: A Panel Discussion," *Technometrics*, 34, 2, 127-161.

Nair, V. N. and Pregibon, D. (1988), "Analyzing Dispersion Effects From Replicated Factorial Experiments," *Technometrics*, 30, 247-257.

Nelder, J. A., and Lee, Y. (1991), "Generalized Linear Models for the Analysis of Taguchi-type Experiments," *Applied Stochastic Models and Data Analysis*, 7, 107-120.

Patnaik, P. B. (1949), "The Non-Central $\chi^2$- and F-Distributions and Their Applications," *Biometrika* 36, 202-232.

Pignatiello, J. J., and Ramberg, J. S. (1985), "Discussion of Off-Line Quality Control, Parameter Design, and the Taguchi Methods," *Journal of Quality Technology*, 17, 198-206.

Satterthwaite, F. (1941), "Synthesis of Variance," *Psychometrika*, Vol. 6, No. 5, 309-316.

---------- (1946), "An Approximate Distribution of Estimates of Variance Components," *Biometric Bulletin* 2, 110-114.

Searle, S. R. (1971), *Linear Models*, John Wiley and Sons, New York.

Shoemaker, A. C., and Tsui, K. L., (1992), "Taguchi's Parameter Design: A Panel Discussion," *Technometrics*, 34, 2, 127-161.

Shoemaker, A. C., Tsui, K. L., and Wu, C. F. J. (1991), "Economical Experimentation Methods for Robust Design," *Technometrics*, 33, 415-427.

Taguchi, G. and Wu, Y. (1985), "Introduction to Off-Line Quality Control," Central Japan Quality Control Association, Nagoya, Japan.

Vining, G. G., and Myers, R. H. (1990), "Combining Taguchi and Response Surface Philosophies: A Dual Response Approach," *Journal of Quality Technology*, 22, 38-45.

Welch, B. L. (1956), "On Linear Combinations of Several Variances," *Journal of the American Statistical Association* 51, 132-148.

Welch, W. J., Yu, T. K., Kang, S. M., and Sacks, J. (1990), "Computer Experiments for Quality Control by Parameter Design," *Journal of Quality Technology*, 22, 15-22.

Wu, C. F. J. (1992), "Taguchi's Parameter Design: A Panel Discussion," *Technometrics*, 34, 2, 127-161.

# Appendix A. Computation of AVGMSE(B) and AVGMSE(U)

Assume the true response-model is

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'B\,\underline{x} + \underline{z}'\underline{\gamma} + \underline{x}'\Delta\,\underline{z} + \epsilon \tag{A.1}$$

and its corresponding process variance model is

$$\mathrm{Var}_{\underline{z}}(y) = \underline{l}'\underline{l} + \sigma_\epsilon^2. \tag{A.2}$$

Also, assume the response-model to be fit is

$$y = \beta_0 + \underline{x}'\underline{\beta}^* + \underline{x}'B^*\,\underline{x} + \underline{z}'\underline{\gamma}^* + \underline{x}'\Delta^*\underline{z} + \epsilon \tag{A.3}$$

where $*$ indicates that one or more parameters may be left out of the response-model. The corresponding process variance model to Equation (A.3) is given by

$$\mathrm{Var}_{\underline{z}}(y) = \underline{l}^{*'}\underline{l}^* + \sigma_\epsilon^2. \tag{A.4}$$

The estimates of this process variance model are

$$\mathrm{PV}_B = \hat{\underline{l}}^{*'}\hat{\underline{l}}^* + \hat{\sigma}_\epsilon^2 \tag{A.5}$$

and

$$\mathrm{PV}_U = \hat{\underline{l}}^{*'}\hat{\underline{l}}^* + [1 - \mathrm{tr}(C^*)]\,\hat{\sigma}_\epsilon^2. \tag{A.6}$$

Then, assuming an orthogonal design, it can be shown that

$$\mathrm{AVGMSE(B)} = \frac{1}{k\sigma_\epsilon^4}\int_R \mathrm{MSE}(\mathrm{PV}_B)\,d\underline{x}$$

$$= 2\,r_z\mathrm{Var}^2(\hat{\gamma}_j^*) + \frac{4}{3}\,r_z\mathrm{Var}(\hat{\gamma}_j^*)\,\mathrm{tr}\{\mathrm{Var}(\hat{\Delta}_j^{*\prime})\}$$

$$+ 2\, r_z\left(\tfrac{1}{5} + \tfrac{r_x - 1}{9}\right) \mathrm{tr}\{\mathrm{Var}^2(\hat{\Delta}_j^{*\prime})\} + \tfrac{1}{\sigma_\epsilon^2}\Big\{4\, \mathrm{Var}(\hat{\gamma}_j^*)\underline{\gamma}^{*\prime}\underline{\gamma}^*$$

$$+ \tfrac{4}{3}\, \mathrm{Var}(\hat{\gamma}_j^*)\, \mathrm{tr}\{\Delta^*\Delta^{*\prime}\} + \tfrac{4}{3}\, \underline{\gamma}^{*\prime}\underline{\gamma}^*\, \mathrm{tr}\{\mathrm{Var}(\hat{\Delta}_j^{*\prime})\}$$

$$+ 4\left(\tfrac{1}{5} + \tfrac{r_x - 1}{9}\right) \mathrm{tr}\{\mathrm{Var}(\hat{\Delta}_j^{*\prime})\Delta^*\Delta^{*\prime}\}\Big\} + \frac{2}{\mathrm{dfE}}$$

$$+ \tfrac{1}{\sigma_\epsilon^4}\Big\{\underline{\gamma}^{*\prime}\underline{\gamma}^*\, \underline{\gamma}^{*\prime}\underline{\gamma}^* - 2\, \underline{\gamma}^{*\prime}\underline{\gamma}^*\underline{\gamma}'\underline{\gamma} + \underline{\gamma}'\underline{\gamma}\ \underline{\gamma}'\underline{\gamma} + \tfrac{4}{3}\, \mathrm{tr}\{\Delta^*\underline{\gamma}^*\underline{\gamma}^{*\prime}\Delta^{*\prime}\}$$

$$- \tfrac{8}{3}\, \mathrm{tr}\{\Delta\underline{\gamma}\,\underline{\gamma}^{*\prime}\Delta^{*\prime}\} + \tfrac{4}{3}\, \mathrm{tr}\{\Delta\underline{\gamma}\,\underline{\gamma}'\Delta'\} + \left(\tfrac{1}{5} + \tfrac{r_x - 1}{9}\right) \mathrm{tr}\{\Delta^*\Delta^{*\prime}\ \Delta^*\Delta^{*\prime}\}$$

$$- 2\left(\tfrac{1}{5} + \tfrac{r_x - 1}{9}\right) \mathrm{tr}\{\Delta\Delta'\ \Delta^*\Delta^{*\prime}\} + \left(\tfrac{1}{5} + \tfrac{r_x - 1}{9}\right) \mathrm{tr}\{\Delta\Delta'\ \Delta\Delta'\}$$

$$+ r_z^2\, \mathrm{Var}^2(\hat{\gamma}_j^*) + r_z^2\left(\tfrac{1}{5} + \tfrac{r_x - 1}{9}\right) \mathrm{tr}\{\mathrm{Var}^2(\hat{\Delta}_j^{*\prime})\} + \tfrac{2}{3}\, \underline{\gamma}^{*\prime}\underline{\gamma}^*\, \mathrm{tr}\{\Delta^*\Delta^{*\prime}\}$$

$$- \tfrac{2}{3}\, \underline{\gamma}^{*\prime}\underline{\gamma}^*\, \mathrm{tr}\{\Delta\Delta'\} - \tfrac{2}{3}\, \underline{\gamma}'\underline{\gamma}\ \mathrm{tr}\{\Delta^*\Delta^{*\prime}\} + \tfrac{2}{3}\, \underline{\gamma}'\underline{\gamma}\ \mathrm{tr}\{\Delta\Delta'\}$$

$$+ 2\, r_z\, \underline{\gamma}^{*\prime}\underline{\gamma}^*\, \mathrm{Var}(\hat{\gamma}_j^*) - 2\, r_z\, \underline{\gamma}'\underline{\gamma}\, \mathrm{Var}(\hat{\gamma}_j^*) + \tfrac{2}{3}\, r_z\, \underline{\gamma}^{*\prime}\underline{\gamma}^*\, \mathrm{tr}\{\mathrm{Var}(\hat{\Delta}_j^{*\prime})\}$$

$$- \tfrac{2}{3}\, r_z\, \underline{\gamma}'\underline{\gamma}\ \mathrm{tr}\{\mathrm{Var}(\hat{\Delta}_j^{*\prime})\} + \tfrac{2}{3}\, r_z\, \mathrm{Var}(\hat{\gamma}_j^*)\, \mathrm{tr}\{\Delta^*\Delta^{*\prime}\} - \tfrac{2}{3}\, r_z \mathrm{Var}(\hat{\gamma}_j^*)\, \mathrm{tr}\{\Delta\Delta'\}$$

$$+ 2\, r_z\left(\tfrac{1}{5} + \tfrac{r_x - 1}{9}\right) \mathrm{tr}\{\mathrm{Var}(\hat{\Delta}_j^{*\prime})\ \Delta^*\Delta^{*\prime}\}$$

$$- 2\, r_z\left(\tfrac{1}{5} + \tfrac{r_x - 1}{9}\right) \mathrm{tr}\{\mathrm{Var}(\hat{\Delta}_j^{*\prime})\ \Delta\Delta'\}$$

$$+ \tfrac{2}{3}\, r_z^2\, \mathrm{Var}(\hat{\gamma}_j^*)\, \mathrm{tr}\{\mathrm{Var}(\hat{\Delta}_j^{*\prime})\}\Big\} \qquad \text{(A.7)}$$

where $\hat{\Delta}_j^{*\prime}$ = j$^{\text{th}}$ row of $\hat{\Delta}$ corresponding to the j$^{\text{th}}$ noise factor where one or more parameters, $\delta_{ij}$, may be left out. All other terms have been defined previously in Chapter 3.

It can also be shown that

$$\text{AVGMSE(U)} = \frac{1}{\text{k}\sigma_\epsilon^4} \int_R \text{MSE}(\text{PV}_U)\, \text{d}\underline{\text{x}}$$

$$= 2\, \text{r}_z \text{Var}^2(\hat{\gamma}_j^*) + \tfrac{4}{3}\, \text{r}_z \text{Var}(\hat{\gamma}_j^*)\, \text{tr}\{\text{Var}(\hat{\Delta}_j^{*\prime})\}$$

$$+ 2\, \text{r}_z \left( \tfrac{1}{5} + \tfrac{r_x - 1}{9} \right) \text{tr}\{\text{Var}^2(\hat{\Delta}_j^{*\prime})\} + \frac{1}{\sigma_\epsilon^2} \Big\{ 4\, \text{Var}(\hat{\gamma}_j^*) \underline{\gamma}^{*\prime} \underline{\gamma}^*$$

$$+ \tfrac{4}{3}\, \text{Var}(\hat{\gamma}_j^*)\, \text{tr}\{\Delta^* \Delta^{*\prime}\} + \tfrac{4}{3}\, \underline{\gamma}^{*\prime} \underline{\gamma}^*\, \text{tr}\{\text{Var}(\hat{\Delta}_j^{*\prime})\}$$

$$+ 4 \left( \tfrac{1}{5} + \tfrac{r_x - 1}{9} \right) \text{tr}\{\text{Var}(\hat{\Delta}_j^{*\prime}) \Delta^* \Delta^{*\prime}\} \Big\}$$

$$+ \frac{2}{\text{dfE}} \Big\{ 1 + \text{r}_z^2\, \text{Var}^2(\hat{\gamma}_j^*) + \text{r}_z^2 \left( \tfrac{1}{5} + \tfrac{r_x - 1}{9} \right) \text{tr}\{\text{Var}^2(\hat{\Delta}_j^{*\prime})\}$$

$$- 2\, \text{r}_z \text{Var}(\hat{\gamma}_j^*) - \tfrac{2}{3}\, \text{r}_z\, \text{tr}\{\text{Var}(\hat{\Delta}_j^{*\prime})\} + \tfrac{2}{3}\, \text{r}_z^2\, \text{Var}\,(\hat{\gamma}_j^*)\, \text{tr}\{\text{Var}(\hat{\Delta}_j^{*\prime})\} \Big\}$$

$$+ \frac{1}{\sigma_\epsilon^4} \Big\{ \underline{\gamma}^{*\prime} \underline{\gamma}^*\, \underline{\gamma}^{*\prime} \underline{\gamma}^* - 2\, \underline{\gamma}^{*\prime} \underline{\gamma}^* \underline{\gamma}' \underline{\gamma} + \underline{\gamma}' \underline{\gamma}\, \underline{\gamma}' \underline{\gamma} + \tfrac{4}{3}\, \text{tr}\{\Delta^* \underline{\gamma}^* \underline{\gamma}^{*\prime} \Delta^{*\prime}\}$$

$$- \tfrac{8}{3}\, \text{tr}\{\Delta \underline{\gamma}\, \underline{\gamma}^{*\prime} \Delta^{*\prime}\} + \tfrac{4}{3}\, \text{tr}\{\Delta \underline{\gamma}\, \underline{\gamma}' \Delta'\} + \left( \tfrac{1}{5} + \tfrac{r_x - 1}{9} \right) \text{tr}\{\Delta^* \Delta^{*\prime} \Delta^* \Delta^{*\prime}\}$$

$$- 2 \left( \tfrac{1}{5} + \tfrac{r_x - 1}{9} \right) \text{tr}\{\Delta \Delta' \Delta^* \Delta^{*\prime}\} + \left( \tfrac{1}{5} + \tfrac{r_x - 1}{9} \right) \text{tr}\{\Delta \Delta' \Delta \Delta'\}$$

$$+ \text{r}_z^2\, \text{Var}^2(\hat{\gamma}_j^*) + \text{r}_z^2 \left( \tfrac{1}{5} + \tfrac{r_x - 1}{9} \right) \text{tr}\{\text{Var}^2(\hat{\Delta}_j^{*\prime})\} + \sigma_\epsilon^4 \text{r}_z^2 \text{Var}^2(\hat{\gamma}_j^*)$$

$$+ \left( \tfrac{1}{5} + \tfrac{r_x - 1}{9} \right) \sigma_\epsilon^4 \, r_z^2 \, \mathrm{tr}\{\mathrm{Var}^2(\hat{\Delta}_j^{*\prime})\} + \tfrac{2}{3} \underline{\gamma}^{*\prime} \underline{\gamma}^* \, \mathrm{tr}\{\Delta^* \Delta^{*\prime}\}$$

$$- \tfrac{2}{3} \underline{\gamma}^{*\prime} \underline{\gamma}^* \, \mathrm{tr}\{\Delta \Delta'\} - \tfrac{2}{3} \underline{\gamma}' \underline{\gamma} \; \mathrm{tr}\{\Delta^* \Delta^{*\prime}\} + \tfrac{2}{3} \underline{\gamma}' \underline{\gamma} \; \mathrm{tr}\{\Delta \Delta'\}$$

$$+ 2 \, r_z \, \underline{\gamma}^{*\prime} \underline{\gamma}^* \, \mathrm{Var}(\hat{\gamma}_j^*) - 2 \, r_z \, \underline{\gamma}' \underline{\gamma} \, \mathrm{Var}(\hat{\gamma}_j^*)$$

$$+ \tfrac{2}{3} \, r_z \, \underline{\gamma}^{*\prime} \underline{\gamma}^* \, \mathrm{tr}\{\mathrm{Var}(\hat{\Delta}_j^{*\prime})\} - \tfrac{2}{3} \, r_z \, \underline{\gamma}' \underline{\gamma} \; \mathrm{tr}\{\mathrm{Var}(\hat{\Delta}_j^{*\prime})\}$$

$$+ 2 \, \sigma_\epsilon^2 \, r_z \mathrm{Var}(\hat{\gamma}_j^*) \underline{\gamma}^{*\prime} \underline{\gamma}^* - 2 \, \sigma_\epsilon^2 \, r_z \, \mathrm{Var}(\hat{\gamma}_j^*) \, \underline{\gamma}' \underline{\gamma}$$

$$+ \tfrac{2}{3} \, \sigma_\epsilon^2 \, r_z \, \underline{\gamma}^{*\prime} \underline{\gamma}^* \, \mathrm{tr}\{\mathrm{Var}(\hat{\Delta}_j^{*\prime})\} - \tfrac{2}{3} \, \sigma_\epsilon^2 \, r_z \, \underline{\gamma}' \underline{\gamma} \, \mathrm{tr}\{\mathrm{Var}(\hat{\Delta}_j^{*\prime})\}$$

$$+ \tfrac{2}{3} \, r_z \, \mathrm{Var}(\hat{\gamma}_j^*) \, \mathrm{tr}\{\Delta^* \Delta^{*\prime}\} - \tfrac{2}{3} \, r_z \, \mathrm{Var}(\hat{\gamma}_j^*) \, \mathrm{tr}\{\Delta \Delta'\}$$

$$+ 2 \, r_z \left( \tfrac{1}{5} + \tfrac{r_x - 1}{9} \right) \mathrm{tr}\{\mathrm{Var}(\hat{\Delta}_j^{*\prime}) \, \Delta^* \Delta^{*\prime}\}$$

$$- \; 2 \, r_z \left( \tfrac{1}{5} + \tfrac{r_x - 1}{9} \right) \mathrm{tr}\{\mathrm{Var}(\hat{\Delta}_j^{*\prime}) \, \Delta \Delta'\}$$

$$+ \tfrac{2}{3} \, \sigma_\epsilon^2 \, r_z \mathrm{Var}(\hat{\gamma}_j^*) \, \mathrm{tr}\{\Delta^* \Delta^{*\prime}\} - \tfrac{2}{3} \, \sigma_\epsilon^2 \, r_z \mathrm{Var}(\hat{\gamma}_j^*) \, \mathrm{tr}\{\Delta \Delta'\}$$

$$+ 2 \, \sigma_\epsilon^2 \, r_z \left( \tfrac{1}{5} + \tfrac{r_x - 1}{9} \right) \mathrm{tr}\{\mathrm{Var}(\hat{\Delta}_j^{*\prime}) \Delta^* \Delta^{*\prime}\}$$

$$- \; 2 \, \sigma_\epsilon^2 \, r_z \left( \tfrac{1}{5} + \tfrac{r_x - 1}{9} \right) \mathrm{tr}\{\mathrm{Var}(\hat{\Delta}_j^{*\prime}) \Delta \Delta'\}$$

$$+ \tfrac{2}{3} \, r_z^2 \, \mathrm{Var}(\hat{\gamma}_j^*) \, \mathrm{tr}\{\mathrm{Var}(\hat{\Delta}_j^{*\prime})\} + 2 \, \sigma_\epsilon^2 \, r_z^2 \mathrm{Var}^2(\hat{\gamma}_j^*)$$

$$+ \frac{4}{3}\,\sigma_\epsilon^2\,\mathrm{r}_z^2 \mathrm{Var}(\widehat{\gamma}_j^*)\,\mathrm{tr}\{\mathrm{Var}(\widehat{\Delta}_j^{*\prime})\} + 2\,\sigma_\epsilon^2\,\mathrm{r}_z^2\,(\,\frac{1}{5} + \frac{r_x - 1}{9}\,)\,\mathrm{tr}\{\mathrm{Var}^2(\widehat{\Delta}_j^{*\prime})\}$$

$$+ \frac{2}{3}\,\sigma_\epsilon^4\,\mathrm{r}_z^2\,\mathrm{Var}(\widehat{\gamma}_j^*)\,\mathrm{tr}\{\mathrm{Var}(\widehat{\Delta}_j^{*\prime})\}\Big\} \qquad \text{(A.8)}$$

These equations can be easily programmed into a computer and calculations can be made in only a few seconds for several process variance models.

# Appendix B. Properties of $s^2(\underline{x})$

## §B.1 Distribution of $s^2(\underline{x})$

The calculation of the sample variance at a particular setting of the control factors, $s^2(\underline{x})$, is given by $\frac{1}{2^{r_z}} \sum_{\underline{z}} (y - \bar{y})^2$. This can also be expressed in matrix notation as

$$s^2(\underline{x}) = \underline{y}' \left[ I - \frac{\underline{j}\,\underline{j}'}{2^{r_z}} \right] \underline{y} \tag{B.1.1}$$

where $\underline{y}$ is a $2^{r_z}$ x 1 vector of responses observed at a particular setting of $\underline{x}$,

I is a $2^{r_z}$ x $2^{r_z}$ identity matrix,

and $\underline{j}$ is a $2^{r_z}$ x 1 vector of 1's.

The distribution of $s^2(\underline{x})$ is given by

$$\frac{\sigma_\epsilon^2}{2^{r_z}} \chi^{2'} \left( 2^{r_z} - 1, \frac{\underline{\mu}' \left[ I - \frac{\underline{j}\,\underline{j}'}{2^{r_z}} \right] \underline{\mu}}{\sigma_\epsilon^2} \right) \tag{B.1.2}$$

where $\underline{\mu}$ represents the vector of true responses at $\underline{x}$. After algebraic simplification, the noncentrality parameter can be expressed as

$$\frac{\underline{\mu}' \left[ I - \frac{\underline{j}\,\underline{j}'}{2^{r_z}} \right] \underline{\mu}}{\sigma_\epsilon^2} = \frac{2^{r_z}\, \underline{l}'\underline{l}}{\sigma_\epsilon^2} . \tag{B.1.3}$$

Thus, the distribution of $s^2(\underline{x})$ can be written

$$\frac{\sigma_\epsilon^2}{2^{r_z}} \chi^{2'} \left( 2^{r_z} - 1, \frac{2^{r_z}\, \underline{l}'\underline{l}}{\sigma_\epsilon^2} \right) \tag{B.1.4}$$

## §B.2 Equality of $s^2(\underline{x})$ and $PV_B$ when the Response-Model is Saturated

Consider an experiment which consists of 2 control factors and 2 noise factors. All factors have only two levels and the noise factors are assumed to be independent and coded to $\pm 1$ corresponding to $\pm\sigma_z$ units (assume $\sigma_{z_1} = \sigma_{z_2}$). The fitted response-model being considered is saturated and is given by

$$\begin{aligned}\hat{y} = {} & \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_{12} x_1 x_2 + \hat{\gamma}_1 z_1 + \hat{\gamma}_2 z_2 + \hat{\gamma}_{12} z_1 z_2 \\ & + \hat{\delta}_{11} x_1 z_1 + \hat{\delta}_{12} x_1 z_2 + \hat{\delta}_{21} x_2 z_1 + \hat{\delta}_{22} x_2 z_2 \\ & + \hat{\eta}_{121} x_1 x_2 z_1 + \hat{\eta}_{122} x_1 x_2 z_2 + \hat{\zeta}_{112} x_1 z_1 z_2 + \hat{\zeta}_{212} x_2 z_1 z_2 + \hat{\phi}_{1212} x_1 x_2 z_1 z_2. \end{aligned} \quad \text{(B.2.1)}$$

Using this response-model, the biased estimate of the process variance is given by

$$\begin{aligned} PV_B = {} & (\hat{\gamma}_1^2 + \hat{\gamma}_2^2 + \hat{\gamma}_{12}^2) + 2\,(\hat{\gamma}_1\hat{\delta}_{11} + \hat{\gamma}_2\hat{\delta}_{12} + \hat{\gamma}_{12}\hat{\zeta}_{112})\, x_1 \\ & + 2\,(\hat{\gamma}_1\hat{\delta}_{21} + \hat{\gamma}_2\hat{\delta}_{22} + \hat{\gamma}_{12}\hat{\zeta}_{212})\, x_2 \\ & + 2\,(\hat{\delta}_{11}\hat{\delta}_{21} + \hat{\delta}_{12}\hat{\delta}_{22} + \hat{\zeta}_{112}\hat{\zeta}_{212} + \hat{\gamma}_1\hat{\eta}_{121} + \hat{\gamma}_2\hat{\eta}_{122} + \hat{\gamma}_{12}\hat{\phi}_{1212})\, x_1 x_2 \\ & + (\hat{\delta}_{11}^2 + \hat{\delta}_{12}^2 + \hat{\zeta}_{112}^2)\, x_1^2 + (\hat{\delta}_{21}^2 + \hat{\delta}_{22}^2 + \hat{\zeta}_{212}^2)\, x_2^2 \\ & + (\hat{\eta}_{121}^2 + \hat{\eta}_{122}^2 + \hat{\phi}_{1212}^2)\, x_1^2 x_2^2 \\ & + 2\,(\hat{\zeta}_{112}\hat{\phi}_{1212} + \hat{\delta}_{11}\hat{\eta}_{121} + \hat{\delta}_{12}\hat{\eta}_{122})\, x_1^2 x_2 \\ & + 2\,(\hat{\zeta}_{212}\hat{\phi}_{1212} + \hat{\delta}_{21}\hat{\eta}_{121} + \hat{\delta}_{22}\hat{\eta}_{122})\, x_1 x_2^2\,. \end{aligned} \quad \text{(B.2.2)}$$

The calculation of $s^2(\underline{x})$ is given by

$$
\begin{aligned}
s^2(\underline{x}) &= \frac{1}{2^r z} \sum_{\underline{z}} (y - \bar{y})^2 \\
&= \frac{1}{2^r z} \sum_{\underline{z}} (y - [\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_{12} x_1 x_2])^2 \\
&= \frac{1}{2^r z} \sum_{\underline{z}} (\hat{\gamma}_1 z_1 + \hat{\gamma}_2 z_2 + \hat{\gamma}_{12} z_1 z_2 + \hat{\delta}_{11} x_1 z_1 + \hat{\delta}_{12} x_1 z_2 + \hat{\delta}_{21} x_2 z_1 + \hat{\delta}_{22} x_2 z_2 \\
&\quad + \hat{\eta}_{121} x_1 x_2 z_1 + \hat{\eta}_{122} x_1 x_2 z_2 + \hat{\zeta}_{112} x_1 z_1 z_2 + \hat{\zeta}_{212} x_2 z_1 z_2 + \hat{\phi}_{1212} x_1 x_2 z_1 z_2)^2 \\
&\vdots \\
&= (\hat{\gamma}_1^2 + \hat{\gamma}_2^2 + \hat{\gamma}_{12}^2) + 2 (\hat{\gamma}_1 \hat{\delta}_{11} + \hat{\gamma}_2 \hat{\delta}_{12} + \hat{\gamma}_{12} \hat{\zeta}_{112}) x_1 \\
&\quad + 2 (\hat{\gamma}_1 \hat{\delta}_{21} + \hat{\gamma}_2 \hat{\delta}_{22} + \hat{\gamma}_{12} \hat{\zeta}_{212}) x_2 \\
&\quad + 2 (\hat{\delta}_{11} \hat{\delta}_{21} + \hat{\delta}_{12} \hat{\delta}_{22} + \hat{\zeta}_{112} \hat{\zeta}_{212} + \hat{\gamma}_1 \hat{\eta}_{121} + \hat{\gamma}_2 \hat{\eta}_{122} + \hat{\gamma}_{12} \hat{\phi}_{1212}) x_1 x_2 \\
&\quad + (\hat{\delta}_{11}^2 + \hat{\delta}_{12}^2 + \hat{\zeta}_{112}^2) x_1^2 + (\hat{\delta}_{21}^2 + \hat{\delta}_{22}^2 + \hat{\zeta}_{212}^2) x_2^2 \\
&\quad + (\hat{\eta}_{121}^2 + \hat{\eta}_{122}^2 + \hat{\phi}_{1212}^2) x_1^2 x_2^2 \\
&\quad + 2 (\hat{\zeta}_{112} \hat{\phi}_{1212} + \hat{\delta}_{11} \hat{\eta}_{121} + \hat{\delta}_{12} \hat{\eta}_{122}) x_1^2 x_2 \\
&\quad + 2 (\hat{\zeta}_{212} \hat{\phi}_{1212} + \hat{\delta}_{21} \hat{\eta}_{121} + \hat{\delta}_{22} \hat{\eta}_{122}) x_1 x_2^2 . \qquad \text{(B.2.3)}
\end{aligned}
$$

The step which is left out carries out the square and sums across all settings of $\underline{z}$. Terms containing odd powers of $z_1$ and/or $z_2$ sum to zero. Terms that are left contain even powers of $z_1$ and/or $z_2$. Once these are summed across the settings of $\underline{z}$ they then equal 1. And, the final product of $s^2(\underline{x})$ equals $PV_B$.

# Appendix C. Approximate Distributions Required for the Various Confidence Regions in the Case of the Extended Response-Model

## §C.1 Confidence Region on $\frac{\underline{l}'\underline{l} + \sigma_\epsilon^2}{\sigma_\epsilon^2}$

Calculation of this region requires the approximate distribution of

$$\hat{\underline{l}}'\hat{\underline{l}} + 2\sum_{j=1}^{r_z}\hat{\psi}_{jj}^2 + \sum\sum_{j<k}\hat{\psi}_{jk}^2 \tag{C.1.1}$$

The exact distribution of $\hat{\underline{l}}'\hat{\underline{l}}$ is known from (6.2.1) as

$$\hat{\underline{l}}'\hat{\underline{l}} = \sum_{j=1}^{r_z}\hat{l}_j^2 \sim \sum_{j=1}^{r_z}\chi^{2'}(1, \lambda_j)\,\sigma_\epsilon^2 c_j \tag{C.1.2}$$

where $\lambda_j = \frac{l_j^2}{\sigma_\epsilon^2 c_j}$ . Similarly, $2\sum_{j=1}^{r_z}\hat{\psi}_{jj}^2 + \sum\sum_{j<k}\hat{\psi}_{jk}^2$ has the exact distribution

$$2\sum_{j=1}^{r_z}\chi^{2'}(1, \lambda_{jj})\sigma_\epsilon^2\mathrm{Var}(\hat{\psi}_{jj}) + \sum\sum_{j<k}\chi^{2'}(1, \lambda_{jk})\sigma_\epsilon^2\mathrm{Var}(\hat{\psi}_{jk}) \tag{C.1.3}$$

where $\lambda_{jj} = \frac{\psi_{jj}^2}{\sigma_\epsilon^2\mathrm{Var}(\hat{\psi}_{jj})}$ . ($\mathrm{Var}(\hat{\psi}_{jj})$ and $\mathrm{Var}(\hat{\psi}_{jk})$ are apart from $\sigma_\epsilon^2$.) These exact distributions need to be approximated by one distribution. To begin, recall from §6.2 the approximate distribution of

$$\hat{\underline{l}}'\hat{\underline{l}} \sim \sigma_\epsilon^2 d\,\chi_\nu^2 \tag{C.1.4}$$

where $d = \dfrac{\sum_{j=1}^{r_z} (c_j \rho_j)^2 \nu_j}{\sum_{j=1}^{r_z} c_j \rho_j \nu_j}$ and $\nu = \dfrac{\left(\sum_{j=1}^{r_z} \sigma_\epsilon^2 c_j \rho_j \nu_j\right)^2}{\sum_{j=1}^{r_z} (\sigma_\epsilon^2 c_j \rho_j)^2 \nu_j}$. This approximation is due to Patnaik (1949) and Satterthwaite (1946) and Welch (1956). Similarly, the distribution of $2\sum_{j=1}^{r_z} \hat{\psi}_{jj}^2 + \sum\sum_{j<k} \hat{\psi}_{jk}^2$ can be approximated by

$$2\sum_{j=1}^{r_z} \sigma_\epsilon^2 \mathrm{Var}(\hat{\psi}_{jj})\, k^* \chi^2{}_{\nu_{jj}} + \sum\sum_{j<k} \sigma_\epsilon^2 \mathrm{Var}(\hat{\psi}_{jk}) k^{**} \chi^2{}_{\nu_{jk}} \qquad \text{(C.1.5)}$$

using Patnaik's (1949) approximation where

$$k^* = \frac{1 + 2\lambda_{jj}}{1 + \lambda_{jj}}, \quad \nu_{jj} = \frac{(1 + \lambda_{jj})^2}{1 + 2\lambda_{jj}}, \quad k^{**} = \frac{1 + 2\lambda_{jk}}{1 + \lambda_{jk}}, \quad \text{and } \nu_{jk} = \frac{(1 + \lambda_{jk})^2}{1 + 2\lambda_{jk}}.$$

Furthermore, this sum of approximations can be approximated by

$$\hat{\underline{l}}'\hat{\underline{l}} + 2\sum_{j=1}^{r_z} \hat{\psi}_{jj}^2 + \sum\sum_{j<k} \hat{\psi}_{jk}^2 \sim \sigma_\epsilon^2 K \chi_\nu^2 \qquad \text{(C.1.6)}$$

using an approximation due to Satterthwaite (1946) and Welch (1956) where

$$K = \frac{d^2\nu_2 + \sum_{j=1}^{r_z} \left\{\mathrm{Var}(\hat{\psi}_{jj})\, k^*\right\}^2 \nu_{jj} + \sum\sum_{j<k} \left\{\mathrm{Var}(\hat{\psi}_{jk})\, k^{**}\right\}^2 \nu_{jk}}{d\nu_2 + \sum_{j=1}^{r_z} \mathrm{Var}(\hat{\psi}_{jj})\, k^* \nu_{jj} + \sum\sum_{j<k} \mathrm{Var}(\hat{\psi}_{jk})\, k^{**} \nu_{jk}}$$

and

$$\nu = \frac{\left\{d\nu_2 + \sum_{j=1}^{r_z} \mathrm{Var}(\hat{\psi}_{jj})\, k^* \nu_{jj} + \sum\sum_{j<k} \mathrm{Var}(\hat{\psi}_{jk})\, k^{**} \nu_{jk}\right\}^2}{d^2\nu_2 + \sum_{j=1}^{r_z} \left\{\mathrm{Var}(\hat{\psi}_{jj})\, k^*\right\}^2 \nu_{jj} + \sum\sum_{j<k} \left\{\mathrm{Var}(\hat{\psi}_{jk})\, k^{**}\right\}^2 \nu_{jk}}.$$

## §C.2 One-Sided Tolerance Interval on Future Observations

Calculation of this region requires the approximate distribution of

$$\hat{\underline{l}}'\hat{\underline{l}} + 2\sum_{j=1}^{r_z}\hat{\psi}_{jj}^2 + \sum\sum_{j<k}\hat{\psi}_{jk}^2 + \hat{\sigma}_\epsilon^2 \,. \tag{C.2.1}$$

This can be derived using the approximate distribution developed in §C.1. It begins by recognizing that the approximate distribution of (C.2.1) is equal to

$$\sigma_\epsilon^2 \mathrm{K}\chi_\nu^2 + \frac{\sigma_\epsilon^2}{\mathrm{dfE}}\chi_{dfE}^2 \tag{C.2.2}$$

where terms have been defined in §C.1. The approximate distribution of (C.2.2) using an approximation due to Satterthwaite (1946) and Welch (1956) is

$$\sigma_\epsilon^2 \mathrm{K}^{**}\chi_{\nu^{**}}^2$$

$$\text{where } \mathrm{K}^{**} = \frac{\mathrm{K}^2\nu + (\frac{1}{\mathrm{dfE}})^2\,\mathrm{dfE}}{\mathrm{K}\nu + (\frac{1}{\mathrm{dfE}})\,\mathrm{dfE}} \quad \text{and} \quad \nu^{**} = \frac{\left\{\mathrm{K}\nu + (\frac{1}{\mathrm{dfE}})\,\mathrm{dfE}\right\}^2}{\mathrm{K}^2\nu + (\frac{1}{\mathrm{dfE}})^2\,\mathrm{dfE}} \,. \tag{C.2.3}$$

## §C.3 Prediction Limits on a Future Response

It is necessary to derive the approximate distribution of the estimated variance of $[y_f(\underline{x}_0) - \hat{\mu}(\underline{x}_0)]$ denoted by

$$\hat{\sigma}^2_{y\text{-}\hat{\mu}} = [\hat{\underline{l}}'\hat{\underline{l}} + 2\sum_{j=1}^{r_z}\hat{\psi}_{jj}^2 + \sum\sum_{j<k}\hat{\psi}_{jk}^2 + \hat{\sigma}_\epsilon^2(1 + \underline{x}'_0(\mathrm{X}'\mathrm{X})^{-1}\underline{x}_0)] \,. \tag{C.3.1}$$

Recall the approximate distribution of

$$\hat{\underline{l}}'\hat{\underline{l}} + 2\sum_{j=1}^{r_z}\hat{\psi}_{jj}^2 + \sum\sum_{j<k}\hat{\psi}_{jk}^2 \sim \sigma_\epsilon^2 \mathrm{K}\chi_\nu^2 \tag{C.3.2}$$

derived earlier in §C.1 This will be used along with the distribution of

$$\hat{\sigma}_{\epsilon}^{2}(1 + \underline{x}'_{0}(X'X)^{-1}\underline{x}_{0}) \sim \sigma_{\epsilon}^{2}\frac{(1 + \underline{x}'_{0}(X'X)^{-1}\underline{x}_{0})}{dfE}\chi^{2}_{dfE} \quad (C.3.3)$$

to determine the distribution of the estimated variance of $[y_f(\underline{x}_0) - \hat{\mu}(\underline{x}_0)]$. It should be clear that its approximate distribution is given by

$$\hat{\sigma}^{2}_{y\text{-}\hat{\mu}} = [\hat{\underline{l}}'\hat{\underline{l}} + 2\sum_{j=1}^{r_z}\hat{\psi}^{2}_{jj} + \sum_{j<k}\sum\hat{\psi}^{2}_{jk} + \hat{\sigma}^{2}_{\epsilon}(1 + \underline{x}'_{0}(X'X)^{-1}\underline{x}_{0})] \sim \sigma^{2}_{\epsilon}\, a^{*}\chi^{2}_{\nu_5} \quad (C.3.4)$$

where $a^{*} = \dfrac{K^{2}\nu + \left\{\frac{1}{dfE}(1 + \underline{x}'_{0}(X'X)^{-1}\underline{x}_{0})\right\}^{2}dfE}{K\nu + \left\{\frac{1}{dfE}(1 + \underline{x}'_{0}(X'X)^{-1}\underline{x}_{0})\right\}dfE}$ and

$$\nu_5 = \frac{\left\{K\nu + \left\{\frac{1}{dfE}(1 + \underline{x}'_{0}(X'X)^{-1}\underline{x}_{0})\right\}dfE\right\}^{2}}{K^{2}\nu + \left\{\frac{1}{dfE}(1 + \underline{x}'_{0}(X'X)^{-1}\underline{x}_{0})\right\}^{2}dfE}.$$

Again, this approximation is due to Satterthwaite (1946) and Welch (1956).

# Appendix D. Derivation for the Lower One-Sided Tolerance Interval in the Presence of Categorical Noise Factors

Recall the probability statement from Definition 7.4.1

$$P_{\hat{\mu},\hat{\sigma}_{\underline{I}}\,y(\underline{x})}[P_{y_f}\{y_f \geq \hat{\mu}(\underline{x}) - k\,\hat{\sigma}_{\underline{I}}\,y(\underline{x})\} \geq p\,] \geq 1-\gamma \qquad (D.1)$$

where $\hat{\mu}(\underline{x}) = [\hat{\beta}_0 + E(\underline{I})'\hat{\underline{\gamma}}] + \underline{x}'[\hat{\underline{\beta}} + \hat{\Delta}\,E(\underline{I})] + \underline{x}'\hat{B}\,\underline{x}$

and $\hat{\sigma}_{\underline{I}}\,y(\underline{x}) = \sqrt{\hat{\underline{l}}\,'(\underline{x})\,V\,\hat{\underline{l}}\,(\underline{x}) + [1 - tr(CV)]\hat{\sigma}_\epsilon^2}$ .

The probability statement can be modified to form a standard normal variate.

$$P_{\hat{\mu},\hat{\sigma}_{\underline{I}}\,y(\underline{x})}[P_{y_f}\{\frac{y_f - E_{\underline{I}}(y)}{\sqrt{\underline{l}\,'V\,\underline{l} + \sigma_\epsilon^2}} \geq \frac{\hat{\mu} - E_{\underline{I}}(y) - k\,\hat{\sigma}_{\underline{I}}\,y(\underline{x})}{\sqrt{\underline{l}\,'V\,\underline{l} + \sigma_\epsilon^2}}\} \geq p\,] \geq 1-\gamma. \qquad (D.2)$$

Denote the standard normal percentage point as $K_p$. Then (D.2) can be rewritten as

$$P_{\hat{\mu},\hat{\sigma}_{\underline{I}}\,y(\underline{x})}[\frac{\hat{\mu} - E_{\underline{I}}(y) - k\,\hat{\sigma}_{\underline{I}}\,y(\underline{x})}{\sqrt{\underline{l}\,'V\,\underline{l} + \sigma_\epsilon^2}} \leq K_p\,] \geq 1-\gamma. \qquad (D.3)$$

Further modification yields

$$P_{\hat{\mu},\hat{\sigma}_{\underline{I}}\,y(\underline{x})}[\frac{\hat{\mu} - E_{\underline{I}}(y)}{\sigma_\epsilon\sqrt{\underline{x}\,'(X'X)^{-1}\underline{x}}} \leq \frac{K_p\sqrt{\underline{l}\,'V\,\underline{l} + \sigma_\epsilon^2} + k\,\hat{\sigma}_{\underline{I}}\,y(\underline{x})}{\sigma_\epsilon\sqrt{\underline{x}\,'(X'X)^{-1}\underline{x}}}\,] \geq 1-\gamma. \qquad (D.4)$$

$$P_{\hat{\mu},\hat{\sigma}_{\underline{I}}\, y(\underline{x})}\left[ Z \;\le\; \lambda + \frac{k\,\hat{\sigma}_{\underline{I}}\, y(\underline{x})}{\sigma_\epsilon \sqrt{\underline{x}'(X'X)^{-1}\underline{x}}} \right] \ge 1\text{-}\gamma \tag{D.5}$$

where $\lambda = \dfrac{K_p \sqrt{\underline{l}'V\,\underline{l} + \sigma_\epsilon^2}}{\sigma_\epsilon \sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}$

$$P_{\hat{\mu},\hat{\sigma}_{\underline{I}}\, y(\underline{x})}\left[ \frac{Z - \lambda}{\sqrt{\dfrac{\hat{\underline{l}}'V\,\hat{\underline{l}} + [1 - \mathrm{tr}(CV)]\hat{\sigma}_\epsilon^2}{\sigma_\epsilon^2}}} \;\le\; \frac{k}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}} \right] \ge 1\text{-}\gamma \tag{D.6}$$

We must now determine the distribution of the denominator of the left hand side. Assume that the $\hat{l}_j$'s are orthogonal to each other. Then, recall the approximate distribution given in Equation (7.3.4)

$$\hat{\underline{l}}'V\,\hat{\underline{l}} \;=\; \hat{\underline{z}}'D\,\hat{\underline{z}} \;\sim\; \sigma_\epsilon^2\, d\, \chi^2_{\nu_2} \tag{D.7}$$

where $d = \dfrac{\sum_{j=1}^{r_z} (d_{jj}\rho_j)^2 \nu_j}{\sum_{j=1}^{r_z} d_{jj}\rho_j \nu_j}$ and $\nu_2 = \dfrac{\left\{\sum_{j=1}^{r_z} d_{jj}\rho_j \nu_j\right\}^2}{\sum_{j=1}^{r_z} (d_{jj}\rho_j)^2 \nu_j}$ .

And, $\dfrac{\hat{\sigma}_\epsilon^2}{\sigma_\epsilon^2} \sim \left(\dfrac{1}{\mathrm{dfE}}\right) \chi^2_{dfE}$ . So, we can approximate $\dfrac{\hat{\underline{l}}'V\,\hat{\underline{l}}}{\sigma_\epsilon^2} + \dfrac{[1 - \mathrm{tr}(CV)]\,\hat{\sigma}_\epsilon^2}{\sigma_\epsilon^2}$ by

$$\left\{ d\chi^2_{\nu_2} + \frac{[1 - \mathrm{tr}(CV)]\,\chi^2 \mathrm{dfE}}{\mathrm{dfE}} \right\} \doteq d^*\, \chi^2_{\nu^*}$$

where $d^* = \dfrac{d^2\nu_2 + \dfrac{[1 - \mathrm{tr}(CV)]^2}{\mathrm{dfE}}}{d\nu_2 + [1 - \mathrm{tr}(CV)]}$ and $\nu^* = \dfrac{\{d\nu_2 + [1 - \mathrm{tr}(CV)]\}^2}{d^2\nu_2 + \dfrac{[1 - \mathrm{tr}(CV)]^2}{\mathrm{dfE}}}$ .

Using this approximation, Equation (D.6) can be rewritten as

$$P_{\hat{\mu},\hat{\sigma}_{\underline{I}}\, y(\underline{x})}\left[ \frac{Z - \lambda}{\sqrt{d^*\chi^2_{\nu^*}}} \;\le\; \frac{k}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}} \right] \ge 1\text{-}\gamma \tag{D.8}$$

which becomes

$$P_{\hat{\mu},\hat{\sigma}_{\underline{I}}\, y(\underline{x})}\left[\frac{Z - \lambda}{\sqrt{\frac{\chi^2_{\nu^*}}{\nu^*}}} \leq \frac{k\sqrt{d^*\nu^*}}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}\right] \geq 1-\gamma \quad . \tag{D.9}$$

Now, we will turn the left hand side into a noncentral t variate by writing Equation (D.9) as

$$P_{\hat{\mu},\hat{\sigma}_{\underline{I}}\, y(\underline{x})}\left[\frac{Z + \lambda}{\sqrt{\frac{\chi^2_{\nu^*}}{\nu^*}}} \geq \frac{-k\sqrt{d^*\nu^*}}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}\right] \geq 1-\gamma \tag{D.10}$$

And, finally

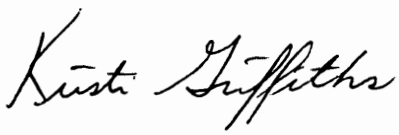
$$P_{\hat{\mu},\hat{\sigma}_{\underline{I}}\, y(\underline{x})}\left[t'_{\nu^*,\,\lambda} \geq m^*\right] \geq 1-\gamma \tag{D.11}$$

or equivalently,

$$P_{\hat{\mu},\hat{\sigma}_{\underline{I}}\, y(\underline{x})}\left[t'_{\nu^*,\,\lambda} \leq m^*\right] \leq \gamma \tag{D.12}$$

where $m^* = \dfrac{-k\sqrt{d^*\nu^*}}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}$. Use the procedure in §7.4 to calculate k.

# Vita

The author was born on July 10, 1968 in Edina, Minnesota. She graduated from St. Louis Park Senior High School, St. Louis Park, Minnesota in 1986. She attended St. Olaf College in Northfield, MN where her interests included trumpet performance as a soloist, in brass ensembles, and in the St. Olaf Concert Band. She graduated in 1990 with a Bachelor of Arts degree in Mathematics with a Concentration in Statistics. She then enrolled in the Statistics graduate program at Virginia Polytechnic Institute and State University in August, 1990. She graduated with a Master of Science degree in December, 1991, and decided to continue her studies by pursuing a Ph.D. degree.

Three years later she graduated with a Doctor of Philosophy degree in Statistics in May, 1995. Her career will begin as a Senior Statistician in the Statistical and Mathematical Sciences Division at Eli Lilly and Company in Indianapolis, Indiana.