Clemson University
TigerPrints

All Dissertations | Dissertations

5-2013

ROBUST PARAMETER DESIGN IN COMPLEX ENGINEERING SYSTEMS:

Gregory Boylan
Clemson University, boylangc@earthlink.net

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

 Part of the Industrial Engineering Commons

Recommended Citation
Boylan, Gregory, "ROBUST PARAMETER DESIGN IN COMPLEX ENGINEERING SYSTEMS:" (2013). *All Dissertations*. 1081.
https://tigerprints.clemson.edu/all_dissertations/1081

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

ROBUST PARAMETER DESIGN IN COMPLEX ENGINEERING SYSTEMS: ANALYTICS, MODELING, AND OPTIMIZATION UNDER ASYMMETRIC AND HIGH-VARIABILITY CONDITIONS

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Industrial Engineering

---

by
Gregory Louis Boylan
May 2013

Accepted by:
Dr. Byung Rae Cho, Committee Chair
Dr. Scott Shappell
Dr. Joel Greenstein
Dr. Brian Melloy

ABSTRACT

Many industrial firms seek the systematic reduction of variability as a primary means for reducing production cost and material waste without sacrificing product quality or process efficiency. Despite notable advancements in quality-based estimation and optimization approaches aimed at achieving this goal, various gaps remain between current methodologies and observed in modern industrial environments. In many cases, models rely on assumptions that either limit their usefulness or diminish the reliability of the estimated results. This includes instances where models are generalized to a specific set of assumed process conditions, which constrains their applicability against a wider array of industrial problems. However, such generalizations often do not hold in practice. If the realities are ignored, the derived estimates can be misleading and, once applied to optimization schemes, can result in suboptimal solutions and dubious recommendations to decision makers. The goal of this research is to develop improved quality models that more fully explore innate process conditions, rely less on theoretical assumptions, and have extensions to an array of more realistic industrial environments. Several key areas are addressed in which further research can reinforce foundations, extend existing knowledge and applications, and narrow the gap between academia and industry. These include the integration of a more comprehensive approach to data analysis, the development of conditions-based approaches to tier-one and tier-two estimation, achieving cost robustness in the face of dynamic process variability, the development of new strategies for eliminating variability at the source, and the integration of trade-off analyses that balance the need for enhanced precision against associated costs. Pursuant to a detailed literature review, various quality models are proposed, and numerical examples are used to validate their use.

DEDICATION

This manuscript is the result of many hours of study, research, and writing; none of which would have been possible without the patience, love, mentorship, and support of several people. Accordingly, this work is dedicated to many, but mostly the following:

- Foremost, my loving wife, Colleen, and my three children, Tully, Abigail, and Emily, whose unwavering confidence, encouragement, patience, and understanding made this achievement not only possible, but a reality.
- My parents, Peter and Kathy Boylan, whose steadfast love and confidence in my abilities, conscious and unconscious mentorship, and limitless support for my desire to pursue this endeavor nourished my own desire and confidence in my abilities to get it done.
- My loyal and faithful dog, Charlie (Chuck), whose many hours spent by my side or under my feet while conducting research or typing furiously at my desk made those hours more enjoyable and served as a constant reminder of what life is really all about: love and respect for others.

ACKNOWLEDGEMENTS

As such works are rarely the product of one person's efforts, I must acknowledge the efforts and contributions of several others that made this endeavor not only possible, but enlightening and enjoyable. They have served as my teachers, coaches, and mentors during my experience at Clemson University and I am indebted to them:

- I must express my most sincere gratitude and respect for my research advisor and Committee Chair, Dr. Byung Rae Cho. Through his guidance and mentorship, he taught me not only how to properly conduct research at the doctoral level and communicate the results to the field, but also how to successfully manage the myriad aspects of research and teaching effectively in a university setting.
- I must also thank my supporting committee members – Dr. Scott Shappell, Dr. Joel Greenstein, and Dr. Brian Melloy – for providing the example, guidance, and direction to be a successful doctoral student and for supporting all of my efforts. Observing these gentlemen interact with and guide their students and research assistants has been a source of inspiration as I prepare to resume my own pursuits in academia.
- Lastly, I must acknowledge the Omar Bradley Fellowship Foundation for funding my research associated with tier-one estimation in ballistic armor applications. This funding enabled me to present my research at various conferences during my time at Clemson University and I am thankful for the opportunity to participate in the fellowship.

TABLE OF CONTENTS

Page

Table of Contents (Continued)

Page

Table of Contents (Continued)

Page

LIST OF TABLES

Table Page

List of Tables (Continued)

Table Page

List of Tables (Continued)

Table Page

LIST OF FIGURES

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

CHAPTER ONE

INTRODUCTION TO RESEARCH

1.1 Introductory Remarks

In the age of global markets and competition, efforts to ensure that products and processes satisfy technical specifications have yielded to a growing focus on ensuring that certain quality characteristics of interest adhere to desired target values with as little variability as possible. The philosophy and methods of Taguchi (1986, 1987), collectively referred to as robust parameter design, or *RPD*, evolved to provide a cost-effective approach for improving product and process quality. In the context of quality improvement, *RPD* is often identified as one of the most critical design approaches currently examined within the research community. Consequently, many industries have implemented robust design techniques to generate improvements in the quality of their products and production methods. Bendell *et al.* (1987), Dehnad (1989), Dobrzanski *et al.* (2007), and Dasgupta *et al.* (2010) provide a number of examples of *RPD* applications to various engineering problems in the automotive, process and health care industries, as well as in the information technology, plastics development, and nanotechnology sectors.

Taking a systems approach to products and processes, Taguchi defined two classes of factors that act upon the system to transform the inputs to outputs: control and noise factors. Control factors represent those variable components of the system that can be readily controlled. Thus, once the control factors are set at specified levels, they do not change and, therefore, do not contribute to undesired system variation. Noise factors,

on the other hand, consist of a variety of system variables that we generally cannot control during normal process operations or product development. Because of their resistance to control, these variables can change quickly and without warning, inducing unwanted variation in system outputs.

The response variables denote the observed system output, or quality characteristic of interest. Taguchi used three classes for response variables: smaller-the-better (*S*-type), larger-the-better (*L*-type), and nominal-the-best (*N*-type). Within this system framework, the primary objective of robust design is to identify the optimum operating conditions for a set of control factors by minimizing the associated variability in the quality characteristic of interest while simultaneously keeping that characteristic's mean value or response at the threshold specified by the customer or decision maker. These solutions are said to be robust in the sense that they render the system relatively impervious to the effects of variation induced by the noise factors.

While much research effort has been placed on developing better and more efficient optimization schemes to determine the best and most robust operating conditions, there are other aspects of the design process that can and should be considered in order to achieve that goal. These include the initial analysis of the collected data, the distributional approaches to modeling realistic process characteristics, the estimation process for process parameters and regression coefficients that facilitate model development and response surface estimation, as well as the approaches to optimization in which we determine an optimal solution to the robust design problem. This dissertation focuses on these critical areas with an overarching objective of creating

quality models and refinements to existing approaches that achieve the best possible *RPD* solutions under realistic process conditions. The deeper aims of this objective, however, are twofold. First is the development of methodologies that facilitate the attainment of enhanced degrees of precision and accuracy than would otherwise be achieved by assuming away true process conditions. The second is to provide engineers and decision makers greater awareness, improved guidelines, and flexibility in applying the best and most appropriate statistical tools and techniques to determine optimal process settings.

1.2 Research Background

Robust design comprises an engineering methodology that aims to optimize product and process conditions in such a way that the outputs are high-quality and cost-efficient products that are minimally sensitive to causes of variation. Within this methodology, there are three principal design steps: system design, parameter design, and tolerance design. The System Design step involves the development of a prototypical design that meets customer requirements, as well as the determination of the materials, parts, components, assembly system, and manufacturing technology required to produce it. Emphasis is placed on using the best available methods and/or technology at the lowest cost to achieve customer specifications through quality function deployment (QFD). Parameter Design focuses specifically on the design variables (or control factors) that affect the functional aspects of a product in terms of a particular "quality characteristic of interest." That is, how should a particular manufacturing process be set up to achieve the desired results or customer specifications. Finally, the Tolerance Design step is applied if the results obtained via parameter design are insufficient. In

short, we look at how we can modify process tolerances to achieve the desired/required results and delineate the associated costs. The emphasis in this step is determining the trade-offs between achieving process targets with further reductions in variability and the associated increases in manufacturing costs.

This dissertation is focused specifically on the robust *parameter* design step. In this step, the objective is to determine the levels of the control factors that are robust to the various influencers of variability that affect product quality, will get the average output (or value) of the quality characteristic at a desired or specified target with minimal variation, and minimize the associated manufacturing costs. This characteristic is often referred to as the response variable, as the value it assumes is in "response" to the settings of the control factors. The resulting control factor settings are referred to as the optimal operating conditions (OOC) for the process.

Determining the optimal process factor settings usually requires an experimental design approach to develop regression models for the average output and the variance associated with it. These models serve to explain the relationship between the various factors and the response, or quality characteristic of interest. Thereafter, we may apply of any one of a considerable suite of alternative optimization methods and schemes to determine the specific levels for the various factors that achieve the desired result which is essentially achieving the least variation in the product's response function relative to the specified or desired target value. The flowchart in Figure 1.1 provides an overview of the traditional *RPD* process.

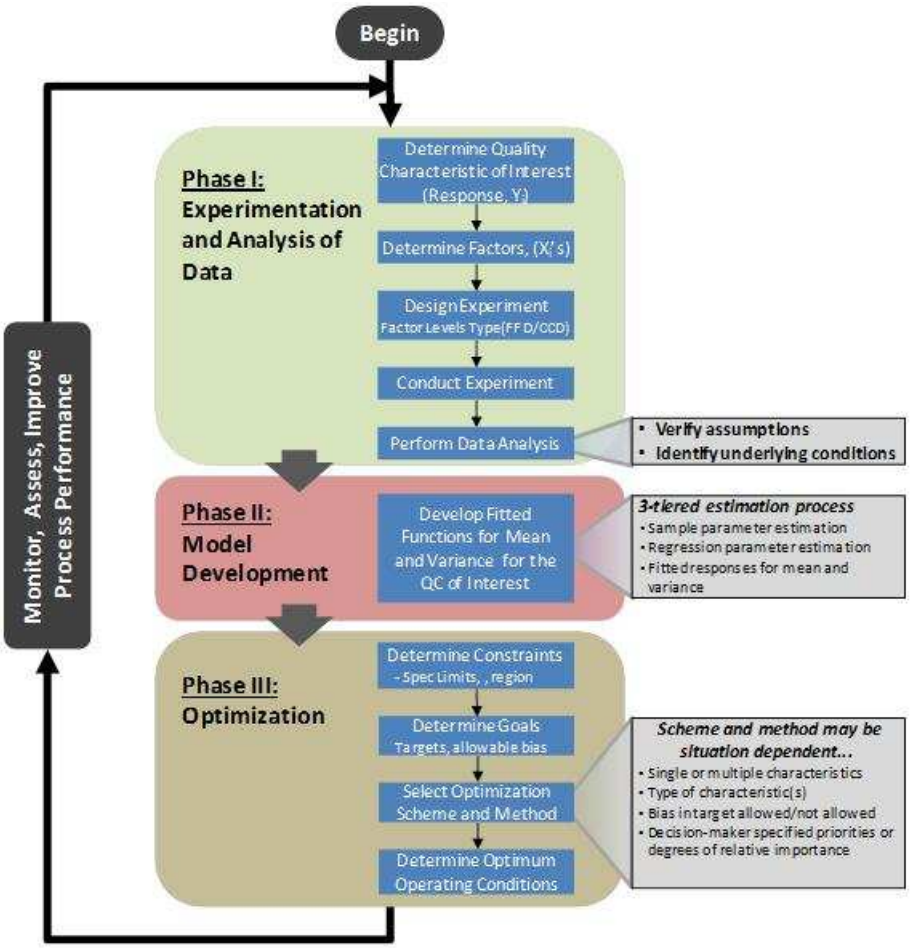


Figure 1.1. Flowchart of the Robust Parameter Design process.

While each of the steps within each phase bears significance, three are more critical than the others. These are the data analysis that completes Phase I, the estimation process used to develop fitted functions that estimate the process parameters (mean and variance), and the selection of appropriate optimization schemes and methods.

Data analysis is intended to be a comprehensive examination of the data collected via experimentation. If done correctly and thoroughly, it can help to answer basic questions about inherent conditions in the process under study, as well as to verify preliminary assumptions made to support the use of various experimental and statistical

analysis methods. We can investigate these questions in a number of ways, both subjectively (using plots, histograms, and other graphical methods) and objectively (using various statistical tests), and each has its advantages and disadvantages, which will be discussed in greater detail in Chapters 2 and 4.

In either event, as the first (and arguably most critical) component in any design of experiments and robust design problem, the data analysis will essentially set the tone for everything that follows. It will serve as the foundation for, and therefore drive all assumptions and approaches to estimation used to develop response surface models in Phase II, as well as the optimal conditions determined in Phase III. Thus, the quality and "structural integrity" of that foundation (and the models and optimization results we build upon it) will only be as good and reliable as the effort committed to doing a proper and thorough data analysis from the outset.

The estimation process that makes up Phase II is basically a three-tiered effort. In the first tier, we must select estimators that will facilitate the estimation of the initial parameters that define the location and scale of the underlying distribution. It is worth noting that the location simply tells us where the preponderance of values in the distribution falls, or where the process is targeted. In the normal (or symmetric) case, this will generally be defined by the sample mean or the center of the distribution; in the asymmetric case, however, it will not. In either instance, the scale is a measure of the dispersion that indicates how much the process varies around that location. It also is important to distinguish between an estimator and an estimate. The former is a mathematical formula that evaluates the available data in a certain way; for example, the

familiar sample mean given by $\bar{y} = \left(\sum_{i=1}^{n} y_i \right) \Big/ n$ where each of the y_i's is a value within an observed set of *n* data points. The latter is simply the numerical result obtained via the estimator.

The initial estimates obtained in tier one form the basis for the estimation method applied in tier two whereby we estimate coefficients of the regression function, denoted by the β's in the function $Y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$. Here, the coefficients define a relationship between the factors (the X_i's) and the response or quality characteristic we're interested in (the *Y*). In short, they help us understand how changes in the factor settings affect process outputs. When combined in the form of the above function, these estimates now comprise the estimator used in the third tier, wherein we use the function to compute the mean or average process response at pre-selected settings for the various factors (X_i's).

What should be clear is that we essentially derive an estimate of distribution parameters to obtain estimates for regression or response surface coefficients, and then obtain estimates for both the mean process response and its variance. Given the nested nature of this estimation process, it is obvious that they are all inter-related. Consequently the quality of the ultimate estimates generated in tier three will depend upon two critical choices: the initial tier-one estimators used to "classify" the process in terms of location and scale, and then the tier-two approach used to estimate the regression coefficients in the response surface functions. Many researchers have developed a considerable assortment of approaches to estimation in each case. But what is not clear is

when, or under what conditions are certain approaches better, or more appropriate, than others? This question requires more attention, particularly since researchers and practitioners have begun to acknowledge that traditional theoretical assumptions made to support the use of certain analytical methods often fail to hold in industrial processes. Chapters 5-8 will address these concerns in detail.

The selection of an optimization scheme and method is similarly driven by a variety of factors. Foremost, we recognize from the first paragraph of this section what it is, exactly, that we are trying to optimize. It boils down to the identification of the best settings for each of the controllable process factors being considered, such that we achieve a mean process result that is as close as possible to some desired target with minimal variation. Beyond this, a number of other criteria can guide the decision. For example, it will depend on the quality characteristic type(s). These can either be smaller-the-best (*S*-type) characteristics, such as the weight of a body armor plate; larger-the-best (*L*-type) characteristics, such as the strength of the body armor plate; or nominal-the-best (*N*-type) characteristics, such as the size or thickness of the body armor plate. Considerations for each of these characteristic types may include the specification limits and/or desired target levels and the degree to which the decision maker is willing to allow deviation from these tolerances in the interests of less variability, or vice versa. As a final example, the selection might depend on the number of quality characteristics under consideration. Sometimes, the decision maker is only interested in a particular quality characteristic. However, given a greater interest in the larger "system of systems", it is often more realistic to consider multiple characteristics simultaneously, and quite often

these are at odds with each other. In the case of the body armor plate, two conflicting characteristics might be the weight of the plate and its strength (in terms of its ability to stop different types of bullets and shrapnel). Cleary, the goal would be to minimize the weight while maximizing the strength. However, making it lighter may suggest less material, thereby reducing its strength. Conversely, making it stronger implies more or denser material, which tends to add weight. When we expand these considerations to the broader system of systems that includes the soldier, the armor plate characteristics must also be considered in the context of the higher level characteristics that impact the soldier's performance, such as his/her ability to shoot, move, and communicate. Clearly, tradeoffs must be considered and so these types of situations require alternative approaches to optimization. These might include developing alternative ways to achieve greater precision in our results, or employing different optimization approaches such as desirability functions, goal programming, or perhaps compromise programming.

Understanding such conditions, as well as the preferences of the decision maker or customer will inform our selection of a particular optimization scheme and/or method. Much like estimation procedures, researchers have developed numerous schemes that employ a variety of different methods to obtain the optimum operating conditions. However, once again, it is not entirely clear in what situations a particular scheme and/or method becomes better than the others. This, too, requires more comprehensive attention.

Ultimately, these three components influence quality in the following way: the level of quality we can achieve stems from the results we obtain through optimization. It

follows, then, that the optimization results depend on the "quality" of the estimates we obtain at each tier in the estimation process. Likewise, the "quality" of these estimates depends upon the appropriateness of the estimators selected, a decision based on the results of the initial data analysis we perform. Thus, if we presume that the definitive purpose of robust parameter design is to allow us as industrial engineers to make reliable recommendations to decision makers that assure the highest levels of quality we can achieve, then the quality and reliability of the results we recommend depend very heavily on the decisions we make regarding the data and approaches to estimation and optimization. Inadequate analysis will lead to poor decisions, which will inevitably translate to suboptimal recommendations and diminished product quality. Following an extensive literature review associated with each component in Chapter 2, the motivation for additional research in these areas is delineated in Chapter 3. Thereafter, proposed methodologies for improvement and the associated analyses are provided in Chapters 4-10. Finally, overarching conclusions and potential areas for further research are provided in Chapter 11.

1.3 Notation and Abbreviations

As a reference, a list of many of the common notations and abbreviations used throughout this manuscript are provided in Table 1.1. Most, if not all, are re-introduced in throughout the dissertation in the context of the chapter in which they are used.

Table 1.1. Common notation and abbreviations used throughout research.

Notation	Meaning / Representation
Y	Quality characteristic of interest (univariate)
$\mathbf{Y}$	Quality characteristics (multivariate), where $\mathbf{Y} = (Y_1, Y_2, ..., Y_w)$
y	Observations made on a quality characteristic
$\bar{y}$	Mean of responses for a quality characteristic
$\mathbf{X}$	Design matrix for least squares regression containing predictor variables
$\mathbf{x}$	Vector of v control factors, where $\mathbf{x} = (X_1, X_2, ..., X_v)$
$\mathbf{x^*}$	Optimal factor settings, where $\mathbf{x^*} = (X_1^*, X_2^*, ..., X_v^*)$
s	Standard deviation of responses for a quality characteristic
$\ln s$	Logarithm of the standard deviation of responses for a characteristic
s^2	Variance of responses for a quality characteristic
s_{ij}	Covariance between the ith and jth characteristics
τ	Desired target value for a characteristic (univariate)
$\boldsymbol{\tau}$	Desired target vector for characteristics (multivariate), where $t= (\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, ..., \boldsymbol{\tau}_w)$
γ_3	sample skewness of responses for a quality characteristic
μ	Process mean for a characteristic (univariate)
μ^*	Optimal process mean for a characteristic (univariate)
$\boldsymbol{\mu}$	Process mean vector for characteristics (multivariate), where $\boldsymbol{\mu}= (\mu_1, \mu_2, ..., \mu_w)$
$\boldsymbol{\mu}^*$	Optimal process mean vector (multivariate), where $\boldsymbol{\mu}^* = (\mu_1^*, \mu_2^*, ..., \mu_w^*)$
$\boldsymbol{\Lambda}$	Covariance matrix
$\hat{\mu}_q(\mathbf{x})$	Response surface design for the mean of the qth characteristic
$\hat{\sigma}_q(\mathbf{x})$	Response surface design for the standard deviation of the qth characteristic
$\widehat{\sigma_q^2}(\mathbf{x})$	Response surface design for the variance of the qth characteristic
$\hat{\sigma}_{ij}(\mathbf{x})$	Response surface design for covariance between the ith and jth characteristics
LSL_q	Lower specification limit for the qth characteristic
USL_q	Upper specification limit for the qth characteristic
n	Number of design points or treatments used in any given experiment
m	Number of replications per design point used in any given experiment
k	Number of parameters used within a regression model

CHAPTER TWO

LITERATURE REVIEW

2.1 Robust Parameter Design

The concepts and methods collectively referred to as robust parameter design (*RPD*) evolved from the works of Taguchi (1986, 1987) as a way to provide a cost-effective approach for improving product and process quality. Although, as several contributors to Nair (1992) point out, the concept of minimizing variability in production process responses has roots dating back more than half a century, it was not until the early 1980's that Taguchi's approach to quality helped the idea go mainstream. Consequently, they have become known as the 'Taguchi method' to achieving quality.

In the context of Taguchi's philosophical stance, the customer's perspective demands that manufacturers consider both the mean (or target specification) and the variability for a particular quality measure in order to improve the quality of the delivered product, as well as the quality of the process used to manufacture it. Taking a systems approach to products and processes, Taguchi defined two classes of factors that act upon the system to transform the inputs to outputs. Control factors represent those variable components of the system that we can readily control. Thus, once set at specified levels, they do not change and therefore do not contribute to undesired system variation. Noise factors, on the other hand, consist of a variety of system variables that generally cannot be controlled during normal process operations or product development. Because of their resistance to control, these variables can change quickly and without warning, inducing

unwanted variation in system outputs. Within this system framework, the goal of *RPD* is to achieve system robustness by selecting control factor settings that help to render the system impervious to the effects of variability.

Taguchi's methodology for realizing this goal incorporated orthogonal crossed-array designs and signal-to-noise ratios (*SNR*) to determine the optimal factor settings. The orthogonal arrays were two: a control, or inner array for the control factors, and a noise, or outer array for the noise factors. From the responses at each control factor setting, the mean and variance were combined to establish the signal-to-noise ratio. The particular *SNR* defined for a response depended on the classification of that response. Taguchi used three classes: smaller-the-better (*S*-type), larger-the-better (*L*-type), and nominal-the-best or specific target value (*N*-type).

Despite the widely recognized contributions of Taguchi's philosophical approaches to quality engineering, various aspects of his methods have received considerable scrutiny (see Box (1988) and Box *et al.* (1988)). In a well-known panel discussion that brought together a number of leading researchers in the field, Nair (1992) codified three general criticisms of Taguchi's methods: 1) his use of signal-to-noise ratios as a basis for analysis, 2) his approaches to experimental design, and 3) his analytical methods. Numerous others, including Tsui (1992), Robinson, *et al.* (2004), Park *et al.* (2006), Arvidsson and Gremyr (2008), Kovach and Cho (2007a), and Hasenkamp *et al.* (2009) echoed portions of this, citing the use of crossed arrays as a particularly disadvantageous practice that leads to a prohibitive number of observations and fails to incorporate the analysis of interaction effects between factor settings, which

is a critical component for accurately estimating optimal operating conditions. Accordingly, much of the research in this field over the past two decades has revolved around developing and improving the analytical methods used to enhance quality. Briefly stated, a number of researchers concluded that Taguchi's methodology merely facilitated process improvement instead of optimization. In response, many of them proposed *RPD* models that investigated a variety of alternative optimization methodologies that were more firmly grounded in well-established approaches to experimental design, most notably response surface methodology, or *RSM*, which predates Taguchi's methods by more than three decades.

2.1.1 Response Surface Methodology

RSM is an experimental methodology that evolved from the work of Box and Wilson (1951) in support of a chemical processing facility. Through experimental designs, they endeavored to determine the optimal levels for various process settings that would, in turn, result in optimal process yield and purity. As part of this, they developed and used the central composite design (*CCD*) concept, which would allow them capture any inherent curvature in a response surface more effectively than was possible using a common factorial design. In particular, by adding axial points to the design space at two levels, and thereby expanding the feasible region from a cuboidal one to a spherical one, the design becomes "rotatable." This is achieved by setting the axial points equal to $\pm(F)^{1/4}$, where F represents the number of factorial points in the experimental framework.

The advantage of rotatability is that it essentially guarantees that the variability in the fitted value obtained at $\mathbf{X}_h$ (the vector of values for the various factors used) will be the same for all fitted values at a given distance from the center, regardless of the point $\mathbf{X}_h$ used or the direction from the center. Accordingly, a design that possesses this property assures the experimenter that the precision in the fitted values depends solely on the distance from the center and not the direction. Figure 2.1 portrays a rotatable *CCD* for three factors in two levels each using coded values, along with four center point replications. The axial point levels are determined by $\alpha = \pm(2^3)^{1/4} = \pm(8)^{1/4} = \pm 1.6818$. It is worth noting here that the number of center point replications for a three-factor *CCD* is typically set to 6, but, as Kutner *et al*. (2005) note, this number may be too large to achieve uniform precision in the estimated mean response. Thus, the actual number may be modified based on experimental conditions.

	Control Factor Levels		
Run	X_1	X_2	X_3
1	1	1	1
2	-1	1	1
3	1	-1	1
4	-1	-1	1
5	1	1	-1
6	-1	1	-1
7	1	-1	-1
8	-1	-1	-1
9	1.682	0	0
10	-1.682	0	0
11	0	1.682	0
12	0	-1.682	0
13	0	0	1.682
14	0	0	-1.682
15	0	0	0
16	0	0	0
17	0	0	0
18	0	0	0

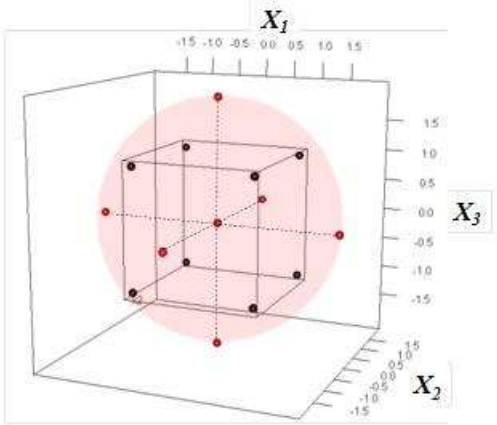


Figure 2.1. Experimental design space for a hypothetical CCD in three factors.

At the time, Box and Wilson considered second-order functions sufficient for modeling response surface designs, as higher-order terms were believed to have negligible effects on a given regression model. Equation (2.1) denotes the standard form of the second-order response surface designs used, in which X_1, X_2, ..., X_{k-1} represents a set of experimental control factors and Y represents the response variable, or quality characteristic of interest for the process under study.

$$\hat{Y} = \beta_0 + \sum_{i=1}^{k-1} \beta_i X_i + \sum_{i=1}^{k-1} \beta_i X_i^2 + \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} \beta_{ij} X_i X_j + \varepsilon \qquad (2.1)$$

Here, the first term represents the intercept, the second represents the main factor effects, the third and fourth correspond to second-order and interaction effects, and ε is the associated error. Since the initial efforts of Box and Wilson, a number of other researchers, including Bose and Carter (1959), Bose and Draper (1959), Box and Behnken (1960a, b), Box and Draper (1959, 1963), and Das (1963) further explored the use of second-order designs as a basis for *RSM*.

While second-order functions are useful in finding optimal solutions in the presence of curvature for most processes, there are some emerging fields in the contemporary engineering and science environment, such as nano-science and molecular biometrics, where significantly elevated levels of precision require alternative estimation methods. Goethals and Cho (2011c) explored this by incorporating higher-order polynomial response surface methods combined with a desirability function approach to achieve higher precision in multi-response optimization problems. Rather than creating higher-order polynomial models simply by adding terms, their proposed method searches

for the best combination of terms based upon the results of various screening criteria. In some cases, the number of model parameters that most appropriately estimates the true response may be less than or equal to the number of terms in the second-order model. By incorporating screening tests to espy potential issues in using higher-order terms along with a procedure using multiple criteria for model selection, they demonstrated increased assurance that increased precision with minimal bias will result.

Early applications of *RSM* tended to be limited to the chemical processing field in which Box and Wilson conceived the approach. Since that time, however, the benefits of *RSM* have become much more widely known, resulting in applications across a large swath of the engineering and other sciences, including physical, social, and biological science disciplines. Within the industrial engineering field, *RSM* approaches have become one of the foremost tools applied in contemporary research efforts. Furthermore, many researchers have investigated ways to improve upon *RSM* methods in all three phases of the *RSM* framework (development of the experimental framework, determination of the estimator functions for the response surface design, and optimization). Among the enhancements in the first two phases, the use of optimal designs and Taguchi's robust parameter design modifications stand out. Enhancements in the third phase stemmed from the aforementioned concerns that many researchers shared regarding Taguchi's approaches. Whereas Taguchi's methods were addressed previously, the remaining two will be discussed in the following paragraphs.

2.1.1.1 Phase I/II RSM Enhancements – The Integration of Optimal Design

The need to consider optimal design strategies arises when certain conditions render the use of traditional experimental approaches inappropriate. Such conditions include instances where time, resource limitations, or some physical restrictions constrain the experimental region. Although the earliest traces of optimal design theory may be attributed to Smith (1918), its contemporary form has its origins in the late 1950's/early 1960's with the works of Kiefer (1959, 1961) and Kiefer and Wolfowitz (1959). Their early efforts, and the extensions that evolved since, have guided the use of optimal designs, establishing a variety criteria as bases for an array of optimal designs. These include the works of numerous researchers, including St. John and Draper (1975), Ash and Hedayat (1978), Silvey (1980), and Atkinson (1982), among others.

The objective for any optimal design is to determine the set of design points from among a list of candidate points that best satisfy the selected criterion. Most often, the optimality criterion used is a function of the variance of the estimated model parameters. Moreover, due to the iterative nature and the complexity involved, these designs typically require the use of sophisticated computer algorithms. Traditional optimality criteria involve invariants of the information matrix $(\mathbf{X}^{\mathbf{T}}\mathbf{X})$, where $\mathbf{X}$ is the design matrix and $\mathbf{X}^{\mathrm{T}}$ is its transposed form. Among these are *A*-optimal designs, which focus on minimizing the trace, or the sum of the elements on the principal diagonal of $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$; *D*-optimal designs, which seek to minimize the determinant of $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$ or, equivalently, to maximize $\left|(\mathbf{X}^{\mathrm{T}}\mathbf{X})\right|$; *E*-optimal designs, which maximize the minimum eigenvalue of the

information matrix by invoking convex minimization methods that use sub-gradients as opposed to gradients at non-differentiable points; and *T*-optimal designs, which maximize the trace of the information matrix. Among these, the *D*-criterion stands out as the most widely used, in large part due to its availability in a suite of contemporary software packages.

In contrast, other optimality criteria focus instead on the variances of predicted responses. This class of optimal designs includes *G*-designs, which seek to minimize the maximum entry in the diagonal of the hat matrix, $\mathbf{H} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}$ and has the consequential effect of minimizing the maximum variance of the predicted values. *I*-optimal designs minimize the average prediction variance across the entire design space, whereas *V*-optimal designs denote a slightly constrained version of *I*-designs in that they minimize the average prediction variance over a specified set of *m* distinct points in the design space.

A number of researchers have addressed the use of optimal designs in their work. Hardin and Sloane (1993) developed an algorithm that combined a modified version of Hooke and Jeeve's pattern search approach with Monte Carlo methods for computing exact calculations in order to identify *A*-, *D*-, *E*-, and *I*-optimal designs for a range of response surface design problems. Kovach and Cho (2007a) evaluated situations in which both experimental constraints exist and noise factors are considered in conjunction with control factors. Seeking an alternative to Taguchi's crossed-array designs and SNR's, they invoked a response model approach involving a combined array design that facilitates the inclusion of both control and noise factors in a single design. Compared to

traditional approaches typically (and sometimes erroneously) applied in these situations, the result is a more efficient methodology that uses a *D*-optimality approach to generating robust solutions.

Ultimately, one of the chief advantages afforded by optimal design is that it provides the researcher with the flexibility to design a tailor-made experiment in an objective way. Atkinson and Donev (1992), Pukelsheim (1995), Montgomery et al. (2002), as well as Kovach and Cho (2007b) provide considerable detail regarding optimal design theory applications.

2.1.1.2 Phase III Enhancements – Integration of the "Dual Response" Approach

Advancements in the Optimization Phase of *RSM* (Phase III) derived from the aforementioned skepticism surrounding the experimental and analytical methods proposed by Taguchi in his *RPD* framework. Vining and Myers (1990) (hereafter referred to as *VM*) were among the first to craft a refinement to Taguchi's method by incorporating the "dual response" approach to multi-response optimization introduced by Myers and Carter (1973) into the *RPD* framework. For a set of control factors, $\mathbf{x} = (X_1, X_2, ..., X_{k-1})$, their dual response approach used separate response function estimates for the process mean and standard deviation given by $\hat{\mu}(x)$ and $\hat{\sigma}(x)$, respectively. Using these estimates, they then proposed to determine optimal factor settings by minimizing the standard deviation while adjusting the process mean to the desired target value, shown by the following scheme:

$$\begin{aligned} &\text{Minimize} \quad \hat{\sigma}(\mathbf{x}) \\ &\text{Subject to:} \quad \hat{\mu}(\mathbf{x}) = \tau \text{ , where } \tau \text{ is the target value} \end{aligned} \tag{2.2}$$

Pursuant to *VM*'s work, Del Castillo and Montgomery (*DM*) (1993) proposed the use of nonlinear programming techniques, specifically the generalized reduced gradient algorithm (*GRG*), to achieve the same results more efficiently. In particular, they showed the *GRG* method as a viable means to formulate problems using a broader array of response surfaces and constraints and thereby, in some cases, achieve improved results.

Cho (1994) and Lin and Tu (*LT*) (1995) focused on *VM*'s modeling approach, arguing that restricting the mean response to a specific value could potentially eliminate better operating conditions from consideration. Accordingly, their extension uses the mean squared error (*MSE*) criterion to allow for bias between the mean response and target value in the following manner:

$$\text{Minimize} \quad MSE = (\hat{\mu}(\mathbf{x}) - \tau) + \hat{\sigma}^2(\mathbf{x}) \tag{2.3}$$

Copeland and Nelson (*CN*) (1996) agreed with the assessment to incorporate bias, but challenged the idea of allowing unlimited deviation from the target. Hence, they modified the *MSE*-based approach by restricting the amount of bias that could be allowed. The generalized result was the following optimization scheme that blended the target-constrained and *MSE*-based schemes:

$$\begin{aligned} &\text{Minimize} \quad \hat{\sigma}(\mathbf{x}) \\ &\text{Subject to:} \quad (\hat{\mu}(\mathbf{x}) - \tau)^2 \le \Delta^2, \text{ where } \Delta \text{ denotes the amount of allowable bias} \end{aligned} \tag{2.4}$$

Like *DM*, *CN* also viewed nonlinear programming as a more efficient way to obtain optimal results using their approach. However, they advocated the Nelder-Mead simplex method as a potentially more effective tool for optimizing dual response schemes in *RPD*.

More recently, Costa (2010) proposed another method to simultaneously optimize the mean and standard deviation of a particular response. Costa's optimization scheme, an extension of the global criterion method developed by Tabucanon (1988), seeks to minimize the deviation of the mean and standard deviation from specified target values as a ratio to the considered range as follows:

$$\text{Minimize} \quad \left(\frac{\left| \hat{u}(\mathbf{x}) - \tau_{\mu} \right|}{U_{\mu} - L_{\mu}} \right)^{\omega_{\mu}} + \left(\frac{\left| \hat{\sigma}(\mathbf{x}) - \tau_{\sigma} \right|}{U_{\sigma} - L_{\sigma}} \right)^{\omega_{\sigma}} \tag{2.5}$$

where τ_i represents the target value and (L , U) denote the lower and upper specification limits for μ and σ, respectively. The incorporation of weights ω_{μ} and ω_{σ} establishes relative priorities between the mean and standard deviation, allowing experimenters to explore varying magnitudes for each, and thus evaluates trade-off analyses between them. Similarly, the inclusion of the ratio *1/ (U - L)* also affords practical flexibility to evaluate trade-offs between different limit settings.

Figure 2.2 shows several of the more commonly used optimization schemes in the literature, delineated where appropriate by the quality characteristic type to which they apply. Many other researchers have developed extensions of these models, contributing to the growing wealth of knowledge in the field. For example, Ding *et al.* (2004) proposed a data-driven approach to determine the weights for use in the Cho/LT *MSE* weighting scheme, which Jeong *et al.* (2005) echoed. Shaibu and Cho (2009) offered another angle to *LT*'s approach that considers the degree to which the standard deviation varies from a specified target in the objective function for all types of responses. These and a number of others, including Kim and Lin (1998, 2006), Tang and Xu (2002), Kim

and Cho (2002), and Koksoy and Yalcinoz (2008), are extensively delineated in Robinson *et al.*(2006), Costa (2010), Shin and Cho (2008, 2009), Shin *et al.* (2011), Kovach and Cho (2007a, 2007b, 2008), Shaibu *et al.* (2009), and Kovach *et al.* (2009).

Research Work	*S*-Type	*N*-Type	*L*-Type
Vining & Myers (1990)	Minimize $\hat{\mu}$ Subject to: $\hat{\sigma} = \tau_\sigma$	Minimize $\hat{\sigma}$ Subject to: $\hat{\mu} = \tau_\mu$	Minimize $\hat{\mu}$ Subject to: $\hat{\sigma} = \tau_\sigma$
Cho (1994) **Lin & Tu (1995)**	Min $(\omega\hat{\mu}^2 + (1-\omega)\hat{\sigma}^2)$	Min $(\omega(\hat{\mu} - \tau_\mu)^2 + (1-\omega)\hat{\sigma}^2)$	Min $(-\omega\hat{\mu}^2 + (1-\omega)\hat{\sigma}^2)$
Copeland & Nelson (1996)	Minimize $\hat{\mu}$ Subject to: $\hat{\sigma} \le U_\sigma$	Minimize $\hat{\sigma}$ Subject to: $(\hat{\mu} - \tau_\mu)^2 \le \Delta^2$	Maximize $\hat{\mu}$ Subject to: $\hat{\sigma} \le U_\sigma$
Kim & Lin (1998)	Maximize λ subject to $d_j(z) \ge \lambda,\ j = \mu, \sigma$ with $d_j(z) = \begin{cases} \frac{\exp(t) - \exp(t\lvert z\rvert)}{\exp(t) - 1}, & t \ne 0 \\ 1 - \lvert z\rvert, & t = 0 \end{cases}$ $z = \frac{\hat{y} - L}{U - L}$ $z = \frac{\hat{y} - \tau}{\tau - L}$ $z = \frac{U - \hat{y}}{U - L}$		
Tang & Xu (2002)	Minimize $(\delta_\mu^2 + \delta_\sigma^2)$ subject to $\hat{\mu} - \omega_\mu\delta_\mu = \tau_\mu$ and $\hat{\sigma} - \omega_\sigma\delta_\sigma = \tau_\sigma$		
Kim & Cho (2002)	Minimize $\omega_\mu^-(g_\mu^-)^2 + \omega_\mu^+(g_\mu^+)^2 + \omega_\sigma(\tau_\sigma - g_\sigma^- - g_\sigma^+)^2$ subject to $\hat{\mu} + g_\mu^- - g_\mu^+ = \tau_\mu$ and $\hat{\sigma} + g_\sigma^- - g_\sigma^+ = \tau_\sigma$ with $g_\mu^- * g_\mu^+ = 0,\ g_\sigma^- * g_\sigma^+ = 0$, and $g_\mu^-, g_\mu^+, g_\sigma^-, g_\sigma^+ \ge 0$		
Shaibu & Cho (2009)	Min $(\hat{\mu} + (\hat{\sigma} - \tau_\sigma)^2)$	Min $((\hat{\mu} - \tau_\mu)^2 + (\hat{\sigma} - \tau_\sigma)^2)$	Min $(-\hat{\mu} + (\hat{\sigma} - \tau_\sigma)^2)$
Costa (2010)	Minimize $\left(\frac{\lvert\hat{u}(\mathbf{x}) - \tau_\mu\rvert}{U_\mu - L_\mu}\right)^{\omega_\mu} + \left(\frac{\lvert\hat{\sigma}(\mathbf{x}) - \tau_\sigma\rvert}{U_\sigma - L_\sigma}\right)^{\omega_\sigma}$		

Figure 2.2. Optimization schemes used in contemporary RPD research.

2.1.2 The Optimal Process Mean (OPM) Component of the *RPD* Problem

A key aspect of the *RPD* problem focuses determining the optimal process mean that results from the identification of the optimal factor settings obtained through optimization. A subtle distinction between this component and traditional *RPD* approaches is the target itself. Whereas traditional *RPD* optimizes based on a customer-specified target, the *OPM* problem seeks to optimize what that target should be. Although intricately linked to the *RPD* framework, determination of the optimal process

mean represents a classical problem in quality control that actually predates it by more than 30 years. C.H. Springer's (Springer, 1951) efforts in determining optimal fill levels for cans in a canning process with given specification limits generally mark the advent of research in this area. Whereas Springer's experiments focused on optimal fill levels given predetermined lower and upper specification limits, Bettes (1962) extended this by fixing one of the specification limits and then optimizing the process mean and tolerance settings for the other limit.

Since these early initiatives took root, many researchers have promulgated and expanded upon the ideas of Springer and Bettes, applying a variety of models to the same types of problems but with alternative methods for defining cost and profit relationships. These researchers include Hunter and Kartha (1977), Bisgaard *et al.* (1984), Golhar and Pollock (1988, 1992), Rahim and Shaibu (2000), and Chen (2004). However, according to the recent review of process mean literature by Tahera *et al.* (2008), most of the research to this point has focused on processes with single quality characteristics. As they note, customers typically judge product quality by more than one characteristic. Thus, models should incorporate the effects of multiple quality characteristics. Several researchers have endeavored to rectify this gap by developing models that address the effects and interactions among multiple quality characteristics. Arcelus and Rahim (1990) are considered among the first to do so. Since then, a number of others have examined situations involving multiple N-type characteristics, including Elsayed and Chen (1993), Teeravaraprug and Cho (2002), Chen and Chou (2003), and Chan and Ibrahim (2004).

For more than thirty years after Springer's initial work, researchers defined quality loss based on whether or not a product conformed to given specification limits. Namely, quality loss did not occur unless the product exceeded the specification limits, implying that any results within the limits were equally good regardless of how far they deviated from the target. Taguchi's introduction of a quadratic loss function in the mid 1980's altered the classical step-loss approach, suggesting that any deviation from the target, regardless of size, should result in a loss of quality from the customer's perspective. Following this lead, a number of researchers developed extensions to Taguchi's approach in order to incorporate the customer's voice more efficiently and effectively. Spiring (1994) applied an inverse normal function to model quality loss, which provided a boundary as a means to associate economic loss with target deviation. Cho and Leonard (1997) developed a class of quasi-convex quality loss functions that viewed loss as proportional to the deviation of the characteristic of interest from its specification limit. Still others, such as Moorhead and Wu (1998) and Chen (2004), have examined the use of asymmetric loss functions to more accurately portray situations in which symmetry is not practical.

In cases involving multiple quality characteristics, overall product quality depends on the consolidated effects of the target deviations for each of the characteristics. However, because of potential interactions between responses (characteristics), the loss incurred may not be simply an additive function of the losses realized for each of the individual responses. Byrne and Taguchi (1987) used a connector and tube example to show that, in the case of conflicting characteristics, evaluating quality loss based on each

separately inadequately accounts for the interactions between them. Pignatiello (1993) considered a multivariate loss function to measure quality based on the interactions among multiple N-type characteristics; Tsui (1999) extended this work to include S- and L-type characteristics. Many others have continued to explore this area, including Kapur and Cho (1996), Teeravaraprug and Cho (2002), and Chan *et al.* (2004, 2005). The amassed effects of both univariate and multivariate quality loss efforts ultimately have resulted in the integration of the customer's perspective into the process mean problem, marking a significant shift from the traditional focus on meeting the manufacturers' needs in conforming to specifications.

When it comes to process variance, it is generally understood that increases in variability translate to an increase in nonconformance and quality characteristic deviation from pre-defined target settings. In the interest of tractability, most studies to date have assumed a known and constant process variance. As Tahera et al. (2008) note, however, many processes actually exhibit some degree of autocorrelation due to various influencers inherent to or impacting on the process. A number of researchers have acknowledged this reality and examined the effects of variance reduction. Schmidt and Pfeifer (1989) made one of the first attempts to explore the effects of variance reduction on expected total cost. Gohlar and Pollock (1992) scoped their approach, looking specifically at the effects of variance reduction on production costs. A number of others have investigated the problem of simultaneously choosing the most economical process mean and variance in continuous production processes, including Rahim and Shaibu (2000), Kim *et al.* (2000), and Chen *et al.* (2002).

2.2 Approaches to Parameter and Regression Estimation

Regardless of the optimization scheme applied to obtain the optimum operating conditions for a given process, the results invariably depend upon the initial estimates used to derive the response surface functions that essentially define the process. In robust design, as in any methodology involving design of experiments and regression analysis, estimation occurs in two tiers. Tier One involves estimation of the location and scale parameters from the data, which serve to describe distributional properties inherent to system outputs. These estimates then form the basis for regression analysis and are used in Tier Two to estimate the coefficients of the regression models for the mean response and the standard deviation or variance, which describe inherent process relationships between the various factors and a particular system response (or quality characteristic of interest). In turn, these models are used in conjunction with a particular optimization scheme to determine the optimal factor settings or operating conditions.

Estimation theory pertains to the branch of statistics that focuses on the use of measured or empirical random data to obtain estimates for parameters of interest. According to Hayter (2006), we can think of parameters as denoting quantities of interest that define certain properties of an underlying distribution governing a particular observation from a larger population. Since the available data almost always comprise a relatively small sample of the population, the underlying distribution and its associated parameters are often unknown. Thus, the initial objective for any experimenter is to ascertain as much as possible about these parameters via estimation, since they provide

the required understanding of the underlying distribution that characterizes the population.

Since robust design focuses on achieving a mean performance level as close to a specified target as possible with minimal variance therein, the parameters we typically seek to estimate from the data are location and scale, as these values will form the basis of our regression analysis. The location parameter refers to a data value that essentially demarks the middle or origin of the dataset or population. Estimates for this parameter typically involve the mean, the median, the mid-range, or perhaps the mode of the data. The scale parameter essentially defines the spread or dispersion of the distribution relative to its location; the larger the scale, the greater the spread. The literature indicates that researchers typically use the standard deviation, the variance, or the logarithmic transformation of the standard deviation to estimate the scale. Goethals *et al.* (2009) investigated the differences between these estimators in terms of the effects they have on the optimal operating conditions. Regardless of the method used, the presence of outliers can significantly affect the estimate, as the squares of the deviations from the mean enter into the calculation and thereby exacerbate the effects of any outliers on the estimate (Ripley, 2004).

Characteristics of a good estimator include its bias, variance, breakdown point, consistency, and efficiency; concepts largely attributed to the seminal work of Fisher (1922). The bias of an estimator denotes the degree to which its expected value differs from the true parameter being estimated. The variance indicates the average degree to which a collection of estimates deviates from the expected values of those estimates. It is

worth noting that a useful relationship exists between the bias, variance, and the mean-square error (*MSE*) of any estimator. Manipulation of the typical rules associated with the expectation of squares yields the following:

$$MSE = E_{\theta}\left[(\theta - \hat{\theta})^2\right] = \left(B_{\hat{\theta}}(\theta)\right)^2 + V_{\hat{\theta}}(\theta) \tag{2.6}$$

Thus, the mean square error of an estimator simplifies to the sum of two terms: one that measures the average difference between the estimator and the true parameter and another that measures the variability of the estimator.

An estimator's breakdown point refers to the proportion of arbitrarily large observations an estimator can handle before giving an arbitrarily large result. Consistency is used to describe how an estimator "acts" as the amount of available data increases. An estimator is said to be consistent if, as more data is collected and new estimates are generated, the distributions of the estimators become increasingly concentrated near the true value of the estimated parameter. Finally, efficiency refers to how accurate a particular estimator is in terms of the number of samples required. That is, an efficient estimator requires fewer samples to achieve an accurate estimate than a less efficient alternative. Often, efficiency is viewed as a relative measure between two estimators. That is,

$$\text{Relative Efficiency} = \frac{MSE\ estimator\ B}{MSE\ estimator\ A}$$

The resulting value is commonly expressed as a percentage and so theoretically ranges between zero and one. However, it can obviously exceed one if the MSE of the numerator is higher. Regardless, higher percentages are more desirable. The following

subsections provide brief historical foundations and a review of contemporary literature associated with tier-one and tier-two estimation approaches.

2.2.1 Tier-one Estimation

Point estimation of a parameter or set of parameters refers to a decision problem in which we observe independent and identically distributed data, X_i, drawn from some probability distribution $p_\theta(x)$ and our goal is to estimate the parameter θ from the data. An "estimator" is any decision rule or function from the data space X^N into the parameter space (i.e., the sample mean, in the Gaussian case). The literature contains myriad approaches and techniques to parameter estimation, as it comprises one of the most fundamental problems in statistics.

Robust design uses the ordinary least squares method to obtain response function estimates for the process (or response) mean and variance. This corresponds to an inherent assumption that the experimental data are normally distributed and relatively devoid of outliers (i.e., uncontaminated). Under these assumptions, the standard methods for estimating the process mean and variance are the sample mean and sample variance given by

$$\bar{y} = 1/n \sum_{i=1}^{m} y_i \quad \text{and} \quad s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1} \tag{2.7}$$

The former simply denotes the arithmetic average of the data observations, while the latter is comprised of the sum of the squared deviances of the data observations about the sample mean.

While the sample mean is the most common method for estimating the location parameter, the most common method for estimating the scale tends to fluctuate between the sample variance, the sample standard deviation, and the logarithm of the sample standard deviation, which tends to be applied when data transformations are required. Different researchers have applied one method or another to demonstrate superior results in comparative studies. Although likely the most common approaches to estimating location and scale, the estimates generated by these sample statistics are known to be vulnerable to outliers and other violations of assumed distributional characteristics.

The method of moments (*MoM*) is another method largely attributed to Karl Pearson in the late 1800's (see Bera and Bilias, 2002). In short, after he developed his self-named class of distributions to fit data that could not be modeled using a normal distribution, he had to develop a new way for estimating the five parameters associated with his model, as methods at the time could not deal with so many parameters. His solution was to equate the sample moments to the population moments and then solve the set of highly non-linear equations sequentially, eliminating one parameter in each step. The first moment is defined to be the mean, the second moment the variance, the third moment is the skewness, and the fourth moment is excess kurtosis. In complex models, with more than one parameter, it can be difficult to solve for these moments directly. Consequently, moment generating functions were developed using sophisticated analytical techniques. These moment generating functions can also be used to estimate their respective moments. Koutrouvelis and Canavos (2000) conducted a comparison of moment-based methods for estimating the log Pearson type 3 distribution, using Monte

Carlo simulation to demonstrate the effectiveness of the method for flood events in high-return periods with large sample size and low-return periods irrespective of sample size. As Hayter (2006) and a number of other statistical textbooks or sources attest, the advantage of this particular method lies in the ease with which estimates may be computed by hand. However, the estimators generated by this method tend to have problems with relatively small samples and are sometimes insufficient in the sense that they fail to account for all relevant information in the data sample.

Probably one of the most well-known and most often used parameter estimation methods is the method of maximum likelihood (*MLE*). Originally developed and introduced by R.A. Fisher in 1922 as a counterweight to Pearson's *MoM*, the *MLE* method simply posits that the desired probability distribution is the one that makes the observed data "most likely" to have occurred. Hence, it is based on maximizing the likelihood function $L(x,\theta)$ where

$$L(x,\theta) = \prod_{i=1}^{n} f(x_i,\theta) \tag{2.8}$$

Typically, this function is maximized by taking the derivative with respect to the parameter of interest, setting it equal to zero, and solving for the parameter estimate. Most often, it is more convenient to take the natural logarithm of the likelihood function, as this yields a monotonic function that is equivalent to simply maximizing $L(x,\theta)$. As Hayter (2006) and Fernandez (2004) note, the *MLE* method generally produces good estimates with good theoretical properties when the sample size is reasonably large. Moreover, it generalizes to situations involving two or more parameters. The literature is

replete with research studies and applications involving the *MLE* method across a broad swath of disciplines and fields of study since its inception nearly a century ago. More contemporary examples include Tsai and Bockenholt (2001) who used MLE for estimating probabilistic paired comparison models, Holmstrom and Petersson (2002) used a weight-based approach with MLE to develop acceptable estimates for applications involving exponential sums, Pisarenko and Sornette's (2004) use of MLE and a "piece-wise" application of *MLE* to compute parameter estimates for deterministically chaotic time series, Bera and Bilias' (2002) review and comparison of *MLE* and five other estimation pedagogies, and Myung's (2003) tutorial on the application of the *MLE* method, to name just a few.

Since Fisher's work in the 1920's, many researchers have developed a variety of extensions of *MLE* aimed at addressing peculiarities of particular situations and achieving better estimators in general. Quasi-likelihood estimation approaches allow greater degrees of variability in data than would typically be allowed or expected from the statistical model used. Within this class are two commonly used estimators which are usually applied to fit binary or count data. The first of these is the marginal quasi-likelihood (*MQL*) method developed by Breslow and Clayton (1993), which Robinson *et al.* (2006) applied to their analysis of data from quality-improvement experiments involving generalized linear models (*GLM*) but found that the method tended to yield large biases when the variance is large. In situations involving binomial *GLM*s, Noh and Lee (2007) showed the penalized quasi-likelihood (*PQL*) method to outperform the *MQL*

estimator. Nonetheless, this method, too, suffered from significant biases in the face of large variance components.

As an alternative to the quasi-likelihood approaches, Lee and Nelder (2006) showed the hierarchical likelihood (H-likelihood) method to be free of such bias issues. Lee and Nelder (1996) initially advocated this method in conjunction with hierarchical generalized linear models (*HGLM*), and have continued to champion its viability and utility since then. The H-likelihood refers to a function of both fixed parameters and random instances of missing data and/or latent variables and then maximizing over the random instances, using an adjusted profile likelihood for predictive estimation. Several have taken issue with the value of the H-likelihood method as a good estimator. Namely, Kuk and Cheng (1999) and Waddington and Thompson (2004) challenged its ability to generate satisfactory estimates in situations involving binary data. Kuk and Cheng (1999) went so far as to challenge the theoretical basis for the method. Nevertheless, Lee and Nelder (2006, 2007) refuted these concerns and showed the method to be free of the various issues affecting the *MQL* and *PQL* methods. They reinforced their findings even more recently in Lee *et al.* (2011), in which they advocated its usefulness in robust parameter designs for quality improvement.

Still other estimation methods, among myriad more may be explored. These include resampling methods, such as bootstrapping or the Jack knife approach which focus on estimates of standard error, Maximum spacing estimation which maximizes the geometric mean of the spacings in the data, and maximum a posteriori estimation.

Overviews and applications of these methods may be found in Shao and Tu (1995), Ranneby (1984), and Degroot (1970), respectively.

Robust estimation methods arose in the 1950's and 1960's out of concerns with the impact of outliers on existing estimation approaches. In particular, researchers had become increasingly aware that many of the most common procedures that were optimized under assumptions of normality were grossly sensitive to even the most minor deviation in the data and from the assumptions. Hence, according to Huber (2009), in this context, robust denotes insensitivity to small deviations from the assumptions. More specifically, the concept pertains to distributional robustness. That is, the shape of the actual underlying distribution exudes only minor deviations from the assumed model. Essentially, robust statistical methods seek to outperform classical statistical methods in the presence of outliers, or, more generally, when underlying parametric assumptions are not quite correct, although more contemporary application of these methods has focused predominantly on handling the effects of outliers.

Examples or robust estimators for the location, or mean, include the sample median and *M*-, *L*-, and *R*-estimators, which actually denote broader classes of estimators. The sample median represents probably the oldest robust method for estimating the location. Defined simply as the numerical value that separates the lower half of a sample or distribution from the higher half, this approach tends to be comparatively impervious to the effects of outliers. That is, whereas the sample mean could be affected significantly by the introduction of a single outlier, the median would remain unchanged. As an example, consider a set of data comprised of ten observations and calculations for

the mean and median as shown in the left column of Table 2.1. The right column contains the same data with the value in the lower right corner changed to reflect an outlier. The results show how the introduction of a single outlier, either through experimental error, process nuances, or data entry errors, induces a 25% shift in the sample mean, which, in the context of robust design, can have enormous impacts on the optimal solution achieved. By comparison, the median is unaffected and therefore considered robust to the effects of the outlier. In terms of contemporary application to robust parameter design, Park and Cho (2003) proposed the use of the median in conjunction with the median absolute deviation method for estimating the scale (discussed in subsequent paragraphs) to achieve more robust solutions in situations involving data that is either contaminated with outliers or violates assumptions of normality.

Table 2.1. Mean and median for datasets with and without a single outlier

No outlier		***Outlier Present***	
34.4	49.7	34.4	49.7
46.7	54.1	46.7	54.1
48.1	54.6	48.1	54.6
48.1	55.3	48.1	55.3
49.4	59.2	49.4	**159.2**
mean:	49.96	**mean:**	59.96
median:	49.55	**median:**	49.55

Among more contemporary approaches for dealing with the effects of outliers, the *M*-estimator proposed by Huber (1964) is one of the first robust methods adopted. The *M* stands for "maximum likelihood type" and these estimators essentially amount to a generalization of *MLE*s. Whereas *MLEs* involve maximizing $\prod_{i=1}^{n} f(x_i)$ or, equivalently,

minimizing $\sum_{i=1}^{n} -\log f(x_i)$, Huber (1964) suggested generalizing this to minimizing $\sum_{i=1}^{n} \rho(x_i)$ where ρ is some arbitrary function. He notes further that when the selection for ρ is $\rho(x;\theta) = -\log f(x;\theta)$, the result is the ordinary maximum likelihood estimate. Thus, *MLE*s are a special instance of *M*-estimators.

L-estimators and *R*-estimators are not as widely used as the *M*-estimator, as they tend to be less flexible and do not generalize as easily to multi-parameter situations. Nonetheless, they present viable alternatives for parameter estimation that should not be overlooked. *L*-estimators correspond to linear combinations of order statistics and include measures such as the median, the trimmed mean, and the "Winsorized" mean. *R*-estimators, on the other hand, are obtained through inverting the rank tests and are often preferred to *L*-based methods due to their global robustness and less restrictive assumptions concerning the underlying distribution(s).

It should be noted that *M*-, *L*-, and *R*- estimators may also be used to obtain robust estimates for the scale, or variability in the data. Huber (2009) provides a detailed overview and theoretical underpinnings for each. Other examples of robust estimators for the scale include the median absolute deviation (*MAD*) method and *S*-estimators. The *MAD* is defined as the median of the absolute deviations from the median of a set of data. Simply stated, after determining the residual between each data point and the median of the data, the *MAD* is the median of the subset of data created by the absolute values of these residuals. This estimator is robust to outliers in the sense that, like the median

previously discussed, the sizes of the residuals associated with a relatively small number of outliers does not really impact the estimate. By comparison, the sample standard deviation and sample variance estimators require that we square the deviations from the mean. This ultimately means that the deviations associated with outliers receive more weight and therefore exert considerable influence on the resulting estimate. As previously noted, Park and Cho (2003) incorporate this scale estimator as part of a broader proposal to achieve better robust design solutions. Lee *et al.* (2007) also examined the combination of the median and the *MAD* in comparison to M-estimators for achieving optimal robust design solutions.

S-estimators were developed by Rousseeuw and Yohai (1984) as a way to streamline the multi-dimensionality associated with M-estimators for scale without sacrificing their inherent flexibility and asymptotic properties. The result is an approach that seeks the line that minimizes the scale of the residuals. In spite of the resistance of this estimator to outliers in both the predictors (leverage points) and the responses, its creators also found it to be comparatively inefficient. Rousseeuw and Yohai (1984) and Rousseeuw and Leroy (1987) provide a detailed overview of the *S*-estimator and various other robust estimators that researchers have developed since 1940. Despite the various approaches to robust estimation heretofore described and/or cited, Huber (2009) points out that *M*-estimators tend to dominate contemporary usage in the field due to their generality, high breakdown point, and efficiency.

Tau estimators for location and scale are still another set of alternative approaches. Developed by Maronna and Zamar (2002), these estimators aim to reduce

the computational complexities commonly associated with most other high-breakdown robust estimators. In short, the location component uses a weighted mean that incorporates median statistics in establishing the weight, whereas the scale estimator uses a truncated standard deviation that invokes the *MAD* as a starting point. The combined effect as a high-efficiency estimator (80% in each case) that is also highly resistant to the effects of outlying data points.

Because distributional assumptions underlying most traditional parametric statistical tests often fail to hold for a variety of reasons (such as asymmetry or excess kurtosis), it may become necessary to invoke non-parametric methods as a recourse. A common complaint regarding such methods is that while they allow for an assessment of statistical significance, they disallow any measurement of effect size in terms of the differences between samples from a data set. The Hodges-Lehmann (1963) estimator for location overcomes this issue by evaluating said differences using pairwise comparisons and median-based statistics. As Serfling (2011) noted, this estimator is highly competitive with the sample mean under normality, can be "infinitely more efficient" under some other symmetric distributions with heavier tails, and is "never much less efficient" at any distribution.

Ultimately, a cursory review of statistics-based literature will highlight the wide variety of estimation methods and further demonstrate that much has been done with respect to estimation techniques, as myriad alternatives have evolved over the last century. Concomitantly, a number of researchers, such as Koutrouvelis and Canavos (2000) and Bera and Bilias (2002) have also examined the differences between the

performances of various estimators. Several of these, including Simpson and Montgomery (1998), Muhlbauer (2009), and Weins and Wu (2010), specifically examined robust estimators, focusing primarily on relative efficiencies, breakdown points, and robustness as means for ascertaining the superiority of one method relative to others. However, the investigated conditions almost singularly focus on cases involving outliers. This is unsurprising, though, as most alternative techniques to the estimation methods used in the *OLS* framework grew out of concern for the impact of outliers on data and subsequent estimates. But a variety of other conditions, such as asymmetry and high variability, can also exist and have equally detrimental effects on the quality of the estimates.

Interestingly, despite of the very large pool of alternative estimators, comparatively few research efforts have explored them in the context of *RPD*, and among these, the only conditions examined include outliers. Table 2.2 summarizes the works pulled from contemporary *RPD* literature.

Table 2.2. Summary of works examining estimator selection in *RPD* problems.

Author(s)	Year	Conditions Examined	Estimators Compared
Park and Cho	2003	Outliers and non-normal symmetric distributions	$(\bar{y}, s^2)$ *vs.* $(\tilde{y}, MAD)$
Lee *et al.*	2007	Outliers and non-normal symmetric distributions	$(\bar{y}, s^2)$ *vs.* $(\tilde{y}, MAD)$ *vs.* Huber 'Proposal 2' estimators
Goethals *et al.*	2009	Impact of variability measure selection under normal conditions	$(\bar{y}, s^2)$ *vs.* $(\bar{y}, s)$ *vs.* $(\bar{y}, \ln(s))$

The works of Park and Cho (2003) and Lee *et al.* (2007) investigate alternative *RPDs* in the face of outliers and non-normal, but symmetric conditions. In particular, the authors examine alternative location and scale estimators to achieve better *RPD* solutions

under the stated conditions. Each of these is discussed in greater detail in Chapter 5. The work by Goethals *et al.* (2009) differs in that it examines and compares three commonly used variability measures under normal conditions to determine which, if any tends to result in better *RPD* solutions when the degree of variability in system outputs changes.

At this point, three things become quite clear: 1) the quality and reliability of the optimization results depend quite heavily on the initial process estimates used, 2) innate process conditions may affect the quality of those estimates, and 3) very little has been done in the RPD research field to explore alternatives to traditional estimation approaches when certain conditions exist. For these reasons, this chapter will outline a conditions-based approach for working with non-standard conditions. In particular, modifications to traditional *RPD* methodology are proposed that incorporate the various aspects of data analysis outlined in Chapter 4 in order to ascertain underlying data conditions and then use this information to drive the selection of tier-one estimators. Numerical examples and Monte Carlo simulation will serve to illustrate how the proposed methodology is used, as well as which estimators tend to perform best under asymmetric and highly variable conditions.

2.2.2 Tier two Estimation

As with tier-one estimators, many alternatives to tier-two estimation exist, as well. The traditional, and by far the best known and most applied approach to estimating the coefficients of a linear regression function is the method of ordinary least squares (*OLS*). Although Carl Friedrich Gauss claimed to have first used the least squares method as

early as 1795, it was actually first published by Legendre in 1805 as an appendix to his work titled *Nouvelles m´ethodes pour la d´etermination des orbites des com`etes*.

The *OLS* method minimizes the sum of squared errors or residuals between the observed responses in a dataset and the responses predicted by the linear approximation, shown as follows:

$$\underset{\hat{\beta}}{Minimize} \sum_{i=1}^{n} \varepsilon_i^{\,2} \tag{2.9}$$

where ε_i denotes the residual of the i^{th} design point and

$$\varepsilon_i = Y_i - \beta_0 - \beta_1 X_{i1} - \ldots - \beta_{p-1} X_{i,p-1}$$

The result is a computationally simple and straightforward estimator that is consistent when the predictor variables are independent and non- multicollinear, and optimal when the residuals possess constant variance (homoscedasticity) and are uncorrelated. When these conditions hold, and the residuals have finite variances, the method of *OLS* provides minimum-variance mean-unbiased estimation.

Once again, the method of maximum likelihood provides an alternative approach to *OLS* when the assumptions required by the latter fail to hold. When applied to a known, fixed dataset with an underlying statistical model, the *MLE* method identifies model parameters that result in the distribution that yields the highest probability for the observed data. The *MLE* method requires distributional assumptions for purposes of assigning a likelihood function that invokes the associated density function and parameters of the assumed distribution.

It is important to note that, under the added assumption of normally distributed residuals, the *MLE* and *OLS* methods are equivalent. In the case of normally distributed response variables and residuals, this can be done analytically; however, in many others, closed-form analytical solutions are not available. This typically requires a generalization of the traditional linear model to accommodate the unique features of various types of data. However, this tends to introduce some key complications; notably, that the mean response may now by nonlinear model, and then the variance may now depend on the mean response. The result of the increased complexity is a necessity to derive estimates numerically using iterative approaches and algorithms, such as the iteratively reweighted least squares approach or the expectation maximization algorithm. In the case of the former, Nelder and Wedderburn (1972) used *IRWLS* in their generalized linear model approach to obtain *MLE*'s of the regression coefficients in situations involving non-normal data and systematic effects requiring some form of linear transformation. Using the latter, Lim (2007) showed that through the expectation maximization algorithm, the *MLE* will tend to produce better results in situations involving missing covariate data.

It is well known that, for a linear model in which the above assumptions hold, the *OLS* estimator yields the best linear unbiased estimator of the coefficients. However, in situations where these assumptions do not hold, it may very likely generate misleading results. Consequently, the *OLS* method is said to be "non-robust" to violations of its underpinning assumptions. In particular, as Huber (1973) points out, *OLS* estimates have been shown to be highly vulnerable to outliers, which refer to observations that deviate

from the patterns established by the rest of the observations. Although this may not pose an issue if the outlier is actually an extreme observation drawn from either tail of a normal distribution, if it stems from non-normal measurement error or some other violation of typical *OLS* assumptions, then the regression results may be compromised. Referring back to the breakdown point previously discussed, since a single outlier can corrupt the least squares line, the *OLS* method is said to have a breakdown point equal to zero and thus is not a robust regression procedure.

In response to the risks posed by certain conditions (namely outliers), numerous researchers starting in the 1970's undertook efforts to develop alternatives to the *OLS* method that would provide greater resistance to the leverage exerted by outlying observations. Enter the concept of "robust regression." Robust regression is a form of regression analysis designed to circumvent the limitations associated with traditional parametric and non-parametric methods. More specifically, the underlying methods are designed such that violations of supporting assumptions yield little impact on the regression results.

One of the simplest alternatives for estimating robust regression coefficients is least absolute deviations (*LAD*). Interestingly, the *LAD* method predates *OLS* by nearly half a century. Boscovich actually introduced *LAD* in 1757 as a means for overcoming inconsistent measurements in estimating the shape of the earth. Despite early interest in *LAD* among researchers, the advent of the *OLS* method supplanted its use and interest waned. Karst (1958) is considered among the first of contemporary researchers to suggest the use of least deviations to overcome the issues of *OLS* regarding the influence

of outliers. Schlossmacher (1973) and a number of others extended Karsts work to yield more precise and robust regression models using the absolute deviations approach. In the last decade, researchers have continued to employ the *LAD* method. Notably, Li and Arce (2004) developed a modified *LAD* approach that invoked a weighted median operation as the basis for a sequence of maximum likelihood estimates of location. Kumar and Shunmugam (2006) proposed the use of *LAD* regression for fitting an engineering reference surface that remains robust to the presence of outliers, namely deep grooves in the surface material. Similarly, Choi and Buckley (2007) used *LAD* to overcome the loss of accuracy caused by the OLS method in their fuzzy regression model. Using simulation, they showed the *LAD* approach to be superior to *OLS* when "fuzzy outliers" exist in the data.

Although the *LAD* method has proved more robust than *OLS*, significant outliers, particularly on the predictor variables, can still bear negatively on the model. Several researchers, including Rousseeuw and Leroy (1987) noted this shortcoming relatively early on, which motivated research into even more robust approaches.

Among the first to spearhead the drive for more robust regression methods, Huber (1973) introduced *M*-estimation for regression, which he basically modeled as an extension of the tier-one *M*-estimation method previously discussed. Various researchers such as Wu (1985) noted the relative efficiency of *M*-estimators compared with other robust alternatives for samples of 40 or more, and that this efficiency increased with the sample size. Thus, the method proved to be a viable estimator that was robust to outliers in the response variable. However, it was also found to be not so resistant to outliers in

the predictor variables, commonly referred to as leverage points. Consequently, in such situations, the method loses most of the advantages it otherwise enjoyed over the *OLS* method. Absent these conditions, however, Andersen (2008) notes that proper formulation of *M*-estimates yields very robust results, particularly in terms of location (or the mean). Notwithstanding, the method's vulnerability to leverage points led a number of researchers including Mallows (1975), Hill (1977), Hampel (1978), Krasker (1980), Marazzi (1993), and Simpson and Montgomery (1998) to contribute to the development of the generalized *M*-estimator, or *GM*-estimator. This modification included a weight function to establish a bound on the influence exerted by any one outlier.

In the 1980s, a number of researchers offered several alternatives to *M*-estimation in an effort to overcome its susceptibility to outliers in the predictor variables. These included least median squares (*LMS*), least trimmed squares (*LTS*), *S*-estimation, and *M*-*M* estimation among. The following paragraphs offer a brief overview of each of these methods; a more thorough discussion of these and those above is provided in Chapter 7. For additional comprehensive discussions and derivations, see Rousseeuw and Leroy (1987), McCullagh and Nelder (1989), and Venables and Ripley (2002).

Rousseeuw (1984) introduced the *LMS* method as an alternative robust regression method that could "resist the effect of nearly 50% of contamination in the data." Rousseeuw showed this method to be very robust to false matches as well as outliers due to bad localization. However, he also showed it to perform poorly in terms of asymptotic efficiency. Accordingly, Rousseeuw (1984) also proposed the *LTS* method. This method has grown in popularity and has been used considerably in the literature. Among more

recent applications, Atkinson and Cheng (1999) used the *LTS* method in conjunction with a forward search algorithm and simulation to determine the amount of data that should be trimmed. Using the commonly-used Brownlee stack loss plant data (Brownlee, 1960), they demonstrated improved efficiency and asymptotic properties relative to the *LMS* method proposed by Rousseeuw. Willems and Van Aelst (2005) pointed out the considerable computational expense of the *LTS* method. To overcome this issue, they proposed augmenting the *LTS* method with a modified bootstrap approach that retained the robustness of the *LTS* estimator while alleviating its otherwise intensive computational requirements.

The fact that both the *LMS* and *LTS* methods involve the minimization of a robust measure of the scatter of the residuals gave rise to a generalization in the form of *S*-estimators. Introduced by Rousseeuw and Yohai (1984) as a means for performing robust regression in time series analysis, this method finds a line (plane or hyperplane) that minimizes a robust estimate of the scale (hence the *S* in its name) of the residuals. This method is highly resistant to leverage points, and is robust to outliers in the response. However, it also was found to be inefficient and computationally expensive (Rocke, 1996).

Yohai (1987) proposed the *M*-*M*-estimator as an improved alternative to *LMS* and *LTS* that would retain the high breakdown points these methods achieved but would also achieve higher efficiency. Yohai's approach essentially blended earlier methods in order to retain the robustness and resistance of *LMS*, *LTS*, and *S*-estimation, while gaining the efficiency of *M*-estimation. This particular method has seen considerable use in the

literature. For example, a study by Simpson and Montgomery (1998) involved several simulations to assess the performance of various robust regression techniques in the face of outliers. Their results confirmed the superiority of the *M-M* estimation method. Ch'ng *et al.* (2005) echoed the findings of Simpson and Montgomery (1998), proposing *M-M* estimation as a preferable means for obtaining optimal operating conditions in dual response surface and multiple response optimizations when responses are non-normal and outliers exist, particularly the latter. It should be noted, however, that although the data examined were non-normal, they were symmetric. Most recently, Maronna (2011) proposed a combination of *M-M*-estimation with ridge regression to overcome inherent susceptibilities of the latter to outliers in the predictor variables (leverage points) and low performance in instances where the number of predictors exceeds the number of observations or design points.

The weighted least squares (*WLS*) method developed by Aitken in 1935 offers yet another way to address the issue of outliers. Here, if the residual of the *i*th point is relatively small then it will be retained in the analysis. Conversely, if the *i*th residual is large, it will be identified as an outlier and then significantly down-weighted to mitigate its influence on the model. The selection of weights is generally a subjective exercise, but typically they are determined to be inversely proportional to the error variance. Thus, the higher the error variance, the lower the associated weight for that observation, and vice versa. *WLS* is also useful in cases in which the error variances are not constant or are heteroscedastic. Heteroscedasticity can result from high degrees of inherent variability in process responses, lack of normality in the data, or in cases of unbalanced

experimental designs in which the number of observations taken at each design point are unequal for a variety of possible reasons. In such situations, the existence of heteroscedasticity suggests that the *OLS* standard errors are potentially wrong, which calls into question any statistical inference based on them, and further suggests that the *OLS* method will not produce the best (or most efficient) estimates for the regression coefficients.

In cases involving non-normality, researchers typically apply transformations or *GLMs*. Transformations typically serve any of four purposes. These include 1) transforming a non-linear model to one that is linear, 2) transforming the predictors (X_i) or the responses (Y_i) to enhance the linear relationships between the two and thereby improve the fit, 3) correcting for heteroscedasticity in the residuals, and 4) overcoming issues with non-normality. Often times, it is preferable to transform *Y* rather than *X*, as changing the latter usually does very little to alter the distribution of the data relative to the regression line. Conversely, transforming *Y* induces changes in both the distribution of data about the regression line and the vertical spacing of the observed values. In either case, transformations require prudence, as the implied need to transform may stem from just a few influential observations. As Ryan (2009) noted, it is preferable that the data set as a whole drive the need for any transformation, rather than a few outlying observations.

*GLM*s, on the other hand, have received considerable attention as a practical alternative to data transformations (see McCullagh and Nelder (1989), Myers and Montgomery (1997), Hamada and Nelder (1997), Venables and Ripley (2002), and Myers *et al.* (2002)). In short, they expand approaches to linear modeling to account for

non-normal response distributions, as well as the need for linear transformations. Consider the linear model $y_i=\mathbf{X}_i\beta + \varepsilon_i$, where $\varepsilon_i \sim N(0,\sigma^2)$. This may be rewritten as $\mu_i=\mathbf{X}_i\beta$, where $y_i \sim N(\mu_i,\sigma^2)$ and $\mu_i=E[y_i]$. A *GLM* extends this by stating $g(\mu_i)=\mathbf{X}_i\beta$, whereby $y_i \sim EF(\mu_i,\phi)$. In this case, $g(\bullet)$ denotes any smooth monotonic link function, $EF(\mu_i,\phi)$ is any exponential-family distribution (such as normal, gamma, Poisson, binomial, etc.), ϕ is a known or unknown scale parameter, and $\mathbf{X}\beta$ is the linear predictor. Typical link functions include log, square root, and/or logit arguments (i.e., $(\log\{\mu_i/(1-\mu_i)\})$. In some ways, $g(\bullet)$ acts like a transformation, except that it transforms $E[y_i]$ rather than y_i itself. Thus, it facilitates transforming the systematic part of a particular model without altering the distribution of the associated random variation.

A number of other estimation regression estimation methods exist. These include maximum entropy methods, nonlinear least squares, and ridge regression, just to name a few. Detailed exposition of these methods may be found in Golan *et al.* (1996), Kutner *et al.* (2005), and Gruber (1990), respectively. However, just as in the case of tier-one estimators, despite the extensive pool of alternative regression approaches, comparatively few quality engineering efforts have explored them as alternatives for obtaining *RPD* solutions, and among these, the only conditions examined include outliers and unbalanced data sets, as once again even non-normal conditions focused on symmetric distributions. Table 2.3 summarizes contemporary research efforts found in the *RPD* literature.

Table 2.3. Summary of works examining estimator selection in *RPD* problems.

Author(s)	Year	Conditions Examined	Regression Estimators Compared
Simpson and Montgomery	1998	Outliers under normality	*OLS*, *M*, *Most-B Robust*, *LTS*, *S*, *M-M*, various versions of *Generalized M*
Lee and Nelder	2003	Non-constant variance and non-identity (Gaussian) link functions	*GLM*
Cho and Park	2005	Unbalanced datasets	*OLS* vs. *WLS*
Ch'ng *et al*.	2005	Non-normal responses and outliers	*OLS* vs. *M-M*
Robinson *et al.*	2006	non-normal (gamma) and batch-to-batch variation (random block design)	*GLMM (gamma with log link)*
Lee *et al.*	2007	Outliers, non-normal symmetric distributions	*OLS* vs. *M-M*
Goethals and Cho	2011	Heteroscedastic conditions and unbalanced data	*OLS* vs. *WLS*

Regarding *GLMs*, Lee and Nelder (2003) examined their use as a generalization of data transformation and *RSM* approaches that allows for "arbitrary variance and link functions." In a more recent effort, Robinson, *et al.* (2006) examined generalized linear mixed models (*GLMM*) specifically in an *RPD* context to address the non-normality encountered with a resistivity quality characteristic, using the known distribution for the response (gamma) combined with a log link. While the results in each of these works clearly demonstrated the benefits of utilizing *GLM*s, they were not necessarily comparative studies, per se.

The remaining works shown in Table 2.3 pertain to more direct comparisons between traditional and robust regression approaches. Simpson and Montgomery (1998) examined alternative regression techniques when dealing with outliers within normally

distributed data. However, this study focused more on statistical estimator performance measures such as efficiency, consistency, and breakdown points rather than any optimal *RPD* solutions obtained through application. Cho and Park (2005) specifically considered *RPD* solutions in the case of unbalanced data and proposed the integration of a *WLS* approach in which the weighting scheme is based upon the number of observations at each design point, as this value is inversely proportional to the variance associated with the response surface functions obtained for the process parameters. In the interests of finding better optimal settings in dual-response surface optimization problems when non-normal conditions and/or outliers exist, Ch'ng *et al.* (2005) performed a study that compared *OLS* to the *M-M* robust estimation technique developed by Yohai (1987), a well-known robust regression technique that will be described in greater detail in Section 6.2.2. In their examination of tier-one estimators in robust designs involving contaminated data, Lee *et al.* (2007) also included a comparison of the *OLS* method to the *M-M* regression technique, but their work investigated the differences in the context of robust design. Goethals and Cho (2011) extended the work of Cho and Park (2005) to the optimal process target problem, considering heteroscedastic conditions in addition to the unbalanced data case.

2.3 The Use of Graphical Methods and Statistical Tests to Support Data Analysis

Just as the initial phases of estimation will affect the outcome of the RPD problem in terms of the solution obtained, the quality and reliability of the estimates used to obtain those solutions will invariably depend upon the quality analysis used to derive those estimates. A recurrent theme in each of the preceding sections is that all of the

applications of the experimental approaches, estimation methods, and optimization schemes heretofore addressed rely upon a variety of assumptions about the process and the underlying conditions of data obtained from it. These include assumptions about the responses, as well as the residual errors obtained through regression. In the case of the responses, most research efforts assume normality, which implies symmetry; moderate to low variability commensurate with a stable process; and, in some cases, such as in specific applications of the optimal process mean problem, fixed process variability. For the residual errors, the basic assumptions that underpin the *OLS* method are that they are independent and identically distributed normal random variables.

In either case, these assumptions are, as Hasenkamp *et al.* (2009) noted, "propositions that are taken for granted, that is, as if they were known to be true." However, as our knowledge and understanding of industrial processes has evolved, it has become apparent that they quite often do not hold in practice. As Hasenkamp *et al.* (2009) further note,

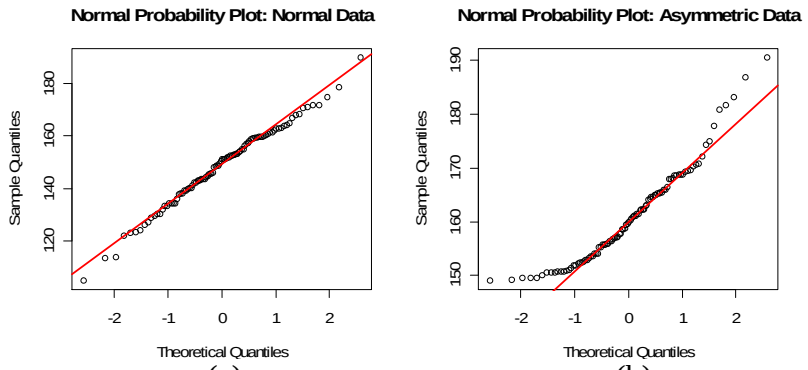
> "Doubtless, the checking of assumptions... cannot be left to some standardized tool. It must be done by engineers who will have to revert to experience and prior knowledge."

This underpins the criticality of data analysis to the *RPD* problem, a methodology that involves a blending of both graphical and objective approaches, but which, based on the engineer's "experience and prior knowledge," depend more heavily on the former.

For many decades, statisticians have invoked a variety of graphical methods for examining data to determine how they are distributed and what sorts of characteristics those distributions have. Such graphical methods have included histograms, half-normal

plots, percentile-percentile (P-P) plots, or quantile-quantile (Q-Q) plots, among others. Of these, Q-Q plots, also known as probability plots, are the most often used means for comparing the actual distribution for a set of data against a hypothesized distribution for the population.

Although such plots can be constructed for any distribution, we most often hear of them in the context of normal probability plots. The reason for this stems from the typical assumptions of normality inherent in many statistical procedures. Beyond this, however, the use of normal probability plots as an initial basis for data analysis makes logical sense. In general, most people, both statisticians and non-statisticians alike, are very familiar with the bell-shaped curve associated with the normal distribution compared with the appearance of other more obscure distributions. Yet, detecting departures from a particular curved shape can be difficult compared to assessing deviations from a straight line. By invoking a straight line to represent normality within a dataset, the use of normal probability plots provides a way to quickly and effectively analyze data in ways that facilitate understanding as to how the data relate to or deviate from a normal distribution. The sample plots in Figure 2.3 provide a comparison of data sets derived from a skew normal distribution with common values for the location and scale (150 and 15, respectively) but different degrees of inherent skew (0 and 15, respectively). As expected, the data in the left plot generally fall along the reference line, indicating normality. By comparison, the data in the right plot clearly deviate from the line, suggesting non-normality.



(a) (b)

Figure 2.3. Normal probability plots for (a) normal and (b) asymmetric data.

Pursuant to constructing a probability plot based on a selected plotting position, many researchers have proposed a variety of techniques for objectively evaluating the information in the plot to assess the fit against the hypothesized distribution. Several have offered the use of the correlation coefficient in support of a goodness-of-fit test statistic derived from the plot. These included Filliben (1975), Shapiro and Francia (1972), Johnson and Wichern (1982), and Looney and Gulledge (1985), among others. However, one of the most powerful tests for normality is the Shapiro-Wilk (1965) test, which considers a ratio of two estimators of σ^2. In short, if the normality assumption is valid, then this ratio will be very close to one. Looney and Gulledge (1985) highlighted a number of other researchers who proposed extensions of other familiar tests, such as the Kolmogorov-Smirnov test and the Lilliefors statistic. However, they also pointed out that these, along with the Shapiro-Wilk test, fail to address the linearity of the probability plot as well as methods that invoke the correlation coefficient. The Anderson-Darling (*A-D*) test provides another objective basis for investigating assumptions of normality. Although, as Vining (2010) pointed out, for small sample sizes this test often fails to

identify issues with such assumptions. Conversely, for large sample sizes in which departures from normality are less consequential, it tends to overcompensate and signal even small deviations from normality. Chi-squared analysis is also a well-known avenue. However, chi-squared analysis is fairly restrictive, as the data can have more departure from normality than the chi-squared analysis is willing to allow.

While such methods for examining the normality of data are time-tested, they do little beyond confirming that a particular set of data are either normally distributed or not. That is, in the cases whereby the data are not normally distributed, they provide no information as to what the underlying distribution might be. We certainly could perform hypothesis tests for other distributions but we would need to have an idea as to what distribution the data might follow. A good and relatively simple alternative for deriving information the distribution of data is the normal probability plot. Wilk and Gnanadesikan (1968) suggested the use of Q-Q plots, P-P plots, and a variety of derivatives of these plots as effective informal statistical methods that are "suggestive and constructive rather than formal procedures to be applied in the light of a tightly specified mathematical model." Since then, a number of other researchers and statisticians, such as Snedecor and Cochran (1980), Nelson (1982), and Johnson and Wichern (1982), have also recommended the use of these plots for assessing the goodness of fit of a hypothesized distribution. Some, including Shapiro and Brain (1981), suggested that any formal goodness-of-fit test include a probability plot simply because visual inspection afforded by the plot usually provides more descriptive information about the data than any single test statistic. Looney and Gulledge (1985) approached this from the opposite

direction, suggesting that probability plots be examined first and then augmented by formal statistical testing based on information deduced from the plot.

Although predominantly used for residual analyses, a number of contemporary research efforts have incorporated these plots in other ways. Yang and Qi (2010) invoked the Q-Q distance calculated from Q-Q plots as the basis for a distribution-free nonparametric quickest detection procedure. Through experimental studies, their use of the Q-Q plot and associated distance measure yielded results comparable to or better than those achieved using classical parametric and nonparametric methods. In their study on processing observational data obtained from global navigation satellite systems, Luo *et al.* (2010) proposed the use of Q-Q plots in conjunction with analysis of the first four sample moments to facilitate statistical inferences drawn to illustrate the discrepancies between the common assumption of normality and what is more realistically found in practice. Azadeh *et al.* (2011) used normal probability plots to identify the presence of outliers as part of two nonparametric efficiency frontier analysis sub-algorithms based on the Artificial Neural Network (ANN) technique and a combination of ANN and Fuzzy C-Means for measuring efficiency. Their methodology incorporated plots as both a visual and objective means to identify the presence of outliers and then select one of the two sub-algorithms to determine the stochastic frontier.

Regardless of the motivations behind its use, the normal probability plot produces a curve that reflects the distribution of the data relative to the normal distribution, which is denoted by a straight line. As DeCarlo (1997) pointed out, typical introductory courses in statistics comment on the ability to characterize a distribution in terms of its “central

tendency, variability, and shape." While most textbooks define and illustrate shape in the context of skewness, kurtosis is another aspect of shape that has equally important implications for the data. In any event, these qualities essentially correspond to the moments of a distribution. Thus, the shape of the curve in the probability plot and the manner in which it deviates from the reference line provide information about the characteristics of the underlying distribution in terms of the first four sample moments.

CHAPTER THREE

RESEARCH MOTIVATION

3.1 Introductory Remarks

Since the introduction of Taguchi's robust parameter design approach to quality, the industrial and manufacturing environments have evolved significantly. In particular, advancements in computing and robotics have expanded many engineering capabilities, including the ability to operate at infinitesimal levels. This, in turn, has led to demands for increased precision and accuracy in all facets of the engineering design process. Irrespective of the levels at which a process operates, the global nature of today's market typically judges its products according to multiple characteristics, each of which are influenced by any number of controllable or uncontrollable factors within the process. While modern industrial practices have needed to adjust somewhat to account for this, the underlying economic objectives have not - the intent remains to reduce production cost and material waste without sacrificing product quality or process efficiency. Consequently, the criticality of quality control and assurance programs in process and product optimization will continue to expand.

While advancements in research have been made in the design of quality optimization models, significant gaps exist between current methodologies and the realities typically observed within contemporary industrial engineering systems. In many cases, engineers base their models upon assumptions that either limit their usefulness or diminish the reliability of the estimated results. This includes instances whereby

engineers generalize the models to a specific set of assumed process conditions, which invariably constrains the applicability of the models against a broader selection of industrial problems.

However, as collective understanding of industrial and other processes has evolved, many such assumptions and generalizations often do not hold in practice. If ignored, the estimates derived in such instances are likely to be problematic and misleading. Moreover, once applied to optimization schemes to determine optimal operating conditions, they cannot help but generate suboptimal solutions and uncertain recommendations to decision makers. Take the sample mean $\overline{y}$, for example. While commonly used to estimate the location or distributional average of observed data, it can be grossly upset by the presence of a single outlier (Ripley, 2004) or if the underlying distribution deviates from the assumed normal one (Huber, 1964). When assumptions of normality are violated, there are in fact rigorous ways to overcome it, including non-parametric approaches or data transformation techniques, such as a logarithm or a Box-Cox transformation. While these approaches work in theory, there is no guarantee that they will yield a result commensurate with the normality assumption. Alternatively, we can attribute outliers to experimental error or other extraneous effects and therefore remove them from the data. In fact, a major reason for discarding data points is that, under the *OLS* method, a fitted regression function is disproportionately pulled toward outlying observations as a consequence of using squared deviations in the approach.

But it is not enough to simply eradicate or assume away existing conditions, as they could very well be inherent process realities that must be accounted for. Park and

Cho (2003) noted that such conditions may actually convey important information, such as interactions with other response variables, and therefore should be retained and addressed. For example, in the case of outliers, Ripley (2004) articulated four cautionary notes against screening data and removing these points:

1. Users, including expert statisticians, quite often do not screen the data.
2. The decision to retain or remove an outlier is wasteful; the analysis would be better-served by incorporating a weighting scheme whereby questionable observations get down-weighted. Completely wrong observations would an exception.
3. Identifying outliers can be difficult, if not impossible in multivariate or highly dimensional data.
4. Removing outliers affects the theorized distribution, which then must be adjusted to avoid underestimating variability measures from the 'cleaned' data.

In many cases, research efforts to date have avoided some of these issues due to the mathematical complexity and associated difficulty in modeling some conditions. The result is extremely limited documentation in the quality engineering literature. However, the fact that increases in computing power and speed have rendered historically complicated and difficult problems comparatively tractable suggests that we should continually seek ways to better understand and implement alternative approaches that facilitate more precise and robust results. The literature review in Section 2 highlights the breadth and depth of the many ways in which researchers have answered the call to improve the statistical underpinnings of and approaches to quality engineering. Yet, in their review of practices for *RPD* methodology, Hasenkamp *et al.* (2009) note that *RPD* research has predominantly focused on developing an array of new statistical techniques rather than a refined understanding of how to use existing techniques better. A direct

consequence of this tool-based research focus is the evolution of different, often conflicting ideas regarding the scope and framework of *RPD*, namely in terms of the preferred application of the various statistical tools. This, in turn, has induced stagnation in the application of *RPD* principles in many industrial settings.

Ultimately, applying a tool without regard for its underlying and motivating practice may very well lead to incorrect or suboptimal applications, and eventually misinformed decisions. There is a need for clarification about the activities required to achieve robust design solutions. Such activities should not be confused with the development of yet more specific tools instructing on how to fulfill or accomplish these activities. Rather, they should maintain and clarify the right purposes and conditions for utilizing existing tools appropriately. Currently, there is a lack of such clearly identified practices as the literature has focused instead on technical details of statistical tools for *RPD*. Thus, the goal of this research is to create more accurate and applicable quality models that more fully explore inherent process conditions, rely less on theoretical scientific assumptions, and have extensions to an array of more realistic process conditions. Pursuant to this, the following sections address key areas wherein further research can reinforce foundations, extend existing knowledge and applications, and perhaps narrow the gap between academia and industry. These include the integration of a more comprehensive approach to data analysis based on normal probability plot analysis, the development of conditions-based approaches to tier-one and tier-two estimation, achieving cost robustness in the face of dynamic process variability, and the

integration of a trade-off analysis that balances the need for enhanced precision against associated costs.

3.2 Improving Data Analysis Using Normal Probability Plots

Ultimately, effective quality improvement programs begin with the collection and subsequent examination of data. Given this, the importance of proper data analysis in the initial stages of cannot be overstated, as this analysis underpins our assumptions for and approaches to process modeling, estimation, and ultimately optimization. However, merely analyzing data is not sufficient from the point of view of making decisions. How and what one interprets from the analyzed data may be more important. Thus, data analysis comprises a decision support system rather than a decision making system. Thus, we need to know, up front, what is happening in the data. This will help to inform us about inherent "peculiarities" in the process (such as skewness or kurtosis), the degree of variability and sources of influence (such as outliers in the responses or predictors), all of which will better inform our underlying assumptions about the data and drive our selection of parameter and regression estimation approaches.

The question becomes how to proceed with data analysis to develop and eventually validate those assumptions. Many statistical tests have evolved over time and are widely used to provide an objective basis for answering this question. These tests are important, as an assumption about a particular distribution affects how we estimate parameters, which in turn influence the estimation of statistical models, and ultimately optimization results. However, these tests essentially only answer one question: the data either are or are not distributed as hypothesized. While we generally hope for the former

result, as it provides statistical support for our assumptions, the latter is problematic simply because it provides no other information as to what the underlying distribution might be.

This dichotomy highlights the utility and importance of graphical methods in understanding what the data are communicating in terms of distributional properties and other underlying conditions. Such graphical methods might include histograms, half-normal plots, percentile-percentile (P-P) plots, or quantile-quantile (Q-Q) plots, among others. Of these, Q-Q plots, also known as probability plots, are used more often as a graphical means for comparing the actual distribution for a set of data against a hypothesized distribution for the population. Despite the variety of graphical techniques available, a full analysis of these plots and the information they can provide about data has yet to be done. Chapter 4 endeavors to fill this gap by providing a comprehensive analysis of the various data characteristics and properties that a fuller use of the normal probability plot can provide. This includes analysis of how efforts to modify system properties with the overarching aim of variance reduction can reveal underlying properties in the data that may have been otherwise undetectable.

3.3 Conditions-Based Approaches to Tier One and Tier Two Estimation

Response surface methodology (*RSM*), which is a key statistical method for modeling robust design problems, explores the functional relationship between an array of predictor variables, or factors, and a response variable, or characteristic of interest. The many research efforts discussed in Chapter 2 demonstrate that, since the 1980's, much of the extensive attention paid to the *RPD* problem in terms of the *RSM*-based dual

response approach has focused on alternative optimization techniques for obtaining robust design solutions. In the interests of promoting various statistical and optimization approaches, the response surface models used with the optimization approaches described in Chapter 2 are typically assumed to be good fits and that ordinary least squares assumptions hold. However, the degree to which a response surface yields a good fit is, in large part, attributable to the quality of the estimates used to develop the fitted models. Thus, in presuming that all procedures preceding the optimization phase have been performed correctly, it is also presumed that the estimates derived in both tiers are good and have been obtained using appropriate estimators.

3.3.1 Conditions-based selection of Tier-One Estimators

Typical *RPD* approaches invoke least squares methods to obtain appropriate response functions for the mean and standard deviation (or variance) by assuming that the experimental data follow a normal distribution and are relatively free of contaminants or outliers. Consequently, the most common estimators used in the initial tier of estimation are the sample mean and sample standard deviation, as they are very good estimators when these assumptions hold. However, when asymmetric conditions persist, they can have tremendous impacts on the estimates and subsequent results if unaccounted for. For example, in cases involving smaller-the-best (*S*-type) or larger-the-best (*L*-type) characteristics, the distribution will almost invariably assume the shape of an asymmetric or skewed distribution, which actually would be preferred. And this could even occur in the case of certain instances of nominal-the-best (*N*-type) characteristics. Two important factors to recognize under these circumstances are that 1) depending on the degree of

variability and skew, the mean will shift away from the central tendency of the distribution; and 2) the standard deviation will not accurately describe the dispersion in the distribution as it will be significantly affected by the "play" in the skewed or long tail of the distribution. If true contamination exists, the effects worsen.

Certainly, in the case of a normal or other symmetric distribution (Laplace, uniform, Cauchy, etc.), the sample mean will correspond to the value that possesses the greatest probability of occurring. However, in asymmetric cases, it will not. Consider a comparison between the probability densities of samples drawn from a skew normal distribution with positive skewness and a normal distribution with the same mean shown in Figure 3.1 below. In the normal case, the mean corresponds to the peak in the density function, indicating that the likelihood of obtaining that value or values very close to it in a random draw exceeds all others in the distribution. In the skew normal or asymmetric case, the mean clearly lies to the right of the preponderance of probabilities, whereas the median of the asymmetric example falls very close to the peak of the distribution. Thus, in the latter instance, while the mean certainly still defines central tendency in terms of the population mean, it does not necessarily correspond to central tendency in terms of the values of greatest likelihood. This demonstrates that, when actual conditions deviate from traditional assumptions, alternatives to the sample mean and variance should be considered for estimating the predominant location and scale of the underlying distribution.

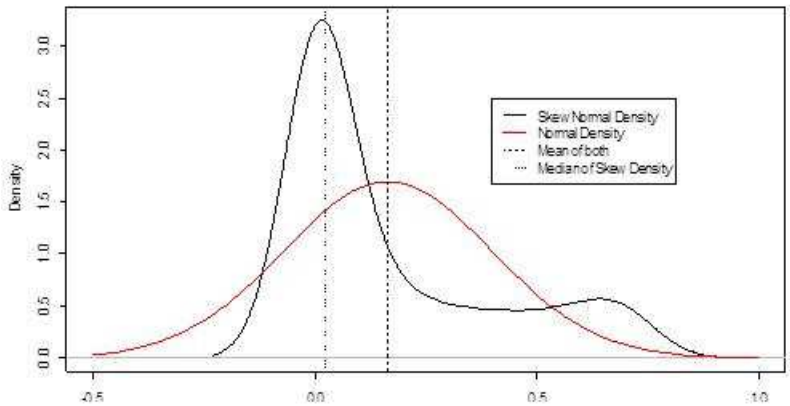


Figure 3.1. Comparison of normal and skew normal densities with a common mean.

The selection of estimators in highly variable conditions follows logic very similar to that described for asymmetric conditions. Consider a situation in which a process possesses inherently high degrees of variability. Under such circumstances, the likelihood of obtaining "extreme" responses increases considerably. Figure 3.2 portrays this reality using densities from two normal distributions with a common mean, but drastically different standard deviations (one low and the other high). From the figure, it is obvious that the propensity for observations in the tails increases with increases in variability.

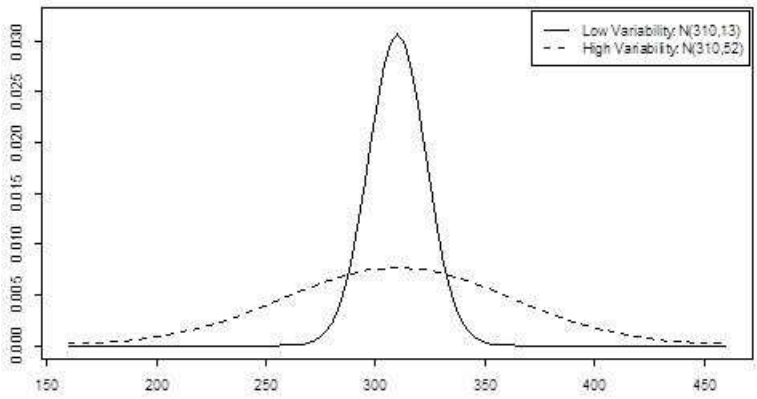


Figure 3.2. Comparison of normal densities with different degrees of variability.

In many cases, because of underlying assumptions of normality with reasonably low to moderate variability, such observations would be treated as outliers and potentially "cleaned" from the dataset. However, in highly variable conditions, there is a much greater probability that such an observation is not only legitimate but a *likely* result and therefore should *not* be omitted from the data, as it changes the underlying distribution of the data. Notwithstanding, such observations could disproportionately influence the sample mean and standard deviation, which could drastically affect the quality of the estimate used.

This is important. Given that most *RPD* problems contain only a handful of observations at each experimental design point, the quality of the estimates used to support regression analysis and thus optimization approaches will be affected considerably by inherent system variability. If this is known to be low, then "extreme" observations might be treated with a greater degree of scrutiny; if it is known to be elevated or high, then they are more likely to be valid. However, this valid point may very well induce an errant shift in the estimate nonetheless. For example, consider the set of four observations in Table 3.1 taken at a particular design point in an *RPD* experiment involving a highly variable process. Three of the four points are clearly grouped relatively close together, with the fourth, Y_4, deviating from this group noticeably. The sizeable standard deviation reflects the elevated variability. Note further the mean of the observations both with this extreme point and without. Lastly, note the median of the four observations.

Table 3.1. Sample of replicates taken from a single design point in a hypothetical experiment.

Y_1	Y_2	Y_3	Y_4
192	186	208	279

Standard Deviation:	42.85
Mean with y_4:	216.25
Mean without y_4:	195.33
Median of all obs.:	200.00

According to Willinger, *et al.* (2004), high variability is a phenomenon by which a set of observations assumes values that vary over orders of magnitude, with most observations taking values that are relatively close together, with a few extreme observations attaining values that deviate considerably from this first group with non-negligible probabilities, and with intermediate-sized observations occurring with appreciable frequencies. Thus, we would expect the tighter grouping of the three observations to be more reflective of the actual location of the process at the factor settings used for this particular design point. While using the mean of these points without including the "extreme" observation provides an estimate better aligned with this expectation, it completely disregards the data point which would be wrong, as it alters the underlying distribution of the data. On the other hand, using the mean for all four observations causes a moderate shift in the estimate due to the leverage imposed by the extreme point. The median, however, seems to provide a compromise between the two that acknowledges the location of the preponderance of observations without discounting the presence of observations in the tails, and therefore may be a better alternative. Once again, this illustrates the point that when actual conditions deviate from traditional assumptions, alternatives to the sample mean and standard deviation should be

considered for estimating the predominant location and scale of the underlying distribution.

3.3.2 Conditions-based selection of Tier-Two Estimators

In the traditional robust parameter design research, researchers usually assume that normality, homoscedasticity, and independence among the residuals are valid within a set of observational data so that the method of ordinary least squares may be used to obtain unbiased, minimum variance estimates. Realistically, however, asymmetry is much more likely to prevail in many industrial settings, particularly when *S*- and *L*-type quality characteristics and certain *N*-type characteristics are involved; effects that can become amplified if coupled with elevated degrees of process variability, which is also relatively common in practice. Under such conditions certain characteristics emerge. Namely, the combination of asymmetry and high variability would increase the propensity for outlying observations in the long tail of the distribution. We do not want to treat these as "errant" observations because they are actually indicative of "typical" process operations given the inherent conditions. Thus, the question becomes one of how to address the process conditions without discounting them altogether via assumptions.

Again, customary approaches to addressing high variability and asymmetry (each of which can also induce heteroscedasticity) include transformations or *GLMs*. However, the use transformations to deal with such situations can be problematic. We can rely on empirically chosen transformations using either logarithmic approaches, or other well-known methods to make the data conform to assumptions. However, as Counsell, *et al.* (2011) noted, identifying a suitable transformation is not always possible, and analyzing

data on a different scale could compromise interpretability. That is, rather than dataset properties informing the statistical analyses, often inappropriate or non-optimal methods that fail to fully exploit the data are used, and violations of key assumptions are deemed unavoidable nuisances rather than inherent aspects of true process conditions. Beyond this, there is simply no guarantee that a particular transformation will satisfy all the required properties for analysis.

*GLM*s, on the other hand, have been shown to be viable *RPD* alternatives under non-normality, but questions arise as to which distribution-link function combination should be used to model the responses. Often, *GLM*-based analyses of non-normal responses assume a known Poisson or gamma distribution, or perhaps an inverse Gaussian distribution. The Poisson works well for count data, but not for continuous data, and is often too inflexible to accommodate asymmetric situations. The latter two, on the other hand, work well for continuous positive data, but may be too different to the normal to be useful. In some cases, particularly in multi-response problems, it is often necessary to account for both the symmetry typically anticipated with *N*-type responses, as well as the asymmetry expected with *S*- and *L*-type responses. While the gamma distribution can be used to model asymmetric circumstances, it suffers in comparison to the normal when zero skewness is observed; and the inverse Gaussian can achieve some semblance of normality only when the variance becomes extremely large.

These potential issues with more traditional approaches suggest a need to consider other, non-traditional alternatives to the *OLS* approach when asymmetry and high variability pervade process outputs. A considerable array of alternative resistant and

robust approaches has evolved over the years, but only a handful of these have been examined in the context of *RPD* problems, which leads to the logical question that asks why. In fact, proponents of these alternative methods have demonstrated their superior performance over *OLS* regression in certain situations (namely those involving outliers). Notwithstanding, their use continues to be limited, particularly in the fields of quality engineering and robust parameter design. Hampel (1986) point out several reasons for this; most notably, that many of the alternative methods are significantly more computationally intensive than the *OLS* approach and that most statistical software packages at the time neglected to include implementations of them. Consequently, engineers and analysts, particularly those with less statistical background, tend to avoid their use. However, the advent and availability of considerably improved computing power coupled with much more refined statistics-based software (such as R) has rendered such concerns more or less irrelevant. Accordingly, engineers and researchers should be more aggressive in actively investigating ways to employ these alternative methods in the interests of finding the best *RPD* solutions possible.

As will be discussed in greater detail in Chapter 5, true normality is actually quite improbable in most applications. Realistically, asymmetry is much more likely to prevail, particularly in situations involving *S*- and *L*-type quality characteristics, and certain cases of *N*-type characteristics; a condition likely to be amplified if occurring concomitantly with high variability in process outputs. Two factors to recognize when asymmetric conditions exist are that the mean will shift away from the central tendency of the distribution, and the standard deviation will no longer accurately describe the dispersion

as a result of the "play" in the long tail of the distribution. Recalling the comparison between a skew normal density and a normal density with a common mean in Figure 3.1, it is clear that in the normal case, the mean and values immediately adjacent to it correspond to the peak in the density function, indicating that the likelihood of those values exceed all others in the distribution. In the asymmetric case, the mean clearly lies to the right of the preponderance of probabilities, whereas the median of the asymmetric example falls very close to the peak of the distribution. Thus, in the latter instance, while the mean certainly still defines central tendency in terms of the population mean, it does not necessarily correspond to central tendency in terms of the values of greatest likelihood. What is more, the non-normality in such cases tends to also correspond to non-constant variance, or heteroscedastic conditions, which complicates the use of *OLS* even further.

Chapter 2 illustrated the wide array of both tier-one and tier-two estimation methods that has evolved and that several researchers have examined the differences between the performances of various estimators. However, the investigated conditions almost singularly focus on situations involving outliers. But this should come as no surprise, as most alternative techniques to the estimation methods used in the *OLS* framework grew out of concern for the impact of outliers on data and subsequent estimates. But a variety of other conditions, such as asymmetry, can also exist and have equally detrimental effects on the quality of the estimates in both tiers. Despite the large pool of potential alternatives, few efforts have explored them in the context of robust

parameter design, and among these, the only conditions examined include outliers (see Tables 2.2 and 2.3).

Depending on which estimator is used in a given situation, different sets of optimum operating conditions will likely result, which complicates comparison studies and data analysis. As such, given the considerable array of available choices, the intrinsic question centers on which method(s) to use and when. More specifically, how do inherent conditions in the data or the actual process influence the selection of estimation approaches in order to achieve the best robust design solutions? In spite of the numerous comparative studies available in statistical literature, these questions have not been adequately addressed in the robust design research community. Accordingly, what is needed for engineers is clarification as to which estimators should be used under certain sets of conditions. This requires analysis that examines conditions beyond just outliers to determine which estimators deliver optimal performance in terms of achieving the best *RPD* solutions. Chapters 6 and 7 will address these questions, the former proposing a conditions-based approach to tier-estimation and the latter focusing on tier-two, or regression estimation.

3.4 Achieving Cost Robustness Under Dynamic Process Variability

The primary goal of robust design is to identify the optimal process settings that achieve desired targets for mean performance while simultaneously minimizing variability in the results. Despite extensive effort across a broad array of topics in *RPD* research, robustness in almost universally considered in the context of variability among process outputs. Sometimes, however, dynamic process variability cannot be avoided. In

such cases, it may be more appropriate to focus attention on stabilizing associated costs. Thus, considerable room for further advancement exists, particularly in the context of attaining cost-robust solutions. There are essentially four motivating factors that will be explored in detail in Chapter 7:

(i) Mixed multiple quality characteristics: Generally, the average customer will examine multiple characteristics to assess the quality of a particular product. While a variety of papers have addressed multiple quality characteristics to incorporate the realities found in application, these efforts have typically confined themselves to like-type characteristics. Consequently, very few have adequately addressed mixed multiple characteristics, which is even more realistic than considering like-type multiples alone.

(ii) Integrating disparate degrees of skewness between multiple characteristics: Because contemporary research has really only examined like-type characteristics in the multiple-characteristic problem, researchers have only addressed skewness in the limited context of the characteristic type they used. That is, they only considered asymmetric or symmetric distributions in their models, but not both. For *N*-type characteristics, virtually all researchers assume a symmetric normal distribution. In *S*- and *L*-type cases, researchers commonly assume distributions that exhibit positive and negative skewness, respectively. For example, Chan *et al.* (2003) employed a bivariate exponential distribution to model the positive skewness associated with S-type responses. However, when considering multiple mixed-type responses, it is important to account both for the asymmetry typically observed with *S*- and *L*-type responses and the symmetry associated with *N*-type responses. This cannot be done using different distributions for each of the

characteristics but rather through a multivariate version of a particular distribution. While various alternatives exist, this paper invokes the multivariate skew normal (*MSN*) distribution which allows us to integrate and examine the varying degrees of symmetry or asymmetry through the skewness parameter.

(iii) Known constant variability vs. range of variability: Nearly all process target efforts to date have assumed prior knowledge of the process mean and variance. While this common assumption certainly enhances the tractability of the problem, in real-world settings engineers rarely know these parameters with certainty. This is particularly true in industrial settings in which phenomena impacting the process cause the mean and variance to shift or drift. These phenomena could include environmental conditions such as temperature, humidity, and pressure; operator arrangement in the workplace; seasonal demand; or perhaps a change in product output. Whatever the causes, it seems more realistic to assume that an engineer would be aware of a *range* of variability and recognize that, as the process transitions through this range, the location of the optimal process target vector, $E[\mathbf{Y}^*]$, that minimizes total cost will likely shift with it.

(iv) Achieving cost robustness: While the process mean literature is replete with discussions about process variance, it is comparatively rare to see this problem discussed in the context of robust design. Pioneered by Taguchi (1984 and 1986), robust design methods focus on improving engineering and manufacturing productivity by actively considering the effects of system variation on process performance. In short, the customer's perspective demands that manufacturers consider both the target value and the variability in the response for a particular quality characteristic in order to improve both

the quality of the delivered product and the process used to deliver it. Many researchers have explored this particular area within quality engineering, focusing predominantly on model estimation and optimization schemes for determining the optimal process settings that achieve specified target settings while minimizing variability. In virtually all of the research in this area, "robustness" is typically assessed in the context of system responses. That is, determining the process settings that will render these responses or quality characteristics of interest insensitive (or *robust*) to influencers of variability. However, we have yet to find an example wherein solutions are sought to achieve robustness in minimum cost.

3.5 Strategies for Improving Precision in Highly Variable Processes

In many cases, including each of the research efforts delineated in Section 2.1.1.2 of Chapter 2, researchers have developed optimization procedures using data sets from highly variable processes. This is not to imply that the developed procedures are faulty, as their proponents have clearly demonstrated improved results through a range of comparative analyses. Rather, it is to point out that the data used are so variable, that with only a handful of replications (typically three or four) these models would likely generate drastically different results spanning the breadth of the design space every time. Moreover, it is almost equally probable in such highly variable situations that any one of the schemes would produce the best result in a particular situation.

Perhaps one of the most widely used data sets for comparing *RPD* optimization schemes is the printing press study data introduced by Box and Draper (1987). Vining and Myers (*VM*) (1990) used the experiment to illustrate the effectiveness of the dual

response surface approach in identifying the optimal factor settings for the process. In the two decades since, a number of other researchers have used the same experimental data to propose a variety of *RPD* extensions to the dual response approach (Del Castillo and Montgomery (*DM*) (1993), Lin and Tu (*LT*) (1995), Copeland and Nelson (*CN*) (1996), Del Castillo *et al.* (1997), Kim and Lin (1998), Fan (2000), Tang and Xu (2002), Ding *et al.*(2004), Jeong *et al.*(2005), Kazemzadeh *et al.*(2008), Koksoy and Yalcinoz (2008), Chen and Ye (2009), and Das and Lee (2010)).

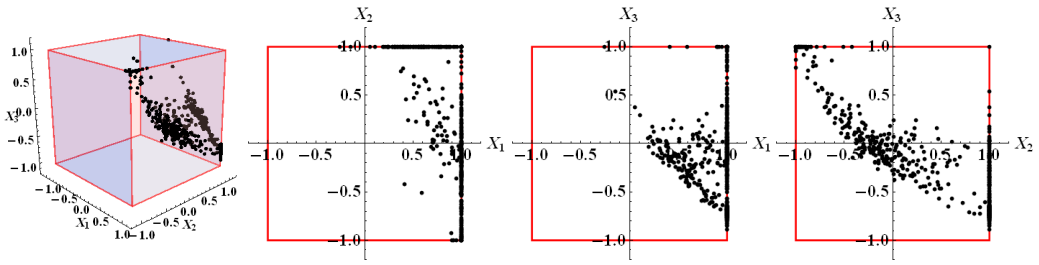
In this experiment, the printing machine's index in applying coloring inks to package labels is a normally distributed *N*-type quality characteristic of interest, *Y*. The control factors known to influence *Y* include the speed (X_1), pressure (X_2), and distance (X_3) settings for the machine, which are evaluated at three levels each. The desired target value for the machine's index is $\tau = 500$. The experiment considers the control factors at three levels for each and consists of three replications at each of the twenty-seven design points of the full factorial design. Table 3.2 displays the printing press data, along with the calculations for the mean and standard deviation at each design point.

Using this example, previous researchers developed second-order response surface designs for the mean and standard deviation and then applied their particular optimization scheme to find optimal solutions $\mathbf{x}^* = (X_1^*, X_2^*, X_3^*)$. Suppose we repeat this experiment 500 times using Monte Carlo simulation, using a random number generator to create observations following the mean and standard deviation specified at each design point in Table 3.2.

Table 3.2. Experimental Design for the Printing Press Study.

	Coded Units			Labeling Index Y (3 replications)			Mean	Standard Deviation
	Speed	Pressure	Distance					
Run	X_1	X_2	X_3	y_1	y_2	y_3	$\bar{y}$	s
1	–1	–1	–1	34	10	28	24.00	12.490
2	0	–1	–1	115	116	130	120.33	8.386
3	1	–1	–1	192	186	263	213.67	42.829
4	–1	0	–1	82	88	88	86.00	3.464
5	0	0	–1	44	188	188	140.00	83.138
6	1	0	–1	322	350	350	340.67	16.166
7	–1	1	–1	141	110	86	112.33	27.574
8	0	1	–1	259	251	259	256.33	4.619
9	1	1	–1	290	280	245	271.67	23.629
10	–1	–1	0	81	81	81	81.00	0.000
11	0	–1	0	90	122	93	101.67	17.673
12	1	–1	0	319	376	376	357.00	32.909
13	–1	0	0	180	180	154	171.33	15.011
14	0	0	0	372	372	372	372.00	0.000
15	1	0	0	541	568	396	501.67	92.500
16	–1	1	0	288	192	312	264.00	63.498
17	0	1	0	432	336	513	427.00	88.606
18	1	1	0	713	725	754	730.67	21.079
19	–1	–1	1	364	99	199	220.67	133.822
20	0	–1	1	232	221	266	239.67	23.459
21	1	–1	1	408	415	443	422.00	18.520
22	–1	0	1	182	233	182	199.00	29.445
23	0	0	1	507	515	434	485.33	44.636
24	1	0	1	846	535	640	673.67	158.210
25	–1	1	1	236	126	168	176.67	55.510
26	0	1	1	660	440	403	501.00	138.935
27	1	1	1	878	991	1161	1010.00	142.454

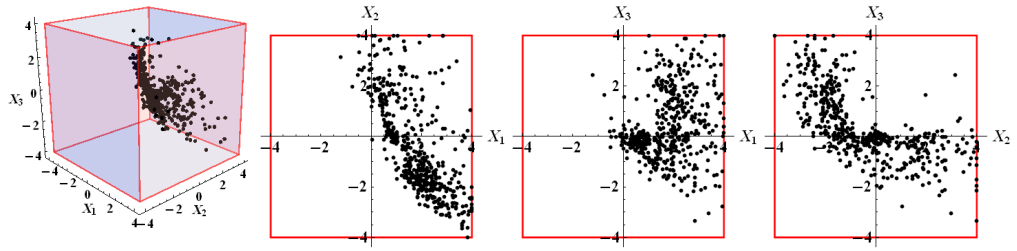
Figure 3.3(a) shows the set of all points **x*** identified using *LT*'s optimization scheme outlined in Equation (8.2). The corresponding figures in (b)-(d) depict the position of this set of solutions as it relates to each cross-section of factors. A random search technique with a large search point array facilitates obtaining global solutions at each simulation iterate.



(a) (b) (c) (d)

Figure 3.3. Identification of x*, Printing Press Study, $-1 \le X_i \le 1$.

As is the case with most of the solutions generated by previous researchers examining the printing process data, the experimental region $-1 \le X_i \le 1$ is found to be binding. When we relax this constraint and repeat the simulation, we readily observe the extent of the extrapolated optimal operating region for this example (see Figure 3.4). It is noteworthy that although this is an extrapolated view of the design space, it is necessary in order to actually see what is transpiring. Nevertheless, in spite of this, it still captures the imprecision among the generated solutions.



(a) (b) (c) (d)

Figure 3.4. Identification of **x***, Printing Press Study, $-4 \le X_i \le 4$.

Suppose now we perform 1,000 iterations of the simulation, evaluating the more commonly used optimization models proposed by *VM* (1990), *LT* (1995), *CN* (1996), and Costa (2010) at each iterate. Assuming each of the researchers uses the same full second-order model, a comparison of the optimal factor settings **x*** is performed using the *MSE*

criterion shown in Equation (8.4). Table 3.3 depicts the results as a proportion of the 1,000 iterations a particular model generated the best solution. It is interesting to note that, while each researcher indicated either an improvement or equally acceptable result in the literature, it is the original model by *VM* (1990) that more consistently results in a lower overall *MSE*. Thus, as these results show, the inherent variability of this particular data set greatly influences the ability of any one optimization model to achieve superior results with any certainty.

Table 3.3. Model Comparison – Printing Press Study.

	Results (% of 1000 iterations with Minimum *MSE*)				
Model	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
Vining and Myers (1990)	55.2%	58.1%	56.7%	55.8%	57.8%
Lin and Tu (1995)	39.2%	37.2%	37.3%	38.7%	37.0%
Copeland and Nelson (1996)	2.5%	0.9%	2.1%	2.2%	2.0%
Costa (2010)	3.1%	3.8%	3.9%	3.3%	3.2%

This brief illustration demonstrates the drawbacks in the design of the *RPD* problem for processes that exhibit elevated degrees of variability. Since the number of replications at each design point is low, repeating the experiment under the same conditions (i.e. same number of replications) can result in the identification of settings that may span the entire length or more of the factor space. This indicates considerable imprecision in the estimated functions used to generate the factor settings. More importantly, it creates a false perception in the superiority of one optimization model over another; the data are so variable that the results should not be trusted with any degree of certainty.

Ultimately, in answering the call for methods and techniques that will help manufacturers and decision makers improve the quality of their processes and products, engineers should examine various trade-offs to best support the decision making process. Certainly, time, costs, and other resource constraints will impact the experimental approaches employed, but this should be weighed against the need for elevated levels of precision. In short, if a model is imprecise in estimating the appropriate response surface, then optimizing that model is going to yield equally imprecise results as they relate to the true optimal settings. Accordingly, the results and recommendations provided to decision makers could generate suboptimal modifications to processes and products alike. Thus, an obvious question arises: once an estimation approach has been selected, what measures can be taken to ensure the most precise result possible, particularly in processes with naturally high variability? Several alternatives exist, such as increased replication, the development of higher order response surface models, or even removing sources of excessive variability. Chapter 8 examines this question in terms of the latter, proposing a technique based on the coefficient of variation as a way to identify and remove influential sources of process variability, thereafter incorporating optimal design concepts to rebalance the experimental framework . Chapter 9 addresses the question of precision using all the alternatives together, proposing the integration of a trade-off analysis that weighs the need for enhanced precision to determine the true optimal process settings against the costs of obtaining it.

CHAPTER FOUR

IMPROVED ANALYTICAL APPROACHES USING PROBABILITY PLOTS

4.1 Introductory Remarks

Effective quality improvement programs begin with the collection and subsequent examination of data. Given this, the importance of proper data analysis in the initial stages of process analysis cannot be overstated, as it underpins all assumptions for and approaches to process modeling, estimation, and ultimately optimization. However, merely analyzing data is not sufficient from the point of view of making decisions. How and what one interprets from the analyzed data may be-more important. Thus, data analysis comprises a decision support system rather than a decision making system; and as with any such system, the quality and reliability of the decisions that emanate from it will be only as good as the quality and depth of the analysis put into it. In particular, effective data analyses can offer the following benefits:

- Structuring the findings from research or other means of data collection
- Decomposition of a macro picture into a micro one
- Acquiring meaningful insights from the dataset that facilitate better understanding of inherent statistical characteristics
- Basing critical decisions from the findings
- Ruling out human bias through proper statistical treatment

Many of the statistical procedures used in quality engineering require the validity of certain assumptions as they relate to the distribution of the data. Recognizing the critical role that such assumptions play in nearly all formal statistical analyses, the

question becomes how to proceed with data analysis to develop and eventually validate those assumptions. Many statistical tests have evolved over time and are widely used to provide an objective basis for answering this question. These tests are important, as an assumption about a particular distribution affects how we estimate parameters, which in turn influence the estimation of statistical models, and ultimately optimization results. However, these tests essentially only answer one question: the data either are or are not distributed as hypothesized. While we generally hope for the former result, as it provides statistical support for our assumptions, the latter result is problematic simply because it provides no other information as to what the underlying distribution might be.

This highlights the utility and importance of graphical methods in understanding what is communicated by the data in terms of how they are distributed and what sorts of characteristics that distribution has. Such graphical methods might include histograms, half-normal plots, percentile-percentile (P-P) plots, or quantile-quantile (Q-Q) plots, among others. Of these, Q-Q plots, also known as probability plots, are used more often as a graphical means for comparing the actual distribution for a set of data against a hypothesized distribution for the population.

Although such plots can be constructed for any distribution, we most often hear of them in the context of normal probability plots. The reason for this stems from the typical assumptions of normality inherent in many statistical procedures. Beyond this, however, the use of normal probability plots as an initial analytical basis for evaluating normality within a set of data makes logical sense. In general, most people, both statisticians and non-statisticians alike, are very familiar with the bell-shaped curve

associated with the normal distribution compared with the appearance of other more obscure distributions. However, detecting departures from a particular curved shape can be difficult compared to assessing deviations from a straight line. Thus, the use of normal probability plots provides a way to quickly and effectively analyze data in ways that facilitate understanding as to how the data relate to or deviate from a normal distribution.

Most of the statistical analyses put forward in the continuous quality improvement literature place little emphasis on the up-front data analysis that provides the basis for their assumptions and ensuing selection of approaches to addressing a particular problem. The focus of this chapter is to digress momentarily from that avenue in the interests of revisiting some of the basic principles of data analysis that buttress virtually all of the assumptions for and approaches to modeling and optimization. Thus, the goal is to provide a comprehensive investigation of shape analyses of normal probability plots that allow a fuller understanding of data and how certain efforts and actions to improve process performance may influence them in terms of revealing or masking their underlying characteristics. The original work for this research is published, with reference Boylan and Cho (2012a).

4.2 Background

4.2.1 Normal Probability Plots

Of the various assumptions required by most statistical procedures used in quality engineering, one of the most common is that process data follow a normal distribution. Consider statistical process control as an example, wherein an assumption of normality

underpins the calculation of control limits in the control charts for both individuals and averages. If analysts use these charts for situations in which the data in fact are not normally distributed, points outside the control limits may be due to the skew of the data and not due to special-cause variation. As is well known, the normal distribution yields values that range from minus infinity to plus infinity. However, it is equally well known that this rarely reflects real life. Fortunately, statistical procedures are fairly robust in the sense that the underlying distribution need not be precisely normal. Instead, they require that the data be "near-normal." So if data need to be near-normal, how does one test for that? For many decades, statisticians have invoked a variety of graphical methods for examining data to determine how they are distributed and what sorts of characteristics those distributions have. Such graphical methods have included histograms, half-normal plots, percentile-percentile (P-P) plots, or quantile-quantile (Q-Q) plots, among others. Of these, Q-Q plots, also known as probability plots, are the most often used means for comparing the actual distribution for a set of data against a hypothesized distribution for the population.

Although such plots can be constructed for any distribution, we most often hear of them in the context of normal probability plots. The reason for this stems from the typical assumptions of normality inherent in many statistical procedures. Beyond this, however, the use of normal probability plots as an initial basis for data analysis makes logical sense. In general, most people, both statisticians and non-statisticians alike, are very familiar with the bell-shaped curve associated with the normal distribution compared with the appearance of other more obscure distributions. Yet, detecting

departures from a particular curved shape can be difficult compared to assessing deviations from a straight line. By invoking a straight line to represent normality within a dataset, the use of normal probability plots provides a way to quickly and effectively analyze data in ways that facilitate understanding as to how the data relate to or deviate from a normal distribution.

Simply stated, a probability plot is a graph of the relative cumulative frequencies of the data, using a specific, and relatively straightforward plotting convention. Let ($y_{(1)}$, $y_{(2)}$, ..., $y_{(n)}$) denote an ordered sample of data from a hypothesized distribution with a cumulative distribution function of the form $F(Y)=[(y-\mu)/\sigma]$, where μ and σ represent the location (mean) and scale (standard deviation) parameters, respectively. Building a probability plot then proceeds by plotting the sample order statistic $y_{(i)}$ on the horizontal axis against $z_{(i)} = F^{-1}(p_i)$ on the vertical axis. The result is a plot that appears as the example shown in Figure 4.1. Data strewn along the 45° reference line in the plot would suggest normally distributed data. Conversely, deviations from this line would suggest otherwise.

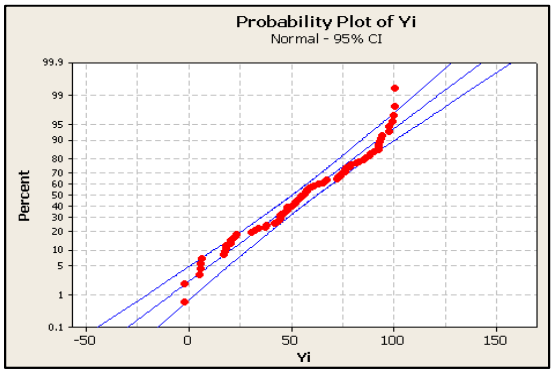


Figure 4.1. Example of a normal probability plot.

The term, p_i represents an estimate of the probability associated with $F[(y_{(i)} - \mu) / \sigma]$ and is commonly referred to as the plotting position. Over the last century, a number of researchers have developed a variety of approaches for establishing plotting positions in probability plots. Those developed by Hazen (1914) and Weibull (1939) are among the earliest and are still commonly used. As Vining (2010) points out, most statisticians continue to use the plotting position developed by Hazen (1914) which is denoted by:

$$p_i = \frac{i - 0.5}{n} \quad \text{where } n \text{ is the sample size} \tag{4.1}$$

Another approach commonly used in various textbooks, such as Kutner *et al.* (2005), is that of Blom (1958), who defined the plotting position for the normal distribution by:

$$p_i = \frac{i - 0.375}{n + 0.25} \tag{4.2}$$

As noted in Chapter 2, a number of researchers examined the topic of plotting positions and developed a variety of approaches to it. Looney and Gulledge (1985) provided an overview of these works as they relate to normal probability plots.

4.2.2 Sample Moments

Regardless of the purposes behind its use in a particular study, the normal probability plot produces a curve that reflects the distribution of the data relative to the normal distribution, which is denoted by a straight line. As DeCarlo (1997) points out, typical introductory courses in statistics comment on the ability to characterize a distribution in terms of its "central tendency, variability, and shape." While most textbooks define and illustrate shape in the context of skewness, kurtosis is another aspect

of shape that has equally important implications for the data. In any event, these qualities essentially correspond to the moments of a distribution. Thus, the shape of the curve in the probability plot and the manner in which it deviates from the reference line provide information about the characteristics of the underlying distribution in terms of the first four sample moments.

From a statistical or mathematical perspective, moments denote quantitative measures that define the shape or density of a set of points. More specifically, moments represent the expected value of various powers of a random variable. Thus, for each random variable *X* and every positive integer, *r*, the r^{th} moment of *X* is denoted by $E[X^r]$. According to DeGroot and Schervish (2002), the r^{th} moment is said to exist if and only if $E[|X|^r] < \infty$. Furthermore, if *X* is bounded by finite numbers *a* and *b* such that the $\Pr(a \le X \le b) = 1$, all moments of *X* must exist.

Central moments differ from ordinary moments in that they examine the spread about the mean. Mathematically, $E[X^r]$ changes such that, for every positive integer *r*, the r^{th} central moment of *X* is given by $E[(X-\mu)^r]$. Said differently, this equation represents the r^{th} moment of *X* about the mean. In general, many prefer the use of central moments as a more useful means for characterizing a distribution, mainly because the higher-order moments then relate solely to the spread and shape of the distribution as opposed to just its location. In general, we consider four central moments to characterize the distribution of a particular population. The first two consist of the population mean and variance, respectively. The third and fourth moments measure the degree to which

the distribution is skewed from or peaked around the mean, more commonly known as skewness and kurtosis. Each will be examined in detail in the following paragraphs.

Analysis of the sample central moments themselves says very little regarding the details of the associated probability density of a distribution. Rather, it provides descriptive information about the characteristics of the density. For a univariate real-valued random variable X with *an* independent observations ($x_1, x_2, ..., a_n$), the sample statistic for the r^{th} sample central moment is defined as:

$$m_r = \left(\sum_{i=1}^{n} (x_i - m)^r \right) \Big/ n, \text{ for } r \geq 2, \tag{4.4}$$

Here, m corresponds to the sample mean and is defined as the expectation of X, or $E[X] = \left(\sum_{i=1}^{n} x_i \right) \Big/ n$ which is an unbiased estimator for the population mean μ. It follows that the first central moment for any distribution must equal 0, as $m_1 = E[(X - \mu)] = E[X] - \mu = \mu - \mu = 0$. Thus, using Equation (4.4), the first sample central moment is given by

$$m_1 = \left(\sum_{i=1}^{n} (x_i - m) \right) \Big/ n = 0 \tag{4.5}$$

The second moment about the mean is the population variance, σ^2. Thus, when $r = 2$ in Equation (4.4), we can exchange the denominator n for $n-1$, due to the fact that μ is unknown, and then easily obtain the unbiased estimator for the population variance which is given by $\tilde{s}^2 = \left(\sum_{i=1}^{n} (x_i - \bar{x})^2 \right) \Big/ n - 1$. This translates to the second sample central moment which is

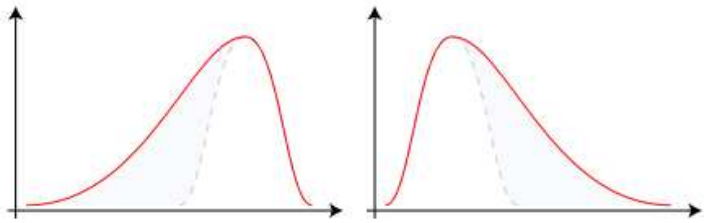
$$m_2 = \left(\sum_{i=1}^{n} (x_i - m)^2 \right) \Big/ n-1. \tag{4.6}$$

Although these first two moments completely characterize the distribution in the context of its location and scale, the third and fourth moments facilitate the examination and understanding of any inherent asymmetry and peakedness in the probability density function.

Skewness(γ) comprises the third central moment and measures the asymmetry of the distribution relative to the mean. It can be positive or negative, or 0 in the case of a normal distribution, which is symmetric. Quantitatively, the sample skewness g_1 is given by:

$$g_1 = \frac{m_3}{m_2^{3/2}} = \frac{\frac{1}{n}\sum_{i=1}^{n} (x_i - m)^3}{\left(\frac{1}{n-1} \sum_{i=1}^{n} (x_i - m)^2 \right)^{3/2}} \tag{4.7}$$

Qualitatively, this implies that when γ > 0, the right tail of the distribution is longer and heavier and the preponderance of values, including the median and the mode, generally lay to the left of the mean. Conversely, γ < 0 implies a heavier and longer left tail, with the bulk of values laying to the right of the mean. Figures 4.2(a) and (b) portray negative and positive skew, respectively.



(a) (b)

Figure 4.2. Portrayals of (a) negative and (b) positive skewness.

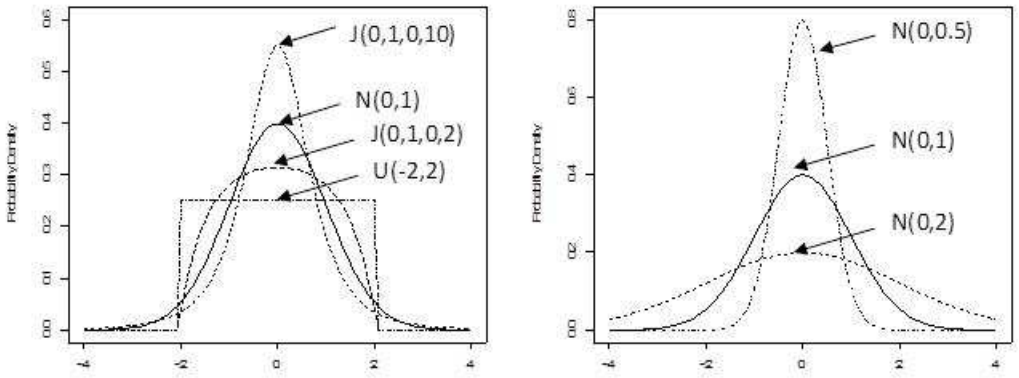
The fourth moment of a distribution is called kurtosis, which represents a classical measure of non-Gaussianity. Sheskin (2007) notes that the primary reason for measuring kurtosis is to determine whether or not sample data derived from a normally distributed population. Whereas skewness refers to the asymmetry in the distribution with respect to the mean, population kurtosis measures the peakedness of the density and is typically defined as the fourth moment about the mean divided by the square of the variance minus three. This is shown mathematically by $Kurtosis = \left[\left(\sum (X-\mu)^4\right) \Big/ N\sigma^4\right] - 3$. As DeCarlo (1997) and a number of others point out, because the natural kurtosis of the normal distribution is 3, the "– 3" in this formula serves to make the "excess" kurtosis of the normal distribution equal to 0. For a sample of size *n*, the measure to estimate population kurtosis g_2 is determined by:

$$g_2 = \frac{m_4}{m_2^{3/2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - m)^4}{\left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i - m)^2\right)^2} \; - \; 3 \tag{4.8}$$

In practical terms, values of $g_2 > 0$ result in a distribution with a high or sharp peak about the mean and heavier tails, formally referred to as a leptokurtic distribution. This translates to a lower probability of achieving values near the mean and a higher probability of extreme values, as compared with a normally distributed random variable. That is, higher excess kurtosis implies that the variance is more heavily influenced by infrequent outliers rather than more frequent and modestly-sized deviations from the mean. Values of $g_2 < 0$ result in platykurtic distributions that possess a lower, flatter peak about the mean and lighter or thinner tails. In this case the comparisons to the normal distribution are precisely the opposite of those in the leptokurtic case just

described. The case of $g_2 = 0$ describes a mesokurtic distribution that possesses zero excess kurtosis, such as the normal distribution.

Figure 4.3(a) depicts several examples of distributions exhibiting varying degrees of kurtosis where *J*, *N*, and *U* denote Johnson, normal, and uniform distributions, respectively. It is important to note that, in each case, the distributions with positive or negative kurtosis cross the normal curve twice on each side of the mean, which is a common trait of distributions with excess kurtosis. For distributions with excess positive kurtosis, such as the *J*(0,1,0,10), this results in a pattern of higher-lower-higher probabilities on each side of the peak, relative to each crossing point. For example, consider the *J*(0,1,0,10) relative to the *N*(0,1) in Figure 4.3(a). For values near the mean, the probabilities are higher for the *J*(0,1,0,10) distribution. Moving left or right from the mean, the densities cross and the relationship shifts to lower probabilities for the J(0,1,0,10) and then shift back to higher probabilities moving further to the left or right. For distributions with negative excess kurtosis, such as the uniform, the pattern becomes lower-higher-lower. This is a very important distinction, as when they cross only once, this is indicative of changes in variability vice kurtosis. The example of the normal distributions with different variances in Figure 4.3(b) depicts this. Although the estimators for sample skewness and kurtosis are biased and therefore tend to underestimate the true values for the population, the consideration of these moments is necessary, as no probability distribution could technically be described as Gaussian normal unless both equal zero.



(a) (b)

Figure 4.3. Graphical representations of (a) distributions possessing varying degrees of excess kurtosis relative to a $N(0,1)$ and (b) three normal distributions with different variances and zero excess kurtosis.

4.3 Proposed Shape Analysis of Probability Plots

4.3.1 Methodology

In order to perform the shape analysis, plots are required using data that possess predetermined values for each of the first four moments about the mean. This would facilitate an investigation of the effects of modifying one parameter at a time, while keeping the other three fixed in order to observe the impact on the probability plots. In turn, this would allow us to develop inferences about the data. Given the need to specify values for each of the sample moments, we used Monte Carlo simulation in the R statistical software package to generate random samples from predetermined distributions. Within R, we used two different distributional packages to examine the effects of skewness and kurtosis.

Due to the inherent relationship between the normal and skew normal distributions, the skew normal package in R (Package "*sn*", Azzalini (2011)) is used to examine the effects of asymmetry or skewness on the probability plots. First introduced

by O'Hagan and Leonhard (1976) and addressed more recently by Azzalini (1985), Azzalini and Valle (1996), and Arellano-Valle, *et al.* (2004), the skew normal extends the normal distribution by invoking a third parameter, α, to account for non-zero skewness. Taking this third parameter into account, the probability density function for the skew normal relative to the normal distribution is given by $f(x) = 2\phi(x)\Phi(\alpha x)$. Adding location ($x$) and scale ($w$) parameters to this density function using the transformation $x \to (x-\xi)/\omega$, yields:

$$\frac{1}{\omega\pi} e^{-\frac{(x-\xi)^2}{2\omega^2}} \int_{-\infty}^{\alpha\left(\frac{x-\xi}{\omega}\right)} e^{-\frac{t^2}{2}} dt \tag{4.9}$$

It can be shown that when $\alpha = 0$, the skew normal distribution reduces to the normal. Using the following syntax in *R*, myriad samples of 100 variates are simulated and stored:

```
require(sn)
x <- rsn(100, 0, 1, 4)
write.csv(x, file="skewdata1.csv")
```

where the command in the second line actually pertains to the simulation of random variates from a skew normal distribution with location = 0, scale = 1, and shape = 4 (i.e., *SN*(0, 1, 4)).

To examine kurtosis, and ultimately the combined effects of variability, skewness, and kurtosis, we turned to the Johnson family of distributions in R. Specifically, we used the JohnsonFit(●) function, which invokes the algorithm of Hill *et al.* (1976) to estimate the Johnson parameters (γ, δ, ξ, λ, type) using the sample moments. The syntax for the function is simply `JohnsonFit(t, moment="xx")` where `t` denotes either an

observation vector, `t=x`, or moment vector, `t=[mean, m2, m3, m4],` and `moment` is a character scalar that specifies how to interpret the vector `t`. As an example, the following R code demonstrates an instance of this function:

```
require(SuppDists)
parms<-JohnsonFit(c(0,1,-5,4),moment="use")
x <- rJohnson(100, parms)
write.csv(x, file="data1.csv")
```

Here, the values of 0, 1, -5, and 4 denote the vector `t.` The `"use"` command specifies this as the vector of moments, which the program then uses to estimate the parameters of the distribution. Thereafter, we simply simulate 100 random variates using these parameters and write these data to an *Excel*-based file for analysis. Using this particular function allows for quickly and easily varying the levels for each of the for sample moments in order to assess the impact on the shape of the curves in the normal probability plots.

In each case of the analysis, simulation is used in R to generate data samples according to the specified distributions and imported them into *Excel* for some preliminary analysis and spreadsheet calculations. Thereafter, they are imported into *Minitab* to generate normal probability plots for visual analysis of the results. It is noteworthy that all of this could have been performed easily within R. However, the use of *Excel* and *Minitab* provide a more practical means for comparative analysis and the simultaneous visual inspection of multiple data sets and plots. The next two sections will investigate skewness and kurtosis, respectively. Thereafter, the combined effects of the four moments on the data shall be examined and the findings summarized.

4.3.2 Investigating the Impacts of Skewness on the Shape of the Probability Plot

The investigation of skewness began with the establishment of a base case from which to deviate to examine the effects on the curves in the probability plots. For this purpose, a sample of 100 random variates was generated from a skew normal distribution with location = 5, scale = 2, and shape = 0. The probability plot associated with these data is shown in Figure 4.4. As expected, with the scale parameter set to zero, the resulting distribution essentially follows a normal distribution with a mean of 5 but with a variance closer to 4 than 2. The associated *A-D* statistic and *p*-value for the plot confirm the normality of the underlying distribution.

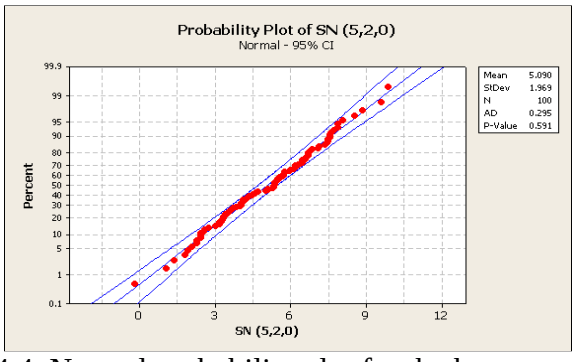
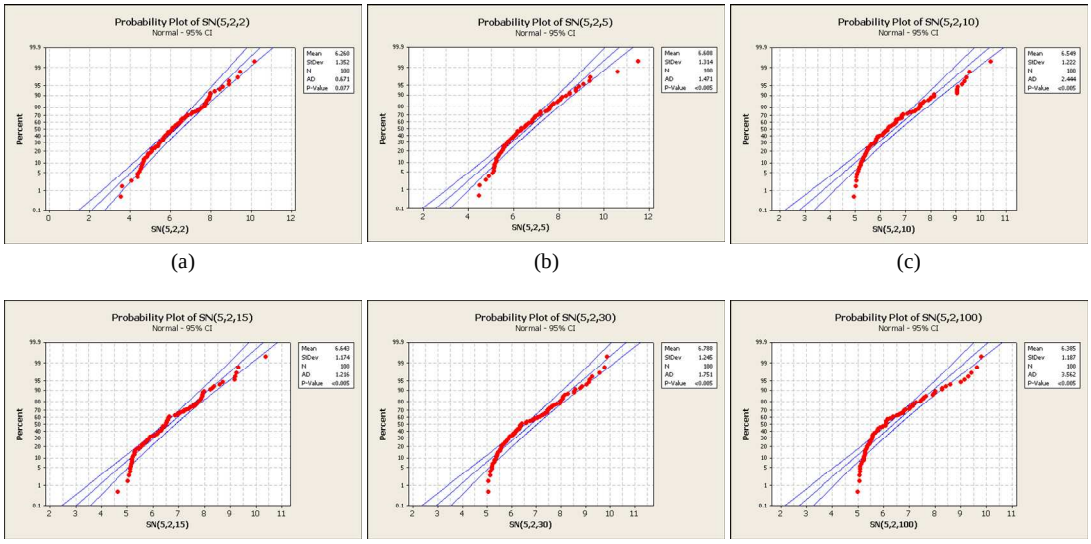


Figure 4.4. Normal probability plot for the base case of data sampled from a *SN* (5, 2, 0).

From here, we can begin to examine how this plot changes as a result of changes to the skewness in the data. Because changing the mean merely shifts the distribution along the x-axis without changing the visual properties of the plot, we only examined two cases. In the first, the location and scale parameters are fixed while varying the shape (or skewness) parameter. In the second, the location and the shape are fixed while varying the scale (or variability).

4.3.2.1 Assessing the Effects of Varying Degrees of Skewness

To examine the effects of the skewness parameter on the probability plot, the location and scale parameters were fixed at 5 and 2, respectively. Thereafter, six levels are considered for both positive and negative skewness: ±2, ±5, ±10, ±15, ±30, and ±100. Pursuant to generating the random variates for each of these twelve distributions, the simulated data were imported into *Minitab* to create the normal probability plots displayed in Figures 4.5 and 4.6; Tables 4.1 and 4.2 contain the sample statistics associated with each, respectively.



(a) (b) (c)

(d) (e) (f)

Figure 4.5. Normal probability plots showing the effects of positive skewness.

Table 4.1. Sample statistics associated with probability plots in Figure 4.5.

	Baseline	Fig 4(a)	Fig 4(b)	Fig 4(c)	Fig 4(d)	Fig 4(e)	Fig 4(f)
	SN (5,2,0)	SN(5,2,2)	SN(5,2,5)	SN(5,2,10)	SN(5,2,15)	SN(5,2,30)	SN(5,2,100)
Mean:	5.090	6.260	6.608	6.549	6.643	6.788	6.385
Standard Dev:	1.969	1.352	1.314	1.222	1.174	1.245	1.187
Sample Skew:	-0.056	0.476	1.036	1.010	0.655	0.587	1.080

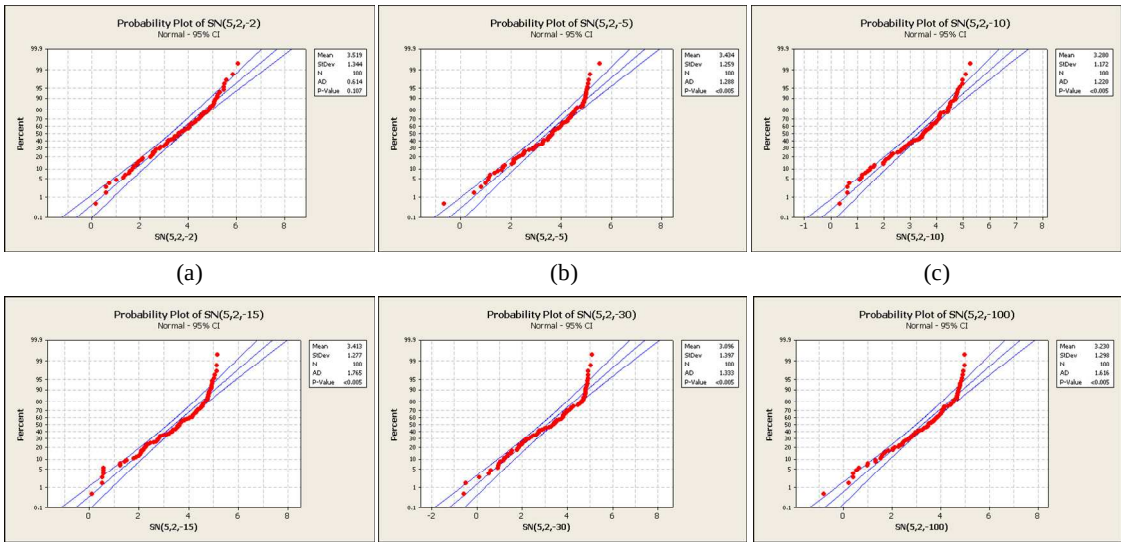


Figure 4.6. Normal probability plots showing the effects of negative skewness.

Table 4.2. Sample statistics associated with probability plots in Figure 4.6.

	Baseline	Fig 5(a)	Fig 5(b)	Fig 5(c)	Fig 5(d)	Fig 5(e)	Fig 5(f)
	SN (5,2,0)	SN(5,2,-2)	SN(5,2,-5)	SN(5,2,-10)	SN(5,2,-15)	SN(5,2,-30)	SN(5,2,-100)
Mean:	5.090	3.519	3.434	3.280	3.413	3.096	3.230
Standard Dev:	1.969	1.344	1.259	1.172	1.277	1.397	1.298
Sample Skew:	-0.056	-0.370	-0.722	-0.591	-0.696	-0.496	-0.786

Focusing on the positively skewed plots in Figure 4.6, several key observations emerge. Foremost, relative to the base case shown in Figure 4.4, as the skewness parameter increases, the associated asymmetry becomes more visibly apparent by the concavity of the plot. This should make sense based on the previous discussion of positive versus negative skewness. That is, since the bulk of the values in the distribution fall to the left of the mean due to the positive skewness, we would expect the percentage (or probability) to rise much more sharply moving toward the mean from left to right, followed by a more tapered increase thereafter.

A close examination of the plots relative to the base case also shows changes in the likely range of values of the samples. Although asymptotic to the *x*-axis in both directions, a shortening and thinning of the left tail of the distribution is observed while the right tail elongates and grows heavier. This is important, particularly when considering that an increasing trend in skewness seems to correspond to a slightly increasing trend in the sample mean coupled with a moderately decreasing trend in the sample standard deviation. Thus, as the right tail grows heavier, the variability grows much more sensitive to extreme observations. The observations for the negatively skewed plots in Figure 4.5 mirror those for Figure 4.6 from precisely the opposite viewpoint.

4.3.2.2. Effects of Increasing/Decreasing Variability in the Presence of Skewness

Investigating the effects of variability on the asymmetry of the data required fixing the location and shape parameters, while incrementally increasing and decreasing the scale or variance parameter. For the purposes of this experiment, a base case was established using a skew normal distribution with a mean of 5, variance of 2, and a skewness of 5. Six variations were then created on this case, three of which involved increasing values for the variance while the remaining three involved decreasing values. After using R to generate 100 random variates for each of these six distributions, *Excel* and *Minitab* are again used to generate sample statistics and normal probability plots. Figure 4.6 shows the six variations examined and the corresponding normal probability plots generated in *Minitab*. Table 4.3 contains the sample statistics for each.

The plots in Figure 4.7 impart several interesting observations. As expected, the positive skew incorporated in the base case in Figure 4.7(a) evinces the characteristic concavity previously described in Section 4.3.2.1, along with a shortening and thinning of the left tail coupled with a corresponding elongation of the right tail. Beyond this, the plots clearly show that changes in variability have a much greater effect on the shape of the plot than changes in skewness. With each exponential increase in variability in Figures 4.7(e), (f), and (g), we observe a corresponding right shift in the mean as the right tale expands and thickens. This shift essentially causes the tell-tale "hump" of the distribution to expand to the right as well, which simultaneously induces a slight lengthening of the left tail. It is also noteworthy that, as we would expect, the residual error in the plots increases with the variability as evidenced by the increasing scale on the x-axis for each subsequent plot.

Figures 4.7(b), (c), and (d) depict plots in which we sequentially reduce the variance from 2, to 0.5, 0.4, and finally to 0.125. In contrast to the examples with increasing variability, here we observe the distribution shorten in both tails as the bulk of the values push in towards the mean of 5. Also note that the variance reduction begins to offset the skewness, inducing a degree of symmetry about the mean. Obviously, as the variance approaches zero, we would expect to observe a straight line, perpendicular to the *x*-axis, situated at the mean. The plots are clearly trending this way, as shown by Figure 4.7(d).

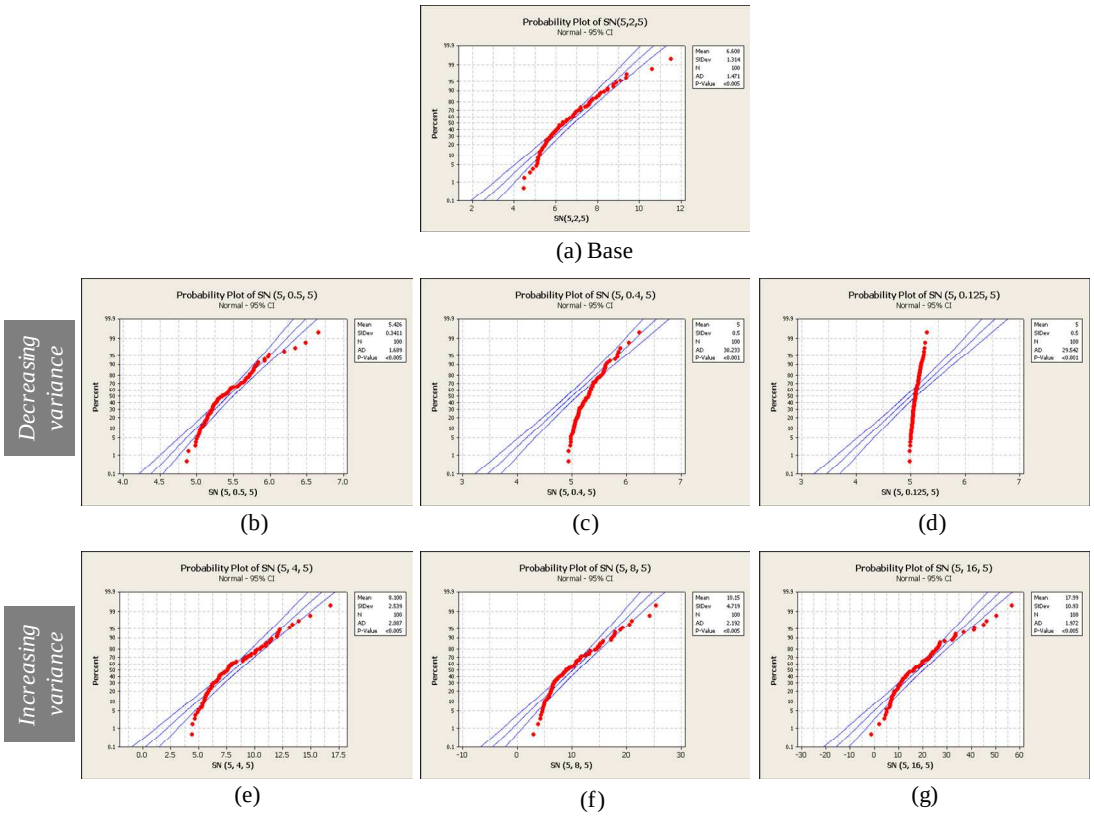


Figure 4.7. Normal probability plots showing the impact of decreasing and increasing variability.

Table 4.3. Sample statistics for the six variants depicted in Figures 4.7(b)-(g).

	Fig 6(b)	Fig 6(c)	Fig 6(d)	Fig 6(e)	Fig 6(f)	Fig 6(g)
Sample Mean:	5.1801	5.3157	5.4263	8.0998	10.1469	17.9946
Sample St Dev:	0.1538	0.2541	0.3411	2.5393	4.7193	10.9273

Notwithstanding, in spite of the evolving symmetry, reducing the variance exacerbates the departure from normality, as demonstrated by the density plot in Figure 4.8, which is associated with the probability plot in Figure 4.7(d). Figure 4.8 also suggests that the variance reduction also increases (or, more correctly, reveals) the presence of excess kurtosis. To examine this, a Monte Carlo simulation was developed in R to generate

1,000 random variates from three skew normal distributions with decreasing variance and then compute the sample kurtosis for each.

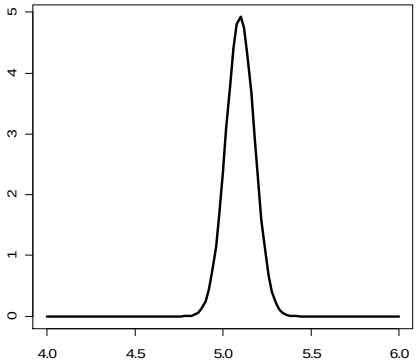


Figure 4.8. Density Plot for *SN*(5, 0.125, 5).

Upon replicating this 25 times for each distribution, the average sample kurtosis is computed across all 25 replications. As Table 4.4 shows, the change in the average sample kurtosis from one distribution to the next is quite negligible, although a slight downward trend is evident. This effect will be addressed in greater detail in the next section.

4.3.3 Investigating the Impacts of Kurtosis on the Shape of the Probability Plot

The investigation of kurtosis necessitated the use of the Johnson distribution due of the ease associated with modifying its characteristics in terms of the first four sample moments. Thereafter, it progressed in much the same way. As with skewness, a base case was first established that followed a normal distribution and possessed zero excess kurtosis. Shown in Figure 4.9(a), a sample of 100 random variates was simulated from a Johnson distribution with parameters (0, 1, 0, 3), which denote the values for each of the first four standardized sample moments, respectively. Given the absence of skewness

and excess kurtosis, shown by the last two parameters, this essentially mirrors a standard normal distribution.

Table 4.4. Simulation results showing the effects of decreasing variability on kurtosis.

	Distribution		
Replication	SN(5, 2, 5)	SN(5, 0.125, 5)	SN(5, 0.000001, 5)
1	3.8443	4.4361	4.6599
2	3.3164	3.8242	3.7289
3	3.6749	3.5820	3.7146
4	4.1430	4.1273	3.5713
5	3.4125	3.2447	3.5153
6	3.7120	4.0332	3.4886
7	3.1549	3.2256	3.7964
8	3.5791	3.9068	3.6698
9	4.0171	3.7396	3.1812
10	3.1560	3.1909	3.9925
11	4.2931	3.6938	3.5839
12	3.7077	4.4886	3.0338
13	4.2205	3.3399	3.7171
14	4.6252	3.7137	4.0824
15	4.1962	3.5054	3.3999
16	3.6051	4.0770	3.7087
17	3.6405	3.8258	3.5915
18	3.3797	3.9733	3.8604
19	3.5710	4.0638	3.6055
20	3.5938	3.4656	3.3789
21	3.3048	3.6741	3.6527
22	4.4158	3.7355	3.8506
23	3.4050	3.5457	3.2789
24	4.4480	3.2999	4.5962
25	3.5878	3.1402	4.1615
Avg Kurtosis	**3.7602**	**3.7141**	**3.7128**

4.3.3.1 Assessing the Effects of Varying Degrees of Kurtosis

To examine the effects of kurtosis, six variants in two categories were subsequently generated from the base case, varying the parameter associated with the fourth moment in each instance. The probability plots shown in Figures 4.9(b), (c), and (d) reflect instances for Category 1 in which increasing values were used for the fourth moment. As is clearly evident, the increased positive excess kurtosis induces an obvious

deviation from the normality established in the base case, resulting in an *S*-shaped curve which becomes visibly more pronounced with each subsequent increase in the fourth standardized moment.

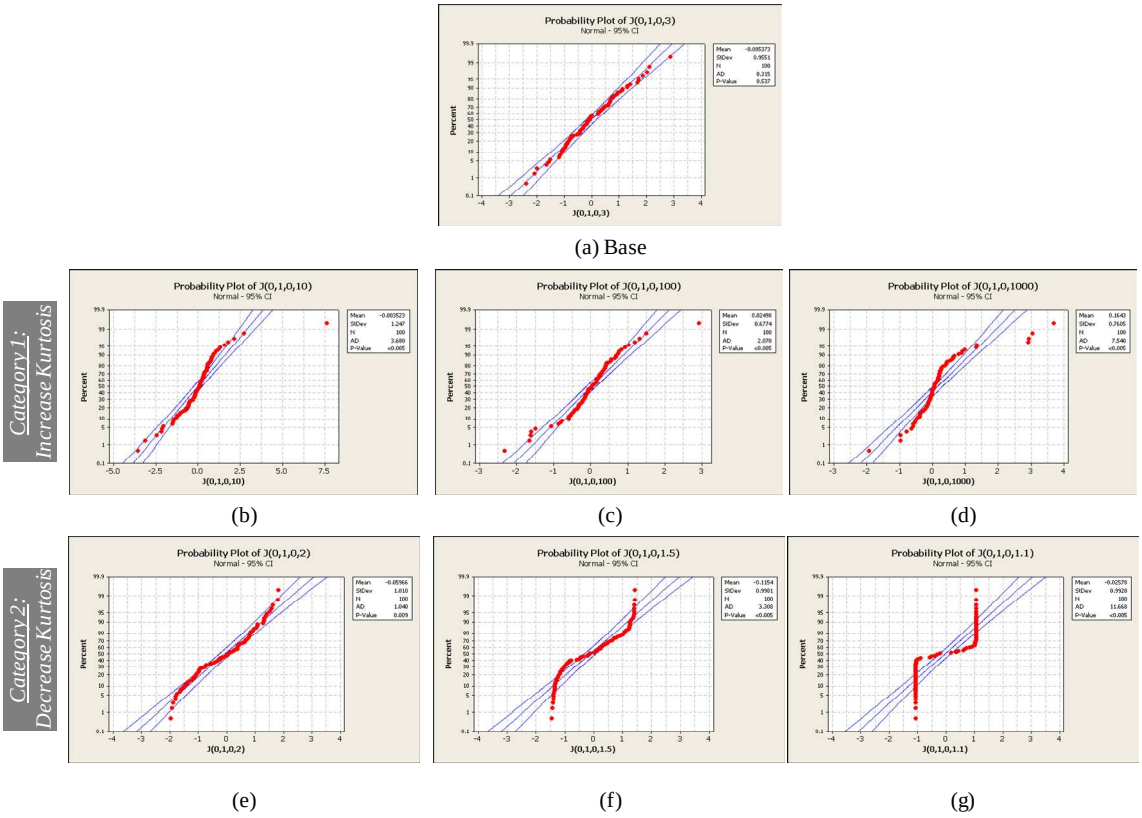


Figure 4.9. Normal probability plots showing the effects of decreased and increased levels of excess kurtosis.

This is interesting and logical, as the shape reflects the manner in which the distribution is peaking about the mean while the tails grow heavier and longer. Consequently, we would expect probabilities to flatten along the tails but to increase very rapidly approaching the mean from the left and then decreasing equally rapidly moving away from the mean to the right. Thus, the sporadic data points that comprise the left and right "tips" of the s-shape indicate the growing influence of outliers or extreme occurrences on

the variability in the data set. Similarly, the angle of the body of the *S*-shaped yields information about the peak in the distribution; the more vertical the line becomes suggests a high, narrow peak about the mean, thereby indicating significant levels of positive excess kurtosis likely exist.

For Category 2, decreasing values were used for the fourth moment to investigate the effects of negative kurtosis on the probability plots. Bearing in mind that this value must exceed 1.0 per the discussion in Section 4.2.2, values of 2, 1.5, and 1.1 were used, respectively. Figures 4.9(e), (f), and (g) depict the probability plots for these three instances. In contrast to the *S*-shaped curve realized in the cases with positive excess kurtosis, the presence of negative excess kurtosis exhibits the opposite effects as the distribution flattens about the mean. That is, we observe a comparatively gradual transition from the mean in either direction that becomes considerably steeper as the distance increases. Furthermore, we note the effect of increasing negative kurtosis on the curves in Figures 4.9(f) and (g), whereby the tails begin to push in toward the mean, becoming nearly vertical in the latter figure. This suggests a trend toward bimodality in the data as the fourth moment approaches 1.0.

4.3.3.2 Effects of Increasing or Decreasing Variability in the Presence of Kurtosis

Generally, we cannot control whether kurtosis exists or not. It is simply a characteristic of the data set that we should consider during analysis. On the other hand, as has been noted several times in Chapters 1-3, efforts to control or eliminate product and process variability are paramount goals for engineers and manufacturers dealing with the robust parameter design problem. With this in mind, an investigation was performed

to determine how decreasing or increasing levels variability in the data would affect the nature of kurtosis in the probability plots, using the examples of extreme positive and negative excess kurtosis from each category (Figures 4.9(f) and (g), respectively) as starting points.

Since typical quality engineering efforts focus on reducing variability, this case was examined first. Figure 4.10(a) shows the probability plots for the extreme case from Category 1 and the corresponding plots for reduced variability and increased variability in Figures 4.10(b) and (c), respectively. The density plots adjacent to each probability plot provide an alternative means to visualize the effects on the data. Figure 4.11 contains comparable plots for Category 2.

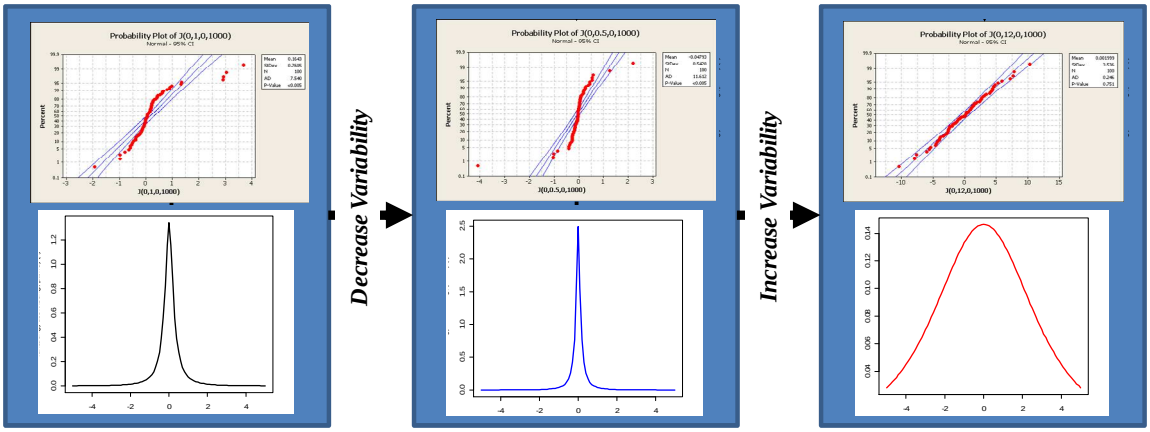
The probability and density plots in Figures 4.10(a) and (b) seem to imply that decreasing variability in the data effectively increases the degree of positive excess kurtosis. This is evidenced by a steepening of the central portion of the plot about the mean and the lengthening of the tails, once again implying the growing influence of extreme observations on the variance. Examination of the density plots confirms this assessment, clearly showing a narrower peak that doubles in height about the mean. However, as DeCarlo (1997) notes, kurtosis represents a "movement of mass in the distribution" that has no bearing on the variance. Likewise, the measure used to describe kurtosis presented in Section 4.2.2 is scaled with respect to the variance and therefore is not affected by it. That is, kurtosis serves to describe the shape of a distribution irrespective of the variance. Notwithstanding, the plots portray a clear relationship.

Thus, the deeper implication appears to be that the kurtosis existed from the start but was masked by the variability.

Given this reaction in the data to decreased variability, it seemed logical that inducing variability in the data would therefore yield the opposite effect. The probability and density plots in Figure 4.10(c) confirm that it does, as the plot forms a nice straight line about the 45 degree index line, thereby implying normality in the data. The *A-D* test statistic of 0.246 and a *p*-value of 0.751 provide a more objective basis to support this implication. However, what is particularly noteworthy here is the significant degree of variability that had to be introduced to compensate for excess positive kurtosis, a twelve-fold increase from 1 to 12. This is very interesting, as it suggests that if we fail to inspect for the presence of excess kurtosis in the initial data analysis, it will not reveal itself until we have induced efforts to reduce variability. Moreover, it implies that visual inspections of plots, absent deeper analysis of all of the distribution's characteristics, are insufficient. Thus, in this particular case, inspection of the plot and the summary statistics accompanying it would have caused us to assume normality in the data. However, once efforts to reduce the significant variability are initiated, it would become apparent that our data may not be normal at all, which could indicate that that methods used to optimize process parameters were inappropriate and thus compromise the results.

Since increasing the variability tended to compensate for positive excess kurtosis, it was surmised that a similar increase in the case of negative excess kurtosis would only exacerbate the effect. Consequently, in dealing with the negative excess kurtosis in category 2, incremental reductions in variability were examined to determine if there is a

point of diminishing returns. In fact, the sequence of plots in Figures 4.10(a)-(c) suggests such a conclusion.



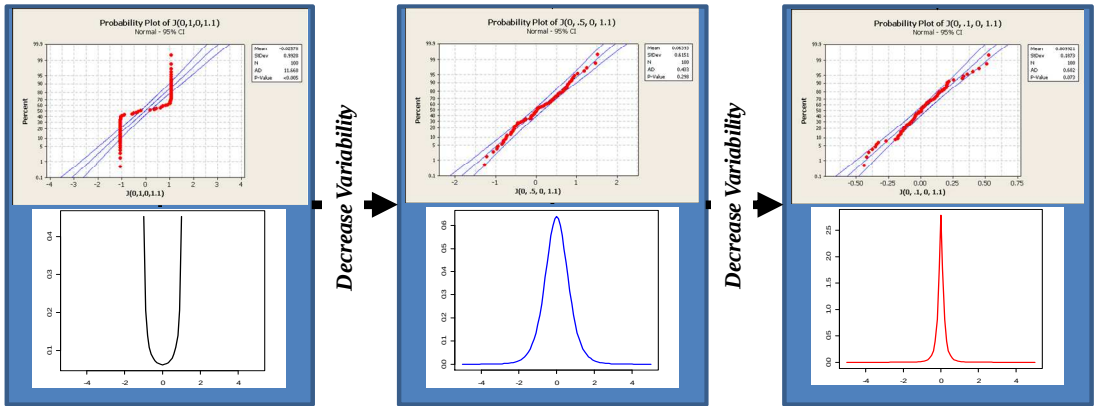
(a) (b) (c)

Figure 4.10. Effects of changing variability on positive excess kurtosis.

The transition from the probability and density plots in Figure 4.11(a) to those in 11(b) clearly shows that a 50% reduction in variability (from 1 to 0.5) mitigates the effects of negative excess kurtosis, resulting in a probability plot that suggests normality. Again, the *A-D* statistic and *p*-value would provide objective support to such a conclusion. It is interesting to note, however, that the appearance of the density plot in Figure 4.11(b) suggests that positive excess kurtosis is now present. A sample kurtosis calculation in R confirms this, yielding a modest value of 1.4. Recall that normally distributed data have zero excess kurtosis.

If the variability is reduced even further, shown in Figure 4.11(c), we readily observe the introduction of positive excess kurtosis through the emergence of the *S*-shaped curve coupled with an elongation of the tails in the probability plot. The *A-D*

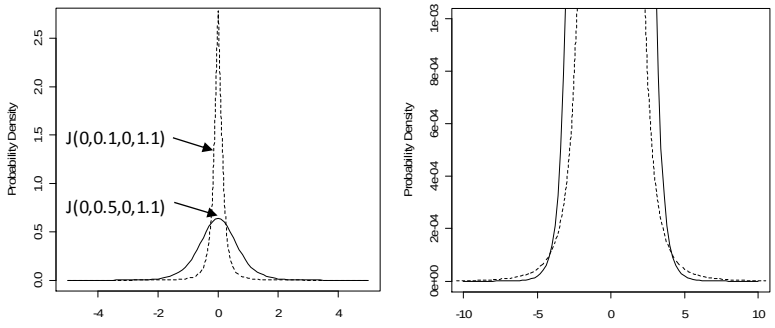
statistic and *p*-value confirm the departure from normality and a calculation for sample kurtosis performed in *R* shows an increase in sample kurtosis to 107.



(a) (b) (c)

Figure 4.11. Effects of changing variability on negative excess kurtosis.

Overlaying the two density plots in Figure 4.12(a) lends further weight to the analysis, showing a peak about the mean that has clearly narrowed in breadth and quadrupled in height with the decreased variability and the increased (albeit slight) heaviness in the tails, which is more clearly evident in Figure 4.12(b). This is not to suggest that we can either induce or remove kurtosis within a data set, but rather that variance reduction measures can reveal its underlying presence in the data.



(a) (b)

Figure 4.12. Overlays of the density plots from Figures 11(b) and (c), revealing increased peakedness (a) and heaviness in the tails (b). Note that scales on both axes have been adjusted to facilitate a "zoomed in" effect in figure (b).

4.3.4 Impacts of Combined Effects of Multiple Moments on the Probability Plot

The final set of experiments sought to ascertain the effects reducing variability on normal probability plots for distributions that possess elevated levels of skewness and kurtosis. As a starting point, 1000 random variates were generated from a $J(10, 8, -5, 100)$ distribution. Thereafter, four additional samples were simulated consisting of 1000 variates each, reducing the variability in each case by 50% from 8 down to 1, and then by 20% from 1 to 0.8. Figure 4.13 displays the resulting probability plots, density plots, and several sample statistics for each of the five cases examined.

Continued variance reductions to 2 and then 1 in Figures 4.13(c) and (d) reinforce the expectations we derived from the analysis in Section 4.3.3.2. That is, once again reductions in variability beyond some threshold actually begin to exacerbate the levels of positive excess kurtosis. In the normal probability plots we see the familiar *S*-shaped curve emerge as the peakedness increases about the mean in each case and the tails grow longer and heavier, particularly to the left. The associated density plots provide a clearer demonstration of these effects, showing both the narrowing of the peak in the distribution coupled with a nearly two-fold increase in its height. The sample statistics for Figures 4.13(c) and (d) certainly show the effects on excess kurtosis, but they also show a clear impact on the asymmetry, as the negative skewness increases by a factor of three in each case. This is further evidenced by the continued, albeit slight leftward shift in the mean.

Reducing the variance to 0.8 yields interesting results as shown in Figure 4.13(e). Here, the normal probability plot clearly evinces the concave shape associated with negative skew discussed in Section 4.3.2.1. Furthermore, we observe the elongation of

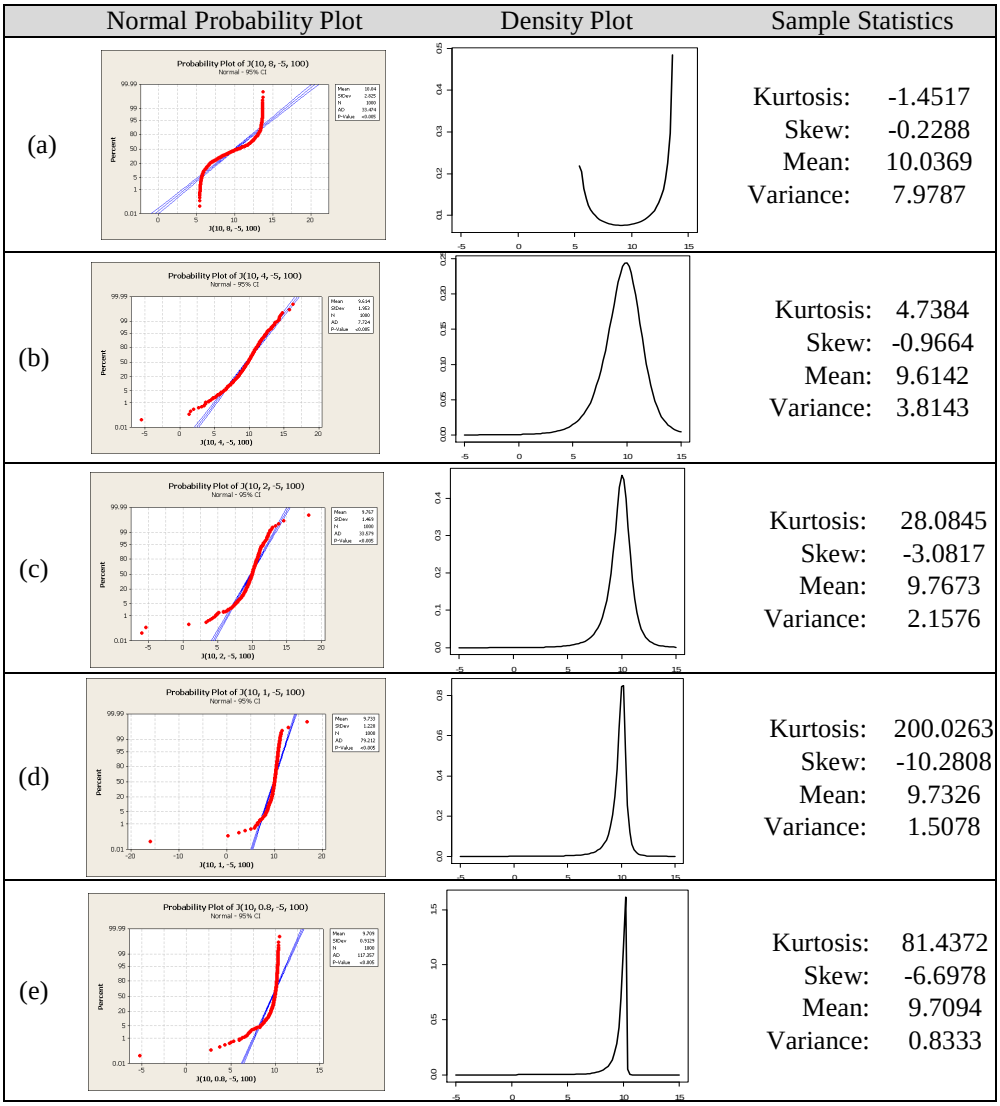


Figure 4.13. Normal probability and density plots portraying the effects of decreasing variability in the presence of non-zero skewness and kurtosis.

the left tail coupled with the shortening of the right tail. The shape of the density plot in 4.13(e) conveys similar information and the sample statistics for skewness and the mean reinforce this, showing a further shift to the left by the mean as the bulk of the distribution piles up to the right. However, the statistics also show that considerable excess positive kurtosis remains, despite the lack of visual evidence in the normal

probability plot. Essentially, the reduced variability has caused the distribution to peak to such a degree about the mean that, in the presence of negative skew, the right tail becomes nearly non-existent.

4.4 Summary of Findings

Normal probability plots play a very important role in the analysis of data. While still a critical part of any good residual analysis of regression models, the investigations and analysis we have heretofore described reinforce their equally critical role in facilitating data analysis in general. A basic understanding of how the characteristics of the data in terms of the four sample moments influence the properties of the probability plot can yield significant information about those data that will, in turn influence certain approaches to model estimation and optimization. We summarize the findings from these investigations in (i)-(iii) below:

(i) In the most basic sense, a detailed visual inspection of the curve in a probability plot will reveal information about the presence of asymmetry and kurtosis along and suggest the degree to which they influence the data and how. Concavity versus convexity will typically indicate the presence of either positive or negative asymmetry or skewness, respectively. Moreover, the nature of the tails provides a hint of how variability affects the data. That is, whereas the nature of the curve can impart knowledge as to skewness about the mean, the nature of the tails can impart knowledge about the degree of variability present in the data. The plots in Figures 4.7(c), (e), and (g) serve as a good example of this, showing a clear elongation of the right tail in the face of increasing variability. However, the nature of the concave or convex curve may also

reveal the presence of kurtosis in the data, suggesting a combined effect of distribution characteristics at work within the data. The probability plot in Figure 4.13(e) provides a stark example of this, depicting the convex curve illustrative of negative asymmetry about the mean, but also a significant degree of positive excess kurtosis suggested by the dense, near-vertical curve in the immediate proximity of the mean.

Similarly, an *S*-shaped curve either with a denser, largely vertical base and elongated horizontal tails or a denser horizontal base with elongated vertical tails provides a very good indication of positive or negative excess kurtosis within the data. This is important on a number of levels. First, density and degree to which the main portion of the curve about the mean are vertical or horizontal coupled the heaviness of the tails yield important information. In one sense, they very quickly show us the likelihood of achieving values close to the mean. In another, particularly in the case of the tails, they provide an idea as to the increased or decreased propensity for outliers to occur and the subsequent degree to which they influence variability in the process. This particular point is noteworthy insofar as positive kurtosis can arise either due to the presence of outliers within data that are otherwise normally distributed, or because the underlying distribution is simply not normal. In the latter case, then, the curve would suggest consideration of heavy tailed non-normal distributions for modeling. Second, as several researchers have noted, including Browne (1984), the presence of kurtosis can have large effects on tests for significance and standard errors associated with parameter estimates. Thus, a probability plot that suggests potential issues with the presence of kurtosis would

likewise suggest the need to consider alternative tests and estimators that are more robust to detecting departures from normality.

(ii) Virtually all process optimization schemes strive for a reduction in variability to improve process performance. The analysis in Sections 4.3.2 and 4.3.3 demonstrates that efforts to either reduce or, in some potentially unique situations, induce variability within a process can affect what we thought or assumed to be true about the data. In the case of skewness, as the analysis in Section 4.3.2.2 shows, a reduction of variance within asymmetric data produces precisely the effect we would expect in terms of an increased departure from normality as the bulk of the distribution begins to stack up and peak about the mean. What is particularly interesting here, however is the plot result shown in Figure 4.7(f), which reinforces the importance of understanding what the data are conveying via the curve. Consider a situation involving a smaller-the-better type quality characteristic. In this case, we would expect a degree of positive asymmetry and so would expect to observe a concave plot similar to the plot in Figure 4.7(b). However, the near-vertical line and mirror-image tails in Figure 4.7(f) imply symmetry rather than asymmetry, which might lead to suspicion about the data, the manner in which they were collected, or about the process itself. But what the analysis shows is that asymmetry is present but is being masked by low levels of variability.

Kutner *et al.* (2005) pointed out that, in the case of analysis of variance, kurtosis generally bears more importance than skewness. This is attributed to general findings that whereas skew has a more profound impact on tests for means, kurtosis has greater effect on tests for variance and covariance. Instances of positive and negative kurtosis

are often considered indications of large or small variance respectively. However, as previously noted, kurtosis neither affects nor is affected by the variance. Nevertheless, as the analysis in Section 4.3.3 suggests, increasing or decreasing variability can reveal or mask its presence. This is important to note, since the concepts associated with kurtosis such as tail weight and peakedness play a critical role in descriptive and inferential statistics. As noted in (i), kurtosis can induce considerable effects on significance tests and standard errors of parameter estimates. Thus, using normal probability plots to decipher the presence and degree of excess kurtosis can also help to indicate whether we need to consider a number of alternative (and possibly more robust) tests and estimators for modeling.

4.5 Conclusion

The knowledge and use of normal probability plots dates back many decades and for a period of time received considerable attention in the literature as a viable and particularly useful tool in data analysis and the substantiation of statistical assumptions. However, little has been mentioned about their role in data analysis beyond the "fat pencil" test used in residual analysis. That is, the use of these plots as a critical analytical tool outside of residual analysis seems to be largely overlooked. This is not to say that researchers or analysts ignore them, but rather because of the subjective and simplistic nature of their application, there is considerable potential for many to take what they offer for granted.

The work set forth in this chapter, the first of its kind in terms of the breadth and scope of the analysis performed, has both reaffirmed and expanded the usefulness of an

old, reliable, and particularly powerful tool for statistical inference and analysis. The contribution is a comprehensive demonstration and analysis of how various statistical properties within a data set can influence the shape and corresponding properties in a normal probability plot. This should, in turn, provide analysts with a better understanding of the ways in which data communicate through the plot. Thus, using this information, analysts can observe a normal probability plot and derive a better and fuller understanding of the underlying properties in the data that are associated with the skewness, the presence and degree of excess kurtosis, the combined effects of both, and the extent to which variability may be influencing the plot. In turn, this will inform them as to other distributions to consider, as well as possible transformations if normality is required.

The criticality of this analysis is that our initial assumptions concerning the actual experimental data will ultimately drive model estimation and any optimization results thereafter. Consequently, it is of particular importance that we understand the data as completely as possible in order to identify the most precise means for estimating parameters that will, in turn, facilitate better estimation of our fitted models. Failure to examine the data could lead to inappropriate assumptions and faulty distribution fitting, which would impact both tier-one and tier-two estimates, and ultimately the optimization results themselves.

CHAPTER FIVE

ANALYZING THE EFFECTS OF VARIABILITY MEASURE SELECTION ON PROCESS AND PRODUCT OPTIMIZATION

5.1 Introductory Remarks

The numerous efforts examining the dual response surface approach or a branch of robust parameter design use one of three variability measures to evaluate process variability – the standard deviation, variance, or logarithm of the standard deviation. In most cases, upon developing a response surface design for their variability measure, the response model is then normalized to the variance. For instance, as indicated in Chapter 1, a review of the literature indicates that many researchers chose to model the standard deviation and square it before applying an optimization scheme. In the last decade, however, a number of researchers have either directly modeled the variance in the experimental framework or have selected an alternative measure such as the logarithm of the standard deviation. They then perform comparison studies with previous work, in order to identify improvements within the research field. Table 5.1 provides a list of several researchers and the measure(s) of variability they used in the context of their work.

When seeking the optimal factor settings using the dual response surface approach, a variety of solutions may result depending on the variability measure used. In addition, a statistical evaluation of each measure's fitted function may indicate different degrees of precision in their respective estimation of the response. As a result, it is

difficult to ascertain the accuracy among comparison studies using the dual response surface approach. Contemporary robust design research has yet to fully address this problem. Thus, the goal of this paper is to identify how and why one variability measure may perform better than another under a given set of conditions, and to suggest a uniform approach to the selection of a variability measure. The original work for this research is published with reference Goethals and Boylan (2011). Following a description of the methodology used to investigate this problem, an analysis of the results is provided and the findings summarized.

Table 5. 1. Variability Measure Selection in contemporary quality-based research.

		Variability Measure		
Researcher(s)	Year	$\hat{\sigma}$	$\widehat{\sigma^2}$	$\widehat{\ln \sigma}$
Vining and Myers	1990	X		
Del Castillo and Montgomery	1993	X		
Cho	1994		X	
Lin and Tu	1995	X		
Copeland and Nelson	1996	X		
Kim and Lin	1998	X		
Kim and Cho	2002		X	
Kim and Rhee	2003	X		
Koksoy and Doganaksoy	2003	X		
Kulkarni and Mariappan	2003		X	X
Cho and Park	2005		X	
Chen *et al.*	2006			X
Anderson and Whitcomb	2007	X		X
Pickle *et al.*	2008		X	X
Giovagnoli and Romano	2008		X	
Kwon *et al.*	2008	X		
Johnson *et al.*	2009		X	
Kovach *et al.*	2009		X	
Costa	2010	X		

5.2 Methodology

5.2.1 Establishing an Experimental Framework

Upon identifying a quality characteristic of interest Y that is influenced by a set of control factors $\mathbf{x} = (X_1, X_2, ..., X_v)$ for a given process, suppose we conduct an experiment

to identify the optimal factor settings $\mathbf{x}^* = (X_1^*, X_2^*, ..., X_v^*)$. The experimental framework consists of n runs with m replications for these factors. Let y_{jk} be the kth response at the jth design point, where $j = 1, 2, ..., n$ and $k = 1, 2, ..., m$. The mean, standard deviation, variance, and logarithm of the standard deviation at the jth design point are found using the formulas at Equation 5.1. Table 5.2 outlines the format for such an experiment.

$$\bar{y}_j = \frac{\sum_{k=1}^{m} y_{jk}}{m} \qquad s_j = \sqrt{\frac{\sum_{k=1}^{m} (y_{jk} - \bar{y}_j)^2}{m-1}} \qquad s_j^2 = \frac{\sum_{k=1}^{m} (y_{jk} - \bar{y}_j)^2}{m-1} \qquad \ln s_j = \ln \sqrt{\frac{\sum_{k=1}^{m} (y_{jk} - \bar{y}_j)^2}{m-1}} \tag{5.1}$$

Using least squares regression, we develop response surface designs for the mean and each variability measure. In particular, the general form of the response surface design for the mean including the intercept with p parameters or $p-1$ predictor variables may be written as:

$$\hat{\mu}(\mathbf{x}) = \mathbf{X}\hat{\beta}_\mu, \text{ where } \hat{\beta}_\mu = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\bar{\mathbf{y}}, \mathbf{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1,p-1} \\ 1 & X_{21} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{n,p-1} \end{bmatrix}, \text{ and } \bar{\mathbf{y}} = [\bar{y}_1, \bar{y}_2, ..., \bar{y}_n]^{\mathrm{T}} \tag{5.2}$$

Table 5.2. Experimental Response Surface Methodology Format

Run	X_1 X_2 …… X_v	*Replications* (Y)	$\bar{y}$	s	s^2	$\ln s$
1	Control Factor Settings	y_{11} y_{12} … y_{1k} … y_{1m}	$\bar{y}_1$	s_1	s_1^2	$\ln s_1$
2		y_{21} y_{22} … y_{2k} … y_{2m}	$\bar{y}_2$	s_2	s_2^2	$\ln s_2$
⋮		⋮ ⋮ ⋮ ⋮	⋮	⋮	⋮	⋮
j		y_{j1} y_{j2} … y_{jk} … y_{jm}	$\bar{y}_j$	s_j	s_j^2	$\ln s_j$
⋮		⋮ ⋮ ⋮ ⋮	⋮	⋮	⋮	⋮
n		y_{n1} y_{n2} … y_{nk}… y_{nm}	$\bar{y}_n$	s_n	s_n^2	$\ln s_n$

We develop response surface designs for each variability measure in a similar fashion:

$$\hat{\sigma}(\mathbf{x}) = \mathbf{X}\hat{\beta}_{\sigma}, \text{ where } \hat{\beta}_{\sigma} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{s}, \text{ and } \mathbf{s} = [s_1, s_2, ..., s_n]^{\mathrm{T}} \tag{5.3}$$

$$\widehat{\sigma^2}(\mathbf{x}) = \mathbf{X}\hat{\beta}_{\sigma^2}, \text{ where } \hat{\beta}_{\sigma^2} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{s^2}, \text{ and } \mathbf{s^2} = [s_1^2, s_2^2, ..., s_n^2]^{\mathrm{T}}$$

$$\widehat{\ln\sigma}(\mathbf{x}) = \mathbf{X}\hat{\beta}_{\ln\sigma}, \text{ where } \hat{\beta}_{\ln\sigma} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\ln\mathbf{s}, \text{ and } \ln\mathbf{s} = [\ln s_1, \ln s_2, ..., \ln s_n]^{\mathrm{T}}$$

To ensure a fair and valid comparison between each variability measure, we use the same full order array of predictor variables to model each measure. We then consider a variety of criteria to evaluate the fit of each response surface design. Since the observations on the variance response are naturally at a higher scale than either the standard deviation or logarithm of the standard deviation, the error sum of squares (*SSE*) for the variance response will naturally be higher. Thus, evaluation criteria such as the prediction error sum of squares (*PRESS*) and Akaike's Information Criterion (*AIC*) that are primarily based upon the value of *SSE* do not serve as useful tools for comparison. The coefficient of determination R^2 for a given model, however, involves a ratio of the regression sum of squares (*SSR*) or *SSE* to the total sum of squares (*SSTO*) and therefore may provide some value in the comparison of designs.

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO} \tag{5.4}$$

5.2.2 Investigating the Variability Measures

In order to identify when and under what conditions one variability measure may outperform another measure, we conduct several experiments. Since the amount of variability in the data or the response surface design itself may affect the performance of a specific measure, examining the results while adjusting these conditions may be useful. In addition, constraining the mean to its target and subsequently optimizing the

variability measure may produce different results than when we optimize the mean and variance simultaneously. Thus, alternative optimization methods should be observed to differentiate among the schemes previously used by researchers. Finally, it is not completely clear what effect the fit of each surface is having on the performance of the variability measure. To consider each of these factors, four experiments are conducted, labeled Parts A, B, C, and D, respectively.

Part A and B: The method used by Vining and Myers (Equation 5.5) is used to compare and contrast solutions. This involves using an iterative Nelder-Mead direct search method within the experimental region of interest to identify a global maximum. In particular, for an *N*-type characteristic, where $\hat{\theta}(\mathbf{x})$ represents the normalized variability measure of interest, the optimization scheme is:

$$\begin{aligned} &\text{Minimize: } \hat{\theta}(\mathbf{x}) \\ &\text{Subject to: } \hat{\mu}(\mathbf{x}) = \tau_N \\ &\qquad \mathbf{x}^{\mathrm{T}}\mathbf{x} \le \rho^2 \text{ (central composite design), or} \\ &\qquad |X_i| \le 1, \text{ for } i = 1, 2, ..., v \text{ (factorial design),} \\ &\qquad \text{where } \hat{\theta}(\mathbf{x}) = \hat{\sigma}^2(\mathbf{x}),\ \widehat{\sigma^2}(\mathbf{x}),\ \text{or } \{\exp[\ln\widehat{\sigma}(\mathbf{x})]\}^2 \end{aligned} \tag{5.5}$$

For Part A, the data are slightly adjusted to support testing the measures against different variability conditions, specifically when $\bar{s} < 1$, $\bar{s} \approx 1$, $\bar{s} \approx 5$, $\bar{s} \approx 10$, $\bar{s} \approx 30$, and where the mean coefficient of variation $\overline{CV} \approx 1$ and $\overline{CV} > 1$. For Part B, the variability in each response surface is adjusted to identify the effect on the measure of interest, specifically when Var(*s*) ≈ 0.05, Var(*s*) ≈ 0.50, Var(*s*) ≈ 2, Var(*s*) ≈ 5, Var(*s*) ≈ 60, and Var(*s*) ≈ 120.

Part C and D: For these experiments, the objective is to simultaneously optimize the process mean and variance. Xu *et al*. (2004) developed a modified desirability

function approach using a goal attainment technique to identify the optimal factor settings when considering several characteristics. The method was specifically designed to provide greater weight to those fitted response surfaces with a higher R^2. Although designed for multi-response problems, one can easily adapt the approach to consider the mean and the variability measure as two separate characteristics. The procedure consists of three distinct steps, outlined in (i)-(iii):

(i) Individual desirability functions are developed for the mean and the normalized variability measure, $d[\hat{\mu}(\mathbf{x})]$ and $d[\hat{\theta}(\mathbf{x})]$, respectively. In particular, considering the mean of an *N*-type characteristic and treating the variability measure as an *S*-type characteristic, we have:

$$d[\hat{\mu}(\mathbf{x})] = \begin{cases} 0 & \text{if } \hat{\mu}(\mathbf{x}) < LSL \text{ or } \hat{\mu}(\mathbf{x}) > USL \\ \left(\dfrac{\hat{\mu}(\mathbf{x}) - LSL}{\tau_N - LSL} \right)^{\gamma_N} & \text{if } LSL \le \hat{\mu}(\mathbf{x}) \le \tau_N \\ \left(\dfrac{\hat{\mu}(\mathbf{x}) - USL}{\tau_N - USL} \right)^{\gamma_N} & \text{if } \tau_N \le \hat{\mu}(\mathbf{x}) \le USL \end{cases} \tag{5.6}$$

$$d[\hat{\theta}(\mathbf{x})] = \begin{cases} 0 & \text{if } \hat{\theta}(\mathbf{x}) > USL \\ \left(\dfrac{\hat{\theta}(\mathbf{x}) - USL}{\tau_S - USL} \right)^{\gamma_S} & \text{if } \tau_S \le \hat{\theta}(\mathbf{x}) \le USL \\ 1 & \text{if } \hat{\theta}(\mathbf{x}) < \tau_S \end{cases}$$

where γ_N and γ_S represent the shape parameters for the individual desirability functions, chosen based upon the degree of importance in obtaining the corresponding target τ_N or

τ_S. The range of each individual desirability function is [0, 1], where values close to or equal to 1 are considered ideal and values at or near 0 are considered undesirable.

(ii) Weights are established for each of the individual desirability functions, denoted *w*, based upon the established fit of the response surface designs. Here, the weighting system proposed by Xu *et al*. (2004) is invoked:

$$w_{\mu,\theta} = \frac{1}{R^2}, \text{ where } 0 \le R^2 \le 1 \tag{5.7}$$

(iii) Given the weights for the mean and variability measure of interest, the objective is then to determine the factor settings $\mathbf{x}^* = (X_1^*, X_2^*, ..., X_v^*)$ that maximize the weighted geometric mean of the individual desirability functions, referred to as the composite desirability function *D*:

$$D = \left\{ d[\hat{\mu}(\mathbf{x})]^{w_\mu} \cdot d[\hat{\theta}(\mathbf{x})]^{w_\theta} \right\}^{1/(w_\mu + w_\theta)} \tag{5.8}$$

As with the individual desirability functions, a composite desirability function with a value close to or equal to 1 is considered ideal, whereas a value at or near zero is completely undesirable.

Similar to the experiment in Part A, Part C examines the performance of the variability measures against different variability conditions, specifically when $\bar{s} < 1$, $\bar{s} \approx 1$, $\bar{s} \approx 5$, $\bar{s} \approx 10$, $\bar{s} \approx 30$, and where the mean coefficient of variation $\overline{CV} \approx 1$ and $\overline{CV} > 1$. And as in Part B, Part D involves adjusting the variability within each response surface, namely when Var(*s*) ≈ 0.05, Var(*s*) ≈ 0.50, Var(*s*) ≈ 2, Var(*s*) ≈ 5, Var(*s*) ≈ 60, and Var(*s*) ≈ 120.

Finally, using the optimal factor settings $\mathbf{x}^*$ identified for each experimental condition, we may calculate the mean square error (*MSE*) using Equation (5.9). This

criterion will serve as the primary tool for comparing the performance among the different variability measures.

$$MSE = [\hat{\mu}(\mathbf{x}^*) - \tau]^2 + \hat{\theta}(\mathbf{x}^*), \text{ where } \hat{\theta}(\mathbf{x}^*) = \hat{\sigma}^2(\mathbf{x}^*), \widehat{\sigma}^2(\mathbf{x}^*), \text{ or } \{\exp[\widehat{\ln}\sigma(\mathbf{x}^*)]\}^2 \quad (5.9)$$

5.3 Numerical Example

5.3.1 Examining a Semiconductor Manufacturing Process

Consider the experiment shown in Table 5.3 where the effect of various factor settings on the coating thickness (Y) of silicon wafers, measured in micrometers, is investigated for a semiconductor manufacturing process. A central composite design is used, consisting of 13 runs among two control factors – the temperature of the molding stage (X_1) measured in degrees Fahrenheit, and the injection flow rate (X_2) measured in pounds per second. As an *N*-type characteristic, a desired target value of $\tau = 77.5$ μm is established for the coating thickness in this experiment, with a lower and upper specification limit of 73.5 μm and 81.5 μm, respectively. Four replications of the experiment are performed; Table 5.3 depicts the calculations for the mean and each variability measure for the case when $\bar{s} < 1$ ($\bar{s} = 0.686$). We then develop a full second-order response surface design for the mean and each variability measure, as noted below:

$$\hat{\mu}(\mathbf{x}) = 75.5701 - 0.6509X_1 - 0.6120X_2 - 0.0125X_1X_2 - 0.3867X_1^2 + 0.2822X_2^2$$

$$\hat{\sigma}(\mathbf{x}) = 0.5540 + 0.0476X_1 - 0.0069X_2 - 0.0556X_1X_2 + 0.1021X_1^2 + 0.1122X_2^2$$

$$\widehat{\sigma}^2(\mathbf{x}) = 0.3396 + 0.0650X_1 - 0.0024X_2 - 0.0958X_1X_2 + 0.1291X_1^2 + 0.1402X_2^2$$

$$\widehat{\ln}\sigma(\mathbf{x}) = -0.6538 + 0.0705X_1 - 0.0164X_2 - 0.0649X_1X_2 + 0.1776X_1^2 + 0.1960X_2^2$$

Table 5.3. Semiconductor Manufacturing Process – Experimental Framework ($\bar{s}$ <1)

	Coded Units									
	Temp	Flow Rate	Coating Thickness (μm)							
Run	X_1	X_2	Y (4 replications)				$\bar{y}$	s	s^2	ln s
1	–1	–1	76.3	78.0	76.5	77.0	76.950	0.759	0.577	–0.275
2	1	–1	74.0	75.5	73.4	74.8	74.425	0.918	0.843	–0.086
3	–1	1	75.2	76.1	77.0	74.9	75.800	0.949	0.900	–0.053
4	1	1	72.0	73.5	74.1	73.3	73.225	0.885	0.782	–0.123
5	–1.41	0	74.9	75.1	76.0	74.8	75.200	0.548	0.300	–0.602
6	1.41	0	75.6	75.9	74.3	74.7	75.125	0.750	0.563	–0.288
7	0	–1.41	78.5	77.0	77.2	76.9	77.400	0.744	0.553	–0.296
8	0	1.41	76.0	75.9	76.1	76.7	75.600	0.594	0.353	–0.520
9	0	0	76.1	77.0	75.9	76.7	76.425	0.512	0.263	–0.669
10	0	0	75.1	75.3	76.1	74.0	74.800	0.572	0.327	–0.559
11	0	0	75.5	76.0	75.8	76.1	75.850	0.265	0.070	–1.330
12	0	0	74.0	75.2	75.1	74.3	74.650	0.592	0.350	–0.525
13	0	0	76.2	75.0	77.0	76.3	76.125	0.830	0.689	–0.186

For Part A and B, we use the Vining and Myers optimization scheme to identify the optimal factor settings for the process. In Part A, the conditions as shown in Table 5.3 where $\bar{s}$ <1 are considered, and in Part B, the conditions are adjusted slightly such that Var(s) ≈ 0.05. Shown in Tables 5.4 and 5.5 are the results for Part A and B, respectively, examining each of these specific cases when the normalized variability measure is used. The most favorable results in terms of *MSE* are highlighted.

For Part C and D, we observe the effects of performing a simultaneous optimization of the mean and variance using a modified desirability function approach. The same conditions used in Parts A and B are applied to Parts C and D, respectively; Tables 5.6 and 5.7 portray the results looking specifically at these specific cases involving the normalized variability measure. The shape parameter is established at $\gamma = 1$ for each of the individual desirability functions.

Table 5.4. Results for Part A

Condition	*Measure*	$\mathbf{x}^*$	$\hat{\mu}(\mathbf{x}^*)$	$\hat{\theta}(\mathbf{x}^*)$	*MSE*
$\bar{s} < 1$	σ	(–0.358864, –1.42299)	77.19	0.576	0.672
	σ^2	(–0.358849, –1.42300)	77.19	0.571	0.667
	ln σ	(–0.358899, –1.42299)	77.19	0.584	0.679
$\bar{s} \approx 1$	σ	(0.639456, 0.157584)	77.50	1.023	1.023
	σ^2	(0.639724, 0.153632)	77.50	1.134	1.134
	ln σ	(0.639589, 0.155577)	77.50	0.822	0.822
$\bar{s} \approx 5$	σ	(0.590287, 0.733351)	77.50	28.88	28.88
	σ^2	(0.579087, 0.695808)	77.50	29.04	29.04
	ln σ	(0.603747, 0.776081)	77.50	28.72	28.72
$\bar{s} \approx 10$	σ	(–0.707010, –1.22480)	77.50	104.14	104.14
	σ^2	(–0.707010, –1.22480)	77.50	104.08	104.08
	ln σ	(–0.707011, –1.22480)	77.50	104.18	104.18
$\bar{s} \approx 30$	σ	(–0.828696, 1.14598)	77.50	872.76	872.76
	σ^2	(–0.828695, 1.14598)	77.50	867.84	867.84
	ln σ	(–0.828696, 1.14598)	77.50	876.98	876.98
$\overline{CV} \approx 1$	σ	(–0.714443, 0.090337)	580.00	3.35e+05	3.35e+05
	σ^2	(1.41287, 0.061710)	580.00	3.42e+05	3.42e+05
	ln σ	(0.140391, 0.464538)	580.00	3.22e+05	3.22e+05
$\overline{CV} > 1$	σ	(–0.111354, –0.099333)	580.00	4.98e+05	4.98e+05
	σ^2	(–0.136141, –0.201816)	580.00	5.08e+05	5.08e+05
	ln σ	(–0.085403, –0.040233)	580.00	4.88e+05	4.88e+05

Table 5.5. Results for Part B

Condition	*Measure*	$\mathbf{x}^*$	$\hat{\mu}(\mathbf{x}^*)$	$\hat{\theta}(\mathbf{x}^*)$	*MSE*
Var(s) ≈ 0.05	σ	(–0.322639, –1.39602)	77.41	0.895	0.903
	σ^2	(–0.322949, –1.39592)	77.41	0.915	0.922
	ln σ	(–0.322926, –1.39594)	77.41	0.877	0.884
Var(s) ≈ 0.5	σ	(–0.650584, 0.060606)	77.50	0.893	0.893
	σ^2	(–0.829865, 0.243073)	77.50	0.955	0.955
	ln σ	(–0.557127, –0.057275)	77.50	0.725	0.725
Var(s) ≈ 2	σ	(–0.408740, –1.35386)	77.50	11.73	11.73
	σ^2	(–0.408740, –1.35386)	77.50	12.65	12.65
	ln σ	(–0.408740, –1.35386)	77.50	11.42	11.42
Var(s) ≈ 5	σ	(–0.234457, 0.428894)	77.50	14.69	14.69
	σ^2	(–0.020587, 0.405397)	77.50	20.36	20.36
	ln σ	(–0.483806, 0.486030)	77.50	7.63	7.63
Var(s) ≈ 60	σ	(1.12756, 0.853585)	77.50	160.93	160.93
	σ^2	(1.12756, 0.853585)	77.50	145.37	145.37
	ln σ	(1.12756, 0.853585)	77.50	161.14	161.14
Var(s) ≈ 120	σ	(–0.801505, –1.16516)	77.50	42.05	42.05
	σ^2	(–0.767158, –1.13274)	77.50	0.011	0.011
	ln σ	(0.541847, 1.30629)	77.50	64.33	64.33

The weighting of the composite desirability function *D* is based upon the coefficient of determination, as described in Section 2. While the response surface design for the mean naturally has the same fit for all three measures, the values of R^2 for the variability measures differ as shown in the table. For this experiment, the quality of the fit for the response surface corresponding to a particular variability measure is of little concern; the comparison of the fit among response surfaces is of greater concern, as they are all designed using the same terms. The most favorable results in terms of the *MSE* and composite desirability (*D*) are highlighted.

Table 5.6. Results for Part C.

Condition	*Measure*	$\mathbf{x}^*$	$\hat{\mu}(\mathbf{x}^*)$	$\hat{\theta}(\mathbf{x}^*)$	*MSE*	R^2	*D*
$\bar{s} < 1$	σ	(–0.631033, –1.00624)	76.72	0.422	1.030	38.1%	0.745
	σ^2	(–0.730602, –1.14028)	76.89	0.466	0.838	38.7%	0.743
	ln σ	(–0.545607, –0.896334)	76.58	0.368	1.214	34.8%	0.754
$\bar{s} \approx 1$	σ	(0.639450, 0.157666)	77.50	1.023	1.023	63.3%	0.893
	σ^2	(0.639724, 0.153633)	77.50	1.134	1.134	75.5%	0.892
	ln σ	(0.639583, 0.155669)	77.50	0.822	0.822	44.1%	0.897
$\bar{s} \approx 5$	σ	(0.590353, 0.733567)	77.50	28.88	28.88	49.9%	0.426
	σ^2	(0.579091, 0.695822)	77.50	29.04	29.04	50.4%	0.424
	ln σ	(0.603748, 0.776092)	77.50	28.72	28.72	49.3%	0.431
$\bar{s} \approx 10$	σ	(–0.707010, –1.22480)	77.50	104.14	104.14	54.5%	0.414
	σ^2	(–0.707010, –1.22480)	77.50	104.08	104.08	55.4%	0.418
	ln σ	(–0.707010, –1.22480)	77.50	104.18	104.18	53.6%	0.410
$\bar{s} \approx 30$	σ	(–0.828695, 1.14598)	77.50	872.76	872.76	51.3%	0.280
	σ^2	(–0.828695, 1.14598)	77.50	867.84	867.84	51.5%	0.289
	ln σ	(–0.828695, 1.14598)	77.50	876.98	876.98	51.0%	0.272
$\overline{CV} \approx 1$	σ	(1.27569, –0.610416)	509.27	2.60e+05	2.65e+05	47.3%	0.598
	σ^2	(0.916154, –1.07734)	497.06	2.38e+05	2.45e+05	50.0%	0.612
	ln σ	(0.298050, –0.189638)	528.44	2.74e+05	2.77e+05	43.9%	0.596
$\overline{CV} > 1$	σ	(–0.111354, –0.099331)	580.00	4.98e+05	4.98e+05	51.9%	0.716
	σ^2	(–0.136142, –0.201827)	580.00	5.08e+05	5.08e+05	51.8%	0.709
	ln σ	(–0.085424, –0.040272)	580.00	4.88e+05	4.88e+05	51.3%	0.722

Table 5.7. Results for Part D.

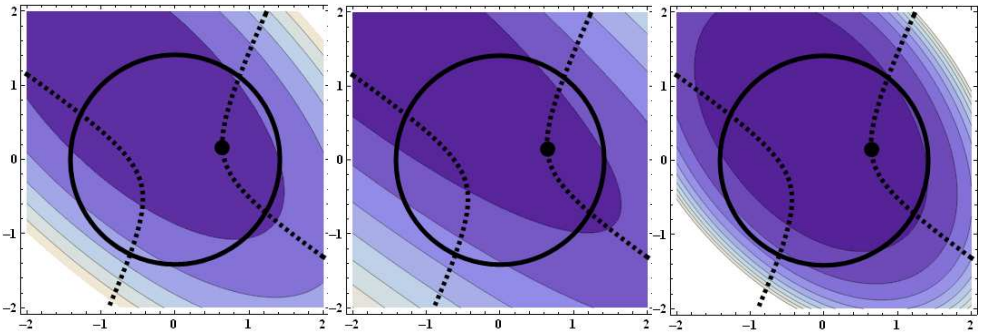
Condition	*Measure*	**x***	$\hat{\mu}(\mathbf{x}^*)$	$\hat{\theta}(\mathbf{x}^*)$	*MSE*	R^2	*D*
Var(*s*) ≈ 0.05	σ	(–0.633500, –0.670144)	76.43	0.425	1.553	48.6%	0.718
	σ^2	(–0.695300, –0.599372)	76.36	0.450	1.744	51.2%	0.701
	ln σ	(–0.557368, –0.707802)	76.47	0.380	1.439	42.5%	0.739
Var(*s*) ≈ 0.5	σ	(–0.649855, 0.059772)	77.50	0.893	0.893	84.4%	0.919
	σ^2	(–0.828640, 0.241932)	77.50	0.955	0.955	92.2%	0.918
	ln σ	(–0.555983, –0.058904)	77.50	0.725	0.725	62.0%	0.922
Var(*s*) ≈ 2	σ	(–0.408740, –1.35386)	77.50	11.73	11.73	35.8%	0.794
	σ^2	(–0.408751, –1.35389)	77.50	12.65	12.65	33.3%	0.771
	ln σ	(–0.408740, –1.35386)	77.50	11.42	11.42	38.6%	0.806
Var(*s*) ≈ 5	σ	(–0.219517, –1.02140)	77.50	11.84	11.84	13.3%	0.728
	σ^2	(–0.410969, –1.32174)	77.50	13.71	13.71	15.0%	0.689
	ln σ	(–0.483759, 0.486018)	77.50	7.63	7.63	15.0%	0.831
Var(*s*) ≈ 60	σ	(1.12756, 0.853585)	77.50	160.93	160.93	61.7%	0.956
	σ^2	(1.12756, 0.853585)	77.50	145.37	145.37	64.6%	0.962
	ln σ	(1.12756, 0.853585)	77.50	161.14	161.14	57.3%	0.954
Var(*s*) ≈ 120	σ	(–0.801510, –1.16516)	77.50	42.05	42.05	64.2%	0.982
	σ^2	(–0.767175, –1.13276)	77.50	0.002	0.002	69.5%	0.999
	ln σ	(–0.801504, –1.16516)	77.50	56.09	56.09	51.1%	0.973

5.3.2 Analysis of Results

Upon examining the results in Tables 5.4-5.7, the most obvious finding is that the results for *MSE* and *D* never indicate the standard deviation as outperforming either the variance or logarithmic transformation as a variability measure. In fact, the performance of the measure always appears to be second among the three choices. The reason for this behavior is best explained by examining the situation graphically. As discussed in (i)-(iii) below, we encounter three separate situations in the results:

(i) Consider the conditions established for the experiment in Part A, specifically the case where $\bar{s} \approx 1$. Figure 5.1 depicts the contour plots for each of the normalized response surfaces, along with the circular experimental region (——) and the target line

(▪▪▪▪▪), indicating all of the points by which the process mean is equal to the target of 77.5 μm.



(a) (b) (c)

Figure 5.1. Comparison of Optimal Factor Settings (•) for Variability Measures (a) Standard Deviation, (b) Variance, and (c) Logarithm of the Standard Deviation

The fitted response surfaces within Figure 5.1 are slightly different. The surfaces for the normalized standard deviation and variance measures are more elongated than that of the logarithmic function, and the latter appears to be centered more within the experimental region of interest. Suppose we overlap the contour surfaces in Figure 5.1 and the minimum for each surface is identified in relation to the target line and the experimental region of interest, as shown in Figure 5.2. The minima within the region of interest are labeled S, V, and L for the standard deviation, variance, and log functions, respectively. For this case, a comparison of the surfaces at an identical contour value indicates that the logarithm function is able to achieve a minimum closer to the target line within the experimental region of interest. The numerical results indicate the same outcome – the logarithm function is able to achieve a smaller *MSE* and the resulting optimal factor settings for each of the variability measures are slightly different.

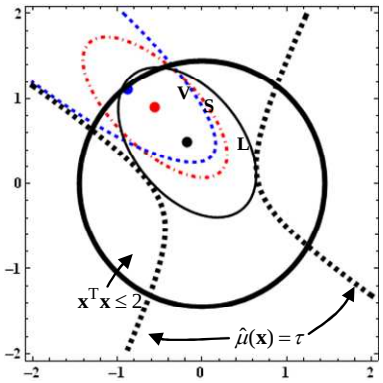
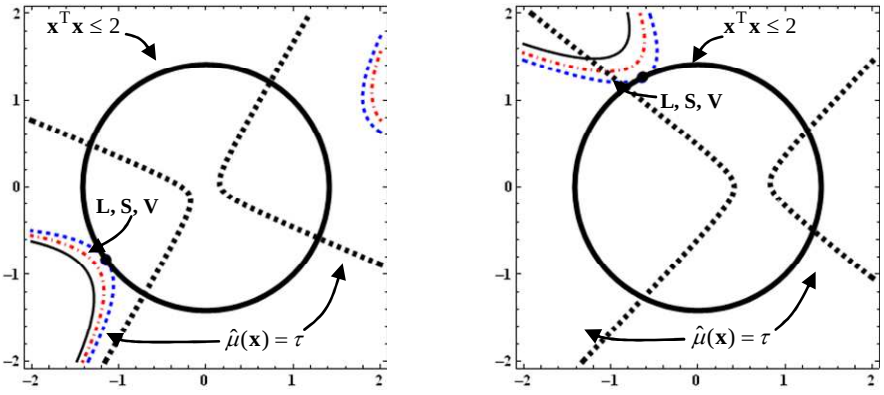


Figure 5.2. Relation between Response Surfaces at Identical Contour Heights and the Target Line ($\bar{s} \approx 1$) for the Standard Deviation (–•–), Variance (▪▪▪▪), and Logarithm of the Standard Deviation (——)

(ii) With some of the experimental conditions, however, the optimal factors identified among the different variability measures are found to be the same, as in Part A for (a) $\bar{s} \approx 10$ and (b) $\bar{s} \approx 30$. In this situation, the minima for the unconstrained surfaces are likely found outside the experimental region of interest; when the surfaces are then constrained, it creates a situation where each minimum within the experimental region of interest aligns at the same point on the boundary.



(a) $\bar{s} \approx 10$ (b) $\bar{s} \approx 30$

Figure 5.3. Relation between Response Surfaces at Identical Contour Heights and the Target Line for the Standard Deviation (–•–), Variance (▪▪▪▪), and Logarithm of the Standard Deviation (——)

As shown in Figure 5.3 (a) and (b), corresponding to the numerical results, the variance estimator actually achieves a value closer to the target line when performing an identical contour height comparison.

(iii) In addition, the numerical output indicates that there are situations, specifically when $\bar{s} < 1$, where the desired target of 77.5 μm is not achievable, and some process bias results. This particular situation occurs when the target line lies outside the experimental region of interest, as observed in Figure 5.4. In this case, given identical contour heights for each of the surfaces, it is the variance estimator which achieves a value closer to the target line.

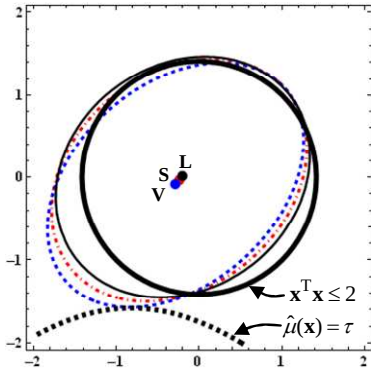


Figure 5.4. Relation between Response Surfaces at Identical Contour Heights and the Target Line ($\bar{s} < 1$) for the Standard Deviation (–•–), Variance (▪▪▪▪), and Logarithm of the Standard Deviation (——)

The analysis in (i), (ii), and (iii), reveals several findings. First, the primary reason that the standard deviation appears to perform unsatisfactorily from a numerical perspective as compared to the variance and logarithm function is due to the scaling associated with each measure. The scale of the functions naturally increases when one transitions from the logarithm function to the standard deviation to the variance. Thus,

the contours of the surface associated with the standard deviation are naturally bounded by the contours of the logarithm function and variance for any condition. Depending on the orientation of the surfaces within the experimental region of interest, it is then either the contours of the logarithm function or variance that are aligned closer to the target line. Figure 5.5 captures this concept graphically; the results from adjusting the target line in the case when $\bar{s}$ <1 are shown in the upper left portion of each graph.

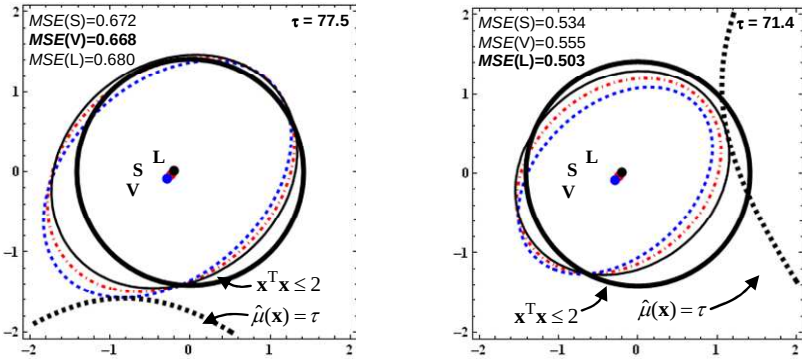


Figure 5.5. Relation between Response Surfaces at Identical Contour Heights and the Target Line ($\bar{s}$ <1) for the Standard Deviation (–•–), Variance (▪▪▪▪), and Logarithm of the Standard Deviation (——)

Another finding from the analysis is the fact that the apparent quality of the fit for each response surface has little effect on the performance of the variability measure in terms of *MSE* or composite desirability *D*. Nearly one-half of the experiments conducted in Part C and D produced results where the measure with the lowest R^2 value produced the lowest *MSE* and highest *D*. In addition, for a given set of conditions, the optimization scheme selected appears to have only a minor effect on the performance of the variability measure – in only one instance did the measure with the lowest *MSE* change upon transitioning from a constrained to an unconstrained mean problem. The simultaneous optimization of the mean and normalized variance, however, can result in optimal factor

settings that are slightly different than in the constrained problem. Finally, there appears to be little to no effect of increasing the mean or variance of the standard deviation on the performance of any one measure – the optimal solutions continue to shift between the variance and logarithm function measures. In sum, the overwhelming factor in the performance of each measure is the location of the target line with respect to the minima for the normalized surface, as the variation among the minima is slight in nearly every case. Thus, the problem is data-driven; the performance of any two measures may change just by examining a different data set.

5.3.3 Further Discussion

While the analysis in Section 5.3.2 offers insight into why one variability measure may outperform another measure, it does not indicate which measure may be more suitable to use. Researchers generally use the logarithm function when the values of the standard deviation are extremely close to zero in order to avoid cases where a fitted function may result in values less than zero. In contrast, they will either use the standard deviation or variance for any other set of experimental conditions, so further discussion is limited to these estimators.

For the purpose of illustration, consider a basic example where r replicated observations are collected on a single factor over multiple levels. Given the calculations for the standard deviation (•) and the variance (▲), first-order fitted functions are developed for s and s^2. Figure 5.6 portrays the general relationship between these functions and the square of the fitted function $\hat{\sigma}(x)$. A second-order fitted function for s^2 is also shown for comparison.

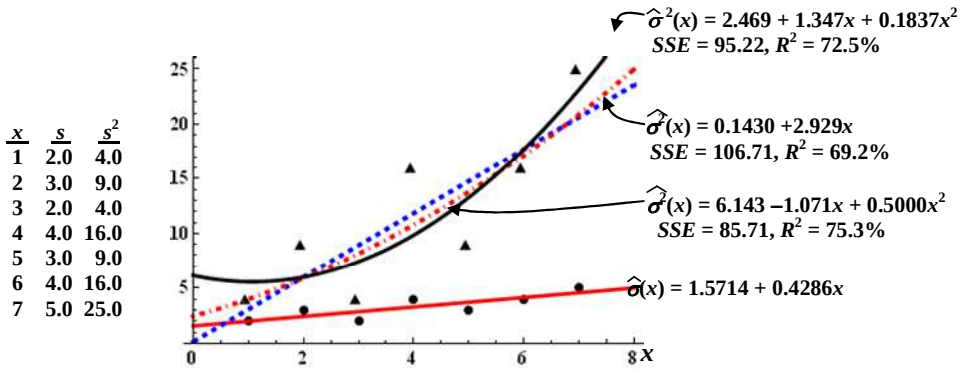


Figure 5.6. Example – Comparison of $\hat{\sigma}^2(x)$ and $\widehat{\sigma^2}(x)$

When the first-order fitted function for the standard deviation $\hat{\sigma}(x)$ is squared, the result is a second-order polynomial. As the data for the variance maintains a quadratic relationship, this function is naturally able to provide a more suitable estimation. The precision of the function $\hat{\sigma}^2(x)$, however, is insufficient when compared to the second-order fitted function for the variance $\widehat{\sigma^2}(x)$. The same general result is observed when the standard deviation and variance are estimated using second-order fitted functions, as shown in Figure 5.7.

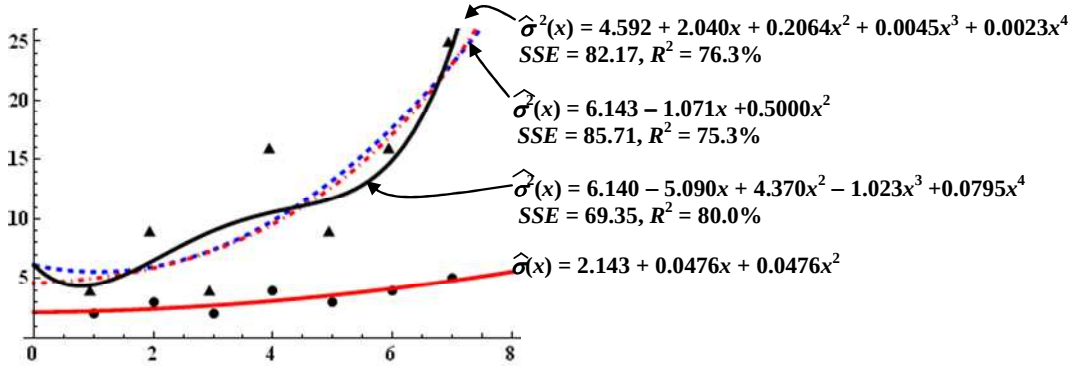


Figure 5.7. Example – Comparison of $\hat{\sigma}^2(x)$ and $\widehat{\sigma^2}(x)$

The example above presents the pattern which is observed when two, three, or higher-factor experiments are performed. When both $\hat{\sigma}^2(x)$ and $\widehat{\sigma^2}(x)$ are of the same order, the latter is able to provide a more suitable estimation of the response, as it is designed to provide the least squares fit for the data. In contrast, when $\hat{\sigma}(x)$ and $\widehat{\sigma^2}(x)$ are of the same order, and $\hat{\sigma}(x)$ is then squared for normalization, $\hat{\sigma}^2(x)$ is able to provide a more suitable estimation. This presents a tradeoff which describes exactly when one variability measure should be used over another. There are additional considerations, however, that make using $\widehat{\sigma^2}(x)$ over $\hat{\sigma}^2(x)$ more difficult.

Consider, for instance, the same conditions outlined within Section 5.3.1; the fitted functions for $\hat{\sigma}^2(\mathbf{x})$ and $\widehat{\sigma^2}(\mathbf{x})$ are shown below:

$$\widehat{\sigma^2}(\mathbf{x}) = 0.3396 + 0.0650X_1 - 0.0024X_2 - 0.0958X_1X_2 + 0.1291X_1^2 + 0.1402X_2^2$$

$$\begin{aligned}\hat{\sigma}^2(\mathbf{x}) = {} & 0.3069 + 0.0527X_1 - 0.0076X_2 - 0.0623X_1X_2 + 0.1154X_1^2 + 0.1243X_2^2 - 0.0067X_1^2X_2 \\ & + 0.0114X_1X_2^2 + 0.0097X_1^3 - 0.0015X_2^3 - 0.0114X_1^3X_2 - 0.0125X_1X_2^3 + 0.0260X_1^2X_2^2 \\ & + 0.0104X_1^4 + 0.0126X_2^4\end{aligned}$$

In this case, the predictive ability of the fourth-order model $\hat{\sigma}^2(\mathbf{x})$ is much greater than the second-order model $\widehat{\sigma^2}(\mathbf{x})$. In order to overcome this problem and continue to achieve a least squares fit of the data, a fourth-order model for $\widehat{\sigma^2}(\mathbf{x})$ is then constructed. Pursuant to a screening test for multi-collinearity between the predictor variables, the variance inflation factors for the terms in $\widehat{\sigma^2}(\mathbf{x})$ all tend to infinity, as the variables are found to be perfectly correlated with one another. Furthermore, the terms X_1X_2, $X_1{}^3X_2$, and $X_1X_2{}^3$ are aliased within the factor space for the fourth-order model, so the columns

of the design matrix **X** are found to be linearly dependent. Even if we remove the fourth-order terms and retain the third-order terms, more than one-half of the predictor variables in the resulting model produce variance inflation factors that tend to infinity. Although a more accurate estimate of σ^2 may be obtained through using $\widehat{\sigma^2}(\mathbf{x})$ the confidence intervals for each of the regression coefficients in the model are extremely wide, thus wrongly influencing the identification of the optimal factor settings **x***. The reason $\hat{\sigma}^2(\mathbf{x})$ is not affected by the same issue is due to the reduced parameter space it contains compared to $\widehat{\sigma^2}(\mathbf{x})$. Given parameters α_0, α_1, ..., α_p for $\hat{\sigma}(x)$ and parameters β_1, β_2, ..., β_p for $\widehat{\sigma^2}(x)$, Table 5.8 presents the general construct of different response surface designs for a single-factor case.

Table 5.8. Comparison of Parameter Space for $\hat{\sigma}^2(x)$ and $\widehat{\sigma^2}(\mathbf{x})$

Basis	*Model*	*Order*	*# Parameters*
Given: $\hat{\sigma}(x) = \alpha_0 + \alpha_1 x$ (First Order)	$\hat{\sigma}^2(x) = \alpha_0^2 + 2\alpha_0\alpha_1 x + \alpha_1^2 x^2$	2	2 (α_0, α_1)
	$\widehat{\sigma^2}(x) = \beta_0 + \beta_1 x$	1	2 (β_0, β_1)
	$\widehat{\sigma^2}(x) = \beta_0 + \beta_1 x + \beta_2 x^2$	2	3 (β_0, β_1, β_2)
Given: $\hat{\sigma}(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2$ (Second Order)	$\hat{\sigma}^2(x) = \alpha_0^2 + 2\alpha_0\alpha_1 x + (2\alpha_0\alpha_2 + \alpha_1^2)x^2 + 2\alpha_1\alpha_2 x^3 + \alpha_2^2 x^4$	4	3 (α_0, α_1, α_2)
	$\widehat{\sigma^2}(x) = \beta_0 + \beta_1 x + \beta_2 x^2$	2	3 (β_0, β_1, β_2)
	$\widehat{\sigma^2}(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4$	4	5 (β_0, β_1, β_2, β_3, β_4)

Thus, $\hat{\sigma}^2(x)$ is able to achieve the higher-order estimation without the over-parameterization that results from developing a higher-order regression model as shown for $\widehat{\sigma^2}(x)$.

In most dual response approach experiments, second-order designs are established for $\hat{\sigma}(\mathbf{x})$ and $\widehat{\sigma^2}(\mathbf{x})$. If the experiment has only two factors, it is not possible for $\widehat{\sigma^2}(\mathbf{x})$ to

upgrade to a fourth-order model without violating the prescribed limits of multi-collinearity. Given a three-factor experiment, some of the terms would likely be aliased and would therefore have to be removed. In this case, it is possible that the resulting fourth-order model for $\widehat{\sigma}^2(\mathbf{x})$ with less terms than $\hat{\sigma}^2(\mathbf{x})$ might serve as a more suitable estimate for the response; given a particular example, each measure would have to be analyzed statistically. Finally, given four or more factors, it is likely that $\widehat{\sigma}^2(\mathbf{x})$ is able to achieve a fourth-order fit without the presence of undue correlation within the model, and thus would be more suitable to use. Thus, in sum, the dimensions of the factor space play the most crucial role in variability measure selection; the variance estimator becomes increasingly more accurate as the number of factors in an experiment increase.

5.4 Conclusion

When using response surface methods to model the process variability in a robust design approach, several different variability measures may be considered. Typically, the standard deviation, variance, or logarithm of the standard deviation is modeled within an experimental framework. The response surfaces for each of these measures differ, however, leading to optimal factor settings that also vary slightly from one another. To further complicate the matter, most researchers will use existing data sets in the literature to demonstrate an improvement of performance in one model over another. Thus, an analysis of variability measure selection is warranted. The analysis conducted in this chapter suggests that the performance of each variability measure is completely driven by the data in an experiment, regardless of the conditions and whether the process mean is constrained or unconstrained. Therefore, in terms of an evaluation criterion such as the

mean square error or desirability index, a researcher may demonstrate an improvement over an existing model just by changing the variability measure or by examining a different data set. It is important to also note that each measure has specific attributes that make it more attractive to use over another under a given set of conditions. Given normalized response surface designs of the same order, the models using the standard deviation and logarithm function can never achieve the least squares fit that a model using the variance can obtain. Fitted functions for the variance, however, face challenges with multi-collinearity when attempting to model a response using higher-order terms. If the tradeoff between these measures is recognized, it becomes more feasible to compare and contrast existing models in robust design.

CHAPTER SIX

DEVELOPING A CONDITIONS-BASED APPROACH TO TIER-ONE ESTIMATION IN ROBUST PARAMETER DESIGN

6.1 Introductory Remarks

Practically every form of statistical analysis relies upon the validity of certain assumptions associated with the inherent conditions of the data being examined. In fact, most of the statistical procedures used by researchers throughout the quality engineering literature are based upon underlying assumptions of normality and only moderate variability among system outputs. Similarly, even analyses absent any distributional context typically require independence between observed data. However, many such assumptions about the data are made in the interests of tractability and time. As an example, consider the ordinary least squares (*OLS*) approach to regression. Huber (1973) noted that Gauss' original justification for *OLS* regression was "somewhat circular." That is, while *OLS* estimates are optimal only if the errors are independent and identically distributed normal random variables with constant variance, Gauss only assumed this about the errors so that he could then use the sample mean, which was (and still is) "generally accepted as a good [tier-one] estimate" for the location and which turns out to be the optimal estimate in the simplest – or normal case.

Plainly, when statisticians and mathematicians of the day began using *OLS* regression they did not have the computational capabilities we have at our disposal today. Thus, what made *OLS* regression particularly advantageous was the simplistic and explicit manner in which it could be computed from the data. As a testament to these

qualities, Rousseeuw and Leroy (1987) noted its continued widespread use after more than 200 years due to "tradition and computational speed;" a fact that remains true to this day.

As our collective understanding of industrial and other processes has evolved, researchers and practitioners have begun to accept that the assumptions of normality, moderate system variability, and constant variance in the data quite often do not hold in practice. Realistically, industrial processes often exhibit elevated levels of variability, particularly in mass production lines, which can confound many of the modeling assumptions behind the robust parameter design models available in the literature. Furthermore, it is also accepted that perfect normality rarely, if ever, exists in practice. Even Walter Shewhart, a notable pioneer in quality engineering, acknowledged the dearth of normally distributed data in industrial settings. Similarly, Pyzdek (1995) observed the following:

> "For instance, most… processes don't produce normal distributions. There are many reasons for this. One important reason is that the objective of most management and engineering activity is to control natural processes tightly, eliminating sources of variation whenever possible."

These efforts to mitigate variation cause distortions in the data. Thus, as Pyzdek (1995) points out, some asymmetry is not only realistic but practically inevitable – particularly in situations involving smaller-the better (*S*-type), larger-the-better (*L*-type), and select instances of nominal-is-best (*N*-type) quality characteristics; a condition that could become even more pronounced under conditions of elevated variability, especially when the number of observations or replicates at each design point is small.

Many contemporary quality assurance programs incorporate *OLS* methods to solve a variety of robust design problems as they endeavor to maintain a competitive edge in the industrial market. Pursuant to this, engineers develop linear or non-linear optimization schemes that most appropriately and accurately portray a given process to determine a set of optimal operating conditions that will assure minimal deviation from a specified mean performance target with minimum variability. However, if the true underlying process conditions are overlooked or otherwise assumed away, the estimates derived using the traditional least squares approach are likely to be problematic and misleading. Moreover, once applied to optimization schemes to determine optimal operating conditions, they may very well generate suboptimal solutions and lead to dubious recommendations to decision makers.

At this point, three things should be clear: 1) the quality and reliability of the optimization results depend quite heavily on the initial process estimates used, 2) innate process conditions may affect the quality of those estimates, and 3) very little has been done in the *RPD* research field to explore alternatives to traditional estimation approaches when certain conditions exist. For these reasons, this chapter will outline a conditions-based approach for working with non-standard conditions. In particular, modifications to traditional *RPD* methodology are proposed that incorporate the various aspects of data analysis outlined in Chapter 4 in order to ascertain underlying data conditions and then use this information to drive the selection of tier-one estimators. Numerical examples and Monte Carlo simulation will serve to illustrate how the proposed methodology is used, as well as which estimators tend to perform best under asymmetric and highly

variable conditions. The original work for the research outlined in and otherwise associated with this chapter is published with references Boylan and Cho (2012b, 2012c, and 2013a).

6.2 Methodology Development

While a variety of conditions could affect estimator selection, this paper focuses on inherent asymmetry and elevated variability in process outputs. A conditions-based approach to the *RPD* problem is proposed, whereby a more comprehensive data analysis is used to illuminate intrinsic process conditions that will, in turn, inform the selection of appropriate tier-one estimation approaches and the development response surface functions for the process mean and variability. The result is a methodology that is based on a fuller understanding of innate process characteristics that will, in turn, facilitate a more realistic and accurate portrayal of the process under investigation. A proposed sequence for implementing this approach is portrayed in Figure 6.1, wherein the highlighted portions in Phase Ib (sub-steps (i) and (ii)) encompass proposed enhancements to the *RPD* methodology.

At this point, it is important to distinguish between the goals of this chapter and the proposed methodology that will evolve based upon the results to be shown. Because this chapter aims to demonstrate the differences between estimator performances given intrinsic process conditions, the integration of the estimation approaches into the *RPD* framework includes all of the candidate estimators under consideration to facilitate comparisons and to support conclusions. However, as the results will eventually show, certain estimators will indeed tend to perform best when certain conditions prevail. Thus,

the proposed methodology actually advocates an in-depth analysis to ascertain those conditions and then the selection of the most appropriate location and scale estimators based upon the identified conditions.

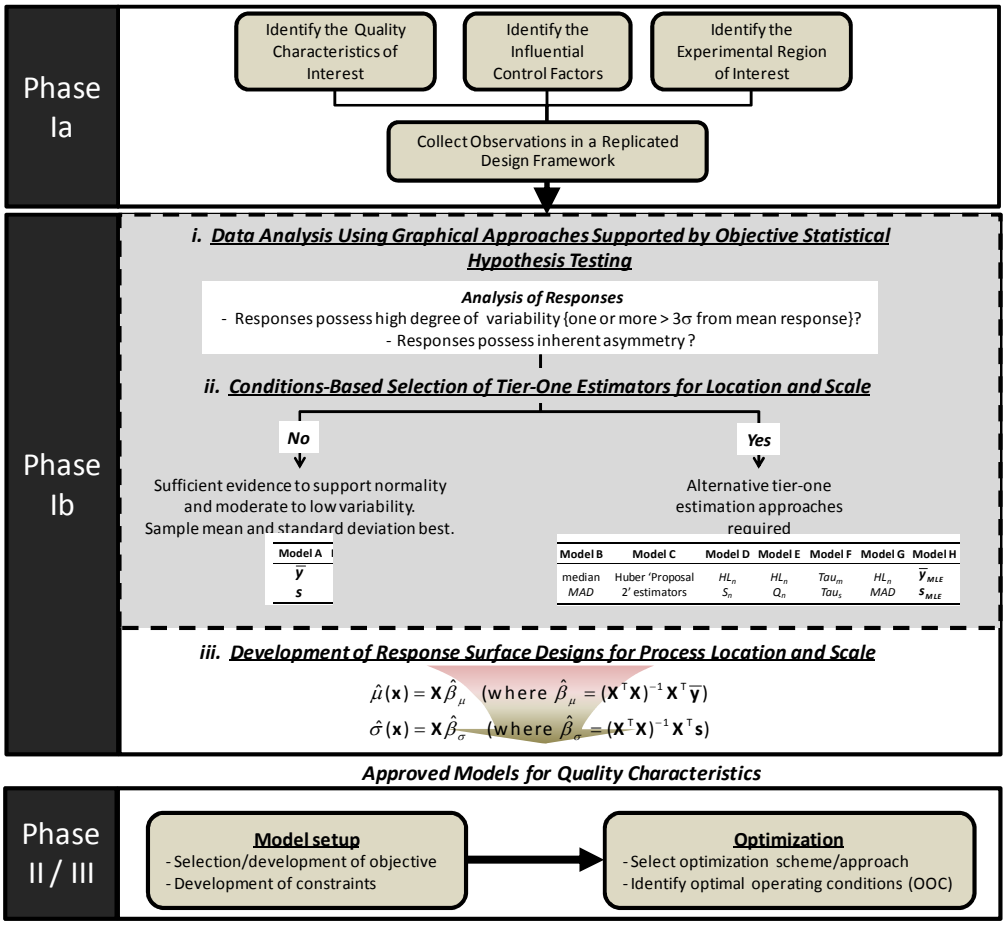


Figure 6.1. Process map of proposed methodology for selecting tier-one estimators

6.2.1 Establishing an Experimental Framework

The procedures described in this chapter focus on the selection of appropriate tier-one estimators when non-standard conditions such as asymmetry and/or high variability prevail. Although addressed in the context of the general univariate problem, these procedures can be extended to the multivariate case, as well as to other conditions.

Hereafter, Section 6.2.1.1 will lay out a general experimental framework for the *RPD* problem. Section 6.2.1.2 will delineate the procedures for selecting appropriate tier-one estimators under asymmetric conditions. Lastly, Section 6.2.1.3 will address procedures for estimator selection when high variability pervades system outputs.

6.2.1.1 The General Univariate Problem

Experimentation and analysis begin with the specification of response variables, factors or predictors influencing the responses, and region of interest for a designed experiment. Suppose the objective is to identify the optimal factor settings $\mathbf{x}^*=(X_1^*, X_2^*, ..., X_k^*)$ that support achieving a mean process performance with minimal deviation from a desired target and with minimum variability in the result. Pursuant to this, consider an experimental framework whereby a quality characteristic, Y, is influenced by a set of control factors $X_1, X_2, \ldots, X_k$. The experiment consists of n design points, or runs, each of which contains m replicates for the observed response. Let y_{qj} denote the jth response at the qth design point, where $q = 1, \ldots, n$ and $j = 1, \ldots, m$. Table 6.1 portrays the framework for such an experiment.

Table 6.1. Experimental response surface methodology framework.

Design Point	X_1 X_2 .... X_k	Replications	$\bar{y}$	s	γ_3
1		y_{11}……y_{1m}	$\bar{y}_1$	s_1	γ_{31}
⋮	Control	⋮ ⋮	⋮	⋮	⋮
q	Factor	y_{q1}……y_{qm}	$\bar{y}_q$	s_q	γ_{3q}
⋮	Settings	⋮ ⋮	⋮	⋮	⋮
n		y_{n1}……y_{nm}	$\bar{y}_n$	s_n	γ_{3n}

The replicates at each design point are then used to obtain parameter estimates for the data. Traditionally, this includes the sample mean ($\bar{y}$) and standard deviation (s) (or

variance, s^2), as these obviously define the assumed normal distribution. However, when asymmetry is suspected or expected, the mean and the standard deviation no longer completely define the distributional properties inherent to the data. Thus, the sample skewness (γ_3) must also be obtained to account for the non-normality in the responses. These estimates for the first three sample moments in the data are found using the following formulas:

$$\bar{y}_q = \frac{\sum_{j=1}^{m} y_{qj}}{m}, \quad s = \frac{\sum_{j=1}^{m} y_{qj}}{m-1}, \text{ and } \quad \gamma_3 = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_{qj} - \bar{y}_q)^3}{\left(\frac{1}{n-1}\sum_{i=1}^{n}(y_{qj} - \bar{y}_q)^2\right)^{3/2}} \tag{6.1}$$

Prior to the estimation of response surface functions for the mean and variability of the process, a comprehensive data analysis of both the responses and the residuals (obtained through preliminary regression analysis using estimators for the mean and standard deviation) must be performed to ascertain the underlying conditions in the data. This includes an investigation of normality and variability in the process responses, as well as the verification of the assumptions of normality, homoscedasticity, and independence in the residuals. These are investigated using both graphical and objective methods, which may include the following approaches briefly discussed in the following paragraphs.

Given the focus on inherent process conditions, namely asymmetry and high variability, a thorough analysis of the process outputs, or observed responses for the quality characteristic of interest, must be performed. This is accomplished in two ways. The first consists of a comprehensive analysis using the graphical techniques outlined in Chapter 4. In particular, these methods will serve to determine whether the distribution is

symmetric or asymmetric, the degree of positive or negative skewness, the degree of variability, as well as whether or not kurtosis is present. The second consists of more objective approaches that serve to reinforce deductions inferred from the graphical analysis. In this particular case, we are interested in asymmetry and high variability, which will drive the choice of the objective methods used. Parts (a) and (b) address each condition, respectively.

(a) Assessing asymmetry. Since most statistical tests relating to distributional contexts focus on whether data are normally distributed or not, testing for asymmetry is best achieved by approaching the hypothesis from the opposite context. That is, by testing for normality. This can be accomplished via a variety of well-established methods, such as the Kolmogorov-Smirnov test for large samples (> 2000), or the Shapiro-Wilk test for small or medium-sized samples. Generally, due to a variety of reasons including experimental costs, the sample sizes obtained in *RPD* experimentation typically fall in the latter category. Thus, for the Shapiro-Wilk test, when the *p* observations made on quality characteristic are sorted in ascending order, the alternative hypotheses $H_0 : Y \in N(\mu, \sigma^2)$ and $H_1 : Y \notin N(\mu, \sigma^2)$ are evaluated using the *W* statistic given by: $W^* = \left[\frac{b}{s\sqrt{p-1}} \right]^2$, where $b = \sum_{l=1}^{\kappa} a_{p-l+1} \left[y_{(p-l+1)} - y_l \right]$. Here, κ is the largest integer that is less than or equal to *p/2*, and *s* denotes the sample standard deviation. For a given significance level α, tables are then used to reference the coefficients *a* and the critical values W_α. Different from most statistical tests, since the critical region lies in the small tail of the distribution, if $W^* > W_\alpha$ then H_0 is concluded (that is, sufficient evidence exists

to suggest the observations follow a normal distribution). It is important to note that one drawback to these objective tests is that if the data are not normally distributed they provide no information regarding the manner in which they are distributed. This highlights the importance of first using the graphical approaches described in Chapter 4 to overcome this shortfall and to provide more salient information about the data.

(b) Assessing high variability. Determining whether a process is highly variable is somewhat of a subjective matter delineated in objective terms. A trademark of high variability data is that their sample standard deviation is in general quite large, implying a "largely uninformative" sample mean that fails to adequately describe the location of the bulk of the observed values. Using this concept, a highly variable process is classified as one in which the range of variability in the responses is noticeably large and where one or more of the responses lies more than three standard deviations ($+/-3\sigma$) from the mean response.

Since *OLS* regression is applied in the development of response surface designs in this Chapter, the necessary assumptions underpinning this particular approach must first be validated. Regardless of the distribution used to model *Y*, the residuals in fitting a response variable should be normally distributed, uncorrelated, and exhibit homoscedasticity or constant variance. As in the analysis of the responses, a variety of graphical methods may be used to ascertain the degree to which the residuals comply with these assumptions and more objective hypotheses tests may be conducted to confirm the feedback from these visual results. As with the responses, normality in the residuals may also be examined using graphical measures such as the normal probability plot,

followed by the more objective methods provided by the Kolmogorov-Smirnov or Shapiro-Wilk tests. To investigate independence, the Durbin-Watson test is usually sufficient to detect a lack of randomness in the residuals. Should remediation be necessary, one can add predictor variables or use transformations in the variables to eliminate interdependencies. Heteroscedasticity, or non-constant variance, is most often investigated graphically using a plot of the residuals against the fitted values, as well as objectively using either the Brown-Forsythe test, which is more robust to departures from normality in the data, or the Breusch-Pagan test. For the purposes of this chapter, it is assumed that these assumptions hold, thereby eliminating the need to illustrate the graphical and objective analyses in detail. Chapter 7 will examine them more closely in the context of tier-two estimation.

6.2.1.2 Modeling Asymmetry

In traditional *RPD* applications, asymmetric conditions typically observed in the univariate *S*- and *L*-type problem are often modeled via a Weibull or gamma distribution. Ideally, however, it would be preferable to use a distributional model capable of supporting both the symmetry usually assumed in the *N*-type model, as well as the asymmetry of the *S*-, *L*-, and certain *N*-type models. Although some common distributions (such as the gamma, Weibull, and unbounded Johnson distributions) can effectively portray processes with innate skewness, they present challenges in modeling normality when zero skewness exists.

Due to its inherent relationship to the normal distribution, the skew normal (*SN)* distribution provides a suitable alternative for modeling both symmetric and asymmetric

situations. First introduced by O'Hagan and Leonhard (1976) and addressed more recently by Azzalini (1985), Azzalini and Dalla-Valle (1998), and Arellano-Valle *et al.* (2002), the skew normal distribution extends the normal by incorporating a third parameter, α, as a shape parameter to account for non-zero skewness. Incorporating this shape parameter, the probability density function for the skew normal relative to the normal distribution is given by:

$$f(x|\alpha) = 2\phi(x)\Phi(\alpha x) \quad , x \in \mathbb{R}$$

where $\phi(x)$ and $\Phi(\alpha x)$ (with $\alpha = 0$) correspond to the probability density and cumulative distribution functions of the normal distribution. Recall that the normal probability density function for some random variable *Z* with parameters μ and σ^2 can be rewritten in terms of the standard normal density function

$$f_z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(z-\mu)^2}{2\sigma^2}\right\} = \frac{1}{\sigma}\phi\left(\frac{z-\mu}{\sigma}\right) \tag{6.2}$$

We can easily extend this by adding location (ξ) and scale (ω) parameters to the density function, using the transformation $x \to (x-\xi)/\omega$. This yields:

$$\begin{aligned} f_x(x) &= \frac{2}{\sigma}\left[\phi\left(\frac{x-\xi}{\omega}\right)\right]\Phi\left[\frac{\alpha(x-\xi)}{\sigma}\right] \\ &= \frac{2}{\omega}\left[\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(x-\xi)^2}{2\omega^2}\right\}\right] \times \frac{1}{2}\left[1+\text{erf}\left\{\frac{\alpha(x-\xi)/\omega}{\sqrt{2}}\right\}\right] \\ &= \frac{1}{\omega\sqrt{2\pi}}\exp\left\{-\frac{(x-\xi)^2}{2\omega^2}\right\} \times \frac{1}{2}\left[1+\text{erf}\left\{\frac{\alpha(x-\xi)}{\omega\sqrt{2}}\right\}\right] \end{aligned} \tag{6.3}$$

It can be shown that when $\alpha = 0$, the skew normal distribution reduces to the normal, making normality a special case of the $SN(\xi, \omega, \alpha)$ distribution. From Azzalini (1985), the mean and standard deviation of a $SN(\xi, \omega, \alpha)$ distribution are given by:

$$E[y] = \hat{\mu} = \xi + \omega\delta\sqrt{2/\pi} \tag{6.4}$$

$$s = \sqrt{\omega^2\left(1 - 2\delta^2/\pi\right)}, \tag{6.5}$$

where $\delta = \alpha/\sqrt{1+\alpha^2}$. The *SN* distribution is a relatively new distribution compared to the more commonly observed family of continuous distributions. Since it derives from the normal distribution, which remains widely used among quality engineers for *N*-type characteristics, its extension to *S* and *L*-type characteristics may help to overcome many of the modeling complexities encountered in asymmetric situations.

As Goethals and Cho [53] showed, modeling system properties with the skew normal distribution can be achieved by first calculating estimates for the first three sample moments (mean, standard deviation, and skewness) for the *q*th design point. Although the location and scale of a skew normal distribution are not equivalent to the mean and standard deviation as they are in the normal case (unless the shape parameter $\alpha = 0$), since the number of replicates *m* at any one design point are usually low (i.e., $m<5$) we may assume normality among the replicates. Under this assumption, the sample mean $\bar{y}_q$ and standard deviation s_q then correspond to the location (ξ_q) and scale (ω_q) parameters at the *q*th design point. Thereafter, these estimates are used to derive estimates for the skew normal process mean and standard deviation by applying them to Equations (4) and (5) as follows:

$$\hat{\mu}_{q(SN)} = \bar{y}_q + s_q \delta_q \sqrt{2/\pi} \tag{6.6}$$

$$s_{q(SN)} = \sqrt{s_q^{\,2}\left(1 - 2\delta_q^2/\pi\right)} \tag{6.7}$$

Here, the parameter δ_q is estimated using the sample skew. In short, using an alternative formulation for sample skew provided by Azzalini (1985), an estimate for δ_q can be derived as follows:

$$\hat{\gamma}_{3q} = \frac{4-\pi}{2}\frac{\left(\delta\sqrt{2/\pi}\right)^3}{\left(1-2\delta^2/\pi\right)^{3/2}} \qquad \Rightarrow \qquad \left|\hat{\delta}_q\right| = \sqrt{\frac{\pi}{2}\cdot\frac{\left|\hat{\gamma}_{3q}\right|^{2/3}}{\left|\hat{\gamma}_{3q}\right|^{2/3} + ((4-\pi)/2)^{2/3}}}$$

where the sign of $\hat{\gamma}_{3q}$ determines the sign of δ_q. For the purposes of simulation, the estimate for δ_q may then be used to estimate the shape parameter directly by rearranging the previously stated relationship in the following way:

$$\delta_q = \frac{\alpha_q}{\sqrt{1+\alpha_q^2}} \qquad \Rightarrow \qquad \hat{\alpha}_q = \frac{\hat{\delta}_q}{\sqrt{1-\hat{\delta}_q^2}}$$

This step is necessary in R, as the shape parameter and sample skew are scaled differently in the context of the skew normal distribution.

As Equations (6.6) and (6.7) suggest, the estimates for the process mean and standard deviation are influenced by the inclusion of the sample skew. Thus, by using this method, we ensure that inherent process skewness is accounted for in the ultimate response surface estimates used for the process mean and standard deviation, and so further ensure that actual process characteristics are more accurately represented.

6.2.1.3 Estimator selection under asymmetric conditions

Statistical literature is replete with myriad estimators for location and scale. A great many of these, commonly known as robust estimators, evolved to address the problems presented by outliers. In this instance, the term robust describes an estimator's ability to overcome the influences or leverage exerted by outliers on the estimate. However, whereas outliers are often viewed as anomalies and errant data, asymmetry can be an inherent and intricate part of a process, particularly in the case of certain types of quality characteristics. But the idea of asymmetry or skewness is not so far removed from the discussion about outliers. In fact, but for the inherent heaviness in the long tail of a skewed distribution, the distribution might not be skewed at all. In examining asymmetric conditions, four location and five scale estimators are considered as potential alternatives to the sample mean and standard deviation (base case), listed in Table 6.2.

Table 6.2. Location and scale estimators examined as potential *RPD* alternatives.

	Location Estimators	**Scale Estimators**
Base Case	sample mean ($\bar{y}$)	sample standard deviation (s)
Alternatives for Comparison	median ($\tilde{y}$) Huber 'Proposal 2' M-estimator ($\tilde{y}_{n,H2}$) Hodges-Lehmann estimator (HL_n) Tau ($\tilde{y}_{n,\tau}$)	median absolute deviation (*MAD*) Huber 'Proposal 2' M-estimator ($s_{n,H2}$) S_n estimator Q_n estimator Tau ($s_{n,\tau}$)

The reasons for selecting this particular set of alternative estimators are quite simple. First, the top two alternatives in each category are the only tier-one estimation alternatives that have been examined in *RPD* settings. Consequently, the prior efforts

involving them, listed in Table 2.3 in Chapter 2, provide a logical basis of comparison for the conditions examined in this paper. The remaining candidates are well-known in statistical circles for their robustness and efficiencies but are, as of yet, untested in the context of *RPD* and so are examined as potential alternatives here. The following paragraphs provide some basic background on these estimators and the reasons for including them; Table 6.3 summarizes them in the context of their mathematical formulations.

Park and Cho (2003) suggested the use of the sample median and the *MAD* as alternatives to the sample mean and standard deviation in *RPD* approaches where outliers contaminate the data. Although these estimators are well-known for their resistivity to the effects of outliers, they are equally well-known for their comparative lack of efficiency under normality. But in the case of the mean, this is not surprising, as the sample mean is known to be the maximum likelihood estimator of the population mean in the Gaussian case. However, it should be noted that this result reflects the *limiting* case and is more correctly referred to as the asymptotic relative efficiency (*ARE*) or the large-sample limit of the ratio of the variances. As Serfling (2011) notes, for smaller sample sizes the relative efficiency of the median to the mean improves. In particular, when n=5 in the normal case, the *ARE* jumps to 95%, and then decreases to 80% when n=10, 70% for n=20, and ultimately to 64% as $n \to \infty$. The *MAD* fairs even worse at 37% efficiency relative to the standard deviation, suggesting an even greater need for considerably larger samples to achieve equivalent performances. In terms of resistivity to the leverage of outliers, however, each possesses a breakdown of 50%, which is the highest possible.

This simply refers to the proportion of outlying observations that can be accommodated before the estimator "breaks down." By comparison, the mean and standard deviation have a breakdown of 0% since even a single outlying observation can induce a significant shift in the estimate.

To overcome the efficiency issues with the median and *MAD*, Lee *et al.*[36] extended the Park and Cho (2003) proposal by incorporating Huber's *M*-estimation approach to provide a highly efficient and resistant alternative for robust design solutions in situations involving outliers. The 'Proposal 2' estimator suggested by Huber (2009) uses a bounded monotone function to incorporate resistivity to outlying influencers without discounting them altogether. This so-called 'psi' function forms the basis of two equations which are solved simultaneously to obtain robust estimates for location and scale parameters. Relative to the sample mean and standard deviation, these estimators have *AREs* of 96% and 80%, respectively, which are marked improvements over the median and *MAD*.

Another well-established estimator yet to be applied in *RPD* problems is the Hodges-Lehmann location estimator (HL_n). Proposed by Hodges-Lehmann (1963) in the early 1960s, the HL_n estimator evolved as a robust nonparametric estimator of a population's location parameter, also referred to as the pseudo-median. Computed as the median of all pairwise averages the sample observations, the HL_n estimator is known for its excellent overall performance in terms of its efficiency and resistivity to outliers. In fact, with $ARE(HL_n, y, N(\theta, \sigma^2)) \approx 96\%$, this estimator is highly competitive with the sample mean under normality, can be "infinitely more efficient" under some other

symmetric distributions with heavier tails, and is "never much less efficient" at any distribution, thereby making it a viable alternative to consider under the asymmetric conditions examined here (Serfling, 2011).

Table 6.3. Estimators and associated formulae compared in hypothetical example.

Estimator	Formula
Sample Mean	$\bar{y}_q = \sum_{j=1}^{m} y_{qj} \Big/ m$
Sample Standard Deviation	$s_q = \sqrt{\sum_{j=1}^{m} (y_{qj} - \bar{y}_q)^2 \Big/ m-1}$
Sample Median	$\tilde{y}_q = \begin{cases} y_{q(k)} & \text{, odd \# data points} \\ \frac{1}{2}(y_{q(k)} + y_{q(k+1)}) & \text{, even \# data points} \end{cases}$ where k = the middle observation of the ordered data (if even, k and k+1 are the middle two data points)
Median Absolute Deviation (*MAD*)	$MAD_q = \underset{1 \le j \le m}{median} \left\{ \left\| y_{qj} - \underset{1 \le q \le m}{median}\left(y_q \right) \right\| \right\}$
Huber Proposal 2 (location and scale)	$\tilde{y}_{q,H2} = \mu_{Loc}$ and $s_{q,H2} = \sigma_{Scale}$ which are obtained by solving (1) and (2) simultaneously: (1) $\sum_{j=1}^{m} \psi\left(\frac{y_j - \mu_{Loc}}{\sigma_{Scale}} \right) = 0$ (2) $\sum_{j=1}^{m} \psi^2\left(\frac{y_j - \mu_{Loc}}{\sigma_{Scale}} \right) = \eta$ Here, ψ = Huber's ψ function, which is: $\psi(t) = \begin{cases} -c, & t < -c \\ x, & \lvert t \rvert \le c \\ c, & t > c \end{cases}$ where c = a user-specified tuning constant that accounts for the degree of efficiency the user is willing to sacrifice in exchange for robustness. $\eta = E\left[\psi^2(Z)\right]$, where $Z \sim N(0,1)$
Tau-m/s (location and scale)	$\tilde{y}_{q,\tau} = \frac{\sum_{j=1}^{m} y_{qj} w_{qj}}{\sum_{j=1}^{m} w_{qj}}$ and $s_{q,\tau} = \sqrt{\frac{s_0^2}{m} \sum_j \rho_{c2}\left(\frac{y_j - \tilde{y}_{q,\tau}}{s_0} \right)}$ where $w_{qj} = W_{c1}\left(\frac{y_{qj} - median(y_q)}{s_0} \right)$, $W_c(y_{qj}) = \left(1 - \left(\frac{y_{qj}}{c} \right)^2 \right)^2$, $\rho_c(y_{qj}) = \min(y_{qj}^{\ 2}, c^2)$ and $s_0 = MAD(y_q)$
HL_n	$HL_q = Median\left\{ \frac{y_{qi} + y_{qj}}{2} \right\}$ which amounts to the median of all pair-wise averages of the sample observations
S_n	$S_n = c \underset{q}{median}\left\{ \underset{j}{median}\left\{ \left\| y_{qi} - y_{qj} \right\| \right\} \right\}$ for j=1, ..., m and where c = a consistency factor that defaults to 1.1926
Qn	$Q_n = d \left\{ \left\| y_{qi} - y_{qj} \right\|; i < j \right\}_{(k)}$ Where d = constant factor, $k = \binom{h}{2} \approx \binom{m}{2} / 4$ And $h = [n/2] + 1$ is roughly half the number of observations

S_n and Q_n estimators for scale were developed by Rousseeuw and Croux (1993) as alternatives to the *MAD* that can be used as initial or ancillary scale estimates but that are

more efficient and not "slanted" toward symmetric distributions. The *MAD*'s partiality toward symmetry lies in the fact that, after estimating a central value (the median), it assigns equal importance to positive and negative deviations from that value. This clearly presents potential issues in asymmetric cases where such deviations may not bear equal importance. The S_n estimator uses median statistics to achieve the same high breakdown of the *MAD* (50%) but with greater efficiency (~58%). The Q_n uses a different order statistic that facilitates an equally high breakdown as the S_n, but with an even better efficiency at Gaussian distributions (~82%). More importantly in the context of this paper, both are equally suited for asymmetric distributions, as well.

Finally, *tau* estimators for location and dispersion were developed by Maronna and Zamar (2002) to reduce the computational complexities and times associated with other high-breakdown robust estimators. The location estimate uses a weighted mean that incorporates median statistics in establishing the weight; the scale estimate uses a truncated standard deviation that invokes the *MAD* as a starting point. These two estimators also combine relatively high efficiency (80% in each case) with high resistivity, but have yet to be examined in *RPD* problems.

6.2.1.3 Estimator selection in highly variable processes

As has been noted previously, elevated degrees of variability increase the potential for extreme observations from one tail or the other of the distribution. While such data points are often perceived as anomalies and therefore errant observations, under highly variable conditions they can be an inherent part of process performance that must be accounted for. This makes robust estimators particularly interesting because they

mitigate the leverage exerted by extreme observations without discounting them altogether. In examining high-variability conditions, five location and six scale estimators are considered as potential alternatives to the sample mean and standard deviation. As Table 6.4 shows, these are precisely the same estimators considered in the asymmetric case with one additional estimator each for the location and the scale. These are the maximum likelihood estimators (*MLE*) for the mean and standard deviation, which are obtained using the *SN* distribution previously described.

Table 6.4. Location and scale estimators examined in the high-variability case.

	Location Estimators	**Scale Estimators**
Base Case	sample mean ($\bar{y}$)	sample standard deviation (s)
Alternatives for Comparison	median (y_{median})	median absolute deviation (*MAD*)
	Huber 'Proposal 2' M-estimator (y_{H2})	Huber 'Proposal 2' M-estimator ($s_{n,H2}$)
	Hodges-Lehmann estimator (HL_n)	S_n estimator
	Tau (y_τ)	Q_n estimator
	Maximum Likelihood Estimator $\left(\bar{y}_{MLE}\right)$	Tau ($s_{n,\tau}$)
		Maximum Likelihood Estimator (s_{MLE})

The inclusion of the *MLE* in this case may seem somewhat redundant and therefore make little sense, recognizing that the location of a normal distribution equates to the population mean, for which the sample mean *is* the *MLE*. However, it is also widely acknowledged that true normality rarely, if ever, exists and that some asymmetry is not only realistic but practically unavoidable; a situation that could become even more pronounced under conditions of elevated variability, particularly when the number of observations or replicates at each design point is small.

As was previously mentioned, it can be shown that when $\alpha = 0$, the skew normal distribution reduces to the normal, making normality a special case of the *SN* distribution (Azzalini (1985), Arellano-Valle (2002)). Thus, using the skew normal distribution works to overcome problems presented by inherent asymmetry by allowing us to model both situations. If the data are "perfectly" normal or even very close, then it will treat the situation as a special case of the *SN* distribution with zero or negligible skewness. If they are not, however, it will capture any inherent skewness (rather than assume it away) and provide a more precise estimate with respect to the location and scale of the distribution. Since no closed form expression for the $SN(\xi, \omega, \alpha)$ parameters exists, they are estimated numerically by maximizing the likelihood function with respect to the components of $\theta = (\xi, \omega, \alpha)$:

$$\log L(\theta) = -\frac{n}{2}\log\frac{\pi\omega^2}{2} - \frac{1}{2\omega^2}\sum_{i=1}^{n}(y_i - \xi)^2 + \sum_{i=1}^{n}\log\Phi\left[\frac{\alpha(y_i - \xi)}{\omega}\right] \tag{6.8}$$

This can be achieved using the `sn.mle()` function from Package (*sn*) in R. This function fits a set of data to the *SN* distribution and then, using the obtained estimates for the three parameters, computes maximum likelihood estimates for the mean, standard deviation, and skewness for the set of observations.

6.2.2 Incorporating estimator selection into the *RPD* framework

Under either set of conditions, the procedures for integrating the estimators into the *RPD* framework are the same. Consider an industrial process involving a nominal-the-best (*N*-type) quality characteristic as the response of interest, *Y*, which depends on the selected levels of *k* control factors, $\mathbf{x} = (x_1, \ldots, x_k)$. Within this system, assume that the

levels of x_i for $i = 1, 2,\ldots,k$ are quantitative, continuous, and can be controlled by the experimenter. Further consider a response surface design consisting of m experimental runs performed at n design points. Let y_{qj} denote the j^{th} response at the q^{th} design point, where $j = 1,\ldots,m$ and $q = 1,\ldots,n$.

Using the formulations shown in Table 6.3 for each of the estimators under consideration, estimates for the location and scale are obtained at the q^{th} design point. Using these estimates, fitted response functions are then developed for the process location and scale using traditional least squares regression. In particular, assuming second-order polynomials for the response functions in each case, the general form of the estimated response functions for the process location and scale with k parameters or $k-1$ predictor variables appear as:

$$\text{Location:}\quad \hat{\mu}(\mathbf{x}) = \hat{\beta}_{\mu,0} + \mathbf{X}^{\mathrm{T}}\hat{\mathbf{b}}_{\mu} + \mathbf{X}^{\mathrm{T}}\hat{\mathbf{B}}_{\mu}\mathbf{X} + \varepsilon_{\mu} \tag{6.9}$$

$$\text{Scale:}\quad \hat{\sigma}\mathbf{x}) = \hat{\beta}_{\sigma,0} + \mathbf{X}^{\mathrm{T}}\hat{\mathbf{b}}_{\sigma} + \mathbf{X}^{\mathrm{T}}\hat{\mathbf{B}}_{\sigma}\mathbf{X} + \varepsilon_{\sigma} \tag{6.10}$$

where $\underset{\mathbf{1\times k}}{\mathbf{X}^{\mathrm{T}}} = \begin{bmatrix} X_1 & X_2 & \cdots & X_{k-1} \end{bmatrix}$, $\underset{\mathbf{k\times 1}}{\hat{\mathbf{b}}} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$, and $\underset{\mathbf{k\times k}}{\hat{\mathbf{B}}} = \begin{bmatrix} \beta_{11} & \beta_{12}/2 & \cdots & \beta_{1k}/2 \\ & \beta_{22} & \cdots & \beta_{2k}/2 \\ & & \ddots & \vdots \\ \textit{sym.} & & & \beta_{kk} \end{bmatrix}$

and where $\hat{\beta}_{\mu,0}$ (and $\hat{\beta}_{\sigma,0}$), $\hat{\mathbf{b}}_{\boldsymbol{\mu}}$ (and $\hat{\mathbf{b}}_{\boldsymbol{\sigma}}$) and $\hat{\mathbf{B}}_{\boldsymbol{\mu}}$ (and $\hat{\mathbf{B}}_{\boldsymbol{\sigma}}$) reflect the estimates of the intercept, linear, and second-order coefficients for the process location and scale, respectively. The term ε_{μ} and ε_{σ} correspond to the residual error for the mean and standard deviation. In similar fashion, fitted response surface functions are developed for each of the estimate vectors containing the supporting information for each parameter.

From here, typical approaches to *RPD* use the sample mean and standard deviation (or variance) as the basis for estimating the regression coefficients in Equations (6.9) and (6.10), respectively. With this in mind, seven additional combinations of the alternative estimators (Models B – H in Table 6.5) are considered for the purposes of performance comparison. Note that Model H is used only in the high-variability case.

Table 6.5. Estimator combinations used in comparative analysis.

Model A	Model B	Model C	Model D	Model E	Model F	Model G	Model H
$\overline{y}$	y_{median}	Huber 'Proposal	HL_n	HL_n	Tau_m	HL_n	$\overline{y}_{MLE}$
s	MAD	2' estimators	S_n	Q_n	Tau_s	MAD	s_{MLE}

To investigate the effects of the fitted functions at estimating a response, the *MSE*-based optimization scheme for an *N*-type quality characteristic developed by Cho (1994) and Lin and Tu (1995), denoted by

$$\text{Minimize} \quad MSE = (\hat{\mu}(\mathbf{x}) - \tau) + \hat{\sigma}^2(\mathbf{x}) \tag{6.11}$$

is applied to either a spherical region of interest (in the case of a central composite design) such that $\mathbf{x}'\mathbf{x} \leq \rho^2$, where ρ defines the spherical experimental region, and $\mathbf{x} \in$ the domain of the control factors, $\boldsymbol{\Omega}$, or a cuboidal region (in the case of a full factorial design) whereby $\mathbf{x}$ is bounded on the range (-1, 1) based on coded units. Using this approach as a framework, appropriate combinations of location and scale estimators are tested in their ability to determine optimal operating conditions for the system.

6.3 Numerical Examples and Simulation

In this section, several examples are used to illustrate both the impacts of inherent conditions on estimator performance and the utility of the proposed methodology as a result. Specifically, Section 6.3.1 will examine the effects of asymmetric conditions on estimator performance, while Section 6.3.2 will extend the problem to high variability processes to ascertain the impacts on estimator selection under this set of conditions. The numerical code used to solve the examples may be found in Appendix A.

6.3.1 Investigating estimator performance under asymmetric conditions

The following analysis of estimator selection under asymmetric conditions involves two numerical examples to motivate and illustrate the concepts discussed in this chapter. The first consists of a hypothetical example examined via simulation, whereas the second is more of a case study using actual industrial data drawn from the literature as a basis for additional simulation. The results from both are used to draw conclusions regarding the effects of asymmetry on estimator performances.

6.3.1.1 Hypothetical example via simulation

A Monte Carlo simulation was constructed using the R software package. At each iterate, two central composite designs (*CCD*) are constructed, one using normal data and the other using skew normal data with a common mean. This approach allows for a more direction comparison between estimator performances in the normal case versus the asymmetric case, thereby facilitating an examination of how the presence of skewness affects the estimators in an *RPD* context. Each *CCD* consists of 18 total design points with 100 replicates for the response at each point.

Replicates at each design point are obtained by generating two sets of random variates from both normal and skew normal distributions. In doing so, an existing asymmetric data set consisting of replicated responses at each design point serves a basis from which to derive the first three sample moments for each design point (which, from Equation (6.1) for the *q*th design point, are the mean ($\overline{y}_q$), standard deviation (s_q), and skewness ($\hat{k}_q$)). To generate normally distributed data, the first two sample moments at the *q*th design point are used to create 100 random variates from a $N(\overline{y}_q, s_q)$ distribution. The generation of skewed data in R requires inputs for the location, scale, and shape parameters. Moreover, the location and scale of a skew normal distribution are not equivalent to the mean and standard deviation as they are in the normal case (unless the shape parameter $\alpha = 0$). Thus, the skew normal approach delineated in Section 6.2.1.2 is applied to generate the appropriate estimates for the mean and standard deviation from the skewed distribution that more accurately account for inherent process asymmetry. Thereafter, response surface functions are developed and the *MSE*-based optimization scheme in Equation (6.11) is applied to obtain optimum operating conditions. Table 6.6 depicts a generalized version of the *CCD* construct used.

A total of 1,500 iterations of the simulation were performed using freshly generated random variates for each *CCD* at each iterate. As the objective of robust parameter design is to determine the optimum operating conditions that facilitate target achievement with minimal bias and variability, comparisons are made using the bias and mean squared error resulting from the applied optimization scheme.

Table 6.6. Central composite design (*CCD*) comprised of eight corner points, six axial points, and four center points, used at each iterate.

Run	X_1	X_2	X_3	*Replications (newly generated at each iteration)*	$\bar{y}$, $\tilde{y}$, $\tilde{y}_{H2}$, $\tilde{y}_{HL}$, $\tilde{y}_{\tau}$	s, MAD, $s_{H2,}$ S_n, Q_n, s_{τ}
1	-1	-1	-1	$y_{1,1}$ ….. $y_{1,100}$		
2	1	-1	-1	$Y_{2,1}$ …. $y_{2,100}$		
3	-1	1	-1	$Y_{3,1}$ …. $y_{3,100}$		
4	1	1	-1			
5	-1	-1	1			
6	1	-1	1			
7	-1	1	1			
8	1	1	1			
9	-1.68	0	0		Location Estimates	Scale Estimates
10	1.68	0	0			
11	0	-1.68	0			
12	0	1.68	0			
13	0	0	-1.68			
14	0	0	1.68			
15	0	0	0			
16	0	0	0			
17	0	0	0			
18	0	0	0	$Y_{18,1}$ …. $y_{18,100}$		

Assuming a customer-specified process target of $t_0 = 350$, Table 6.7 shows the estimated bias and *MSE* of the optimal mean response $\hat{m}(\mathbf{x^*})$. The mean squared error (*MSE*) for each model is also included for comparison purposes.

Table 6.7. Simulation results showing the average *MSE* and bias of the optimal mean response and the average squared residual error for each model across 1,500 iterations.

		Model A	**Model B**	**Model C**	**Model D**	**Model E**	**Model F**	**Model G**
		$\bar{y}$ s	y_{median} MAD	Huber 'Proposal 2'	HL_n S_n	HL_n Q_n	HL_n MAD	Tau_m Tau_s
Normal	MSE of $m(\mathbf{x^*})$	1541.433	1745.370	1597.847	1640.447	1573.517	1703.918	1573.214
	Bias of $m(\mathbf{x^*})$	5.862	7.244	5.935	5.998	5.885	6.284	6.362
	Model Error	761.195	821.726	766.193	767.326	767.326	767.326	797.567
Skew Normal	MSE of $m(\mathbf{x^*})$	665.009	641.212	586.543	602.040	592.576	620.946	565.565
	Bias of $m(\mathbf{x^*})$	8.592	7.167	7.910	7.635	7.503	7.410	7.318
	Model Error	847.503	869.485	838.530	839.215	839.215	839.215	852.689

From the results in Table 6.7, two observations are immediately clear. First, as expected, under normal conditions, the sample mean and standard deviation outperform

the other models in all three comparison categories. It is noteworthy that the median/*MAD* combination in Model B performed the worst. Given the previous discussion concerning efficiencies of these two estimators under normality, this result is consistent with expectations. Similarly, recognizing the improved efficiencies of the other estimators in Models C-G relative to those in Model A, the results reaffirm the idea that these estimators will perform nearly as well as the sample mean and standard deviation under normal conditions.

Second, when asymmetric conditions prevail, it is clear that Model A performs considerably worse in the first two comparison categories relative to all other models considered. Interestingly, while the median/*MAD* combination (Model B) does not result in the smallest *MSE* for the optimal response on average, it does tend to produce the least bias among the alternative models. Moreover, it is clear from an examination of the results from all 1,500 iterations (not shown) that no one model is the best, although the tau estimators in Model G tended to yield the best results in both optimization criteria categories. This suggests that tradeoff analyses could be performed between bias and variability to narrow the estimator selection field even further. In order to facilitate a more transparent comparison between models in terms of *MSE* and bias, we graphically portray the results in Figure 6.2.

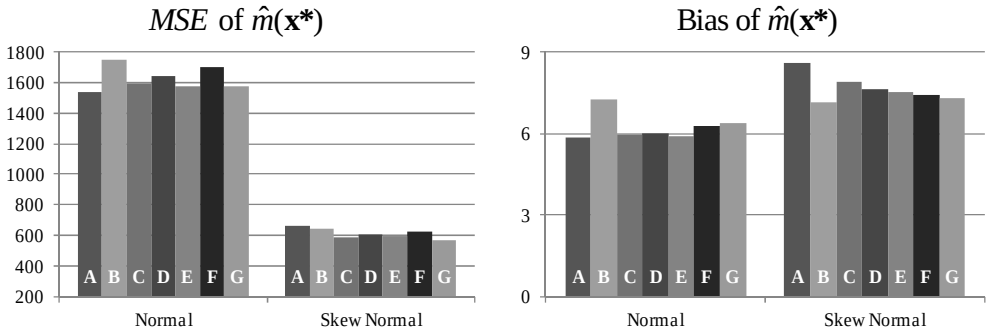


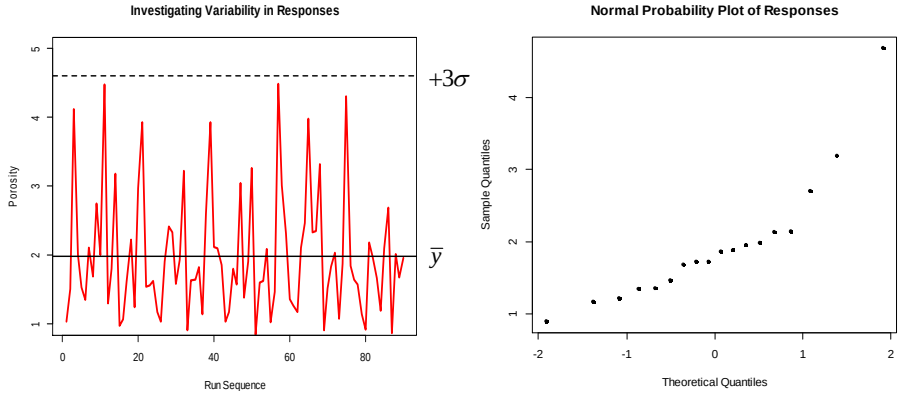
Figure 6.2. Histograms facilitating visual comparison of estimated *MSE* and bias of the optimal mean response for all models considered under normal and skew normal conditions. Model associations are shown across the base of each histogram.

6.3.1.2. Case Study – A Ceramic Coating Process

Tillman *et al.* (2010) used experimental design to determine the optimal settings for a carbide coating process designed to strengthen various ceramic materials. The coating mechanism consisted of a high-velocity oxygen fuel spray technique, which generated different results depending on the kerosene level (X_1), oxygen level (X_2), and stand-off distance (X_3) used. To maintain adequate levels of resistivity in the fiber, they specifically sought the factor settings that produced optimal responses with respect to the material hardness (Y_1) and the porosity (Y_2). As we are concerned about the univariate case in this particular paper, we will only address the latter. Porosity is an *S*-type characteristic with an upper acceptable threshold of 3% and an associated objective of minimizing as close to 0% as possible. Note that this could also be treated as an *N*-type characteristic with lower and upper limits of 0% and 3%, respectively, and a desired target, τ=0.

Since Tillman *et al.* were concerned solely with the mean response, an adaptation of the experiment developed by Goethals and Cho (2011d) that includes replication is

used to facilitate estimates for the variability and skewness of the responses as well. Pursuant to the methodology described in Section 6.2.1.1, a graphical analysis of the data is performed to investigate the assumptions regarding normality and variability in the responses. While Figure 6.3(a) indicates a reasonable, albeit elevated degree of variability in the responses, the curved plot in Figure 6.3 (b) clearly suggests a fair degree of asymmetry. Note that the spacing between the points in the plot in Figure 6.3(b), which appears to widen moving from left to right, reinforces the elevated variability inferred from Figure 6.3(a).



(a) (b)

Figure 6.3. Graphical investigation of variability (a) and normality (b) in the responses.

Application of the Shapiro-Wilk test yields $W^* = 0.8186$ vs. $W_\alpha = 0.946$ which supports the alternative hypothesis H_1 (that sufficient evidence exists to suggest non-normality), and further reinforces the graphical analysis.

Following the approach outlined in the previous section, the first three sample moments (mean ($\bar{y}$), standard deviation (s), and skew (k)) are then used to obtain estimates for the location, scale, and shape parameters of the skew normal distribution. These estimates, in turn, are used to collect ten replicates at each design point. Table 6.8

shows the central composite design used for this experiment, including the calculations for the sample moments and the generated replicates at each design point.

Table 6.8. Adapted experimental design for the ceramic coating process

Coded Units				Sample Moments used to estimate (ξ,ω,α) parameters			Randomly Generated Replicates for Porosity (Y, %) from $SN(\xi,\omega,\alpha)$									
Run	X_1	X_2	X_3	$\bar{y}$	s	k	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
1	–1	–1	–1	1.22	0.19	0.36	1.04	1.25	1.14	1.02	1.08	1.17	1.35	1.11	1.10	1.19
2	–1	–1	1	1.73	0.42	0.58	1.51	2.97	2.60	1.47	1.88	1.70	1.25	1.84	1.94	1.73
3	–1	1	–1	4.68	0.39	–0.67	4.12	3.93	3.93	4.49	4.31	4.50	4.38	4.60	4.66	3.99
4	–1	1	1	2.14	0.51	0.59	2.00	1.54	2.12	3.02	1.85	2.41	2.25	2.13	1.88	2.61
5	1	–1	–1	1.96	0.42	0.58	1.53	1.56	2.10	2.32	1.65	2.29	1.48	2.32	2.66	1.81
6	1	–1	1	1.73	0.47	0.66	1.35	1.63	1.86	1.36	1.57	1.37	1.86	1.61	2.14	1.41
7	1	1	–1	1.35	0.34	0.67	2.11	1.17	1.04	1.27	1.14	1.00	1.23	2.34	1.08	1.73
8	1	1	1	1.17	0.24	0.37	1.69	1.04	1.17	1.17	0.92	1.64	1.02	1.00	1.53	0.90
9	–2	0	0	1.99	0.45	0.52	2.75	1.91	1.80	2.11	2.18	1.81	2.19	1.67	2.11	1.85
10	2	0	0	1.89	0.46	0.27	2.00	2.41	1.57	2.47	1.97	2.06	1.61	1.87	1.57	1.84
11	0	–2	0	3.19	1.05	–0.53	4.48	2.33	3.05	3.98	1.67	4.13	1.67	3.19	4.18	3.14
12	0	2	0	1.47	0.43	0.44	1.30	1.58	1.38	2.33	1.19	1.13	1.40	1.55	1.21	1.27
13	0	0	–2	1.87	0.21	0.17	1.80	1.93	1.91	2.35	2.10	1.93	2.43	1.97	2.17	1.95
14	0	0	2	2.70	0.58	–0.71	3.18	3.23	3.27	3.32	2.69	3.04	2.98	2.46	2.32	2.35
15	0	0	0	0.90	0.06	0.53	0.97	0.91	0.85	0.91	0.87	0.87	0.97	1.01	0.96	0.89
16	0	0	0	1.36	0.43	0.25	1.07	1.64	1.60	1.52	2.01	1.13	1.30	1.08	1.63	1.84
17	0	0	0	1.69	0.24	0.37	1.61	1.65	1.63	1.84	1.68	1.76	1.65	1.55	1.51	1.82
18	0	0	0	2.15	0.51	0.70	2.23	1.83	2.09	2.04	1.96	2.14	2.20	1.63	2.15	1.84

The replicates $y_{q1}, ..., y_{q10}$ are then used to compute estimates for the sample mean and standard deviation ($\bar{y}_q$ and s_q), the median and the *MAD* (y_{median} and *MAD*), Huber's Proposal 2 estimates for location and scale ($y_{q,H2}$ and $s_{q,H2}$), the Hodges-Lehmann location estimator ($y_{q,HL}$), the $S_{q,n}$ and $Q_{q,n}$ scale estimators, and the tau estimators for location and scale ($y_{q,\tau}$ and $s_{q,\tau}$) for the qth design point; Table 6.9 contains the estimates obtained in R.

Using the same seven models previously defined for estimator combinations, full second order polynomial response surface models were developed for each model, followed in turn by the application of the *MSE*-based optimization scheme (Equation (6.5)) to identify the optimum operating conditions. Table 6.10 provides a summary of

the results for each model, including the estimated regression coefficients of the mean response ($\mu(\mathbf{x})$), the optimal factor settings ($\mathbf{x}^*$), and the optimal mean response ($\hat{\mu}(\mathbf{x}^*)$), standard deviation ($\hat{\sigma}(\mathbf{x}^*)$), and *MSE*. Note that, given $\tau = 0$, the value for $\hat{\mu}(\mathbf{x}^*)$ also denotes the corresponding bias.

Table 6.9. Estimates generated using the various estimators in Table 6.3.

$\bar{y}$	s	y_{median}	*MAD*	y_{H2}	s_{H2}	y_{HL}	S_n	Q_n	y_τ	s_τ
1.15	0.10	1.12	0.08	1.14	0.10	1.14	0.10	0.10	1.12	0.09
1.89	0.52	1.78	0.32	1.85	0.51	1.80	0.43	0.38	1.71	0.37
4.29	0.28	4.34	0.36	4.29	0.32	4.29	0.35	0.30	4.31	0.28
2.18	0.42	2.13	0.39	2.16	0.43	2.13	0.41	0.44	2.12	0.40
1.97	0.42	1.96	0.54	1.97	0.47	1.94	0.61	0.44	1.96	0.41
1.62	0.27	1.59	0.33	1.60	0.27	1.61	0.30	0.35	1.59	0.26
1.41	0.48	1.20	0.20	1.39	0.49	1.27	0.23	0.25	1.15	0.25
1.21	0.30	1.10	0.20	1.21	0.34	1.17	0.29	0.22	1.06	0.25
2.04	0.31	2.01	0.26	2.00	0.26	2.00	0.37	0.28	1.97	0.25
1.94	0.32	1.92	0.34	1.94	0.36	1.94	0.43	0.37	1.91	0.32
3.18	1.03	3.16	1.33	3.18	1.17	3.16	1.18	1.29	3.22	1.02
1.43	0.35	1.34	0.21	1.37	0.21	1.37	0.21	0.23	1.33	0.20
2.05	0.20	1.96	0.14	2.05	0.22	2.03	0.17	0.21	1.96	0.16
2.88	0.39	3.01	0.42	2.88	0.45	2.86	0.35	0.36	2.96	0.40
0.92	0.05	0.91	0.07	0.92	0.06	0.92	0.06	0.06	0.92	0.05
1.48	0.33	1.56	0.40	1.48	0.37	1.47	0.40	0.35	1.50	0.32
1.67	0.11	1.65	0.10	1.67	0.12	1.66	0.12	0.12	1.66	0.11
2.01	0.19	2.06	0.18	2.02	0.20	2.02	0.20	0.19	2.06	0.18

Per the results in Table 6.10, Model B outperforms the other models in terms of both the bias and *MSE* of the optimization result. It is also noteworthy that, once again, the models using the HL_n location estimator and the tau estimators (Models D-G) follow in very close succession to Model B, further suggesting the superiority of median-based estimators under asymmetric conditions.

To verify these findings, we performed 500 iterations of the experiment using newly generated random data from the same $SN(\xi,\omega,\alpha)$ distribution at each iteration. Table 6.11 contains the average performance measures across the 500 iterations for the

MSE and bias (relative to $\tau = 0$). The results reinforce our deductions, demonstrating the superior performance of the median-based estimators, on average.

Table 6.10. Summary of optimization results for the ceramic coating process.

	Model A	**Model B**	**Model C**	**Model D**	**Model E**	**Model F**	**Model G**
β_0	1.537	1.564	1.539	1.536	1.536	1.536	1.551
β_1	-0.254	-0.270	-0.247	-0.254	-0.254	-0.254	-0.264
β_2	-0.035	-0.055	-0.042	-0.046	-0.046	-0.046	-0.067
β_3	-0.039	-0.019	-0.041	-0.039	-0.039	-0.039	-0.028
β_{11}	0.092	0.067	0.087	0.083	0.083	0.083	0.060
β_{22}	0.206	0.169	0.195	0.187	0.187	0.187	0.178
β_{33}	0.263	0.252	0.262	0.253	0.253	0.253	0.243
β_{12}	-0.551	-0.601	-0.555	-0.575	-0.575	-0.575	-0.617
β_{13}	0.101	0.136	0.108	0.133	0.133	0.133	0.143
β_{23}	-0.337	-0.326	-0.333	-0.324	-0.324	-0.324	-0.312
x_1	1.682	1.682	1.682	1.682	1.682	1.682	1.682
x* x_2	1.670	1.577	1.660	1.682	1.681	1.682	1.427
x_3	1.309	1.154	1.237	1.167	1.117	1.000	0.878
$\hat{\mu}(\mathbf{x}^*)$	0.222	0.024	0.179	0.093	0.083	0.062	0.045
$\hat{\sigma}(\mathbf{x}^*)$	0.377	0.096	0.358	0.233	0.205	0.132	0.202
MSE	0.191	0.010	0.160	0.063	0.049	0.021	0.043

Table 6.11. Average results across 500 iterations of simulated replications.

Model	**A**	**B**	**C**	**D**	**E**	**F**	**G**
	$\bar{y}$ s	y_{median} MAD	Huber 'Proposal 2'	HL_n S_n	HL_n Q_n	HL_n MAD	Tau_m Tau_s
Avg *MSE* for $\hat{\mu}(\mathbf{x}^*)$	0.094	0.076	0.095	0.086	0.087	0.082	0.069
Avg Bias for $\hat{\mu}(\mathbf{x}^*)$	0.162	0.130	0.156	0.147	0.145	0.145	0.123
Proportion of Runs Best MSE Achieved	24.70%	20.80%	9.90%	13.00%	9.90%	9.40%	12.30%

6.3.2 Investigating estimator performance under high-variability conditions

As in the previous section, the numerical example used to illustrate the impacts of high-variability conditions on estimator performance is approached in two parts. In the first, a pilot study using highly variable data from an actual industrial process is used to motivate the concepts and suggestions heretofore discussed. The second serves to

expand the analysis via simulation in order to facilitate the broader comparisons between estimators under the stated conditions.

6.3.2.1 Part 1 – Pilot Study Using the Box and Draper (1987) Printing Press Study

Perhaps one of the most widely used data sets for comparing *RPD* optimization schemes is the printing press study data introduced by Box and Draper (1987) and discussed in Chapter 3. In this experiment, the printing machine's index in applying coloring inks to package labels is a normally distributed *N*-type quality characteristic of interest, *Y*. The control factors known to influence *Y* include the speed (X_1), pressure (X_2), and distance (X_3) settings; the desired target value for the machine's index is $\tau = 500$. The experiment considers the control factors at three levels and consists of three replicates at each of the twenty-seven design points of the full factorial design. Table 6.12 displays the original printing press data, including the calculations for the mean and standard deviation at each design point. It should be noted that, in order to evaluate the performance of the *tau* estimators without incurring a value error in R, minor adjustments were required in the data. Specifically, in cases where two or more observations within a particular design point assumed the same value, we adjusted one down by 0.01 and another up by 0.01. The design points requiring these modifications are indicated by those having real numbers vice integer values in the y_1, y_2, and y_3 columns.

Using this example, previous researchers developed second-order response surface designs for the mean and standard deviation and then applied their particular optimization scheme to find optimal solutions $\mathbf{x}^* = (X_1^*, X_2^*, X_3^*)$. The preliminary

estimates for the sample mean and standard deviation shown for each design point in Table 6.12 allow for the assessment of high variability in the system.

Table 6.12. Experimental design for the printing press study.

	Coded Units			Labeling Index (Y) (3 replications)				Standard
	Speed	Pressure	Distance				Mean	Deviation
Run	X_1	X_2	X_3	y_1	y_2	y_3	$\bar{y}$	s
1	-1	-1	-1	34	10	28	24	12.49
2	0	-1	-1	115	116	130	120.33	8.39
3	1	-1	-1	192	186	263	213.67	42.83
4	-1	0	-1	82	87.5	88.5	86.00	3.46
5	0	0	-1	44	187.99	188.01	140	83.14
6	1	0	-1	322	349.99	350.01	340.67	16.17
7	-1	1	-1	141	110	86	112.33	27.57
8	0	1	-1	258.99	251	259.01	256.33	4.62
9	1	1	-1	290	280	245	271.67	23.63
10	-1	-1	0	81	80.99	81.01	81	0
11	0	-1	0	90	122	93	101.67	17.67
12	1	-1	0	319	375.99	376.01	357	32.91
13	-1	0	0	179.99	180.01	154	171.33	15.01
14	0	0	0	372	371.99	372.01	372	0
15	1	0	0	541	568	396	501.67	92.50
16	-1	1	0	288	192	312	264	63.50
17	0	1	0	432	336	513	427	88.61
18	1	1	0	713	725	754	730.67	21.08
19	-1	-1	1	364	99	199	220.67	133.82
20	0	-1	1	232	221	266	239.67	23.46
21	1	-1	1	408	415	443	422	18.52
22	-1	0	1	181.99	233	182.01	199	29.44
23	0	0	1	507	515	434	485.33	44.64
24	1	0	1	846	535	640	673.67	158.21
25	-1	1	1	236	126	168	176.67	55.51
26	0	1	1	660	440	403	501	138.94
27	1	1	1	878	991	1161	1010	142.45

This idea reveals itself in two ways. First, casual inspection of the standard deviations in the right-most column highlights a number of fairly large values, pointing to elevated variability within the design points. Second, observe the average and standard deviation of the means themselves, noting in particular the relatively large value of the latter measure, which points to considerable variability between the design points. It should be noted in the latter case that, although each design point technically denotes a unique population due to the different control factor settings, we would still reasonably expect

only a moderate degree of variability within the response vector. As Figure 6.4 shows below, the process associated with the print press study qualifies as a highly variable process based on the first aspect of the definition provided in Section 4.

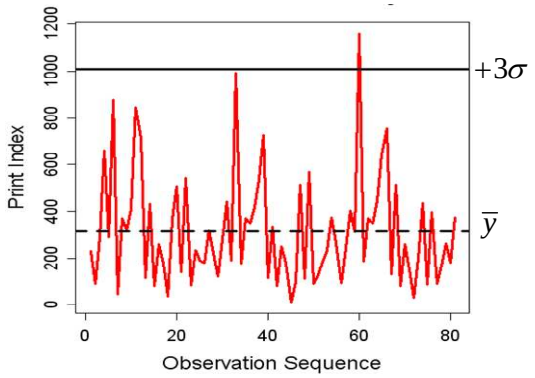


Figure 6.4. Variability of Observations in the Printing Press Study.

The three replicates at each design point are used to compute estimates using each of the various location and scale estimators. Table 6.13 contains the estimates obtained via R. Using the eight models defined for estimator combinations, full second order polynomial response surface models are developed for each and then identified the optimum operating conditions by applying the *MSE*-based optimization scheme (Equation 6.11) on a cuboidal region of interest in which each of the factor settings can assume a value on the range of (-1, 1). Table 6.14 provides a summary of the results for each model, including the estimated coefficients of the mean response ($\hat{m}(\mathbf{x})$), the optimal factor settings ($\mathbf{x}^*$), the optimal mean response $\hat{m}(\mathbf{x}^*)$, standard deviation $\hat{s}(\mathbf{x}^*)$, and *MSE*.

Table 6.13. Estimates generated using the various estimators in Table 6.3.

$\bar{y}$	s	y_{median}	MAD	y_{H2}	s_{H2}	y_{HL}	S_n	Q_n	y_τ	s_τ	$\bar{y}_{MLE}$	s_{MLE}
24.00	12.49	28.00	8.90	24.00	14.16	25.00	13.25	13.23	27.94	11.37	22.85	8.45
120.33	8.39	116.00	1.48	120.33	9.51	119.25	2.21	2.21	115.53	1.85	123.99	0.00
213.67	42.83	192.00	8.90	213.67	48.54	208.25	13.25	13.23	189.15	11.10	220.20	26.00
86.00	3.50	87.50	1.48	86.00	3.97	86.38	2.21	2.21	87.97	1.85	85.49	2.29
140.00	83.14	187.99	0.03	140.00	94.23	152.00	0.04	0.04	188.00	0.04	125.07	47.70
340.67	16.17	349.99	0.03	340.67	18.32	343.00	0.04	0.04	350.00	0.04	337.95	9.17
112.33	27.57	110.00	35.58	112.33	31.25	111.75	52.98	52.92	111.61	23.42	113.54	20.88
256.33	4.62	258.99	0.03	256.33	5.23	257.00	0.04	0.04	259.00	0.04	255.35	2.78
271.67	23.63	280.00	14.83	271.67	26.78	273.75	22.08	22.05	281.74	18.71	269.13	15.83
81.00	0.01	81.00	0.01	81.00	0.01	81.00	0.02	0.02	81.00	0.01	81.00	0.01
101.67	17.67	93.00	4.45	101.67	20.03	99.50	6.62	6.62	91.58	5.55	104.10	10.72
357.00	32.91	375.99	0.03	357.00	37.30	361.75	0.04	0.04	376.00	0.04	351.28	18.80
171.33	15.01	179.99	0.03	171.33	17.01	173.50	0.04	0.04	180.00	0.04	168.95	8.40
372.00	0.01	372.00	0.01	372.00	0.01	372.00	0.02	0.02	372.00	0.01	372.00	0.01
501.67	92.50	541.00	40.03	501.67	104.84	511.50	59.60	59.54	553.82	49.97	489.68	59.35
264.00	63.50	288.00	35.58	264.00	71.97	270.00	52.98	52.92	296.96	44.49	256.68	41.98
427.00	88.61	432.00	120.09	427.00	100.43	428.25	178.81	178.61	428.41	75.25	427.00	72.35
730.67	21.08	725.00	17.79	730.67	23.89	729.25	26.49	26.46	726.59	18.39	732.42	14.73
220.67	133.82	199.00	148.26	220.67	151.67	215.25	220.75	220.50	211.53	114.02	228.70	98.57
239.67	23.46	232.00	16.31	239.67	26.59	237.75	24.28	24.26	231.79	20.85	242.09	16.04
422.00	18.52	415.00	10.38	422.00	20.99	420.25	15.45	15.44	412.39	12.98	424.06	12.18
199.00	29.44	182.01	0.03	199.00	33.37	194.75	0.04	0.04	182.00	0.04	204.07	16.77
485.33	44.64	507.00	11.86	485.33	50.59	490.75	17.66	17.64	510.80	14.81	478.83	27.46
673.67	158.21	640.00	155.67	673.67	179.31	665.25	231.79	231.53	655.72	135.62	685.02	113.79
176.67	55.51	168.00	62.27	176.67	62.91	174.50	92.72	92.61	173.10	47.28	179.90	40.96
501.00	138.94	440.00	54.86	501.00	157.47	485.75	81.68	81.59	422.44	68.47	519.57	88.40
1010.00	142.45	991.00	167.53	1010.00	161.46	1005.25	249.45	249.17	1002.89	121.18	1017.65	106.12

Table 6.14. Results of pilot study using print press data.

	Model A	**Model B**	**Model C**	**Model D**	**Model E**	**Model F**	**Model G**	**Model H**
β_0	328.123	347.350	329.506	332.930	332.930	351.076	332.930	322.285
β_1	177.000	177.026	176.519	177.007	177.007	175.177	177.007	177.013
β_2	109.426	108.945	109.574	109.306	109.306	105.235	109.306	109.631
β_3	131.278	120.196	131.130	128.507	128.507	125.740	128.507	134.773
β_{11}	31.630	30.862	31.815	31.438	31.438	47.085	31.438	31.766
β_{22}	-22.759	-37.222	-23.574	-26.375	-26.375	-48.809	-26.375	-18.010
β_{33}	-28.870	-43.972	-29.685	-32.646	-32.646	-53.150	-32.646	-24.020
β_{12}	66.028	62.918	66.028	65.250	65.250	64.373	65.250	67.174
β_{13}	75.472	75.042	75.472	75.365	75.365	75.916	75.365	75.721
β_{23}	43.583	36.667	43.361	41.854	41.854	41.216	41.854	45.907
x_1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\mathbf{x}^*$ x_2	0.060	-0.097	0.037	-0.079	-0.079	-0.135	-0.071	0.016
x_3	-0.243	-0.203	-0.236	-0.165	-0.165	-0.247	-0.155	-0.172
$\hat{\mu}(\mathbf{x}^*)$	494.657	497.473	493.547	493.478	493.491	497.942	496.916	496.964
Bias	5.343	2.527	6.453	6.522	6.509	2.058	3.084	3.036
$\hat{\sigma}(\mathbf{x}^*)$	44.596	14.539	48.122	24.570	24.547	15.958	17.118	29.548
MSE	2017.325	217.774	2357.385	646.243	644.901	258.885	302.541	882.277

Overall, the results from this singular instance suggest that considerable improvements stand to be gained by using the estimator combinations in Models B or G. Not only to they produce results with less deviation from the desired target, but they yield marked reductions in variability of the optimal mean response in terms of the *MSE*. The regression coefficients and optimal factor settings, **x***, are included to show the subtle differences in the results obtained for each model. Note that the differences in the optimal factor settings (coded units) apply only to the x_2 and x_3 factors and that these differences are measured, for the most part, in hundredths of a point. Notwithstanding, the differences are enough to induce added precision that results in significant reductions in bias and *MSE*.

It is important to note here that, recognizing the inherent elevated variability in the print press study, it is quite likely that the results obtained Table 6.14 are simply the circumstance of the data used. That is, under similar conditions with a different set of data, the results could be different in terms of which models perform best. Thus, to verify our findings, and to more clearly determine the impacts of high versus low levels of variability on estimator selection, we performed Monte Carlo simulations to facilitate comparisons of estimator performances in obtaining robust design solutions. The next section describes this process and the results.

6.3.2.2 Part 2 – Simulation to identify trends

The purpose of the simulation was to facilitate a broader comparison between estimator performances in both high-variability and low-variability situations that would serve to verify the inferences derived from the case study. This would, in turn, facilitate

more definitive conclusions regarding the impact of variability conditions on estimator selection by allowing us to analyze trends over multiple iterations of simulated conditions.

The simulation was constructed using R. At each iterate, two full-factorial designs (*FFD*) of three factors at three levels each are created using the original print press data as a source data from which to generate normally distributed random variates (using the `rnorm()` function in R). The first of these consists of a high-variability design in which the original sample means and standard deviations at each design point are used to generate normal random variates. The second is a low-variability design that is created in the same way except that the standard deviations are multiplied by 0.25 to generate normal random variates with 75% less variability.

The simulation experiment examined 15 different combinations of design point replicates and simulation iterations for both the high- and low-variability cases using freshly generated random data for each simulation run. This would enable analysis of the impacts of increases in observations per design point and the number of iterations on estimator performance trends. Thus, the number of random variates generated in each case depended on the number of replicates, *m*, being examined in a particular instance of the experiment. In every case, however, the *m* replicates within each of the design points in both *FFD*s are then used to obtain estimates using each of the location and scale estimators under consideration. Table 6.15 depicts a generalized version of the *FFD* framework used, consisting of 27 total design points. The experimental framework for the simulation is portrayed in Figure 6.5.

Table 6.15. Full factorial design (*FFD*) framework used in the pilot study.

Run	X_1	X_2	X_3	*Replications (newly generated at each iteration)*	$\bar{y}, y_{median}, y_{H2}, y_{HL}, y_{\tau}, \bar{y}_{MLE}$	s, MAD, s_{H2}, S_n, Q_n, s_τ, s_{MLE}
1	-1	-1	-1	$y_{1,1}$ ….. $y_{1,m}$		
2	0	-1	-1	$Y_{2,1}$ …. $y_{2,m}$		
3	1	-1	-1	$Y_{3,1}$ …. $y_{3,m}$		
4	-1	0	-1			
5	0	0	-1			
6	1	0	-1			
7	-1	1	-1			
8	0	1	-1			
9	1	1	-1		Location Estimates	Scale Estimates
10	-1	-1	0			
11	0	-1	0			
12	1	-1	0			
13	-1	0	0			
14	0	0	0			
15	1	0	0			
16	-1	1	0			
17	0	1	0			
18	1	1	0			
19	-1	-1	1			
20	0	-1	1			
21	1	-1	1			
22	-1	0	1			
23	0	0	1			
24	1	0	1			
25	-1	1	1			
26	0	1	1			
27	1	1	1	$Y_{27,1}$ …. $Y_{27,m}$		

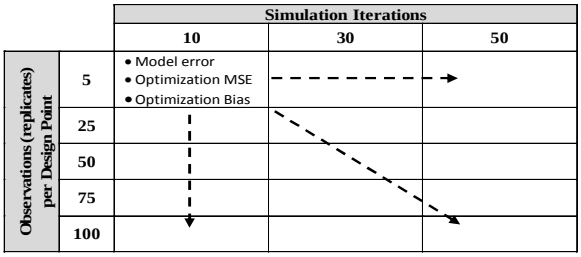


Figure 6.5. Experimental framework used at both levels of variability in the simulation.

Thereafter, response surface designs are developed and the *MSE*-based optimization scheme is applied using a target value of $\tau = 500$ to obtain optimum operating conditions in each iteration. The results for model error, optimization *MSE*, and optimization bias are tallied across the various numbers of iterations and then averaged to yield trends for each experimental combination of design point replicates and simulation iterations. Figure 6.6 contains the output for both the high and low variability scenarios.

# Sim Reps	Obs/ DP	Perf. Measure	High Variability Case Model A	B	C	D	E	F	G	H	Low Variability Case Model A	B	C	D	E	F	G	H
10	5	Mod Err	6877.10	7413.49	6895.77	6881.11	6881.11	7673.95	6881.11	6936.28	6618.43	6582.23	6621.51	6589.56	6589.56	7089.07	6589.56	6655.00
		MSE	426.26	283.87	555.73	525.33	560.69	432.72	288.52	310.30	60.52	37.36	74.52	73.88	69.72	59.53	36.96	41.00
		Bias	2.33	1.56	3.06	3.23	3.40	2.19	1.59	1.75	0.21	0.16	0.27	0.27	0.29	0.23	0.16	0.18
	25	Mod Err	6483.86	6729.93	6500.82	6510.58	6510.58	7134.15	6510.58	6458.81	6517.25	6551.62	6524.42	6523.91	6523.91	7021.18	6523.91	6523.44
		MSE	859.47	722.40	835.78	898.60	889.74	1275.75	715.40	829.01	101.99	98.83	108.33	108.03	106.65	143.74	98.77	101.48
		Bias	3.82	3.06	3.74	3.86	3.80	4.45	3.03	3.69	0.42	0.35	0.42	0.41	0.42	0.45	0.36	0.41
	50	Mod Err	6633.30	6744.14	6667.10	6689.27	6689.27	7204.21	6689.27	6645.69	6511.45	6524.10	6518.24	6518.90	6518.90	7008.36	6518.90	6512.71
		MSE	748.23	771.18	787.14	734.84	764.40	1210.25	765.30	727.88	89.86	94.34	92.51	94.31	90.38	141.34	94.43	89.34
		Bias	3.64	3.40	3.71	3.52	3.63	4.56	3.42	3.56	0.41	0.40	0.42	0.42	0.41	0.50	0.40	0.41
	75	Mod Err	6719.83	6699.83	6688.24	6714.98	6714.98	7144.92	6714.98	6729.32	6492.23	6479.89	6485.49	6483.65	6483.65	6955.17	6483.65	6492.72
		MSE	805.94	783.71	837.35	833.06	831.38	1290.24	792.07	795.06	54.32	57.19	55.31	54.18	54.77	84.52	57.26	53.66
		Bias	3.77	3.66	3.88	3.88	3.87	4.97	3.69	3.72	0.24	0.23	0.24	0.24	0.24	0.30	0.23	0.24
	100	Mod Err	6674.22	6743.62	6682.89	6669.01	6669.01	7190.85	6669.01	6674.29	6569.43	6565.17	6568.49	6570.42	6570.42	7045.37	6570.42	6570.68
		MSE	842.11	782.32	853.14	849.67	846.02	1350.19	782.81	836.05	55.20	58.70	57.90	57.34	56.99	89.10	58.75	54.79
		Bias	3.75	3.51	3.71	3.67	3.73	4.95	3.51	3.72	0.25	0.24	0.25	0.25	0.25	0.32	0.24	0.24
30	5	Mod Err	6903.73	7220.55	6904.97	6831.82	6831.82	7658.04	6831.82	7009.19	6714.47	6775.37	6713.72	6740.73	6740.73	7267.67	6740.73	6707.45
		MSE	783.95	415.78	973.62	634.05	691.47	607.85	415.93	505.18	97.07	63.97	105.73	98.98	93.59	81.69	63.49	57.85
		Bias	3.30	1.83	4.10	2.80	3.04	2.30	1.82	2.22	0.37	0.28	0.43	0.41	0.41	0.29	0.28	0.24
	25	Mod Err	6592.70	6720.41	6585.54	6606.83	6606.83	7159.01	6606.83	6577.89	6576.05	6591.43	6580.01	6586.28	6586.28	7076.42	6586.28	6574.49
		MSE	841.69	747.23	870.65	880.22	860.81	1222.63	751.62	817.92	90.88	79.13	99.24	94.61	93.02	135.91	79.36	88.38
		Bias	3.77	3.17	3.87	3.81	3.83	4.37	3.15	3.68	0.42	0.34	0.43	0.41	0.40	0.50	0.34	0.40
	50	Mod Err	6532.35	6552.73	6532.90	6542.17	6542.17	7024.29	6542.17	6528.92	6562.02	6550.89	6559.46	6558.07	6558.07	7026.00	6558.07	6559.45
		MSE	847.96	820.26	837.19	909.01	864.35	1332.92	818.12	830.88	92.82	95.03	92.67	95.11	94.76	146.39	95.16	91.64
		Bias	3.93	3.75	3.93	4.06	3.94	5.11	3.74	3.87	0.42	0.41	0.42	0.42	0.42	0.53	0.41	0.41
	75	Mod Err	6625.73	6634.99	6613.11	6615.71	6615.71	7098.08	6615.71	6623.69	6536.51	6532.27	6535.78	6534.72	6534.73	7003.50	6534.73	6537.18
		MSE	733.16	695.98	734.09	725.43	726.02	1167.89	696.54	726.27	50.21	50.87	50.51	51.98	50.83	80.80	50.92	49.50
		Bias	3.60	3.30	3.54	3.49	3.51	4.65	3.30	3.56	0.24	0.23	0.24	0.24	0.24	0.31	0.24	0.24
	100	Mod Err	6680.17	6713.85	6664.68	6672.82	6672.82	7160.99	6672.82	6677.29	6560.25	6546.28	6556.78	6557.53	6557.53	7024.51	6557.53	6560.13
		MSE	805.31	774.17	812.15	782.61	800.04	1292.56	772.07	799.86	51.80	52.84	53.43	51.95	51.89	83.95	52.92	51.57
		Bias	3.69	3.56	3.68	3.61	3.66	4.86	3.56	3.67	0.24	0.24	0.24	0.24	0.24	0.32	0.24	0.24
50	5	Mod Err	7210.38	7436.24	7215.91	7176.52	7176.52	7983.83	7176.52	7305.01	6589.07	6644.11	6593.26	6593.65	6593.65	7110.51	6593.65	6591.00
		MSE	656.22	351.53	792.90	579.09	551.63	520.68	357.71	439.93	78.32	46.62	96.87	67.54	71.20	69.00	45.05	52.01
		Bias	2.93	1.58	3.58	2.68	2.41	2.18	1.60	1.95	0.29	0.21	0.38	0.31	0.30	0.25	0.21	0.20
	25	Mod Err	6684.14	6745.37	6687.47	6684.16	6684.16	7217.11	6684.16	6696.35	6582.30	6587.25	6582.29	6585.42	6585.42	7069.66	6585.42	6585.20
		MSE	851.07	807.77	864.85	926.15	880.42	1238.81	813.65	836.47	90.70	88.95	95.15	97.82	94.99	137.02	88.27	88.14
		Bias	3.66	3.19	3.67	3.70	3.60	4.26	3.25	3.59	0.39	0.38	0.41	0.42	0.41	0.49	0.38	0.39
	50	Mod Err	6557.69	6638.67	6569.63	6574.12	6574.12	7099.57	6574.12	6564.21	6534.04	6546.62	6534.18	6533.69	6533.69	7009.86	6533.69	6530.69
		MSE	793.20	798.85	820.43	828.14	819.51	1258.27	797.43	789.53	87.42	82.04	86.79	88.30	85.65	138.24	81.95	86.20
		Bias	3.73	3.62	3.79	3.79	3.79	4.84	3.62	3.69	0.40	0.39	0.40	0.41	0.40	0.51	0.39	0.39
	75	Mod Err	6602.21	6611.14	6604.88	6608.39	6608.39	7092.02	6608.39	6594.45	6545.26	6541.48	6544.98	6545.87	6545.87	7017.32	6545.87	6545.92
		MSE	836.84	820.22	851.51	851.31	844.63	1333.38	820.91	830.38	53.57	55.53	55.49	55.31	55.27	85.92	55.53	53.26
		Bias	3.78	3.68	3.84	3.83	3.81	4.89	3.68	3.75	0.24	0.25	0.25	0.25	0.25	0.31	0.25	0.24
	100	Mod Err	6607.08	6666.27	6619.77	6613.28	6613.28	7130.93	6613.28	6605.88	6555.91	6551.75	6554.58	6555.69	6555.69	7026.37	6555.69	6556.48
		MSE	791.08	757.49	783.56	794.21	789.08	1254.83	755.43	782.91	51.14	51.99	51.31	52.11	51.30	83.15	52.03	50.72
		Bias	3.68	3.60	3.70	3.71	3.70	4.86	3.60	3.65	0.24	0.24	0.24	0.24	0.24	0.31	0.24	0.23

Figure 6.6. Simulation output depicting the average modeling and optimization results across the experimental framework for High and Low Variability cases.

What is immediately clear from the overall results in Figure 6.6 is the superior performance delivered by Models B and G in the vast majority of experimental combinations. In fact, it is not until both the number of design point replicates and the number of simulation iterations are increased that the traditional approach in Model A and the *MLE*-based approach in Model H surpass them. Thus, although the results reflect

tendencies across the simulation runs, they clearly suggest that the degree of variability inherent to the system or process impacts estimator performance in terms of the *RPD* solutions attained. The following paragraphs delve into the results of each variability case more closely.

(i) <u>High variability case</u>: In this case, Models B and G outperform all others in nearly every experimental combination. For fewer observations, the trends separating these models from the others are the most pronounced. Again, given the high variability conditions, extreme observations incurred in smaller sample sizes will create a "perception" of asymmetry and thus cause considerable shift in the sample mean and standard deviation, whereas the median-based location estimators of B and G and the *MAD* will tend to resist such leverage and remain closer to the central tendency of the bulk of the observations. As the number of observations and the number of iterations are increased, the gap does appear to narrow somewhat; notwithstanding, as is made clear in the (100, 50) combination of observations and iterations, the trends clearly identify Models B and G as the preferred set of estimators in highly variable conditions, irrespective of the number of design point replicates obtained or the number of simulation iterations performed.

(ii) <u>Low variability case</u>: Again, at low numbers of observations per design point, Models B and G tend to outperform the others, including the traditional approach in Model A, although only slightly. This makes sense, however, given previous discussions about efficiency at lower sample sizes and estimator breakdown. Additionally, although the chance of extreme values is far less in this case relative to the

high-variability instance, such occurrences can and do happen, which would induce the shift in the sample mean and standard deviation previously discussed.

However, as observations are increased, irrespective of the number of iterations, the traditional approach (Model A), as well as the *MLE* approach in Model H improve considerably, and ultimately surpass the performance of Models B and G in the (100, 50) combination of design point observations and simulation iterations. Again, recognizing the deterioration in relative efficiencies of the median and *MAD* as the sample size increases serves as a partial explanation for this. Acknowledging once again the rarity of "true" normality helps to further explain the superior performance of the *MLE* estimates in Model H over those in Model A.

6.4 Summary of Findings

The findings from the analyses of both conditions are summarized in paragraphs (i) and (ii) below:

(i) <u>Asymmetric conditions</u>: Based upon the tallied optimization results using the hypothesized datasets in Section 6.3.1.1, as well as the results using actual process data in Section 6.3.1.2, it is clear that all of the alternative estimators tend to outperform the traditional sample mean and standard deviation under asymmetric conditions, both in terms of deviation from the target (bias) and variability in the result (*MSE*). We compared our proposed models (Models B-G) with the traditional approach of Model A through the use of Monte Carlo simulations. The simulation results plainly demonstrate that, under asymmetric conditions, estimators incorporating median-based approaches

tend to work best. In particular, the *Tau* estimators for location and scale will tend to produce the best *RPD* results when such conditions are prevalent.

(ii) <u>High-variability conditions</u>: It is clear from the discussion in Section 6.3.2 and the results portrayed in Figure 6.6 that the degree of variability impacts estimator performance in *RPD* problems and therefore should bear influence on the specific estimators selected when such conditions prevail. In low variability conditions, the results demonstrate that the traditional approach will, in the long run, tend to perform best. However, at lower levels of observations (which, we might add, most often tends to be the case due to experimental costs and time constraints), the models that use combinations of the median or HL_n location estimators with the *MAD* scale estimator (models B and G) can yield preferable results. As variability increases, it is clear that estimator preference shifts exclusively to Models B and G, as these tend to produce better *RPD* results in terms of optimization bias and *MSE*.

It is important to note that the results obtained under both sets of conditions denote *tendencies* in performance rather than conclusive evidence supporting the selection of one particular combination of estimators over all others. That is, while a particular combination of estimators may perform best on average, examination of all simulation iterations would show that this combination performed best in most iterations, worst in others, and in the middle of the pack in others still. Thus, there is some variability. In spite of this, though, it remains clear that, over the long run, the alternatives examined here produce better results as noted above.

It is also important to note that full second-order models were used in all of the experiments. In reality, the likelihood that such a model would be appropriate in every case is low. In fact, it could very well be that higher-order models would be required to achieve a sufficient fit in the response surface design, particularly in cases involving large residual errors. Thus, a more in-depth analysis would be necessary to identify the proper subset of terms for modeling the quality characteristic and to then evaluate and compare the results of the model estimation using a variety of well-established criteria. In the end, however, it is unlikely that this would drastically affect the results and may, in fact, provide even greater support for them.

6.5 Concluding remarks

The purpose of this chapter was to investigate alternative approaches to tier-one estimation in order to attain better robust parameter design solutions when inherent process conditions are either asymmetric or highly variable. Pursuant to this, *RPD* models have been developed that outperform traditional approaches in terms of minimizing bias and variability in the optimal mean response while retaining qualities of efficiency and resistance to influential observations. The proposed models (Models B - H) have been cross-analyzed with the traditional approach of Model A through the use of Monte Carlo simulations. The simulation results plainly demonstrate that, under either condition, estimators incorporating median-based approaches tend to work best. The ultimate result is an improved methodology for solving *RPD* problems; one that integrates the comprehensive data analysis methods proposed in Chapter 4 and combines them with a conditions-based selection of tier-one estimators to drive the development of

response surface designs and ultimately optimal *RPD* solutions. Future work in this area should invoke similar analysis to investigate the impacts of other conditions, such as heteroscedasticity, data imbalances, process stability, etc. on tier-one estimator selection. Chapter 6 will extend this work to examine how tier-two (regression) estimation methods might also be affected under such conditions.

In the end, the selection of one estimator or the other will require a trade-off analysis between bias, variability, efficiency, and perhaps robustness. But above all, it will require a firm understanding of the conditions inherent to the process under study. Pyzdek (1995) amplifies this point in the following passage:

> "You'll find that assuming normality hampers your efforts at continuous improvement. If the process distribution is skewed, the optimal setting (or target) will be somewhere other than the center of the engineering tolerance, but you'll never find it if you assume normality. Your quality-improvement plan must begin with a clear understanding of the process and its distribution."

Hoaglin *et al.* (1983) point out that potential targets for location – including the mean, median, and perhaps others not yet mentioned – are all likely to differ depending on what we are trying to measure. In the robust parameter design context, the measurement of interest corresponds to the quality characteristic of interest. Accordingly, this measure, coupled with a base understanding of the conditions inherent in the data, will dictate what our potential targets should/will be, which will, in turn, inform our selection of the appropriate estimator.

CHAPTER SEVEN

EXTENDING THE CONDITIONS-BASED APPROACH TO REGRESSION ESTIMATION IN ROBUST PARAMETER DESIGN

7.1 Introduction

Chapter 6 examined the impacts of asymmetric and high-variability conditions on tier-one estimator performance, demonstrating that the presence of such conditions does indeed matter in terms of the robust design solutions obtained and that certain estimators tend to perform better than others when these conditions prevail. It is important to note that that ordinary least squares (*OLS*) regression was used in the tier-two estimation phase for each of the models examined in Chapter 6 in order to derive response surface designs for the process location and scale. That is, while the tier-one estimators were varied to investigate performance differences between them, the tier-two estimation approach was fixed so as not to confound the comparisons. To enable this experimental approach, it was assumed that residual errors were independent and identically distributed $N(\mu, \sigma^2)$ random variables. That is, that the three assumptions regarding residual errors that underpin *OLS* application held. While this is acceptable, it is also important to recognize that the presence of significant asymmetry and high variability in the responses can also induce heteroscedasticity in the residual errors, which violates one of these assumptions. When this occurs, it can render *OLS* estimates sub-optimal and necessitate remedial action.

Our evolved understanding of contemporary industrial manufacturing processes has made it increasingly apparent that the assumptions of normality and moderate system

variability quite often do not hold in practice. Realistically, industrial processes often exhibit elevated levels of variability, particularly in mass production lines, which can confound many of the modeling assumptions behind the robust parameter design models available in the literature. Furthermore, it is also accepted that perfect normality rarely, if ever, exists and that some asymmetry is not only realistic but practically inevitable – particularly in situations involving smaller-the-better (*S*-type), larger-the-better (*L*-type), and select instances of nominal-is-best (*N*-type) quality characteristics; a condition that could become even more pronounced under conditions of elevated variability, especially when the number of observations or replicates at each design point is small. As noted above, this can induce the added effect of non-constant variance in the residuals.

Again, what has made *OLS* regression the method of choice over the 200 years of its existence is the simplistic and explicit nature by which it could be derived from the data. Consequently, it has become the flagship approach in the vast majority of applied linear statistics texts and robust design literature. In fact, many contemporary quality assurance programs incorporate *OLS* methods to solve a variety of robust design problems as they endeavor to maintain a competitive edge in the industrial market. Pursuant to this, engineers develop linear or non-linear optimization schemes that most appropriately and accurately portray a given process to determine a set of optimal operating conditions that will assure minimal deviation from a specified target with minimum variability. However, if the true underlying process conditions are overlooked or otherwise assumed away, the estimates derived using the traditional least squares approach are likely to be problematic and misleading. Moreover, once applied to

optimization schemes to determine optimal operating conditions, they may very well generate suboptimal solutions and lead to dubious recommendations to decision makers.

As noted in Chapter 2, myriad alternative approaches to regression estimation have evolved over the years. These include generalized linear models (*GLM*) and weighted least squares (*WLS*), as well as an assortment of resistant and/or robust regression techniques such as least trimmed squares (*LTS*), least median of squares (*LMS*), least absolute deviation (*LAD*), *M*-estimation, *M*-*M* estimation, and *S*-estimation, among others. These robust methods in regression evolved to allow engineers to fit equations to the majority of the data in the presence of outliers, without discounting the outliers altogether. This becomes particularly important in asymmetric and high-variability conditions in which extreme observations in the tail(s) become more likely and therefore wield more influence on regression estimates. A selection of these methods will be discussed in greater detail in Section 7.2.2; for a more comprehensive discussion, however, see Venables and Ripley (2002) or Rousseeuw and Leroy (1987)

While many of these methods have seen considerable attention throughout statistics and regression literature, they are less well-known among engineers endeavoring to apply statistical methods in the interests of quality improvement, particularly those with limited statistics backgrounds. In fact, the list of available estimation methods is so long, that many engineers may have questions about with method or approach to use and when. The objectives of this chapter, then, are threefold:

1) Propose a conditions-based approach to the *RPD* problem, whereby data analysis is used to illuminate intrinsic process conditions that will, in turn, inform the selection

of appropriate regression estimation methods for modeling process response surfaces.

2) Integration of the skew normal distribution to facilitate a fuller and more accurate representation of asymmetric system properties via the inclusion of a shape or skewness parameter.
3) Through experimental investigations and simulation, determine which regression approaches tend to perform best in the context of producing the best *RPD* results under the examined conditions.

Pursuant to these objectives, a variety of alternative regression approaches are examined via experimental analysis and simulation to determine which methods will tend to produce the best robust parameter design (*RPD*) solutions when such conditions exist. Section 7.2, delineates the conditions-based methodology for modeling asymmetry and selecting estimation alternatives. In Section 7.3, numerical demonstrations are presented via two case studies and Monte Carlo simulation. Finally, in Section 7.4, the results obtained are summarized and analyzed. The original work associated with this research has been submitted for publication with reference Boylan and Cho (2013b)

7.2 Proposed Modeling and Optimization Procedures

The discussion in Section 3.3.2 of Chapter 3 illustrated that, when actual conditions deviate from the traditional assumptions behind least squares regression (for both responses and residual errors), alternatives to the *OLS* approach may be necessary for estimating the response functions for the process parameters. Given the host of choices, the intrinsic question centers on which estimator(s) to use and when, as it is very likely that different sets of optimum operating conditions will result depending on the estimator used. More specifically, when certain conditions pervade process operations

and outputs, which estimation approaches should be selected in order to ensure the best *RPD* solutions are achieved?

While a variety of conditions could affect estimator selection, this chapter focuses on inherent asymmetry and elevated variability in process outputs. The conditions-based approach incorporated into the proposed *RPD* methodology for tier-one estimation in Chapter 6 is expanded to include tier-two estimation. Thus, the same comprehensive data analysis is used to illuminate intrinsic process conditions that will, in turn, inform the selection of approaches to regression estimation and the development response surface functions for the process mean and variability. The result is a further refined methodology that entails a fuller understanding of the conditions inherent to the process and the associated data. As before, this will facilitate a more realistic and accurate portrayal of the process being investigated and ultimately allow engineers to achieve better and more precise *RPD* solutions. Figure 7.1 shows the refined methodology, where Phase Ib reflects both refinements made to the methodology proposed in Chapter 6 as well as to the traditional *RPD* methodology. The numerical case studies and associated Monte Carlo simulations discussed in Section 7.3 provide clarification as to which estimators should be considered for selection in Step (ii) of Phase Ib as process conditions evolve.

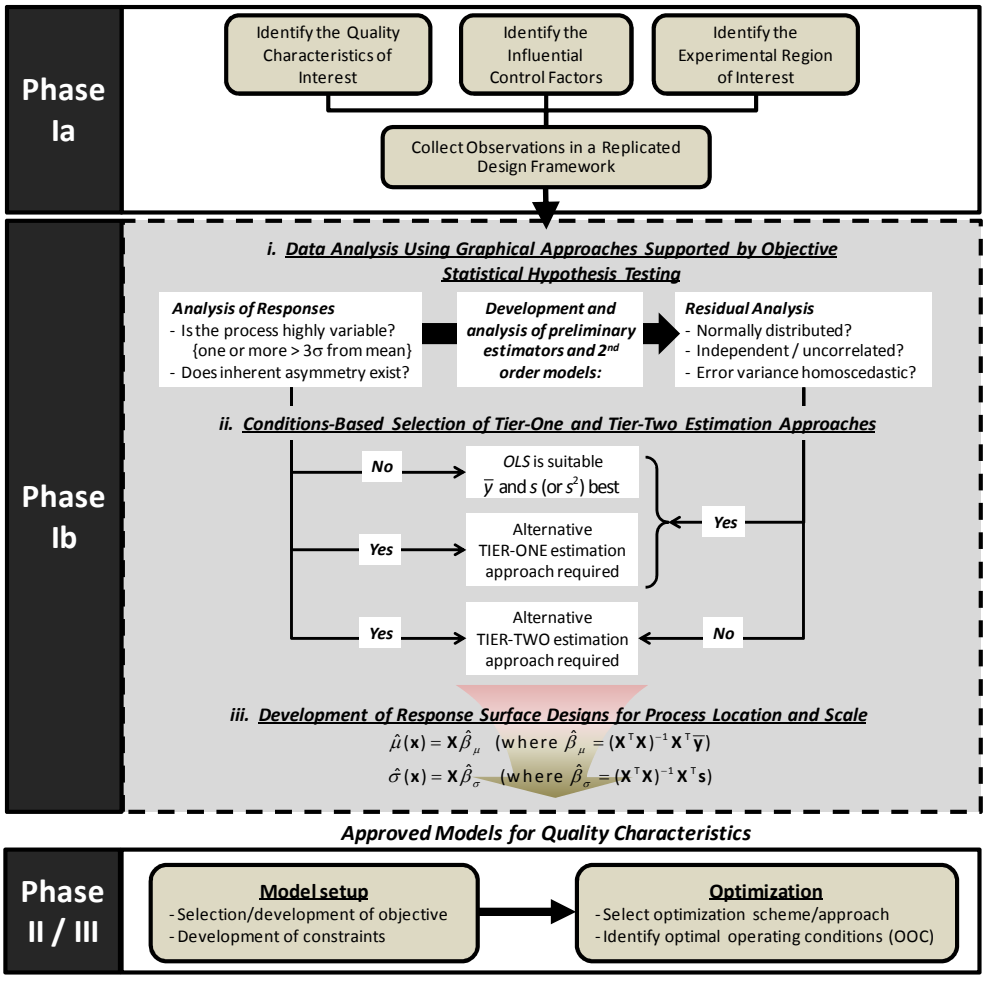


Figure 7.1. Enhanced conditions-based methodology process map.

7.2.1 Experimentation and Analysis

Consider a situation beginning with the specification of response variables, factors or predictors influencing the responses, and region of interest for a designed experiment. Suppose the engineer's objective is to identify the optimal factor settings $\mathbf{x}^*=(X_1^*,X_2^*, ..., X_k^*)$ that support achieving a mean process performance with minimal deviation from a desired target and with minimum variability in the result. As in Chapter 6, an experimental framework involving a set of control factors (X_1, X_2, …, X_k) and a

response, Y, is once again established. The experiment consists of n design points, each of which contains m replicates for the observed response. Let y_{qj} denote the jth response at the qth design point, where $q = 1, \ldots, n$ and $j = 1, \ldots, m$. Table 7.1 portrays the framework for such an experiment.

Table 7.1. Experimental response surface methodology framework.

Design Point	X_1 X_2 .... X_k	Replications	$\bar{y}$	s	γ_3
1	Control Factor Settings	y_{11}……y_{1m}	$\bar{y}_1$	s_1	γ_{31}
⋮		⋮ ⋮	⋮	⋮	⋮
q		y_{q1}……y_{qm}	$\bar{y}_q$	s_q	γ_{3q}
⋮		⋮ ⋮	⋮	⋮	⋮
n		y_{n1}……y_{nm}	$\bar{y}_n$	s_n	γ_{3n}

The replicates at each design point are then used to obtain parameter estimates for the data. As explained in Chapter 6, while this typically includes only the sample mean and standard deviation (or variance), the presence of asymmetry necessitates that the sample skewness also be obtained to account for the non-normality in the responses. The formulas used to obtain these are restated here for convenience:

$$\bar{y}_q = \frac{\sum_{j=1}^{m} y_{qj}}{m}, \quad s = \frac{\sum_{j=1}^{m} y_{qj}}{m-1}, \text{ and } \quad \gamma_3 = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_{qj} - \bar{y}_q)^3}{\left(\frac{1}{n-1}\sum_{i=1}^{n}(y_{qj} - \bar{y}_q)^2\right)^{3/2}} \tag{7.1}$$

Prior to the estimation of response surface functions for the mean and variability of the process, a comprehensive data analysis of both the responses and the residuals must be performed to ascertain the underlying conditions in the data. As before, this includes an investigation of normality and variability in the responses, as well as the verification of the assumptions of normality, homoscedasticity, and independence in the residuals. Whereas Chapter 6 focused solely on the responses while presuming that

assumptions associated with residual errors held, in this chapter no such presumptions is made and conditions in both the responses and residuals are investigated. This is done using both graphical and objective methods, which may include the following approaches briefly discussed in (i) and (ii) below:

(i) Assessment of normality and variability in the responses: The approaches used to investigate inherent asymmetry and elevated variability in the responses are the same as those delineated in Chapter 6. Again, these combine the in-depth graphical approaches discussed in Chapter 4 with more objectively-based methods such as the Shapiro-Wilk test for normality. Refer to Section 6.2.1.1 for the details on the objective approaches used to examine each condition.

(ii) Residual analysis. As with the responses, normality in the residuals may also be examined using graphical measures such as the normal probability plot, followed by the more objective methods provided by the Kolmogorov-Smirnov or Shapiro-Wilk tests. To investigate independence, the Durbin-Watson test is usually sufficient to detect a lack of randomness in the residuals. Should remediation be necessary, one can add predictor variables or use transformations in the variables to eliminate interdependencies. Lastly, heteroscedasticity, or non-constant variance, is most often investigated graphically using a plot of the residuals against the fitted values, as well as objectively using either the Brown-Forsythe test, which is more robust to departures from normality in the data, or the Breusch-Pagan test. The Breusch-Pagan test assumes independence and normality among the residuals, but further assumes a relationship for the error variance σ_q^2 among the k regression coefficients and k-1 predictor variables that takes the form $\log_e s^2_q = \gamma_0 +$

$\gamma_1 X_{q1} + \ldots + \gamma_{k-1} X_{q,k-1}$, and implies that the error variance fluctuates up or down with **x**, based on the sign of the associated coefficients. Since constant error variance corresponds to the instance in which each of the coefficients contained in response function equals 0, the alternative hypotheses H_0: $\gamma_1 = \gamma_2 = \ldots = \gamma_{k-1} = 0$versus H_1:not all $\gamma_i = 0$ are tested using the statistic:

$$X^2_{BP} = \frac{SSR^*}{2} \div \left(\frac{SSE}{Nm} \right)^2$$

in which *Nm* denotes the total number of experimental observations, *SSR** is the regression sum of squares obtained by regressing the squared residuals, e_j^2, against one or more of the predictor variables, and *SSE* is error sum of squares obtained for the full regression model. Comparing this statistic to the chi-square distribution with *k*-1 degrees of freedom, if $\chi^2_{BP} > \chi^2_{(1-\alpha),k-1}$ then we reject H_0 and conclude that sufficient evidence exists to suggest non-constant variance. In processes with inherently high variability or asymmetry in the responses, the assumption of constant variance in the residuals would most likely not hold and would thereby necessitate the use of remedial measures. These will be addressed in the next section.

7.2.2 Modeling Symmetry and Asymmetry

Asymmetry in the tier-two case is once again modeled using the skew normal distribution presented in Chapter 6. As discussed in Section 6.2.1.2, the examination of the asymmetric case via simulation requires a different approach in order to account for the inherent skewness in the process. Accordingly, precisely the same methodology is applied in the modeling of process responses (see Appendix B.2 for actual R code).

Simulated replicates are then used to obtain tier-one estimates for the mean, median, standard deviation, and *MAD* at each design point which are, in turn, used to develop response surface functions using each of the ten regression models under consideration. The *MSE*-based optimization scheme is then applied to obtain optimum operating conditions for comparison and analysis.

7.2.3 Estimator Selection

Pursuant to a comprehensive data analysis, a regression estimation approach is selected based upon the inherent conditions identified through the analysis. As has been noted, the focus of this paper is to determine which approaches tend to perform best in terms of yielding optimal *RPD* solutions under asymmetric and highly variable conditions. Accordingly, the investigation considers the alternative regression methods listed in Table 7.2.

Table 7.2. Regression estimators examined as potential *RPD* alternatives.

	Regression Estimators
Base Case	*OLS* ($\bar{y}$ and s as the tier-one basis for regression)
Alternatives for Comparison	1) *GLM* (gamma or inverse Gaussian model) 2) *WLS* ($\bar{y}$ and *s*) 3) *WLS* (median and *MAD*) 4) Least absolute deviation (*LAD*) 5) Least median squares (*LMS*) 6) Least trimmed squares (*LTS*) 7) *M*-estimation (Huber Proposal 2) 8) *M-M* estimation 9) *S*-estimation

Application of the *GLM* procedure first requires the specification of a distributional model for the linear predictor. Given the focus on inherent asymmetry, we examine *GLM*s assuming either a gamma or inverse Gaussian distribution, both of which are

suitable for modeling varying degrees of asymmetry involving positive continuous data in which the variance tends to increase with the mean response. The probability density function for each is given by:

$$\text{Gamma: } f(y:\omega,\psi)=\left(\frac{y}{\omega}\right)^{\psi-1}\left[\frac{\exp\left(\frac{-y}{\omega}\right)}{\omega\Gamma(\psi)}\right] \text{ for } y>0, \text{ where } \omega=\text{scale}, \ \psi=\text{shape}$$

$$\text{Inverse Gaussian: } f(y:\mu,\lambda)=\left[\frac{\lambda}{2\pi y^3}\right]^{-1/2}\exp\left(\frac{-\lambda(y-\mu)^2}{2\mu^2 y}\right) \text{ for } y>0, \text{ where } \mu\text{=mean}, \ \lambda\text{=shape}$$

Pursuant to model selection, an appropriate link function $g(\bullet)$ is also selected, which transforms the expected value of the response to the linear predictor and takes the form:

$$g(\mu_i)=\eta_i=\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+\cdots+\beta_k X_{ik} \tag{7.2}$$

One benefit of using *GLM*s is that the selection of the link function can be separate from the distributional assumption. Although a variety of commonly used link functions has been established, our focus on the gamma and inverse Gaussian distribution limits the selection to the few delineated in Table 7.3.

Table 7.3. Applicable link functions for the gamma and inverse Gaussian distributions.

Link Function	$\eta_i = g(\mu_i)$	Gamma	Inverse Gaussian
Identity	μ_i	X	X
Log	$\log_e \mu_i$	X	X
Inverse	μ_i^{-1}	X*	X
Inverse-square	μ_i^{-2}		X*

* denotes the default link used by the `glm()` function in R.

The selection of the appropriate link function for each distribution depends upon inherent conditions, and although each has default (or canonical) links that typically produce

preferable mathematical and numerical properties, these functions may not be the best choice for a particular set of data. For example, if the response, *Y* increases somewhat linearly with X_i and the variance appears to increase with the square of the mean, then the gamma distribution could be paired with the identity link rather than the default inverse link. Since the response deviance is scaled the same for all models generated using the *GLM* approach, determining the link that will generate the best fit can be achieved by selecting the *GLM* model that yields the lowest residual deviance or Akaike Information Criterion (*AIC*) value.

The remaining alternatives in Table 7.2, dubbed robust regression approaches, were developed specifically to address the influence exercised by outliers on the response surface functions. In this context, the term "robust" describes an estimator's ability to overcome the influences or leverage exerted by outliers on the generated estimate. Furthermore, whereas outlying responses are often viewed as anomalies or potentially contaminated data, asymmetry and elevated degrees of variability can be inherent and intricate parts of manufacturing processes that increase the likelihood of observing such responses. This makes robust estimators particularly interesting because they mitigate the leverage of extreme observations without discounting them altogether.

The inherent process asymmetry and high variability that are the focus of this paper also tend to induce heteroscedasticity (Kutner *et al.* (2005)). The existence of heteroscedasticity suggests that the *OLS* standard errors are potentially wrong, which calls into question any statistical inference based on them, and further suggests that the *OLS* method will not produce the best estimates for the regression coefficients. The

method of weighted least squares (*WLS*) method, developed by Aitken in 1935, is among the first alternatives to remediate such situations. Here, if the residual of the *q*th point is relatively small then it will be retained in the analysis. Conversely, if the *q*th residual is large, it will be removed as an outlier. Mathematically, the modification to the classical *OLS* model is straightforward:

$$\underset{\hat{\beta}}{\text{Minimize}} \sum_{q=1}^{n} w_q \varepsilon_q^{\,2} \tag{7.3}$$

where ε_q denotes the residual associated with the *q*th design point given by $y_q - \mathbf{x}_q\hat{\boldsymbol{\beta}}$, and where y_q refers to the sample mean in the case of multiple replications per design point.

The use of the median and *MAD* as alternative starting points in the *WLS* method merits some mention. Certainly, in the case of a normal or other symmetric distribution (Laplace, uniform, Cauchy, etc.), the sample mean will correspond to the value that possesses the greatest probability of occurring. However, in asymmetric cases, it will not. Consider the comparison of probability densities of samples drawn from a skew normal distribution and a normal distribution with the same mean shown in Figure 7.2.

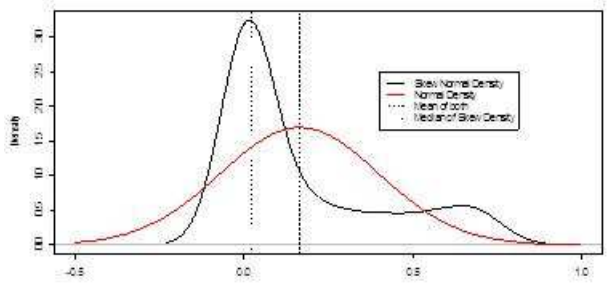


Figure 7.2. Comparison of normal and skew normal densities with common sample mean.

In the normal case, the mean corresponds to the peak in the density function, indicating that the likelihood of that value (and values near it) exceeds all others in that distribution. Conversely, in the skew normal case, the mean lies to the right of the preponderance of probabilities, whereas the median falls very close to the peak of the distribution. Thus, while the mean certainly still defines central tendency in terms of the population mean in the skewed case, it does not necessarily correspond to central tendency in terms of the values of greatest likelihood. When dealing with asymmetric conditions, two factors bear considerable importance: 1) depending on the degree of variability and skew, the mean will shift away from the central tendency of the distribution; and 2) the standard deviation will not accurately describe the dispersion in the distribution as it will be significantly affected by the "play" in the skewed or long tail of the distribution. If contamination exists, the effects worsen. Accordingly, we elected to reexamine both the *OLS* and *WLS* methods using these tier-one estimators.

Based on the conditions examined in this chapter, the weights for the *WLS* method are determined in a manner similar to that used by Goethals and Cho (2011a). That is, observations possessing less variance receive greater weight. Recognizing that $\boldsymbol{\varepsilon}$ denotes the vector of residuals, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ in the general form of the standard regression model $\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}$ in which $\mathbf{Y}$ is the vector of responses, $\mathbf{X}$ is the design matrix and $\hat{\boldsymbol{\beta}}$ is the vector of estimated regression coefficients, in the case of non-constant error variances, we may rewrite the $n \times n$ covariance matrix as:

$$Var(\boldsymbol{\varepsilon}) = \begin{bmatrix} Var(\varepsilon_1) & 0 & \cdots & 0 \\ 0 & Var(\varepsilon_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Var(\varepsilon_n) \end{bmatrix}$$

Since for each of *n* components of $\boldsymbol{\varepsilon}$, $E(\varepsilon_q) = 0$ and therefore $Var(\varepsilon_q) = E(\varepsilon_q^2)$, we may use the vector of squared residuals to estimate the error variance. Then, after regressing the squared residuals against the predictors in **X**, we may use the fitted values of the resulting variance vector $\hat{\boldsymbol{\varphi}}$ to establish weights, w_q, for each of the *n* design points. In order to mitigate the effects of high variability exerted by large residuals, we can then define the weights as inversely proportional to the error variance by $w_q = 1/\hat{\varphi}_q^2$. As such, the higher the error variance, the lower the associated weight for that observation, and vice versa. It can be shown that the *WLS* estimator becomes $\hat{\boldsymbol{\beta}}_w = \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$, suggesting equivalence to the *OLS* estimator. To ensure minimal model error, our procedure uses an iterative approach to reweight the model using subsequent estimation of the error variance. Convergence is achieved and the iteration stops when the difference between the standard error associated with each of the estimated coefficients in $\hat{\boldsymbol{\beta}}_w$ is less than 0.05 relative to the standard errors obtained in the previous iteration.

In response to the risks posed by certain conditions (namely outliers), a variety of so-called robust regression alternatives to the *OLS* method evolved to provide greater resistance to the leverage exerted by outlying observations. Robust regression approaches are designed to circumvent the limitations associated with traditional parametric and non-parametric methods. More specifically, the underlying methods are

designed such that violations of supporting assumptions yield little impact on the regression results.

One of the simplest alternatives for estimating robust regression coefficients is least absolute deviations (*LAD*), which actually predates *OLS* regression. Karst (1958) is considered among the first of contemporary researchers to revive the *LAD* approach in the wake of *OLS* methods, suggesting its use to overcome the susceptibilities of *OLS* regarding the influence of outliers. The *LAD* method basically works to minimize the sum of the absolute values of the residuals or errors between points generated by the regression function and corresponding data points:

$$\underset{\hat{\beta}}{\text{Minimize}} \sum_{q=1}^{n} | \varepsilon_q | \tag{7.4}$$

Schlossmacher (1973) and a number of others extended Karsts work to yield more precise and robust regression models using the absolute deviations approach. Although the *LAD* method has proved more robust than *OLS*, significant outliers can still bear negatively on the model. Several researchers noted this shortcoming relatively early on, which motivated research into even more robust approaches.

Among the first to spearhead the drive for more robust regression methods, Huber (1973) introduced *M*-estimation for regression, which he basically modeled as an extension of his robust parameter for tier-one estimation. Mathematically, the method applied in tier-two estimation focuses on the residuals and takes the following form:

$$\underset{\hat{\beta}}{\text{Minimize}} \sum_{q=1}^{n} \rho(\frac{\varepsilon_q}{s}) + n \log s \tag{7.5}$$

where ρ = some symmetric function with a unique minimum at 0. If we presume *s* is known and set $\psi = \rho'$, then the MLE $\hat{\beta}$ of the regression coefficients solves the non-linear system of equations

$$\sum_{q=1}^{n} \mathbf{x}_i \psi\left(\frac{\varepsilon_q}{s}\right) = 0$$

where ψ represents Huber's bounded monotone ψ function. After some modification, this becomes:

$$\sum_{q=1}^{n} \psi^2\left(\frac{\varepsilon_q}{s}\right) = (n-p)\gamma$$

where γ is selected for consistency at normality and the embedded tier-one estimates for location and scale are obtained using Huber's Proposal 2 estimators, which result from solving the following equations simultaneously for μ and σ (see Huber (1973)):

$$\sum_{q=1}^{n} \psi\left(\frac{y_q - \mu}{\sigma}\right) = 0$$
$$\sum_{q=1}^{n} \psi^2\left(\frac{y_q - \mu}{\sigma}\right) = \eta$$

This method has proven to be a viable and efficient estimator that is robust to outliers in the response variable. However, it was also found to lack resistivity to outliers.

Rousseeuw (1984) proposed the least trimmed squares (*LTS*) method to overcome efficiency shortcomings with a previous method (least median of squares (LMS), Rousseeuw (1984)). The objective in this approach involves minimizing the sum of squared residuals over a subset, *q*, of the complete set of *n* points:

$$\underset{\hat{\beta}}{\text{Minimize}} \sum_{q=1}^{q} \left|\varepsilon_q\right|^2_{q:n} \tag{7.6}$$

In short, the residuals are squared and then sorted in ascending order. Of the n residuals in the full set, the $(n - q)$ largest are "trimmed" so that only the residuals from the remaining q points are included in the regression. Thus, the $(n - q)$ largest points which are not used do not influence the fit. The result is a fit that retains the resistivity properties of the LMS method, but is more efficient.

The fact that both the *LMS* and *LTS* methods involve the minimization of a robust measure of the scatter of the residuals gave rise to *S*-estimation as a generalization of the two. Introduced by Rousseeuw and Yohai (1984) as a means for performing robust regression in time series analysis, this method finds a plane or hyperplane in which the coefficients are selected to identify a solution to

$$\sum_{q=1}^{n} \chi\left(\frac{y_q - \mathbf{x}_q\hat{\boldsymbol{\beta}}}{c_0 s}\right) = (n-p)\varphi \tag{7.7.}$$

such that it minimizes the scale, s. In this context, p corresponds to the k-1 predictors; χ is typically denoted by the integral of Tukey's bisquare function given by

$$\chi(u) = \begin{cases} u^6 - 3u^4 + 3u^2 & |u| \le 1 \\ 1 & |u| > 1 \end{cases}$$

and c_0=1.548 and φ = 0.5 are selected for consistency at the normal distribution. This method is highly resistant to leverage points, robust to outliers in the response, and more efficient than the *LTS* method.

Yohai (1987) proposed the *M*-*M*-estimator as an improved alternative to *LMS* and *LTS* that would retain the high breakdown points these methods achieved but would also achieve higher efficiency. Yohai's approach essentially blended earlier methods in order to retain the robustness and resistance of *LMS*, *LTS*, and *S*-estimation, while gaining the efficiency of *M*-estimation. The method proceeds in three stages. The first involves an initial estimation of regression coefficients (in the case of Yohai (1987), using either the *LMS* or *LTS* method). In the second, a highly robust and resistant *S*-estimate is computed that minimizes an *M*-estimate of the scale of the residuals. In the final stage, the estimated scale is then held constant while a nearby *M*-estimate of the regression coefficients is determined.

7.2.4 Integrating the Estimators into the *RPD* Framework

Considering the framework delineated in Section 2.1 for an industrial process involving a nominal-the-best (*N*-type) quality characteristic as the response of interest, we assume that the levels of x_i for $i = 1, 2,\ldots,k$ are both quantitative and continuous, and can be controlled by the experimenter. Pursuant to the identification of underlying conditions, tier-one estimates for the location and scale are obtained at the q^{th} design point using $\bar{y}$ and $\tilde{y}$ for the location, and *s* and *MAD* for the scale. Again, as was discussed in Section 6.2.1.2, the estimates for the process mean and standard deviation are influenced by the inclusion of the sample skew in their derivation. Thus, inherent process skewness is accounted for in the ultimate response surface estimates used for the process mean and standard deviation, which ensures that actual process characteristics are more accurately represented.

will be discussed in Section 7.3.2, this will require a more indirect approach when addressing asymmetric conditions. Fitted response functions (tier-two estimates) are then developed for the process location and scale for each of the regression methods outlined in Section 2.2. In particular, assuming second-order polynomials for the response functions in each case, the general form of the estimated response functions for the process location and scale with k parameters or $k-1$ predictor variables appears as:

$$\text{Location:} \quad \hat{\mu}(\mathbf{x}) = \hat{\beta}_{\mu,0} + \mathbf{X}^{\mathrm{T}}\hat{\mathbf{b}}_{\mu} + \mathbf{X}^{\mathrm{T}}\hat{\mathbf{B}}_{\mu}\mathbf{X} + \varepsilon_{\mu} \tag{7.8}$$

$$\text{Scale:} \quad \hat{\sigma}(\mathbf{x}) = \hat{\beta}_{\sigma,0} + \mathbf{X}^{\mathrm{T}}\hat{\mathbf{b}}_{\sigma} + \mathbf{X}^{\mathrm{T}}\hat{\mathbf{B}}_{\sigma}\mathbf{X} + \varepsilon_{\sigma} \tag{7.9}$$

where $\underset{\mathbf{1\times k}}{\mathbf{X}^{\mathrm{T}}} = \begin{bmatrix} X_1 & X_2 & \cdots & X_{k-1} \end{bmatrix}$, $\underset{\mathbf{k\times 1}}{\hat{\mathbf{b}}} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$, and $\underset{\mathbf{k\times k}}{\hat{\mathbf{B}}} = \begin{bmatrix} \beta_{11} & \beta_{12}/2 & \cdots & \beta_{1k}/2 \\ & \beta_{22} & \cdots & \beta_{2k}/2 \\ & & \ddots & \vdots \\ \textit{sym.} & & & \beta_{kk} \end{bmatrix}$

and where $\hat{\beta}_{\mu,0}$ (and $\hat{\beta}_{\sigma,0}$), $\hat{\mathbf{b}}_{\boldsymbol{\mu}}$ (and $\hat{\mathbf{b}}_{\boldsymbol{\sigma}}$) and $\hat{\mathbf{B}}_{\boldsymbol{\mu}}$ (and $\hat{\mathbf{B}}_{\boldsymbol{\sigma}}$) once again correspond to the estimates of the intercept, linear, and second-order coefficients of the response surface functions for the location and scale, respectively. The term ε_{μ} and ε_{σ} correspond to the residual error for the mean and standard deviation, respectively. In similar fashion, fitted response surface functions are developed for each of the estimate vectors containing the supporting information for each parameter. The *MSE*-based optimization scheme given by

$$\text{Minimize} \quad MSE = (\hat{\mu}(\mathbf{x}) - \tau) + \hat{\sigma}^2(\mathbf{x}) \tag{7.10}$$

is again used on either a spherical region of interest (central composite design) such that $\mathbf{x}'\mathbf{x} \le \rho^2$, or a cuboidal region of interest (full factorial design) whereby $\mathbf{x}$ is bounded by

(-1,1) . Using this approach as a framework, the various regression estimators are evaluated in their ability to determine optimal operating conditions for the system.

7.3 Numerical Demonstration via Simulation

In this section, we examine two cases using commonly applied experimental data sets as bases for Monte Carlo simulation. The overarching purpose of the simulation is to demonstrate the degree to which underlying process asymmetry and variability (and, consequently, heteroscedasticity) affect estimator performance in the context of *RPD* solutions, which should ultimately serve to inform engineers as to which estimators tend to perform best under a particular set of conditions..

The first study involves normally distributed data with moderately low variability in which all of the base assumptions concerning the data hold. Through experimentation, we then investigate the impact of increasing variability on estimator performance. In the second, we expand the investigation to include asymmetry. We examine asymmetric data with low variability and heteroscedasticity using four scenarios derived from combinations of high/low asymmetry with high/low variability. Within each case study, initial results are obtained from the base data and observations are then drawn to assess estimator performance. Thereafter, 1,000 iterations of each simulation scenario are conducted to enable the analysis of performance trends and thereby develop fuller assessments regarding estimator performance under the evaluated conditions.

Simulations were developed in the statistical computing environment R version 2.14.1 (see reference R Core Team (2012)). For the purposes of estimator comparison, each simulation involves several key settings that are applied to each estimation model:

1) Using the actual experimental data obtained for each case, simulated data are derived using the skew normal approach introduced in Section 6.2.1.2 and as discussed in Section 7.2.4
2) Full second-order response surface functions are developed for the location (mean or median) and scale (standard deviation or *MAD*) response surface functions using Equations 7.8 and 7.9.
3) Optimization results are obtained for each estimation model using the *MSE*-based optimization scheme (Eqn. 7.10) developed by Cho (1994) / Lin and Tu (1995).

Pursuant to (3) above, estimation approaches are then evaluated based on the optimization results they generate in terms of the *MSE* and deviation from the established process target (i.e., target bias, or simply bias).

7.3.1 Case Study A - Metal Cutting Process

In this experiment, adapted from Shin, *et al.* (2011), the metal removal rate (mm^3/min) of a metal cutting machine is a normally distributed *N*-type quality characteristic of interest, *Y*. The control factors known to influence *Y* include the cutting speed (X_1), cutting depth (X_2), and cutting feed rate (X_3) settings for the machine. The desired target value for the machine's removal rate is $\tau = 57.5$ mm^3/min. The experimental framework displayed in Table 7.4 is a *CCD* comprised of eight factorial points, six axial points, and six center points, with the calculations for the mean and standard deviation at each design point.

Table 7.4. Experimental framework for the metal cutting study.

	Coded Units								Metal Removal Rate (mm^3/min)		
	Cut Speed	Cut Feed	Cut Depth	Observed Responses (simulated)							
Run	X_1	X_2	X_3	Y_1	Y_2	Y_3	Y_4	Y_5	$\bar{Y}$	s	γ_3
1	-1	-1	-1	44.6	52.5	57.4	52.4	57.8	53.2	3.82	0.19
2	1	-1	-1	63.9	60.3	64.7	65.8	67.5	62.9	3.51	-0.77
3	-1	1	-1	45.6	51.5	45.4	62.0	52.8	53.4	3.67	-0.24
4	1	1	-1	67.1	64.5	61.6	58.6	55.5	62.6	3.24	-0.32
5	-1	-1	1	59.4	55.6	51.4	57.7	59.5	57.3	3.10	-0.11
6	1	-1	1	67.6	64.6	64.3	71.8	67.4	67.9	4.31	-0.21
7	-1	1	1	65.5	60.8	60.5	57.2	55.6	59.8	4.47	0.46
8	1	1	1	67.4	66.5	71.8	68.2	72.0	67.8	3.21	-0.85
9	-1.682	0	0	58.2	56.1	61.3	65.0	47.3	59.1	4.73	-1.13
10	1.682	0	0	69.5	63.2	59.3	73.0	61.0	65.9	4.46	0.73
11	0	-1.682	0	63.2	60.4	59.0	61.0	65.8	60	3.55	-0.16
12	0	1.682	0	59.5	62.6	61.7	57.3	59.9	60.7	3.10	-0.17
13	0	0	-1.682	51.7	66.3	57.2	61.9	64.4	57.4	4.29	-1.13
14	0	0	1.682	65.3	66.1	61.4	72.5	64.2	63.2	5.04	1.32
15	0	0	0	60.3	56.5	64.1	61.1	60.5	59.2	3.87	-0.03
16	0	0	0	59.2	66.9	56.7	62.7	57.8	60.4	3.74	-1.18
17	0	0	0	58.5	59.0	61.2	56.4	57.2	59.1	3.95	-0.08
18	0	0	0	62.4	53.0	59.6	64.0	56.6	60.6	3.71	0.22
19	0	0	0	64.8	63.3	60.9	54.9	66.3	60.8	4.00	0.64
20	0	0	0	53.4	60.5	60.9	64.7	59.9	58.9	3.92	-0.51

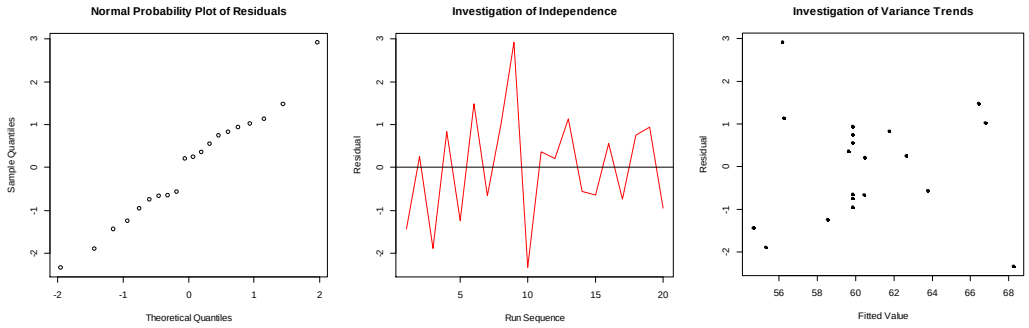
7.3.1.1 Preliminary Data Analysis

The initial graphical analysis of the responses suggests that symmetry and moderately low variability are inherent to this particular set of data. Apart from what appear to be a few "stray" responses, the observations in the normal probability plot in Figure 3(a) generally fall along the reference line, implying normality. A more objective assessment using the Shapiro-Wilk test the statistic yields $W^* = 0.943$ vs. $W_\alpha \approx 0.951$. Since $W^* < W_\alpha$ we conclude H_0, or that sufficient evidence exists to support normality, which reinforces the graphical analysis. Regarding process variability, Figure 7.3(b) shows the deviations between the observations and the mean response to be small, and well within the 3σ threshold previously defined.

(a) (b)

Figure 7.3. Analysis of variability (a) and normality (b) in the metal cutting responses.

After performing a preliminary regression for a full second-order model for the mean response using the *OLS* approach, a graphical analysis of the residuals (Figure 7.4) suggests that the assumptions of normality and independence hold, but that heteroscedasticity may exist.



(a) (b) (c)

Figure 7.4. Investigation of assumptions on a) normality, b) independence, and c) constant variance in the residuals.

To confirm the variability trends observed in the residual plots for full ten-parameter second-order model, we apply the Breusch-Pagan hypothesis test in which the calculated statistic tests H_0: $\gamma_1 = \gamma_2 = \ldots = \gamma_{10} = 0$ versus H_1: Not all $\gamma = 0$. This yields the test result

$$\chi^2_{BP} = \frac{SSR^*}{2} \div \left(\frac{SSE}{Nm} \right)^2 = \frac{78}{2} \div \left(\frac{31.1}{20} \right)^2 = 16.1 \quad < \quad \chi^2(.95,9) = 16.9$$

which, since 16.1<16.9, suggests constant variance and thus disputes the deduction suggested by the plot in Figure 7.3(c). However, unless more than thirty-five observations are available, it would be difficult to find a data set with non-constant variance; it is often the case with smaller sample sizes that the objective test results fail to capture the presence of non-constant variance. Thus, to investigate this more fully, since the mean responses in Table 7.3 derive from evaluating multiple observations at each design point, we can gain much more qualitatively by considering all of the individual observations rather than just the mean. The Breusch-Pagan test still compares the differences in the residuals across the fitted line; it just happens to have more data available at each setting of the fitted line to compare. Pursuant to this, we use the sample statistics for each design point in Table 7.3 as a basis for simulating five observations at each point for a total of 100 experimental observations in R. Thereafter, a full second-order regression is performed once again and the reiterated Breusch-Pagan test yields $\chi^2_{BP} = 11.9$, which is clearly less than $\chi^2(0.95,9) = 16.919$ and suggests that the error variance is, in fact, constant. Note that the patterns in the residual plots for the 100-observation case in Figure 7.5 closely resemble the patterns in Figure 4 but that the greater number of observations makes it much more obvious that the base residual assumptions hold.

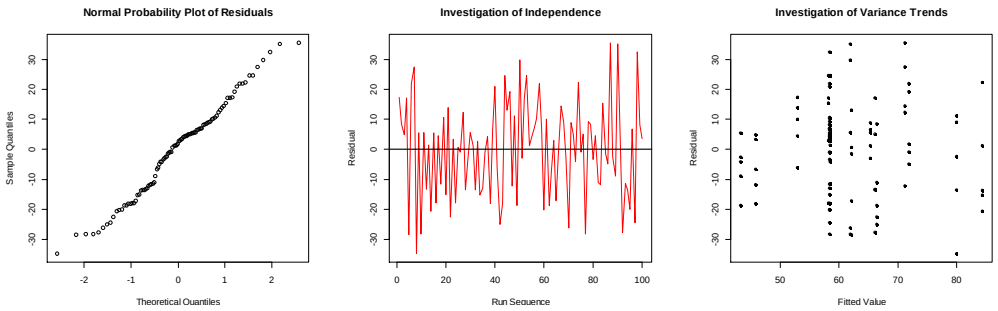


Figure 7.5. Residual analysis based on all 100 observations in the metal cutting study.

Taken together, the results of the data analysis suggest that the experimental data meet all requisite provisions for the application of *OLS* regression. This implies that *OLS* would be the best approach given that this method is known to produce the best linear unbiased estimates (*BLUE*) for the process location and scale, or dispersion when these conditions hold.

7.3.1.2 Simulation Results with Original Experimental Data

As a starting point for this study, a single run of the metal cutting experiment is performed to motivate the discussion on conditions-based selection of tier-two estimators. Additionally, to demonstrate the benefits of using the skew normal distribution to model system properties, the *OLS* method using traditional tier-one estimators under the assumption of zero skewness is also applied. The results in Figure 6 show the optimal operating conditions $\mathbf{x}^* = (x_1^*, x_2^*, x_3^*)$ obtained under each regression model using the *MSE*-based optimization scheme, the associated optimal process mean and standard deviation, and the resulting target bias and *MSE*. For the *GLM* approach, the Gaussian-identity (default) distribution-link combination was used, which essentially mirrors the *OLS* counterpart and is appropriate when traditional assumptions hold.

	OLS (Traditional)	*OLS (SN)*	*WLS Mean/s*	*WLS Median./MAD*	*LTS*	*S*	*LAD*	*MM*	*Huber Prop 2*	*GLM*
x_1	-0.380	-1.682	-0.071	-0.515	0.056	0.023	-0.215	0.101	-1.682	-1.682
x_2	-1.682	-1.682	1.682	-1.682	-1.682	-1.682	1.682	-1.682	-1.682	-1.682
x_3	-0.170	1.014	-1.682	1.682	-1.682	-1.682	-1.682	-1.682	1.155	1.014
$\hat{\mu}(\mathbf{x}^*)$	57.172	57.615	58.825	57.540	57.466	57.490	58.430	57.459	57.586	57.615
bias	0.328	0.115	1.325	0.040	0.034	0.010	0.930	0.041	0.086	0.115
$\hat{\sigma}(\mathbf{x}^*)$	2.649	2.086	2.733	0.449	2.369	2.169	2.563	2.928	2.064	2.086
MSE	7.127	4.365	9.225	0.203	5.612	4.703	7.431	8.575	4.266	4.365

Figure 7.6. Optimization results of single run with 5 simulated observations (darker shading indicates a better *RPD* solution in terms of *MSE*).

From the results in Figure 6, a single run of the experiment suggests two things: first, that accounting for even low degrees of asymmetry can produce better *RPD* solutions than the traditional approach to *OLS* estimation; and second, that *OLS* regression (under the *SN* approach) is still suitable, although the use of the median-based *WLS* method can achieve superior results. However, recognizing that these solutions are, in fact estimates, it is therefore quite likely that subsequent implementations of the experiment could yield different sets of optimal coordinates. Moreover, the objective is to examine trends to develop a better sense of how the estimators perform on average, which cannot be achieved via a single run. Accordingly, 1,000 iterations of the simulation were executed, generating fresh random data at each iterate. At the conclusion of each, the *MSE* and target bias were recorded for the optimal mean response and then averaged across all iterations to observe trends. Figure 7.7contains the simulation results, along with the proportion of iterations in which a particular estimation approach yielded the smallest *MSE* and bias.

	OLS (Traditional)	*OLS (SN)*	*WLS Mean/s*	*WLS Median/MAD*	*LTS*	*S*	*LAD*	*MM*	*Huber Prop 2*	*GLM*
Avg MSE	4.029	3.250	3.349	2.379	2.636	2.593	3.359	3.205	3.292	3.250
% Best MSE	4.90%	3.50%	5.10%	30.30%	20.70%	20.50%	9.20%	5.50%	1.90%	3.50%
Avg Bias	0.300	0.266	0.260	0.243	0.205	0.200	0.220	0.220	0.259	0.266
% Best Bias	7.20%	4.80%	6.70%	19.50%	23.10%	19.90%	9.70%	7.70%	3.90%	4.80%

Figure 7.7. Simulation results under low-variability conditions.

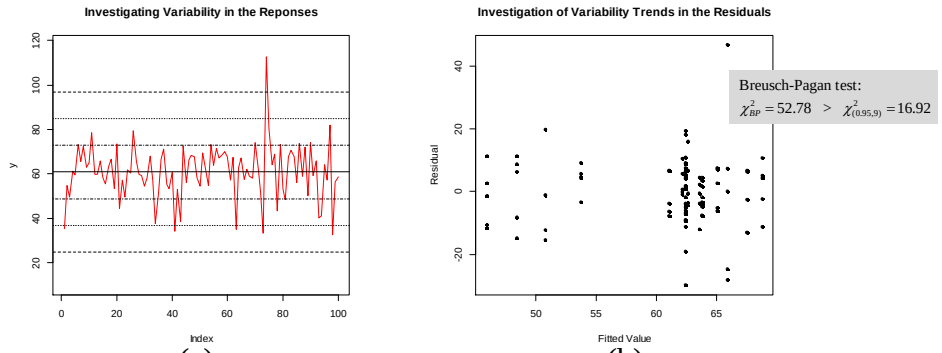
Noting that all nine of the alternative estimation approaches outperformed the traditional *OLS* approach in Figure 7.7, it is clear that despite approximate symmetry/normality in the process data, there is enough inherent skewness to affect the optimization results. In the most basic sense, this is illustrated by comparing the first two columns in Figure 7 (*OLS-Traditional* vs. *OLS-SN*), which suggests that by accounting for even slight levels of non-zero skewness, better *RPD* results can be obtained. Beyond this, the fact that median-based approaches (*WLS*(median/*MAD)*, *LTS*, and *S*-estimation) yielded the best results in terms of both average performance and consistency suggests that these methods are preferable when any degree of asymmetry exists. That the *WLS* procedure produced better results (on average) is most likely the result of down-weighting those observations with higher variability, thus demonstrating the viability of using that method to exert greater control over sources of process variation.

7.3.1.3 Investigating the Effects of High Variability Conditions

To examine the effects of variability, a simple modification to the simulation in R was incorporated that would induce a greater degree of variability in the process. Whereas before the sample standard deviations (*s*) in Table 7.3 was used to generate random variates in the base scenario, in this instance a randomly sampled integer from a range of 2 to 5 is obtained at each design point to serve as a factor that would then be

multiplied by the original values for *s*. Hence, the variability at each design point would be increased by a factor of 2 to 5 times.

The idea here was that simply multiplying the *s* vector by a single common factor would not have any impact on the results other than to scale them by that factor. That is, there would certainly be more variability in the responses, but the proportional change in each design point would be the same and would negate any real effects on the results as the underlying conditions regarding base assumptions would still hold. Thus, the objective was to inject variability not only horizontally within each design point, but also vertically across the vector of sample standard deviations. This would challenge system performance and very likely upend the underlying assumptions of response variability and heteroscedasticity. As the plots in Figure 7.8 show, this is precisely what occurs, as several observations exceed the 3σ threshold (Figure 7.8(a)), and the variability trends coupled with the Breusch-Pagan results in Figure 7.8 (b) ($\chi^2_{BP} = 52.78 > \chi^2_{(0.95,9)} = 16.92$) clearly indicate non-constant variance in the residuals. As noted previously, the presence of such conditions inhibits the use of *OLS* and suggests the need for either remedial measures, such as transformations, or alternative estimation approaches.



(a) (b)

Figure 7.8. Analysis of responses (a) and residuals (b) under high-variability conditions

Once again, in the interests of establishing performance trends among the estimators, we used the aforementioned approaches for increasing variability and simulating data to generate five observations from a skew normal distribution at each design point. In this instance, the *GLM* approach is also modified to account for differences in the data. Specifically, further analysis of the residual data for response surface functions for both the mean and variation suggested the need to consider either a gamma or inverse Gaussian distribution to correct for non-constant variance. After preliminary modeling using each distribution and their respective options for link functions in the *GLM* approach, it was determined that the gamma-identity and inverse Gaussian-log combinations would produce the best fit for the mean and standard deviation response surface functions, respectively, as these yielded the lowest residual deviance and *AIC* values. To establish performance trends among the various estimation approaches, 1,000 iterations of the experiment were performed under the simulated high-variability conditions. Results for the *MSE* and bias were then averaged across all 1,000 iterations, and performance proportions were calculated to produce the results shown in Figure 7.9.

The results indicate several things. First, all nine of the alternatives once again produced a better result than the traditional *OLS* method, reinforcing the benefit of using the skew normal approach for modeling process asymmetry. Second, it is clear that the increased variation induces a change in estimator performance such that the *GLM* method using the gamma-identity and inverse Gaussian-log combinations outperforms all others on average, both in terms of the resulting *MSE* and target bias. While the next-best

performers (LTS, *WLS* (median/*MAD*), and S-estimation methods) performed relatively well, they all achieved an average *MSE* nearly 6 times larger than the *GLM* method.

	OLS (Traditional)	*OLS (SN)*	*WLS Mean/s*	*WLS Median/MAD*	*LTS*	*S*	*LAD*	*MM*	*Huber Prop 2*	*GLM*
Avg MSE	35.604	30.398	28.261	20.194	19.714	20.277	32.284	28.514	31.253	3.643
% Best MSE	4.90%	3.70%	7.30%	18.30%	17.40%	16.10%	8.60%	5.60%	3.40%	16.00%
Avg Bias	2.225	2.079	2.087	1.790	1.579	1.694	2.089	2.037	2.112	0.166
% Best Bias	5.20%	2.80%	6.40%	15.10%	15.90%	14.40%	8.10%	4.90%	3.50%	24.80%

Figure 7.9. Simulation results under high-variability conditions.

Although differences in the generated data can be a contributing factor, the reasons behind these results can also be attributed to the increased likelihood of extreme observations in either tail. And if an extreme observation from one tail is not counter-balanced by an extreme point from the other, then the resulting sample could very well appear skewed, despite being generated from a normal distribution. Obviously, when the data are approximately normal, then the mean and the median will assume nearly the same value. However, as the data become skewed due to the occasion of one or more extreme observations, mean-based estimators deteriorate in their ability to provide the best estimate of central tendency due to the influenced of outlying data points. Similarly, the standard deviation no longer provides the best measure of the true dispersion in the distribution. The median and the *MAD*, on the other hand, retain their properties and are resistant to extreme observations, thereby making them preferable when such conditions exist. In addition to this, high variability typically will also induce heteroscedasticity, which invariably creates situations whereby less than optimal solutions result using *OLS* regression. As the plots in Figure 7.8 show, this is precisely what is occurring in this scenario, and serves to explain why the robust and *GLM* approaches perform well.

7.3.2 Case Study B –Ceramic Coating Process

The investigation of asymmetric conditions involves the ceramic coating experiment described in Chapter 6 (Section 6.3.1.2) which was obtained from Tillman *et al.* (2010). As before, the univariate case is examined using porosity as the quality characteristic of interest. Recall that porosity is an *S*-type characteristic with an upper acceptable threshold of 3% and an associated objective of minimizing as close to 0% as possible. Note that this could also be treated as an *N*-type characteristic with lower and upper limits of 0% and 3%, respectively, and a desired target, τ=0. As in Chapter 6, a replication-based adaption is used to facilitate tier-one estimation at each design point. Table 7.5 shows the same *CCD* used for this experiment in Chapter 6.

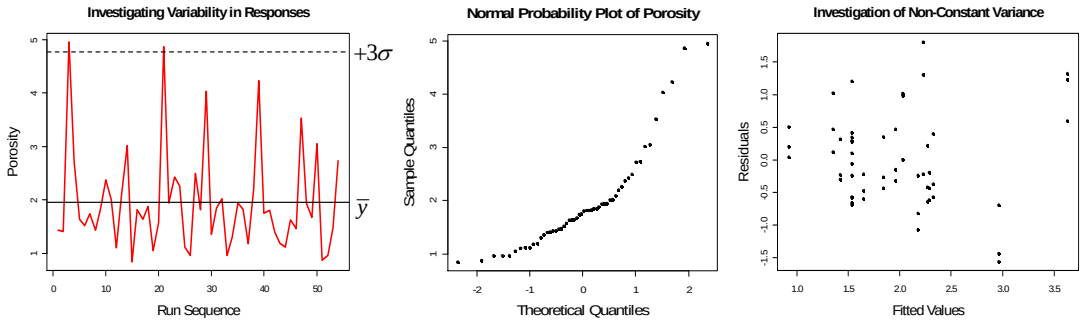
7.3.2.2 Preliminary Data analysis

As in the first case study, a graphical analysis of the data is performed to investigate the assumptions regarding normality and variability in the responses and homoscedasticity in the residuals (along with the assumptions of normality and independence). The plots in Figure 7.10 show the process to be highly variable (10a), asymmetric (10b), and that error variance is non-constant (10c). Application of the Shapiro-Wilk test yields $W^* = 0.837$ vs. W_α=0.946 which supports the alternative hypothesis H_1 (that sufficient evidence exists to suggest non-normality), and further reinforces the graphical analysis. Similarly, the Breusch-Pagan test using full second-order regression yields $X^2_{BP} = 25.705 < \chi^2(0.95,9) = 16.919$, reinforcing the deduction that the error variance is non-constant. As before, the existence of these conditions nullifies several of the

assumptions of the *OLS* approach and demonstrates the need to consider alternative estimation techniques for this particular process.

Table 7.5. Adapted experimental design for the ceramic coating process.

	Control Factor Settings (Coded)			Replications (% Porosity)			Sample Moments Used to Estimate (ξ,ω,α)		
Run	X_1	X_2	X_3	y_1	y_2	y_3	$\bar{y}$	s	γ_3
1	–1	–1	–1	1.43	1.05	1.18	1.22	0.19	0.89
2	–1	–1	1	1.41	1.58	2.2	1.73	0.42	1.41
3	–1	1	–1	4.95	4.86	4.23	4.68	0.39	-1.63
4	–1	1	1	2.72	1.95	1.75	2.14	0.51	1.44
5	1	–1	–1	1.64	2.43	1.81	1.96	0.42	1.41
6	1	–1	1	1.52	2.27	1.4	1.73	0.47	1.61
7	1	1	–1	1.74	1.12	1.19	1.35	0.34	1.65
8	1	1	1	1.43	0.96	1.12	1.17	0.24	0.90
9	–1.682	0	0	1.85	2.49	1.63	1.99	0.45	1.27
10	1.682	0	0	2.38	1.82	1.47	1.89	0.46	0.67
11	0	–1.682	0	2.01	4.03	3.53	3.19	1.05	-1.30
12	0	1.682	0	1.11	1.36	1.94	1.47	0.43	1.08
13	0	0	–1.682	2.09	1.85	1.67	1.87	0.21	0.42
14	0	0	1.682	3.02	2.03	3.05	2.70	0.58	-1.73
15	0	0	0	0.85	0.97	0.88	0.90	0.06	1.29
16	0	0	0	1.82	1.3	0.96	1.36	0.43	0.61
17	0	0	0	1.64	1.95	1.48	1.69	0.24	0.90
18	0	0	0	1.88	1.83	2.74	2.15	0.51	1.71



(a) (b) (c)

Figure 7.10. Graphical investigation of variability (a) and normality (b) in the responses and homoscedasticity in the residuals (c).

7.3.2.3 Simulation Development and Results

In this particular experiment, our objective was to determine estimator performance trends given varying degrees of both asymmetry and variability in the simulated process. Pursuant to this, we examined four scenarios that coupled high/low

degrees of asymmetry with high/low degrees of variability. Similar to the manner in which we adjusted variability in the first case study, in this case we used factors to increase the levels of both conditions, as depicted in Figure 7.11. Note that the Low-Low scenario served as the base case and was based on the natural process skewness and variability depicted in the original experimental data.

		Variability	
		Low	High
Asymmetry	Low	Scale (ω) * 1 Shape (α) * 1	Scale(ω) * 1 Shape (α) * 1.5
	High	Scale(ω) * factor (random integer drawn from (3 - 6) Shape (α)* 1	Scale(ω) * factor (random integer drawn from (3 - 6) Shape (α) * 1.5

Figure 7.11. Simulation scenarios examined for Case Study B (base settings highlighted).

It should be noted that, whereas all of the other approaches remained unchanged between scenarios, the application of *GLM*s evolved with the conditions. In particular, as the conditions changed between scenarios, each of the possible distribution-link combinations was re-evaluated to determine which among them would produce the best fit in a given scenario. Using the residual deviances and *AIC* values as selection criteria resulted in the *GLM* combinations for each simulation scenario shown in Table 7.6.

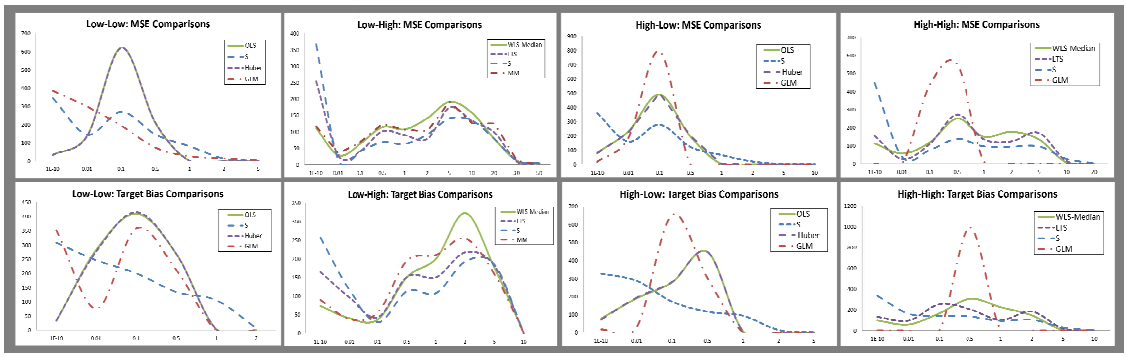
Table 7.6. Distribution-link combinations used for *GLM* method in each scenario.

Scenario	Response Surface Function	Distribution	Link Function
Low-Low	Mean Variation	Inverse Gaussian Gamma	Inverse square Inverse
Low-High	Mean Variation	Gamma Gamma	Identity Identity
High-Low	Mean Variation	Inverse Gaussian Gamma	Inverse square Identity
High-High	Mean Variation	Gamma Gamma	Inverse Identity

As before, 1,000 iterations of each scenario were performed, recording the *MSE* and target bias associated with each regression approach at each iterate, and then averaging across all iterations to facilitate trend analysis. Since the benefits of the *SN* approach were demonstrated in the first case study, the traditional *OLS* approach is omitted and only the *OLS* method using the *SN* approach for tier-one estimation is used. Figure 7.12 contains the numerical results of the simulation experiment in terms of the *MSE* and bias of the optimization results, including the proportion of the 1,000 runs each method achieved the lowest *MSE* and bias. Figure 7.13 contains frequency plots of the four top-performing regression methods for *MSE* and bias in each scenario. These plots essentially reflect the distribution of the results for each measure under each method shown.

Scenario	Perf. Measure	*OLS (SN App.)*	*WLS Mean/s*	*WLS Median/MAD*	*LTS*	*S*	*LAD*	*MM*	*Huber Prop 2*	*GLM*
Low – Low	*Avg MSE*	0.06	0.14	0.12	0.35	0.12	0.39	0.23	0.06	0.07
	% Best	4.80%	1.30%	3.20%	8.50%	31.10%	1.40%	3.20%	3.90%	45.60%
	% Worst	0.60%	3.70%	7.10%	24.20%	8.80%	36.40%	13.80%	0.60%	5.40%
Low – High	*Avg MSE*	4.15	4.51	3.42	3.26	3.03	4.77	3.72	4.21	4.38
	% Best	5.00%	6.50%	11.40%	19.90%	31.00%	13.80%	6.90%	5.50%	3.60%
	% Worst	7.30%	14.00%	11.80%	13.90%	16.00%	19.70%	5.00%	7.50%	10.80%
High – Low	*Avg MSE*	0.06	0.13	0.10	0.23	0.07	0.37	0.14	0.06	0.03
	% Best	11.00%	2.20%	5.60%	14.30%	39.80%	1.40%	9.30%	9.20%	14.60%
	% Worst	0.70%	5.10%	7.10%	17.50%	9.60%	49.30%	8.60%	0.80%	2.00%
High – High	*Avg MSE*	2.362	1.002	0.871	0.956	0.523	2.842	1.838	2.452	0.141
	% Best	0.90%	6.52%	13.15%	13.37%	41.80%	2.58%	3.03%	0.90%	18.31%
	% Worst	19.33%	2.81%	5.73%	8.54%	6.07%	41.80%	11.46%	22.13%	0.34%
Low – Low	Avg Bias	0.08	0.25	0.20	0.38	0.13	0.47	0.23	0.08	0.05
	% Best	7.40%	1.20%	2.00%	9.00%	35.70%	1.60%	9.80%	7.20%	31.70%
	% Worst	0.90%	5.62%	7.53%	24.38%	7.53%	37.19%	15.51%	0.79%	1.35%
Low – High	Avg Bias	1.14	1.26	1.09	0.96	0.91	1.12	1.00	1.14	1.27
	% Best	5.00%	6.50%	10.70%	19.90%	31.90%	14.10%	7.40%	5.60%	2.20%
	% Worst	7.30%	13.93%	13.60%	15.06%	16.63%	17.08%	4.04%	7.30%	10.90%
High – Low	Avg Bias	0.12	0.26	0.20	0.28	0.10	0.49	0.14	0.12	0.08
	% Best	9.20%	1.30%	3.10%	15.50%	41.90%	1.10%	21.90%	9.10%	3.90%
	% Worst	1.24%	6.52%	6.29%	16.40%	8.20%	50.67%	8.76%	1.35%	1.80%
High – High	Avg Bias	1.061	0.504	0.454	0.454	0.322	1.176	0.785	1.088	0.310
	% Best	1.46%	7.98%	12.25%	20.56%	45.39%	3.26%	4.49%	1.12%	4.16%
	% Worst	19.89%	2.92%	5.28%	8.54%	6.63%	39.55%	11.01%	21.80%	2.70%

Figure 7.12. Average performance results for each simulation scenario (darker shading indicates superior performance under the stated scenario conditions).



(a) (b) (c) (d)

Figure 7.13. Frequency plots of *MSE* and Bias results for the top-performing approaches for the (a) Low-Low, (b) Low-High, (c) High-Low, and (d) High-High scenarios.

The results in Figure 7.12 show a clear shift away from traditional *OLS* approaches when various degrees of both asymmetry and variability coexist. Each of the scenarios is discussed in paragraphs (i) – (iv) below.

(i) Low-Low Scenario: Under low degrees of both asymmetry and variability, the *OLS*, Huber, and *GLM* methods all achieve the best average performance in terms of both *MSE* and bias (although the *GLM* method edges the others out in bias. However, it is clear that the *GLM* method achieves the best results far more often than the other top-performers, thereby making it more preferable under these conditions. The reason for this is that while low/moderate levels of asymmetry/variability allow the *OLS* and Huber methods to perform well, there is enough of each inherent to this process (see plots in Figure 7.10) to allow the *GLM* method with the right distribution-link combination to achieve better results. Moreover, the frequency plots in Figure 7.13(a) reinforce this. In particular, the results from the *GLM* approach (and, to a degree, the *S*-estimation method)

assume a more exponential appearance, suggesting a much higher likelihood of very low *MSE*/bias values in the *RPD* solutions obtained.

(ii) Low-High Scenario: When the degrees of variability become significantly elevated, median-based regression approaches become preferable. Specifically, the *S*-estimation, *LTS*, and *WLS*(median/*MAD*) methods outperform the others in terms of both measures. In this instance, the *S*- method becomes especially preferable simply because it achieves the best results far more frequently than any of the other approaches. While the frequency plots in Figure 7.13(b) depict a wide range for potential results for each of top-performing approaches, they also show that the *S*- method provides a much greater chance of achieving an *MSE*/bias result that is less than 0.01, and a lower chance of achieving *MSE*/bias results greater than 0.01 (that is, to the left of 0.01, the *S*-method curve is higher than the others, whereas to the right it is lower).

(iii) High-Low Scenario: Although the *GLM* method produces the lowest *MSE* and bias on average, the *S*-estimation method is preferred under these conditions. The reasons are two-fold: first, the average results using this method are only slightly higher than those achieved using the *GLM* method; second, the S-estimation method achieves the best results far more frequently and with much greater likelihood (see Figure 7.13(c)). Beyond this, the high-low scenario produces interesting results in that the *OLS* method performs quite well despite the presence of significant process skewness. The reason for this can be explained by the fact that the relatively low levels of variability translate to reductions in both the length and heaviness of the long tail of the skewed distribution. Thus, the influence typically associated with observations from the long tail is now

reduced. This effect is shown in Figure 7.14, whereby the low degree of variability counteracts the elevated levels of asymmetry by shortening and thinning the tail while causing the distribution to peak around its central values. The net effect is some semblance of symmetry that is close enough to preserve the performance of the *OLS* method, at least when the suggested *SN* approach to tier-one estimation is applied.

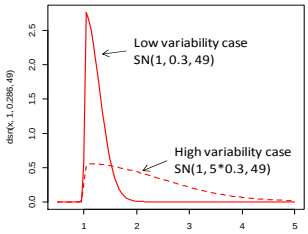


Figure 7.14. Comparison of probability densities for the High-Low and High-High scenarios.

(iv) High-High Scenario: As in the High-Low scenario, the *GLM* method produces the best average results for MSE and bias, followed closely by the *S*-estimation method. The *S*- method, however, once again produces the best results far more often than all of the other approaches. Nevertheless, while the frequency plots in Figure 7.13(d) show the same exponential appearance for the *S*-estimation approach, particularly for *MSE*, which suggests a much greater likelihood of low results, the curves for the *GLM* approach suggest a much greater degree of predictability. That is, while this method rarely achieves the low levels for either *MSE* or bias that the S-method can achieve, 98.5% of the iterations resulted in an *MSE* between 0.01 and 0.5 and 99.5% resulted in a bias within the same range. Couple this with the fact that, based on the results in Figure 7.12 and the plots in Figure 7.13(d), the *S*- method achieves *MSE* values as high as 20 (compared to 1 for the *GLM* approach), and that the *S*- method is also ~17 times more

likely than the *GLM* approach to achieve the worst *MSE*, makes the *GLM* approach preferable under these conditions.

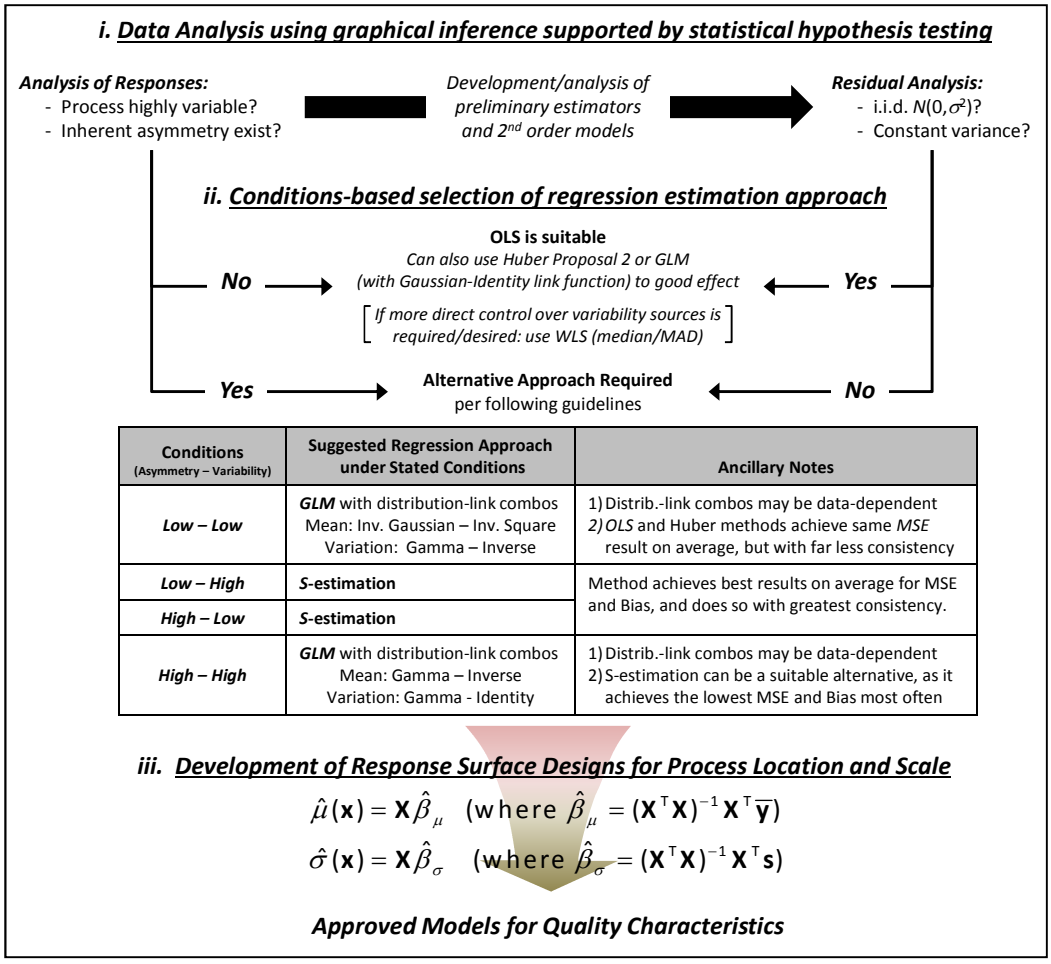
7.4 Summary of Findings

The numerical results in each of the case studies illustrate several key insights for solving the *RPD* problem in asymmetric and highly variable conditions, which are summarized in (i)-(iii) below:

(i) Most importantly, the simulation results across all scenarios clearly demonstrate that as process variability increases, alternative approaches to the traditional *OLS* method are not only necessary, but preferable. When coupled with asymmetric conditions, the effects become even more pronounced, particularly when the levels of both conditions are high. The key question is why? As previously discussed, once elevated degrees of variability and inherent asymmetry shift the data from assumed normality, the performance of traditional approaches to estimation suffers as a result of the influence exerted by extreme observations from the long tail of the skewed distribution. The alternative methods examined (namely the *GLM*, *S*-, *LTS*, and *WLS* (median/*MAD*)) tend to overcome those influences most effectively. As the results have shown, the *GLM* approach tended to perform very well, if not best, in all of the examined scenarios. But it is important to recognize that this is predicated on the identification of the right distribution-link combination, which is data-dependent and so constitutes another required step in the application of that particular method. However, viable alternatives to this are the *WLS* (median/*MAD*), *LTS*, and *S*-estimation methods, which also performed markedly better than traditional *OLS* and *WLS* approaches in high variability and high

asymmetry-high variability situations. Thus, in view of the aims of this paper, the pressing question for engineers is which approach to use and when. Based on the analysis of the presented results, the answer is depicted in Figure 7.15, which shows the modification to Phase Ib of the original process map (from Figure 7.1).

Two additional points should be made. First, some might suggest that high variability should not pose an issue, as it could be overcome by simply increasing the sample size required for estimation. Added replication at each experimental design point could ameliorate potential issues and would be preferred. However, this is often not feasible due to time and cost constraints, as well as other limitations on resources required for experimentation. Second, the results obtained in both numerical examples denote performance trends rather than definitive conclusions as to the certainty of one estimator's performance versus another's. Albeit, what they do demonstrate is that when elevated degrees of process variability and asymmetry exist, estimator selection matters in terms of achieving the best *RPD* solution. This echoes the importance of a detailed analysis in the early stages of experimentation to ascertain the degree to which such underlying conditions exist in the data, which in turn will influence the selection of the most appropriate estimation approach to use for response surface modeling and optimization.



Conditions (Asymmetry – Variability)	Suggested Regression Approach under Stated Conditions	Ancillary Notes
Low – Low	***GLM*** with distribution-link combos Mean: Inv. Gaussian – Inv. Square Variation: Gamma – Inverse	1) Distrib.-link combos may be data-dependent 2) *OLS* and Huber methods achieve same *MSE* result on average, but with far less consistency
Low – High	***S*-estimation**	Method achieves best results on average for MSE and Bias, and does so with greatest consistency.
High – Low	***S*-estimation**	
High – High	***GLM*** with distribution-link combos Mean: Gamma – Inverse Variation: Gamma - Identity	1) Distrib.-link combos may be data-dependent 2) S-estimation can be a suitable alternative, as it achieves the lowest MSE and Bias most often

Figure 7.15. Conditions-based selection guidelines for regression estimation approaches in asymmetric and/or high-variability process conditions.

(ii) Use of the skew normal distribution facilitates more accurate modeling of the inherent distributional properties associated with a particular set of data and, based on the results of Case Study A, can produce better *RPD* results in terms of minimal bias and variability. Most notably, because normality is a special case, the skew normal can very easily capture both symmetric and asymmetric properties, and thereby account for the presence of either in process outputs. Thus, by using the first three moments to replicate experimental observations in each Case Study, we were able to more accurately portray

process characteristics such as skewness, regardless of how minute, than otherwise would have been possible had we assumed normality. This is very important, as elevated variability and asymmetry are, in reality, quite probable in many industrial and manufacturing processes. Hence, use of the skew normal distribution provides the capability to model either situation simultaneously and thereby allows for a more accurate accounting of innate system properties.

(iii) The ease and explicitness associated with the *OLS* approach has helped to solidify its position as the basis for regression estimation for more than two hundred years; and it continues to see the preponderance of use throughout the literature and in applied statistics texts. Moreover, what tend to steer engineers away from considering realistic process conditions (i.e., asymmetry) and many alternative estimation methods are the computational complexities associated with them. But with today's high-speed computing power and myriad readily-available software platforms such as R, the computational complexity of alternative estimation methods should no longer be avoided. As our results show, these methods can make a significant difference in the quality of the results achieved when certain conditions exist. But the reality is that these conditions actually exist more in practice than otherwise; and when they do, the necessary assumptions that underpin *OLS* regression no longer hold. If used in spite of this reality, the *OLS* method may likely yield suboptimal solutions.

7.5 Concluding Remarks

High variability and asymmetry are conditions that occur quite often across a broad range of industrial applications and so must be given special consideration in the

experimental process regarding the selection of appropriate approaches to response surface estimation. To that end, the focus of this chapter has been to examine various robust regression approaches and specific implementations of GLM approaches as possible alternatives to *OLS* regression in the *RPD* framework such conditions prevail. The results and analysis demonstrate that, as process conditions evolve (i.e., variability and/or asymmetry increase), the estimator selection process should evolve, as well, to achieve the best solutions possible. In particular, the results have shown that the *GLM*, *S-*, *LTS*, and median-based *WLS* methods tend to yield the best *RPD* solutions. While such methods are fairly well-known in statistical circles, their use by engineers in robust design applications is comparatively rare. As Hasenkamp, et al. (2009) noted:

> "The majority of past [robust design methodology] research has traditionally been carried out by statisticians targeting an audience with good insights on statistics. When, instead, targeting engineers with less statistical knowledge as the major audience, clearly other demands are put on guidelines and tools."

One such demand is a clearer understanding of which tools to use and when. The methodology and analysis offered in this chapter, coupled with the analysis and results from Chapter 5, should help to correct this disparity by providing engineers with some clarification as to which estimation approaches will tend to provide the best *RPD* solution when certain conditions exist.

The analysis in this chapter is based upon controlled experimentation, the replication of observations made on a specified quality characteristic of interest under highly variable and asymmetric conditions, and the implementation of the skew normal distribution to effectively model both symmetric and asymmetric instances. Future

research may expand the investigation to include additional conditions, as well as processes that involve multiple quality characteristics of interest. Additionally, the development of skew normal-based link functions for use with *GLM* approaches would also add benefit. In the end, proper accounting for the inherent conditions in the data will allow engineers to more accurately model the processes they endeavor to optimize, which will invariably translate to better *RPD* solutions and more reliable recommendations to decision makers.

CHAPTER EIGHT

ACHIEVING COST ROBUSTNESS IN PROCESSES WITH MIXED MULTIPLE QUALITY CHARACTERISTICS AND DYNAMIC VARIABILITY

8.1 Introductory Remarks

The deteriorating economic conditions over the past several years increasingly threaten the economic viability of many companies. Recognizing that efficiencies in production and product quality tend to drive profits and costs, many companies are seeking alternative means for reducing costs without surrendering quality in order to ensure their survival. One of the more common techniques for achieving high quality at minimal cost is to apply the principles of robust parameter design to identify the ideal setting for the process mean among various quality characteristics or responses while simultaneously minimizing the degree of variability associated with each. In this particular case, the basic objective is to determine the process parameter settings that will minimize total expected cost for the quality characteristics.

As noted in Chapters 4 and 5, a variety of assumptions are made on the data to facilitate analysis. In the case of the optimal process mean component of the *RPD* problem, most studies to data have presumed a known (and constant) process mean and variance among an array of other parameters. Additionally, much of the efforts aimed at addressing this problem have focused on situations involving either a single response or multiple responses of the same type. However, prior knowledge of process parameters is rare in actual industrial settings, particularly because the process distribution may

expand, contract, or otherwise shift due to chance or assignable causes. Furthermore, just as most customers do not judge a product based on a single quality characteristic, it is equally unlikely that the multiple characteristics they consider are of the same type. Consider, for instance, a ballistic armor plate for military ballistic armor vests. Three characteristics for the plate might very well be the weight (an S-type characteristic), the ballistic strength (an L-type characteristic), and the thickness (an N-type characteristic). Thus, given smaller-the-better (*S*-type), nominal-the-best (*N*-type), and larger-the-better (*L*-type) characteristics, many production systems seek optimal conditions in terms of a combination of such characteristics. In particular, for a bivariate mixed-multiple case, the problem may involve the joint effects of different characteristic types, such as *SN*, *SL*, or *LN*. Furthermore,

Acknowledging these circumstances, the question becomes one of how to identify the optimal process parameter settings under the more realistic conditions of multiple mixed-type characteristics and dynamic process variability. Adding to this, how do we achieve robustness, not only in terms of the quality characteristics themselves, but in the context of *cost minimization*, as well? The obvious implication is that if process variability changes over time, then the optimal process target vector associated with minimum total cost will shift with it. If manufacturers are going to maintain a competitive edge, they will need to understand how to adapt their procedures to maintain the most profitable process target settings.

This chapter explores these questions, proposing a methodology that addresses the realities of mixed multiple characteristics and process variability in order to provide a

more robust solution in terms of minimizing total costs. In the following paragraphs, a proposed methodology is outlined for solving the process target problem under the conditions heretofore described. Thereafter, numerical examples are used to facilitate illustration of the approach and the associated benefits. The original work associated with this research is published, with reference Boylan and Goethals (2011).

8.2 Methodology Development

This section comprises a proposed methodology for solving the optimal process mean problem involving mixed multiple quality characteristics. Upon selecting a model to represent the quality characteristics of interest, an appropriate loss function and cost structure are delineated. Thereafter, an optimization framework is applied to suggest the ideal location of the process mean vector.

8.2.1. Selecting a Model for Quality Characteristic Representation

The distribution selected to model the quality characteristics of interest should account for the asymmetry typically observed in *S*- and *L*-type responses, as well as the symmetry of the *N*-type response. Suppose that the process characteristics form a *w*-dimensional vector $\mathbf{Y} = (Y_1, Y_2, ..., Y_w)^T$ from a *MSN* distribution with location vector $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_w)^T$, covariance matrix $\boldsymbol{\Lambda}$, and skewness vector $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_w)^T$. Furthermore, let $\mathbf{y} = (y_1, y_2, ..., y_w)^T$ be the vector of responses observed for the process. The form of the multivariate distribution developed by Azzalini and Capitanio (1999) is utilized to model the process. Given that ϕ_w is the probability density function for the *w*-dimensional normal distribution and $\Phi(\cdot)$ is the corresponding cumulative distribution

function, this distribution can be further defined in terms of the error function (erf), as indicated in Equation (8.1) below:

$$f_{\mathbf{Y}}(\mathbf{y}) = 2 \cdot \frac{1}{|\boldsymbol{\Lambda}|^{1/2}} \cdot \left[\phi_w\left(\boldsymbol{\Lambda}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\right)\right] \Phi\left[\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\right] \tag{8.1}$$

$$= 2 \cdot \frac{1}{|\boldsymbol{\Lambda}|^{1/2}} \left[\frac{1}{(2\pi)^{w/2}} \exp\left\{-\frac{1}{2}\left\|\boldsymbol{\Lambda}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\right\|^2\right\}\right] \cdot \frac{1}{2}\left[1+\operatorname{erf}\left\{\frac{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})}{\sqrt{2}}\right\}\right]$$

$$= \left[\frac{1}{(2\pi)^{w/2}|\boldsymbol{\Lambda}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\}\right] \cdot \left[1+\operatorname{erf}\left\{\frac{\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})}{\sqrt{2}}\right\}\right]$$

where $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_w \end{bmatrix}$, $\boldsymbol{\Lambda} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1w} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2w} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{w1} & \sigma_{w2} & \cdots & \sigma_w^2 \end{bmatrix}$ with $\sigma_{ij} = \sigma_{ji}\ \forall i \neq j$, and $\boldsymbol{\gamma} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_w \end{bmatrix}$

In the design of the optimization framework, the process mean $E[Y_q]$ for the qth characteristic is assumed to lie within a given specification region, i.e. $LSL_q \leq E[Y_q] \leq USL_q$, constrained by the lower and upper specification limits, *LSL* and *USL*, respectively. In order to simplify the formulas for the mean and variance of the *MSN* distribution, an additional parameter vector $\boldsymbol{\lambda}$ is typically used, as shown in Equation (8.2). Using the moment generating function provided by Azzalini and Capitanio (1999), we may calculate the mean and variance of the *MSN* distribution as follows

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \boldsymbol{\lambda}\sqrt{\frac{2}{\pi}} \quad \text{and} \quad \operatorname{Var}[\mathbf{Y}] = \boldsymbol{\Lambda} - \frac{2}{\pi}\boldsymbol{\lambda}\boldsymbol{\lambda}^{\mathrm{T}}, \text{ where } \boldsymbol{\lambda} = \frac{\boldsymbol{\Lambda}\boldsymbol{\gamma}}{\sqrt{1+\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{\gamma}}} \tag{8.2}$$

8.2.2. Identifying an Appropriate Quality Loss Function

A number of researchers have used the multivariate quadratic loss function in examining the optimal process mean problem, including Kapur and Cho (1996),

Teeravaraprug and Cho (2002), and Chan and Ibrahim (2004). Using the quality loss coefficients for the qth characteristic, k_{qq}, and between the qth and jth characteristics, k_{qj}, a penalty is imposed for observations deviating from the w-dimensional target vector $\boldsymbol{\tau} = (\tau_1, \tau_2, ..., \tau_w)^{\mathrm{T}}$. Using the positive definite $w \times w$ quality loss coefficient matrix $\mathbf{K}$, we may then express the multivariate quadratic loss function $L(\mathbf{Y}, \boldsymbol{\tau})$ in the form:

$$L(\mathbf{Y},\boldsymbol{\tau}) = (\mathbf{Y}-\boldsymbol{\tau})^{\mathrm{T}}\mathbf{K}(\mathbf{Y}-\boldsymbol{\tau}), \text{ with } \mathbf{K} = \begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1w} \\ k_{12} & k_{22} & \cdots & k_{2w} \\ \vdots & \vdots & \ddots & \vdots \\ k_{1w} & k_{2w} & \cdots & k_{ww} \end{bmatrix} \tag{8.3}$$

where the diagonal elements of $\mathbf{K}$ represent the penalties for each respective characteristic being off-target, while the off-diagonal elements represent the penalties for two characteristics being off-target simultaneously. Teeravaraprug and Cho (2002) discuss the necessary conditions for the quality loss coefficients such that convexity in the loss function is maintained. Since $\mathbf{Y} \sim MSN(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\gamma})$ as in Equation (8.2), we can derive the expected loss in quality using the mean and variance previously established:

$$\begin{aligned} E[L(\mathbf{Y},\boldsymbol{\tau})] &= E[(\mathbf{Y}-\boldsymbol{\tau})^{\mathrm{T}}\mathbf{K}(\mathbf{Y}-\boldsymbol{\tau})] \\ &= \mathrm{Tr}\{\mathbf{K}[\mathrm{Var}(\mathbf{Y})]\} + [E(\mathbf{Y})-\boldsymbol{\tau}]^{\mathrm{T}}\mathbf{K}[E(\mathbf{Y})-\boldsymbol{\tau}] \\ &= \mathrm{Tr}\{\mathbf{K}[\boldsymbol{\Lambda} - \frac{2}{\pi}\boldsymbol{\lambda}\boldsymbol{\lambda}^{\mathrm{T}}]\} + \left[\boldsymbol{\mu} + \boldsymbol{\lambda}\sqrt{\frac{2}{\pi}} - \boldsymbol{\tau}\right]^{\mathrm{T}} \mathbf{K} \left[\boldsymbol{\mu} + \boldsymbol{\lambda}\sqrt{\frac{2}{\pi}} - \boldsymbol{\tau}\right] \end{aligned} \tag{8.4}$$

When specifically considering the bivariate case ($w = 2$), we can expand Equation (8.3) to:

$$\begin{aligned} L(\mathbf{Y},\boldsymbol{\tau}) &= (\mathbf{Y}-\boldsymbol{\tau})^{\mathrm{T}}\mathbf{K}(\mathbf{Y}-\boldsymbol{\tau}) \\ &= \begin{bmatrix} y_1-\tau_1, y_2-\tau_2 \end{bmatrix} \begin{bmatrix} k_{11} & k_{12} \\ k_{12} & k_{22} \end{bmatrix} \begin{bmatrix} y_1-\tau_1 \\ y_2-\tau_2 \end{bmatrix} = k_{11}(y_1-\tau_1)^2 + 2k_{12}(y_1-\tau_1)(y_2-\tau_2) + k_{22}(y_2-\tau_2)^2, \end{aligned}$$

where the expected loss becomes:

$$E[L(\mathbf{Y},\boldsymbol{\tau})] = E[(\mathbf{Y}-\boldsymbol{\tau})^{\mathrm{T}}\mathbf{K}(\mathbf{Y}-\boldsymbol{\tau})] \tag{8.5}$$

$$= \mathrm{Tr}\left\{\begin{bmatrix} k_{11} & k_{12} \\ k_{12} & k_{22} \end{bmatrix}\left(\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} - \frac{2}{\pi}\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}\begin{bmatrix} \lambda_1, \lambda_2 \end{bmatrix}\right)\right\} + \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}\cdot\sqrt{\frac{2}{\pi}} - \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix}\right)^{\mathrm{T}}\begin{bmatrix} k_{11} & k_{12} \\ k_{12} & k_{22} \end{bmatrix}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}\cdot\sqrt{\frac{2}{\pi}} - \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix}\right)$$

$$= k_{11}\left\{\sigma_1^2 - \frac{2}{\pi}\lambda_1^2 + \left[\mu_1 + \lambda_1\sqrt{\frac{2}{\pi}} - \tau_1\right]^2\right\} + 2k_{12}\left\{\sigma_{12} - \frac{2}{\pi}\lambda_1\lambda_2 + \left[\mu_1 + \lambda_1\sqrt{\frac{2}{\pi}} - \tau_1\right]\left[\mu_2 + \lambda_2\sqrt{\frac{2}{\pi}} - \tau_2\right]\right\}$$

$$+k_{22}\left\{\sigma_2^2 - \frac{2}{\pi}\lambda_2^2 + \left[\mu_2 + \lambda_2\sqrt{\frac{2}{\pi}} - \tau_2\right]^2\right\}$$

In this form, the average loss of quality loss takes into account the position of the mean for both an asymmetric and symmetric process. For instance, note that if we assume zero skewness for the bivariate process (i.e. $\boldsymbol{\gamma} = \mathbf{0}$, and hence $\boldsymbol{\lambda} = \mathbf{0}$) the average loss in Equation (8.5) reduces to the familiar formula noted in previous research examining the multiple-*N*-type problem:

$$E[L(\mathbf{Y},\boldsymbol{\tau})] = E[(\mathbf{Y}-\boldsymbol{\tau})^{\mathrm{T}}\mathbf{K}(\mathbf{Y}-\boldsymbol{\tau})]$$

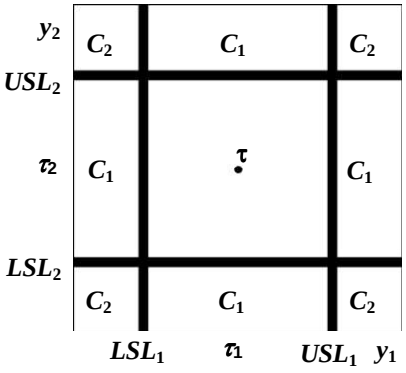
$$= k_{11}\{\sigma_1^2 + [\mu_1 - \tau_1]^2\} + 2k_{12}\{\sigma_{12} + [\mu_1 - \tau_1][\mu_2 - \tau_2]\} + k_{22}\{\sigma_2^2 + [\mu_2 - \tau_2]^2\}$$

For bivariate mixed characteristics, one or both of the quality characteristics may exhibit some level of skewness depending on its type. Whereas Equation (8.5) denotes the average loss when both Y_1 and Y_2 are asymmetric, such as with the *SL* mixed bivariate characteristic model, the expansion in Equation (8.6) denotes the average loss for cases where only one of the characteristics (Y_1) is asymmetric, such as with *SN* and *LN* models.

$$E[L(\mathbf{Y},\boldsymbol{\tau})] = k_{11}\left\{\sigma_1^2 - \frac{2}{\pi}\lambda_1^2 + \left[\mu_1 + \lambda_1\sqrt{\frac{2}{\pi}} - \tau_1\right]^2\right\} + 2k_{12}\left\{\sigma_{12} + \left[\mu_1 + \lambda_1\sqrt{\frac{2}{\pi}} - \tau_1\right]\left[\mu_2 - \tau_2\right]\right\} + k_{22}\left\{\sigma_2^2 + \left[\mu_2 - \tau_2\right]^2\right\} \tag{8.6}$$

8.2.3. Identifying an Appropriate Cost Structure

Manufacturing costs (*MC*) typically arise when observed values fail to meet pre-defined specification limits. These costs may take the form of additional labor to rework a product, service costs for the discard of material, or the cost to purchase additional materials. The degree of cost for processing may depend on a particular characteristic or even a specification limit that is violated. Hence, it is not uncommon to find processes where a trade-off of processing costs exists. Many researchers examining the multivariate *N*-type model, such as Chan and Ibrahim (2004), considered a symmetric cost structure, whereby the processing costs depend on the number of non-conforming specification limits. Using this structure, the number of potential nonconformance costs $C_1, C_2, ..., C_w$, is equal to the number of quality characteristics. Given that $R_1, R_2, ..., R_w$, represent the various regions for which a particular non-conforming cost is applied, a symmetric cost structure usually applied to the bivariate *N*-type model is shown in Figure 8.1.



$$R_1 = \{(Y_1,Y_2): [(Y_1 < LSL_1)\cap(LSL_2 \le Y_2 \le USL_2)]\cup [(Y_1 > USL_1)\cap(LSL_2 \le Y_2 \le USL_2)]\cup [(LSL_1 \le Y_1 \le USL_1)\cap(Y_2 < LSL_2)]\cup [(LSL_1 \le Y_1 \le USL_1)\cap(Y_2 > USL_2)]\}$$

$$R_2 = \{(Y_1,Y_2): [(Y_1 < LSL_1)\cap(Y_2 < LSL_2)]\cup [(Y_1 > USL_1)\cap(Y_2 > USL_2)]\cup [(Y_1 > USL_1)\cap(Y_2 < LSL_2)]\cup [Y_1 < LSL_1)\cap(Y_2 > USL_2)]\}$$

Figure 8.1. Bivariate *N*-type model cost structure (symmetric nonconformance).

For mixed multiple characteristics, unlike the multivariate *N*-type model, at least one of the characteristics possesses a one-sided limit with respect to the tolerance region and the target vector $\boldsymbol{\tau}$ is positioned based upon the characteristic type. The symmetry of the cost structure depends on the priorities associated with an observation nonconforming to a set of characteristics. For instance, Figure 8.2 shows the *S* and *L*-type bivariate model where (i) the costs of nonconformance are identical for Y_1 and Y_2, and (ii) where the cost of nonconforming to Y_1 may be greater or less than Y_2.

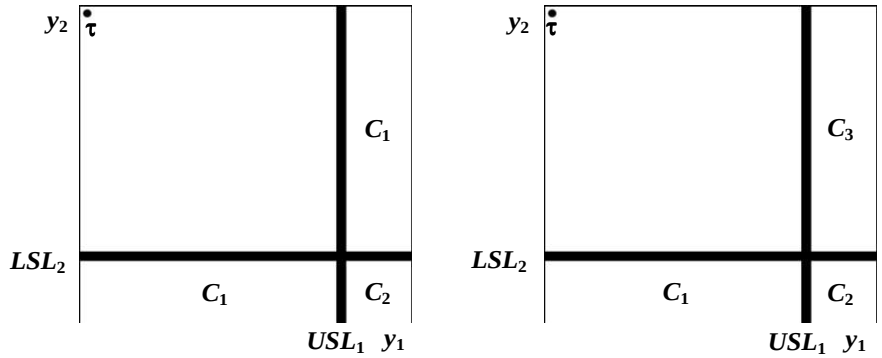


Figure 8.2. Bivariate *S* and *L*-type model cost structures for (i) symmetric nonconformance, and (ii) asymmetric nonconformance.

To provide a generalization among the various symmetric and asymmetric cost structures that can result, we assume a total of *v* regions $R_1, R_2, ..., R_v$ exist, whereby up to *u* costs of nonconformance $C_1, C_2, ..., C_u$, $u \leq v$, may apply. Since we desire to minimize the expected cost of nonconformance, $E[NC]$, which depends on the probability of an observation lying in the *m*th region, $\Pr[R_m]$, these cost constraints must be incorporated into the overall optimization framework.

Finally, when working specifically with *S* and *L*-type characteristics, it is common to include an additional cost beyond the penalties of nonconformance or deviating from a

target value. The *S*- and *L*-type characteristic target values are frequently established at 0 and ∞, respectively, making them intrinsically more difficult to achieve. At the cost of adding new technologies or more labor, however, it may be possible to reduce the process bias in the characteristic observations substantially. For this reason, a number of researchers have incorporated a production cost function, $P(\mathbf{Y}, \boldsymbol{\tau})$, to account for this constraint. The form of $P(\mathbf{Y}, \boldsymbol{\tau})$ can vary widely depending on the manufacturing application. Chase and Parkinson (1991), Shin *et al*. (2005), as well as Chen and Huang (2011), used linear functions to represent the cost of achieving tighter tolerances in various production systems. Many researchers, such as Michael and Siddall (1981), Fang and Wu (2000), Shin and Cho (2007), and Peng *et al*. (2008), have employed a form of the exponential cost function to serve this same purpose. Particularly when solving problems using *S*- and *L*-type characteristics, any hybrid of the exponential cost functions previously examined may be useful. Equation (8.7) displays the general form of a target-based exponential production cost for *S*- and *L*-type characteristics, respectively:

$$P_S(\mathbf{Y}, \boldsymbol{\tau}) = \sum_{q=1}^{w} c_q \exp\{-\theta\left[y_q - \tau_q\right]\} \text{ and } P_L(\mathbf{Y}, \boldsymbol{\tau}) = \sum_{q=1}^{w} c_q \exp\{\theta\left[y_q - \tau_q\right]\}, \tag{8.7}$$

where c and θ are constant coefficients established based upon the perceived difficulty in the characteristic observations achieving a desired target value. While these functions are primarily used in assessing production costs associated with *S*- and *L*-type characteristics, we can also formulate these equations to account for observations made on *N*-type characteristics.

8.2.4. Applying an Optimization Framework

When seeking solutions to the optimal process mean problem, we typically consider the costs associated with product material waste, processing, and a loss in product quality. The integration of the multivariate quadratic loss function from Section 8.2.2 enables the manufacturer to penalize a process for any deviation from a desired target vector. Section 8.2.3 accounts for the expected loss corresponding to a given nonconformance cost structure $E[NC]$ in order to reduce processing costs corresponding to products that must be discarded or reworked. Similar to previous research examining the optimal process mean problem, the generation of solutions requires several assumptions. First, known and fixed nonconformance costs C for a process are assumed, recognizing that we would have to reapply the optimization framework if these costs are adjusted at some future point in time. We also presume to know the quality loss coefficients k for $L(\mathbf{Y}, \boldsymbol{\tau})$, reflecting the priority given to maintaining a particular characteristic on target.

Unlike previous research involving multiple characteristics, however, a known measure of the variance and covariance is *not* assumed, as it is more realistic that an engineer will have some knowledge of the variability range for a characteristic rather than a specific setting at a given time. Thus, the problem is formulated utilizing this range of process variability. Thereafter, an iterative programming routine is implemented using random number generators for each range of variability to identify the resulting position of the optimal process mean vector, $E[\mathbf{Y}^*]$, that minimizes the total cost (TC). The end result is the identification of a potential range in the shift of $E[\mathbf{Y}^*]$ depending on

the feasible variability conditions that could arise. The formulation of the optimization framework used to generate an array of acceptable solutions for mixed multiple characteristic problems follows:

Given:

1. Range of variability for the quality characteristics of interest:
$\alpha_1 \le \sigma_1^2 \le \beta_1,\ \alpha_2 \le \sigma_2^2 \le \beta_2, ..., \alpha_w \le \sigma_w^2 \le \beta_w,\ \alpha_{12} \le \sigma_{12} \le \beta_{12},\ \alpha_{13} \le \sigma_{13} \le \beta_{13}, ..., \alpha_{ij} \le \underset{i<j}{\sigma_{ij}} \le \beta_{ij}, ...,$
$\alpha_{w-1,w} \le \sigma_{w-1,w} \le \beta_{w-1,w}$
2. Observations made on $\mathbf{Y}$ that form a w-dimensional vector, $\mathbf{y} = (y_1, y_2, ..., y_w)^{\mathrm{T}}$
3. Desired target vector, $\boldsymbol{\tau} = (\tau_1, \tau_2, ..., \tau_w)^{\mathrm{T}}$
4. Skewness vector, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_w)^{\mathrm{T}}$
5. Quality loss coefficient matrix, $\mathbf{K}$
6. Tolerance region, T, defined by $LSL_q \le y_q \le USL_q$, for $q = 1, 2, ..., w$
7. Nonconformance costs $C_1, C_2, ..., C_u$ for S, N, or L-type characteristics and production cost functions $P(\mathbf{Y}, \boldsymbol{\tau})$ for S or L-type characteristics

For: Iterate $n = 1, 2, ..., \varphi$, where $\sigma_1, \sigma_2, ..., \sigma_w, \sigma_{12}, \sigma_{13}, ..., \sigma_{ij}, ..., \sigma_{w\text{-}1,w}$ are randomly generated,

Minimize: $E[TC] = E[NC] + E[L(\mathbf{Y}, \boldsymbol{\tau})] + E[P(\mathbf{Y})]$

$$= \sum_{m=1}^{u} C_m \Pr[R_m] + \int_{LSL_w}^{USL_w} \int_{LSL_{w-1}}^{USL_{w-1}} \cdots \int_{LSL_1}^{USL_1} L(\mathbf{Y}, \boldsymbol{\tau}) \cdot f_{\mathbf{Y}}(\mathbf{y}) \mathrm{d}y_1 \mathrm{d}y_2 \cdots \mathrm{d}y_w$$

$$+ \int_{LSL_w}^{USL_w} \int_{LSL_{w-1}}^{USL_{w-1}} \cdots \int_{LSL_1}^{USL_1} P(\mathbf{Y}, \boldsymbol{\tau}) \cdot f_{\mathbf{Y}}(\mathbf{y}) \mathrm{d}y_1 \mathrm{d}y_2 \cdots \mathrm{d}y_w, \text{ where}$$

$$f_{\mathbf{Y}}(\mathbf{y}) = \left[\frac{1}{(2\pi)^{w/2} |\boldsymbol{\Lambda}|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Lambda}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\} \right] \cdot \left[1 + \operatorname{erf} \left\{ \frac{\boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{\Lambda}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})}{\sqrt{2}} \right\} \right],$$

with $\mathbf{y} \in \Re^w$, $L(\mathbf{Y}, \boldsymbol{\tau}) = (\mathbf{Y} - \boldsymbol{\tau})^{\mathrm{T}} \mathbf{K} (\mathbf{Y} - \boldsymbol{\tau})$ and $P(\mathbf{Y}, \boldsymbol{\tau}) = \sum_{q=1}^{w} c_q \exp\{\pm\theta [y_q - \tau_q]\}$

Subject to: $E[\mathbf{Y}] = \boldsymbol{\mu} + \boldsymbol{\lambda} \sqrt{\frac{2}{\pi}} \in T$, where $\boldsymbol{\lambda} = \frac{\boldsymbol{\Lambda}\boldsymbol{\gamma}}{\sqrt{1 + \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{\Lambda} \boldsymbol{\gamma}}}$

Find: Optimal process mean vector, $E[\mathbf{Y}^*]_n = (E[Y_{1n}^*], E[Y_{2n}^*], ..., E[Y_{wn}^*])^{\mathrm{T}}$

Iterating the optimization routine a significant number of times (φ) results in an optimal process mean region that corresponds to the estimated range of variability. Using the multivariate uniform distribution, we may then generate appropriate values for σ_q and σ_{ij} at each iterate. For the univariate case, an interval for the optimal process mean is approximated, whereas for the multivariate case, the region for the optimal process mean vector is cuboidal. Depending on the specific values of variability, it is likely that we can subsequently divide the interval or region for $E[\mathbf{Y}^*]$ into low and high variability densities. In turn, a manufacturer or engineer might use this tool to provide greater prediction in the setting of $E[\mathbf{Y}^*]$, given that certain production stages or cycles are inherently known to possess lower or higher process variability.

With the trade-off of nonconformance costs, production costs, and the cost of deviating from a desired target value, the objective function is typically unimodal for the process mean problem. The nonlinear optimization algorithm `nlminb()` in R, which utilizes a quasi-Newton method under bound constraints, can generate solutions to these problems in a timely and efficient manner. Given the convex function f and a vector of parameters $\mathbf{z}$ over which the minimization of f occurs, the algorithm involves an approximation using the second-order Taylor expansion about some point $\mathbf{z}_m$:

$$f(\mathbf{z}) = f(\mathbf{z}_m) + (\mathbf{z}-\mathbf{z}_m)^{\mathrm{T}}\nabla f(\mathbf{z}_m) + \tfrac{1}{2}(\mathbf{z}-\mathbf{z}_m)^{\mathrm{T}}\mathbf{H}(\mathbf{z}_m)(\mathbf{z}-\mathbf{z}_m),$$

where $\nabla f(\mathbf{z}_m)$ is the gradient of f evaluated at the point $\mathbf{z}_m$ and $\mathbf{H}$ represents the Hessian matrix of second derivatives. Since $\nabla f(\mathbf{z})$ with respect to $\mathbf{z}$ is equal to $\nabla f(\mathbf{z}_m) + \mathbf{H}(\mathbf{z}_m)(\mathbf{z}-\mathbf{z}_m)$, setting this equal to $\mathbf{0}$ and solving for $\mathbf{z}$ gives the solution $\mathbf{z} = \mathbf{z}_m - \mathbf{H}^{-1}(\mathbf{z}_m)\nabla f(\mathbf{z}_m)$. Unless $\mathbf{z}$ denotes a final solution reached within a given tolerance, a new

iterate $\mathbf{z}_{m+1}$ is found using a pre-defined step size ξ, $\mathbf{z}_{m+1} = \mathbf{z}_m - \xi\mathbf{H}^{-1}(\mathbf{z}_m)\nabla f(\mathbf{z}_m)$. In the context of the "nlminb" optimization algorithm, the default minimum step size is ξ = 2.2e-14 and the default tolerance for convergence is 1e-20.

8.3 Numerical Example

In order to illustrate the proposed methodology, a modification of an experiment examined by Kovach and Cho (2008) is used, whereby the objective is to determine the optimal process mean vector for mixed multiple characteristics. Their experiment considered a chemical filtration process, wherein dosages were prepared within vials and measurements were taken with regard to the purity of each sample. The cost of processing each vial sample was found to be directly related to the quantities of three different quality characteristics – the filtration time (Y_1) measured in seconds, the filtration volume (Y_2) measured in milliliters (mL), and the filtration purity (Y_3) measured as a percentage. Since the system objectives are to minimize filtration time and maximize purity while obtaining an ideal filtration volume of 10 mL, Y_1, Y_2, and Y_3 may be considered *S*-, *L*- and *N*-type characteristics, respectively. Table 8.1 displays the specification limits, nonconformance costs, and desired target values for Y_1, Y_2, and Y_3.

Table 8.1. Established conditions for the chemical filtration process.

Characteristic	Specification Limits	Nonconformance Costs*	Target
Y_1, Time (sec)	$y_1 \leq 2.5$	$C_1 = 80$	0
Y_2, Volume (mL)	$9.5 \leq y_2 \leq 10.5$	$C_{2(LSL)} = 60$, $C_{2(USL)} = 100$	10
Y_3, Purity (%)	$y_3 \geq 94$	$C_3 = 100$	100

*A cost of $C_4 = 125$ is administered for observations that fail more than one specification limit.

The very nature of minimizing or maximizing the response on a particular characteristic implies that we might observe some degree of positive or negative skewness in the distribution of observations. Many researchers, such as Vannman and Albing (2007), Feng and Kapur (2009), Abbasi (2009), Liao (2010), have attested to the non-normal tendencies of *S*- and *L*-type characteristics. Suppose multiple observations have been recorded for each of the characteristics in this example under typical process settings. Based upon these results and some knowledge of the process, the distribution of observations for Y_1 and Y_3 are found to be positively and negatively skewed with $\gamma_1 = 5$ and $\gamma_3 = -6$, respectively, whereas the distribution of observations on Y_2 has zero skewness, i.e. $\gamma_2 = 0$. It is also generally known that the range of variability for Y_1, Y_2, and Y_3, is $0.4 \leq \sigma_1^2 \leq 1.0$, $0.1 \leq \sigma_2^2 \leq 0.6$, and $1.0 \leq \sigma_3^2 \leq 3.0$, respectively, and the range of covariance is negligible at $-0.1 \leq \sigma_{ij} \leq 0.1$. Finally, the quality loss coefficient settings are established for the characteristics such that the same penalty is applied for observations deviating from their respective target value. The production costs for each characteristic, however, are defined based upon the tolerance-cost relationships for each mixed case.

In each of the following sections, the feasible range of $E[\mathbf{Y}^*]$ is identified for each pair of bivariate mixed characteristics. The effect of skewness is also examined through a comparison of bivariate normal and bivariate skew normal processes. For any one of the problems discussed in Sections 3.1, 3.2, and 3.3, a sensitivity analysis regarding the specific settings for the tolerance, cost, or loss function may be performed to gain additional information on the robustness of the optimal process mean vector.

8.3.1. Solving the Bivariate *S*- and *N*-Type Problem

Using a random number generator, the variability settings within each prescribed interval for Y_1 and Y_2 are established based upon a uniform distribution at each iterate. The nonlinear constrained optimization algorithm contained in function `nlminb()` in R is used to solve the optimal process mean vector problem, repeating the program routine 100 times to capture the full range of the shift in the $E[\mathbf{Y}^*]$. An outline of the optimization routine follows:

Given: $0.4 \le \sigma_1^2 \le 1.0$, $0.1 \le \sigma_2^2 \le 0.6$, $-0.1 \le \sigma_{12} \le 0.1$, $\boldsymbol{\tau} = \begin{bmatrix} 0 \\ 10 \end{bmatrix}$, $\boldsymbol{\gamma} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$, $\mathbf{K} = \begin{bmatrix} 0.25 & 0.15 \\ 0.15 & 0.25 \end{bmatrix}$,
with $USL_1 = 2.5$, $LSL_2 = 9.5$, $USL_2 = 10.5$, and Costs $C_1 = 80$, $C_{2(LSL)} = 60$, $C_{2(USL)} = 100$, $C_4 = 125$

Minimize:

$$E[TC] = C_1 \int_{USL_1}^{\infty} \int_{LSL_2}^{USL_2} f_{\mathbf{Y}}(\mathbf{y}) dy_2 dy_1 + C_{2(LSL)} \int_{-\infty}^{LSL_2} \int_{-\infty}^{USL_1} f_{\mathbf{Y}}(\mathbf{y}) dy_1 dy_2 + C_{2(USL)} \int_{USL_2}^{\infty} \int_{-\infty}^{USL_1} f_{\mathbf{Y}}(\mathbf{y}) dy_1 dy_2$$

$$+ C_4 \left[\int_{-\infty}^{LSL_2} \int_{USL_1}^{\infty} f_{\mathbf{Y}}(\mathbf{y}) dy_1 dy_2 + \int_{USL_2}^{\infty} \int_{USL_1}^{\infty} f_{\mathbf{Y}}(\mathbf{y}) dy_1 dy_2 \right] + \int_{LSL_2}^{USL_2} \int_{0}^{USL_1} L(y_1, y_2) f_{\mathbf{Y}}(\mathbf{y}) dy_1 dy_2$$

$$+ \int_{LSL_2}^{USL_2} \int_{0}^{USL_1} P(y_1, y_2) f_{\mathbf{Y}}(\mathbf{y}) dy_1 dy_2 ,$$

where: $f_{\mathbf{Y}}(\mathbf{y}) = \left[\frac{1}{\sqrt{2\pi}|\boldsymbol{\Lambda}|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Lambda}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\} \right] \cdot \left[1 + \operatorname{erf} \left\{ \frac{\boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{\Lambda}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})}{\sqrt{2}} \right\} \right]$, with $\mathbf{y} \in \Re^w$,

$L(y_1, y_2) = 0.25 y_1^2 + 0.25(y_2 - 10)^2 + 0.15 y_1 (y_2 - 10)$, and $P(y_1, y_2) = 500 \exp\{-5 y_1 - 0.10 y_2\}$

Subject to: $E[\mathbf{Y}] = \boldsymbol{\mu} + \boldsymbol{\lambda} \sqrt{\frac{2}{\pi}} \in T$, where $\boldsymbol{\lambda} = \frac{\boldsymbol{\Lambda}\boldsymbol{\gamma}}{\sqrt{1 + \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{\Lambda} \boldsymbol{\gamma}}}$

Find: Optimal process mean vector, $E[\mathbf{Y}^*]_n = (E[Y_{1n}^*], E[Y_{2n}^*])^{\mathrm{T}}$ for $\varphi = 100$ iterations

---

Upon completion of the program, each of the φ optimal process mean vectors is plotted so that the complete range of $E[\mathbf{Y}^*]$ is visibly apparent. Figure 8.3 (i) shows the results using the bivariate skew normal distribution whereby the tolerance region for Y_1 and Y_2, where identical contours for the process at low (–) and high variability (–) are

highlighted. The contours of the quality loss function $L(y_1, y_2)$ are shown using a dashed line (---), and the target vector is identified with a large point (•). The same legend is used in Figure 8.3 (ii), where the bivariate normal distribution is used, i.e. the distribution of observations is assumed to possess zero skewness. In addition, we observe the differences in the mean and median of the objective function values, $\overline{E[TC]}$ and $\tilde{E}[TC]$.

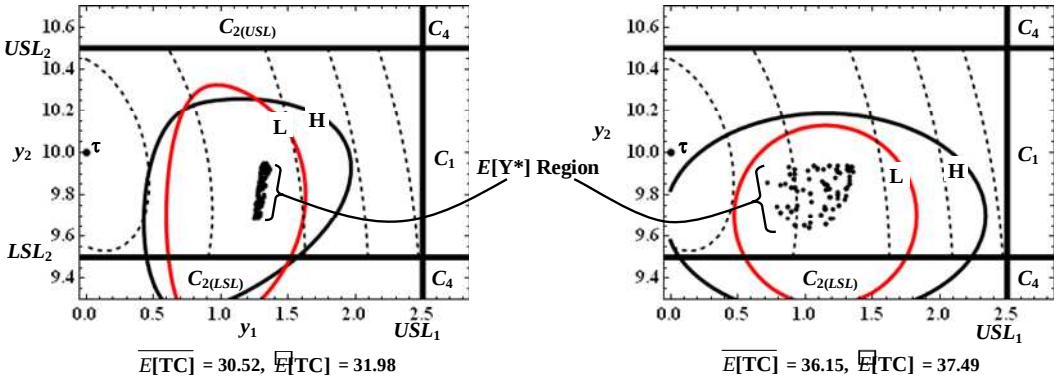


Figure 8.3. Optimization results – identical contours for **Y**, where

(i) Feasible Region for $E[\mathbf{Y}^*]$ for $SN(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\gamma})$, $\boldsymbol{\gamma} = (5, 0)^T$, from Low (L) to High (H) Variability.
(ii) Feasible Region for $E[\mathbf{Y}^*]$ for $N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, i.e. $\boldsymbol{\gamma} = (0, 0)^T$, from Low (L) to High (H) Variability.

8.3.2. Solving the Bivariate *S*- and *L*-Type Problem

In this section, the joint effect of the filtration time (Y_1) and purity (Y_3) are considered, where the distributions of observations are positively and negatively skewed, respectively. The formulation of the optimization routine for the bivariate *S*- and *L*-type process mean vector problem appears as:

Given: $0.4 \le \sigma_1^2 \le 1.0$, $1.0 \le \sigma_3^2 \le 3.0$, $-0.1 \le \sigma_{13} \le 0.1$, $\boldsymbol{\tau} = \begin{bmatrix} 0 \\ 100 \end{bmatrix}$, $\boldsymbol{\gamma} = \begin{bmatrix} 5 \\ -6 \end{bmatrix}$, $\mathbf{K} = \begin{bmatrix} 0.25 & 0.15 \\ 0.15 & 0.25 \end{bmatrix}$,
with $USL_1 = 2.5$, $LSL_3 = 94$, and Costs $C_1 = 80$, $C_3 = 100$, $C_4 = 125$

Minimize: $E[TC] = C_1\int_{LSL_3}^{\infty}\int_{USL_1}^{\infty} f_{\mathbf{Y}}(\mathbf{y})dy_1dy_3 + C_3\int_{-\infty}^{LSL_3}\int_{-\infty}^{USL_1} f_{\mathbf{Y}}(\mathbf{y})dy_1dy_3 + C_4\int_{-\infty}^{LSL_3}\int_{USL_1}^{\infty} f_{\mathbf{Y}}(\mathbf{y})dy_1dy_3$

$$+\int_{LSL_3}^{100}\int_{0}^{USL_1} L(y_1, y_3)\, f_{\mathbf{Y}}(\mathbf{y})dy_1dy_3 + \int_{LSL_3}^{100}\int_{0}^{USL_1} P(y_1, y_3)\, f_{\mathbf{Y}}(\mathbf{y})dy_1dy_3,$$

where: $f_{\mathbf{Y}}(\mathbf{y}) = \left[\frac{1}{\sqrt{2\pi}|\mathbf{\Lambda}|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{\Lambda}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\}\right]\cdot\left[1+\mathrm{erf}\left\{\frac{\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{\Lambda}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})}{\sqrt{2}}\right\}\right]$, with $\mathbf{y} \in \Re^w$,

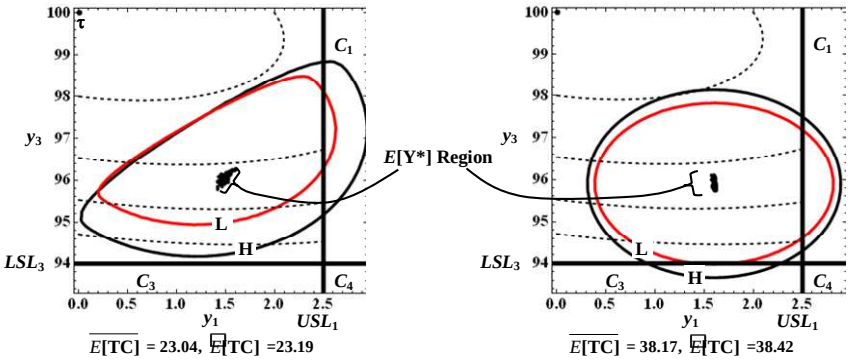
$L(y_1, y_3) = 0.25y_1^2 + 0.25(y_3 - 100)^2 + 0.15y_1(y_3 - 100)$, and

$P(y_1, y_3) = 1000[\exp\{-10y_1\} + \exp\{2(y_3 - 100)\}]$

Subject to: $E[\mathbf{Y}] = \boldsymbol{\mu} + \boldsymbol{\lambda}\sqrt{\frac{2}{\pi}} \in T$, where $\boldsymbol{\lambda} = \frac{\mathbf{\Lambda}\boldsymbol{\gamma}}{\sqrt{1+\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{\Lambda}\boldsymbol{\gamma}}}$

Find: Optimal process mean vector, $E[\mathbf{Y}^*]_n = (E[Y_{1n}^*], E[Y_{3n}^*])^{\mathrm{T}}$ for $\varphi = 100$ iterations

As in Section 8.3.1, this optimization scheme is iterated to generate solutions to the optimal process mean vector problem. Figure 8.4 (i) shows the graphical results using the bivariate skew normal distribution, which we may then compare to the results in Figure 8.4 (ii), in which the bivariate normal distribution is used as a model for both characteristics.



(i) (ii)

Figure 8.4. Optimization results – identical contours for **Y**, where

(i) Feasible Region for $E[\mathbf{Y}^*]$ for $SN(\boldsymbol{\mu}, \mathbf{\Lambda}, \boldsymbol{\gamma})$, $\boldsymbol{\gamma} = (5, -6)^{\mathrm{T}}$, from Low (L) to High (H) Variability.
(ii) Feasible Region for $E[\mathbf{Y}^*]$ for $N(\boldsymbol{\mu}, \mathbf{\Lambda})$, i.e. $\boldsymbol{\gamma} = (0, 0)^{\mathrm{T}}$, from Low (L) to High (H) Variability.

8.3.3. Solving the Bivariate L and N-Type Problem

Finally, the joint effect of the filtration volume (Y_2) and purity (Y_3) is examined whereby the distribution of observations is symmetric and negatively skewed, respectively. The formulated optimization routine for the bivariate *L*- and *N*-type characteristics is as follows:

Given: $0.1 \le \sigma_2^2 \le 0.6$, $1.0 \le \sigma_3^2 \le 3.0$, $-0.1 \le \sigma_{23} \le 0.1$, $\boldsymbol{\tau} = \begin{bmatrix} 10 \\ 100 \end{bmatrix}$, $\boldsymbol{\gamma} = \begin{bmatrix} 0 \\ -6 \end{bmatrix}$, $\mathbf{K} = \begin{bmatrix} 0.25 & 0.15 \\ 0.15 & 0.25 \end{bmatrix}$,

with $LSL_2 = 9.5$, $USL_2 = 10.5$, $LSL_3 = 94$, and Costs $C_{2(LSL)} = 60$, $C_{2(USL)} = 100$, $C_3 = 100$, $C_4 = 125$

Minimize:

$$E[TC] = C_{2(LSL)} \int_{-\infty}^{LSL_2} \int_{LSL_3}^{\infty} f_{\mathbf{Y}}(\mathbf{y}) \mathrm{d}y_3 \mathrm{d}y_2 + C_{2(USL)} \int_{USL_2}^{\infty} \int_{LSL_3}^{\infty} f_{\mathbf{Y}}(\mathbf{y}) \mathrm{d}y_3 \mathrm{d}y_2 + C_3 \int_{-\infty}^{LSL_3} \int_{LSL_2}^{USL_2} f_{\mathbf{Y}}(\mathbf{y}) \mathrm{d}y_2 \mathrm{d}y_3$$

$$+ C_4 \left[\int_{-\infty}^{LSL_2} \int_{-\infty}^{LSL_3} f_{\mathbf{Y}}(\mathbf{y}) \mathrm{d}y_3 \mathrm{d}y_2 + \int_{USL_2}^{\infty} \int_{-\infty}^{LSL_3} f_{\mathbf{Y}}(\mathbf{y}) \mathrm{d}y_3 \mathrm{d}y_2 \right] + \int_{LSL_2}^{USL_2} \int_{LSL_3}^{100} L(y_2, y_3) f_{\mathbf{Y}}(\mathbf{y}) \mathrm{d}y_3 \mathrm{d}y_2$$

$$+ \int_{LSL_2}^{USL_2} \int_{LSL_3}^{100} P(y_2, y_3) f_{\mathbf{Y}}(\mathbf{y}) \mathrm{d}y_3 \mathrm{d}y_2 ,$$

where: $$f_{\mathbf{Y}}(\mathbf{y}) = \left[\frac{1}{\sqrt{2\pi} |\boldsymbol{\Lambda}|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Lambda}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\} \right] \cdot \left[1 + \mathrm{erf} \left\{ \frac{\boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{\Lambda}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})}{\sqrt{2}} \right\} \right]$$, with $\mathbf{y} \in \Re^w$,

$L(y_2, y_3) = 0.25(y_2 - 10)^2 + 0.25(y_3 - 100)^2 + 0.15(y_2 - 10)(y_3 - 100)$, and

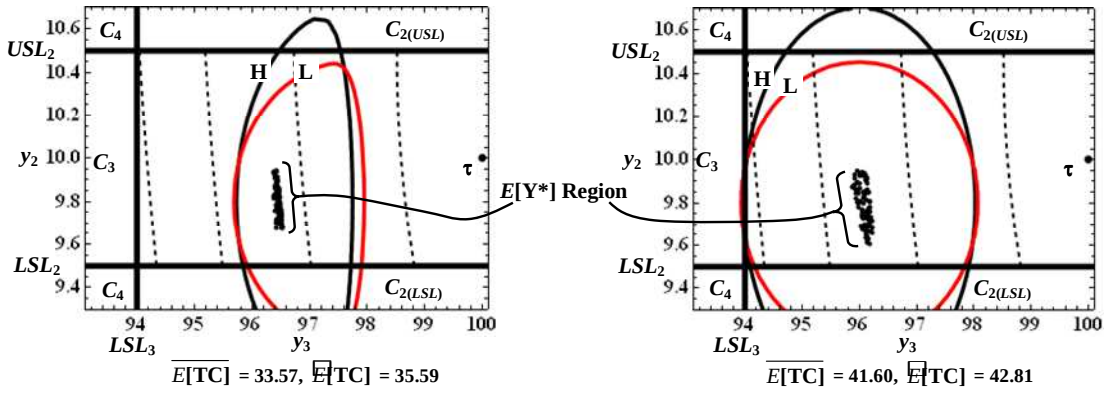
$P(y_2, y_3) = 1000 \exp\{2(y_3 - 100)\}$

Subject to: $E[\mathbf{Y}] = \boldsymbol{\mu} + \boldsymbol{\lambda} \sqrt{\frac{2}{\pi}} \in T$, where $\boldsymbol{\lambda} = \frac{\boldsymbol{\Lambda}\boldsymbol{\gamma}}{\sqrt{1 + \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{\Lambda} \boldsymbol{\gamma}}}$

Find: Optimal process mean vector, $E[\mathbf{Y}^*]_n = (E[Y_{2n}^*], E[Y_{3n}^*])^{\mathrm{T}}$ for $\varphi = 100$ iterations

---

The optimization results for the *L*- and *N*-type problem are shown graphically in Figures 8.5(i) and 8.5(ii), respectively.



(i) (ii)

Figure 8.5. Optimization results – identical contours for **Y**, where

(i) Feasible Region for $E[\mathbf{Y}^*]$ for $SN(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\gamma})$, $\boldsymbol{\gamma} = (-6, 0)^T$, from Low (L) to High (H) Variability
(ii) Feasible Region for $E[\mathbf{Y}^*]$ for $N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, i.e. $\boldsymbol{\gamma} = (0, 0)^T$, from Low (L) to High (H) Variability.

8.3.4. Summary of Results

Analysis of the numerical results in Sections 3.1, 3.2, and 3.3 leads to several conclusions regarding the proposed approach, which are summarized in (i)-(iii) below:

(i) Effect of Variability. Whether a process is considered in statistical control or not, there is likely some range of variability inherent in the distribution of characteristic observations. As a result of the trade-off of non-conformance costs and the cost to achieve a customer-defined target value, the ideal setting of the process mean is likely to shift as the variability increases or decreases. For the processes modeled in Figures 8.3 (i) and 8.4 (i), the corresponding shift in the optimal process mean vector is nearly one-fourth the width of the tolerance region. The setting of $E[\mathbf{Y}^*]$ may also be more robust to changes in variability, as is the case with the process modeled in Figure 8.5 (i).

Rather than providing an exact position of the optimal process mean vector based upon values established for the variance or covariance among characteristics, the feasible region for $E[\mathbf{Y}^*]$ offers more predictive capability for the manufacturer. It is not

unrealistic to assume that an engineer may know precisely when process variability tends to be low or high. The influence of various conditions, such as the time of day, environmental conditions, the array of operators in the workplace, seasonal demand, or a change in product output, may all affect process stability in different ways. Thus, with some knowledge of the feasible region for $E[\mathbf{Y}^*]$, the engineer is provided with a "road map" to the most profitable target setting. For instance, given the example previously outlined, dramatic changes in process variability appear to affect the target position for the filtration volume (Y_2) more than either the filtration time (Y_1) or purity (Y_3). Adjusting the setting between 9.7 mL and 9.9 mL ensures that the minimum processing costs for this problem are achieved.

(ii) Effect of Skewness. To investigate the effect of skewness on solutions to the optimal process mean vector problem, a comparison between the bivariate skew normal and bivariate normal distribution was provided. With the only modification being the distribution used to model the characteristics, solutions for the *SN*, *SL*, and *LN*-type mixed characteristic problems were generated. Pursuant to this, the mean and median of the objective function values were calculated for the 100 iterations and then used as a basis to examine the effect of skewness. For each mixed case, accounting for the asymmetry typically observed in *S*- and *L*-type characteristics resulted in a 16% to 40% reduction in the mean and median total cost. The difference is most significant in the *SL* case, in which both positive and negative skewness exists due to the nature of the characteristics.

(iii) Effect of Altering the Conditions. An analysis of the sensitivity of the optimal process mean vector can provide important information as to the robustness of the solutions. It may also suggest alternative feasible and acceptable options for reducing the overall production cost. The effects of adjusting (1) the nonconformance cost $C_{2(USL)}$, and (2) the tolerance setting for Y_1, are displayed in Figure 8.6 for the *SN*-type problem.

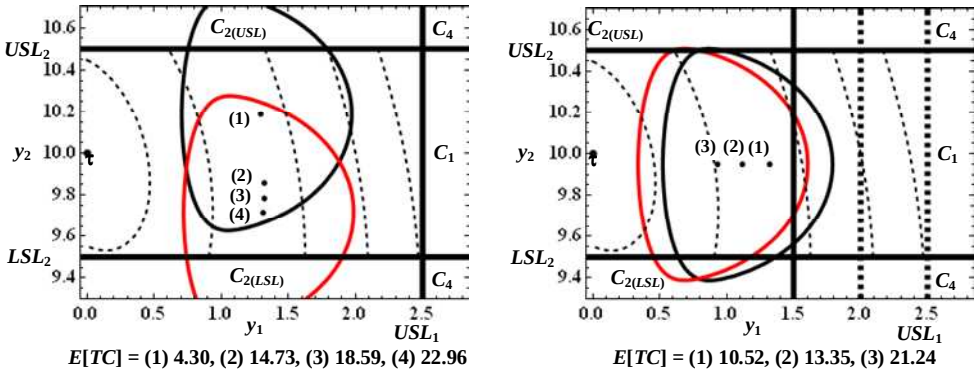


Figure 8.6. Sensitivity analysis – shift in $E[\mathbf{Y}^*]$ for altering conditions

(i) Change in Nonconformance with Y_2 *USL*, $C_{2(USL)} = 10$ to $C_{2(USL)} = 1000$
(ii) Change in Tolerance for Y_1, $USL_1 = 2.5$ to $USL_1 = 1.5$.

As observed, the expected total processing cost appears to be more robust to modifications in the tolerance for Y_1 than changes in the cost of nonconforming to the filtration volume's upper specification limit. With this information, some flexibility may be gained by the manufacturer in altering the tolerance limits.

8.4. Concluding Remarks

Within industry and experimental research settings, statistical assumptions in the analysis of observations play an important role in achieving optimal solutions. Situations involving multiple mixed-type quality characteristics and dynamic process variability

occur regularly across a wide array of industrial and manufacturing applications and therefore merit special consideration. However, under previous or more traditional research efforts associated with the optimal process mean problem, multiple characteristic problems typically only address like-type characteristics or cases that involve strictly asymmetric or symmetric responses, but very rarely both. More importantly, researchers usually assume foreknowledge of the process mean and variance prior to seeking solutions to the optimization problem. Although sound in theoretical principal, these methodologies invoke assumptions that often fail to adequately represent the realities found in industrial settings. A more realistic approach to the process mean problem is to integrate mixed-type quality characteristics, adequately addressing the expected skewness associated with each, and to incorporate a range of variability, which is far more readily available to engineers than a known constant variance.

Much remains open for exploration in this area. In particular, follow-on research might investigate multivariate cases beyond the bivariate examples provided here. While the extensions have been theorized, the more difficult work of deriving and applying them is yet to be done. Recognizing that such conditions exist as heretofore described, future work may look at employing alternative techniques to achieve greater accuracy, flexibility, and robustness in their solutions. More specifically, as the results in this chapter show, such techniques will allow us to accurately predict the location of the optimal process mean vector as it shifts in response to changes in process variability. Such knowledge of a feasible region can help to achieve cost robustness by providing a quantitative “road map” for maintaining the most profitable process target settings.

CHAPTER NINE

DEVELOPING AN ALTERNATIVE STRATEGY FOR OVERCOMING SOURCES OF VARIABILITY TO ACHIEVE ENHANCED PRECISION

9.1 Introduction

Contemporary efforts in engineering design, manufacturing engineering and quality control/assurance, embrace variability reduction as a primary means to improve product quality in terms of achieving performance targets while reducing defects. In fact, many firms today seek as a primary engineering goal the continuous and systematic reduction of variability across key process and product dimensions. Take, for example, the development and procurement of body armor for use by military personnel. According to the U.S. Army's Board on Army Science and Technology and the National Research Council (see Board on Army Science and Technology *et al.*, 2012),

> "Uncertainty and variation in the manufacture, testing, and employment of body armor, as well as the natural concern for protecting personnel, tend to result in conservative decision-making, which in turn can result in body armor overdesign and/or overmanufacture."

In short, variation in manufacturing process can result in armor plates that are heavier and perhaps larger than required to reduce the chances of producing plates that fail to conform to established specifications and meet prescribed testing levels.

Typically associated with unwanted process conditions, variation may be defined the difference between a current reality in the context of system performance and a desired end-state. Clearly, nearly all manufacturing and product performance processes possess some degree of variation. Variation occurs in all natural and man-made

processes, even those in which it may appear non-existent. If variation cannot be measured, it is most likely because the measurement systems being used do not possess sufficient capabilities in terms of precision and accuracy. Certainly, some processes will tend to possess more than others, depending on the type of process, what is being measured, and who/what is measuring it. Notwithstanding, irrespective of the degree of variability that exists, managing and reducing it requires that it be traced back to its source.

In solving this problem, a variety of assumptions are made in order to facilitate the application of least squares regression, which is the focus of Chapters 5-7. These include normality, independence, and constant variance (or homoscedasticity) in the residuals. By extension, the validity of these assumptions, namely the last regarding constant variance, also presume moderate to low degrees of inherent process variability or that the process is stable. As demonstrated in Chapter 7, when these assumptions fail, alternative methods to *OLS* regression become necessary. In fact, virtually every facet of a manufacturing process (materials, work methods, machinery, measurement, human interaction, etc.) possesses inherent variability. In many cases, such variability can be comparatively high, which in turn can introduce non-constant variance in the residuals of system outputs. When heteroscedasticity exists, the most common approach is the application of weighted least squares regression. While effective and time-tested, the results obtained via the *WLS* method nevertheless can be negatively affected when high process variability exists. Moreover, this method is invoked in the post experimentation-phase, after the development of fitted response surface estimators.

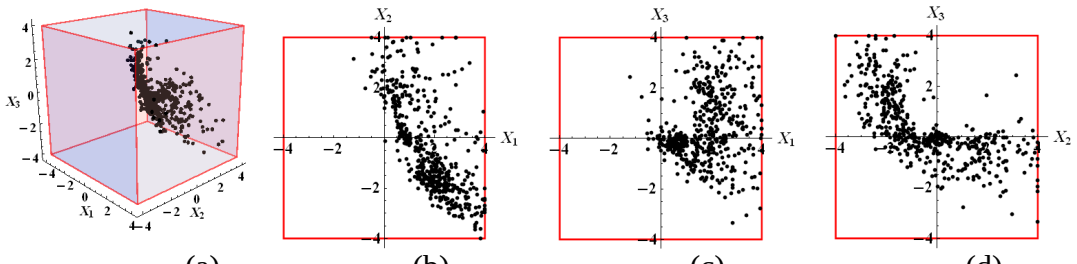
This chapter proposes an alternative method for dealing with inherently high process variability in the experimental phase of the *RPD* framework. In particular, the proposed method employs graphical and objective measures to identify sources of variability and then uses the coefficient of variation as a means for distinguishing the relative degrees of influence these sources exert among the various design points. Thereafter, the method advocates removing sources of variability and then applying optimal design theory to rebalance the experimental framework. Thereafter, commonly known *RPD* optimization schemes may be applied to obtain more precise optimal operating conditions with less variability and bias. The original work associated with the research presented in this chapter is published (currently in press) with reference Boylan and Cho (2013c).

9.2 Concept Motivation – Box and Draper's (1987) Printing Press Study

In the traditional *RPD* research, assumptions are typically made on the process parameters and underlying distribution, as well as on the condition of the data obtained through process operation and experimentation. In particular, it is usually assumed that moderate variability in system responses, as well as normality, homoscedasticity, and independence among the residuals are valid within a set of observational data to support the application of ordinary least squares regression in obtaining unbiased, minimum variance estimates. While this may greatly simplify a problem, it is often the case that such conditions do not hold in practice, as high variability is not uncommon in many industrial settings. This can complicate the search for optimal solutions using the *OLS* approach and highlights the need for alternative approaches.

As a motivation for this concept, reconsider the pilot study presented in Chapter 3 involving the Box and Draper (1987) printing press experiment. The experiment concerns a printing machine's index in applying coloring inks to package labels, which is presumed to be a normally distributed *N*-type quality characteristic of interest, *Y*. The desired target value for the machine's index is $\tau = 500$. Table 3.2 in Chapter 3 contains the experimental framework for this study.

Using this example, previous researchers developed second-order response surface designs for the mean and standard deviation and then applied their particular optimization scheme to find optimal operating conditions $\mathbf{x}^* = (X_1^*, X_2^*, X_3^*)$. As illustrated in Chapter 3, repeating this experiment 500 times within an extrapolated experimental region via Monte Carlo simulation reveals a considerable degree of imprecision in the optimal solutions. The plots in Figure 9.1 depict the position of these solutions in three dimensional space, as well in relation to each cross-section of factor pairs.



(a) (b) (c) (d)

Figure 9.1. Identification of $\mathbf{x}^*$, Printing Press Study, $-4 \le X_i \le 4$ (500 iterations).

This highlights vulnerabilities in the traditional approach to the *RPD* problem when high variability permeates system performance. In particular, such conditions can result in the identification of "optimal settings" that may span the entire length or more of the factor

space, a problem that can be exacerbated when the number of replicates at each design point is low. This indicates significant imprecision in the estimated functions used to generate the factor settings and further suggests that because the data are so variable, the results should not be trusted with any degree of certainty; conditions that ultimately can lead to more conservative decision-making and the associated problems with overdesign and overmanufacture described in Section 9.1.

We could use additional replications at each design point to mitigate issues with high variability but constraints on time and other resources typically preclude it. Thus, alternative approaches should be considered first to address more directly the issues caused by excessive variability. Certainly, time, costs, and a variety of other resource constraints may impact the experimental approaches employed, but this should be weighed against the need for elevated levels of precision. In short, as has been noted in previous chapters, if our model is imprecise in estimating the appropriate response surface, then optimizing that model is going to yield equally imprecise results as they relate to the true optimal factor settings. Accordingly, the results and recommendations provided to decision makers would very likely lead to the implementation of suboptimal modifications to processes and products. The purpose of this chapter, then, is to propose an alternative technique that focuses on identifying and extracting sources of variability, followed by the application of optimal design theory to restructure the experimental approach, in order to obtain more precise *RPD* solutions.

9.3 Methodology Development

9.3.1 Experimentation and Analysis

Experimentation and analysis begin with the specification of response variables, factors or predictors influencing the responses, and region of interest for a designed experiment. Suppose the objective is to identify the optimal factor settings $\mathbf{x}^*=(X_1^*, X_2^*, \ldots, X_k^*)$ that support achieving a mean process performance with minimal deviation from a desired target and with minimum variability in the result. Pursuant to this, consider an experimental framework whereby a quality characteristic, Y, is influenced by a set of control factors $X_1, X_2, \ldots, X_k$. The experiment consists of n design points, or runs, each of which contains m replicates for the observed response. Let y_{qj} denote the jth response at the qth design point, where $q = 1, \ldots, n$ and $j = 1, \ldots, m$. Table 9.1 portrays the framework for such an experiment.

Table 9.1. Experimental *RSM* framework.

Design Point	X_1 X_2 .... X_k	Replications	$\bar{y}$	s
1	Control Factor Settings	y_{11} ..... y_{1m}	$\bar{y}_1$	s_1
⋮		⋮ ⋮	⋮	⋮
q		y_{q1} ..... y_{qm}	$\bar{y}_q$	s_q
⋮		⋮ ⋮	⋮	⋮
n		y_{n1} ..... y_{nm}	$\bar{y}_n$	s_n

The replicates at each design point are then used to obtain parameter estimates for the mean and standard deviation using the following formulas:

$$\bar{y}_q = \frac{\sum_{j=1}^{m} y_{qj}}{m_q}, \quad s_q = \sqrt{\frac{\sum_{j=1}^{m} \left(y_{qj} - \bar{y}_q \right)^2}{m_q - 1}} \tag{9.1}$$

Prior to the final estimation of response surface functions for the mean and variability of the process, a comprehensive data analysis of both the responses and the residuals must be performed to ascertain the underlying conditions in the data. This includes an investigation of normality and variability in the process responses, as well as the verification of the assumptions of normality, homoscedasticity, and independence in the residuals. These are investigated using both graphical and objective methods, which may include the following approaches briefly discussed in (1) - (3) below:

(1) To assess deviations from normality in the responses, researchers tend to use either the Kolmogorov-Smirnov test for large samples (> 2,000), or the Shapiro-Wilk test for small or medium-sized samples. Generally, the sample sizes obtained in *RPD* experimentation typically fall in the latter category. Thus, for the Shapiro-Wilk test, when the *p* observations made on quality characteristic are sorted in ascending order, the alternative hypotheses $H_0 : Y \in N(\mu, \sigma^2)$ and $H_1 : Y \notin N(\mu, \sigma^2)$ are evaluated using the *W* statistic given by:

$$W^* = \left[\frac{b}{s\sqrt{p-1}} \right]^2, \quad \text{where } b = \sum_{l=1}^{\kappa} a_{p-l+1} \left[y_{(p-l+1)} - y_l \right]$$

Here, κ is the largest integer that is less than or equal to *p/2*, and *s* denotes the sample standard deviation. For a given significance level α, tables are then used to reference the coefficients *a* and the critical values W_α. Different from most statistical tests, since the critical region lies in the small tail of the distribution, if $W^* > W_\alpha$ then H_0 is concluded

(that is, sufficient evidence exists to suggest the observations follow a normal distribution).

(2) A trademark of high variability data is that their sample standard deviation is in general quite large, implying a "largely uninformative" sample mean that fails to adequately describe the location of the bulk of the observed values. Willinger, *et al.* (2004) associated high variability with situations in which a set of observations assumes values that vary over orders of magnitude, with most observations taking values that are relatively close together, with a few extreme observations attaining values that deviate considerably from this first group with non-negligible probabilities, and with intermediate-sized observations occurring with appreciable frequencies. Using this concept, a highly variable process is classified as one in which the range of variability in the responses is noticeably large and where one or more of the responses lies more than three standard deviations (+/-3σ) from the mean response.

(3) The analysis of residual errors is predicated on the development of preliminary regression models for the mean and standard deviation using the *OLS* approach. This yields the response surface models for each which take the following forms:

$$\begin{aligned} \hat{\mu}(\mathbf{x}) &= \mathbf{X}\hat{\beta}_{\mu} \quad \text{where } \hat{\beta}_{\mu} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\bar{\mathbf{y}} \\ \text{and } \hat{\sigma}(\mathbf{x}) &= \mathbf{X}\hat{\beta}_{\sigma} \quad \text{where } \hat{\beta}_{\sigma} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{s} \end{aligned} \tag{9.2}$$

$$\text{and where } \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1,k-1} \\ 1 & x_{21} & \cdots & x_{2,k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{n,k-1} \end{bmatrix}, \ \bar{\mathbf{y}} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_n \end{bmatrix}, \text{ and } \mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}$$

Using these models, namely that for the mean, as a basis for analysis, the residuals may be examined to verify underlying assumptions of normality, independence, and homoscedasticity. The procedures for each are described in (a) – (c):

(a) As with the responses, normality in the residuals may also be examined using graphical measures such as the normal probability plot, followed by the more objective methods provided by the Kolmogorov-Smirnov or Shapiro-Wilk tests.

(b) Independence in the residuals may be examined using the Durbin-Watson test, which is typically sufficient to detect a lack of randomness therein. If dependence is detected, then remedial measures may be performed, such as adding predictor variables or using transformations in the variables to eliminate interdependencies.

(c) The investigation of heteroscedasticity (non-constant variance), typically involves the use of graphical measures, such as a plot of the residuals against the fitted values, as well as objective hypothesis testing using either the Brown-Forsythe test, which is more robust to departures from normality in the data, or the Breusch-Pagan test. The Breusch-Pagan (B-P) test assumes independence and normality among the residuals, but further assumes a relationship for the error variance σ_q^2 among the *k* regression coefficients and *k*-1 predictor variables. This relationship appears as $\log_e \sigma_q^2 = \gamma_0 + \gamma_1 X_{q1} + \cdots + \gamma_{k-1} X_{q,k-1}$, and implies that the error variance fluctuates up or down with **x**, based on the sign of the associated coefficients. Since constant error variance corresponds to the instance in which each of the coefficients contained in response function equals 0, the alternative hypotheses $\mathrm{H}_0 : \gamma_1 = \gamma_2 = \cdots = \gamma_{k-1} = 0$ versus H_1 : not all $\gamma_i = 0$ are tested using the B-P statistic:

$$X_{BP}^2 = \frac{SSR^*}{2} \div \left(\frac{SSE}{Nm} \right)^2 \tag{9.3}$$

where *Nm* denotes the total number of experimental observations, *SSR** is the regression sum of squares obtained by regressing the squared residuals, ε_j^2, against one or more of the predictor variables, and *SSE* is error sum of squares obtained for the full regression model. Comparing this statistic to the chi-square distribution with *k*-1 degrees of freedom, if $\chi_{BP}^2 > \chi_{(1-\alpha),k-1}^2$ then we reject H_0 and conclude that sufficient evidence exists to suggest non-constant variance. In processes with inherently high variability or asymmetry in the responses, the assumption of constant variance in the residuals would most likely not hold and would thereby necessitate the use of remedial measures.

9.3.2 Identifying and Overcoming Sources of Variability

9.3.2.1 The Weighted Least Squares (*WLS*) Approach

Traditionally, when heteroscedastic conditions prevail, investigators use one of two alternative approaches to obtaining improved estimators for the mean and standard deviation response surface functions. These include either a transformation of the response variable *Y*, which essentially induces constant variance by equalizing the error variation across all predictor variables, or through the application of a weighted least squares (*WLS*) approach in the estimation process to mitigate the effects of unequal error deviations. While either method may work well in a broad variety of applications, the *WLS* approach seems to be the most common choice, as it avoids the possibility of an inappropriate regression relationship which can result from transformations of *Y*.

Since we rarely have the good fortune of knowing the error variance, we consider two approaches for developing estimates of the variance σ_j^2. First, it can be shown that $\sigma_j^2 = E\left[\varepsilon_j^2\right]$. If we let $E\left[\varepsilon_j^2\right] = e_j^2$, wherein e_j^2 denotes the squared residual for the *j*th design point, then we can say that e_j^2 is an unbiased estimator of σ_j^2. As Kutner *et al.* (2005) point out, these relationships allow us to develop an estimate of the variance function by first fitting the regression model using ordinary unweighted least squares and then regressing the squared residuals against the appropriate factors, or predictor variables. Pursuant to estimating the standard deviation function, shown in Equation (9.2), we then use this resulting variance function $\hat{s}_j$ to obtain fitted values, for each of the *N* design points, which in turn allow us to compute the estimated weights for each point, *j*, as follows:

$$w_j = 1 \Big/ (\hat{s}_j)^2 \tag{9.4}$$

Using the estimated weights, we can then construct the weight matrix

$$\underset{\mathbf{N\times N}}{\mathbf{W}} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_N \end{bmatrix}$$

We next apply standard regression procedures to the weighted regression models. Notably, for a regression model with k parameters, the *WLS* estimators of the regression coefficients $\mathbf{b}_w = (b_{1w}, b_{2w}, ..., b_{kw})^T$, whereby $\mathbf{b}_w = \left(\mathbf{X}_w^T\mathbf{X}_w\right)^{-1}\mathbf{X}_w^T\mathbf{Y}_w$, may be obtained via the weight matrix, **W** as follows:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\mathbf{w}} = \mathbf{b}_{w} &= \left[\mathbf{W}^{1/2}\mathbf{X})^{T}\mathbf{W}^{1/2}\mathbf{X}\right]^{-1}\left(\mathbf{W}^{1/2}\mathbf{X}\right)^{T}\mathbf{W}^{1/2}\mathbf{Y} \\ &= (\mathbf{X}^{T}\mathbf{W}^{1/2}\mathbf{W}^{1/2}\mathbf{X})^{-1}\mathbf{X}^{T}\mathbf{W}^{1/2}\mathbf{W}^{1/2}\mathbf{Y} \\ &= (\mathbf{X}^{T}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{T}\mathbf{W}\mathbf{Y}\end{aligned} \tag{9.5}$$

Obviously, the elimination or reduction of the influences of variability is achieved by assigning the greatest weight to those terms with the smallest error variance. The second approach involves cases in which, given sufficient replication at each experimental design point, we can simply use the sample variances or sample standard deviations of the observations at each design point to estimate the error deviations. These, in turn, allow us to compute estimates for the various weights, w_j, using the ratio in Equation (9.4). The key in this approach is determining what number of replications at each design point is sufficient. It is worth noting that either approach for estimating the weights can be particularly helpful in cases where the error variances differ significantly; that is, in highly variable processes. However, when the differences are relatively modest, the value of using these approximation methods in the *WLS* approach diminishes.

The iteration in the *WLS* approach occurs through a comparison between the standard errors of the *WLS* coefficients to the *OLS* coefficients. If the regression coefficients obtained using this method are significantly different than those obtained using ordinary least squares regression, the model may be reweighted using a subsequent re-estimation of the variance. This process continues in an iterative fashion, whereby revised weights are recomputed and then used to ensure convergence and the reduction of model error. Pursuant to this approach, fitted response functions are developed for the mean, and standard deviation. In particular, using the mean response $\overline{y}$ and standard deviation *s*, the general form of the estimated response functions for the process mean

and standard deviation with *k* parameters or *k-1* predictor variables changes from the formulations in Equation (9.2) with the inclusion of **W**, as given by:

$$\begin{aligned} \hat{\mu}(\mathbf{x})_{WLS} &= \mathbf{X}\hat{\beta}_{\mu} \quad \text{where } \hat{\beta}_{\mu} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\overline{\mathbf{y}} \\ \text{and } \hat{\sigma}(\mathbf{x})_{WLS} &= \mathbf{X}\hat{\beta}_{\sigma} \quad \text{where } \hat{\beta}_{\sigma} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{s} \end{aligned} \tag{9.6}$$

and where **X**, $\overline{\mathbf{y}}$, and **s** are defined precisely the same as in Equation (9.2).

9.3.2.2 The Proposed Coefficient of Variation Technique

Whereas the *WLS* method addresses variability in the post-experimentation phase by modifying the estimated models, an alternative approach to eliminating variability can be applied in the experimental phase itself using the coefficient of variation (*CV*) to scope the design space and then applying optimal designs to rebalance the experiment. Obviously, this component in our methodology implies that some experimentation has been performed and it has been determined 1) the process under study possesses elevated levels of inherent variability, 2) that the assumption of homoscedasticity fails to hold and, consequently, 3) that the estimated model is a poor fit. Thus, invoking this particular approach further implies that we would revisit the experiment in order to redefine the experimental parameters and the associated design space to reduce variability and achieve greater precision.

The *CV* for a quality characteristic aims to describe the dispersion of the characteristic in a way that does not depend on its measurement unit. That is, it describes data variability in terms of the relative sizes of the squared residuals and outcome values. The *CV* for the *q*th design point is determined using the following ratio:

$$CV_q = \frac{s_q}{\overline{y}_q} \tag{9.7}$$

Using this ratio, a higher *CV* indicates greater the dispersion in the response. Conversely, a lower *CV* corresponds to smaller residuals relative to the predicted value, suggesting a sufficient relationship exists between the estimated and true responses.

Unfortunately, no definitive threshold exists for the *CV* that indicates elevated or high variation, as it depends on the process and responses under examination. For our purposes, we use the *CV* to identify sources of variation that unduly influence or exert leverage upon the response. Thus, the examination focuses on relative degrees of influence exerted by any one or set of individual design points on the process. As such, the CV threshold used will likely vary between manufacturing processes.

Once points are identified as particularly influential due to excessive variability, they may be removed from consideration. As a consequence of extracting such points, however, a non-standard design likely results, which raises the need for alternative experimental designs. The need to consider optimal design strategies arises either when certain conditions render the use of traditional experimental approaches inappropriate. Such conditions include instances where time, resource limitations, or some physical restrictions constrain the experimental region. In this situation, however, the identification and elimination of sources of process variability equates to the imposition of constraints on the design space in the interest of obtaining enhanced precision in the response surface models and, ultimately, the optimization results. While an assortment of optimal designs exist for consideration, the objective is not to discuss the full range of alternatives in this chapter. Nevertheless, in the interests of not confining the

methodology to a singular design approach, several of the more commonly used optimal designs (e.g., *D*-, *A*-, and *G*-optimal designs) are included in the discussion.

D-optimal designs are particularly useful when precision is the primary factor in model estimation. The goal of *D*-optimality is to improve the experimental design through the careful selection of candidate points, which is done using the *D*-criterion. The *D*-criterion is based on the ellipsoidal joint confidence region for the parameters in the standard regression model. This region is defined by the set of coefficient vectors $\hat{\boldsymbol{\beta}}$, that satisfies

$$\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{X}^T \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{pMSE} \leq F(1-\alpha; p, N-p) \tag{9.8}$$

where $\mathbf{X}$ denotes the design matrix. Among others, two properties that make *D*-optimal designs appealing to analysts involve their ability to minimize both the generalized variance of the estimated parameters $\hat{\boldsymbol{\beta}}$ and the volume of the ellipsoidal confidence region. Using the notation previously described, we can determine the variance of the estimates in $\hat{\boldsymbol{\beta}}$ through:

$$V(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \tag{9.9}$$

Simple deduction shows that the variance of $\hat{\boldsymbol{\beta}}$ depends on the generalized variance $(\mathbf{X}^T \mathbf{X})^{-1}$. The larger this value, the poorer the estimation of $\hat{\boldsymbol{\beta}}$. Consequently it becomes equally clear that a large value for $\left|\mathbf{X}^T \mathbf{X}\right|$ will serve to minimize $(\mathbf{X}'\mathbf{X})^{-1}$ and thereby minimize $V(\hat{\boldsymbol{\beta}})$. Thus, the objective of the *D*-criterion is to maximize the determinant of the information matrix.

Unlike the *D*-optimal design, *A*-optimal designs do not consider the covariance among regression coefficients, focusing instead on the individual variances of the coefficients themselves. Albeit, the overarching goal remains focused on minimizing the variance of estimators. *A*-optimal designs achieve this by minimizing the trace of the inverse of the information matrix, whereby the trace of an *N x N* square matrix, call it **M**, is defined to be the sum of the elements on the diagonal of **M**. That is,

$$trace(\underset{\mathbf{N}\times\mathbf{N}}{\mathbf{M}}) = m_{11} + m_{22} + \ldots + m_{NN} = \sum_{j=1}^{N} m_{jj} \tag{9.10}$$

Thus, the *A*-criterion used is shown by Minimize $trace(\mathbf{X}^T\mathbf{X})^{-1}$, which once again is associated with the generalized variance. The result is the minimization of the average variance of the regression coefficient estimates.

Whereas *D*- and *A*-optimal designs employ invariants of the information matrix in their optimality criteria, *G*-optimal designs (along with *I*- and *V*-optimal designs) focus on the variance of predictions. Considering the mean response, we can express the variance of predictions, $Var[\hat{\mu}(\mathbf{x})]$ as:

$$Var[\hat{\mu}(\mathbf{x})] = MSE\ [\mathbf{X}_\mathbf{m}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}_\mathbf{m}]$$

Where the vector $\mathbf{X}_\mathbf{m}^T = \begin{bmatrix} 1 & X_{m1} & \cdots & X_{m,p-1} \end{bmatrix}$ represents the location within the design space where a prediction is being made. Typically, the prediction variance is scaled as shown in Equation (9.11) to facilitate comparison between various designs.

$$\frac{NVar[\hat{\mu}(\mathbf{x})]}{MSE} = N\mathbf{X}_\mathbf{m}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}_\mathbf{m} \tag{9.11}$$

In contrast to *D*- and *A*-optimality criteria that comprise a single value, the *G*-optimality criterion depends on the location $\mathbf{X_m}$ at which predictions are being made. Specifically, this criterion seeks to minimize the maximum entry in the main diagonal of the (*N x N)* hat matrix, $\mathbf{H} = \mathbf{X(X}^T\mathbf{X)}^{-1}\mathbf{X}^T$. The generalized effect is to minimize the maximum variance of the predicted values.

The objective for any optimal design is to determine the set of design points from among a list of candidate points that best satisfy the selected criterion. Most often, the optimality criterion used is a function of the variance of the estimated model parameters. Moreover, due to the iterative nature and the complexity involved, these designs typically require the use of sophisticated computer algorithms. A variety of methods exist for determining the best set of candidate points. The proposed approach uses the so-called Federov exchange method.

In a broader sense, a general exchange method involves the simultaneous addition and removal of a pair of a design point in the following manner:

$$\left(\mathbf{X}_{\mathbf{new}}^T\mathbf{X}_{\mathbf{new}}\right) = (\mathbf{X}_{\mathbf{old}}^T\mathbf{X}_{\mathbf{old}}) - (X_{\mathbf{i}}X_i^T) + (X_{\mathbf{j}}X_j^T)$$

where X_i and X_j denote the point being considered for removal and the point being considered for inclusion respectively. In contrast, the method developed by Federov (1972) incorporates the interaction of the variance functions of the candidate pair into the calculation of the new determinant such that

$$\left(\mathbf{X}_{\mathbf{new}}^T\mathbf{X}_{\mathbf{new}}\right) = (\mathbf{X}_{\mathbf{old}}^T\mathbf{X}_{\mathbf{old}})(1 + \Delta(X_i, X_j) \qquad (9.12)$$

Essentially, the algorithm iteratively examines the list of candidate points, seeking to replace an existing point with another one that yields a higher determinant. Thus, at each iterate, the goal is to generate a new vector of included design points that will increase the determinant of the information matrix and ultimately generate the best design. At each stage, one of the design points is exchanged with another design point so that the ratio of the new determinant (containing the new design point) to the previous determinant is maximized. When the ratio falls below some tolerance level, the previous design is taken as the optimal design.

In R, the `optFederov()` function from the AlgDesign Package (Wheeler, 2004) applies Federov's exchange algorithm at Equation (9.12) to compute "an exact or approximate algorithmic design" for each of the aforementioned criteria. Thereafter, a best-subsets regression is performed to determine the best, or most appropriate mix of predictors for the refined experimental design. This also may be performed in R using the `leaps()` function from the Leaps Package (Lumley, 2009). In short, this function executes an exhaustive search for the best subsets of predictors using an efficient branch-and-bound algorithm. Models therein are compared using four well-known criteria, including the adjusted coefficient of determination $R_{a,p}^2$, Mallow's C_p, the prediction sum of squares ($PRESS_p$), and the mean square error (MSE_p). Table 9.2 summarizes these criteria.

Experimental designs, such as those heretofore described, are viable alternatives for a variety of situations in which standard experimental approaches do not apply or cannot achieve adequate precision. Generally, they provide an efficient and cost-

effective way to select design points for experimentation. We should note, however, that the use of optimal designs should not serve as a direct substitute for standard experimental approaches, but rather as an alternative approach when standard methods are inappropriate or otherwise insufficient.

Table 9.2. Evaluation Criteria for the Model with p Parameters.

Measure (with Formula)	Description / Interpretation
$R^2_{a,p} = 1 - \left(\frac{N-1}{N-v}\right)\frac{SSE_p}{SSTO}$	High criterion values are sought. Here, the number of parameters is taken into account through the corresponding degrees of freedom. Thus, this measure is not influenced by an increase in predictor variables.
$C_p = \frac{SSE_p}{MSE(X_1,\ldots,X_{p-1})} - (n-2p)$	The criterion is concerned with the total *MSE* of the n fitted values for the subset regression model. The objective is to identify subsets for which C_p is small (low total *MSE*) but is near p (which indicates low bias).
$PRESS_p = \sum_{j=1}^{N}\left(\frac{e_j}{1-h_{jj}}\right)^2$	Here, e_j represents the jth residual and h_{jj} represents the jth diagonal element of the hat matrix **H**. For models with large h_{jj} , the prediction sum of squares will be large, resulting in more influential points within a regression model. Models with smaller $PRESS_p$ values contain less overall prediction error in their estimation.
$MSE_p = \frac{SSE_p}{N-p}$	Measures the average of the square of the error for a model with p parameters. Low values are sought.

9.3.3 Determination of Approved Models and Optimization

Considering the framework delineated in Section 9.3.1 for an industrial process involving a nominal-the-best (N-type) quality characteristic as the response of interest, we assume that the levels of x_i for $i = 1, 2,\ldots,k$ are both quantitative and continuous, and can be controlled by the experimenter. Fitted response functions are developed for the process mean and standard deviation using the techniques described in Section 9.3.2. In particular, assuming second-order polynomials for the response functions in each case, the general form of the estimated response functions for the process mean and standard deviation with k parameters or $k-1$ predictor variables appears as:

$$\hat{\mu}(\mathbf{x}) = \hat{\beta}_0 + \mathbf{x}^T\hat{\mathbf{b}}_\mu + \mathbf{x}^T\hat{\mathbf{B}}_\mu\mathbf{x} + \varepsilon_\mu \tag{9.13}$$

$$\text{and} \quad \hat{\sigma}(\mathbf{x}) = \hat{\gamma}_0 + \mathbf{x}^T\hat{\mathbf{b}}_\sigma + \mathbf{x}^T\hat{\mathbf{B}}_\sigma\mathbf{x} + \varepsilon_\sigma \tag{9.14}$$

$$\text{where} \quad \underset{\mathbf{1\times k}}{\mathbf{x}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{n,k-1} \end{bmatrix}, \ \underset{\mathbf{k\times 1}}{\hat{\mathbf{b}}} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, \text{and} \ \underset{\mathbf{k\times k}}{\hat{\mathbf{B}}} = \begin{bmatrix} \beta_{11} & \beta_{12}/2 & \cdots & \beta_{1k}/2 \\ & \beta_{22} & \cdots & \beta_{2k}/2 \\ & & \ddots & \vdots \\ sym. & & & \beta_{kk} \end{bmatrix},$$

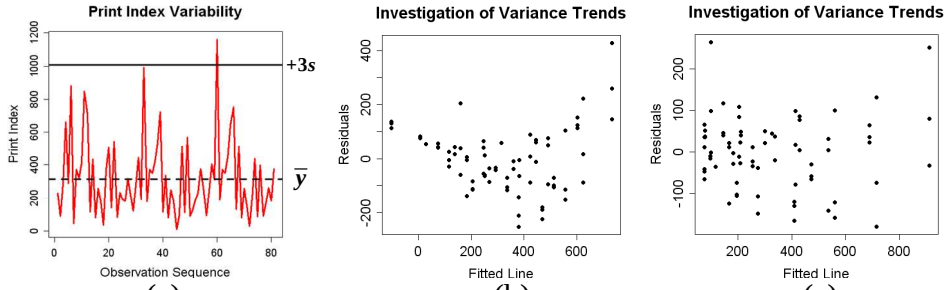
As noted previously, β_0 (and γ_0), $\hat{\mathbf{b}}_\mu$ (and $\hat{\mathbf{b}}_\sigma$) and $\hat{\mathbf{B}}_\mu$ (and $\hat{\mathbf{B}}_\sigma$) reflect the estimates of the intercept, linear, and second-order coefficients, respectively, and ε_μ and ε_σ denote the residual error for the mean and standard deviation. Cho's (1994)/Lin and Tu's (1995) *MSE*-based optimization scheme is used on either a spherical region of interest (in the case of a central composite design) such that $\mathbf{x}'\mathbf{x} \le \rho^2$, where ρ defines the spherical experimental region, and $\mathbf{x} \in$ the domain of the control factors, $\mathbf{\Omega}$; or a cuboidal region of interest bounded by (-1,1) in the case of factorial design.

9.4 Numerical Example – Printing Press Study Revisited

In this section, the printing press study introduced in Section 9.2 is revisited in two parts to facilitate comparison between the proposed technique and traditional approaches. In the first part, the three methods (*OLS*, *WLS*, and the proposed *CV* technique) are examined in the context of the original experiment and draw inferences from the performances of each. In the second, Monte Carlo simulation is used to more fully investigate performance trends and precision; observations are made and conclusions drawn regarding the simulation results.

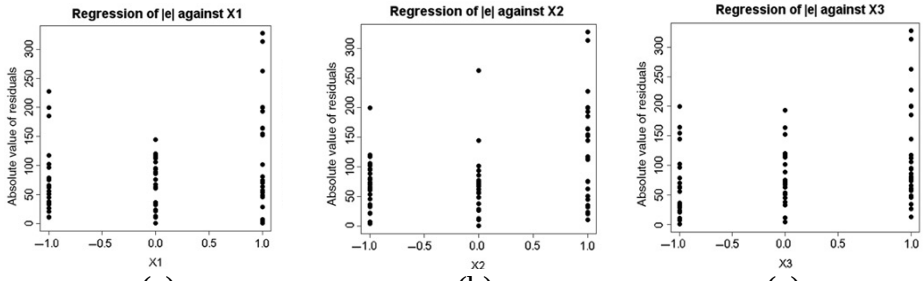
9.4.1 Initial Data Analysis

Pursuant to a thorough analysis of the original data associated with the print press experiment, two observations emerge. First, based on the criteria established in Section 4.1, the time-sequence plot in Figure 9.2(a) shows that the process qualifies as highly variable since one or more observations exceeds the +/- 3σ threshold. Second, as the residual plots for both first and second-order models in Figures 9.3(b) and (c) show, regression of the residuals against the fitted line suggests evidence of non-constant variance. After regressing the absolute values of the residuals against each of the predictor variables for the fitted line separately, the evidence for non-constant variance becomes more conclusive for each of the primary factors (see Figure 9.3).



(a) (b) (c)

Figure 9.2. Investigation of assumptions on (a) variability in the responses and (b)-(c) constant error variance for first and second order models, respectively.



(a) (b) (c)

Figure 9.3. Regression of squared residuals against (a) X_1, (b) X_2, and (c) X_3.

To confirm the variability trends observed in the residual plots, the Breusch-Pagan hypothesis test is applied. In the first order model, the calculated statistic (1) tests H_0: $\gamma_1 = \gamma_2 = \gamma_3 = 0$ versus H_1: Not all $\gamma = 0$, whereas in the second order model (2), the hypotheses are extended for $p = 10$ parameters.

$$(1)\quad \chi^2_{BP} = \frac{7.95\text{e+}09}{2} \div \left(\frac{1012557}{27(3)} \right)^2 = 25.45 \qquad (2)\quad \chi^2_{BP} = \frac{2.28\text{e+}09}{2} \div \left(\frac{539518}{81} \right)^2 = 25.72$$

Since $25.447 > \chi^2(.95, 3) = 7.81$ and $25.720 > \chi^2(.95, 9) = 16.92$ for the first and second-order models, respectively, we conclude H_1; that is, the error variance is not constant. The results of these tests reinforce the inferences drawn from the graphical analysis and further suggest the need for remedial measures in the development of fitted functions. Such remedial actions are performed using both the usual *WLS* approach and the proposed *CV* technique, which are described in the ensuing paragraphs.

9.4.2. Comparison of Methods via a Single Iteration of the Original Experiment

The initial comparison of the three methods (*OLS*, *WLS*, and the proposed *CV* technique) was performed using the original data for the printing press experiment shown in Table 3.2 on page 79. Applying the *OLS* regression method first yields the following second-order response surface models, which are commensurate with the models obtained by previous researchers:

$$\hat{\mu}(\mathbf{x})_{OLS} = 327.63 + 177.01x_1 + 109.42x_2 + 131.47x_3 + 32.01x_1^2 - 22.39x_2^2 - 29.06x_3^2 + 66.03x_1x_2 + 75.46x_1x_3 + 43.58x_2x_3$$

$$\hat{\sigma}(\mathbf{x})_{OLS} = 34.884 + 11.529x_1 + 15.327x_2 + 29.191x_3 + 4.206x_1^2 - 1.318x_2^2 + 16.776x_3^2 + 7.724x_1x_2 + 5.115x_1x_3 + 14.083x_2x_3$$

As with previous research using this particular data set, these models were developed under the premise that underlying assumptions regarding variability and the residual errors held.

The development of second order fitted response surface functions for the mean and standard deviation was performed using an iteratively reweighted least squares regression program constructed in R (refer to Appendix E for the associated programming code). At each iterate in the *WLS* procedure, the standard error for each regression coefficient is observed and compared to the standard error obtained in the previous trial. Once the estimated coefficients stabilize, the iterative process ceases and the fitted function obtained in the last trial is selected. For the purposes of simulation, the process was set to iterate up to 50 times to ensure convergence; convergence was achieved when the difference between standard error for each of the estimated coefficients was less than 0.05 when compared to the corresponding errors obtained in the previous iterate. Table 9.3 depicts the results of this process for the mean response surface function, which converged after 11 iterations; the corresponding process for the standard deviation converged after 23 iterations.

Table 9.3. Standard errors for mean regression coefficients in each *WLS* iterate.

Iteration	**$s(b_{0w})$**	**$s(b_{1w})$**	**$s(b_{2w})$**	**$s(b_{3w})$**	**$s(b_{4w})$**	**$s(b_{5w})$**	**$s(b_{6w})$**	**$s(b_{7w})$**	**$s(b_{8w})$**	**$s(b_{9w})$**
1	22.646	13.099	17.355	14.210	22.274	20.926	21.880	23.649	22.351	24.224
2	21.578	13.341	14.767	15.296	20.580	20.743	21.657	20.571	22.450	21.863
3	23.927	13.777	14.306	15.576	21.939	19.933	23.026	18.404	20.352	19.151
4	23.382	12.729	14.307	15.506	21.472	19.999	22.257	17.083	18.996	19.152
5	22.791	11.793	14.215	15.046	21.400	19.086	21.588	17.247	18.684	18.733
6	22.726	11.502	14.200	14.827	21.413	18.860	21.391	17.350	18.401	18.623
7	22.773	11.374	14.150	14.669	21.412	18.756	21.301	17.364	18.212	18.486
8	22.830	11.307	14.117	14.575	21.426	18.714	21.241	17.399	18.080	18.418
9	22.849	11.260	14.092	14.511	21.429	18.673	21.204	17.412	18.001	18.366
10	22.862	11.229	14.075	14.472	21.432	18.649	21.179	17.423	17.948	18.335
11	22.868	11.209	14.063	14.447	21.433	18.631	21.163	17.428	17.914	18.315

The *OLS* and *WLS* procedures resulted in the following estimated response surface functions for the mean and standard deviation for each method, respectively:

$$\hat{\mu}(\mathbf{x})_{WLS} = 317.09 + 171.23x_1 + 132.43x_2 + 125.67x_3 + 25.27x_1^2 - 11.43x_2^2 - 11.76x_3^2 + 37.63x_1x_2 + 46.69x_1x_3 + 60.66x_2x_3$$

$$\hat{\sigma}(\mathbf{x})_{WLS} = 45.014 + 21.461x_1 + 25.435x_2 + 34.658x_3 - 5.041x_1^2 - 19.516x_2^2 + 22.875x_3^2 - 21.45x_1x_2 + 21.177x_1x_3 + 27.421x_2x_3$$

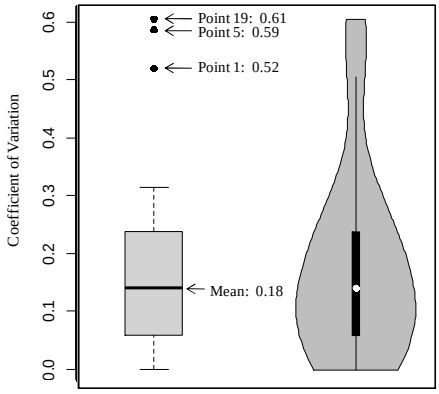
As expected, a comparison of the fitted response surface models developed using the *WLS* procedure to those obtained via *OLS* regression shows clear improvements in the precision of the models for both the mean and standard deviation. In particular, as shown in Table 9.4, a significant reduction in the error for each of the coefficients is observed when comparing the mean and standard deviation response models obtained through the *WLS* approach against those calculated using traditional *OLS* regression.

Table 9.4. Comparison of response models obtained via the *OLS* and *WLS* methods.

		Standard Errors of Regression Coefficients									
		$s(b_0)$	$s(b_1)$	$s(b_2)$	$s(b_3)$	$s(b_4)$	$s(b_5)$	$s(b_6)$	$s(b_7)$	$s(b_8)$	$s(b_9)$
Mean	*OLS*	38.75	17.94	17.94	17.94	31.07	31.07	31.07	21.97	21.97	21.97
	WLS	22.87	11.21	14.06	14.45	21.43	18.63	21.16	17.43	17.91	18.32
Reduction (%)		47.03	43.91	29.63	27.70	38.08	46.17	38.86	28.79	26.80	25.17
Standard Deviation	*OLS*	22.31	10.33	10.33	10.33	17.89	17.89	17.89	12.65	12.65	12.65
	WLS	17.87	9.51	9.57	10.08	14.43	13.80	13.77	11.48	12.41	12.83
Reduction (%)		58.61	52.39	52.12	49.56	58.30	60.12	60.21	53.08	49.30	47.60

The application of the proposed *CV* technique begins with the identification of potential sources of variability or instability in the process. As the plot in Figure 9.2(a) showed, a handful of responses either approach or exceed the 3σ threshold used to delineate a high-variability process. Computing the *CV* for each design point (see Figure

9.4) reveals three likely sources of excessive variability at design points 1, 5, and 19. The box plot in Figure 9.4 provides a more visual portrayal of these points, indicating clearly the extent to which they deviate from the rest of the data. The corresponding violin plot illustrates the influence these points exert on the rest of the data and the corresponding degree to which they infuse asymmetry into process outputs.



Design Point	Mean	Standard Deviation	CV	Design Point	Mean	Standard Deviation	CV
1	24	12.49	0.52	15	501.7	92.5	0.18
2	120.3	8.39	0.07	16	264	63.5	0.24
3	213.7	42.8	0.20	17	427	88.61	0.21
4	86	3.46	0.04	18	730.7	21.08	0.03
5	136.7	80.41	0.59	19	220.7	133.8	0.61
6	340.7	16.17	0.05	20	239.7	23.46	0.10
7	112.3	27.57	0.25	21	422	18.52	0.04
8	256.3	4.62	0.02	22	199	29.45	0.15
9	271.7	23.63	0.09	23	485.3	44.64	0.09
10	81	0	0.00	24	673.7	158.2	0.23
11	101.7	17.67	0.17	25	176.7	55.51	0.31
12	357	32.91	0.09	26	501	138.9	0.28
13	171.3	15.01	0.09	27	1010	142.5	0.14
14	372	0	0.00				

Figure 9.4. Box plot and table of *CV* values for the printing press study.

Per the methodology outlined in Section 9.4, these three points are then removed from the experiment, which results in a non-standard experimental region. Optimal design theory is then applied using the `optFederov()` function in R to rebalance the design while retaining 27 design points. This yields the design in Table 9.5, including the points retained from the original experiment, the points to be replicated, and the values for each of the optimal design criteria. An additional output in R is the lower bound on *D*-efficiency, which is determined by $D_e \geq e^{(1-1/G_e)}$, where G_e denotes the *G*-efficiency, which is an available standard of optimal design quality since the theoretical optimum value of the *G*-criterion is known.

Table 9.5. Optimal design results generated using the `optFederov()` function in R.

	# Times to Replicate	X_1	X_2	X_3
1	1	0	-1	-1
2	1	1	-1	-1
3	1	-1	0	-1
4	1	1	0	-1
5	2	-1	1	-1
6	1	0	1	-1
7	2	1	1	-1
8	2	-1	-1	0
9	1	1	-1	0
10	1	0	0	0
11	1	1	0	0
12	1	-1	1	0
13	1	0	1	0
14	1	0	-1	1
15	2	1	-1	1
16	1	-1	0	1
17	1	0	0	1
18	1	1	0	1
19	2	-1	1	1
20	1	0	1	1
21	2	1	1	1

Criterion	Value
D-	0.458
A-	3.708
G-	0.998
I-	9.709
Lower Bound on D- efficiency	0.998

Interestingly, as Table 9.5 shows, this procedure results in the extraction of three additional points from the original experiment, namely design points 11, 13, and 18. Thus the optimal design consists of 21 experimental points, six of which are replicated twice to achieve the 27 runs desired. Figure 9.5 depicts the new experimental region for the experiment after removing the identified sources of variability.

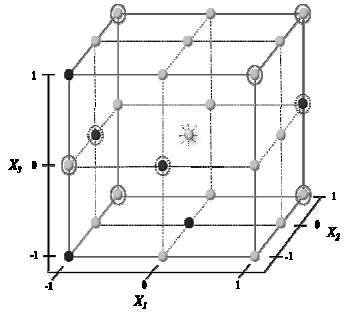


Figure 9.5. Optimal experimental design generated in R, where (●) denotes points extracted based on their *CV*, (◉) denotes additional points omitted through the optimal design process, and (◎) denotes replicated points in the new design space.

Thereafter, a best-subsets regression procedure is performed using the `leaps()` function in R, which yields the results shown in Table 9.6. Based on the four evaluation criteria used, the preferred model excludes the pure quadratic effect for x_1^2, resulting in k=9 parameters for the modified experimental design (Model 15 in Table 9.6).

Table 9.6. Results of the best-subsets regression procedure performed in R.

Model	x_1	x_2	x_3	x_1^2	x_2^2	x_3^2	x_1x_2	x_1x_3	x_2x_3	$R^2_{p,adj}$	C_p	$PRESS_p$	MSE_p
1	X									39.9%	230.14	995279.0	34066.90
2			X							22.8%	302.26	1279200.6	43773.53
3	X		X							62.4%	131.04	654308.6	21313.93
4	X							X		50.3%	180.03	883956.0	28181.29
5	X	X	X							75.4%	76.18	468233.1	13922.96
6	X		X					X		68.9%	101.56	554024.7	17635.95
7	X	X	X					X		84.0%	42.23	322833.1	9058.63
8	X	X	X				X			77.8%	65.19	451590.9	12569.76
9	X	X	X					X	X	89.7%	21.48	215416.1	5844.10
10	X	X	X				X	X		87.7%	28.57	267707.2	6980.80
11	X	X	X				X	X	X	93.1%	10.31	145201.3	3922.04
12	X	X	X			X		X	X	90.7%	18.50	194087.4	5298.24
13	X	X	X		X		X	X	X	93.7%	9.05	143351.0	3550.65
14	X	X	X			X	X	X	X	93.6%	9.38	138457.8	3608.01
15	X	X	X		X	X	X	X	X	94.3%	8.16	138370.0	3207.90
16	X	X	X	X	X		X	X	X	93.5%	10.87	155710.0	3714.19
17	X	X	X	X	X	X	X	X	X	94.1%	10.00	149750.9	3364.48

Applying the *OLS* regression procedure to the new experimental design yields the following response surface functions for the mean and standard deviation, respectively:

$$\hat{\mu}(\mathbf{x})_{CV} = 371.53 + 182.17x_1 + 103.74x_2 + 104.06x_3 - 46.22x_2^2 - 45.71x_3^2 + 50.07x_1x_2 + 114.26x_1x_3 + 72.24x_2x_3$$

$$\hat{\sigma}(\mathbf{x})_{CV} = 50.39 + 20.63x_1 + 27.67x_2 + 19.06x_3 - 5.38x_2^2 - 9.60x_3^2 - 2.67x_1x_2 + 13.48x_1x_3 + 23.49x_2x_3$$

Using the estimated response surface functions obtained via each of the three methods, a non-linear optimization scheme using the *MSE*-based model from Cho (1994)/Lin and Tu (1995) is employed in R to identify the optimal operating conditions for which total *MSE*

and bias are minimized. Table 9.7 contains the consolidated results for each of the three methods. These include comparisons of the standard errors for each of the coefficients in Table 9.7(a); overall model performance based on the $PRESS_p$ and $R^2_{p,adj}$ for each inTable 9.7(b); and the optimization results in terms of the total *MSE* and bias for the optimal operating conditions in Table 9.7(c).

It is immediately clear from the results in Table 9.7 that both the *WLS* and *CV* methods exhibit significant improvements over the traditional *OLS* approach. Between the two, the performances are comparable, with each outperforming the other based on certain evaluation criteria. In Table 9.7(a), the standard errors for the *WLS* and *CV* mean response surface models are equally split, whereas for the standard deviation model, the *CV* technique clearly outperforms the *WLS* alternative. If we examine the overall model evaluation criteria in Table 9.7(b), the *CV* technique demonstrates much better predictive capability in terms of the $PRESS_p$ criterion. However, the values for the $R^2_{p,adj}$ criterion obtained via the *WLS* and *CV* methods suggest that the model produced by the *WLS* technique achieves a slightly better fit.

In terms of the optimization performances shown in Table 9.7(c), the *WLS* and *CV* approaches yield a virtual tie for both total *MSE* and bias in the optimization results. Thus, the overall results indicate that in some respects, the proposed *CV* technique is, at the very least comparable to the more traditional *WLS* approach, but in some ways potentially superior in terms of the predictive capability it achieves.

Table 9.7. Consolidated results for the *OLS*, *WLS*, and proposed *CV* methods based on a single iteration of the original printing press experiment.

(a)

		Standard Errors of Regression Coefficients									
		$s(b_0)$	$s(b_1)$	$s(b_2)$	$s(b_3)$	$s(b_4)$	$s(b_5)$	$s(b_6)$	$s(b_7)$	$s(b_8)$	$s(b_9)$
Mean	***OLS***	38.75	17.94	17.94	17.94	31.07	31.07	31.07	21.97	21.97	21.97
	WLS	22.87	11.21	14.06	14.45	21.43	18.63	21.16	17.43	17.91	18.32
	CV	28.24	13.55	13.73	13.83	----	25.18	26.26	15.49	15.44	16.00
Standard Deviation	***OLS***	22.31	10.33	10.33	10.33	17.89	17.89	17.89	12.65	12.65	12.65
	WLS	17.87	9.51	9.57	10.08	14.43	13.80	13.77	11.48	12.41	12.83
	CV	13.86	6.65	6.74	6.79	----	12.36	12.89	7.60	7.58	7.85

(b)

	Process Mean			**Process Standard Deviation**		
	OLS-m	***WLS-m***	***CV-m***	***OLS-s***	***WLS-s***	***CV-s***
Avg *PRESS*$_p$	337,730.1	243,597.9	138,370	93,049.95	99,770.42	28,900.63
Avg $R^2_{p,adj}$	0.888	0.990	0.943	0.165	0.736	0.625

(c)

		OLS	***WLS***	***CV***
Optimization Results	**Avg *MSE***	406.349	1.73E-19	1.03E-16
	Avg Bias	2.386	1.24E-10	8.61E-09

Nevertheless, analyzing the actual optimal operating conditions obtained using each method raises questions regarding precision. Figure 9.6 shows the relative proximity of the $\left(X_1^*, X_2^*, X_3^*\right)$ point obtained for each method, and the disparities between them are obvious. Recalling the considerable dispersion in the simulated results from the pilot study shown in Figures 9.1 and 9.2, it is practically impossible to determine from these results which method is more reliable and precise.

Hence, while the comparisons to this point demonstrate the potential for comparable or improved results using the *CV* technique, the real question is one of precision. That is, can we expect a greater degree of precision in the results obtained via

the *CV* technique compared to the other two? In order to examine this aspect, a Monte Carlo simulation approach in R is used.

OOCs obtained using each method (single run)

OLS	1.20	-0.32	-0.08
WLS	1.15	-0.12	-0.14
CV	1.41	1.55	-0.62

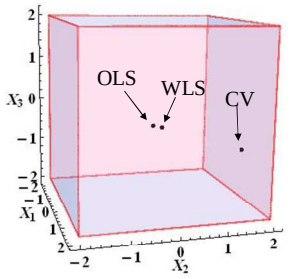


Figure 9.6. Comparison of optimal operating conditions obtained via the *OLS*, *WLS*, and proposed *CV* methods using a single run of the original experimental data.

9.4.3 Comparison of Methods via Monte Carlo Simulation

As the results in Figures 9.1 and 9.2 showed from the pilot study, simulation of the printing press experiment revealed considerable imprecision in the optimization results obtained for the 500 runs. Given the results in Section 9.4.2, it was presumed that the two alternative methods would likewise possess some degree of imprecision in the results. The questions were how much imprecision and to what extent did one method outperform the others? To this end, the overarching purpose of the simulation was to facilitate an investigation of performance trends between the three methods to determine whether better and more precise solutions could be expected via the proposed *CV* technique in the presence of high system variability.

Two simulations were constructed in R, one for the *OLS* and *WLS* approaches and one for the proposed *CV* technique, which required a unique approach to the former based on the aspects of optimal design and best-subsets regression used in the procedure. In each iterate of the simulation, fresh observations are generated using the values for the

mean and standard deviation at each design point in Table 9.2 to generate normal random variates consistent with the original experiment. The same generated data are used in each of the three simulations and are then used to obtain estimates for the sample mean and standard deviation at each design point. The derived estimates are used to perform the appropriate regression procedure to obtain fitted response surface functions for the mean and standard deviation. The resuling fitted functions are used with the *MSE*-based optimization framework to determine the optimal operating conditions. The results at each iterate are recorded, and output is generated to facilitate analysis.

Two trials of the simulation were performed in which both the number of observations generated per design point and the number of iterations executed were varied. To remain consistent with the original experiment, the first trial involved three simulated observations per design point. The experiment was iterated 1,000 times, yielding the averaged results across all 1,000 iterations shown in Table 9.8.
The results from the first simulation trial demonstrate once again that the *CV* technique performs at least comparable to, but in several ways better than the *WLS* approach when examined in the context of model errors and overall predictive capabilities. However, the optimization results in Table 9.8(c) suggest superiority in the proposed method. In particular, the proposed CV techniques achieves reductions of up to 50% for both the *MSE* and target bias. Furthermore, comparisons of the plots in Figure 9.7 illustrate clear gains in precision using the *CV* technique.

Table 9.8. Comparison of consolidated results for Trial 1 (1,000 iterations; 3 obs/design point) for (a) coefficient errors, (b) model performance, and (c) optimization results.

(a)

		Standard Errors of Regression Coefficients									
		$s(b_0)$	$s(b_1)$	$s(b_2)$	$s(b_3)$	$s(b_4)$	$s(b_5)$	$s(b_6)$	$s(b_7)$	$s(b_8)$	$s(b_9)$
Mean	*OLS*	43.17	19.98	19.98	19.98	34.61	34.61	34.61	24.47	24.47	24.47
	WLS	26.27	15.82	16.17	17.41	23.44	24.31	24.32	21.24	22.69	22.01
	CV	33.67	16.15	16.37	16.49	----	30.02	31.31	18.47	18.40	19.08
% Error Reduction	*OLS* → *WLS*	39.15	20.82	19.09	12.88	32.29	29.76	29.75	13.22	7.31	10.08
	OLS → *CV*	22.02	19.18	18.09	17.46	----	13.27	9.55	24.55	24.80	22.04
Standard Deviation	*OLS*	24.63	11.40	11.40	11.40	19.75	19.75	19.75	13.96	13.96	13.96
	WLS	17.19	9.37	9.03	10.04	14.17	14.82	14.10	11.27	12.04	11.62
	CV	18.57	8.91	9.03	9.10	----	16.56	17.27	10.19	10.15	10.53
% Error Reduction	*OLS* → *WLS*	60.19	53.10	54.80	49.78	59.05	57.19	59.26	53.93	50.80	52.54
	OLS → *CV*	56.98	55.42	54.82	54.47	----	52.16	50.11	58.38	58.52	56.99

(b)

	Process Mean			Process Standard Deviation		
	OLS-m	***WLS-m***	***CV-m***	***OLS-s***	***WLS-s***	***CV-s***
Avg PRESS	407,375.30	323,176.10	194,641.40	116,673.80	106,025.60	56,619.01
Avg $R^2_{p,adj}$	0.86	0.94	0.92	0.11	0.41	0.40

(c)

		OLS	***WLS***	***CV***
Optimization Results	Avg MSE	220.63	143.35	74.94
	Avg Bias	0.91	0.63	0.37

Specifically, as (X_1^*, X_2^*, X_3^*) are plotted for each of the 1,000 iterations of the simulation, a tighter grouping of the results is observed in the plots corresponding to the *CV* technique (Figure 9.7(c)). By comparison, the plots associated with the *OLS* and *WLS* methods (Figures 9.7(a) and (b), respectively) continue to reflect results that span a considerable portion of the design space.

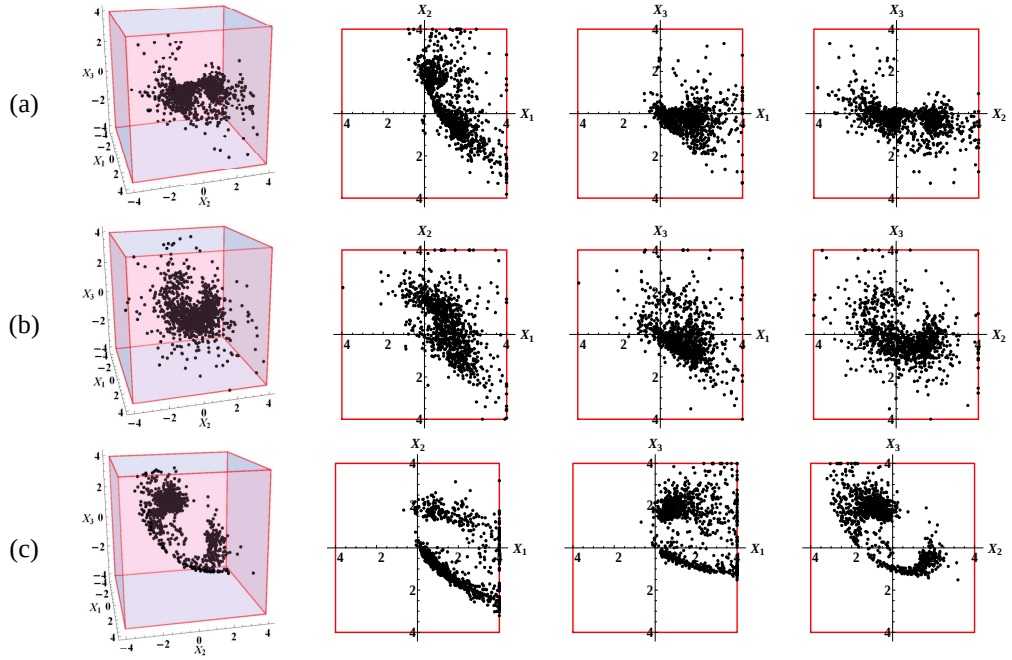


Figure 9.7. Identification of x* for Simulation Trial 1, $-4 \le X_i \le 4$; (a) x* for *OLS* method, (b) x* for *WLS* method, (c) x* for *CV* technique.

In the second simulation trial, the number of observations per design point are increased to 25 in order to examine the effects of added replication on the simulation results. Additionally, the number if iterations is increased to 2,000. These modifications in trial 2 yielded the results in Table 9.9.

As with the results in Trial 1, the results in Trial 2 demonstrate comparable performances between the *WLS* and *CV* approaches for both individual coefficient errors, overall model errors, and predictive capabilities. Once again, both alternatives drastically outperform the traditional *OLS* approach, but between the two the *CV* technique performs at the level of or slightly better than the *WLS* approach, at least in the context of predicting process responses. In terms of the optimization results, however, the gap

between the two methods closes considerably, with the *CV* technique yielding, on average, the result with the least variability and bias.

Table 9.9. Comparison of consolidated results for Trial 2 (2,000 iterations; 25 observations per design point) for (a) coefficient errors, (b) model performance, and (c) optimization results.

(a)

		Standard Errors of Regression Coefficients									
		***s*(b_0)**	***s*(b_1)**	***s*(b_2)**	***s*(b_3)**	***s*(b_4)**	***s*(b_5)**	***s*(b_6)**	***s*(b_7)**	***s*(b_8)**	***s*(b_9)**
Mean	*OLS*	39.28	18.18	18.18	18.18	31.5	31.5	31.5	22.27	22.27	22.27
	WLS	24.56	13.49	14.11	15.45	21.7	21.46	22.2	18.3	20.36	19.2
	CV	28.93	13.88	14.06	14.17	----	25.79	26.9	15.87	15.81	16.4
% Error Reduction	*OLS* → *WLS*	43.11	32.49	29.39	22.69	37.31	38.01	35.86	25.22	16.8	21.57
	OLS → *CV*	33	30.56	29.63	29.09	----	25.48	22.29	35.18	35.39	33.01
Standard Deviation	*OLS*	22.57	10.45	10.45	10.45	18.09	18.09	18.09	12.79	12.79	12.79
	WLS	17.36	9.78	9.88	10.5	14.58	14.04	14.18	12.2	12.67	12.88
	CV	14.44	6.93	7.02	7.08	----	12.88	13.43	7.92	7.89	8.19
% Error Reduction	*OLS* → *WLS*	59.78	51.08	50.54	47.47	57.89	59.43	59.04	50.13	48.22	47.38
	OLS → *CV*	66.55	65.33	64.87	64.59	----	62.8	61.2	67.64	67.74	66.56

(b)

	Process Mean			**Process Standard Deviation**		
	OLS-m	***WLS-m***	***CV-m***	***OLS-s***	***WLS-s***	***CV-s***
Avg PRESS	345,490.00	268,920.00	144,710.00	95,499.00	102,320.00	31,709.00
Avg $R^2_{p,adj}$	0.89	0.96	0.94	0.16	0.70	0.60

(c)

		OLS	***WLS***	***CV***
Optimization Results	Avg MSE	515.76	32.21	31.56
	Avg Bias	1.61	0.17	0.14

Whereas the gap between the *WLS* and *CV* techniques regarding the optimization results appears to close with greater replication at each design point, the opposite is observed to occur in terms of the precision in the results. As the plots in Figures 9.8(a) – (c) show, the x* obtained via the *CV* technique continue to consolidate even more tightly in the (+, □, +) quadrant of the experimental region. Thus, while the average

optimization results between the two appear nearly equal, the precision achieved by the *CV* technique clearly exceeds the improvements attained through the *WLS* approach.

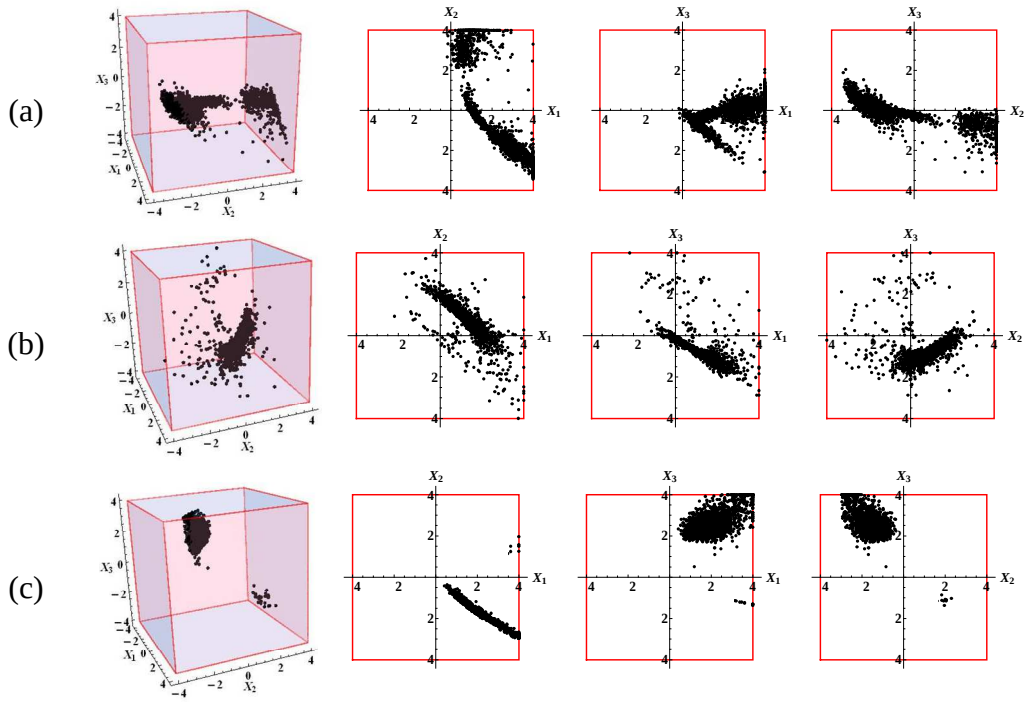


Figure 9.8. Identification of x* for Simulation Trial 2, $-4 \le X_i \le 4$; (a) x* for *OLS* method, (b) x* for *WLS* method, (c) x* for *CV* technique.

9.5 Summary of Findings

The results of the numerical example and associated simulations demonstrate that, while both the *WLS* method and proposed *CV* technique provide viable alternatives to traditional *OLS* regression approaches when inherently high variability pervades process operations, the proposed method can deliver significant improvements in the quality and precision of the *RPD* solutions. Whereas the *WLS* approach affords some flexibility in iteratively down-weighting sources of variability, it does not eliminate them altogether. The proposed *CV* technique, on the other hand, does. Consequently, removing sources of

variability followed by a restructuring of the experimental design through a combination of optimal designs and best-subsets regression can have three significant positive impacts, which are summarized in (i) – (iii) below:

(i) First, it tends to produce a model with a much greater degree of predictive capability regarding system outputs or responses (Y_i). In each aspect of the numerical example, the proposed technique resulted in a $PRESS_p$ that was up 46% less than the *WLS* approach for the mean response model, and up to 70% less for the standard deviation.

(ii) Second, the *CV* technique tends to generate the best *RPD* solutions in terms of minimal *MSE* and bias in the optimization results. Although the gap between the *WLS* and proposed *CV* approaches closes with the introduction of additional replication, it would still be fair to suggest the *CV* technique would be preferred due to the added costs that would be incurred to achieve such increases in experimental replication.

(iii) Finally, after 3,000 iterations of the experiment, it is clear that the *CV* technique achieves greater precision. That is, judging by the plots shown in Figures 9.7(c) and 9.8(c) relative to the corresponding plots for the *OLS* and *WLS* methods in each, there is a considerable enhancement in the degree of precision attained using the proposed method. This translates to a greater degree of certainty that the generated results are indeed optimal.

9.6 Conclusion

As has been a recurring theme in previous chapters, the presence of high-variability conditions can invalidate several of the assumptions that underpin the application of ordinary least squares regression. The pilot study presented in Chapter 3

and used as a motivating basis in this chapter illustrates that, if such variability is assumed away or otherwise unaccounted for, the application of traditional approaches can lead to significant imprecision both in the estimates of the fitted response surface models and ultimately the optimal operating conditions derived from them. To overcome this, an alternative technique based on the *CV* has been proposed in which engineers identify and then remove influential sources of variability and then apply optimal design techniques to restructure and balance the experiment in order to obtain more precise estimators and ultimately determine the true optimal operating conditions.

While the results of the proposed method demonstrate improvements, there appears to be room for yet greater improvements in the overall degree of precision achieved. Thus, future extensions to this method might explore the incorporation of higher-order modeling, or perhaps alternative experimental designs that would facilitate increased replication. In either case however, additional costs in time and resources are inevitable. As such, trade-offs between these costs and the degree of precision possible should be considered. This would ultimately help analysts to develop recommendations that offer decision makers a greater degree of flexibility and control in the decision process. Chapter 10 addresses these concepts in detail.

CHAPTER TEN

INTEGRATING A TRADE-OFF ANALYSIS BETWEEN COST AND PRECISION IN HIGHLY VARIABLE PROCESSES

10.1 Introductory Remarks

In the age of global markets and competition, efforts to ensure product and process adherence to technical specifications have yielded to a growing focus on ensuring that their associated characteristics adhere to desired target values with as little variability as possible. As noted in the outset of Chapter 9, variation and uncertainty in manufacturing and testing processes can have significant impacts in terms of costs to both the manufacturer and the user/vendor. Thus, precision achieved through variance reduction must be the paramount goal. Notwithstanding, in all problems of this nature, there exists a trade-off between precision and the costs required to achieve it. Typical costs are expressed in terms of the manufacturer's perspective and may include capital, equipment, operating costs, testing/experimentation costs, manufacturing throughput, etc. However, there are also potential costs for *not* achieving sufficient precision that affect the customer and so, by extenstion, the manufacturer. In particular, they include product overdesign and overmanufacture that stem from conservative decision-making in the manufacturing process, which can ultimately lead to customer dissatisfaction with the level of quality and loss of vendor confidence.

Taguchi's (1986, 1987) robust parameter design (*RPD*) approach evolved to provide a cost-effective approach for improving product and process quality via variability reduction. As has been noted in several chapters, the goal of *RPD* is to

achieve robustness by selecting control factor settings that help to render the system impervious to the effects of process variability. In the context of Taguchi's philosophical approach, the customer's perspective demands that manufacturers consider both the mean (or target specification) and the variability for a particular quality measure in order to improve the quality of the delivered product, as well as the quality of the process used to manufacture it. Thus, since the 1980's, the ideas and associated methods for considering both the process mean and variance have received much attention. Indeed, a great portion of this attention has focused on alternative methods for efficiently determining optimal process settings or conditions. In the interests of promoting various optimization approaches, the models used are assumed to be good fits and ordinary least squares assumptions hold.

As manufacturers seek improved methods for ensuring quality in resource-constrained environments, engineers should examine trade-offs to achieve the levels of precision that best support their decision making. In contrast to previous research, this chapter proposes a trade-off analysis between enhanced precision in the generated solutions and the costs associated with achieving it. Several techniques are considered in the early stages of experimental design, using Monte Carlo simulation as a tool, for revealing potential options to the decision maker. To this end, only situations involving controllable factors are considered for the purposes of comparison and analysis. While the inclusion of noise factors would more closely mirror reality, the literature lacks consensus on the most appropriate method for doing so. Accordingly, the focus of this chapter has been scoped. Notwithstanding, this research may represent the first study to

show the avenue which may lead to more effective robust parameter design models with the optimal combination of cost constraints and desired precision of solutions. The original research associated with this chapter is published with reference Boylan *et al.* (2013).

10.2 Brief Review of Optimization Schemes

Chapter 2 addressed some of the controversy surrounding Taguchi's methods, which led a number of researchers to conclude that Taguchi's methodology merely facilitated process improvement rather than optimization. Consequently, many of them proposed *RPD* models that investigated a variety of alternative optimization methodologies and were more firmly grounded in well-established approaches to design of experiments and response surface methodology (*RSM*). A summary of various optimization schemes observed in the contemporary *RPD* literature may be reviewed in Figure 2.2 on page 23. For the purposes of this chapter, shortened summaries of several of the more commonly applied schemes are readdressed in the following paragraphs.

Initial efforts by Vining and Myers (1990) focused on a priority-based scheme rooted in the dual-response component of *RSM*. Their approach constrained results to strict adherence to desired targets in the spirit of Taguchi's philosophy and assumed the following form:

$$\begin{aligned} &\text{Minimize} \quad \hat{\sigma}(\mathbf{x}) \\ &\text{Subject to:} \quad \hat{\mu}(\mathbf{x}) = \tau \text{ , where } \tau \text{ is the target value} \end{aligned} \tag{10.1}$$

Subsequent research by Cho (1994) and Lin and Tu (1995) demonstrated that allowing for some bias in process targets could facilitate even further reductions in variability.

The results were *MSE*-based schemes that allowed an unspecified degree of bias in process targets in the interests of reducing variability:

$$\text{Minimize} \quad MSE = (\hat{\mu}(\mathbf{x}) - \tau) + \hat{\sigma}^2(\mathbf{x}) \tag{10.2}$$

This sparked some debate regarding the amount of bias that could or should be allowed, which led Copeland and Nelson (1996) to propose a model that constrained allowable bias to some fixed level, Δ:

$$\begin{aligned} &\text{Minimize} \quad \hat{\sigma}(\mathbf{x}) \\ &\text{Subject to:} \quad (\hat{\mu}(\mathbf{x}) - \tau)^2 \leq \Delta^2 \text{, where } \Delta \text{ denotes the amount of allowable bias} \end{aligned} \tag{10.3}$$

Most recently, Costa (2010) proposed another method to simultaneously optimize the mean and standard deviation of a particular response based on the global criterion method developed by Tabucanon (1988):

$$\text{Minimize} \quad \left(\frac{\left| \hat{u}(\mathbf{x}) - \tau_\mu \right|}{U_\mu - L_\mu} \right)^{\omega_\mu} + \left(\frac{\left| \hat{\sigma}(\mathbf{x}) - \tau_\sigma \right|}{U_\sigma - L_\sigma} \right)^{\omega_\sigma} \tag{10.4}$$

where τ_i represents the target value and (L , U) denote the (lower, upper) specification limits for μ and σ. The incorporation of weights ω_μ and ω_σ establishes relative priorities between the mean and standard deviation, allowing experimenters to explore varying magnitudes for each, and thus evaluates trade-off analyses between them. Similarly, the inclusion of the ratio *1/ (U - L)* also affords practical flexibility to evaluate trade-offs between different limit settings. As noted in Chapter 2, many other researchers have developed extensions of these models, contributing to the growing wealth of knowledge in the field.

10.3 Concept Motivation – Revisiting the Printing Press Pilot Study

Recall from the motivating factors illustrated in Section 3.5 of Chapter 3 that many researchers have developed optimization procedures using data sets from highly variable processes. More importantly in the context of this chapter is that the developers of each of the subsequent optimization schemes outlined above claimed to achieve superior results compared to those achieved using previous models. Again, this is not to question the validity of the proposed procedures, as their proponents have clearly demonstrated improved results through a range of comparative analyses. However, it is noteworthy that the data used are so variable, that with only a handful of replications (typically three or four) a simulation employing these models would likely generate drastically different results spanning the breadth of the design space every time. Moreover, it is almost equally probable in such highly variable situations that any one of the schemes would produce the best result in a particular situation.

The numerical example provided in Chapter 3 serves to motivate these ideas. In particular, recall the simulation results from Section 3.5, which are recaptured below in Figures 10.1 and 10.2 and summarized in Table 10.2.

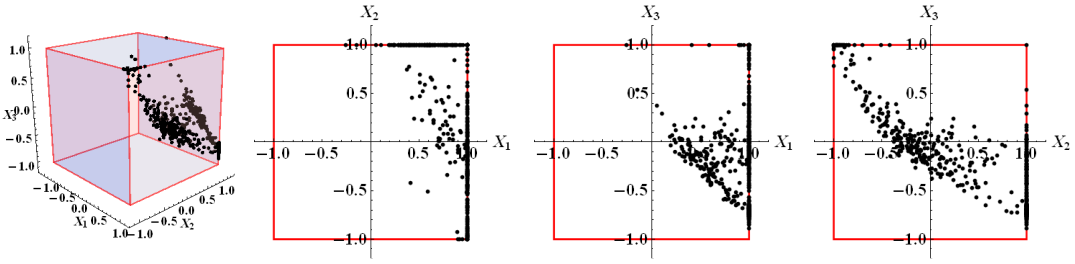
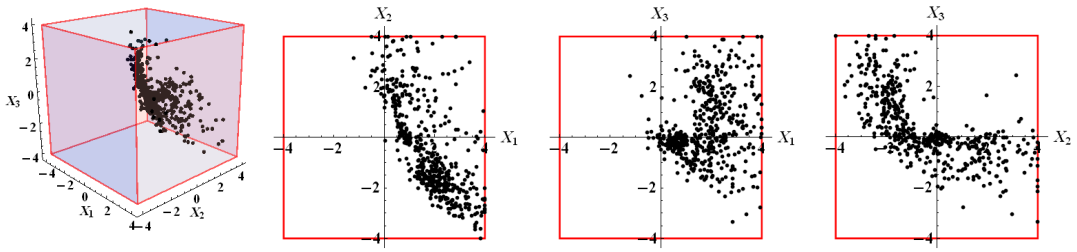


Figure 10.1. Identification of **x***, Printing Press Study, $-1 \le X_i \le 1$ (500 iterations).

Clearly, the imposed constraint of the experimental region is binding. However, expanding this region four-fold in each direction provides an extrapolated view that reveals far more in terms of the imprecision among the generated solutions. As noted, the solutions extend across a considerable portion of the feasible region. Repeating this simulation 1,000 times for each of the optimization schemes yields the results in Table 10.1, demonstrating that the inherent variability of this particular data set greatly influences the ability of any one optimization model to achieve superior results with any certainty.



(a) (b) (c) (d)

Figure 10.2. Identification of **x***, Printing Press Study, $-4 \le X_i \le 4$ (500 iterations).

Table 10.1. Model Comparison – Printing Press Study.

	Results (% of 1,000 iterations with Minimum *MSE*)				
Model	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
Vining and Myers (1990)	55.2%	58.1%	56.7%	55.8%	57.8%
Lin and Tu (1995)	39.2%	37.2%	37.3%	38.7%	37.0%
Copeland and Nelson (1996)	2.5%	0.9%	2.1%	2.2%	2.0%
Costa (2010)	3.1%	3.8%	3.9%	3.3%	3.2%

Once again, this brief illustration demonstrates the drawbacks in the design of the *RPD* problem for processes that possess high variability. Since the number of replications at each design point is low, repeating the experiment under the same

conditions (i.e., same number of replications) can result in the identification of settings that may cover the entire length or more of the feasible factor setting space. This indicates considerable imprecision in the estimated functions used to generate the optimal factor settings. Equally important, it creates a false perception in the superiority of one optimization model over another; the data are so variable that the results should not be trusted with any degree of certainty.

10.4 Development of Proposed Methodology

Focusing on processes exhibiting high degrees of variability, the methodology offered in this chapter expands upon the experimental design and model development phases of the *RPD* framework. Specifically, the development of preliminary estimators and models are considered as outputs from Phase I. Thereafter, a trade-off analysis is incorporated in Phase II, wherein the conditions of the experiment and the resulting preliminary models are evaluated to determine whether alternative measures are necessary to achieve greater precision. The end result is a three-phased procedure depicted in Figure 10.3. The specific components of each phase are discussed in the ensuing paragraphs, focusing in particular on Phases I and II.

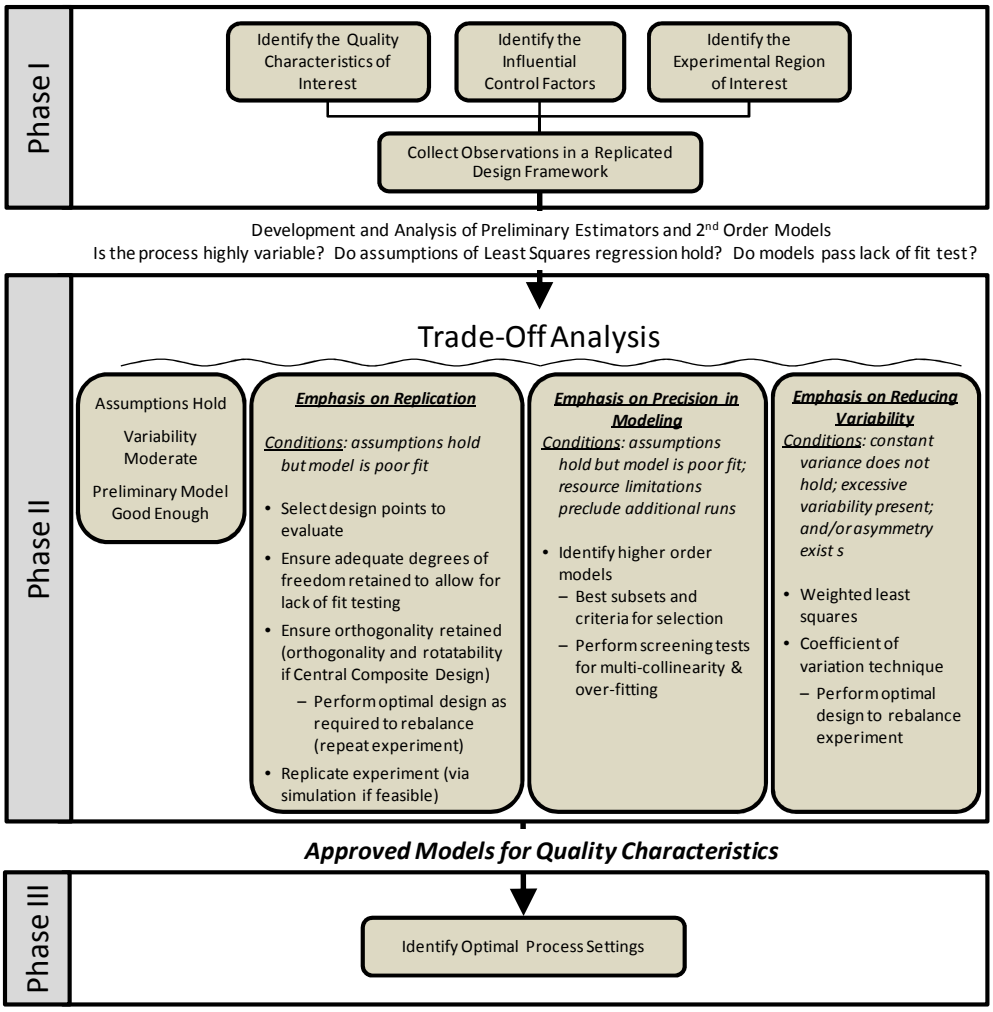


Figure 10.3. Proposed methodology for integrating trade-off analyses.

10.4.1 Phase I – Design Preparation and Experimentation

Consider an experiment in which observations are made on a particular quality characteristic of interest Y, which is influenced by a set of k control factors $\mathbf{x} = (X_1, X_2, ..., X_k)$ evaluated at d levels each. In traditional fashion, the experiment consists of N experimental runs or design points whereby $N=d^k$ for a full factorial design (i.e. d^k factorial points + dk axial points +c center points) for a central composite design or *CCD*. A total of m replications are performed at each design point. Let y_{jq} be the qth

observation at the *j*th design point for the quality characteristic, where $j = 1, 2, ..., N$, and $q = 1, 2, ..., m$. Values for the mean and variance are computed at the *j*th design point using the following formulas:

$$\bar{y}_j = \frac{\sum_{q=1}^{m} y_{jq}}{m} \qquad s_j = \sqrt{\frac{\sum_{q=1}^{m} (y_{jq} - \bar{y}_j)^2}{m-1}} \tag{10.5}$$

A format for this experimental design is shown in Table 10.2.

Table 10.2. Experimental Format.

Design Point	X_1 X_2 .... X_k	Replications	$\bar{y}$	s
1	Control Factor Settings	y_{11} ..... y_{1m}	$\bar{y}_1$	s_1
⋮		⋮ ⋮	⋮	⋮
q		y_{q1} ..... y_{qm}	$\bar{y}_q$	s_q
⋮		⋮ ⋮	⋮	⋮
N		y_{N1} ..... y_{Nm}	$\bar{y}_N$	s_N

Following experimentation and the collection of observations, preliminary estimators are developed for μ and σ in the form of second-order regression models which assume the same form shown in previous chapters, which is:

$$\hat{\mu}(\mathbf{x}) = \beta_0 + \mathbf{x}'\hat{\mathbf{b}}_\mu + \mathbf{x}'\hat{\mathbf{B}}_\mu\mathbf{x} + \varepsilon_\mu \tag{10.6}$$

$$\text{and} \quad \hat{\sigma}(\mathbf{x}) = \gamma_0 + \mathbf{x}'\hat{\mathbf{b}}_\sigma + \mathbf{x}'\hat{\mathbf{B}}_\sigma\mathbf{x} + \varepsilon_\sigma \tag{10.7}$$

where $\underset{\mathbf{1\times k}}{\mathbf{x}'} = \begin{bmatrix} X_1 & X_2 & \cdots & X_{k-1} \end{bmatrix}$, $\underset{\mathbf{k\times 1}}{\hat{\mathbf{b}}} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$, and $\underset{\mathbf{k\times k}}{\hat{\mathbf{B}}} = \begin{bmatrix} \beta_{11} & \beta_{12}/2 & \cdots & \beta_{1k}/2 \\ & \beta_{22} & \cdots & \beta_{2k}/2 \\ & & \ddots & \vdots \\ sym. & & & \beta_{kk} \end{bmatrix}$

At this point, with preliminary estimates formed, the experimental results and the preliminary models are analyzed to 1) determine if the process is significantly variable, 2)

verify that the ordinary least squares assumptions of independent normal model errors with constant variance hold, and 3) perform lack-of-fit testing on the estimators in Equations (10.6) and (10.7). The answers to these three questions will guide what direction we pursue in phase II. The approach to answering these three questions is described in the following paragraphs.

As has been noted in previous chapters, determining whether a process is highly variable is somewhat of a subjective matter delineated in objective terms. Recognizing that the observed variability stems from the agglomeration of a variety of sources throughout the process, both controlled and uncontrolled types are considered. Controlled variation pertains to a stable process exhibiting a pattern of random fluctuation about a constant level. Here, a stable process is one in which the process mean remains constant with consistent and predictable variation. Conversely, uncontrolled variation occurs in more structured patterns that change over time and is thereby unpredictable, rendering associated processes unstable. Using these characterizations of variability, we a highly variable process is considered to be either a stable one in which the controlled variation deviates $\pm 3\sigma$ from the mean response or more, or an unstable one in which the assumption of constant variance does not hold.

The validation of ordinary least squares assumptions follows traditional approaches. Specifically, normal plots of the residuals are used to verify their normality and independence, as well as variability plots of the residuals against the fitted regression line to verify homoscedasticity. Finally, to evaluate the appropriateness of the response

surface fit an *F*-test for lack of fit, a well-established method in the literature, is applied. In particular, for the mean response, the following hypotheses are tested:

$$\mathrm{H}_0 : \hat{\mu}(\mathbf{x}) = \beta_0 + \mathbf{x}'\hat{\mathbf{b}}_\mu + \mathbf{x}'\hat{\mathbf{B}}_\mu\mathbf{x}$$
$$\mathrm{H}_1 : \hat{\mu}(\mathbf{x}) \neq \beta_0 + \mathbf{x}'\hat{\mathbf{b}}_\mu + \mathbf{x}'\hat{\mathbf{B}}_\mu\mathbf{x}$$

using the well-known *F* test statistic: $F^* = \dfrac{SS_{LF}}{c-p} \div \dfrac{SS_{PE}}{N-c} = \dfrac{MS_{LF}}{MS_{PE}}$

in which SS_{LF} and SS_{PE} denote the lack of fit sums of squares and pure error sums of squares, respectively; MS_{LF} and MS_{PE} denote the lack of fit mean square and pure error mean square respectively; and where c = the number of distinct design points, and p = the number of parameters in the evaluated model. Using this test statistic, the appropriate decision rule is:

if $F^* \leq F(1-\alpha; c-p, N-c)$, fail to reject H_0 (conclude model fit is appropriate)
if $F^* > F(1-\alpha; c-p, N-c)$, reject H_0 (conclude model fit is not appropriate)

10.4.2 Phase II – Trade-off Analysis and Model Development

Typical *RPD* experiments assume that full second-order models will adequately estimate the response surface. Moreover, experimenters assume that a relatively few number of replications at each design point will produce sufficiently precise estimates and that the assumptions of normality and moderate variability in the responses and independent and identically distributed normal model errors hold. This is not to suggest that experimenters casually refrain from additional replications. In reality, limited resources and/or time may preclude additional experimental runs. However, in the context of that reality, the assumption is made out of necessity and used to underpin the adequacy of the obtained estimates. Given these assumptions, as evidenced by the

considerable emphasis placed upon optimization in the literature, researchers tend to proceed with their results into the optimization phase. While this may be done with little consequence in a stable process with low variability, many of the assumptions do not hold in cases involving more significant variability. Over the years, researchers and practitioners have observed that, in real-world settings, the assumptions of normality and moderate system variability quite often do not hold. On the contrary, industrial and manufacturing processes often possess elevated variability, particularly in mass production lines, which can confound many of the modeling assumptions behind the robust parameter design approaches. Furthermore, as has been noted in previous chapters, normality rarely exists and some asymmetry is not only realistic but practically inevitable – particularly in situations involving *S*-type, *L*-type, and certain instances of *N*-type quality characteristics; a condition that could become even more pronounced under conditions of elevated variability, especially when the number of observations or replicates at each design point is small. What is more, such conditions can induce the added effect of non-constant variance in the residuals. Thus, the use of generated models that do not account for these conditions may be problematic.

To overcome these potential pitfalls, we propose a trade-off analysis in which we evaluate the benefits of alternative means to achieve more precision and thereby less variability against the costs associated with doing so. We perform this analysis within three areas of emphasis which correspond to the answers to the three questions derived at the end of phase I. These areas include replication, precision in modeling, and

removing/reducing the influences of variability, which we describe in the subsections below.

10.4.2.1 Emphasis on Replication

Ideally, an experiment would retain all design points while including a sufficient number of replications to ensure a good model fit with minimal variability in the residuals. Usually, however, this is impractical due to time and other resource constraints. As such, when the second-order model developed pursuant to Phase I fulfills error assumptions of ordinary least squares yet is a poor statistical fit, we can consider reducing the number of design points evaluated in the experiment while increasing the number of replicates for each of the remaining points to obtain more precision. Figure 10.4 conceptualizes some of the potential alternatives.

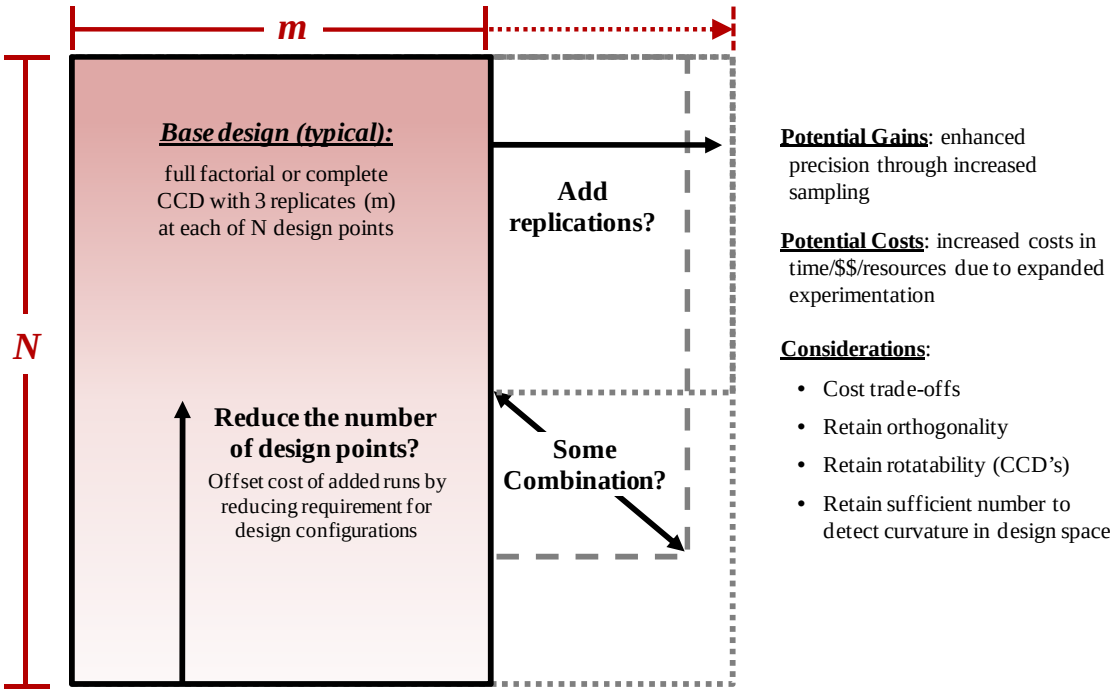


Figure 10.4. Potential alternatives for experimental designs.

For example, consider the printing press problem previously described. In this case, the experiment consisted of three replications of 27 design points in the full factorial design, for a total of 81 experimental runs. Alternatively, we could reduce the number of design points to 15 while doubling the number of replications to 6 by removing the edge points. Although this sacrifices the evaluation of these points and thus places greater weight on the corner and face-centered points, we are able to incorporate twice the replication at a net cost of 9 additional experimental runs. With more replication, we would expect to obtain greater accuracy in finding the true position of the optima as they relate to the factor space. Thus, the trade-off under consideration is the precision gained through increased replication versus the cost of the additional nine runs.

It is important to note that the selection of design points is not done arbitrarily. In fact, it must be done in a way such that the resulting design retains the properties of orthogonality and, in the case of *CCD*s, rotatability. Additionally, we would want to ensure that we retain sufficient points to adequately detect curvature in the design space.

In evaluating the trade-off between precision and cost, it would be beneficial to have an idea as to how much improvement in precision increased replication will yield before we commit to incurring the cost. Two questions arise from this: what level of precision is desired and how many replications are required to attain it? If we specify a target range for precision (or variability about the mean), we can use Monte Carlo simulation to aid in determining the number of replications required. Likewise, we can use simulation both to facilitate additional replications (at least in certain situations in which we can be reasonably assured of the distributional fit of the data at a particular

design point), as well as to better understand the precision of the generated solutions in terms of the process variability.

Within the field of simulation, Monte Carlo methods comprise a class of computational algorithms that use repeated sampling to compute simulated results. Monte Carlo simulation is considered a sampling method because the inputs are randomly generated from probability distributions to simulate the process of sampling from an actual population. Accordingly, it is important to choose a distribution for the inputs that most closely matches data already on hand, or best represents what is currently known about the process under examination.

Once we have refitted the second-order model, we would again examine the quality of the fit and the validity of the least squares assumptions. If these prove satisfactory, we can approve the model and proceed to optimization. Otherwise, additional trade-offs should be considered in the other areas of emphasis.

10.4.2.2 Emphasis on Precision in Modeling

The circumstances leading to this particular avenue in the trade-off analysis could be twofold. Either we have pursued increased precision through added replication and are still unsatisfied with the degree achieved through a second-order model, or we are satisfied that the number of replicates is sufficient but still require increased precision to correct a poor statistical fit. In response to either case, this area of emphasis examines the use of higher-order fitted response surface designs for the mean, standard deviation.

All of the models and extensions referenced in Section 10.2 are based on second-order fitting functions for modeling a response variable of interest. While second-order

functions are useful in finding optimal solutions in the presence of curvature for most processes, there are some emerging fields in the contemporary engineering and science environment, such as nano science and molecular biometrics, where significantly elevated levels of precision require alternative estimation methods. Goethals and Cho (2011) addressed this need, incorporating higher-order polynomial response surface methods combined with a desirability function approach to achieve higher precision in multi-response optimization problems. Rather than creating higher-order polynomial models simply by adding terms, their proposed method searches for the best combination of terms based upon the results of various screening criteria. In some cases, the number of model parameters that most appropriately estimates the true response may be less than or equal to the number of terms in the second-order model. By incorporating screening tests to espy potential issues in using higher-order terms along with a procedure using multiple criteria for model selection, they demonstrated increased assurance that increased precision with minimal bias will result.

Using the mean response vector $\bar{\mathbf{y}}$, the general form of the estimator for the mean with p regression coefficients or p–1 predictors may be written in terms of a design matrix $\mathbf{X}$ as:

$$\hat{\mu}(\mathbf{x}) = \mathbf{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{\mu}}, \text{ where } \hat{\boldsymbol{\beta}}_{\boldsymbol{\mu}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\bar{\mathbf{y}} \text{ and } \mathbf{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1,p-1} \\ 1 & X_{21} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{N1} & \cdots & X_{N,p-1} \end{bmatrix} \quad (10.8)$$

Using the standard deviation response vector **s**, we apply the same procedures to develop a higher-order fitted response surface design for the standard deviation.

$$\hat{\sigma}(\mathbf{x}) = \mathbf{X}\hat{\boldsymbol{\beta}}_{\sigma}, \text{ where } \hat{\boldsymbol{\beta}}_{\sigma} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{s} \tag{10.9}$$

We may use the hat matrix **H** with each of the fitted response surface designs to facilitate the identification of influential observations in the data as a result of their location within the factor space. Recognizing that the procedures are the same for both response functions, the mean response is used to demonstrate:

$$\hat{\mu}(\mathbf{x}) = \mathbf{X}\hat{\boldsymbol{\beta}}_{\mu} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\bar{\mathbf{y}} = \mathbf{H}\bar{\mathbf{y}}, \tag{10.10}$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is an *N* x *N* matrix. Additionally, denoting the vector of residual terms for the response surface function as **e**, we can express this vector as a linear combination of the response vector as follows:

$$\mathbf{e} = \bar{\mathbf{y}} - \hat{\mu}(\mathbf{x}) = \bar{\mathbf{y}} - \mathbf{H}\bar{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\bar{\mathbf{y}}, \tag{10.11}$$

where **I** is the identity matrix. Furthermore, since the idempotent properties of the hat matrix show that (**I**–**H**)(**I**–**H**) = **I**–**H**, we can express the error sum of squares (*SSE*) in the analysis of variance for a given response surface model as $SSE = \mathbf{e}'\mathbf{e} = \bar{\mathbf{y}}'(\mathbf{I} - \mathbf{H})\bar{\mathbf{y}}$ with *N*–*p* degrees of freedom. By incorporating a unity matrix **J** and using the fact that $(\hat{\boldsymbol{\beta}}_{\mu})'\mathbf{X}' = (\mathbf{X}\hat{\boldsymbol{\beta}}_{\mu})' = (\mathbf{H}\bar{\mathbf{y}})' = (\bar{\mathbf{y}})'\mathbf{H}$, it can also be shown that the regression sum of squares (*SSR*) and the total sums of squares (*SSTO*) may be written as:

$$SSR = (\hat{\boldsymbol{\beta}}_{\mu})'\mathbf{X}'\bar{\mathbf{y}} - \left(\frac{1}{N}\right)(\bar{\mathbf{y}})'\mathbf{J}\bar{\mathbf{y}} = (\bar{\mathbf{y}})'\left[\mathbf{H} - \left(\frac{1}{N}\right)\mathbf{J}\right]\bar{\mathbf{y}} \tag{10.12}$$

$$SSTO = SSE + SSR = (\bar{\mathbf{y}})'(\mathbf{I} - \mathbf{H})\bar{\mathbf{y}} + (\bar{\mathbf{y}})'\left[\mathbf{H} - \left(\frac{1}{N}\right)\mathbf{J}\right]\bar{\mathbf{y}} = (\bar{\mathbf{y}})'\left[\mathbf{I} - \left(\frac{1}{N}\right)\mathbf{J}\right]\bar{\mathbf{y}}$$

each with *p*–1 and *N*–1 degrees of freedom, respectively.

For the resulting response surface design, we retain only those terms that contribute to the regression for further analysis. In order to determine the proper subset of terms and compare the results of the estimation, we can examine an array of evaluation criteria. For our purposes, we use five well-established criteria to evaluate the model with p parameters. These include the coefficient of determination R_p^2, the adjusted coefficient of determination $R_{a,p}^2$, the prediction sum of squares $PRESS_p$, the mean square error MSE_p, and the F-test for lack of fit. Table 10.3 below summarizes these criteria, several of which were addressed previously in Chapter 9.

Table 10.3. Evaluation Criteria for the Model with p Parameters.

Measure (with Formula)	Description / Interpretation
$R_p^2 = \frac{SSR_p}{SSTO}$	High values are sought. Although the values can never decrease as predictor variables are added to the model, it helps to identify the point where adding more variables produces no meaningful effect.
$R_{a,p}^2 = 1 - \left(\frac{N-1}{N-v}\right)\frac{SSE_p}{SSTO}$	High values are sought. In contrast to the R^2 measure above, this measure is not influenced by an increase in the number of predictor variables.
$PRESS_p = \sum_{j=1}^{N}\left(\frac{e_j}{1-h_{jj}}\right)^2$	Low values are sought. Models that have smaller $PRESS_p$ values contain less overall prediction error in their estimation.
$MSE_p = \frac{SSE_p}{N-p}$	Low values are sought. Measures the average of the square of the error for a model with p parameters.
$F^* = \frac{MSLF}{MSPE}$	Tests whether the multiple regression model is an appropriate response surface. If $F^* \leq F(1-\alpha; c-p, N-c)$ then conclude that the model is a good fit. Otherwise, the model is deemed a poor fit.

Potential issues in using higher-order polynomials, such as multi-collinearity and the increased presence of outliers within a given factor space, have tended to steer researchers away from investigating the use of higher-order designs. Thus, once we identify a suitable model based upon the results of the evaluation criteria, we must perform two screening tests. The first determines if multi-collinearity exists between the

predictor variables as a result of using a higher-order model. The variance inflation factor (*VIF*) is a frequently used measure for this. When considering standardized estimated regression coefficients, it can be shown that the vth diagonal element of the covariance matrix of regression coefficients, for $v = 1, 2, ..., p-1$, is the variance inflation factor given by $(VIF)_v = 1/(1-R_v^2)$, where R_v^2 is the coefficient of multiple determination, obtained when one predictor variable is regressed against the other $p-2$ variables in the model. If no linear relation exists between a predictor variable and the other variables in the model, then $R_v^2 = 0$ and $(VIF)_v = 1$; if the predictor variable has a perfect linear association with the other variables in the model, then $R_v^2 = 1$, and $(VIF)_v$ is unbounded. Generally, it is accepted that if several *VIF* values are greater than 10 for a given model, then multi-collinearity exists between the predictor variables and presents an issue. Should this occur, we may remove one or more of the correlated predictor variables, use a centering technique, or employ remedial measures such as ridge regression.

The second screening test determines if influential observations exist within the factor space. In particular, we examine the hat matrix **H** for the selected model with p parameters to identify if certain X observations acting as outliers possess excessive leverage for influencing the response surface functions. Here, larger values for the diagonal elements of the hat matrix, h_{jj}, indicate a smaller difference between the fitted value and the observed value. Leverage values greater than $2p/N$, however, point to the presence of X outliers and merit the use of robust regression procedures. Once we implement remedial measures to lessen undue influence in the factor space, the approved model is ready for optimization.

10.4.2.3 Emphasis on Removing Influences of Variability

Increasing replication and employing higher-order response surfaces to achieve greater precision likely will induce reduced variability, as well. However, it may happen that reducing influences of variability is a primary concern. As such, approaches to eliminating variability, such as those discussed in Chapter 9, are considered as a unique area of emphasis by which to obtain more precise results, particularly in situations in which inherent process variability is high and/or error variances are shown to be non-constant.

As discussed in Chapter 8, high variability can be assessed based on the mean and standard deviation of the responses. If one or more responses falls 3σ or farther from the mean response, the process is classified as highly variable. Non-constant variance is typically assessed in several ways. From a graphical standpoint, as has been illustrated in several chapters previously, plotting the residuals against the fitted line or plotting the squared residuals against each of the separate factors will aid in identifying inconsistencies in the residual variances. More objective techniques such as the Breusch-Pagan test provide a way to evaluate the error variance for constancy using hypothesis testing and test statistics. This test assumes independence and normality among the residuals, but further assumes a relationship between the error variance σ_j^2 and the level of *X* that, using the second-order model at Equation (10.8), takes the form

$$\log_e \sigma_j^2 = \mathbf{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{\mu}} \tag{10.13}$$

which implies that the error variance fluctuates up or down with X, depending upon the sign of the associated coefficients in $\hat{\boldsymbol{\beta}}_{\mu}$. Since constant error variance corresponds to the instance in which each of the coefficients contained in $\hat{\boldsymbol{\beta}}_{\mu}$ equal 0, we establish the following hypothesis test for evaluation:

$$\mathrm{H}_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$$
$$\mathrm{H}_1 : \text{ not all } \beta_v = 0$$

The test statistic used to evaluate the hypothesis test consists of

$$\mathrm{X}^2_{BP} = \frac{SSR^*}{2} \div \left(\frac{SSE}{Nm} \right)^2 \qquad (10.14)$$

in which *Nm* denotes the total number of experimental observations and we obtain the regression sum of squares (*SSR**) by regressing the squared residuals, e_j^2, against *X* and the error sum of squares (*SSE*) by regressing the mean response on *X*. Comparing this statistic to the chi-square distribution with φ degrees of freedom (where φ = the number of predictor variables in the model) the decision rule for a significance level of α=0.05 is as follows:

if $\chi^2_{BP} \le \chi^2_{(1-\alpha),\varphi}$ then fail to reject H_0 and conclude that error variance is constant
if $\chi^2_{BP} \le \chi^2_{(1-\alpha),\varphi}$ then reject H_0 and conclude nonconstant error variance

In processes with inherently high variability, the assumption of constant variance in the residuals would most likely not hold and would thereby necessitate the use of remedial measures.

Chapter 9 provided considerable detail regarding approaches for dealing with high-variability conditions, and so those particulars will not be repeated here. As

discussed, investigators traditionally invoke one of two basic approaches to obtaining improved estimators in situations involving heteroscedastic error variances: transformation of the response variable, *Y*, to equalize the error variation across all predictor variables, or using weighted least squares (*WLS*) in the estimation process to account for unequal error deviations. While either method may work well in a broad variety of applications, the *WLS* approach is most often selected in order to avoid the possibility of an inappropriate regression relationship which can result from transformations of *Y*.

In addition to the *WLS* method, the coefficient of variation technique proposed in Chapter 9 is considered once again here for the purposes of examining the trade-offs between enhanced precision and the costs associated with achieving it. An interesting distinction between the *CV* technique and the *WLS*/transformation approaches discussed above is that while the latter are invoked *after* the experiment has been conducted and fitted models have been estimated based on a pre-specified experimental framework, the *CV* technique is applied *during* the experimental phase. Thus, the engineer is afforded a greater degree of flexibility in shaping experimental conditions to indentify and mitigate influential sources of variability early on. Using this technique, the mean and standard deviation of the responses at each design point (presuming multiple replicates are obtained at each) are used to determine the *CV* by:

$$CV_j = \frac{s_j}{\bar{y}_j},$$

which in turn is used to identify the most influential sources of variability. The *CV* for a quality characteristic aims to describe the dispersion of the characteristic in a way that does not depend on its measurement unit. That is, it describes data variability in terms of the relative sizes of the squared residuals and outcome values; the higher the *CV*, the greater the dispersion in the response. Conversely, the lower the *CV*, the smaller the residuals relative to the predicted value, suggesting a sufficient relationship exists between the estimated and true responses.

As noted in Chapter 9, no definitive threshold exists for the *CV* that indicates elevated or high variability. Instead, it will depend on and vary with the process being examined. For the purposes set forth for this research, the *CV* is used to identify sources of variation that unduly influence the response. The associated design points are removed from consideration, effectively scoping the design space, where after optimal design is then applied to rebalance the experiment in the most efficient way possible.

The need to consider optimal design strategies typically arises either when certain conditions render the use of traditional experimental approaches inappropriate or non-standard experimental regions exist. Such conditions include instances where time, resource limitations, or some physical restrictions constrain the experimental region. This particular situation, however, involves a trade-off analysis between the number of design points evaluated and increased replication to eliminate sources of variability and thereby achieve improved precision in modeling. Hence, we essentially are imposing constraints on the design space in the interest of placing greater emphasis on replication. While an assortment of optimal designs exists for consideration, the objective is not to

discuss or explore the full range of alternatives in this chapter. Nevertheless, in the interests of not confining the methodology to a singular design approach, *D*-, *A*-, and *G*-optimal designs are included in the discussion later in this chapter.

In some ways, as shown in Figure (10.3), this component in the methodology, implies that we have already performed some experimentation and determined 1) that the assumptions of moderate to low variability and homoscedasticity fails to hold and, consequently, 2) that the estimated model is a poor fit. Thus, invoking this particular approach further implies that we would revisit the experiment in order to redefine the experimental parameters and associated design space to reduce variability and achieve greater precision. Experimental designs, such as those heretofore described, are a viable alternative for a variety of situations in which standard experimental approaches do not apply or cannot achieve adequate precision. Generally, they provide an efficient and cost-effective way to select design points for experimentation. We should note, however, that the use of optimal designs should not serve as a direct substitute for standard experimental approaches, but rather as an alternative approach when standard methods are inappropriate or otherwise insufficient.

The final decision as to which optimal design to use may become subjective based on the preferences of the investigator and the computational complexities involved. Whatever the approach selected, efficiency ratios can then be used as a tool in the final selection of a particular design.

10.4.2.4 Final Determination of Approved Models

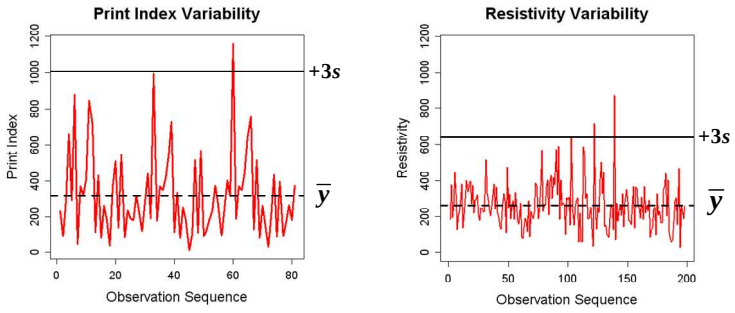
It is important to note that the areas of emphasis described may be considered singularly or by way of some combination. That is, an initial emphasis on replication does not necessarily preclude the inclusion of higher order response surfaces to further enhance precision, or vice versa, as the numerical examples in Section 10.5 will demonstrate. Regardless of the trade-off approach used, the three questions posed in Phase I would need to be reevaluated to ensure assumptions hold and that the generated models procure a sufficient statistical fit. If they do, then they are ready for optimization. If they do not, then additional remedial measures may be needed.

10.4.3 Phase III – Optimization

Using the design of experiments-based methodology described in Phase I and the trade-off analysis described Phase II, the approved model may be optimized using the scheme of choice discussed in the literature review to determine optimal process settings.

10.5 Numerical Examples

Using the methodology outlined in Section 2, the compromise or trade-off of various designs in solving *RPD* problems is investigated in this section. The first example revisits the printing press study discussed previously. Thereafter, a more recent experiment is investigated involving the manufacture of semiconductor wafers. As Figure 10.5 shows below, each of the processes associated with these examples qualified as a highly variable process based on the definition provided in Section 10.4.1. While these graphs only address the variability in terms of $\pm 3\sigma$ from the mean, the issue of non-constant variance in each will be addressed later.



(a) (b)

Figure 10.5. Variability observed in the printing press (a) and semiconductor (b) studies.

In each case, a thorough sensitivity analysis is performed to facilitate the selection of an alternative design; several techniques are implemented to achieve higher precision without significantly altering the general framework of the experiment. Monte Carlo simulation, developed using a combination of *MS Excel* and *Mathematica* [49], is used to compare and contrast the solution space for each example. In short, the regressions are performed for each iteration in excel, using the *Excel*-based random number generator to simulate observations from a normal distribution based upon the mean and standard deviation for each design point in Table 10.1. Thereafter, the generated regression functions for the mean and standard deviation are pulled into *Mathematica* using a built-in feature that allows an interface with *Excel*, and the optimization schemes are executed in each iteration and the results recorded.

10.5.1 The Printing Press Study Revisited

As previously documented, numerous researchers have used the printing press study to advance a variety of extensions to *RPD* optimization models and techniques. While the focus of research effort has primarily been given to suggesting different algorithms or models for achieving optimality, this investigation seeks alternative

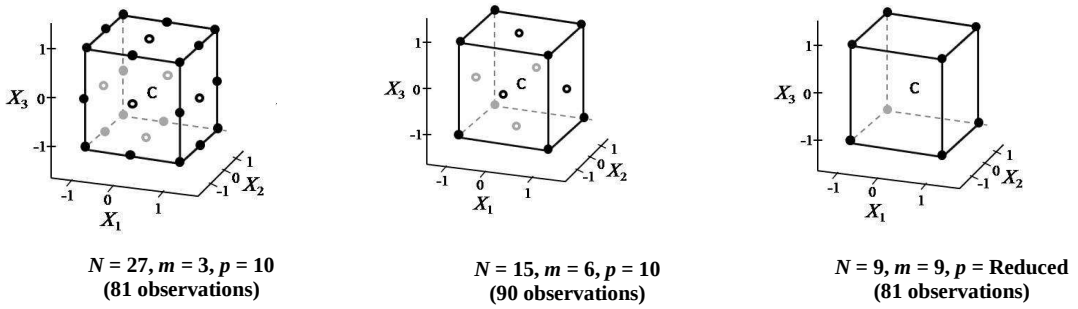
constructs of the design matrix **X** in the early phases of experimentation. In the following subsections, we discuss several variations of the experimental design in the context of the three areas of emphasis described in Section 10.4.

10.5.1.1 Emphasis on Replication

The general full factorial experiment in *VM* (1990) captures all possible combinations of the three factors at three distinct levels. With three replications at each design point, the experiment evaluates a total of 81 observations. However, as previously shown in Figures 10.1 and 10.2, using simulation to repeat the experiment 500 times generates optimal solutions that are dispersed throughout all four quadrants of the factor space. It is noteworthy that the objective function values also range from $1.75 \leq MSE \leq 4790.67$ with a mean of 985.75. To reduce the error incurred by performing only three replications at each design point, we can make several modifications. As it is still possible to detect curvature in the response using the center, factorial, and face points, one modification is to eliminate the edge points and perform the experiment with six replications. Alternatively, we can eliminate the face and edge points and perform nine replications at each of the design points. These modifications in the experimental design of the factor space are shown in Figures 10.6 (b) and (c), as they compare to the original design (a).

With added emphasis on replication, we would expect greater accuracy in finding the true position of the optima relative to the factor space. In each of the alternative designs, we retain the principle of orthogonality as it relates to the linear and interaction terms within the information matrix $\mathbf{X'X}$. While a trade-off is established between the

number of design point evaluations and the number of replications at each design point, the number of total observations remains relatively stable (81-90).



(a) (b) (c)

Figure 10.6. Original Design (a), versus Alternative Design 1 (b), and Alternative Design 2 (c) (corner and edge points (•), face points (**o**), and center point (**c**)).

Using the same full second-order model (p = 10) established for the original design, alternative design 1 examines doubling the number of replications and reducing the design point evaluations to the corner, face, and center points. Using Monte Carlo simulation, 500 iterations of the modified experiment are performed, using a random number generator to create six random observations that follow the mean and standard deviation specified for each design point in Table 10.1. Thus, each iterate consists of 15 x 6=90 experimental runs compared to the original 81. At a cost of nine additional runs per iterate, the solution space appears to confine itself primarily to the (+, +, –) quadrant within the expanded experimental region (see Figure 10.7).

Increasing the amount of replication, while maintaining some degrees of freedom in the design of the regression models, requires us to seek reduced fitted functions. Vining and Myers (1990) and other researchers recognized the inefficiencies in the statistical fit of the mean and standard deviation for the printing press study; their

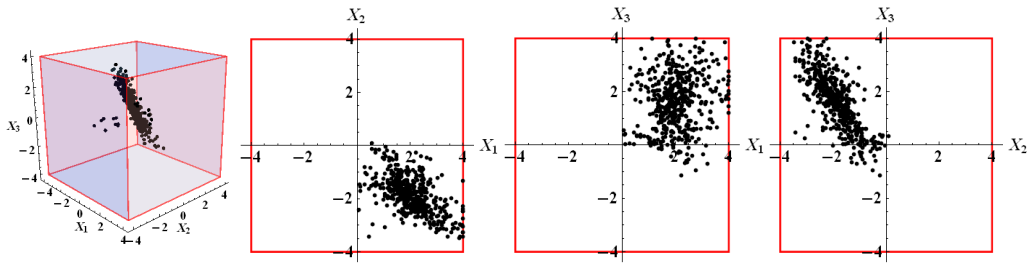


Figure 10.7. Identification of **x***, Alternative Design 1, $-4 \le X_i \le 4$ (500 iterations).

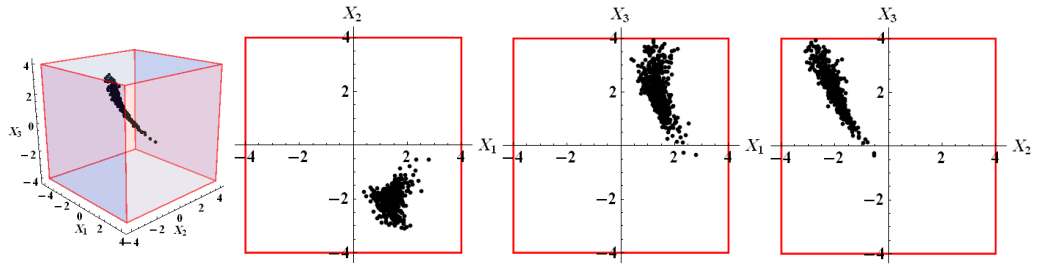
objective was to demonstrate the effectiveness of a particular optimization framework or algorithm in determining the optimal factor settings.

An analysis of the full second order model for the mean indicates that the pure quadratic terms are insignificant with respect to the fit of the surface. Furthermore, none of the terms pertaining to the standard deviation response are considered worthy of retaining; the variability of the observations reduces the effectiveness of ordinary least squares regression in estimating the measure.

In order to implement alternative design 2, which seeks even greater emphasis on replication, the models for the mean and standard deviation are modified, retaining only the linear and interaction terms. As demonstrated using the four criteria in Table 10.4, some additional precision is lost in terms of the predictive capability of each model. However, the trade-off is in ensuring a low deviation between the true response and the observed response obtained experimentally. Without duplicating a design point and increasing *N* for the printing press study, a standard lack of fit test is not feasible for this example. The results of the simulation using Alternative Design 2 are shown in Figure 10.8.

Table 10. 4. Model Comparison (Trade-off between Precision and Replication).

Response	*Model*	*N*	*p*	R_p^2 (%)	$R_{a,p}^2$ (%)	$PRESS_p$	MSE_p
Mean	Previous Research (2nd order)	27	10	92.7	88.8	337789	5794
	Reduced (2nd order)	9	7	91.8	64.6	2869038	29332
Standard Deviation	Previous Research (2nd order)	27	10	45.4	16.5	93691	1921
	Reduced (2nd order)	9	7	50.4	0.0	382458	4946



(a) (b) (c) (d)

Figure 10. 8. Identification of **x***, Alternative Design 2, $-4 \le X_i \le 4$ (500 iterations).

10.5.1.2 Emphasis on Precision

As noted, even when using the full second-order polynomial to model the mean and standard deviation, the statistical fit is not desirable. To improve the fit of the estimate, we can apply higher-order polynomial functions. In order to select an appropriate higher-order model, we must first perform several screening tests. A design matrix is constructed with special attention given to preventing the aliasing of terms, so that the columns of **X** remain linearly independent. Variance inflation factors are calculated for each of the predictor variables to ensure no multi-collinearity exists in a particular model. The diagonals of the hat matrix **H** are also analyzed to prevent values that negatively influence the regression function from exerting excessive leverage. A "best subsets" analysis using the resulting terms as candidates then yields a fourth-order and third-order model for the mean and standard deviation, respectively:

$$\hat{\mu}(\mathbf{x}) = 333.40 + 171.40X_1 + 162.67X_2 + 155.63X_3 + 75.21X_1X_2 + 85.71X_1X_3 + 49.71X_2X_3 - 69.07X_2^2 - 76.37X_1^2X_2 + 82.79X_1X_2X_3 + 42.04X_1^2X_2^2$$

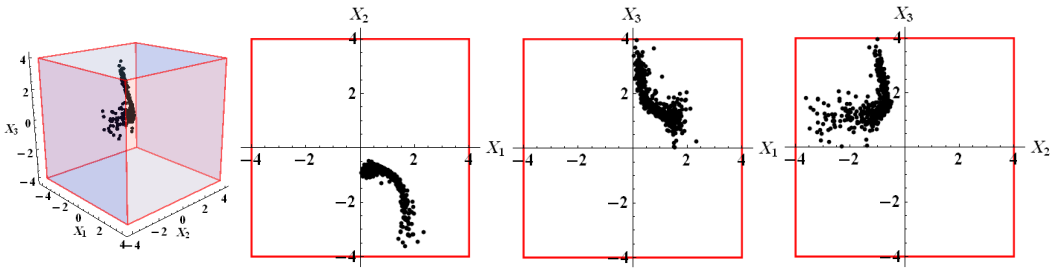
$$\hat{\sigma}(\mathbf{x}) = 53.04 + 38.74X_1 + 35.47X_2 + 20.80X_3 + 21.00X_1X_2 - 30.28X_1^2X_2 - 38.99X_1X_3^2 + 29.57X_1X_2X_3$$

With all $VIF \leq 5.00$, and the diagonals of the hat matrix $h_{jj} \leq 2p/N$, we compare the higher-order polynomials to the full second-order models used by previous researchers (shown in Table 10.5). The results reflect clear improvement in precision with the higher order model.

Table 10.5. Comparison of Models for the Mean and Standard Deviation.

Response	*Model*	*p*	R_p^2 (%)	$R_{a,p}^2$ (%)	$PRESS_p$	MSE_p
Mean	Prev. Research (2nd order)	10	92.7	88.8	337789	5794
	Higher-Order (4th order)	11	99.6	98.5	0	871
Standard Deviation	Prev. Research (2nd order)	10	45.4	16.5	93691	1921
	Higher-Order (3rd order)	8	73.2	46.3	21882	1079

With greater emphasis on precision in estimating each response, the models are used in conjunction with alternative design 1; the results of the Monte Carlo simulation are shown in Figure 10.9.



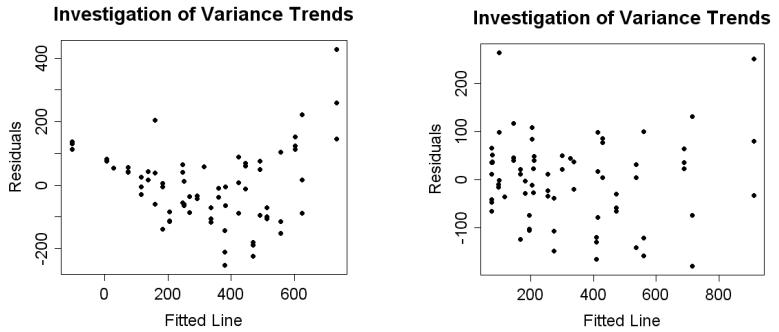
(a) (b) (c) (d)

Figure 10.9. Identification of **x***, Higher-Order with Alternative Design 1, $-4 \leq X_i \leq 4$ (500 iterations).

In contrast to the same design used previously in 3.1.1, the higher-order design appears to further refine the optimal solution space in the (+, +, –) quadrant.

10.5.1.3 Emphasis on Removing Variability Influences

Recall that *RPD* researchers examining this particular case study primarily sought improvements in the optimization scheme or numerical algorithm used to generate solutions. Since least squares regression is the driving tool in acquiring response surface models for further analysis, it is implied that the residuals are uncorrelated and exhibit relatively constant variance. An investigation of these assumptions for the printing press study, depicted in the residual plots in Figure 10.10, reveals evidence of non-constant variance.



(a) (b)

Figure 10.10. Investigating residual variances in the (a) first-order (b) second-order models.

To confirm the variability trends observed in the residual plots for both the first and second order models, we can apply the Breusch-Pagan hypothesis test. In the first order model, the calculated statistic (1) tests H_0: $\gamma_1 = \gamma_2 = \gamma_3 = 0$ versus H_1: Not all $\gamma = 0$, whereas in the second order model (2), the hypotheses are extended for $p = 10$ parameters.

$$(1)\ \chi^2_{BP} = \frac{7.953e{+}09}{2} \div \left(\frac{1012557}{27(3)} \right)^2 = 25.447 \qquad (2)\ \chi^2_{BP} = \frac{2.282e{+}09}{2} \div \left(\frac{539518}{81} \right)^2 = 25.720$$

Since $25.447 > \chi^2(.95, 3) = 7.81$ and $25.720 > \chi^2(.95, 9) = 16.92$ for the first and second-order models, respectively, we conclude H_1; that is, the error variance is not constant. The results of these tests suggest the need for remedial measures in the development of fitted functions for this example.

To remove the influence of non-constant variance, we invoke a weighted least squares approach to develop second order models for both the mean and standard deviation. This yields the following:

$$\hat{\mu}(\mathbf{x}) = 314.21 + 174.32X_1 + 136.49X_2 + 127.17X_3 + 37.49X_1X_2 + 47.76X_1X_3 + 67.12X_2X_3 + 28.46X_1^2 - 8.56X_2^2 - 13.95X_3^2$$

$$\hat{\sigma}(\mathbf{x}) = 44.79 + 16.09X_1 + 24.19X_2 + 24.78X_3 - 5.75X_1X_2 + 7.77X_1X_3 + 20.73X_2X_3 + 1.68X_1^2 - 5.15X_2^2 + 8.84X_3^2$$

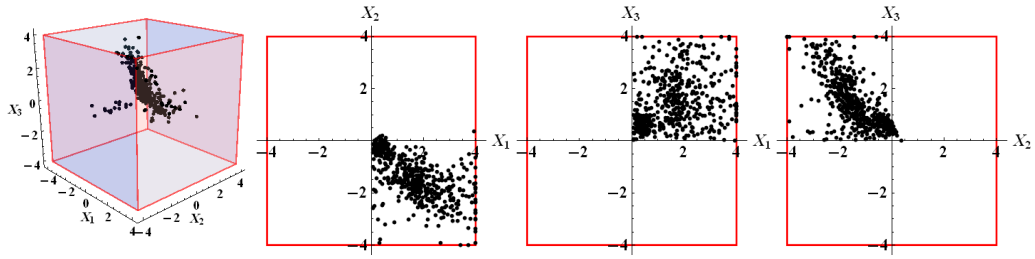
The result, generated using ANOVA and shown in Table 10.6, is a considerable reduction in the error associated with the coefficients of these models.

Table 10.6. Comparison of *OLS* and *WLS* Estimates (Coefficient Error).

Model	$s(b_0)$	$s(b_1)$	$s(b_2)$	$s(b_3)$	$s(b_4)$	$s(b_5)$	$s(b_6)$	$s(b_7)$	$s(b_8)$	$s(b_9)$
Mean - *OLS*	38.72	17.92	17.92	17.92	21.95	21.95	21.95	31.04	31.04	31.04
Mean - *WLS*	23.82	13.56	14.12	15.44	18.11	20.08	18.85	21.75	19.78	22.87
Std Dev - *OLS*	22.31	10.33	10.33	10.33	12.65	12.65	12.65	17.89	17.89	17.89
Std Dev - *WLS*	17.62	8.71	8.45	9.42	9.48	11.21	10.63	12.53	15.34	13.94

Once again using alternative design 1 to run the simulation, a random seed is used to generate models based upon the weighted least squares approach. With each iterate, the observations are randomized in accordance with the mean and standard deviation values in Table 10.1 and then *WLS* models are constructed. When the experiment is repeated

500 times using the second order models, the result is the solution space shown in Figure 10.11.



(a) (b) (c) (d)

Figure 10.11. Identification of **x***, Alternative Design 1 (*WLS*), $-4 \le X_i \le 4$ (500 iterations).

Although the results, in terms of the solution space shown in Figure 10.11, appear comparable to those previously discussed in Figure 10.7 from Section 10.5.1.1, they still are considerably better than those presented for the original study in Figure 10.2.

10.5.2 Examining a Semiconductor Manufacturing Study

Another example, first introduced by Robinson *et al.* (2006) and recently revisited by Lee *et al.* (2011) and Anderson and Whitcomb (2008), is a manufacturing study involving the fabrication of wafers in a semiconductor etching process. As wafers produced over time vary highly in terms of resistivity (Y), the objective of the experiment was to determine the factor settings that support achieving a mean resistivity of 350 with little to no variation. The experiment examined three control factors known to influence Y, specifically the gas flow rate (X_1), the temperature (X_2), and pressure (X_3). In order to develop models for the mean and standard deviation of resistivity, a replications-based experiment was performed using a central composite design with four center points.

Table 10.7 delineates the experimental framework, outlining the calculations for the mean and standard deviation at each design point.

As observed in most *RPD* studies, suppose the experiment involves three replications at each design point. The goal is to determine the optimal factor settings that minimize the estimated *MSE* described in Equation (10.2). Using the calculations established in Table 10.7, a simulation is used once again to repeat the randomized collection of observations at each design point based upon the characteristic mean and standard deviation. As in the previous example, the experimental region, $\mathbf{x}'\mathbf{x} \leq 3$, established by previous researchers and used in the simulation proves to be binding. Figure 10.12 shows the result when the region is expanded to observe the entire solution space for the optimal factor settings.

Table 10.7. Experimental Design for the Semiconductor Manufacturing Study.

	Coded Units				
	Flow Rate	Temperature	Pressure	Mean	Standard Deviation
Run	X_1	X_2	X_3	$\bar{y}$	s
1	–1	–1	–1	263.986	107.424
2	1	–1	–1	390.029	96.107
3	–1	1	–1	205.776	66.990
4	1	1	–1	292.526	110.058
5	–1	–1	1	290.104	141.334
6	1	–1	1	302.320	147.236
7	–1	1	1	164.656	80.455
8	1	1	1	160.369	82.632
9	–1.6818	0	0	211.039	57.150
10	1.6818	0	0	272.076	53.421
11	0	–1.6818	0	293.782	68.930
12	0	1.6818	0	147.133	39.404
13	0	0	–1.6818	418.555	221.964
14	0	0	1.6818	273.064	193.890
15	0	0	0	272.013	62.773
16	0	0	0	236.457	81.860
17	0	0	0	250.016	73.985
18	0	0	0	315.559	99.106

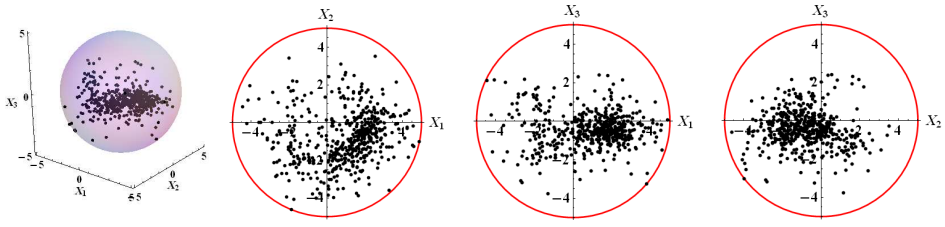


Figure 10.12. Identification of **x***, 3 replications with original design, $\mathbf{x}'\mathbf{x} \leq 25$ (500 iterations).

The variability associated with the design points for the semiconductor manufacturing study is extremely high. Whereas the design points in the printing press study had a range in the coefficient of variation of $0 \leq CV \leq 0.61$ with a mean of $\overline{CV} = 0.18$, the same statistics for this example are $0.20 \leq CV \leq 0.71$ with a mean of $\overline{CV} = 0.37$. In both cases, the median and mean *CV*s are close in approximation; the presence of outliers in the upper quantile of design points is impacting the ability to generate consistency in solutions with a small number of replications. In the semiconductor study, we note that the most significant outlier with respect to *CV* occurs when X_3 (i.e., the etching pressure) is at its highest level (1.6818). For a visual comparison, Figure 10.13 presents the *CV* box and whisker plots with the median highlighted using a dashed line.

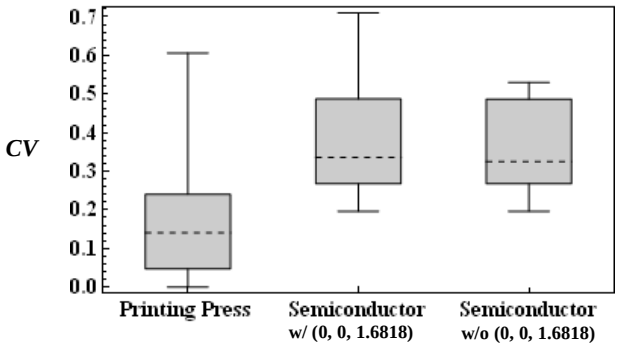


Figure 10.13. Box and whisker plots for *CV* - printing press and semiconductor studies.

Given the degree of variability in the solutions and the fact that $N = 18$ for this example, we would most likely gain very little by reducing the design for the sake of additional replications. The benefit of rotatability in using the central composite design is a desirable property that should be retained if possible. Simply resorting to the use of higher order models in estimating the mean and standard deviation can also be problematic. Although the lack of fit test indicates some precision gained in estimating each parameter, it is not substantial (see Table 10.8). Thus, the emphasis for this example may be toward combining various techniques to reduce or remove the negative influence of data variability.

Table 10.8. Comparison of models for the mean and standard deviation.

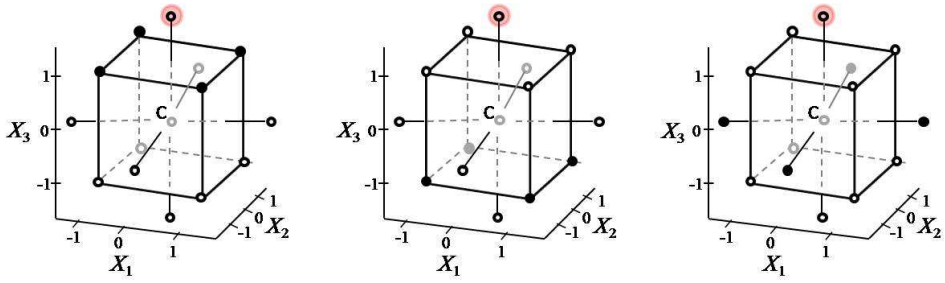
Response	*Model*	*N*	*p*	R_p^2 (%)	$R_{a,p}^2$ (%)	$PRESS_p$	MSE_p	*Lack of Fit Test* Statistic (P-value)
Mean	Prev. Research (2nd ord.)	18	10	94.0	87.3	18962.7	655.0	$F = 0.27$ (0.90)
	Higher-Order (3rd ord.)	18	10	94.3	87.9	21860.0	623.7	$F = 0.23$ (0.93)
Standard Deviation	Prev. Research (2nd ord.)	18	10	93.6	86.3	16426.1	326.4	$F = 1.64$ (0.36)
	Higher-Order (3rd ord.)	18	12	97.7	93.5	5837.0	154.8	$F = 0.33$ (0.81)

More importantly, like the printing press study examined in Section 3.1, the residuals for the semiconductor study parameters also exhibit non-constant variance for the full second-order model when the Breusch-Pagan hypothesis test is performed. In fact, using the data in Table 10.7 with three replications, it is not until we consider a reduced model ($p = 9$) that the assumption of constant variance holds. For the first order model (1) and the second order model (2) where the pure quadratic term X_3^2 is removed, we have:

$$(1)\ \chi^2_{BP} = \frac{9.658\text{e+}08}{2} \div \left(\frac{494666}{54} \right)^2 = 5.75 \qquad (2)\ \chi^2_{BP} = \frac{1.084\text{e+}09}{2} \div \left(\frac{407619}{54} \right)^2 = 9.51,$$

where, with $5.75 < \chi^2(.95, 3) = 7.81$ and $9.51 < \chi^2(.95, 8) = 15.51$, H_0 is concluded for both models; that is, the error variance is constant. To remove the influence of additional variability inherent within the data, an alternative technique is to consider the methodology described in Section 10.4, whereby we perform a screening test for *CV* to identify and remove specific design points for consideration. We then use optimal designs to facilitate minimizing the variability of the coefficients in the regression model.

If we render the design point (0, 0, 1.6818) infeasible, is an irregularly-shaped experimental region results. Using the remaining seventeen candidate design points and the selection of the second order model with $p = 9$ parameters, we examine various experimental designs with $N = 18$, $m = 3$, and $p = 9$ (see Figure 10.14).



(a) (b) (c)

Figure 10.14. (a) Alternative Design 1, (b) Alternative Design 2, and (c) Alternative Design 3 Infeasible Point (Shaded), Single Run Design Point (o), Double Run Design Point (•).

As shown in Table 10.9, the *D*, *A*, and *G* criteria support selecting the same design for the semiconductor manufacturing study.

Table 10.9. Criteria information for Alternative Designs 1, 2, and 3.

Design	*D*-Criterion Maximize $\lvert(\mathbf{X}'\mathbf{X})\rvert$	*A*-Criterion Minimize $tr(\mathbf{X}'\mathbf{X})^{-1}$	*G*-Criterion Maximize Avg $h_{jj}/h_{jj(\max)}$
1	9.537e+09	1.15868	0.80575
2	8.608e+09	1.21221	0.75300
3	6.977e+09	1.23670	0.74791

Using the various optimality criteria, the relative efficiency of designs 2 and 3 are computed and evaluated against design 1 in Table 10.10. While the efficiencies are close in approximation, designs 2 and 3 would have to be replicated more than once in order to achieve a confidence region for $\hat{\boldsymbol{\beta}}$ that is as small as that observed with design 1.

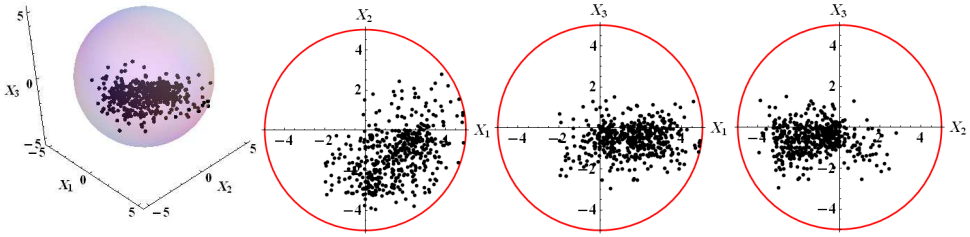
Table 10.10. Relative Efficiency of Designs 2 and 3 to Design 1.

	Relative Efficiency (%) to Design 1		
Design	*D*-Criterion	*A*-Criterion	*G*-Criterion
2	0.989	0.995	0.993
3	0.966	0.993	0.992

Table 10.11 depicts the optimal design – Design 1, along with the new run order and calculations for both the mean and standard deviation. The observations for the new design points are generated using the Monte Carlo simulation technique at the mean and standard deviation previously defined. Simulating the experiment 500 times yields the solution space shown in Figure 10.15. Removing only one design point from the original experiment reduces the variability in the generated solutions considerably (comparison with Figure 10.14). However, the solution space is likely not refined enough to provide consistent results when the experiment is repeated. Hence, it may be only at the cost of adding replications that a final result is deemed acceptable.

Table 10.11. Optimal design for resistivity in semiconductor study.

Run Order	Flow Rate X_1	Coded Units Temperature X_2	Pressure X_3	Mean $\bar{y}$	Standard Deviation s
9	–1.6818	0	0	263.986	107.424
11	0	–1.6818	0	390.029	96.107
10	1.6818	0	0	205.776	66.990
12	0	1.6818	0	292.526	110.058
1	–1	–1	–1	290.104	141.334
2	1	–1	–1	302.320	147.236
3	–1	1	–1	164.656	80.455
4	1	1	–1	160.369	82.632
14	0	0	0	211.039	57.150
5	–1	–1	1	272.076	53.421
8	1	1	1	293.782	68.930
6	1	–1	1	147.133	39.404
7	–1	1	1	418.555	221.964
13	0	0	–1.6818	309.740	157.628
5	–1	–1	1	248.210	163.293
8	1	1	1	168.057	63.470
6	1	–1	1	143.945	57.912
7	–1	1	1	272.013	62.773



(a) (b) (c) (d)

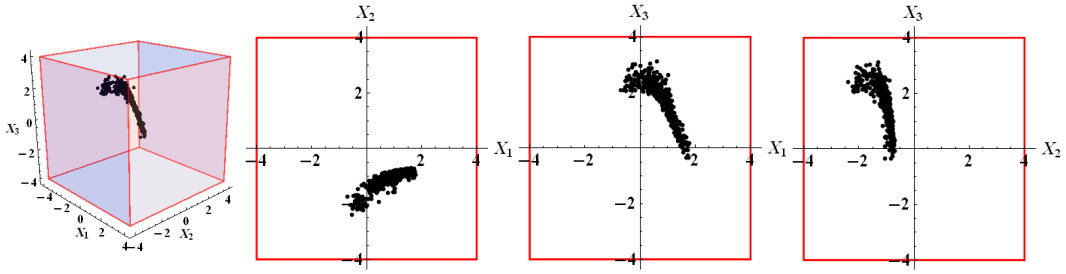
Figure 10.15. Identification of **x***, 3 reps with Alternative Design 1, $\mathbf{x}'\mathbf{x} \leq 25$ (500 iterations).

10.5.3 Summary of Findings

For processes exhibiting elevated degrees of variability, several trade-offs may exist that will allow experimenters to evaluate the benefits of replication or precision against costs. Despite efforts to increase precision in model formulation or reduce the design points for the sake of adding replications, the degree of process variability may

still necessitate additional measures. In Section 10.5.1.3, a weighted least squares approach was used to allocate less weight to those observations acting as outliers. In Section 10.5.2, a new technique was applied using the coefficient of variation as a screening mechanism to completely remove the influence of outlier design points. The research findings for these examples are discussed in (i)-(iii) below:

(i) Suppose the cost and time to perform the printing press experimental study in Section 10.5.1 are not an issue, and so we can maximize the number of replications using the full set of design points. Figure 10.16 portrays the solution space for such a case, wherein we perform 20 replications at each design point. A weighted least squares approach is used to remove the influence of heteroscedasticity and formulate a higher order model for both the mean and standard deviation. Although this output is not practical in terms of the cost to conduct the experiment, it represents the "best case" solution space in terms of precision for the given data set.



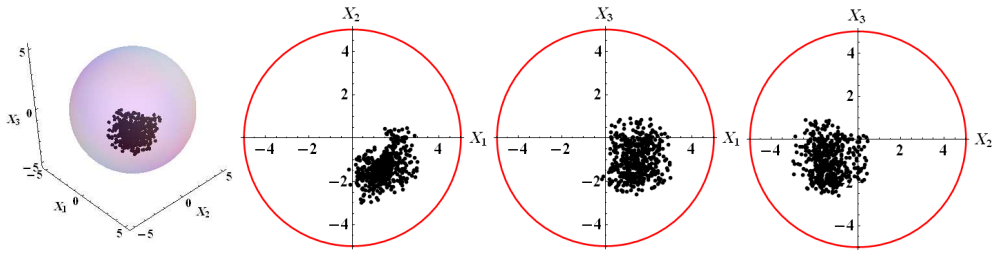
(a) (b) (c) (d)

Figure 10.16. Identification of **x***, 20 reps with original design (*WLS*), $\mathbf{x'x} \leq 25$ (500 iterations).

Comparing the results of Section 10.5.1 with Figure 10.16, it is clear that any one of the alternative designs or techniques used (results in Figures 10.7- 10.11) is much closer to the "best case" scenario than the original attempt in Section 10.3. It is worth noting that,

of the results in Section 10.5.1, only those in Figure 10.8 involve a combination of the approaches described in the trade-off analysis (reduced experimental framework with additional replication and higher order models). This is interesting, as these results (Figure 10.8) most closely match those in the "best case" scenario above despite a slight alteration in the solution space configuration. For this reason, a combined emphasis on replication and precision in modeling may provide the best estimation.

The same procedure using Monte Carlo simulation can be used to compare the generation of solutions for the semiconductor manufacturing study in Section 10.5.2. Suppose we perform 20 replications at each design point, use the second order response surface design, and then implement the *CV* technique to remove the effects of non-constant variance. Figure 10.17 shows the "best case" for this example.



(a) (b) (c) (d)

Figure 10.17. Identification of **x***, 20 reps with original design (*CV*), $\mathbf{x}'\mathbf{x} \leq 25$ (500 iterations).

Again, at no additional cost to the experimenter, the model framework used to generate the solution space in Figure 10.15 is much closer to the "best case" solution space in Figure 10.17.

(ii) Section 10.3 included a comparison of solutions generated using four optimization schemes from the literature. When the experiment was repeated many

times, the optimal solution varied widely among the different models. If the variability of the data at each design point is stable a comparison of *RPD* optimization models can be effective. But for processes with elevated degrees of variability, several remedial measures may be necessary before a valid comparison can be made. Monte Carlo simulation is one technique that can be used to identify a potentially problematic set of experimental data.

(iii) When the intent is to utilize response surface methods and the assumption of constant variance is violated, weighted least squares regression is typically used as a remedial procedure. This establishes a weight matrix that places greater weight on smaller residuals in establishing a fitted function. This effectively discounts observations that result in higher residuals, thus minimizing their influence on the formulation of the regression model. An alternative proposed technique that may relieve heteroscedasticity without weighting each of the design points is the *CV* technique outlined in Chapter 8 and applied once again in Section 10.5.2. At no additional cost, a minimal number of variability-inducing design points are identified for exclusion and optimal designs are then used to select the appropriate experimental framework. Any one of the primary optimal designs - *D*, *A*, and *G*, may be used in the final selection of the model that supports minimizing the variability of the regression coefficients.

10.6 Concluding Remarks

This chapter has examined *RPD* in highly variable processes, pointing out that, barring any consideration of the trade-offs involved, traditional approaches can lead to imprecise estimates that likely will yield imprecise and potentially suboptimal solutions.

To overcome this, the integration of a trade-off analysis into *RPD* approaches is proposed that can be implemented either singularly or multilaterally to obtain more precise estimators. These, in turn, will facilitate determining the true optimal operating conditions. Ultimately, the goal of most contemporary development efforts is to achieve high levels of precision at the lowest cost possible. As manufacturers and decision makers seek improved methods for ensuring quality in resource-constrained environments, engineers must examine and present the trade-offs between precision and the associated costs in order to best support decision making. While the analysis in this chapter contributes both to a broader and deeper understanding of the implications of various decisions in experimental design and optimization, the approaches discussed herein could very well be extended to consider other types of response surface estimators such as those discussed in Chapter 6, as well as other response surface design methods, such as goal programming, desirability function, or compromise programming.

CHAPTER ELEVEN

CONCLUSION AND FUTURE STUDY

The motivating factors for the research presented in this dissertation stemmed from several of the more recent reviews of the *RPD* literature, including Robinson *et al.* (2004), Murphy *et al.* (2005), and Hasenkamp *et al.* (2009). A thematic shortcoming that permeates these reviews is that *RPD* research has largely focused on the development of new statistical tools and techniques, which has created gaps in knowledge and understanding and led to stagnation in practical application in industry. In short, while this tool-based focus has spurred much significant advancement regarding the *RPD* framework, it remains unclear to many engineers and practitioners as to how and when these various tools should be applied. Complicating this, nearly all of the available methods rely on a variety of assumptions about underlying process characteristics and conditions that often deviate from the realities observed in industrial settings. Most notably, many methods fail to consider elevated degrees of variation and/or asymmetry present in a particular process, deferring instead to assumptions of moderate system variability and symmetry to facilitate the application of traditional statistical methods. This "disconnect" can lead to several problems. First, it further clouds understanding among engineers as to which tools to use and when. Second, if such realities are ignored, the derived estimates can be problematic and misleading and, once applied to optimization schemes, will likely result suboptimal solutions and dubious recommendations to decision makers.

Hence, as noted in Chapter 1, the overarching objective of this research was to create quality models and refinements to existing approaches that achieve the best possible *RPD* solutions under realistic process conditions. The deeper aims of this objective, however, were twofold. The first was to facilitate the attainment of enhanced degrees of precision and accuracy than would otherwise be achieved by assuming away true process conditions. The second was to provide engineers and decision makers greater awareness, improved guidelines, and flexibility in applying the best and most appropriate statistical tools and techniques to determine optimal process settings. The research and results in Chapters 4-10 have served to achieve these aims. In Chapter 4, a comprehensive analysis of the ways in which the normal probability plot can be used to support data and process analysis and understanding was presented. Following an investigation of the impacts of variability measure selection under various degrees of process variation in Chapter 5, the examination of inherent process conditions was expanded in Chapter 6 to deduce the impacts of both high-variability and asymmetry on the selection of tier-one estimators. The work in Chapter 7 extended the conditions-based analysis to tier-two estimation in order to determine which estimation approaches performed best under an array of realistic process conditions. Chapter 8examined the effects of dynamic variability in the context of multiple mixed-type quality characteristics, proposing a methodology that will allow for accurately predicting the location of the optimal process mean vector as it shifts in response to dynamic process variability. In Chapter 9, a new technique was suggested as a means of removing sources of process, or system variability in order to achieve greater precision in the

optimal process settings. Chapter 10 extended this even further by proposing the integration of a trade-off analysis that examines a variety of ways to achieve enhancements in precision, accuracy, and predictability against the associated costs as a means of providing engineers, manufacturers, and decision makers greater flexibility in the decision-making process.

Ultimately, this research has focused on providing improved methodologies that will assist engineers in applying existing tools and techniques more effectively and in ways that more fully and accurately account for the realities observed in practice. Thus, whether attempting to reduce process and/or product variation or searching for more effective approaches to minimize processing costs, the research methodologies offered through this dissertation may provide considerable benefit for engineers who apply them, manufacturers and decision makers who guide manufacturing priorities, and customers who have the ultimate stake in the end product.

While the work presented in this dissertation should serve to bridge several gaps that currently exist in the *RPD* field, there are a number of areas that may (and should) be further explored. Most notably, the preponderance of research heretofore discussed has focused on the univariate case - that is, on situations involving only one quality characteristic of interest. Chapter 8 is the only exception. However, most products are judged according to the simultaneous performance of multiple quality characteristics. Moreover, it is often the case that the various objectives for each of characteristics is at odds with the others, which adds to the complexity. Such mixed-type multi-response

situations require attention, particularly in the context of high-variability and asymmetric processes.

Studies examining the multi-response problem could examine the use of the multi-variate skew normal distribution to simultaneously capture/model disparate degrees of both variability and asymmetry associated with the various quality characteristics. In short, the inclusion of a skewness vector would facilitate an even more realistic portrayal of the process response surface. Likewise, the inclusion of a covariance matrix would enable engineers to more readily capture interactions between competing quality characteristics.

The discussion regarding trade-offs could also be further examined in the context of manufacturing priorities using certain optimization approaches, such as goal programming. Such efforts could only help to bridge the gap between academia and industry even further and provide engineers and manufacturers with even better guidelines, more flexible and realistic approaches to solving *RPD* problems, and ultimately greater confidence in the results and recommendations presented.

APPENDICES

A: Supporting Minitab Output for Chapter 5

A.1 First-Order Regression Analysis: m, s, s2, ln(s) versus x1, x2

Response Surface Regression: m versus x1, x2

```
The analysis was done using coded units.

Estimated Regression Coefficients for m

Term         Coef  SE Coef       T      P
Constant  78.3259    1.744  44.912  0.000
x1         0.1259    1.950   0.065  0.950
x2         0.9279    1.950   0.476  0.649
x1*x1     -3.9994    2.957  -1.353  0.218
x2*x2      2.9206    2.957   0.988  0.356
x1*x2     -4.0638    3.899  -1.042  0.332

S = 3.89969    PRESS = 425.983
R-Sq = 39.26%  R-Sq(pred) = 0.00%  R-Sq(adj) = 0.00%


Analysis of Variance for m

Source          DF   Seq SS   Adj SS  Adj MS     F      P
Regression       5   68.814   68.814  13.763  0.90  0.527
  Linear         2    3.508    3.508   1.754  0.12  0.893
  Square         2   48.782   48.782  24.391  1.60  0.267
  Interaction    1   16.524   16.524  16.524  1.09  0.332
Residual Error   7  106.453  106.453  15.208
  Lack-of-Fit    3   46.796   46.796  15.599  1.05  0.464
  Pure Error     4   59.658   59.658  14.914
Total           12  175.267

Estimated Regression Coefficients for m using data in uncoded units

Term           Coef
Constant    78.3259
x1        0.0890259
x2         0.656212
x1*x1      -2.00033
x2*x2       1.46072
x1*x2      -2.03250
```

Response Surface Regression: s versus x1, x2

```
The analysis was done using coded units.

Estimated Regression Coefficients for s

Term          Coef  SE Coef       T      P
Constant   0.77399   0.1922   4.027  0.005
x1        -0.05162   0.2149  -0.240  0.817
x2        -0.43791   0.2149  -2.038  0.081
x1*x1      0.70091   0.3259   2.151  0.069
x2*x2      0.73091   0.3259   2.243  0.060
x1*x2      0.75977   0.4296   1.768  0.120
```

```
S = 0.429769   PRESS = 5.06295
R-Sq = 69.41%  R-Sq(pred) = 0.00%  R-Sq(adj) = 47.56%


Analysis of Variance for s

Source          DF  Seq SS  Adj SS  Adj MS     F      P
Regression       5  2.9338  2.9338  0.5868  3.18  0.082
  Linear         2  0.7778  0.7778  0.3889  2.11  0.192
  Square         2  1.5784  1.5784  0.7892  4.27  0.061
  Interaction    1  0.5776  0.5776  0.5776  3.13  0.120
Residual Error   7  1.2929  1.2929  0.1847
  Lack-of-Fit    3  0.5484  0.5484  0.1828  0.98  0.485
  Pure Error     4  0.7445  0.7445  0.1861
Total           12  4.2267


Estimated Regression Coefficients for s using data in uncoded units

Term            Coef
Constant    0.773988
x1         -0.0365030
x2          -0.309694
x1*x1        0.350562
x2*x2        0.365566
x1*x2        0.380000
```

Response Surface Regression: s2 versus x1, x2

```
The analysis was done using coded units.

Estimated Regression Coefficients for s2

Term         Coef  SE Coef       T      P
Constant   0.7460   0.4364   1.709  0.131
x1        -0.2565   0.4878  -0.526  0.615
x2        -1.4128   0.4878  -2.896  0.023
x1*x1      1.4550   0.7398   1.967  0.090
x2*x2      1.9650   0.7398   2.656  0.033
x1*x2      2.4643   0.9755   2.526  0.039

S = 0.975763   PRESS = 35.4939
R-Sq = 77.97%  R-Sq(pred) = 0.00%  R-Sq(adj) = 62.23%


Analysis of Variance for s2

Source          DF  Seq SS  Adj SS  Adj MS     F      P
Regression       5  23.585  23.585  4.7170  4.95  0.029
  Linear         2   8.248   8.248  4.1240  4.33  0.060
  Square         2   9.260   9.260  4.6302  4.86  0.047
  Interaction    1   6.076   6.076  6.0762  6.38  0.039
Residual Error   7   6.665   6.665  0.9521
  Lack-of-Fit    3   4.520   4.520  1.5067  2.81  0.172
  Pure Error     4   2.145   2.145  0.5362
Total           12  30.250
```

Estimated Regression Coefficients for s2 using data in uncoded units

Term	Coef
Constant	0.745970
x1	-0.181375
x2	-0.999133
x1*x1	0.727740
x2*x2	0.982817
x1*x2	1.23250

Response Surface Regression: lns versus x1, x2

The analysis was done using coded units.

Estimated Regression Coefficients for lns

Term	Coef	SE Coef	T	P
Constant	-0.39401	0.2274	-1.733	0.127
x1	-0.01870	0.2542	-0.074	0.943
x2	-0.29562	0.2542	-1.163	0.283
x1*x1	0.78016	0.3855	2.024	0.083
x2*x2	0.69016	0.3855	1.790	0.117
x1*x2	0.48485	0.5083	0.954	0.372

S = 0.508441 PRESS = 4.57342
R-Sq = 55.51% R-Sq(pred) = 0.00% R-Sq(adj) = 23.73%

Analysis of Variance for lns

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	5	2.2578	2.2578	0.4516	1.75	0.242
Linear	2	0.3510	0.3510	0.1755	0.68	0.538
Square	2	1.6715	1.6715	0.8358	3.23	0.101
Interaction	1	0.2352	0.2352	0.2352	0.91	0.372
Residual Error	7	1.8096	1.8096	0.2585		
Lack-of-Fit	3	0.3147	0.3147	0.1049	0.28	0.838
Pure Error	4	1.4949	1.4949	0.3737		
Total	12	4.0674				

Estimated Regression Coefficients for lns using data in uncoded units

Term	Coef
Constant	-0.394011
x1	-0.0132270
x2	-0.209067
x1*x1	0.390196
x2*x2	0.345183
x1*x2	0.242500

A.2 Second-Order Polynomial Analysis

Regression Analysis: m versus x1, x2, x1x2, x11, x22

The regression equation is

```
m = 78.3 + 0.09 x1 + 0.66 x2 - 2.03 x1x2 - 2.00 x11 + 1.46 x22


Predictor    Coef  SE Coef      T      P    VIF
Constant   78.326    1.744  44.91  0.000
x1          0.089    1.379   0.06  0.950  1.000
x2          0.656    1.379   0.48  0.649  1.000
x1x2       -2.032    1.950  -1.04  0.332  1.000
x11        -2.000    1.479  -1.35  0.218  1.017
x22         1.461    1.479   0.99  0.356  1.017


S = 3.89969   R-Sq = 39.3%   R-Sq(adj) = 0.0%

PRESS = 425.983   R-Sq(pred) = 0.00%


Analysis of Variance

Source          DF      SS     MS     F      P
Regression       5   68.81  13.76  0.90  0.527
Residual Error   7  106.45  15.21
Total           12  175.27


Source  DF  Seq SS
x1       1    0.06
x2       1    3.44
x1x2     1   16.52
x11      1   33.95
x22      1   14.84
```

Best Subsets Regression: m versus x1, x2, x1x2, x11, x22

```
Response is m
                                          x
                                          1 x x
                      Mallows           x x x 1 2
Vars  R-Sq  R-Sq(adj)       Cp       S  1 2 2 1 2
   1  19.4       12.0      0.3  3.5843        X
   2  28.8       14.6      1.2  3.5327      X X
   3  37.3       16.3      2.2  3.4954      X X X
   4  39.2        8.8      4.0  3.6489    X X X X
   5  39.3        0.0      6.0  3.8997  X X X X X
```

Regression Analysis: s versus x1, x2, x1x2, x11, x22

```
The regression equation is
s = 0.774 - 0.037 x1 - 0.310 x2 + 0.380 x1x2 + 0.351 x11 + 0.366 x22

Predictor     Coef  SE Coef      T      P    VIF
Constant    0.7740   0.1922   4.03  0.005
x1         -0.0365   0.1520  -0.24  0.817  1.000
x2         -0.3097   0.1520  -2.04  0.081  1.000
x1x2        0.3800   0.2149   1.77  0.120  1.000
x11         0.3506   0.1630   2.15  0.069  1.017
x22         0.3656   0.1630   2.24  0.060  1.017
```

```
S = 0.429769   R-Sq = 69.4%   R-Sq(adj) = 47.6%

PRESS = 5.06295   R-Sq(pred) = 0.00%


Analysis of Variance

Source          DF      SS      MS     F      P
Regression       5  2.9338  0.5868  3.18  0.082
Residual Error   7  1.2929  0.1847
Total           12  4.2267


Source  DF  Seq SS
x1       1  0.0107
x2       1  0.7672
x1x2     1  0.5776
x11      1  0.6491
x22      1  0.9293
```

Best Subsets Regression: s versus x1, x2, x1x2, x11, x22

```
Response is s
                                                 x
                                                 1 x x
                         Mallows               x x x 1 2
Vars  R-Sq  R-Sq(adj)       Cp        S        1 2 2 1 2
   1  18.2       10.7      9.7  0.56081          X
   2  37.3       24.8      7.3  0.51462              X X
   3  55.5       40.7      5.2  0.45719          X   X X
   4  69.2       53.7      4.1  0.40367          X X X X
   5  69.4       47.6      6.0  0.42977        X X X X X
```

Regression Analysis: s2 versus x1, x2, x1x2, x11, x22

```
The regression equation is
s2 = 0.746 - 0.181 x1 - 0.999 x2 + 1.23 x1x2 + 0.728 x11 + 0.983 x22


Predictor     Coef  SE Coef      T      P    VIF
Constant    0.7460   0.4364   1.71  0.131
x1         -0.1814   0.3450  -0.53  0.615  1.000
x2         -0.9991   0.3450  -2.90  0.023  1.000
x1x2        1.2325   0.4879   2.53  0.039  1.000
x11         0.7277   0.3700   1.97  0.090  1.017
x22         0.9828   0.3700   2.66  0.033  1.017


S = 0.975763   R-Sq = 78.0%   R-Sq(adj) = 62.2%

PRESS = 35.4939   R-Sq(pred) = 0.00%


Analysis of Variance
```

```
Source          DF       SS      MS     F      P
Regression       5  23.5848  4.7170  4.95  0.029
Residual Error   7   6.6648  0.9521
Total           12  30.2496


Source  DF  Seq SS
x1       1  0.2631
x2       1  7.9849
x1x2     1  6.0762
x11      1  2.5439
x22      1  6.7166
```

Best Subsets Regression: s2 versus x1, x2, x1x2, x11, x22

```
Response is s2
                                                x
                                                1 x x
                          Mallows             x x x 1 2
Vars  R-Sq  R-Sq(adj)       Cp         S      1 2 2 1 2
   1  26.4       19.7     14.4   1.4227        X
   2  46.5       35.8     10.0   1.2723        X X
   3  64.9       53.2      6.1   1.0858        X X   X
   4  77.1       65.6      4.3  0.93059        X X X X
   5  78.0       62.2      6.0  0.97576      X X X X X
```

Regression Analysis: lns versus x1, x2, x1x2, x11, x22

```
The regression equation is
lns = - 0.394 - 0.013 x1 - 0.209 x2 + 0.243 x1x2 + 0.390 x11 + 0.345 x22


Predictor     Coef  SE Coef      T      P    VIF
Constant   -0.3940   0.2274  -1.73  0.127
x1         -0.0132   0.1798  -0.07  0.943  1.000
x2         -0.2091   0.1798  -1.16  0.283  1.000
x1x2        0.2425   0.2542   0.95  0.372  1.000
x11         0.3902   0.1928   2.02  0.083  1.017
x22         0.3452   0.1928   1.79  0.117  1.017


S = 0.508441   R-Sq = 55.5%   R-Sq(adj) = 23.7%

PRESS = 4.57342   R-Sq(pred) = 0.00%


Analysis of Variance

Source          DF      SS      MS     F      P
Regression       5  2.2578  0.4516  1.75  0.242
Residual Error   7  1.8096  0.2585
Total           12  4.0674

Source  DF  Seq SS
x1       1  0.0014
x2       1  0.3496
x1x2     1  0.2352
```

```
x11      1  0.8430
x22      1  0.8285
```

A.3 Higher-Order Polynomial Analysis

Regression Analysis: m versus x1, x2, ...

```
The regression equation is
m = 78.3 + 1.82 x1 + 2.25 x2 - 2.03 x1x2 - 2.29 x11 + 1.17 x22 - 3.18 x11x2
    - 3.46 x1x22 + 1.14 x11x22


Predictor    Coef  SE Coef      T      P    VIF
Constant   78.326    1.727  45.35  0.000
x1          1.821    1.931   0.94  0.399  2.000
x2          2.245    1.931   1.16  0.310  2.000
x1x2       -2.032    1.931  -1.05  0.352  1.000
x11        -2.286    1.616  -1.41  0.230  1.239
x22         1.175    1.616   0.73  0.507  1.239
x11x2      -3.178    2.731  -1.16  0.309  2.000
x1x22      -3.464    2.731  -1.27  0.273  2.000
x11x22      1.143    2.731   0.42  0.697  1.385


S = 3.86191   R-Sq = 66.0%   R-Sq(adj) = 0.0%

PRESS = *   R-Sq(pred) = *%


Analysis of Variance

Source          DF      SS     MS     F      P
Regression       8  115.61  14.45  0.97  0.553
Residual Error   4   59.66  14.91
Total           12  175.27


Source  DF  Seq SS
x1       1    0.06
x2       1    3.44
x1x2     1   16.52
x11      1   33.95
x22      1   14.84
x11x2    1   20.20
x1x22    1   23.99
x11x22   1    2.61


Unusual Observations

                                                 St
Obs     x1      m    Fit  SE Fit  Residual  Resid
  1  -1.00  78.90  78.90    3.86      0.00      * X
  2   1.00  79.68  79.68    3.86      0.00      * X
  3  -1.00  81.10  81.10    3.86     -0.00      * X
  4   1.00  73.75  73.75    3.86     -0.00      * X
  5  -1.41  71.18  71.18    3.86      0.00      * X
  6   1.41  76.33  76.33    3.86     -0.00      * X
```

```
  7   0.00  77.50  77.50    3.86      0.00      * X
  8   0.00  83.85  83.85    3.86      0.00      * X

X denotes an observation whose X value gives it large leverage.
```

Regression Analysis: s versus x1, x2, ...

```
The regression equation is
s = 0.774 + 0.152 x1 - 0.484 x2 + 0.380 x1x2 + 0.326 x11 + 0.341 x22
    + 0.349 x11x2 - 0.377 x1x22 + 0.100 x11x22


Predictor     Coef  SE Coef      T      P    VIF
Constant    0.7740   0.1929   4.01  0.016
x1          0.1521   0.2157   0.70  0.520  2.000
x2         -0.4844   0.2157  -2.25  0.088  2.000
x1x2        0.3800   0.2157   1.76  0.153  1.000
x11         0.3256   0.1805   1.80  0.146  1.239
x22         0.3406   0.1805   1.89  0.132  1.239
x11x2       0.3494   0.3051   1.15  0.316  2.000
x1x22      -0.3771   0.3051  -1.24  0.284  2.000
x11x22      0.0998   0.3051   0.33  0.760  1.385

S = 0.431428   R-Sq = 82.4%   R-Sq(adj) = 47.2%

PRESS = *   R-Sq(pred) = *%

Analysis of Variance

Source          DF      SS      MS     F      P
Regression       8  3.4822  0.4353  2.34  0.215
Residual Error   4  0.7445  0.1861
Total           12  4.2267

Source  DF  Seq SS
x1       1  0.0107
x2       1  0.7672
x1x2     1  0.5776
x11      1  0.6491
x22      1  0.9293
x11x2    1  0.2442
x1x22    1  0.2843
x11x22   1  0.0199

Unusual Observations
                                              St
Obs     x1      s    Fit  SE Fit  Residual  Resid
  1  -1.00  2.280  2.280   0.431     0.000      * X
  2   1.00  1.070  1.070   0.431     0.000      * X
  3  -1.00  1.250  1.250   0.431     0.000      * X
  4   1.00  1.560  1.560   0.431     0.000      * X
  5  -1.41  1.210  1.210   0.431     0.000      * X
  6   1.41  1.640  1.640   0.431    -0.000      * X
  7   0.00  2.140  2.140   0.431     0.000      * X
  8   0.00  0.770  0.770   0.431     0.000      * X

X denotes an observation whose X value gives it large leverage.
```

Regression Analysis: s2 versus x1, x2, ...

```
The regression equation is
s2 = 0.746 + 0.435 x1 - 1.41 x2 + 1.23 x1x2 + 0.665 x11 + 0.920 x22
     + 0.823 x11x2 - 1.23 x1x22 + 0.252 x11x22

Predictor     Coef  SE Coef      T      P    VIF
Constant    0.7460   0.3275   2.28  0.085
x1          0.4349   0.3662   1.19  0.301  2.000
x2         -1.4109   0.3662  -3.85  0.018  2.000
x1x2        1.2325   0.3661   3.37  0.028  1.000
x11         0.6647   0.3064   2.17  0.096  1.239
x22         0.9198   0.3064   3.00  0.040  1.239
x11x2       0.8234   0.5178   1.59  0.187  2.000
x1x22      -1.2324   0.5178  -2.38  0.076  2.000
x11x22      0.2520   0.5179   0.49  0.652  1.385

S = 0.732243   R-Sq = 92.9%   R-Sq(adj) = 78.7%

PRESS = *   R-Sq(pred) = *%

Analysis of Variance

Source          DF       SS      MS     F      P
Regression       8  28.1048  3.5131  6.55  0.044
Residual Error   4   2.1447  0.5362
Total           12  30.2496

Source  DF  Seq SS
x1       1  0.2631
x2       1  7.9849
x1x2     1  6.0762
x11      1  2.5439
x22      1  6.7166
x11x2    1  1.3557
x1x22    1  3.0373
x11x22   1  0.1270

Unusual Observations
                                                 St
Obs     x1     s2    Fit  SE Fit  Residual  Resid
  1  -1.00  5.200  5.200   0.732    -0.000      * X
  2   1.00  1.140  1.140   0.732    -0.000      * X
  3  -1.00  1.560  1.560   0.732    -0.000      * X
  4   1.00  2.430  2.430   0.732    -0.000      * X
  5  -1.41  1.460  1.460   0.732    -0.000      * X
  6   1.41  2.690  2.690   0.732    -0.000      * X
  7   0.00  4.580  4.580   0.732     0.000      * X
  8   0.00  0.590  0.590   0.732     0.000      * X

X denotes an observation whose X value gives it large leverage.
```

Regression Analysis: lns versus x1, x2, ...

```
The regression equation is
lns = - 0.394 + 0.106 x1 - 0.361 x2 + 0.243 x1x2 + 0.367 x11 + 0.322 x22
      + 0.303 x11x2 - 0.239 x1x22 + 0.092 x11x22
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	-0.3940	0.2734	-1.44	0.223	
x1	0.1061	0.3057	0.35	0.746	2.000
x2	-0.3607	0.3057	-1.18	0.303	2.000
x1x2	0.2425	0.3057	0.79	0.472	1.000
x11	0.3671	0.2558	1.44	0.225	1.239
x22	0.3221	0.2558	1.26	0.276	1.239
x11x2	0.3032	0.4323	0.70	0.522	2.000
x1x22	-0.2386	0.4323	-0.55	0.610	2.000
x11x22	0.0923	0.4323	0.21	0.841	1.385

S = 0.611335 R-Sq = 63.2% R-Sq(adj) = 0.0%

PRESS = * R-Sq(pred) = *%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	8	2.5724	0.3216	0.86	0.605
Residual Error	4	1.4949	0.3737		
Total	12	4.0674			

Source	DF	Seq SS
x1	1	0.0014
x2	1	0.3496
x1x2	1	0.2352
x11	1	0.8430
x22	1	0.8285
x11x2	1	0.1838
x1x22	1	0.1138
x11x22	1	0.0170

Unusual Observations

Obs	x1	lns	Fit	SE Fit	Residual	St Resid	
1	-1.00	0.820	0.820	0.611	-0.000	*	X
2	1.00	0.070	0.070	0.611	-0.000	*	X
3	-1.00	0.220	0.220	0.611	0.000	*	X
4	1.00	0.440	0.440	0.611	0.000	*	X
5	-1.41	0.190	0.190	0.611	-0.000	*	X
6	1.41	0.490	0.490	0.611	-0.000	*	X
7	0.00	0.760	0.760	0.611	0.000	*	X
8	0.00	-0.260	-0.260	0.611	0.000	*	X

X denotes an observation whose X value gives it large leverage.

Best Subsets Regression: s versus x1, x2, ...

```
Response is s
                                                          x
                                                      x x 1
                                                  x     1 1 1
                                                  1 x x 1 x x
                      Mallows             x x x 1 2 x 2 2
Vars  R-Sq  R-Sq(adj)       Cp        S   1 2 2 1 2 2 2 2
   1  18.2       10.7      9.6  0.56081     X
   2  37.3       24.8      7.2  0.51462       X X
   3  55.5       40.7      5.1  0.45719     X X X
   4  69.2       53.7      4.0  0.40367     X X X X
   5  74.9       57.0      4.7  0.38903     X X X X X
   6  79.7       59.5      5.6  0.37791     X X X X X X
   7  81.9       56.6      7.1  0.39101   X X X X X X X
   8  82.4       47.2      9.0  0.43143   X X X X X X X X
```

Best Subsets Regression: s2 versus x1, x2, ...

```
Response is s2
                                                          x
                                                      x x 1
                                                  x     1 1 1
                                                  1 x x 1 x x
                      Mallows             x x x 1 2 x 2 2
Vars  R-Sq  R-Sq(adj)       Cp        S   1 2 2 1 2 2 2 2
   1  26.4       19.7     32.5   1.4227     X
   2  46.5       35.8     23.2   1.2723     X X
   3  64.9       53.2     14.8   1.0858     X X   X
   4  77.1       65.6      9.9  0.93059     X X X X
   5  85.5       75.2      7.2  0.79137     X X X X   X
   6  90.0       80.0      6.6  0.71042     X X X X X X
   7  92.5       82.0      7.2  0.67405   X X X X X X X
   8  92.9       78.7      9.0  0.73224   X X X X X X X X
```

Best Subsets Regression: lns versus x1, x2, ...

```
Response is lns
                                                          x
                                                      x x 1
                                                  x     1 1 1
                                                  1 x x 1 x x
                      Mallows             x x x 1 2 x 2 2
Vars  R-Sq  R-Sq(adj)       Cp        S   1 2 2 1 2 2 2 2
   1  20.7       13.5     -0.4  0.54141       X
   2  41.1       29.3     -0.6  0.48947       X X
   3  49.7       32.9      0.5  0.47682     X   X X
   4  55.5       33.2      1.8  0.47579     X X X X
   5  60.0       31.4      3.4  0.48213     X X X X X
   6  61.7       23.4      5.2  0.50940     X X X X X X
   7  62.8       10.8      7.0  0.54990   X X X X X X X
   8  63.2        0.0      9.0  0.61133   X X X X X X X X
```

B: Supporting R Code for Chapter 6

B.1 Section 6.3.1: Simulation comparing normal to skew normal

```
# Script for Tier-1 Estimation: simulation comparing normal to skewnormal using simulated data
# Simulation to investigate differences between estimators used
# Conditions: skewed data with some outliers
# data points generated (from a skew normal distribution)

require(MASS)
require(robustbase)
require(ICSNP)
require(sn)
require(moments)

# generate a dataset of observations.  This will be based on a CCD with three
# factors (x1,x2,x3)in three levels each (-1,1), six axial points (alpha =
# 1.682), and four center points.  The manufactured data set will contain
# skewed and contaminated data (outliers)

# Call in the semiconductor dataset and use the 11 observations at each
# design point to obtain MLEs for the location, scale, and skew. Note in the
# final line of the "for" loop that I induce additional skewness to ensure a
# skewed data set for the simulation.

options(digits=5)

xmat<-read.table("xmat_ccd.txt", col.names=c("x1","x2","x3"))
simdata<-read.table("SemiConductor.txt",header=T)
data <- as.matrix(simdata[,4:14])
data2 <- matrix(NA,ncol=3,nrow=length(data[,1]),dimnames=list
            (c(1:length(data[,1])),c("Mean","Std Dev","Skew")))

for (i in 1:length(data[,1])){
   data2[i,] = c(mean(data[i,]),     #sn.mle(y=data[i,],plot.it=F)$cp[1],
                 sd(data[i,]),       #sn.mle(y=data[i,],plot.it=F)$cp[2],
                 skewness(data[i,])) #sn.mle(y=data[i,],plot.it=F)$cp[3])
}

delta <- function(x1){
    (sqrt((pi/2)*(abs(sskew)^(2/3))/(abs(sskew)^(2/3)+((4-pi)/2)^(2/3))))}

parms <- matrix(NA,nrow=length(data2[,1]),ncol=4,dimnames=list
            (c(1:length(data2[,1])),c("d","location","scale","shape")))

for (i in 1:length(data2[,1])){
      sskew = min(data2[i,3],0.9952717)  # associate with sample skew column
      d = delta(sskew)
      w = sqrt((data2[i,2]^2)/(1-(2*d^2)/pi))
      l = data2[i,1]-w*d*sqrt(2/pi)
      a = d/sqrt(1-d^2)
      if (a<9406) a=5*a
      parms[i,1] = d
      parms[i,2] = l
      parms[i,3] = w
      parms[i,4] = a
}
```

```
parms

#Simulation parameters
despts  = 18        # number of design points
obs     = 100       # number of replicates at each design point
num_est = 11        # number of alternative estimators being examined
reps    = 500       # number of simulation iterations
num_mods= 7               # number of models (combinations of estimators)

# Initialize Arrays
normdata <- matrix(NA, nrow=despts,ncol=obs)
skewdata <- matrix(NA, nrow=despts, ncol=obs)

RegModN <- matrix(NA,nrow=10,ncol=num_mods,dimnames=list(c("Bo","B1","B2",
              "B3","B11","“B22","B33","B12","B13","B23"),c("ModA","ModB",
              "ModC","ModD","ModE","ModF","ModG")))
RegModS <- matrix(NA,nrow=10,ncol=num_mods,dimnames=list(c("Bo","B1","B2",
              "B3","B11","B22","B33","B12","B13","B23"),c("ModA","ModB",
              "ModC","ModD","ModE","ModF","ModG")))
Mod_MSE_N  <- matrix(NA,ncol=length(RegModN[1,]),nrow=reps,
               dimnames=list(c(1:reps),c("ModA","ModB","ModC","ModD",
               "ModE","ModF","ModG")))
Mod_MSE_S  <- matrix(NA,ncol=length(RegModS[1,]),nrow=reps,
               dimnames=list(c(1:reps),c("ModA","ModB","ModC","ModD",
               "ModE","ModF","ModG")))
MSE_TallyN <- matrix(NA,ncol=length(RegModN[1,]),nrow=reps,
               dimnames=list(c(1:reps),c("ModA","ModB","ModC","ModD",
               "ModE","ModF","ModG")))
MSE_TallyS <- matrix(NA,ncol=length(RegModS[1,]),nrow=reps,
               dimnames=list(c(1:reps),c("ModA","ModB","ModC","ModD",
               "ModE","ModF","ModG")))
Bias_TallyN <-matrix(NA,ncol=length(RegModN[1,]),nrow=reps,
               dimnames=list(c(1:reps),c("ModA","ModB","ModC","ModD",
               "ModE","ModF","ModG")))

Bias_TallyS <-matrix(NA,ncol=length(RegModS[1,]),nrow=reps,
               dimnames=list(c(1:reps),c("ModA","ModB","ModC","ModD",
               "ModE","ModF","ModG")))
```

 # Simulation start

```
for (j in 1:reps){
```

 # Now use the parameter estimates in "data2" and "parms" MLEs for location, scale, and skew
 # to generate data from both normal and skewnormal distributions. This results in two separate
 # matrices of responses, one normal and one skewed.

```
for(i in 1:despts){
    normdata[i,] <- rnorm(obs, data2[i,1],data2[i,2])
    skewdata[i,] <- rsn(obs, parms[i,2], parms[i,3], parms[i,4])
  }

r1 = as.integer(runif(1,1,18))
c1 = as.integer(runif(1,1,obs))
skewdata[r1,c1] = max(skewdata[r1,])+(1*sd(skewdata[r1,]))

r2 = as.integer(runif(1,1,18))
c2 = as.integer(runif(1,1,obs))
skewdata[r2,c2] = max(skewdata[r2,])+(1*sd(skewdata[r2,]))
```

```
r3= as.integer(runif(1,1,18))
c3 = as.integer(runif(1,1,obs))
skewdata[r3,c3] = max(skewdata[r3,])+(1*sd(skewdata[r3,]))

r4 = as.integer(runif(1,1,18))
c4 = as.integer(runif(1,1,obs))
skewdata[r4,c4] = max(skewdata[r4,])+(1*sd(skewdata[r4,]))

r5 = as.integer(runif(1,1,18))
c5 = as.integer(runif(1,1,obs))
skewdata[r5,c5] = max(skewdata[r5,])+(1*sd(skewdata[r5,]))

r6 = as.integer(runif(1,1,18))
c6 = as.integer(runif(1,1,obs))
skewdata[r6,c6] = max(skewdata[r6,])+(1*sd(skewdata[r6,]))

r7 = as.integer(runif(1,1,18))
c7 = as.integer(runif(1,1,100))
skewdata[r7,c7] = max(skewdata[r7,])+(1*sd(skewdata[r7,]))


estimatorNorm <- matrix(NA,ncol=num_est,nrow=despts,dimnames=list
                     (c(1:despts), c("ybar","s","Median","MAD","Huber2M",
                      "Huber2s","H.L","Sn","Qn","Tau.m","Tau.s")))

estimatorSkew <- matrix(NA,ncol=num_est,nrow=despts,dimnames=list
                    (c(1:despts),c("ybar","s","Median","MAD","Huber2M",
                     "Huber2s","H.L","Sn","Qn","Tau.m","Tau.s")))
```

 # obtain estimates using the various alternative estimators

```
for (i in 1:despts){
     estimatorNorm[i,1]= mean(normdata[i,])
     estimatorNorm[i,2]= sd(normdata[i,])
     estimatorNorm[i,3]= median(normdata[i,])
     estimatorNorm[i,4]= MAD(normdata[i,])
     estimatorNorm[i,5]= hubers(normdata[i,])$mu
     estimatorNorm[i,6]= hubers(normdata[i,])$s
     estimatorNorm[i,7]= hl.loc(normdata[i,])
     estimatorNorm[i,8]= Sn(normdata[i,])
     estimatorNorm[i,9]= Qn(normdata[i,])
     estimatorNorm[i,10]= scaleTau2(normdata[i,],mu.too=T)[1]
     estimatorNorm[i,11]= scaleTau2(normdata[i,],mu.too=T)[2]
     estimatorSkew[i,1]= mean(skewdata[i,])
     estimatorSkew[i,2]= sd(skewdata[i,])
     estimatorSkew[i,3]= median(skewdata[i,])
     estimatorSkew[i,4]= MAD(skewdata[i,])
     estimatorSkew[i,5]= hubers(skewdata[i,])$mu
     estimatorSkew[i,6]= hubers(skewdata[i,])$s
     estimatorSkew[i,7]= hl.loc(skewdata[i,])
     estimatorSkew[i,8]= Sn(skewdata[i,])
     estimatorSkew[i,9]= Qn(skewdata[i,])
     estimatorSkew[i,10]= scaleTau2(skewdata[i,],mu.too=T)[1]
     estimatorSkew[i,11]= scaleTau2(skewdata[i,],mu.too=T)[2]
}

NORM <- data.frame(cbind(xmat,estimatorNorm))
SKEW <- data.frame(cbind(xmat,estimatorSkew))
```

```
ybar1 <-NORM$ybar
s1    <-NORM$s
med1  <-NORM$Median
MAD1  <-NORM$MAD
H2m1  <-NORM$Huber2M
H2s1  <-NORM$Huber2s
HL1   <-NORM$H.L
sn1   <-NORM$Sn
qn1   <-NORM$Qn
taum1 <-NORM$Tau.m
taus1 <-NORM$Tau.s
ybar2 <-SKEW$ybar
s2    <-SKEW$s
med2  <-SKEW$Median
MAD2  <-SKEW$MAD
H2m2  <-SKEW$Huber2M
H2s2  <-SKEW$Huber2s
HL2   <-SKEW$H.L
sn2   <-SKEW$Sn
qn2   <-SKEW$Qn
taum2 <-SKEW$Tau.m
taus2 <-SKEW$Tau.s

x1   <- NORM$x1
x2   <- NORM$x2
x3   <- NORM$x3
x11  <- NORM$x1*NORM$x1
x22  <- NORM$x2*NORM$x2
x33  <- NORM$x3*NORM$x3
x1x2 <- NORM$x1*NORM$x2
x1x3 <- NORM$x1*NORM$x3
x2x3 <- NORM$x2*NORM$x3
z1   <- SKEW$x1
z2   <- SKEW$x2
z3   <- SKEW$x3
z11  <- SKEW$x1*SKEW$x1
z22  <- SKEW$x2*SKEW$x2
z33  <- SKEW$x3*SKEW$x3
z1z2 <- SKEW$x1*SKEW$x2
z1z3 <- SKEW$x1*SKEW$x3
z2z3 <- SKEW$x2*SKEW$x3
```

 # perform regressions using normal data

```
regN1  <-lm(ybar1 ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=NORM)
regN2  <-lm(s1    ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=NORM)
regN3  <-lm(med1  ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=NORM)
regN4  <-lm(MAD1  ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=NORM)
regN5  <-lm(H2m1  ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=NORM)
regN6  <-lm(H2s1  ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=NORM)
regN7  <-lm(HL1   ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=NORM)
regN8  <-lm(sn1   ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=NORM)
regN9  <-lm(HL1   ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=NORM)
regN10 <-lm(qn1   ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=NORM)
regN11 <-lm(HL1   ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=NORM)
regN12 <-lm(MAD1  ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=NORM)
regN13 <-lm(taum1 ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=NORM)
regN14 <-lm(taus1 ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=NORM)
```

#perform regressions using skew-normal data

```
regS1  <-lm(ybar2 ~z1+z2+z3+z11+z22+z33+z1z2+z1z3+z2z3,data=SKEW)
regS2  <-lm(s2    ~z1+z2+z3+z11+z22+z33+z1z2+z1z3+z2z3,data=SKEW)
regS3  <-lm(med2  ~z1+z2+z3+z11+z22+z33+z1z2+z1z3+z2z3,data=SKEW)
regS4  <-lm(MAD2  ~z1+z2+z3+z11+z22+z33+z1z2+z1z3+z2z3,data=SKEW)
regS5  <-lm(H2m2  ~z1+z2+z3+z11+z22+z33+z1z2+z1z3+z2z3,data=SKEW)
regS6  <-lm(H2s2  ~z1+z2+z3+z11+z22+z33+z1z2+z1z3+z2z3,data=SKEW)
regS7  <-lm(HL2   ~z1+z2+z3+z11+z22+z33+z1z2+z1z3+z2z3,data=SKEW)
regS8  <-lm(sn2   ~z1+z2+z3+z11+z22+z33+z1z2+z1z3+z2z3,data=SKEW)
regS9  <-lm(HL2   ~z1+z2+z3+z11+z22+z33+z1z2+z1z3+z2z3,data=SKEW)
regS10 <-lm(qn2   ~z1+z2+z3+z11+z22+z33+z1z2+z1z3+z2z3,data=SKEW)
regS11 <-lm(HL2   ~z1+z2+z3+z11+z22+z33+z1z2+z1z3+z2z3,data=SKEW)
regS12 <-lm(MAD2  ~z1+z2+z3+z11+z22+z33+z1z2+z1z3+z2z3,data=SKEW)
regS13 <-lm(taum2 ~z1+z2+z3+z11+z22+z33+z1z2+z1z3+z2z3,data=SKEW)
regS14 <-lm(taus2 ~z1+z2+z3+z11+z22+z33+z1z2+z1z3+z2z3,data=SKEW)

mod.errN <- cbind(anova(regN1)["Residuals","Mean Sq"],
                  anova(regN3)["Residuals","Mean Sq"],
                  anova(regN5)["Residuals","Mean Sq"],
                  anova(regN7)["Residuals","Mean Sq"],
                  anova(regN9)["Residuals","Mean Sq"],
                  anova(regN11)["Residuals","Mean Sq"],
                 anova(regN13)["Residuals","Mean Sq"])

mod.errS <- cbind(anova(regS1)["Residuals","Mean Sq"],
                  anova(regS3)["Residuals","Mean Sq"],
                  anova(regS5)["Residuals","Mean Sq"],
                  anova(regS7)["Residuals","Mean Sq"],
                  anova(regS9)["Residuals","Mean Sq"],
                  anova(regS11)["Residuals","Mean Sq"],
                  anova(regS13)["Residuals","Mean Sq"])

coeffN <- data.frame(cbind(coef(regN1),coef(regN2),coef(regN3),coef(regN4),
                coef(regN5),coef(regN6),coef(regN7),coef(regN8),coef(regN9),
                coef(regN10),coef(regN11),coef(regN12),coef(regN13),
                coef(regN14)))

coeffS <- data.frame(cbind(coef(regS1),coef(regS2),coef(regS3),coef(regS4),
                coef(regS5),coef(regS6),coef(regS7),coef(regS8),coef(regS9),
                coef(regS10),coef(regS11),coef(regS12),coef(regS13),
                coef(regS14)))

RegModN[,1] <- coef(regN1)
RegModN[,2] <- coef(regN3)
RegModN[,3] <- coef(regN5)
RegModN[,4] <- coef(regN7)
RegModN[,5] <- coef(regN9)
RegModN[,6] <- coef(regN11)
RegModN[,7] <- coef(regN13)

RegModS[,1] <- coef(regS1)
RegModS[,2] <- coef(regS3)
RegModS[,3] <- coef(regS5)
RegModS[,4] <- coef(regS7)
RegModS[,5] <- coef(regS9)
RegModS[,6] <- coef(regS11)
RegModS[,7] <- coef(regS13)

SE_ModN <-cbind(SE(regN1),SE(regN3),SE(regN5),SE(regN7),SE(regN9),SE(regN11),
                SE(regN13))
```

```
SE_ModS <-cbind(SE(regS1),SE(regS3),SE(regS5),SE(regS7),SE(regS9),SE(regS11),
               SE(regS13))

OOCN <- matrix(NA, nrow=3, ncol=length(RegModN[1,]), dimnames=list
             (c("x1","x2","x3"),c(1:length(RegModN[1,]))))
optmnN <- matrix(NA,nrow=1,ncol=length(RegModN[1,]), dimnames=list
             (c("m(x*)"),c(1:length(RegModN[1,]))))
biasN <- matrix(NA,nrow=1,ncol=length(RegModN[1,]), dimnames =list
             (c("bias"),c(1:length(RegModN[1,]))))
optsdN <- matrix(NA,nrow=1,ncol=length(RegModN[1,]), dimnames=list
             (c("sd(x*)"),c(1:length(RegModN[1,]))))
objN <- matrix(NA,nrow=1,ncol=length(RegModN[1,]), dimnames=list
             (c("MSE"),c(1:length(RegModN[1,]))))
OOCS <- matrix(NA, nrow=3, ncol=length(RegModS[1,]), dimnames=list
             (c("x1","x2","x3"),c(1:length(RegModS[1,]))))
optmnS <- matrix(NA,nrow=1,ncol=length(RegModS[1,]), dimnames=list
              (c("m(x*)"),c(1:length(RegModS[1,]))))
biasS <- matrix(NA,nrow=1,ncol=length(RegModS[1,]), dimnames =list
              (c("bias"),c(1:length(RegModS[1,]))))
optsdS <- matrix(NA,nrow=1,ncol=length(RegModS[1,]), dimnames=list
              (c("sd(x*)"),c(1:length(RegModS[1,]))))
objS <- matrix(NA,nrow=1,ncol=length(RegModS[1,]), dimnames=list
             (c("MSE"),c(1:length(RegModS[1,]))))
```

 # Determine Optimum Operating Conditions using MSE-based optimization scheme.

```
for (k in 1:7) {

   m1 <-function(x) {
      x1<-x[1]
      x2<-x[2]
      x3<-x[3]
      coeffN[1,(k+k-1)]+coeffN[2,(k+k-1)]*x1+coeffN[3,(k+k-1)]*x2
      +coeffN[4,(k+k-1)]*x3+coeffN[5,(k+k-1)]*x1^2+coeffN[6,(k+k-1)]*x2^2
      +coeffN[7,(k+k-1)]*x3^2+coeffN[8,(k+k-1)]*x1*x2+coeffN[9,(k+k-1)]*x1*x3
      +coeffN[10,(k+k-1)]*x2*x3}

   s1 <-function(x) {
      x1<-x[1]
      x2<-x[2]
      x3<-x[3]
      coeffN[1,(2*k)]+coeffN[2,(2*k)]*x1+coeffN[3,(2*k)]*x2
      +coeffN[4,(2*k)]*x3+coeffN[5,(2*k)]*x1^2+coeffN[6,(2*k)]*x2^2
      +coeffN[7,(2*k)]*x3^2+coeffN[8,(2*k)]*x1*x2+coeffN[9,(2*k)]*x1*x3
      +coeffN[10,(2*k)]*x2*x3}

   objectiveN<-function(x) {
      x1<-x[1]
      x2<-x[2]
      x3<-x[3]
      (s1(c(x1,x2,x3)))^2 + (m1(c(x1,x2,x3))-350)^2 }

   settingsN <- nlminb(start=c(0,0,0), objectiveN, gradient = NULL, hessian =
                       NULL, lower = c(-1.682,-1.682,-1.682), upper =
                       c(1.682,1.682,1.682))

   OOCN[,k] <- settingsN$par
   optmnN[,k] <- m1(settingsN$par)
   biasN[,k] <- abs(m1(settingsN$par)-350)
```

```
   optsdN[,k]   <- s1(settingsN$par)
   objN[,k] <- settingsN$objective

   m2 <-function(z) {
      z1<-z[1]
      z2<-z[2]
      z3<-z[3]
      coeffS[1,(k+k-1)]+coeffS[2,(k+k-1)]*z1+coeffS[3,(k+k-1)]*z2
      +coeffS[4,(k+k-1)]*z3+coeffS[5,(k+k-1)]*z1^2+coeffS[6,(k+k-1)]*z2^2+
      +coeffS[7,(k+k-1)]*z3^2+coeffS[8,(k+k-1)]*z1*z2+coeffS[9,(k+k-1)]*z1*z3
      +coeffS[10,(k+k-1)]*z2*z3}

   s2 <-function(z) {
      z1<-z[1]
      z2<-z[2]
      z3<-z[3]
      coeffS[1,(2*k)]+coeffS[2,(2*k)]*z1+coeffS[3,(2*k)]*z2
      +coeffS[4,(2*k)]*z3 coeffS[5,(2*k)]*z1^2+coeffS[6,(2*k)]*z2^2+
      +coeffS[7,(2*k)]*z3^2+ coeffS[8,(2*k)]*z1*z2+coeffS[9,(2*k)]*z1*z3
      +coeffS[10,(2*k)]*z2*z3}

   objectiveS<-function(z) {
      z1<-z[1]
      z2<-z[2]
      z3<-z[3]
      (s2(c(z1,z2,z3)))^2 + (m2(c(z1,z2,z3))-350)^2 }

 settingsS <- nlminb(start=c(0,0,0), objectiveS, gradient = NULL, hessian =
                       NULL, lower = c(-1.682,-1.682,-1.682), upper =
                       c(1.682,1.682,1.682))

   OOCS[,k] <- settingsS$par
   optmnS[,k] <- m2(settingsS$par)
   biasS[,k] <- abs(m2(settingsS$par)-350)
   optsdS[,k]   <- s2(settingsS$par)
   objS[,k] <- settingsS$objective
}

NORMresults <- as.matrix(rbind(RegModN,OOCN,optmnN,biasN,optsdN,objN))
SKEWresults <- as.matrix(rbind(RegModS,OOCS,optmnS,biasS,optsdS,objS))

Mod_MSE_N[j,]  <- mod.errN
Mod_MSE_S[j,]  <- mod.errS
MSE_TallyN[j,] <- NORMresults["MSE",]
MSE_TallyS[j,] <- SKEWresults["MSE",]
Bias_TallyN[j,]<- NORMresults["bias",]
Bias_TallyS[j,]<- SKEWresults["bias",]

}
Avg.Mod.MSE.N  <- as.vector(colMeans(Mod_MSE_N))
NormModMSERes  <- rbind(Mod_MSE_N,Avg.Mod.MSE.N)
Avg.Mod.MSE.S  <- as.vector(colMeans(Mod_MSE_S))
SkewModMSERes  <- rbind(Mod_MSE_S,Avg.Mod.MSE.S)
Avg_MSE_N <- as.vector(colMeans(MSE_TallyN))
NormMSERes   <- rbind(MSE_TallyN,Avg_MSE_N)
Avg_Bias_N<- as.vector(colMeans(Bias_TallyN))
NormBiasRes  <- rbind(Bias_TallyN,Avg_Bias_N)
Avg_MSE_S <- as.vector(colMeans(MSE_TallyS))
SkewMSERes   <- rbind(MSE_TallyS,Avg_MSE_S)
Avg_Bias_S<- as.vector(colMeans(Bias_TallyS))
```

```
SkewBiasRes   <- rbind(Bias_TallyS,Avg_Bias_S)

Res <- rbind(Avg.Mod.MSE.N,Avg_MSE_N,Avg_Bias_N,Avg.Mod.MSE.S,Avg_MSE_S,
            Avg_Bias_S)
Res

Res2 <-rbind(NormModMSERes,"X",NormMSERes,"X",NormBiasRes,"XX",SkewModMSERes,
            "X",SkewMSERes,"X",SkewBiasRes)
write.csv(Res2,file="Tallied_res.csv")
```

B.2 Section 6.3.2. Numerical example based on Ceramic Coating Process

```
Require(mass)
Require(robustbase)
Require(icsnp)
Require(sn)
Require(moments)

Options(digits=5)

Data1 <-read.table("goethtilman.txt", header=t)
Data2 <- subset(data1,select= -c(x1,x2,x3,y11,y12,y13,y1bar,s1,k1,y21,y22,y23))
Xmat  <- subset(data1,select=c(x1,x2,x3))

Delta <- function(x1){
    (sqrt((pi/2)*(abs(sskew)^(2/3))/(abs(sskew)^(2/3)+((4-pi)/2)^(2/3))))}


Parms <- matrix(na,nrow=length(data2[,1]),ncol=4,dimnames=list
            (c(1:length(data2[,1])),c("d","location","scale","shape")))

For (i in 1:length(data2[,1])){
      sskew = min(data2[i,3],0.9952717)    # associate with sample skew column
      d = delta(sskew)
      if (sskew<0) d=-1*d
      w = sqrt((data2[i,2]^2)/(1-(2*d^2)/pi))
      l = data2[i,1]-w*d*sqrt(2/pi)
      a = d/sqrt(1-d^2)
      parms[i,1] = d
      parms[i,2] = l
      parms[i,3] = w
      parms[i,4] = a
}

Parms
Despts  = 18
Obs     = 10
Num_est = 11
Reps    = 500
Num_mods= 7
# initialize arrays

Skewdata <- matrix(na, nrow=despts, ncol=obs)
Regmods <- matrix(na,nrow=10,ncol=num_mods,dimnames=list(c("bo","b1","b2",
                "b3","b11","b22","b33","b12","b13","b23"),c("moda","modb",
                "modc","modd","mode","modf","modg")))
Mod_mse_s  <- matrix(na,ncol=length(regmods[1,]),nrow=reps,
```

```
               dimnames=list(c(1:reps),c("moda","modb","modc","modd",
               "mode","modf","modg")))
Mse_tallys <- matrix(na,ncol=length(regmods[1,]),nrow=reps,
               dimnames=list(c(1:reps),c("moda","modb","modc","modd",
               "mode","modf","modg")))
Bias_tallys <-matrix(na,ncol=length(regmods[1,]),nrow=reps,
               dimnames=list(c(1:reps),c("moda","modb","modc","modd",
               "mode","modf","modg")))

For (j in 1:reps){
  for(i in 1:despts){
       skewdata[i,] <- rsn(obs, parms[i,2], parms[i,3], parms[i,4])}

R1 = as.integer(runif(1,1,18))
C1 = as.integer(runif(1,1,obs))
Skewdata[r1,c1] = max(skewdata[r1,])+(1*sd(skewdata[r1,]))

R2 = as.integer(runif(1,1,18))
C2 = as.integer(runif(1,1,obs))
Skewdata[r2,c2] = max(skewdata[r2,])+(1*sd(skewdata[r2,]))

R3= as.integer(runif(1,1,18))
C3 = as.integer(runif(1,1,obs))
Skewdata[r3,c3] = max(skewdata[r3,])+(1*sd(skewdata[r3,]))

R4 = as.integer(runif(1,1,18))
C4 = as.integer(runif(1,1,obs))
Skewdata[r4,c4] = max(skewdata[r4,])+(1*sd(skewdata[r4,]))

Estimatorskew <- matrix(na,ncol=num_est,nrow=despts,dimnames=list
                     (c(1:despts), c("ybar","s","median","mad","huber2m",
                     "huber2s","h.l","sn","qn","tau.m","tau.s")))

For (i in 1:despts){
     estimatorskew[i,1]= mean(skewdata[i,])
     estimatorskew[i,2]= sd(skewdata[i,])
     estimatorskew[i,3]= median(skewdata[i,])
     estimatorskew[i,4]= mad(skewdata[i,])
     estimatorskew[i,5]= hubers(skewdata[i,])$mu
     estimatorskew[i,6]= hubers(skewdata[i,])$s
     estimatorskew[i,7]= hl.loc(skewdata[i,])
     estimatorskew[i,8]= sn(skewdata[i,])
     estimatorskew[i,9]= qn(skewdata[i,])
     estimatorskew[i,10]= scaletau2(skewdata[i,],mu.too=t)[1]
     estimatorskew[i,11]= scaletau2(skewdata[i,],mu.too=t)[2]
}

Skew <- data.frame(cbind(xmat,estimatorskew))

Ybar2 <-skew$ybar
S2    <-skew$s
Med2  <-skew$median
Mad2  <-skew$mad
H2m2  <-skew$huber2m
H2s2  <-skew$huber2s
Hl2   <-skew$h.l
Sn2   <-skew$sn
Qn2   <-skew$qn
Taum2 <-skew$tau.m
Taus2 <-skew$tau.s
```

```
X1   <- skew$x1
X2   <- skew$x2
X3   <- skew$x3
X11  <- skew$x1*skew$x1
X22  <- skew$x2*skew$x2
X33  <- skew$x3*skew$x3
X1x2 <- skew$x1*skew$x2
X1x3 <- skew$x1*skew$x3
X2x3 <- skew$x2*skew$x3

Reg1  <-lm(ybar2 ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=skew)
Reg2  <-lm(s2    ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=skew)
Reg3  <-lm(med2  ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=skew)
Reg4  <-lm(mad2  ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=skew)
Reg5  <-lm(h2m2  ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=skew)
Reg6  <-lm(h2s2  ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=skew)
Reg7  <-lm(hl2   ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=skew)
Reg8  <-lm(sn2   ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=skew)
Reg9  <-lm(hl2   ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=skew)
Reg10 <-lm(qn2   ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=skew)
Reg11 <-lm(hl2   ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=skew)
Reg12 <-lm(mad2  ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=skew)
Reg13 <-lm(taum2 ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=skew)
Reg14 <-lm(taus2 ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=skew)

Mod.errs <- cbind(anova(reg1)["residuals","mean sq"],
                  anova(reg3)["residuals","mean sq"],
                  anova(reg5)["residuals","mean sq"],
                  anova(reg7)["residuals","mean sq"],
                  anova(reg9)["residuals","mean sq"],
                  anova(reg11)["residuals","mean sq"],
                  anova(reg13)["residuals","mean sq"])

Coeff <- data.frame(cbind(coef(reg1),coef(reg2),coef(reg3),coef(reg4),
            coef(reg5),coef(reg6),coef(reg7),coef(reg8),coef(reg9),coef(reg10),
            coef(reg11),coef(reg12),coef(reg13),coef(reg14)))

Regmods[,1] <- coef(reg1)
Regmods[,2] <- coef(reg3)
Regmods[,3] <- coef(reg5)
Regmods[,4] <- coef(reg7)
Regmods[,5] <- coef(reg9)
Regmods[,6] <- coef(reg11)
Regmods[,7] <- coef(reg13)

Se_mods <- cbind(se(reg1),se(reg3),se(reg5),se(reg7),se(reg9),se(reg11),
               se(reg13))
Oocs <- matrix(na, nrow=3, ncol=length(regmods[1,]), dimnames=list
            (c("x1","x2","x3"),c(1:length(regmods[1,]))))
Optmns <- matrix(na,nrow=1,ncol=length(regmods[1,]), dimnames=list
            (c("m(x*)"),c(1:length(regmods[1,]))))
Biass <- matrix(na,nrow=1,ncol=length(regmods[1,]), dimnames =list
            (c("bias"),c(1:length(regmods[1,]))))
Optsds <- matrix(na,nrow=1,ncol=length(regmods[1,]), dimnames=list
            (c("sd(x*)"),c(1:length(regmods[1,]))))
Objs <- matrix(na,nrow=1,ncol=length(regmods[1,]), dimnames=list
            (c("mse"),c(1:length(regmods[1,]))))
```

 # determine OOCs using *MSE*-based optimization model

```
For (k in 1:7) {
      m2 <-function(x) {
         x1<-x[1]
         x2<-x[2]
         x3<-x[3]
         coeff[1,(k+k-1)]+coeff[2,(k+k-1)]*x1+coeff[3,(k+k-1)]*x2+
         coeff[4,(k+k-1)]*x3+coeff[5,(k+k-1)]*x1^2+coeff[6,(k+k-1)]*x2^2+
         coeff[7,(k+k-1)]*x3^2+coeff[8,(k+k-1)]*x1*x2+coeff[9,(k+k-1)]*x1*x3+
         coeff[10,(k+k-1)]*x2*x3}

      s2 <-function(x) {
         x1<-x[1]
         x2<-x[2]
         x3<-x[3]
         coeff[1,(2*k)]+coeff[2,(2*k)]*x1+coeff[3,(2*k)]*x2+
         coeff[4,(2*k)]*x3+coeff[5,(2*k)]*x1^2+coeff[6,(2*k)]*x2^2+
         coeff[7,(2*k)]*x3^2+coeff[8,(2*k)]*x1*x2+coeff[9,(2*k)]*x1*x3+
         coeff[10,(2*k)]*x2*x3}

      objectives<-function(x) {
         x1<-x[1]
         x2<-x[2]
         x3<-x[3]
         (s2(c(x1,x2,x3)))^2 + (m2(c(x1,x2,x3)))^2 }

      settingss <- nlminb(start=c(0,0,0), objectives, gradient = null,
                          hessian = null, lower = c(-1.682,-1.682,-1.682),
                          upper = c(1.682,1.682,1.682))

      Oocs[,k]   <- settingss$par
      optmns[,k] <- m2(settingss$par)
      biass[,k]  <- abs(m2(settingss$par)-0)
      optsds[,k] <- s2(settingss$par)
      objs[,k]   <- settingss$objective
}

Skewresults    <- as.matrix(rbind(regmods,oocs,optmns,biass,optsds,objs))
Mod_mse_s[j,]  <- mod.errs
Mse_tallys[j,] <- skewresults["mse",]
Bias_tallys[j,]<- skewresults["bias",]

}
Avg.mod.mse.s  <- as.vector(colmeans(mod_mse_s))
Skewmodmseres  <- rbind(mod_mse_s,avg.mod.mse.s)
Avg_mse_s      <- as.vector(colmeans(mse_tallys))
Skewmseres     <- rbind(mse_tallys,avg_mse_s)
Avg_bias_s     <- as.vector(colmeans(bias_tallys))
Skewbiasres    <- rbind(bias_tallys,avg_bias_s)

Res <- rbind(avg.mod.mse.s,avg_mse_s,avg_bias_s)
Res
```

C: Supporting R Programming Code for Chapter 7

C.1 R simulation code for the Metal Cutting Experiment

```
# SCRIPT FOR TIER-2 ESTIMATION SIMULATION: Normal data and effects of variability
# SIMULATED NORMAL DATA (METAL CUTTING DATA (Shin et al. 2011) AS BASIS)
# Simulation to investigate differences between estimators used
# Conditions: normal data with low and then high variability)
# data points generated from a normal distribution

# Call up required packages or libraries in R
  require(MASS)
  require(robust)
  require(robustbase)
  require(quantreg)
  require(sn)
  require(qualityTools)
  require(AlgDesign)
  require(ICSNP)
  require(moments)

options(digits=5)         # specify format output values
despts  = 20              # number of experimental design points
obs     = 5               # number of observations per design point
num_est = 10              # number of estimation models examined
reps    = 1000            # number of simulation iterations

simdata<-read.table("metalcutting_skew.txt",header=T)
xmat  <- as.matrix(simdata[,1:3])
data2 <- subset(simdata, select=c(y2,s2,g2))

delta <- function(x){
    (sqrt((pi/2)*(abs(x)^(2/3))/(abs(x)^(2/3)+((4-pi)/2)^(2/3))))}


parms <- matrix(NA,nrow=length(data2[,1]),ncol=4,dimnames=list
             (c(1:length(data2[,1])),c("d","location","scale","shape")))

for (i in 1:length(data2[,1])){
      sskew = min((abs(data2[i,3])+0.2),0.9952717) #associate with sample skew
      d = delta(sskew)
      if (sskew<0) d=-1*d
      w = sqrt((data2[i,2]^2)/(1-(2*d^2)/pi))
      l = data2[i,1]-w*d*sqrt(2/pi)
      a = d/sqrt(1-d^2)
      parms[i,1] = d
      parms[i,2] = l
      parms[i,3] = w
      parms[i,4] = a*1      # to increase skew, change 1 to 1.5
}


# Initialize Arrays that will be repopulated in each iteration
```

```
normdata <- matrix(NA, nrow=despts, ncol=obs)
RegMod <- matrix(NA,nrow=10,ncol=num_est,dimnames=list(c("Bo","B1",
             "B2","B3","B11","B22","B33","B12","B13","B23"), c("OLS",
             "WLS-Mean","WLS-Median","LTS","S","LAD","MM","Robust
             Huber",”GLM”)))
Mod_bias  <- matrix(NA,ncol=num_est,nrow=reps,dimnames=list(c(1:reps),
                   c("OLS","WLS-Mean","WLS-Median","LTS","S","LAD",
                   "MM","Robust Huber",”GLM”)))
MSE_Tally <- matrix(NA,ncol=num_est,nrow=reps, dimnames=list(c(1:reps),
                   c("OLS","WLS-Mean","WLS-Median","LTS","S","LAD",
                   "MM","Robust Huber",”GLM”)))
```

 # Start the simulation

```
for (j in 1:reps){
```

 # Now use FOR loop and parameter estimates in "data2" to generate normal data.
 # SHOWN: High Variability scenario code.
 # For Low Variability scenario: change sample(2:6) to sample(1:1)

```
for(i in 1:despts){
    normdata [i,] <- rsn(obs, parms[i,2], parms[i,3], parms[i,4])
  }
```

 # Compute estimates for the mean, median, standard deviation, and MAD for each row in NORMDATA.

```
ybar <- matrix(apply(normdata ,1,mean),ncol=1,dimnames=list(c(1:despts),
                   c("ybarO")))
s    <- matrix(apply(normdata ,1,sd),ncol=1,dimnames=list(c(1:despts),c("sO")))
skew <- matrix(apply(normdata,1,skewness),ncol=1,dimnames=list(c(1:despts),
                   c("skew")))
med  <- matrix(apply(normdata ,1,median),ncol=1,dimnames=list(c(1:despts),
                   c("median")))
MAD  <- matrix(apply(normdata ,1,mad),ncol=1,dimnames=list(c(1:despts),
                   c("MAD")))
ybarSN <- matrix(NA,ncol=1,nrow=despts,dimnames=list(c(1:despts),c("ybarSN")))
sSN    <- matrix(NA,ncol=1,nrow=despts,dimnames=list(c(1:despts),c("sSN")))
```

 # This FOR loops facilitates the incorporation of more or less variability into the simulated data.
 # For high variability scenarios, use sample(2:5, size=1)

```
for (i in 1:despts){
      ybarSN[i,] <- ybar[i,1]+s[i,1]*(skew[i,1]/sqrt(1+skew[i,1]^2))*sqrt(2/pi)
      sSN[i,]    <- sqrt(s[i,1]^2*(1-2*((skew[i,1]^2)/(1+skew[i,1]^2))/pi))
      factor = sample(1:1,size=1)
      s[i,]  = s[i,]*factor
      sSN[i,]= sSN[i,]*factor
      MAD[i,]= MAD[i,]*factor
      }
newdat <- data.frame(cbind(xmat,ybar,med,s,MAD))  # complete design matrix

ybarO <- newdat$ybarO
ybar <- newdat$ybarSN
sO   <- newdat$sO
s    <- newdat$sSN
med  <- newdat$median
```

```
MAD  <- newdat$MAD
x1   <- newdat$x1
x2   <- newdat$x2
x3   <- newdat$x3
x11  <- newdat$x1*newdat$x1
x22  <- newdat$x2*newdat$x2
x33  <- newdat$x3*newdat$x3
x1x2 <- newdat$x1*newdat$x2
x1x3 <- newdat$x1*newdat$x3
x2x3 <- newdat$x2*newdat$x3
```
 # Build regression models for the mean and standard deviation (or MAD in select
 # cases) using each of the regression methods

```
regmorig <- lm(ybarO~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat)
regsorig <- lm(sO~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat)
regmo <-lm(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat) #OLS-mean
reg1 <- regmo                                                    #WLS-Mean
  for (i in 1:10){
    std.err.m <- SE(reg1)
    rm   <-(residuals(reg1))^2
    regm <-lm(rm~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat)
    vmx  <-fitted(regm)
    wm   <-abs(1/vmx)
    reg1 <-lm(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,weights=wm)
    critm <- abs(SE(reg1)-std.err.m)
```

 # establish iteration criteria – if differences in SE's are < 0.05 it stops

```
    if (critm[1]<0.05 && critm[2]<0.05 && critm[3]<0.05 && critm[4]<0.05 &&
        critm[5]<0.05 && critm[6]<0.05 && critm[7]<0.05 && critm[8]<0.05 &&
        critm[9]<0.05 && critm[10]<0.05) break }

regso <-lm(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat)      #OLS-std dev
reg2 <- regso                                                     #WLS-std dev
  for (i in 1:10){
    std.err.s <- SE(reg2)
    rs   <-(residuals(reg2))^2
    regs <-lm(rs~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat)
    vsx  <-fitted(regs)
    ws   <-abs(1/vsx)
    reg2 <-lm(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,weights=ws)
    crits <- abs(SE(reg2)-std.err.s)
    if (crits[1]<0.05 && crits[2]<0.05 && crits[3]<0.05 && crits[4]<0.05 &&
        crits[5]<0.05 && crits[6]<0.05 && crits[7]<0.05 && crits[8]<0.05 &&
        crits[9]<0.05 && crits[10]<0.05) break }

regmed<-lm(med~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat)     #OLS-Median
reg3 <- regmed                                                     #WLS-Median
  for (i in 1:10){
    std.err.med <- SE(reg3)
    rmed   <-(residuals(reg3))^2
    regmd <-lm(rmed~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat)
    vmedx  <-fitted(regmd)
    wmed   <-abs(1/vmedx)
    reg3 <-lm(med~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,weights=wmed)
    critmed <- abs(SE(reg3)-std.err.med)
    if (critmed[1]<0.05 && critmed[2]<0.05 && critmed[3]<0.05 &&
critmed[4]<0.05
        && critmed[5]<0.05 && critmed[6]<0.05 && critmed[7]<0.05 &&
```

```
        critmed[8]<0.05 && critmed[9]<0.05 && critmed[10]<0.05) break }

regmad<-lm(MAD~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat)      #OLS-MAD
reg4 <- regmad                                                      #WLS-MAD
  for (i in 1:10){
    std.err.mad <- SE(reg4)
    rmad   <-(residuals(reg4))^2
    regma <-lm(rmad~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat)
    vmadx  <-fitted(regma)
    wmad   <-abs(1/vmadx)
    reg4 <-lm(MAD~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,weights=wmad)
    critmad <- abs(SE(reg4)-std.err.mad)
    if (critmad[1]<0.05 && critmad[2]<0.05 && critmad[3]<0.05 &&
critmad[4]<0.05
        && critmad[5]<0.05 && critmad[6]<0.05 && critmad[7]<0.05 &&
        critmad[8]<0.05 && critmad[9]<0.05 && critmad[10]<0.05) break }

reg9 <-lqs(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,method="lts")
reg10<-lqs(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,method="lts")
reg11<-lqs(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,method="S",
           nsamp="best")
reg12<-lqs(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,method="S",
            nsamp="best")
reg13<-rq(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,tau=0.5)
reg14<-rq(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat, tau=0.5)
reg15<-lmrob(ybar ~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,maxit=75)
reg16<-lmrob(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,maxit=75)
reg17<-rlm(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,method="M",
           scale.est="proposal 2",maxit=50)
reg18<-rlm(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,method="M",
           scale.est="proposal 2",maxit=50)
reg19 <-glm(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,
            family=Gamma (link="identity"))
reg20 <-glm(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,
            family=inverse.gaussian(link="log"))
```

 # Place estimated coefficients for each regression model into data frame

```
coeff <-
data.frame(cbind(coef(regmorig),coef(regsorig),coef(regmo),coef(regso),coef(reg
1),coef(reg2),coef(reg3),coef(reg4),

coef(reg9),coef(reg10),coef(reg11),coef(reg12),coef(reg13),coef(reg14),

coef(reg15),coef(reg16),coef(reg17),coef(reg18),coef(reg19),coef(reg20)))

RegMod[,1] <- regmorig$coefficients     #OLS-orig
RegMod[,2] <- regmo$coefficients        #OLS-SN approach
RegMod[,3] <- reg1$coefficients         #WLS-mean-s
RegMod[,4] <- reg3$coefficients         #WLS-median-MAD
RegMod[,5] <- reg9$coefficients         #LTS
RegMod[,6] <- reg11$coefficients        #S
RegMod[,7] <- reg13$coefficients        #LAD
RegMod[,8] <- reg15$coefficients        #MM
RegMod[,9] <- reg17$coefficients        #Huber Prop 2
RegMod[,10]<- reg19$coefficients        #GLM with Gauss(identity)

OOC <- matrix(NA, nrow=3, ncol=num_est, dimnames=list(c("x1","x2","x3"),
```

```
             c(1:num_est)))
OOC2<- matrix(NA, ncol=(3*num_est), nrow=1)
optmn <- matrix(NA,nrow=1, ncol=num_est,dimnames=list(c("m(x*)"),
               c(1:num_est)))
bias <- matrix(NA,nrow=1,ncol=num_est,dimnames =list(c("bias"),c(1:num_est)))
optsd <- matrix(NA,nrow=1,ncol=num_est,dimnames=list(c("s(x*)"),
               c(1:num_est)))
obj <- matrix(NA,nrow=1,ncol=num_est, dimnames=list(c("MSE"),c(1:num_est)))
```

 # FOR loop to Determine OOCs for each of the regression models using Cho's/Lin and
 # Tu's MSE-based optimization scheme.

```
for (k in 1:num_est) {
      m1 <-function(x) {       # function for mean response surface model
         x1<-x[1]
         x2<-x[2]
         x3<-x[3]
         coeff[1,(k+k-1)] + coeff[2,(k+k-1)]*x1 + coeff[3,(k+k-1)]*x2 +
         coeff[4,(k+k-1)]*x3 + coeff[5,(k+k-1)]*x1^2 + coeff[6,(k+k-1)]*x2^2
         + coeff[7,(k+k-1)]*x3^2 + coeff[8,(k+k-1)]*x1*x2 + coeff[9,(k+k-
         1)]*x1*x3 + coeff[10,(k+k-1)]*x2*x3}
      s1 <-function(x) {      # function for std deviation response surface model
         x1<-x[1]
         x2<-x[2]
         x3<-x[3]
         coeff[1,(2*k)] + coeff[2,(2*k)]*x1 + coeff[3,(2*k)]*x2 +
         coeff[4,(2*k)]*x3+coeff[5,(2*k)]*x1^2 + coeff[6,(2*k)]*x2^2 +
         coeff[7,(2*k)]*x3^2 + coeff[8,(2*k)]*x1*x2+ coeff[9,(2*k)]*x1*x3 +
         coeff[10,(2*k)]*x2*x3}

      objective<-function(x) {
         x1<-x[1]
         x2<-x[2]
         x3<-x[3]
         (s1(c(x1,x2,x3)))^2 + (m1(c(x1,x2,x3))-57.5)^2 }
      settings <- nlminb(start=c(0,0,0), objective, gradient = NULL, hessian
                         = NULL, lower = c(-4,-4,-4), upper = c(4,4,4))
```

 # Collect results

```
      OOC[,k] <- settings$par
      OOC2[1,(3*k-2):(3*k)] <- as.matrix(settings$par,nrow=3,ncol=1)
      optmn[,k] <- m1(settings$par)
      bias[,k] <- abs(m1(settings$par)-57.5)
      optsd[,k]   <- s1(settings$par)
      obj[,k] <- settings$objective
}
```

 # Tally results for each iteration

```
results <- as.matrix(rbind(RegMod,OOC,optmn,bias,optsd,obj))
OOC_Tally[j,] <- OOC2
MSE_Tally[j,] <- results["MSE",]
Mod_bias[j,] <- results["bias",]
}
```

 # Compute average performance measures across all iterations

```
Avg_MSE  <- as.vector(colMeans(MSE_Tally))
```

```
Avg_Bias <- as.vector(colMeans(Mod_bias))
MSE_Res  <- rbind(MSE_Tally,Avg_MSE)
Bias_Res <- rbind(Mod_bias,Avg_Bias)

Fig_Res <- rbind(Avg_MSE,Avg_Bias)
Fig_Res

write.csv(MSE_Tally,file="MSE_TallyMC.csv")
```

C.2 Select components of the R Simulation Code for the Ceramic Coating Process

```
data1 <-read.table("GoethTilman.txt", header=T) # read in base data from
                                                # Goethals and Cho (2011)
data2 <- subset(data1,select= c(y2bar,s2,k2))
xmat  <- subset(data1,select=c(x1,x2,x3))
```

 # Define function to determine estimate for δ using Equation (6.8) from Chapter 6

```
delta <- function(x){
    (sqrt((pi/2)*(abs(x)^(2/3))/(abs(x)^(2/3)+((4-pi)/2)^(2/3))))}


parms <- matrix(NA,nrow=length(data2[,1]),ncol=4,dimnames=list
            (c(1:length(data2[,1])),c("d","location","scale","shape")))
```

 # Using FOR loop, use base data to determine estimates for δ (using delta function above) and for Skew
 # Normal Parameters (ξ,α,ω)
 # SHOWN: low asymmetry and low variability codings;
 # to inject high variability conditions, multiply w by sample(3:6, size=1)
 # to increase skew, multiply a by 1.5

```
for (i in 1:length(data2[,1])){
    sskew = min((abs(data2[i,3])+0.2),0.9952717)# associate with sample skew
    d = delta(sskew)
#   if (sskew<0) d=-1*d
    w = sqrt((data2[i,2]^2)/(1-(2*d^2)/pi))
    l = data2[i,1]-w*d*sqrt(2/pi)
    a = d/sqrt(1-d^2)
    parms[i,1] = d
    parms[i,2] = l
    parms[i,3] = w*sample(4:6,size=1)
    parms[i,4] = a*2                          # to increase asymmetry by factor of 2
}

parms
despts  = 18
obs     = 5
num_est = 10
reps    = 1000
```

 # Initialize Arrays to be populated during each iteration

```
skewdata <- matrix(NA, nrow=despts, ncol=obs)

RegMod <- matrix(NA,nrow=10,ncol=num_est,dimnames=list(c("Bo","B1","B2","B3",
```

```
           "B11","B22","B33","B12","B13","B23"),c("OLS","WLS-Mean","WLS-
Median",
           "LTS","S","LAD","MM","Robust Huber","GLM")))
Mod_bias  <- matrix(NA,ncol=num_est,nrow=reps,dimnames=list(c(1:reps),
               c("OLS","WLS-Mean","WLS-Median","LTS","S","LAD","MM",
               "Robust Huber","GLM")))
MSE_Tally <- matrix(NA,ncol=num_est,nrow=reps, dimnames=list(c(1:reps),
               c("OLS","WLS-Mean","WLS-Median","LTS","S",
               "LAD","MM","Robust Huber","GLM")))
```

 # Start simulation

```
for (j in 1:reps){
```

 # Generate data from a SN(ξ,α,ω) distribution using the parameters estimated above.

```
  for(i in 1:despts){
       skewdata[i,] <- rsn(obs, parms[i,2], parms[i,3], parms[i,4])
  }

ybar <- matrix(apply(skewdata,1,mean),ncol=1,dimnames=list(c(1:18),c("ybar")))
s    <- matrix(apply(skewdata,1,sd),ncol=1,dimnames=list(c(1:18),
               c("s")))
skew <- matrix(apply(skewdata,1,skewness),ncol=1,dimnames=list(c(1:18),
               c("skew")))
med  <- matrix(apply(skewdata,1,median),ncol=1,dimnames=list(c(1:18),
               c("median")))
MAD  <- matrix(apply(skewdata,1,mad),ncol=1,dimnames=list(c(1:18),c("MAD")))
ybarSN <- matrix(NA,ncol=1,nrow=despts,dimnames=list(c(1:despts),c("ybar")))
sSN    <- matrix(NA,ncol=1,nrow=despts,dimnames=list(c(1:despts),c("s")))
```

 # Compute estimates for SKEW NORMAL mean and standard deviation

```
for (i in 1:despts){
      ybarSN[i,] <- ybar[i,1]+s[i,1]*(skew[i,1]/sqrt(1+skew[i,1]^2))*sqrt(2/pi)
      sSN[i,]    <- sqrt(s[i,1]^2*(1-2*((skew[i,1]^2)/(1+skew[i,1]^2))/pi))}
```

 # Generate random factor to increase degree of variability at each design point

```
for (i in 1:despts){
     factor=sample(3:6,size=1)
     s[i,]=s[i,]*factor
     MAD[i,] = MAD[i,]*factor}
```

 # from this point, the simulation code mirrors the code shown in B.1 with the exception that GLMs use distribution family-link combinations specified in Table 7.6

C.3 Sample of R programming code used to evaluate all distribution-link combinations for the application of GLMs.

 # GLM models for Metal Cutting Data #

```
require(MASS)
require(robust)
require(robustbase)
require(quantreg)
require(sn)
require(qualityTools)
```

```
require(AlgDesign)
require(ICSNP)
require(moments)

options(digits=5)
despts  = 20
obs     = 15
num_est = 9

simdata<-read.table("metalcutting.txt",header=T)
xmat  <- as.matrix(simdata[,1:3])
data1 <- subset(simdata, select=c(y2,s22))
data2 <- matrix(NA,ncol=3,nrow=despts,dimnames=list(c(1:despts),
               c("y2","factor","s22")))

for (i in 1:despts){
     data2[i,1] <- data1[i,1]
     factor <- sample(2:5,size=1)           # incorporate more or  less variability
     data2[i,2] <- factor                   #  into the simulated data.
     data2[i,3] <- factor*sqrt(data1[i,2])} # High variability: use
                                            # sample(2:5, size=1)
data2
data3 <- subset(data2,select=c(y2,s22))
data3
```

 # Initialize Arrays

```
normdata <- matrix(NA, nrow=despts, ncol=obs)

for(i in 1:despts){
    normdata [i,] <- rnorm(obs, data3[i,1], data3[i,2])}

ybar <- matrix(apply(normdata,1,mean),ncol=1,dimnames=list(c(1:despts),
              c("ybar")))
s    <- matrix(apply(normdata ,1,sd),ncol=1,dimnames=list(c(1:despts),c("s")))

newdat <- data.frame(cbind(xmat,ybar,med,s,MAD))

ybar <- newdat$ybar
s    <- newdat$s
x1   <- newdat$x1
x2   <- newdat$x2
x3   <- newdat$x3
x11  <- newdat$x1*newdat$x1
x22  <- newdat$x2*newdat$x2
x33  <- newdat$x3*newdat$x3
x1x2 <- newdat$x1*newdat$x2
x1x3 <- newdat$x1*newdat$x3
x2x3 <- newdat$x2*newdat$x3

reg1  <- lm(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,newdat)
reg1b <- lm(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat)
reg2  <- glm(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,family=Gamma)
reg2b <- glm(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,family=Gamma)
reg3  <- glm(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,
             family=Gamma(link="identity"))
reg3b <- glm(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,
             family=Gamma(link="identity"))
reg4  <- glm(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,
```

```
              family=Gamma(link="log"))
reg4b <- glm(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,
              family=Gamma(link="log"))
reg5  <- glm(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,
              family=inverse.gaussian)
reg5b <- glm(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,
              family=inverse.gaussian)
reg6  <- glm(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,
              family=inverse.gaussian(link="log"))
reg6b <- glm(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,
              family=inverse.gaussian(link="log"))
reg7  <- glm(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,
              family=inverse.gaussian(link="identity"))
reg7b <- glm(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,
              family=inverse.gaussian(link="identity"))
reg8  <- glm(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,
              family=inverse.gaussian(link="inverse"))
reg8b <- glm(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=newdat,
              family=inverse.gaussian(link="inverse"))

coeff <- data.frame(cbind(coef(reg1),coef(reg1b),coef(reg2),coef(reg2b),
          coef(reg3),coef(reg3b),coef(reg4),coef(reg4b), coef(reg5),
          coef(reg5b), coef(reg6),coef(reg6b), coef(reg7),coef(reg7b),
          coef(reg8),coef(reg8b)))

tab1 <- coef(summary(reg1))[,c(1,2)]
tab2 <- coef(summary(reg2))[,c(1,2)]
tab3 <- coef(summary(reg3))[,c(1,2)]
tab4 <- coef(summary(reg4))[,c(1,2)]
tab5 <- coef(summary(reg5))[,c(1,2)]
tab6 <- coef(summary(reg6))[,c(1,2)]
tab7 <- coef(summary(reg7))[,c(1,2)]
tab8 <- coef(summary(reg8))[,c(1,2)]
tab <- cbind(tab1,tab2,tab3,tab4,tab5,tab6,tab7,tab8)
tab
```

 # Compute Akaike Information Criterion (AIC) for mean and standard deviation models.
 # Minimums determine model that yields best fit to the data

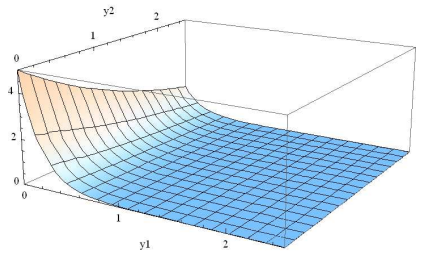
```
p<- AIC(reg2,reg3,reg4,reg5,reg6,reg7,reg8)
q<- AIC(reg2b,reg3b,reg4b,reg5b,reg6b,reg7b,reg8b)
p
q

min(p$AIC)
min(q$AIC)

par(mfrow=c(1,1))
e1 <- resid(reg1)
f1 <- fitted(reg1)
plot(log(abs(e1))~log(predict(reg1)),xlab="log(predicted)",ylab="log(resid)",ce
x=1.1,pch=16,cex.lab=1.3)
w <- glm(log(abs(e1))~log(predict(reg1)))
abline(w)
abline(v=c(3.8,4.0),col=2,lty=2)
abline(h=c(1.13,1.38),col=2,lty=2)
```

D: Supporting Programming Code and Supplementary Results for Chapter 8

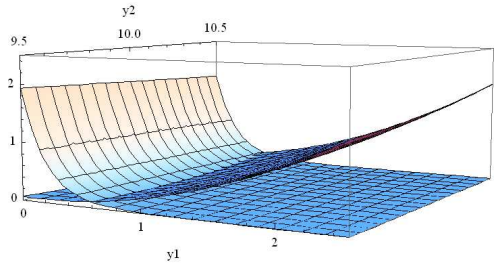
D.1 Mathematica programming code graphical outputs.

```
Needs["MultivariateStatistics`"]
△={{.7,0},{0,.2}};
μ[x1_,x2_]:={{x1}, {x2}};

f[x1_,x2_,y1_,y2_]:=PDF[MultinormalDistribution[{x1,x2},{{.7,0},{0,.2}}],
                    {y1,y2}]*(1+Erf[({{0, 1}}).MatrixPower[△,-1/2].({y1,y2}-μ[x1,x2])/√2 ]);

obj[x1_,x2_,y1_,y2_]:=N[8000*∫_2.5^9.5 ∫_2.5^9.5 f[x1, x2, y1, y2] dy1 dy2 +6000*∫_9.5^□ ∫_2.5^□ f[x1, x2, y1, y2] dy1 dy2 +
        10000*∫_10.5^□ ∫_2.5^□ f[x1, x2, y1, y2] dy1 dy2 +12500*(∫_9.5^□ ∫_2.5^□ f[x1, x2, y1, y2] dy1 dy2 +
        ∫_10.5^□ ∫_2.5^□ f[x1, x2, y1, y2] dy1 dy2 )+
        ∫_9.5^10.5 ∫_0^2.5 0.25 y1^2 0.2 y2^ 10^2 0.15 y1 y2^ 10 f[x1, x2, y1, y2] dy1 dy2 ];

NMinimize[{obj[x1,x2,y1,y2],x1>0&&x1<2.5&&x2>94&&x2<100},{x1,x2}]

N[∫_9.5^10.5 ∫_0^2.5 f[y1, y2] dy1 dy2 ]
{{1.03638×10^-21}}

△={{.07277407^2,.002520596},{0.002520596,.1193343^2}};
μ=({{.4656215},{9.996127}});
Ψ=({{.4027909},{-.3894773}});

Transpose[Ψ].△
{{0.00115149,-0.00453115}}

m1[x1_,x2_,x3_]:=0.5764+0.0008*x1+0.2118*x2-0.0388*x3+0.3708*x1^2+0.2995*x2^2+0.4739*x3^2-
        0.4900*x1*x2+0.0825*x1*x3+0.1983*x2*x3;
s1[x1_,x2_,x3_]:=0.23149-0.03264*x1+0.00142*x2+0.01276*x3-0.04617*x1^2-0.05523*x2^2-0.05738*x3^2-
0.00678*x1*x2-
        0.00489*x1*x3+0.04654*x2*x3;
g1[x1_,x2_,x3_]:=0.5527-0.06067*x1-0.01197*x2+0.16415*x3-0.07908*x1^2-0.04884*x2^2-0.11364*x3^2-
0.12814*x1*x2-
        0.03131*x1*x3-0.08179*x2*x3;
m2[x1_,x2_,x3_]:=10.0181+0.04059*x1+0.03509*x2+0.07718*x3-0.01385*x1^2+0.00971*x2^2-
        0.00443*x3^2+0.00083*x1*x2-0.02333*x1*x3-0.02667*x2*x3;
s2[x1_,x2_,x3_]:=0.10960-0.00292*x1-0.00305*x2+0.00289*x3-0.01156*x1^2-0.01173*x2^2-
        0.00726*x3^2+0.02525*x1*x2+0.00464*x1*x3-0.00649*x2*x3;
g2[x1_,x2_,x3_]:=-0.1819-0.08458*x1+0.14570*x2+0.10559*x3+0.12298*x1^2-0.01346*x2^2+0.02515*x3^2-
        0.11104*x1*x2+0.10554*x1*x3+0.07178*x2*x3;
m3[x1_,x2_,x3_]:=96.6941+0.3483*x1+0.1088*x2+0.2748*x3-1.0088*x1^2-0.1546*x2^2-0.3284*x3^2-
0.0629*x1*x2-
        0.5179*x1*x3-0.2713*x2*x3;
s3[x1_,x2_,x3_]:=0.2552-0.02873*x1+0.01041*x2+0.11675*x3-0.0155*x1^2+0.0281*x2^2+0.0085*x3^2-
0.0659*x1*x2-
        0.0103*x1*x3+0.0091*x2*x3;
g3[x1_,x2_,x3_]:=-0.2186+0.21982*x1-0.05266*x2-0.06542*x3+0.0088*x1^2-0.1929*x2^2+0.0897*x3^2-
0.1821*x1*x2-
        0.0377*x1*x3+0.0135*x2*x3;
```

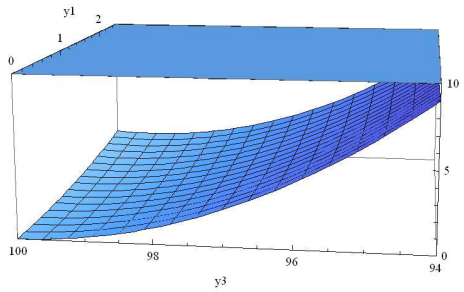
```
c12[x1_,x2_,x3_]:=0.009087-0.002237*x1-0.000247*x2+0.000142*x3-0.002076*x1^2-0.003025*x2^2-
0.003848*x3^2-
        0.000582*x1*x2-0.000368*x1*x3+0.000215*x2*x3;
c13[x1_,x2_,x3_]:=0.00600-0.004558*x1-0.000282*x2+0.001347*x3+0.005347*x1^2-0.002103*x2^2-
0.002696*x3^2-
        0.004535*x1*x2-0.000910*x1*x3+0.002449*x2*x3;
c23[x1_,x2_,x3_]:=0.01449-0.002783*x1+0.003221*x2+0.007672*x3-0.007610*x1^2-
        0.000075*x2^2+0.001800*x3^2+0.000447*x1*x2-0.001439*x1*x3+0.000797*x2*x3;
m1[0,0,0]
m2[0,0,0]
m3[0,0,0]
s1[0,0,0]^2
c12[0,0,0]
c13[0,0,0]
s2[0,0,0]^2
c23[0,0,0]
s3[0,0,0]^2
g1[0,0,0]
g2[0,0,0]
g3[0,0,0]
0.5764
10.0181
96.6941
0.0535876
0.009087
0.006
0.0120122
0.01449
0.065127
0.5527
-0.1819
-0.2186

Needs["MultivariateStatistics`"]
△={{.0535876,.009087,.006},{.009087,.0120122,.01449},{.006,.01449,.065127}};
μ=({{.5764},{10.0181},{96.6941}});

f[y1_,y2_,y3_]:=PDF[MultinormalDistribution[{0.5764,10.0181,96.6941},{{.0535876,.009087,.006},
                {.009087,.0120122,.01449},{.006,.01449,.065127}}],
                {y1,y2,y3}]*(1+Erf[({{.5527, -.1819, -.2186}  }).MatrixPower[△,-1/2].({y1,y2,y3}-μ)/√2 ]);

N[∫_94^100 ∫_9.5^10.5 ∫_0^2.5 f[y1, y2, y3] dy1 dy2 dy3]
{{0.998895}}

f[λ1_,λ2_,y1_,y2_]:=λ1*Exp[-λ1*y1-λ2*y2];
Plot3D[f[5,.5,y1,y2],{y1,0,2.5},{y2,0,2.5},AxesLabel→{y1,y2},PlotRange→{{0,2.5},{0,2.5},{0,5}}]
```



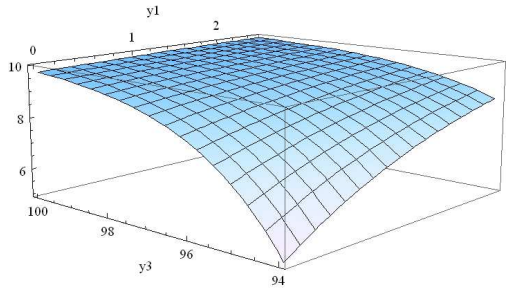

```
(*Plot for Y1 and Y2*)
g[y1_,y2_]:=.25*y1^2+.25*(y2-10)^2+.15*(y1*(y2-10));
f[λ1_,λ2_,y1_,y2_]:=(λ1)*Exp[(-λ1*y1-λ2*(y2))];
Show[Plot3D[f[5,.1,y1,y2],{y1,0,2.5},{y2,9.5,10.5},AxesLabel→{y1,y2},PlotRange→{{0,2.5},{9.5,10.5},{0,2.5}}],
        Plot3D[g[y1,y2],{y1,0,2.5},{y2,9.5,10.5},AxesLabel→{y1,y2}]]
```



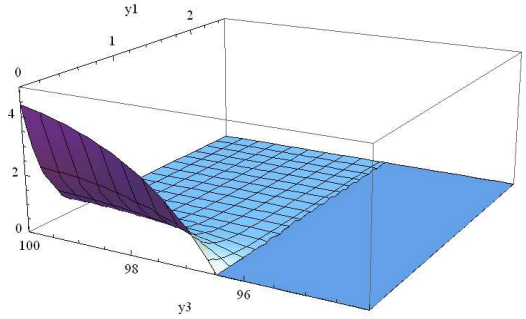

```
(*Plot for Y1 and Y3*)
g[y1_,y3_]:=.25*y1^2+.25*(y3-100)^2+.15*(y1*(y3-100));
f[λ1_,λ3_,y1_,y3_]:=100+(λ1)*(λ3)*Exp[(-λ1*(y1)-λ3*(y3))];
Show[Plot3D[f[1,1,y1,y3],{y1,0,2.5},{y3,94,100},AxesLabel→{y1,y3},PlotRange→{{0,2.5},{94,100},{0,10}}],
        Plot3D[g[y1,y3],{y1,0,2.5},{y3,94,100}]]
```



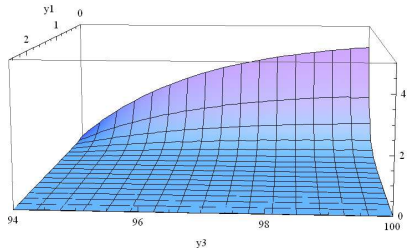

```
f[λ1_,λ3_,y1_,y3_]:=10-(λ1)*(λ3)*Exp[(-λ1*(y1)-λ3*(y3-100))];
Show[Plot3D[f[.5,.5,y1,y3],{y1,0,2.5},{y3,94,100},AxesLabel→{y1,y3}]]
```



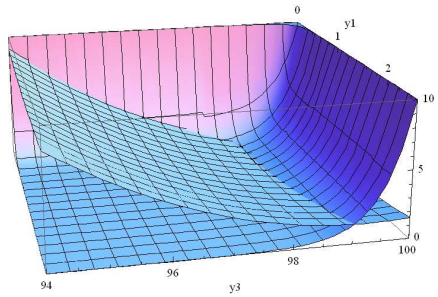
```
f[λ1_,λ3_,y1_,y3_]:=(λ1)*Exp[(-λ1*y1)] -(λ3)*Exp[(-λ1*y1-λ3*(y3-100))];
Show[Plot3D[f[5,.6,y1,y3],{y1,0,2.5},{y3,94,100},AxesLabel→{y1,y3},PlotRange→{{0,2.5},{94,100},{0,5}}]]
```



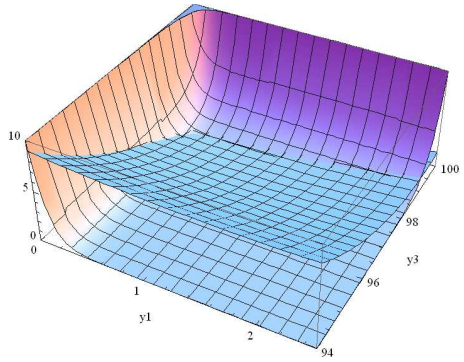

```
f[λ1_,λ3_,y1_,y3_]:=(λ1)*Exp[(-λ1*y1)]-(λ3)*Exp[(-λ1*y1-λ3*(y3-100))];
Show[Plot3D[f[4.5,.4,y1,y3],{y1,0,2.5},{y3,94,100},AxesLabel→{y1,y3},PlotRange→{{0,2.5},{94,100},{0,5}}]]
```



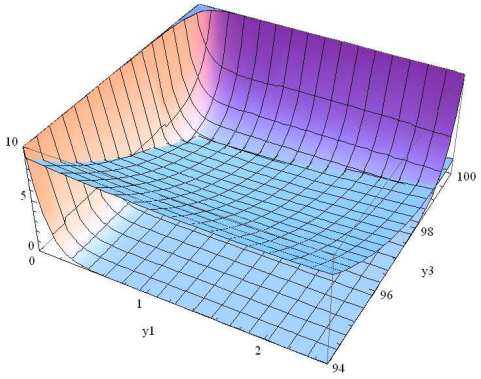

```
g[y1_,y3_]:=.25*y1^2+.25*(y3-100)^2+.15*(y1*(y3-100));
f[λ1_,λ3_,y1_,y3_]:=(λ1)*Exp[(-λ1*y1)]+10/Exp[(-λ3*(y3-100))];
Show[Plot3D[f[10,2,y1,y3],{y1,0,2.5},{y3,94,100},AxesLabel→{y1,y3},PlotRange→{{0,2.5},{94,100},{0,10}}],
        Plot3D[g[y1,y3],{y1,0,2.5},{y3,94,100}]]
```




```
g[y1_,y3_]:=.25*y1^2+.25*(y3-100)^2+.15*(y1*(y3-100));
f[λ1_,λ3_,y1_,y3_]:=(λ1)*Exp[(-λ1*y1)]+10*Exp[(λ3*(y3-100))];
Show[Plot3D[f[10,2,y1,y3],{y1,0,2.5},{y3,94,100},AxesLabel→{y1,y3},PlotRange→{{0,2.5},{94,100},{0,10}}],
        Plot3D[g[y1,y3],{y1,0,2.5},{y3,94,100}]]
```



```
g[y2_,y3_]:=.25*(y2-100)^2+.25*(y3-10)^2+.15*(y1*(y3-100));
f[λ1_,λ3_,y1_,y3_]:=(λ1)*Exp[(-λ1*y1)]+10*Exp[(λ3*(y3-100))];
Show[Plot3D[f[10,2,y1,y3],{y1,0,2.5},{y3,94,100},AxesLabel→{y1,y3},PlotRange→{{0,2.5},{94,100},{0,10}}],
        Plot3D[g[y1,y3],{y1,0,2.5},{y3,94,100}]]
```



D.2 R programming code for simulation of each scenario.

D.2.1 Optimization of Y1 and Y2 (*S*- and *N*-type QCs)

```
library(car)
library(sn)
library(NORMT3)
count<-1
```

 # Start simulation (while loop)

```
while(count<101){
  print(count)
  v1<-runif(1,.4,1)
  v2<-runif(1,.1,.6)
  v12<-runif(1,-.1,.1)
  print(v1)
  print(v2)
  print(v12)
  mu<-function(x) {
     x1<-x[1]
     x2<-x[2]
     c(x1,x2)}
  Omega<-matrix(c(v1,v12,v12,v2),ncol=2,nrow=2)
  alpha<-c(0,0)
```

 # Create functions for each double integral in the objective function

```
 one<-function(x) {
     x1<-x[1]
     x2<-x[2]
     x3<-x[3]
     x4<-x[4]
     integrate(function(x3) { sapply(x3, function(x3) { integrate(function(x4)
{
        sapply(x4, function(x4) dmsn(c(x3,x4),mu(c(x1,x2)), Omega, alpha))},
9.5,
        10.5)$value })}, 2.5, Inf)$value}
```

```
  two<-function(x) {
     x1<-x[1]
     x2<-x[2]
     x3<-x[3]
     x4<-x[4]
     integrate(function(x4) { sapply(x4, function(x4) { integrate(function(x3)
{
        sapply(x3, function(x3) dmsn(c(x3,x4),mu(c(x1,x2)), Omega, alpha))}, -
Inf,
        2.5)$value })}, -Inf, 9.5)$value}
  three<-function(x) {
     x1<-x[1]
     x2<-x[2]
     x3<-x[3]
     x4<-x[4]
     integrate(function(x4) { sapply(x4, function(x4) { integrate(function(x3)
{
        sapply(x3, function(x3) dmsn(c(x3,x4),mu(c(x1,x2)), Omega, alpha))}, -
Inf,
        2.5)$value })}, 10.5, Inf)$value}
  four<-function(x) {
     x1<-x[1]
     x2<-x[2]
     x3<-x[3]
     x4<-x[4]
     integrate(function(x4) { sapply(x4, function(x4) { integrate(function(x3)
{
        sapply(x3, function(x3) dmsn(c(x3,x4),mu(c(x1,x2)), Omega, alpha))},
2.5,
        Inf)$value })}, -Inf, 9.5)$value}
  five<-function(x) {
     x1<-x[1]
     x2<-x[2]
     x3<-x[3]
     x4<-x[4]
     integrate(function(x4) { sapply(x4, function(x4) { integrate(function(x3)
{
        sapply(x3, function(x3) dmsn(c(x3,x4),mu(c(x1,x2)), Omega, alpha))},
2.5,
        Inf)$value })}, 10.5, Inf)$value}
  six<-function(x) {
     x1<-x[1]
     x2<-x[2]
     x3<-x[3]
     x4<-x[4]
     integrate(function(x4) { sapply(x4, function(x4) { integrate(function(x3)
{
        sapply(x3, function(x3) (.25*x3^2+.25*(x4-10)^2+.15*x3*(x4-
        10))*dmsn(c(x3,x4),mu(c(x1,x2)), Omega, alpha))}, 0, 2.5)$value })},
9.5,
        10.5)$value}
  seven<-function(x) {
     x1<-x[1]
     x2<-x[2]
     x3<-x[3]
     x4<-x[4]
     integrate(function(x4) { sapply(x4, function(x4) { integrate(function(x3)
{
        sapply(x3, function(x3) (5*exp(-5*x3-
.1*x4))*dmsn(c(x3,x4),mu(c(x1,x2)),
```

```
        Omega, alpha))}, 0, 2.5)$value })}, 9.5, 10.5)$value}
  objective1<-function(x) {
     x1<-x[1]
     x2<-x[2]
     x3<-x[3]
     x4<-x[4]
     80*(one(c(x1,x2,x3,x4)))+
60*(two(c(x1,x2,x3,x4)))+100*(three(c(x1,x2,x3,x4)))+
        125*(four(c(x1,x2,x3,x4))+five(c(x1,x2,x3,x4)))+six(c(x1,x2,x3,x4))+
        100*seven(c(x1,x2,x3,x4))}

  print(r<-nlminb(start=c(2,10,1,10), objective1, gradient = NULL, hessian =
NULL,
                  lower = c(0,9.5,-Inf,-Inf), upper = c(2.5,10.5,Inf,Inf)))
  lambda<-drop(1/sqrt(1+t(alpha)%*%Omega%*%alpha))*(Omega%*%alpha)
  print("Lambda =")
  print(lambda)
  print("Optimal Process Mean =")
  print(mu(c(r$par[1],r$par[2]))+lambda*sqrt(2/pi))
  print("")
  print("")
  count<-count+1
  }
```

 # End simulation while loop

D.2.2 Optimization of Y1 and Y3 (*S*- and *L*-type QCs)

```
library(car)
library(sn)
library(NORMT3)
count<-1
```

#Start simulation (while loop)

```
while(count<101){
print(count)
v1<-runif(1,.4,1)
v3<-runif(1,1,3)
v13<-runif(1,-.1,.1)
print(v1)
print(v3)
print(v13)
mu<-function(x) {
   x1<-x[1]
   x2<-x[2]
   c(x1,x2)}
Omega<-matrix(c(v1,v13,v13,v3),ncol=2,nrow=2)
alpha<-c(0,0)
one<-function(x) {
   x1<-x[1]
   x2<-x[2]
   x3<-x[3]
   x4<-x[4]
   integrate(function(x4) { sapply(x4, function(x4) { integrate(function(x3) {
      sapply(x3, function(x3) dmsn(c(x3,x4),mu(c(x1,x2)), Omega, alpha))}, 2.5,
Inf)$value })}, 94, Inf)$value}
two<-function(x) {
   x1<-x[1]
   x2<-x[2]
```

```
   x3<-x[3]
   x4<-x[4]
   integrate(function(x4) { sapply(x4, function(x4) { integrate(function(x3) {
      sapply(x3, function(x3) dmsn(c(x3,x4),mu(c(x1,x2)), Omega, alpha))}, -
Inf, 2.5)$value })}, -Inf, 94)$value}
three<-function(x) {
   x1<-x[1]
   x2<-x[2]
   x3<-x[3]
   x4<-x[4]
   integrate(function(x4) { sapply(x4, function(x4) { integrate(function(x3) {
      sapply(x3, function(x3) dmsn(c(x3,x4),mu(c(x1,x2)), Omega, alpha))}, 2.5,
Inf)$value })}, -Inf, 94)$value}
four<-function(x) {
   x1<-x[1]
   x2<-x[2]
   x3<-x[3]
   x4<-x[4]
   integrate(function(x4) { sapply(x4, function(x4) { integrate(function(x3) {
      sapply(x3, function(x3) (.25*x3^2+.25*(x4-100)^2+.15*x3*(x4-
100))*dmsn(c(x3,x4),mu(c(x1,x2)), Omega, alpha))}, 0, 2.5)$value })}, 94,
100)$value}
five<-function(x) {
   x1<-x[1]
   x2<-x[2]
   x3<-x[3]
   x4<-x[4]
   integrate(function(x4) { sapply(x4, function(x4) { integrate(function(x3) {
      sapply(x3, function(x3) (10*exp(-10*x3)+10*exp(2*(x4-
100)))*dmsn(c(x3,x4),mu(c(x1,x2)), Omega, alpha))}, 0, 2.5)$value })}, 94,
100)$value}
objective1<-function(x) {
   x1<-x[1]
   x2<-x[2]
   x3<-x[3]
   x4<-x[4]
   80*(one(c(x1,x2,x3,x4)))+
100*(two(c(x1,x2,x3,x4)))+125*(three(c(x1,x2,x3,x4)))+(four(c(x1,x2,x3,x4))+100
*five(c(x1,x2,x3,x4)))}
print(r<-nlminb(start=c(1.5,97,1,100), objective1, gradient = NULL, hessian =
NULL, lower = c(0,94,-Inf,-Inf), upper = c(2.5,100,Inf,Inf)))
lambda<-drop(1/sqrt(1+t(alpha)%*%Omega%*%alpha))*(Omega%*%alpha)
print("Lambda =")
print(lambda)
print("Optimal Process Mean =")
print(mu(c(r$par[1],r$par[2]))+lambda*sqrt(2/pi))
print("")
print("")
count<-count+1}
```

 # End simulation while loop

D.2.3 Optimization of Y2 and Y3 (*N*- and *L*-type QCs)

```
library(car)
library(sn)
library(NORMT3)
count<-1
```

#Start simulation (while loop)

```
while(count<101){
print(count)
v2<-runif(1,.1,.6)
v3<-runif(1,1,3)
v23<-runif(1,-.1,.1)
print(v2)
print(v3)
print(v23)
mu<-function(x) {
   x1<-x[1]
   x2<-x[2]
   c(x1,x2)}
Omega<-matrix(c(v3,v23,v23,v2),ncol=2,nrow=2)
alpha<-c(0,0)
one<-function(x) {
   x1<-x[1]
   x2<-x[2]
   x3<-x[3]
   x4<-x[4]
   integrate(function(x4) { sapply(x4, function(x4) { integrate(function(x3) {
      sapply(x3, function(x3) dmsn(c(x3,x4),mu(c(x1,x2)), Omega, alpha))}, 94,
Inf)$value })}, -Inf, 9.5)$value}
two<-function(x) {
   x1<-x[1]
   x2<-x[2]
   x3<-x[3]
   x4<-x[4]
   integrate(function(x4) { sapply(x4, function(x4) { integrate(function(x3) {
      sapply(x3, function(x3) dmsn(c(x3,x4),mu(c(x1,x2)), Omega, alpha))}, 94,
Inf)$value })}, 10.5, Inf)$value}
three<-function(x) {
   x1<-x[1]
   x2<-x[2]
   x3<-x[3]
   x4<-x[4]
   integrate(function(x3) { sapply(x3, function(x3) { integrate(function(x4) {
      sapply(x4, function(x4) dmsn(c(x3,x4),mu(c(x1,x2)), Omega, alpha))}, 9.5,
10.5)$value })}, -Inf, 94)$value}
four<-function(x) {
   x1<-x[1]
   x2<-x[2]
   x3<-x[3]
   x4<-x[4]
   integrate(function(x4) { sapply(x4, function(x4) { integrate(function(x3) {
      sapply(x3, function(x3) dmsn(c(x3,x4),mu(c(x1,x2)), Omega, alpha))}, -
Inf, 94)$value })}, -Inf, 9.5)$value}
five<-function(x) {
   x1<-x[1]
   x2<-x[2]
   x3<-x[3]
   x4<-x[4]
   integrate(function(x4) { sapply(x4, function(x4) { integrate(function(x3) {
      sapply(x3, function(x3) dmsn(c(x3,x4),mu(c(x1,x2)), Omega, alpha))}, -
Inf, 94)$value })}, 10.5, Inf)$value}
six<-function(x) {
   x1<-x[1]
   x2<-x[2]
   x3<-x[3]
```

```
   x4<-x[4]
   integrate(function(x4) { sapply(x4, function(x4) { integrate(function(x3) {
      sapply(x3, function(x3) (.25*(x3-100)^2+.25*(x4-10)^2+.15*(x3-100)*(x4-
10))*dmsn(c(x3,x4),mu(c(x1,x2)), Omega, alpha))}, 94, 100)$value })}, 9.5,
10.5)$value}
seven<-function(x) {
   x1<-x[1]
   x2<-x[2]
   x3<-x[3]
   x4<-x[4]
   integrate(function(x4) { sapply(x4, function(x4) { integrate(function(x3) {
      sapply(x3, function(x3) (10*exp(2*(x3-100)))*dmsn(c(x3,x4),mu(c(x1,x2)),
Omega, alpha))}, 94, 100)$value })}, 9.5, 10.5)$value}
objective1<-function(x) {
   x1<-x[1]
   x2<-x[2]
   x3<-x[3]
   x4<-x[4]
   60*(one(c(x1,x2,x3,x4)))+
100*(two(c(x1,x2,x3,x4)))+100*(three(c(x1,x2,x3,x4)))+125*(four(c(x1,x2,x3,x4))
+five(c(x1,x2,x3,x4)))+six(c(x1,x2,x3,x4))+100*seven(c(x1,x2,x3,x4))}
print(r<-nlminb(start=c(97.5,9.75,100,10), objective1, gradient = NULL, hessian
= NULL, lower = c(94,9.5,-Inf,-Inf), upper = c(100,10.5,Inf,Inf)))
lambda<-drop(1/sqrt(1+t(alpha)%*%Omega%*%alpha))*(Omega%*%alpha)
print("Lambda =")
print(lambda)
print("Optimal Process Mean =")
print(mu(c(r$par[1],r$par[2]))+lambda*sqrt(2/pi))
print("")
print("")
count<-count+1
}
```

 # End simulation

D.3 Simulation Output.

Table D.1. Simulation output from each of the 100 iterations performed in the multi-variate cost robustness simulation.

Output includes the optimal process mean vector for each scenario (Y1*,Y2*; Y1*,Y3*; and Y2*,Y3*) and the Expected Total Cost (E[TC]) at each iterate. The mean and median E[TC] are displayed at the head of each scenario and correspond to the outputs in Figures 8.3, 8.4, and 8.5.

	Figure 8.3 (i)			Figure 8.3 (ii)			Figure 8.4 (i)			Figure 8.4 (ii)			Figure 8.5 (i)			Figure 8.5 (ii)		
	Y1, Y2 - with Skew			Y1, Y2 - Without Skew			Y1, Y3 - with Skew			Y1, Y3 - Without Skew			Y2, Y3 - with Skew			Y2, Y3 - Without Skew		
	Mean TC:		30.59	Mean TC:		36.15	Mean TC:		23.04	Mean TC:		38.17	Mean TC:		33.57	Mean TC:		41.60
	Median TC:		31.98	Median TC:		37.49	Median TC:		23.19	Median TC:		38.42	Median TC:		35.59	Median TC:		42.80
Sim Run	Y1*	Y2*	E[TC]	Y1*	Y2*	E[TC]	Y1*	Y3*	E[TC]	Y1*	Y3*	E[TC]	Y2*	Y3*	E[TC]	Y2*	Y3*	E[TC]
1	1.28	9.69	42.67	0.00	9.62	46.07	1.48	96.10	20.86	1.60	96.08	24.37	96.46	9.84	33.00	96.05	9.68	55.06
2	1.25	9.69	40.65	1.15	9.66	44.99	1.48	95.93	26.55	1.61	95.85	45.67	96.50	9.74	39.68	95.88	9.87	45.10
3	1.31	9.94	13.92	1.35	9.93	19.88	1.50	95.95	29.26	1.61	95.77	41.19	96.50	9.71	41.47	96.12	9.93	20.98
4	1.32	9.95	12.84	1.18	9.94	21.36	1.45	95.97	22.35	1.59	95.83	47.93	96.48	9.80	35.58	96.17	9.81	35.33
5	1.31	9.83	33.53	0.00	9.79	38.47	1.47	96.09	19.04	1.61	95.80	48.21	96.45	9.86	28.61	96.16	9.69	47.04
6	1.25	9.70	40.88	1.00	9.65	45.55	1.52	96.00	28.09	1.61	96.02	32.55	96.42	9.70	39.90	95.95	9.81	49.96
7	1.28	9.81	32.48	1.24	9.80	37.58	1.50	96.08	25.05	1.61	95.94	34.87	96.46	9.86	29.81	95.99	9.89	35.04
8	1.26	9.70	40.86	0.93	9.65	45.56	1.59	96.24	22.29	1.60	96.06	25.95	96.47	9.76	37.67	95.97	9.78	51.29
9	1.31	9.77	39.15	0.00	9.71	42.39	1.51	96.18	17.48	1.62	95.75	44.90	96.47	9.74	38.26	96.16	9.63	51.89
10	1.28	9.81	32.62	1.18	9.79	38.06	1.61	96.26	23.15	1.60	96.06	33.36	96.39	9.84	30.72	96.01	9.87	38.78
11	1.32	9.80	36.77	0.00	9.76	40.48	1.53	96.17	24.97	1.60	96.08	25.96	96.47	9.78	36.50	96.12	9.65	51.28
12	1.29	9.86	28.91	1.14	9.85	35.85	1.42	95.78	20.91	1.61	96.02	31.53	96.45	9.70	40.28	96.11	9.82	38.57
13	1.31	9.86	29.21	0.92	9.84	35.67	1.56	96.23	26.27	1.60	96.10	30.02	96.50	9.74	39.54	96.13	9.75	44.18
14	1.29	9.77	38.25	0.00	9.71	42.71	1.43	96.00	17.00	1.61	95.97	38.41	96.44	9.89	24.69	96.11	9.70	49.73
15	1.29	9.87	27.34	1.16	9.85	33.75	1.57	96.23	26.40	1.62	95.85	42.53	96.39	9.83	30.83	96.08	9.84	37.74
16	1.29	9.76	38.69	0.00	9.70	43.02	1.56	96.14	28.72	1.61	96.05	31.72	96.50	9.75	38.77	96.10	9.70	50.21
17	1.27	9.72	40.13	0.00	9.65	45.03	1.46	96.08	19.82	1.60	95.88	34.94	96.37	9.93	16.70	96.04	9.75	50.00
18	1.28	9.77	37.29	0.87	9.74	43.06	1.57	96.24	26.71	1.61	95.96	36.31	96.50	9.73	40.13	96.09	9.74	47.65
19	1.29	9.74	40.02	0.00	9.68	43.97	1.46	95.93	23.91	1.62	95.82	44.46	96.41	9.79	34.89	96.07	9.70	52.32
20	1.29	9.83	31.88	1.04	9.81	37.92	1.46	96.04	19.58	1.60	95.94	40.00	96.44	9.86	30.40	96.08	9.80	42.69
21	1.25	9.69	41.10	1.03	9.64	45.69	1.42	95.96	16.05	1.59	96.02	36.43	96.41	9.91	20.83	95.93	9.82	49.48
22	1.29	9.86	27.29	1.33	9.85	31.93	1.44	95.95	18.29	1.62	95.85	39.50	96.50	9.73	40.11	96.01	9.94	24.49
23	1.28	9.76	37.83	0.89	9.72	42.98	1.49	96.02	27.54	1.61	96.02	30.70	96.51	9.70	41.86	96.04	9.76	48.75
24	1.26	9.73	38.97	1.12	9.70	43.90	1.54	96.13	27.07	1.61	95.78	40.28	96.46	9.81	34.06	95.95	9.84	44.76
25	1.27	9.77	35.89	1.25	9.75	40.26	1.48	96.04	24.74	1.60	95.93	40.69	96.46	9.78	35.89	95.91	9.91	35.72
26	1.30	9.91	20.64	1.24	9.91	28.27	1.47	95.94	24.71	1.62	95.91	39.12	96.47	9.82	34.06	96.15	9.85	32.54
27	1.26	9.71	39.73	1.19	9.69	44.63	1.47	95.91	23.90	1.59	96.09	33.02	96.42	9.82	32.26	95.91	9.87	42.81
28	1.28	9.77	37.46	0.93	9.73	42.76	1.51	96.00	27.67	1.61	95.93	36.23	96.48	9.78	36.72	96.04	9.77	47.61
29	1.36	9.92	21.65	0.00	9.91	27.86	1.45	95.89	21.22	1.61	95.82	46.90	96.42	9.94	16.60	96.20	9.67	44.74
30	1.31	9.86	29.65	0.90	9.84	35.95	1.43	96.08	12.62	1.60	96.07	30.68	96.43	9.70	40.26	96.13	9.74	44.24
31	1.34	9.92	21.55	0.89	9.90	29.10	1.52	96.16	20.60	1.62	95.76	43.40	96.39	9.87	26.53	96.18	9.72	42.75
32	1.34	9.95	13.74	1.07	9.94	22.38	1.48	96.13	18.53	1.60	96.05	27.66	96.42	9.72	39.36	96.19	9.76	38.62
33	1.28	9.76	38.16	0.82	9.72	43.18	1.52	96.13	27.24	1.59	95.79	37.64	96.48	9.78	37.45	96.05	9.74	50.06
34	1.31	9.93	17.09	1.23	9.92	24.56	1.52	96.09	27.05	1.61	95.99	34.46	96.46	9.70	40.25	96.15	9.85	32.46
35	1.29	9.85	28.25	1.30	9.83	33.11	1.51	95.95	27.76	1.58	96.06	35.46	96.46	9.86	30.15	96.00	9.93	27.85
36	1.30	9.93	16.43	1.31	9.92	22.94	1.49	96.02	26.04	1.61	95.99	37.05	96.51	9.69	41.36	96.12	9.90	26.28
37	1.28	9.84	30.51	1.18	9.83	36.58	1.49	95.91	28.50	1.62	95.77	48.09	96.41	9.91	20.30	96.06	9.85	37.39
38	1.35	9.94	16.72	0.88	9.93	24.69	1.42	95.84	19.57	1.59	95.96	40.49	96.51	9.70	41.97	96.21	9.72	41.28
39	1.32	9.87	28.43	0.82	9.85	34.82	1.58	96.25	25.97	1.62	95.77	44.32	96.41	9.84	30.14	96.15	9.72	44.81
40	1.32	9.79	37.87	0.00	9.74	40.90	1.45	96.13	11.83	1.58	96.00	39.30	96.43	9.90	23.80	96.18	9.61	51.39
41	1.33	9.88	27.63	0.72	9.86	33.94	1.53	96.11	28.14	1.59	95.82	35.81	96.47	9.80	34.91	96.17	9.70	44.71
42	1.34	9.93	19.18	0.93	9.92	26.92	1.49	96.09	22.28	1.60	95.83	46.59	96.46	9.86	29.49	96.20	9.73	41.02
43	1.26	9.69	41.70	0.00	9.60	46.33	1.46	95.93	21.97	1.59	95.87	45.98	96.51	9.70	42.05	96.04	9.73	51.53
44	1.28	9.82	31.78	1.24	9.81	37.00	1.55	96.21	24.12	1.59	95.88	44.38	96.38	9.94	15.41	96.00	9.89	34.66
45	1.31	9.95	11.67	1.31	9.94	19.10	1.45	96.10	13.81	1.59	95.93	29.69	96.47	9.77	36.80	96.14	9.89	26.57
46	1.26	9.73	39.13	1.04	9.70	44.20	1.56	96.22	21.15	1.61	96.00	35.72	96.51	9.69	42.24	95.98	9.81	47.02
47	1.31	9.94	14.84	1.24	9.94	22.61	1.46	96.08	16.66	1.62	95.77	48.58	96.52	9.67	42.85	96.16	9.85	31.71
48	1.26	9.75	37.35	1.21	9.73	42.05	1.53	96.19	22.12	1.61	95.81	47.64	96.50	9.71	41.61	95.92	9.89	39.67
49	1.30	9.84	31.94	0.94	9.82	38.07	1.51	96.05	28.64	1.60	95.98	37.09	96.42	9.83	31.52	96.11	9.76	44.18
50	1.26	9.70	40.83	1.04	9.67	46.02	1.47	96.02	21.05	1.60	95.88	44.74	96.46	9.85	30.55	95.98	9.80	48.25
51	1.32	9.85	32.02	0.00	9.81	37.03	1.56	96.20	25.40	1.58	95.99	39.71	96.43	9.91	22.40	96.16	9.69	46.43
52	1.29	9.74	40.04	0.00	9.69	43.63	1.53	96.09	30.10	1.60	95.90	33.03	96.48	9.70	40.82	96.10	9.67	52.34
53	1.33	9.93	18.44	1.02	9.92	26.34	1.52	95.95	30.15	1.62	95.74	47.52	96.46	9.74	38.46	96.18	9.76	39.79
54	1.33	9.88	27.51	0.85	9.85	34.16	1.42	95.82	20.54	1.59	95.82	47.95	96.43	9.91	23.60	96.15	9.72	44.53
55	1.30	9.76	39.28	0.00	9.71	42.56	1.53	96.09	30.94	1.62	95.82	41.47	96.47	9.82	33.39	96.13	9.64	52.06
56	1.33	9.84	33.08	0.00	9.81	37.37	1.45	96.11	14.06	1.61	95.92	36.86	96.41	9.92	19.77	96.15	9.66	48.75
57	1.27	9.69	42.43	0.00	9.62	46.14	1.51	96.15	21.15	1.60	96.05	25.56	96.45	9.81	33.88	96.03	9.70	55.09
58	1.29	9.84	31.82	0.99	9.82	38.47	1.56	96.19	25.55	1.60	95.81	48.21	96.44	9.77	36.38	96.12	9.76	43.25
59	1.27	9.80	33.59	1.19	9.79	38.94	1.46	95.84	24.78	1.59	95.88	44.60	96.44	9.87	29.06	96.00	9.87	38.35
60	1.30	9.92	17.58	1.32	9.92	23.79	1.52	96.16	21.19	1.61	95.88	42.43	96.43	9.72	39.19	96.12	9.91	25.18
61	1.29	9.71	42.36	0.00	9.64	45.15	1.45	96.10	15.38	1.62	95.81	42.00	96.41	9.91	20.90	96.14	9.61	54.58
62	1.32	9.93	17.71	1.09	9.92	25.67	1.50	96.06	25.71	1.60	95.85	35.51	96.41	9.78	35.59	96.18	9.78	38.00
63	1.26	9.76	36.82	1.15	9.75	42.10	1.45	95.89	24.16	1.62	95.78	45.80	96.46	9.75	37.86	95.98	9.85	42.09
64	1.28	9.79	34.20	1.29	9.78	38.30	1.49	96.04	22.61	1.60	95.90	43.07	96.38	9.82	32.40	95.92	9.93	31.15

Table D.1 (continued.

Sim Run	Y1*	Y2*	E[TC]	Y1*	Y2*	E[TC]	Y1*	Y3*	E[TC]	Y1*	Y3*	E[TC]	Y2*	Y3*	E[TC]	Y2*	Y3*	E[TC]
65	1.31	9.93	14.60	1.33	9.93	21.10	1.48	95.90	24.33	1.61	95.96	38.20	96.47	9.83	33.06	96.13	9.91	23.41
66	1.29	9.82	30.98	1.31	9.80	35.28	1.50	95.96	26.49	1.57	96.10	33.81	96.46	9.84	30.86	95.95	9.95	27.69
67	1.31	9.91	19.02	1.35	9.91	24.26	1.45	96.01	18.58	1.61	95.75	50.19	96.41	9.94	16.09	96.09	9.94	20.22
68	1.33	9.85	32.45	0.00	9.81	37.39	1.53	96.17	20.37	1.61	95.96	39.18	96.44	9.76	37.13	96.14	9.68	48.22
69	1.30	9.88	23.82	1.33	9.87	28.74	1.56	96.21	18.71	1.62	95.76	48.91	96.51	9.69	42.23	96.04	9.94	22.98
70	1.29	9.88	26.26	1.16	9.85	33.01	1.53	96.13	27.69	1.58	96.09	33.69	96.44	9.86	30.37	96.09	9.84	37.93
71	1.33	9.93	16.60	1.09	9.93	24.72	1.50	96.17	16.65	1.61	95.91	39.56	96.42	9.76	36.60	96.18	9.78	37.93
72	1.30	9.91	19.01	1.33	9.92	26.05	1.46	95.97	20.99	1.60	96.01	28.53	96.43	9.91	22.03	96.12	9.91	24.66
73	1.29	9.88	24.68	1.23	9.88	31.56	1.51	96.04	27.39	1.59	95.80	36.40	96.05	9.70	40.29	96.11	9.86	33.54
74	1.29	9.84	29.52	1.31	9.84	34.78	1.59	96.24	21.19	1.60	95.88	33.89	96.48	9.79	36.33	96.00	9.92	27.23
75	1.31	9.75	40.27	0.00	9.70	43.09	1.45	95.91	21.35	1.61	95.86	42.65	96.47	9.76	37.41	96.13	9.62	53.33
76	1.30	9.89	22.84	1.33	9.89	28.24	1.59	96.25	26.42	1.61	96.05	31.73	96.47	9.83	32.83	96.07	9.93	23.37
77	1.26	9.70	39.85	1.21	9.69	44.53	1.53	96.10	28.09	1.60	95.90	34.50	96.47	9.76	37.76	95.89	9.88	41.82
78	1.28	9.69	42.88	0.00	9.62	45.73	1.56	96.21	19.39	1.58	95.98	26.34	96.50	9.76	38.14	96.11	9.63	55.18
79	1.26	9.76	37.07	1.16	9.73	41.95	1.50	96.12	22.75	1.59	95.81	48.46	96.44	9.90	24.30	95.95	9.86	42.25
80	1.32	9.85	32.38	0.00	9.81	37.32	1.48	96.12	16.37	1.60	96.09	24.33	96.50	9.75	39.39	96.14	9.69	47.93
81	1.30	9.78	38.04	0.00	9.73	41.79	1.48	96.03	21.92	1.60	95.88	44.39	96.48	9.76	37.57	96.12	9.67	50.83
82	1.30	9.88	25.47	1.12	9.86	32.33	1.46	96.11	15.34	1.60	95.98	28.82	96.51	9.72	40.98	96.11	9.82	38.63
83	1.33	9.90	24.27	0.92	9.87	31.56	1.45	95.91	23.23	1.61	95.98	36.82	96.50	9.74	39.88	96.16	9.73	43.17
84	1.28	9.84	30.45	1.17	9.83	36.43	1.54	96.20	25.59	1.62	95.75	45.79	96.42	9.68	41.13	96.06	9.85	37.83
85	1.29	9.71	41.88	0.00	9.66	44.79	1.52	95.96	29.19	1.59	96.06	23.80	96.40	9.75	37.10	96.10	9.64	54.60
86	1.34	9.90	24.95	0.00	9.87	31.12	1.53	96.11	29.63	1.61	95.84	45.10	96.50	9.73	40.48	96.22	9.68	44.09
87	1.28	9.83	31.56	1.15	9.82	38.16	1.51	96.02	28.67	1.60	96.02	29.18	96.39	9.86	27.78	96.08	9.83	39.09
88	1.31	9.88	26.11	1.03	9.87	33.14	1.53	96.03	29.53	1.61	95.98	34.09	96.42	9.94	16.37	96.15	9.78	40.56
89	1.26	9.76	36.62	1.14	9.73	41.61	1.58	96.23	20.61	1.59	95.94	29.46	96.46	9.85	30.74	95.96	9.86	42.67
90	1.27	9.74	39.35	0.00	9.67	44.40	1.42	95.97	17.02	1.59	95.94	29.26	96.46	9.85	30.48	96.05	9.74	49.27
91	1.29	9.79	36.70	0.79	9.75	41.90	1.45	96.07	17.27	1.59	95.85	34.17	96.42	9.91	24.20	96.08	9.73	48.82
92	1.30	9.90	21.55	1.25	9.90	29.23	1.55	96.21	18.57	1.60	95.82	36.28	96.52	9.68	42.41	96.14	9.86	31.73
93	1.31	9.92	19.33	1.14	9.92	27.38	1.51	96.12	21.61	1.61	95.82	46.92	96.47	9.81	33.90	96.18	9.80	36.68
94	1.28	9.71	41.67	0.00	9.64	45.32	1.54	96.12	26.65	1.60	95.95	39.34	96.46	9.84	32.35	96.08	9.68	53.24
95	1.27	9.80	33.68	1.24	9.78	38.50	1.53	96.02	32.07	1.61	95.91	38.44	96.46	9.86	30.02	95.96	9.90	35.58
96	1.36	9.93	19.22	0.00	9.92	25.72	1.53	96.19	18.07	1.62	95.87	40.20	96.41	9.89	24.33	96.22	9.69	42.79
97	1.33	9.91	22.03	0.92	9.90	29.41	1.49	96.01	23.69	1.62	95.85	42.40	96.51	9.72	40.83	96.18	9.73	41.93
98	1.36	9.91	23.60	0.00	9.89	28.85	1.43	95.93	20.33	1.60	95.74	40.67	96.49	9.68	41.78	96.22	9.66	44.77
99	1.28	9.80	34.67	1.05	9.76	40.29	1.60	96.26	26.06	1.62	95.80	45.14	96.38	9.83	31.27	96.03	9.81	44.52
100	1.28	9.79	35.86	0.94	9.75	41.34	1.51	96.09	25.32	1.61	96.03	30.45	96.47	9.77	36.52	96.05	9.77	47.24

E: Supporting Programming Code and Supplementary Results for Chapter 9

E.1. Simulation components for *OLS* and *WLS* methods (R programming code).

OLS and WLS Methods with FULL 2ND ORDER MODELS

 # Load required package libraries

```
require(AlgDesign)
require(ICSNP)
require(rgl)
```

Call in source data (original printing press experiment, Box-Draper (1987))

```
Origprintdat <- read.table("printOrig.txt",header=T)
Origprintdat

xmat <- subset(Origprintdat,select=c(x1,x2,x3))
```

 # Establish simulation parameters

```
obs = 3      # number of simulated observations per design point
reps = 1     # number of iterations of the experiment to perform
```

 # Function created to compute the Prediction Error Sums of Squares (PRESS)

```
PRESS <- function(object){
   press.resid = resid(object)/(1-hatvalues(object))
   sum(press.resid^2)}
```

 # Declare matrices to be filled during simulation execution

```
printobs <- matrix(NA,nrow=27,ncol=obs)
estimator <- matrix(NA,nrow=27,ncol=2,dimnames=list(c(1:27),c("m","s")))

Press <- matrix(NA,nrow=reps,ncol=4,dimnames=list(c(1:reps),c("OLS-m","WLS-
m","OLS-s","WLS-s")))
R2adj <- matrix(NA,nrow=reps,ncol=4,dimnames=list(c(1:reps),c("OLS-m","WLS-
m","OLS-s","WLS-s")))
MSEp  <- matrix(NA,nrow=reps,ncol=4,dimnames=list(c(1:reps),c("OLS-m","WLS-
m","OLS-s","WLS-s")))

ModErr_Tally_mean <- matrix(NA,nrow=reps,ncol=20,dimnames=list(c(1:reps),

c("OLSInt","x1","x2","x3","x11","x22","x33","x1x2","x1x3","x2x3",

"WLSInt","x1","x2","x3","x11","x22","x33","x1x2","x1x3","x2x3")))
ModErr_Tally_stdev <- matrix(NA,nrow=reps,ncol=20,dimnames=list(c(1:reps),

c("OLSInt","x1","x2","x3","x11","x22","x33","x1x2","x1x3","x2x3",

"WLSInt","x1","x2","x3","x11","x22","x33","x1x2","x1x3","x2x3")))
Mod_bias  <- matrix(NA,nrow=reps,ncol=2,dimnames=list(c(1:reps),c("OLS
Bias","WLS Bias")))
MSE_Tally <- matrix(NA,nrow=reps,ncol=2,dimnames=list(c(1:reps),c("OLS
MSE","WLS MSE")))
OOC_Tally <- matrix(NA,nrow=reps,ncol=6,dimnames=list(c(1:reps),
             c("ax1","ax2","ax3","bx1","bx2","bx3")))
```

 # Start simulation

```
for (j in 1:reps){
```

 # Generate experimental data using original data as basis for random variate generation

```
  for (i in 1:27){
```

```
      printobs[i,] <- rnorm(obs,Origprintdat[i,4],Origprintdat[i,5])
   }

  for (i in 1:27){
      estimator[i,1] <- mean(printobs[i,]) # compute mean and std deviation
      estimator[i,2] <- sd(printobs[i,])   # at each design point
   }

   print <- cbind(xmat,estimator) # for SINGLE ITERATION of orig. experiment,
                                  # use print <- Origprintdat
   ybar <- print$m
   s    <- print$s
   x1   <- print$x1
   x2   <- print$x2
   x3   <- print$x3
   x11  <- print$x1*print$x1
   x22  <- print$x2*print$x2
   x33  <- print$x3*print$x3
   x1x2 <- print$x1*print$x2
   x1x3 <- print$x1*print$x3
   x2x3 <- print$x2*print$x3
```

 # perform regression using OLS and WLS methods; full second order models for each

```
   regmo <-lm(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=print) # OLS-mean
   reg1 <- regmo                                                  # WLS-mean
     for (i in 1:50){
       std.err.m <- SE(reg1)
       rm   <-(residuals(reg1))^2
       regm <-lm(rm~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=print)
       vmx  <-fitted(regm)
       wm   <-abs(1/vmx)
       reg1 <-
lm(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=print,weights=wm)
       print(SE(reg1))     #prints out standard errors for each iteration at
the end of the simulation
       print(anova(reg1)[10,"Mean Sq"])
       critm <- abs(SE(reg1)-std.err.m)
       if (critm[1]<0.05 && critm[2]<0.05 && critm[3]<0.05 && critm[4]<0.05
           && critm[5]<0.05 && critm[6]<0.05 && critm[7]<0.05 &&
           critm[8]<0.05 && critm[9]<0.05 && critm[10]<0.05)
       break }

   regso <-lm(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=print) # OLS-std dev

   reg2 <- regso                                               # WLS-std dev
     for (i in 1:50){
       std.err.s <- SE(reg2)
       rs   <-(residuals(reg2))^2
       regs <-lm(rs~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=print)
       vsx  <-fitted(regs)
       ws   <-abs(1/vsx)
       reg2 <-lm(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=print,weights=ws)
       print(SE(reg2))
       crits <- abs(SE(reg2)-std.err.s)
       if (crits[1]<0.05 && crits[2]<0.05 && crits[3]<0.05 && crits[4]<0.05
           && crits[5]<0.05 && crits[6]<0.05 && crits[7]<0.05 &&
           crits[8]<0.05 && crits[9]<0.05 && crits[10]<0.05)
         break }
```

```
coeff <- data.frame(cbind(coef(regmo),coef(regso),coef(reg1),coef(reg2)))
OrigSE_Mod_mean <- cbind(SE(regmo),SE(reg1))
OrigSE_Mod_stdev<- cbind(SE(regso),SE(reg2))

OOC  <- matrix(NA, nrow=3, ncol=2, dimnames=list(c("x1","x2","x3"),
             c("OLS","WLS")))
OOC2 <- matrix(NA, nrow=1, ncol=6)
optmn<- matrix(NA,nrow=1,ncol=2, dimnames=list(c("m(x*)"),
               c("OLS","WLS")))
bias <- matrix(NA,nrow=1,ncol=2, dimnames =list(c("bias"),c("OLS","WLS")))
optsd<- matrix(NA,nrow=1,ncol=2, dimnames=list(c("var(x*)"),
               c("OLS","WLS")))
obj  <- matrix(NA,nrow=1,ncol=2, dimnames=list(c("MSE"),c("OLS","WLS")))
```

 # Determine OOCs using Cho's/Lin and Tu's MSE-based Model

```
for (k in 1:2) {

   m1 <-function(x) {
      x1<-x[1]
      x2<-x[2]
      x3<-x[3]
      coeff[1,(k+k-1)]+coeff[2,(k+k-1)]*x1+coeff[3,(k+k-1)]*
            x2+coeff[4,(k+k-1)]*x3+coeff[5,(k+k-1)]*x1^2+
             coeff[6,(k+k-1)]*x2^2+ coeff[7,(k+k-1)]*x3^2+
            coeff[8,(k+k-1)]*x1*x2+coeff[9,(k+k-1)]*x1*x3+
             coeff[10,(k+k-1)]*x2*x3}

      s1 <-function(x) {
      x1<-x[1]
      x2<-x[2]
      x3<-x[3]
      coeff[1,(2*k)]+coeff[2,(2*k)]*x1+coeff[3,(2*k)]*x2+coeff[4,(2*k)]*
            x3+coeff[5,(2*k)]*x1^2+coeff[6,(2*k)]*x2^2+ coeff[7,(2*k)]*
            x3^2+coeff[8,(2*k)]*x1*x2+coeff[9,(2*k)]*x1*x3+
             coeff[10,(2*k)]*x2*x3}

      objective<-function(x) {
      x1<-x[1]
      x2<-x[2]
      x3<-x[3]
      (s1(c(x1,x2,x3)))^2 + (m1(c(x1,x2,x3))-500)^2 }
```

 # Optimize using nlminb() function

```
      settings <- nlminb(start=c(0,0,0), objective, gradient = NULL, hessian
                         = NULL, lower = c(-4,-4,-4), upper = c(4,4,4))

   OOC[,k] <- settings$par
   OOC2[1,(3*k-2):(3*k)] <- as.matrix(settings$par,nrow=3,ncol=1)
   optmn[,k] <- m1(settings$par)
   bias[,k] <- abs(m1(settings$par)-500)
   optsd[,k]   <- s1(settings$par)
   obj[,k] <- settings$objective
}
```

 # capture iteration results

```
Press[j,] <- c(PRESS(regmo),PRESS(reg1),PRESS(regso),PRESS(reg2))
R2adj[j,] <- c(summary(regmo)$adj.r.squared, summary(reg1)$adj.r.squared,
```

```
                  summary(regso)$adj.r.squared, summary(reg2)$adj.r.squared)
   MSEp[j,]  <- c(anova(regmo)[10,"Mean Sq"],anova(reg1)[10,"Mean Sq"],
                  anova(regso)[10,"Mean Sq"],anova(reg2)[10,"Mean Sq"])

   results <- as.matrix(rbind(OOC,optmn,bias,optsd,obj))
   ModErr_Tally_mean[j,] <- OrigSE_Mod_mean
   ModErr_Tally_stdev[j,] <- OrigSE_Mod_stdev
   OOC_Tally[j,] <- OOC2
   MSE_Tally[j,] <- results["MSE",]
   Mod_bias[j,] <- results["bias",]
}
```

 # consolidate results and average

```
AvgModErrMean<- as.vector(colMeans(ModErr_Tally_mean))
AvgModErrSD  <- as.vector(colMeans(ModErr_Tally_stdev))
Avg_PRESS    <- as.vector(colMeans(Press))
Avg_R2adj    <- as.vector(colMeans(R2adj))
Avg_MSEp     <- as.vector(colMeans(MSEp))
Avg_MSE      <- as.vector(colMeans(MSE_Tally))
Avg_Bias     <- as.vector(colMeans(Mod_bias))
Avg_Results2 <-
matrix(rbind(Avg_PRESS,Avg_R2adj,Avg_MSEp,Avg_MSE,Avg_Bias),nrow=5,ncol=4,
                       dimnames=list(c("Avg PRESS","Avg R2adj","Avg MSEp","Avg
MSE","Avg Bias"),
                       c("OLS-m","WLS-m","OLS-s","WLS-s")))
Avg_Results2
AvgModErrMean
AvgModErrSD
```

 # Export results to MS Excel

```
write.csv(MSE_Tally,file="MSE_TallyPrintOLSWLS.csv")
```

E.2. Programming code for best subsets regression in R (proposed CV technique).

```
require(leaps)
```

#Origprintdat <- read.table("printOrig.txt",header=T)

```
printCV1 <- read.table("printCV.txt",header=T)

printdat <- printCV1

   ybar <- printdat$m
   s    <- printdat$s
   x1   <- printdat$x1
   x2   <- printdat$x2
   x3   <- printdat$x3
   x11  <- printdat$x1*printdat$x1
   x22  <- printdat$x2*printdat$x2
   x33  <- printdat$x3*printdat$x3
   x1x2 <- printdat$x1*printdat$x2
   x1x3 <- printdat$x1*printdat$x3
   x2x3 <- printdat$x2*printdat$x3

X <- cbind(x1,x2,x3,x11,x22,x33,x1x2,x1x3,x2x3)
best1m <- leaps(X,ybar,method="adjr2",nbest=2,names=colnames(X))
best2m <- leaps(X,ybar,method="Cp",nbest=2, names=colnames(X))
```

```
best1s <- leaps(X,s,method="adjr2",nbest=2,names=colnames(X))
best2s <- leaps(X,s,method="Cp",nbest=2, names=colnames(X))

PRESS <- function(object){
   press.resid = resid(object)/(1-hatvalues(object))
   sum(press.resid^2)}

regm2 <- lm(ybar~x1+x2+x3+x22+x33+x1x2+x1x3+x2x3,printdat)
regs2 <- lm(s~x1+x2+x3+x22+x33+x1x2+x1x3+x2x3,printdat)

regmo <-lm(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=printdat)

reg1 <- regmo  # WLS-Mean
  for (i in 1:10){
    std.err.m <- SE(reg1)
    rm   <-(residuals(reg1))^2
    regm <-lm(rm~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=printdat)
    vmx  <-fitted(regm)
    wm   <-abs(1/vmx)
    reg1 <-
lm(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=printdat,weights=wm)
    critm <- abs(SE(reg1)-std.err.m)
    if (critm[1]<0.05 && critm[2]<0.05 && critm[3]<0.05 && critm[4]<0.05 &&
        critm[5]<0.05 && critm[6]<0.05 && critm[7]<0.05 && critm[8]<0.05 &&
        critm[9]<0.05 && critm[10]<0.05)
    break }

regso <-lm(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=printdat)

reg2 <- regso  # WLS-s
  for (i in 1:10){
    std.err.s <- SE(reg2)
    rs   <-(residuals(reg2))^2
    regs <-lm(rs~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=printdat)
    vsx  <-fitted(regs)
    ws   <-abs(1/vsx)
    reg2 <-lm(s~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,data=printdat,weights=ws)
    crits <- abs(SE(reg2)-std.err.s)
    if (crits[1]<0.05 && crits[2]<0.05 && crits[3]<0.05 && crits[4]<0.05 &&
        crits[5]<0.05 && crits[6]<0.05 && crits[7]<0.05 && crits[8]<0.05 &&
        crits[9]<0.05 && crits[10]<0.05)
    break }

res1m  <- cbind(best1m$which,best1m$adjr2)
Cpm    <- cbind(best2m$Cp)
res1s  <- cbind(best1s$which,best1s$adjr2)
Cps    <- cbind(best2s$Cp)

reg1  <- lm(ybar~x1,printdat)
reg2  <- lm(ybar~x3,printdat)
reg3  <- lm(ybar~x1+x3,printdat)
reg4  <- lm(ybar~x1+x1x3,printdat)
reg5  <- lm(ybar~x1+x2+x3,printdat)
reg6  <- lm(ybar~x1+x3+x1x3,printdat)
reg7  <- lm(ybar~x1+x2+x3+x1x3,printdat)
reg8  <- lm(ybar~x1+x2+x3+x1x2,printdat)
reg9  <- lm(ybar~x1+x2+x3+x1x3+x2x3,printdat)
reg10 <- lm(ybar~x1+x2+x3+x1x2+x1x3,printdat)
reg11 <- lm(ybar~x1+x2+x3+x1x2+x1x3+x2x3,printdat)
```

```
reg12 <- lm(ybar~x1+x2+x3+x33+x1x3+x2x3,printdat)
reg13 <- lm(ybar~x1+x2+x3+x22+x1x2+x1x3+x2x3,printdat)
reg14 <- lm(ybar~x1+x2+x3+x33+x1x2+x1x3+x2x3,printdat)
reg15 <- lm(ybar~x1+x2+x3+x22+x33+x1x2+x1x3+x2x3,printdat)
reg16 <- lm(ybar~x1+x2+x3+x11+x22+x1x2+x1x3+x2x3,printdat)
reg17 <- lm(ybar~x1+x2+x3+x11+x22+x33+x1x2+x1x3+x2x3,printdat)

press <- rbind(PRESS(reg1),PRESS(reg2),PRESS(reg3),PRESS(reg4),
               PRESS(reg5),PRESS(reg6),PRESS(reg7),PRESS(reg8),
               PRESS(reg9),PRESS(reg10),PRESS(reg11),PRESS(reg12),
               PRESS(reg13),PRESS(reg14),PRESS(reg15),PRESS(reg16),
               PRESS(reg17))
MSEp  <- rbind(anova(reg1)[2,"Mean Sq"],anova(reg2)[2,"Mean Sq"],
               anova(reg3)[3,"Mean Sq"],anova(reg4)[3,"Mean Sq"],
               anova(reg5)[4,"Mean Sq"],anova(reg6)[4,"Mean Sq"],
               anova(reg7)[5,"Mean Sq"],anova(reg8)[5,"Mean Sq"],
               anova(reg9)[6,"Mean Sq"],anova(reg10)[6,"Mean Sq"],
               anova(reg11)[7,"Mean Sq"],anova(reg12)[7,"Mean Sq"],
               anova(reg13)[8,"Mean Sq"],anova(reg14)[8,"Mean Sq"],
               anova(reg15)[9,"Mean Sq"],anova(reg16)[9,"Mean Sq"],
               anova(reg17)[10,"Mean Sq"])

BESTm <- matrix(cbind(res1m,Cpm,press,MSEp),nrow=17,ncol=13,dimnames=
                list(c(1:17),c("x1","x2","x3","x11","x22","x33","x1x2",
                "x1x3","x2x3","R2adj","Cp","PRESS","MSEp")))
BESTs <- matrix(cbind(res1s,Cps),nrow=17,ncol=11,dimnames= list(c(1:17),
               c("x1","x2","x3","x11","x22","x33","x1x2","x1x3","x2x3",
               "R2adj","Cp")))
BESTm
BESTs
```

E.3. R simulation code for proposed CV technique.

```
### CV TECHNIQUE USING OPTIMAL DESIGN 1 (R GENERATED)
### Based on best subsets models (x11 omitted from both m and s functions)

require(AlgDesign)
require(ICSNP)
require(rgl)

#GENERATION OF OPTIMAL DESIGN IN R
dat2 <- gen.factorial(3,3)
dat2

# remove points 1, 5, and 19 due to high CV
dat3 <- dat2[-c(1,5,19),]
dat3

# Execute optimal design to determine which points to iterate in order to retain 27 runs
# Output provides D-, A-, G-, and I- criteria for optimal design

desD <- optFederov(~quad(X1,X2,X3),dat3,nTrials=27,eval=T,approximate=T)
desD

# Optimal design manipulated in excel to align design points with original
# data for mean and std deviation; Excel data transferred to notepad as
# "printCV.txt"

# Call in source data file
```

```
printdat <- read.table("printCV.txt",header=T)
printdat

xmat <- subset(printdat,select=c(x1,x2,x3))
```

 # Establish simulation parameters

```
obs = 3
reps = 1

PRESS <- function(object){
   press.resid = resid(object)/(1-hatvalues(object))
   sum(press.resid^2)}

printobs <- matrix(NA,nrow=27,ncol=obs)
estimator <- matrix(NA,nrow=27,ncol=2,dimnames=list(c(1:27),c("m","s")))

Press <- matrix(NA,nrow=reps,ncol=2,dimnames=list(c(1:reps),
                c("CV-m","CV-s")))
R2adj <- matrix(NA,nrow=reps,ncol=2,dimnames=list(c(1:reps),
                c("CV-m","CV-s")))
MSEp  <- matrix(NA,nrow=reps,ncol=2,dimnames=list(c(1:reps),
                c("CV-m","CV-s")))

CVModErr_Tally_mean <- matrix(NA,nrow=reps,ncol=9,dimnames=list(c(1:reps),
                              c("CVInt","x1","x2","x3","x22","x33","x1x2",
                              "x1x3","x2x3")))
CVModErr_Tally_stdev <- matrix(NA,nrow=reps,ncol=9,dimnames=list(c(1:reps),
                               c("CVInt","x1","x2","x3","x22","x33","x1x2",
                               "x1x3","x2x3")))
Mod_bias  <- matrix(NA,ncol=1,nrow=reps,dimnames=list(c(1:reps),c("Bias")))
MSE_Tally <- matrix(NA,ncol=1,nrow=reps,dimnames=list(c(1:reps),c("MSE")))
OOC_Tally <- matrix(NA,nrow=reps,ncol=3,dimnames=list(c(1:reps),
                    c("x1*","x2*","x3*")))
```

 # Start Simulation

```
for (j in 1:reps){

  for (i in 1:27){
       printobs[i,] <- rnorm(obs,printdat[i,4],printdat[i,5])
   }

  for (i in 1:27){
       estimator[i,1] <- mean(printobs[i,])
       estimator[i,2] <- sd(printobs[i,])
   }

 print <- cbind(xmat,estimator)  # for single run using orig. data use
                                 # print<-printdat
   ybar <- print$m
   s    <- print$s
   x1   <- print$x1
   x2   <- print$x2
   x3   <- print$x3
   x11  <- print$x1*print$x1
   x22  <- print$x2*print$x2
   x33  <- print$x3*print$x3
   x1x2 <- print$x1*print$x2
   x1x3 <- print$x1*print$x3
```

```
x2x3 <- print$x2*print$x3

reg1 <-lm(ybar~x1+x2+x3+x22+x33+x1x2+x1x3+x2x3,data=print) # mean
reg2 <-lm(s~x1+x2+x3+x22+x33+x1x2+x1x3+x2x3,data=print)    # std deviation

coeffm <- data.frame(coef(reg1))
coeffs <- data.frame(coef(reg2))
CVSE_Mod_mean <- cbind(SE(reg1))
CVSE_Mod_stdev<- cbind(SE(reg2))

OOC <- matrix(NA, nrow=3, ncol=1, dimnames=list(c("x1","x2","x3"),c(1:1)))
OOC2 <- matrix(NA, nrow=1, ncol=3)
optmn <- matrix(NA,nrow=1,ncol=1, dimnames=list(c("m(x*)"),c(1:1)))
bias <- matrix(NA,nrow=1,ncol=1, dimnames =list(c("bias"),c(1:1)))
optsd <- matrix(NA,nrow=1,ncol=1, dimnames=list(c("var(x*)"),c(1:1)))
obj <- matrix(NA,nrow=1,ncol=1, dimnames=list(c("MSE"),c(1:1)))
```

 # Determine OOCs using Cho's/Lin and Tu's MSE Model

```
for (k in 1:1) {

   m1 <-function(x) {
      x1<-x[1]
      x2<-x[2]
      x3<-x[3]
      coeffm[1,k]+coeffm[2,k]*x1+coeffm[3,k]*x2+coeffm[4,k]*x3+
      coeffm[5,k]*x2^2+ coeffm[6,k]*x3^2+coeffm[7,k]*x1*x2+
      coeffm[8,k]*x1*x3+ coeffm[9,k]*x2*x3}
   s1 <-function(x) {
      x1<-x[1]
      x2<-x[2]
      x3<-x[3]
      coeffs[1,k]+coeffs[2,k]*x1+coeffs[3,k]*x2+coeffs[4,k]*x3+
      coeffs[5,k]*x2^2+coeffs[6,k]*x3^2+coeffs[7,k]*x1*x2+
      coeffs[8,k]*x1*x3+coeffs[9,k]*x2*x3}
   objective<-function(x) {
      x1<-x[1]
      x2<-x[2]
      x3<-x[3]
      (s1(c(x1,x2,x3)))^2 + (m1(c(x1,x2,x3))-500)^2 }

   settings <- nlminb(start=c(0,0,0), objective, gradient = NULL,
                  hessian = NULL, lower = c(-4,-4,-4), upper = c(4,4,4))

   OOC[,k] <- settings$par
   OOC2[1,] <- settings$par
   optmn[,k] <- m1(settings$par)
   bias[,k] <- abs(m1(settings$par)-500)
   optsd[,k]   <- s1(settings$par)
   obj[,k] <- settings$objective
}

Press[j,] <- c(PRESS(reg1),PRESS(reg2))
R2adj[j,] <- c(summary(reg1)$adj.r.squared, summary(reg2)$adj.r.squared)
MSEp[j,]  <- c(anova(reg1)[9,"Mean Sq"],anova(reg2)[9,"Mean Sq"])

results <- as.matrix(rbind(OOC,optmn,bias,optsd,obj))
CVModErr_Tally_mean[j,] <- CVSE_Mod_mean
CVModErr_Tally_stdev[j,] <- CVSE_Mod_stdev
OOC_Tally[j,] <- OOC2
```

```
   MSE_Tally[j,] <- results["MSE",]
   Mod_bias[j,] <- results["bias",]
}
```
 # End simulation
 # Consolidate and average results

```
AvgModErrMean <- as.vector(colMeans(CVModErr_Tally_mean))
AvgModErrSD   <- as.vector(colMeans(CVModErr_Tally_stdev))
Avg_PRESS <- as.vector(colMeans(Press))
Avg_R2adj <- as.vector(colMeans(R2adj))
Avg_MSEp  <- as.vector(colMeans(MSEp))
Avg_MSE       <- as.vector(colMeans(MSE_Tally))
Avg_Bias      <- as.vector(colMeans(Mod_bias))
row_names <- c("Avg PRESS","Avg R2adj","Avg MSEp","Avg MSE","Avg Bias")
Avg_Results1 <-
matrix(rbind(Avg_PRESS,Avg_R2adj,Avg_MSEp,Avg_MSE,Avg_Bias),nrow=5,ncol=2,dimna
mes=list
              (c("Avg PRESS","Avg R2adj","Avg MSEp","Avg MSE","Avg
Bias"),c("CV-m","CV-s")))
Avg_Results1
AvgModErrMean
AvgModErrSD

Comb_Res <-
cbind(Avg_Results2[,1],Avg_Results2[,2],Avg_Results1[,1],Avg_Results2[,3],Avg_R
esults2[,4],Avg_Results1[,2])
Comb_Res

write.csv(MSE_Tally,file="MSE_TallyPrintCV.csv")
```

F: Supporting Mathematica Programming Code for Chapter 10 Simulations

F.1 Mathematica code for pilot study simulation.

```
(* Establish link with MS® Excel *)
Needs["ExcelLink`"]
Excel["L3"]

28.159

n = 1;
one = 0;
two = 0;
three = 0;
four = 0;

(* Define response surface functions using simulated data called from MS® Excel
*)
m[x1_, x2_, x3_] := Excel["W3"] + Excel["W4"]*x1 + Excel["W5"]*x2 +
Excel["W6"]*x3 +
     Excel["W7"]*x1*x2 + Excel["W8"]*x1*x3 + Excel["W9"]*x2*x3 +
Excel["W10"]*x1^2 +
     Excel["W11"]*x2^2 + Excel["W12"]*x3^2;
s[x1_, x2_, x3_] := Excel["Y3"] + Excel["Y4"]*x1 + Excel["Y5"]*x2 +
Excel["Y6"]*x3 +
     Excel["Y7"]*x1*x2 + Excel["Y8"]*x1*x3 + Excel["Y9"]*x2*x3 +
Excel["Y10"]*x1^2 +
     Excel["Y11"]*x2^2 + Excel["Y12"]*x3^2;

(* Start simulation – 1000 iterations *)

While[n < 1001, Print["Iteration ", n] &&

(* Define Vining and Myers (1990) priority-based optimization scheme *)
  Print[NMinimize[{(s[x1, x2, x3])^2,
     m[x1, x2, x3] == 500 && x1 >= -4 && x1 <= 4 && x2 >= -4 &&
      x2 <= 4 && x3 >= -4 && x3 <= 4}, {x1, x2, x3},
    WorkingPrecision -> 10]] &&

  Print[v = ((m[x1, x2, x3] - 500)^2 + (s[x1, x2, x3])^2) /.
     Last[NMinimize[{(s[x1, x2, x3])^2,
        m[x1, x2, x3] == 500 && x1 >= -4 && x1 <= 4 && x2 >= -4 &&
         x2 <= 4 && x3 >= -4 && x3 <= 4}, {x1, x2, x3},
       WorkingPrecision -> 10]]] &&

(* Define Lin and Tu (1995) MSE-based optimization scheme *)
  Print[NMinimize[{(s[x1, x2, x3])^2 + (m[x1, x2, x3] - 500)^2,
     x1 >= -4 && x1 <= 4 && x2 >= -4 && x2 <= 4 && x3 >= -4 &&
      x3 <= 4}, {x1, x2, x3}, WorkingPrecision -> 10]] &&

  Print[l = ((m[x1, x2, x3] - 500)^2 + (s[x1, x2, x3])^2) /.
     Last[NMinimize[{(s[x1, x2, x3])^2 + (m[x1, x2, x3] - 500)^2,
        x1 >= -4 && x1 <= 4 && x2 >= -4 && x2 <= 4 && x3 >= -4 &&
         x3 <= 4}, {x1, x2, x3}, WorkingPrecision -> 10]]] &&

(* Define Copeland and Nelson (1996) constrained bias scheme *)
  Print[NMinimize[{(s[x1, x2, x3])^2,
     m[x1, x2, x3] >= 499 && m[x1, x2, x3] <= 501 && x1 >= -4 &&
```

```
      x1 <= 4 && x2 >= -4 && x2 <= 4 && x3 >= -4 && x3 <= 4}, {x1, x2,
      x3}, WorkingPrecision -> 10]] &&

  Print[e = ((m[x1, x2, x3] - 500)^2 + (s[x1, x2, x3])^2) /.
     Last[NMinimize[{(s[x1, x2, x3])^2,
        m[x1, x2, x3] >= 499 && m[x1, x2, x3] <= 501 && x1 >= -4 &&
         x1 <= 4 && x2 >= -4 && x2 <= 4 && x3 >= -4 && x3 <= 4}, {x1,
         x2, x3}, WorkingPrecision -> 10]]] &&
(* Define Costa (2010) optimization scheme *)
  Print[NMinimize[{Abs[m[x1, x2, x3] - 500]/(1010 - 24) +
      Abs[s[x1, x2, x3]]/(158.2 - 0),
     x1 >= -4 && x1 <= 4 && x2 >= -4 && x2 <= 4 && x3 >= -4 &&
      x3 <= 4}, {x1, x2, x3}, WorkingPrecision -> 10]] &&

 Print[c = ((m[x1, x2, x3] - 500)^2 + (s[x1, x2, x3])^2) /.
     Last[NMinimize[{Abs[m[x1, x2, x3] - 500]/(1010 - 24) +
         Abs[s[x1, x2, x3]]/(158.2 - 0),
        x1 >= -4 && x1 <= 4 && x2 >= -4 && x2 <= 4 && x3 >= -4 &&
         x3 <= 4}, {x1, x2, x3}, WorkingPrecision -> 10]]] &&

  If[c == Min[v, l, e, c], four++,
   If[e == Min[v, l, e, c], three++,
    If[l == Min[v, l, e, c], two++, one++]]] &&
  Print[one, " ", two, " ", three, " " , four ] &&
  Print[m[x1, x2, x3]] && Print[s[x1, x2, x3]] && ExcelCalculate[] &&
  ClearSystemCache[];

 n++]
```

F.2 Mathematica code for Printing Press Study

F.2.1 Simulation coding for all models examined in the Printing Press numerical example.

```
Needs["ExcelLink`"]
Excel["L3"]

(*Original*)
m[x1_, x2_, x3_] :=
   Excel["W3"] + Excel["W4"]*x1 + Excel["W5"]*x2 + Excel["W6"]*x3 +
   Excel["W7"]*x1*x2 + Excel["W8"]*x1*x3 + Excel["W9"]*x2*x3 +
   Excel["W10"]*x1^2 + Excel["W11"]*x2^2 + Excel["W12"]*x3^2;
s[x1_, x2_, x3_] :=
  Excel["Y3"] + Excel["Y4"]*x1 + Excel["Y5"]*x2 + Excel["Y6"]*x3 +
   Excel["Y7"]*x1*x2 + Excel["Y8"]*x1*x3 + Excel["Y9"]*x2*x3 +
   Excel["Y10"]*x1^2 + Excel["Y11"]*x2^2 + Excel["Y12"]*x3^2;

(*Original - Higher*)
(*m[x1_,x2_,x3_]:=Excel["AF3"]+Excel["AF4"]*x1+Excel["AF5"]*x2+Excel[\
      "AF6"]*x3+Excel["AF7"]*x1*x2+Excel["AF8"]*x1*x3+Excel["AF9"]*x2*x3+\
      Excel["AF10"]*x2^2+Excel["AF11"]*x3^2+Excel["AF12"]*x2*x1^2+Excel[\
      "AF13"]*x3*x1^2+Excel["AF14"]*x2*x3^2+Excel["AF15"]*x1*x2*x3+Excel[\
      "AF16"]*(x1^2)*(x2^2);
 s[x1_,x2_,x3_]:=Excel["AH3"]+Excel["AH4"]*x1+Excel["AH5"]*x2+Excel[\
      "AH6"]*x3+Excel["AH7"]*x1*x2+Excel["AH8"]*x1*x3+Excel["AH9"]*x2*x3+\
      Excel["AH10"]*x3^2+Excel["AH11"]*x2*x1^2+Excel["AH12"]*x3*x1^2+Excel[\
      "AH13"]*x1*x2^2+Excel["AH14"]*x3*x2^2+Excel["AH15"]*x1*x2*x3;*)
```

```
(*Proposed*)
(*m[x1_,x2_,x3_]:=Excel["AB3"]+Excel["AB4"]*x1+Excel["AB5"]*x2+Excel[\
      "AB6"]*x3+Excel["AB7"]*x1*x2+Excel["AB8"]*x1*x3+Excel["AB9"]*x2*x3+\
      Excel["AB10"]*x1^2+Excel["AB11"]*x2^2+Excel["AB12"]*x3^2;
 s[x1_,x2_,x3_]:=Excel["AD3"]+Excel["AD4"]*x1+Excel["AD5"]*x2+Excel[\
      "AD6"]*x3+Excel["AD7"]*x1*x2+Excel["AD8"]*x1*x3+Excel["AD9"]*x2*x3+\
      Excel["AD10"]*x1^2+Excel["AD11"]*x2^2+Excel["AD12"]*x3^2;*)

(*Proposed - Higher Order*)
(*m[x1_,x2_,x3_]:=Excel["AC3"]+Excel["AC4"]*x1+Excel["AC5"]*x2+Excel[\
      "AC6"]*x3+Excel["AC7"]*x1*x2+Excel["AC8"]*x1*x3+Excel["AC9"]*x2*x3+\
      Excel["AC10"]*x2^2+Excel["AC11"]*x2*x1^2+Excel["AC12"]*x1*x2*x3+Excel[\
      "AC13"]*(x1^2)*(x2^2);
 s[x1_,x2_,x3_]:=Excel["AE3"]+Excel["AE4"]*x1+Excel["AE5"]*x2+Excel[\
      "AE6"]*x3+Excel["AE7"]*x1*x2+Excel["AE8"]*x2*x1^2+Excel["AE9"]*x1*x3^\
      2+Excel["AE10"]*x1*x2*x3;*)

(*Proposed - 2 - 9 reps*)
(*m[x1_,x2_,x3_]:=Excel["AF3"]+Excel["AF4"]*x1+Excel["AF5"]*x2+Excel[\
      "AF6"]*x3+Excel["AF7"]*x1*x2+Excel["AF8"]*x1*x3+Excel["AF9"]*x2*x3;
 s[x1_,x2_,x3_]:=Excel["AH3"]+Excel["AH4"]*x1+Excel["AH5"]*x2+Excel[\
      "AH6"]*x3+Excel["AH7"]*x1*x2+Excel["AH8"]*x1*x3+Excel["AH9"]*x2*x3;*)\

(*WLS*)
(*m[x1_,x2_,x3_]:=Excel["Y74"]+Excel["Y75"]*x1+Excel["Y76"]*x2+Excel[\
      "Y77"]*x3+Excel["Y78"]*x1*x2+Excel["Y79"]*x1*x3+Excel["Y80"]*x2*x3+\
      Excel["Y81"]*x1^2+Excel["Y82"]*x2^2+Excel["Y83"]*x3^2;
 s[x1_,x2_,x3_]:=Excel["Y92"]+Excel["Y93"]*x1+Excel["Y94"]*x2+Excel[\
      "Y95"]*x3+Excel["Y96"]*x1*x2+Excel["Y97"]*x1*x3+Excel["Y98"]*x2*x3+\
      Excel["Y99"]*x1^2+Excel["Y100"]*x2^2+Excel["Y101"]*x3^2;*)

(*Original - 20 reps, HO*)
(*m[x1_,x2_,x3_]:=Excel["BG3"]+Excel["BG4"]*x1+Excel["BG5"]*x2+Excel[\
      "BG6"]*x3+Excel["BG7"]*x1*x2+Excel["BG8"]*x1*x3+Excel["BG9"]*x2*x3+\
      Excel["BG10"]*x2^2+Excel["BG11"]*x3^2+Excel["BG12"]*x2*x3^2+Excel[\
      "BG13"]*x1*x2*x3+Excel["BG14"]*(x1^2)*(x2^2);
 s[x1_,x2_,x3_]:=Excel["BI3"]+Excel["BI4"]*x1+Excel["BI5"]*x2+Excel[\
      "BI6"]*x3+Excel["BI7"]*x2*x3+Excel["BI8"]*x3^2+Excel["BI9"]*x2*x1^2+\
      Excel["BI10"]*x1*x2^2+Excel["BI11"]*x1*x2*x3;

j:=NMinimize[{(s[x1,x2,x3])^2+(m[x1,x2,x3]-500)^2,x1>=-4&&x1<=4&&x2>=-\
      4&&x2<=4&&x3>=-4&&x3<=4},{{x1,1,4},{x2,-4,1},{x3,1,4}},Method->{\
      "RandomSearch","SearchPoints"->1000}];*)

(*Original - BEST CASE*)
(*m[x1_,x2_,x3_]:=Excel["AK153"]+Excel["AK154"]*x1+Excel["AK155"]*x2+\
      Excel["AK156"]*x3+Excel["AK157"]*x1*x2+Excel["AK158"]*x1*x3+Excel[\
      "AK159"]*x2*x3+Excel["AK160"]*x2^2+Excel["AK161"]*x3^2+Excel["AK162"]*\
      x2*x3^2+Excel["AK163"]*x1*x2*x3+Excel["AK164"]*(x1^2)*(x2^2);
 s[x1_,x2_,x3_]:=Excel["AK183"]+Excel["AK184"]*x1+Excel["AK185"]*x2+\
      Excel["AK186"]*x3+Excel["AK187"]*x2*x3+Excel["AK188"]*x3^2+Excel[\
      "AK189"]*x2*x1^2+Excel["AK190"]*x1*x2^2+Excel["AK191"]*x1*x2*x3;*)

(*j:=NMinimize[{(s[x1,x2,x3])^2+(m[x1,x2,x3]-500)^2,x1>=-4&&x1<=4&&x2>\
      =-4&&x2<=4&&x3>=-4&&x3<=4},{x1,x2,x3},Method->{"RandomSearch",\
      "SearchPoints"->1000}];*)

(*j:=NMinimize[{(s[x1,x2,x3])^2+(m[x1,x2,x3]-500)^2,x1>=-4&&x1<=4&&x2>\
      =-4&&x2<=4&&x3>=-4&&x3<=4},{{x1,1,4},{x2,-4,1},{x3,1,4}},Method->{\
```

```
       "NelderMead","ShrinkRatio"->0.99999999,"ContractRatio"->0.99999999,
       "ReflectRatio"->20}];*)

(* Vining and Myers (1990) scheme *)
v := NMinimize[{(s[x1, x2, x3])^2,
     m[x1, x2, x3] == 500 && x1 >= -4 && x1 <= 4 && x2 >= -4 &&
     x2 <= 4 && x3 >= -4 && x3 <= 4}, {x1, x2, x3},
     WorkingPrecision -> 10];

(* Lin and Tu (1995) scheme *)
l := NMinimize[{(s[x1, x2, x3])^2 + (m[x1, x2, x3] - 500)^2,
     x1 >= -4 && x1 <= 4 && x2 >= -4 && x2 <= 4 && x3 >= -4 &&
     x3 <= 4}, {x1, x2, x3}, WorkingPrecision -> 10];

(* Copeland and Nelson (1996) scheme *)
e := NMinimize[{(s[x1, x2, x3])^2,
     Abs[(m[x1, x2, x3] - 500)] <= 10 && x1 >= -4 && x1 <= 4 &&
     x2 >= -4 && x2 <= 4 && x3 >= -4 && x3 <= 4}, {x1, x2, x3},
     WorkingPrecision -> 10];

(* Costa (2010) scheme *)
c := NMinimize[{Abs[m[x1, x2, x3] - 500]/(1010 - 24) +
     Abs[s[x1, x2, x3]]/(158.2 - 0),
     x1 >= -4 && x1 <= 4 && x2 >= -4 && x2 <= 4 && x3 >= -4 &&
     x3 <= 4}, {x1, x2, x3}, WorkingPrecision -> 10];

n = 1;

(*While[n<5,ExcelCalculate[]&&Print["Iteration",n]&&Print[v]&&Print[l]&&Print[e
]
       &&Print[c]&&Print[m[x1,x2,x3]]&&Print[s[x1,x2,x3]]&&ClearSystemCache[];
       n++]*)

While[n < 5,ExcelCalculate[] && Print["Iteration ", n] && Print[m[x1, x2, x3]]
&&
       Print[s[x1, x2, x3]] && Print[v] && Print[l] && Print[e] && Print[c] &&
       ClearSystemCache[];
       n++]

(*While[n<501,ExcelCalculate[]&&Print[j];n++]*)

m[x1_, x2_, x3_] := Excel["V3"] + Excel["V4"]*x1 + Excel["V5"]*x2 +
Excel["V6"]*x3 +
       Excel["V7"]*x1*x2 + Excel["V8"]*x1*x3 + Excel["V9"]*x2*x3 +
       Excel["V10"]*x1^2 + Excel["V11"]*x2^2 + Excel["V12"]*x3^2;
s[x1_, x2_, x3_] := Excel["X3"] + Excel["X4"]*x1 + Excel["X5"]*x2 +
Excel["X6"]*x3 +
       Excel["X7"]*x1*x2 + Excel["X8"]*x1*x3 + Excel["X9"]*x2*x3 +
       Excel["X10"]*x1^2 + Excel["X11"]*x2^2 + Excel["X12"]*x3^2;

j := Minimize[(s[x1, x2, x3])^2 + (m[x1, x2, x3] - 500)^2,
       x1 >= -4 && x1 <= 4 && x2 >= -4 && x2 <= 4 && x3 >= -4 && x3 <= 4, {x1,
       x2,   x3}];

n = 1;
While[n < 101, ExcelCalculate[] && Print["Iteration ", n] && Print[j] &&
       Print[m[x1, x2, x3]] && Print[s[x1, x2, x3]] &&
       Print["Mean R2 and R2adj = ", Excel["W18"], " and ", Excel["W19"],
       ". SD R2 and R2adj = ", Excel["Y18"], " and ", Excel["Y19"],
       ".\n"];
```

```
n++]
```

F.2.1 Sample simulation output for Printing Press Study (NOTE: output corresponds to the first modeling group shown under (* Original *); only first 5 out of 100 iterations provided).

```
Iteration 1

{4.84327*10^-25,{x1->1.36491,x2->-0.645829,x3->0.14103}}
321.92+175.158 x1+24.058 x1^2+112.491 x2+48.2313 x1x2-43.7109 x2^2+114.137 x3+
      84.4349x1x3+47.9605 x2x3-20.0554 x3^2
20.5352+7.01733 x1-11.0244 x1^2+10.0868 x2+8.55947 x1 x2+5.16785 x2^2+23.8378
x3-
      1.08974 x1x3+15.3198 x2x3+28.7738 x3^2
Mean R2 and R2adj = 0.900541 and 0.847886. SD R2 and R2adj = 0.476855 and
0.199895.

Iteration 2

{2.34933*10^-26,{x1->-0.461313,x2->2.00615,x3->1.71429}}
343.865+187.728 x1+14.605 x1^2+113.556 x2+69.1254 x1 x2-53.6827 x2^2+
      120.995 x3+96.0085 x1 x3+55.1961 x2 x3-10.0308 x3^2
48.7925+20.6917 x1+6.7302 x1^2+9.06667 x2-8.19823 x1 x2-27.4884 x2^2+18.4101
x3+
      20.152 x1 x3+2.93436 x2 x3+6.28158 x3^2
Mean R2 and R2adj = 0.942765 and 0.912464. SD R2 and R2adj = 0.241316 and 0..

Iteration 3

{5.1276*10^-27,{x1->2.37372,x2->-2.2908,x3->2.70274}}
333.613+183.618 x1+35.5828 x1^2+118.723 x2+76.9318 x1 x2-17.3313 x2^2+
      140.996 x3+81.3407 x1 x3+50.8558 x2 x3-37.8868 x3^2
53.1862+3.72398 x1-3.97903 x1^2+5.75439 x2+4.67083 x1 x2-20.6721 x2^2+23.2486
x3-
      9.69796 x1 x3-1.29858 x2 x3+13.5253 x3^2
Mean R2 and R2adj = 0.916449 and 0.872217. SD R2 and R2adj = 0.199336 and 0..

Iteration 4

{3.2211*10^-26,{x1->2.06247,x2->-1.1309,x3->-0.106415}}
319.892+174.849 x1+47.0668 x1^2+105.079 x2+88.8005 x1 x2-21.0468 x2^2+
      136.312 x3+79.2145 x1 x3+35.9897 x2 x3-25.5838 x3^2
40.8505+31.5963 x1-1.77939 x1^2+29.1066 x2+16.3613 x1 x2-13.4615 x2^2+
      44.2703 x3+41.1292 x1 x3+26.9841 x2 x3+29.7985 x3^2
Mean R2 and R2adj = 0.864035 and 0.792053. SD R2 and R2adj = 0.613965 and
0.409593.

Iteration 5

{4.27809*10^-23,{x1->0.591542,x2->1.60699,x3->-0.676838}}
302.76+175.221 x1+32.2337 x1^2+109.941 x2+67.4136 x1 x2+8.43851 x2^2+
      127.908 x3+68.7526 x1 x3+49.604 x2 x3-26.6021 x3^2
19.964+7.01227 x1+15.7854 x1^2+0.332636 x2+5.39662 x1 x2-8.33449 x2^2+26.076
x3-
      0.630013 x1 x3+3.40454 x2 x3+15.9851 x3^2
Mean R2 and R2adj = 0.904482 and 0.853914. SD R2 and R2adj = 0.563255 and
0.332037.
```

F.3 Mathematica code for Semiconductor Study.

F.3.1 Simulation code for all model variants for the Semiconductor study.

```
Needs["ExcelLink`"]
Excel["L3"]

308.892

(*Original - 11 Reps, 2nd Order*)
(*m[x1_,x2_,x3_]:=Excel["AL3"]+Excel["AL4"]*x1+Excel["AL5"]*x2+Excel[\
       "AL6"]*x3+Excel["AL7"]*x1*x2+Excel["AL8"]*x1*x3+Excel["AL9"]*x2*x3+\
       Excel["AL10"]*x1^2+Excel["AL11"]*x2^2+Excel["AL12"]*x3^2;
s[x1_,x2_,x3_]:=Excel["AN3"]+Excel["AN4"]*x1+Excel["AN5"]*x2+Excel[\
       "AN6"]*x3+Excel["AN7"]*x1*x2+Excel["AN8"]*x1*x3+Excel["AN9"]*x2*x3+\
       Excel["AN10"]*x1^2+Excel["AN11"]*x2^2+Excel["AN12"]*x3^2;*)

(*Original - 3 Reps, 2nd Order*)
(*m[x1_,x2_,x3_]:=Excel["V3"]+Excel["V4"]*x1+Excel["V5"]*x2+Excel[\
       "V6"]*x3+Excel["V7"]*x1*x2+Excel["V8"]*x1*x3+Excel["V9"]*x2*x3+Excel[\
       "V10"]*x1^2+Excel["V11"]*x2^2+Excel["V12"]*x3^2;
s[x1_,x2_,x3_]:=Excel["X3"]+Excel["X4"]*x1+Excel["X5"]*x2+Excel["X6"]*\
       x3+Excel["X7"]*x1*x2+Excel["X8"]*x1*x3+Excel["X9"]*x2*x3+Excel["X10"]*\
       x1^2+Excel["X11"]*x2^2+Excel["X12"]*x3^2;*)

(*Original - 5 Reps Higher*)
(*m[x1_,x2_,x3_]:=Excel["X3"]+Excel["X4"]*x1+Excel["X5"]*x2+Excel[\
       "X6"]*x3+Excel["X7"]*x1*x2+Excel["X8"]*x1*x3+Excel["X9"]*x2*x3+Excel[\
       "X10"]*x1*x2*x3;
s[x1_,x2_,x3_]:=Excel["Z3"]+Excel["Z4"]*x1+Excel["Z5"]*x2+Excel["Z6"]*\
       x3+Excel["Z7"]*x2*x3+Excel["Z8"]*x2^2+Excel["Z9"]*x1*x2*x3;*)


(*Optimal - 3 Reps, Full 2nd Order*)
(*m[x1_,x2_,x3_]:=Excel["V3"]+Excel["V4"]*x1+Excel["V5"]*x2+Excel[\
       "V6"]*x3+Excel["V7"]*x1*x2+Excel["V8"]*x1*x3+Excel["V9"]*x2*x3+Excel[\
       "V10"]*x1^2+Excel["V11"]*x2^2+Excel["V12"]*x3^2;
s[x1_,x2_,x3_]:=Excel["X3"]+Excel["X4"]*x1+Excel["X5"]*x2+Excel["X6"]*\
       x3+Excel["X7"]*x1*x2+Excel["X8"]*x1*x3+Excel["X9"]*x2*x3+Excel["X10"]*\
       x1^2+Excel["X11"]*x2^2+Excel["X12"]*x3^2;*)

(*Original - Higher Order, 20 Reps*)
(*m[x1_,x2_,x3_]:=Excel["BD3"]+Excel["BD4"]*x1+Excel["BD5"]*x2+Excel[\
       "BD6"]*x3+Excel["BD7"]*x1*x3+Excel["BD8"]*x2*x3+Excel["BD9"]*x1^2+\
       Excel["BD10"]*x2^2+Excel["BD11"]*x3^2+Excel["BD12"]*x3*x1^2;
s[x1_,x2_,x3_]:=Excel["BF3"]+Excel["BF4"]*x1+Excel["BF5"]*x2+Excel[\
       "BF6"]*x3+Excel["BF7"]*x1*x2+Excel["BF8"]*x1*x3+Excel["BF9"]*x2*x3+\
       Excel["BF10"]*x1^2+Excel["BF11"]*x2^2+Excel["BF12"]*x3^2+Excel["BF13"]\
       *x2*x1^2+Excel["BF14"]*x3*x1^2+Excel["BF15"]*x1*x3^2+Excel["BF16"]*x1*\
       x2*x3+Excel["BF17"]*(x1^2)*(x2^2);*)

(*Original - Higher Order, 5 Reps*)
(*m[x1_,x2_,x3_]:=Excel["Z3"]+Excel["Z4"]*x1+Excel["Z5"]*x2+Excel[\
       "Z6"]*x3+Excel["Z7"]*x1*x3+Excel["Z8"]*x2*x3+Excel["Z9"]*x1^2+Excel[\
       "Z10"]*x2^2+Excel["Z11"]*x3^2+Excel["Z12"]*x3*x1^2;
s[x1_,x2_,x3_]:=Excel["AB3"]+Excel["AB4"]*x1+Excel["AB5"]*x2+Excel[\
       "AB6"]*x3+Excel["AB7"]*x1*x2+Excel["AB8"]*x2*x3+Excel["AB9"]*x1^2+\
       Excel["AB10"]*x2^2+Excel["AB11"]*x3^2+Excel["AB12"]*x2*x1^2+Excel[\
```

```
      "AB13"]*x3*x1^2+Excel["AB14"]*x1*x2*x3;*)

(*Optimal Design minus 3 points, 3 reps*)
(*m[x1_,x2_,x3_]:=Excel["V3"]+Excel["V4"]*x1+Excel["V5"]*x2+Excel[\
      "V6"]*x3+Excel["V7"]*x1*x2+Excel["V8"]*x1*x3+Excel["V9"]*x2*x3+Excel[\
      "V10"]*x1^2+Excel["V11"]*x2^2+Excel["V12"]*x3^2;
s[x1_,x2_,x3_]:=Excel["X3"]+Excel["X4"]*x1+Excel["X5"]*x2+Excel["X6"]*\
      x3+Excel["X7"]*x1*x2+Excel["X8"]*x1*x3+Excel["X9"]*x2*x3+Excel["X10"]*\
      x1^2+Excel["X11"]*x2^2+Excel["X12"]*x3^2;*)

(*Final-Final Optimal Design, minus 1 point, 3 reps*)
(*m[x1_,x2_,x3_]:=Excel["U3"]+Excel["U4"]*x1+Excel["U5"]*x2+Excel[\
      "U6"]*x3+Excel["U7"]*x1*x2+Excel["U8"]*x1*x3+Excel["U9"]*x2*x3+Excel[\
      "U10"]*x1^2+Excel["U11"]*x2^2;
s[x1_,x2_,x3_]:=Excel["W3"]+Excel["W4"]*x1+Excel["W5"]*x2+Excel["W6"]*\
      x3+Excel["W7"]*x1*x2+Excel["W8"]*x1*x3+Excel["W9"]*x2*x3+Excel["W10"]*\
      x1^2+Excel["W11"]*x2^2;*)

(*Best Case*)
m[x1_, x2_, x3_] := Excel["BW3"] + Excel["BW4"]*x1 + Excel["BW5"]*x2 +
Excel["BW6"]*x3
      + Excel["BW7"]*x1*x2 + Excel["BW8"]*x1*x3 + Excel["BW9"]*x2*x3 +
      Excel["BW10"]*x1^2 + Excel["BW11"]*x2^2;
s[x1_, x2_, x3_] := Excel["BY3"] + Excel["BY4"]*x1 + Excel["BY5"]*x2 +
Excel["BY6"]*x3
      + Excel["BY7"]*x1*x2 + Excel["BY8"]*x1*x3 + Excel["BY9"]*x2*x3 +
      Excel["BY10"]*x1^2 + Excel["BY11"]*x2^2;

j := NMinimize[{(s[x1, x2, x3])^2 + (m[x1, x2, x3] - 350)^2,x1^2 + x2^2 + x3^2
<= 25},
      {{x1, 0, 3.5}, {x2, -3.3, 0.5}, {x3, -1.5, 0.5}}, Method ->
      {"RandomSearch", "SearchPoints" -> 4000}];

(*j:=NMinimize[{(s[x1,x2,x3])^2+(m[x1,x2,x3]-500)^2,x1>=-4&&x1<=4&&x2>=-4&&x2
      <=4&&x3>=-4&&x3<=4},{{x1,1,4},{x2,-4,1},{x3,1,4}},Method->{"NelderMead",
      "ShrinkRatio"->0.99999999,"ContractRatio"->0.99999999,"ReflectRatio"-
>20}];*)
n = 1;
While[n < 501, ExcelCalculate[] && Print["Iteration ", n] && Print[j] &&
      Print[m[x1, x2, x3]] && Print[s[x1, x2, x3]] && ClearSystemCache[];
 n++]
```

F.3.2 Sample simulation output for Semiconductor Study (NOTE: output corresponds to the last modeling group shown under (* Best Case *); only first 5 out of 500 iterations provided).

```
Iteration _1

{3.15853×10^-32,{x1->1.27762,x2->-1.37755,x3->0.0582241}}
292.446_+12.3984 x1+0.28171 x1^2-62.5876 x2+11.0145 x1 x2-11.4636 x2^2-41.0637 x3-
51.5439 x1 x3-29.9336 x2 x3
148.515_-26.1509 x1-29.3011 x1^2-27.489 x2+31.4633 x1 x2-25.3502 x2^2-6.21366 x3-
1.68813 x1 x3+14.6574 x2 x3

Iteration _2

{1.50993×10^-31,{x1->0.877149,x2->-0.637961,x3->-1.62737}}
```

```
296.791_+29.8329 x1-15.9887 x1^2-47.3445 x2+30.4843 x1 x2-45.8036 x2^2-0.0956902
x3-62.9794 x1 x3-43.5523 x2 x3
100.648_+2.17183 x1+4.3849 x1^2-31.4991 x2-15.8847 x1 x2-8.7892 x2^2+58.9451
x3+7.8501 x1 x3-23.3123 x2 x3

Iteration _3

{1.97215×10^-31,{x1->1.60189,x2->-0.548164,x3->0.843913}}
269.932_+29.6196 x1+9.77208 x1^2-53.2132 x2-17.8765 x1 x2-6.78595 x2^2-44.144 x3-
9.21232 x1 x3-31.181 x2 x3
111.971_+4.63101 x1-24.0916 x1^2-17.1828 x2+17.1317 x1 x2-16.9687 x2^2-16.1623 x3-
19.818 x1 x3+13.8685 x2 x3

Iteration _4

{4.93038×10^-32,{x1->-0.631603,x2->-2.18276,x3->2.18132}}
283.026_+17.9429 x1-14.4193 x1^2-45.8424 x2+6.94334 x1 x2-21.3807 x2^2-36.0937 x3-
28.1558 x1 x3-24.4118 x2 x3
112.782_-2.01519 x1-18.8686 x1^2+3.83621 x2-9.58974 x1 x2-12.5017 x2^2-19.6082 x3-
16.2286 x1 x3+1.04105 x2 x3

Iteration _5

{0.,{x1->1.71306,x2->-1.89407,x3->-1.29547}}
284.619_+34.9854 x1-10.6786 x1^2-46.022 x2-0.356845 x1 x2-34.0439 x2^2-52.1134 x3-
40.0234 x1 x3-34.9432 x2 x3
124.159_+3.43006 x1-6.85392 x1^2-18.7038 x2+11.5846 x1 x2-18.0312 x2^2+3.64483 x3-
1.19138 x1 x3-16.7075 x2 x3
```

F.3.3 Secondary simulation routine and sample output (first 5 out of 100 iterations).

```
m[x1_, x2_, x3_] := Excel["V3"] + Excel["V4"]*x1 + Excel["V5"]*x2 +
Excel["V6"]*x3 +
      Excel["V7"]*x1*x2 + Excel["V8"]*x1*x3 + Excel["V9"]*x2*x3 +
      Excel["V10"]*x1^2 + Excel["V11"]*x2^2 + Excel["V12"]*x3^2;
s[x1_, x2_, x3_] := Excel["X3"] + Excel["X4"]*x1 + Excel["X5"]*x2 +
Excel["X6"]*x3 +
      Excel["X7"]*x1*x2 + Excel["X8"]*x1*x3 + Excel["X9"]*x2*x3 +
            Excel["X10"]*x1^2 + Excel["X11"]*x2^2 + Excel["X12"]*x3^2;

j := Minimize[(s[x1, x2, x3])^2 + (m[x1, x2, x3] - 500)^2,
      x1 >= -4 && x1 <= 4 && x2 >= -4 && x2 <= 4 && x3 >= -4 && x3 <=
4,{x1,x2,x3}];

n = 1; While[n < 101, ExcelCalculate[] && Print["Iteration ", n] && Print[j] &&
      Print[m[x1, x2, x3]] && Print[s[x1, x2, x3]] && Print["Mean R2 and R2adj
      = ", Excel["W18"], " and ", Excel["W19"], ". SD R2 and R2adj = ",
      Excel["Y18"], " and ", Excel["Y19"], ".\n"]
; n++]
```

(* Begin Output – Iterations 1-5 *)

```
Iteration _1

{4.84327×10^-25,{x1->1.36491,x2□-0.645829,x3->0.14103}}
```

321.92_+175.158 x1+24.058 $x1^2$+112.491 x2+48.2313 x1 x2-43.7109 $x2^2$+114.137 x3+84.4349 x1 x3+47.9605 x2 x3-20.0554 $x3^2$
20.5352_+7.01733 x1-11.0244 $x1^2$+10.0868 x2+8.55947 x1 x2+5.16785 $x2^2$+23.8378 x3-1.08974 x1 x3+15.3198 x2 x3+28.7738 $x3^2$
Mean R2 and R2adj = 0.900541 and 0.847886. SD R2 and R2adj = 0.476855 and 0.199895.

Iteration 2

{2.34933×10^{-26},{x1->-0.461313,x2->2.00615,x3->1.71429}}
343.865+187.728 x1+14.605 $x1^2$+113.556 x2+69.1254 x1 x2-53.6827 $x2^2$+120.995 x3+96.0085 x1 x3+55.1961 x2 x3-10.0308 $x3^2$
48.7925+20.6917 x1+6.7302 $x1^2$+9.06667 x2-8.19823 x1 x2-27.4884 $x2^2$+18.4101 x3+20.152 x1 x3+2.93436 x2 x3+6.28158 $x3^2$
Mean R2 and R2adj = 0.942765 and 0.912464. SD R2 and R2adj = 0.241316 and 0..

Iteration 3

{5.1276×10^{-27},{x1->2.37372,x2->-2.2908,x3->2.70274}}
333.613+183.618 x1+35.5828 $x1^2$+118.723 x2+76.9318 x1 x2-17.3313 $x2^2$+140.996 x3+81.3407 x1 x3+50.8558 x2 x3-37.8868 $x3^2$
53.1862+3.72398 x1-3.97903 $x1^2$+5.75439 x2+4.67083 x1 x2-20.6721 $x2^2$+23.2486 x3-9.69796 x1 x3-1.29858 x2 x3+13.5253 $x3^2$
Mean R2 and R2adj = 0.916449 and 0.872217. SD R2 and R2adj = 0.199336 and 0..

Iteration 4

{3.2211×10^{-26},{x1->2.06247,x2->-1.1309,x3->-0.106415}}
319.892+174.849 x1+47.0668 $x1^2$+105.079 x2+88.8005 x1 x2-21.0468 $x2^2$+136.312 x3+79.2145 x1 x3+35.9897 x2 x3-25.5838 $x3^2$
40.8505+31.5963 x1-1.77939 $x1^2$+29.1066 x2+16.3613 x1 x2-13.4615 $x2^2$+44.2703 x3+41.1292 x1 x3+26.9841 x2 x3+29.7985 $x3^2$
Mean R2 and R2adj = 0.864035 and 0.792053. SD R2 and R2adj = 0.613965 and 0.409593.

Iteration 5

{4.27809×10^{-23},{x1->0.591542,x2->1.60699,x3->-0.676838}}
302.76+175.221 x1+32.2337 $x1^2$+109.941 x2+67.4136 x1 x2+8.43851 $x2^2$+127.908 x3+68.7526 x1 x3+49.604 x2 x3-26.6021 $x3^2$
19.964+7.01227 x1+15.7854 $x1^2$+0.332636 x2+5.39662 x1 x2-8.33449 $x2^2$+26.076 x3-0.630013 x1 x3+3.40454 x2 x3+15.9851 $x3^2$
Mean R2 and R2adj = 0.904482 and 0.853914. SD R2 and R2adj = 0.563255 and 0.332037.

WORKS CITED

Abbasi, B. (2009) "A neural network applied to estimate process capability of non-normal processes", *Journal of Expert Systems with Applications*, Vol. 36, No. 2, pp. 3093-3100.

Anderson, R. (2008). *Modern Methods for Robust Regression: Quantitative Applications in the Social Sciences*. Sage, Los Angeles, CA.

Arcelus, F.J., Rahim, M.A. (1990) "Optimum process levels for the joint control of variables and attributes", *European Journal of Operations Research*, Vol. 45, Nos. 2-3, pp. 224-230.

Arellano-Valle, R.B., Gomez, H., Quintana, F.A.(2004), "A new class of skew normal distributions", *Communications in Statistics – Theory and Methods* 2004; Vol. 33, pp. 1465-1480.

Arvidsson, M., Gremyr, I. (2008), "Principles of robust design methodology", *Quality and Reliability Engineering International*, Vol. 24, No. 1, pp. 23-35.

Ash, A., Hedayat, A. (1978) "An introduction to design optimality with an overview of the literature." *Communications in Statistics, Part A - Theory and Methods, Vol.* 7, No. 1, pp. 1295-1325.

Atkinson, A.C. (1982) "Developments in the Design of Experiments." *International Statistical Review*, Vol. 50, No. 2, pp. 161-177.

Atkinson, A.C., Cheng, T.-C. (1999), "Computing least trimmed squares regression with the forward search", *Statistics and Computing*, Vol. 9, pp. 251-263.

Atkinson, A.C., Dovev A.N. (1992) *Optimum Experimental Designs*. Oxford University Press, Oxford, England.

Azadeh A, Saberi M, Anvari M. (2011), "An integrated artificial neural network fuzzy c-means normalization algorithm for performance assessment of decision-making units: the cases of auto industry and power plant", *Computers and Industrial Engineering*, Vol. 60, pp. 328-340.

Azzalini, A. (1985), "A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, Vol. 12, pp. 171–178.

Azzalini, A. (2011). R Package "sn": the skew normal and skew-t distributions (version 0.4-17); URL http://azzalini.stat.unipd.it/SN.

Azzalini, A., Capitanio, A. (1999), "Statistical applications of the multivariate skew normal distribution", *Journal of the Royal Statistical Society B*, Vol. 61, No. 3, pp. 579-602.

Azzalini, A., Valle, D. (1996), "The multivariate skew-normal distribution", *Biometrika*, Vol. 83, No. 4, pp. 715-726.

Ballestero, E. (2007), "Compromise programming: a utility-based linear-quadratic composite metric from the trade-off between achievement and balanced (non-corner) solutions", *European Journal of Operations Research*, Vol. 182, No. 3, pp. 1369-1382.

Baykasoglu, A., Owen, S., Gindy, N. (1999), "Solution of goal programming models use a basic taboo search algorithm", *Journal of the Operational Research Society*, Vol. 50, pp. 960-973.

Bendell, A., Disney, J., Pridmore, W.A. (1987), *Taguchi Methods: Applications in World Industry*. IFS Publications, London.

Bera, A.K., Bilias, Y. (2002, "The MM, ME, ML, EL, EF, and GMM approaches to estimation: a synthesis", *Journal of Econometrics*, Vol. 107, pp. 51-86.

Bettes, D.C. (1962) "Finding an optimum target value in relation to a fixed lower limit and arbitrary upper limit", *Applied Statistics*, Vol. 11, No. 3, pp. 202-210.

Bisgaard, T.O., Hunter, W.C., Pallensen, L. (1984) "Economic selection of quality of manufactured products", *Technometrics*, Vol. 26, No. 1, pp. 9-18.

Blom G. (1958), *Statistical Estimates and Transformed Beta Variables*. Wiley, New York, NY.

Bose, R.C., Carter, R.L. (1959) "Complex Representation in the Construction of Rotatable Designs." *The Annals of Mathematical Statistics*, Vol. 30, No. 3, pp. 771-780.

Bose, R.C., Draper, N.R. (1959) "Second Order Rotatable Designs in Three Dimensions." *The Annals of Mathematical Sciences, Vol.* 30, No. 4, pp. 1097-1112.

Box, G.E.P. (1988), "Signal-to-noise ratios, performance criteria, and transformations", *Technometrics*, Vol. 30, pp. 1-17.

Box, G.E.P., Behnken, D.W. (1960a) "Simplex-Sum Designs: A Class of Second Order Rotatable Designs Derivable from Those of First Order." *The Annals of Mathematical Sciences*, Vol. 31, No. 4, pp. 838-864.

Box, G.E.P., Behnken, D.W. (1960b) "Some New Three-Level Designs for the Study of Quantitative Variables." *Technometrics*, Vol. 2, No. 4, pp. 455-475.

Box, G.E.P., Bisgaard, S., Fung, C. (1988), "An explanation and critique of Taguchi's contributions to quality engineering", *International Journal of Reliability Management*, Vol. 4, pp. 123-131.

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

Box, G.E.P., Draper, N.R. (1963) "The Choice of a Second Order Rotatable Design." *Biometrika*, Vol. 50, Nos. 3-4, pp. 335-352.

Box, G.E.P., Draper, N.R. (1987), *Empirical Model Building and Response Surfaces*, Wiley, New York, NY.

Box, G.E.P., Wilson, K.B. (1951) "On the Experimental Attainment of Optimum Conditions." *Journal of the Royal Statistical Society, Series B* (*Methodological*) Vol. 13, No. 1, pp. 1-45.

Boylan, G.L., Goethals, P.L. (2011), "Achieving Cost Robustness in Processes with Mixed Multiple Quality Characteristics and Dynamic Variability," *International Journal of Experimental Design and Production Operations*, Vol. 2, No. 3, pp. 243-264.

Boylan, G.L., Cho, B.R. (2012a), "The Normal Probability Plot as a Tool for Understanding Data: A Shape Analysis from the Perspective of Skewness, Kurtosis, and Variability", *Quality and Reliability Engineering International*, Vol. 28, No. 3, pp. 249-264.

Boylan, G.L., Cho, B.R. (2012b), "Studies on the Effects of Estimator Selection in Robust Parameter Design under Asymmetric Conditions", *Quality and Reliability Engineering International*, [in press], DOI: 10.1002/qre.1406.

Boylan, G.L., Cho, B.R. (2012c), "Robust Design in Ballistic Armor Applications: Estimation under Asymmetric Conditions", *Proceedings of the Annual IIE Systems Engineering Research Conference*, 19-23 May 2012.

Boylan, G.L., Cho, B.R. (2013a), "Comparative Studies on the High-Variability Embedded Robust Parameter Design from the Perspective of Estimators", *Computers & Industrial Engineering*, Vol. 64, No. 1, pp. 442-452.

Boylan, G.L., Cho, B.R. (2013b), "Robust Parameter Design in asymmetric and highly variable processes: simulation-based investigations of alternative approaches to regression estimation", *Journal of Statistical Computation and Simulation*, [in review].

Boylan, G.L., Cho, B.R., (2013c) "Robust Parameter Design in Embedded High-Variability Production Processes: An Alternative Approach to Mitigating Sources of Variability", *International Journal of Production Research* [in press].

Boylan, G.L., Goethals, P.L., Cho, B.R. (2013), "Robust Parameter Design in Resource-Constrained Environments: An investigation of Trade-offs between Costs and Precision within Variable Processes", *Applied Mathematical Modeling*, Vol. 37, pp. 2394-2416.

Breslow, N.E., Clayton, D.G. (1993), "Approximate inference in generalized linear mixed models", *Journal of the American Statistical Association*, Vol. 88, pp. 9-25.

Browne, M.W. (1984), "Asymptotically distribution-free methods for the analysis of covariance structures", *British Journal of Mathematical and Statistical Society*, Vol. 37, pp. 62-83.

Brownlee, K.A. (1960). *Statistical Theory and Methodology in Science and Engineering*. Wiley: New York, pp. 491-500.

Byrne, D.M., Taguchi, S. (1987) "The Taguchi approach to parameter design", *Quality Progress*, Vol. 20, No. 12, pp. 19-26.

Ch'ng, C.K., Quah, S.H., Low, H.C. (2005), "The MM-Estimator in response surface methodology," *Quality Engineering*, Vol. 17, No. 4, pp. 561-565.

Chan, W.M., Ibrahim, R.N. (2004) 'Evaluating the quality level of a product with multiple quality characteristics', *International Journal of Advanced Manufacturing Technology*, Vol. 24, Nos. 9-10, pp. 738-742.

Chan, W.M., Ibrahim, R.N., and Lochert, P.B. (2004) "Evaluating the product quality level under multiple L-type quality characteristics", *International Journal of Advanced Manufacturing Technology*, Vol. 27, Nos. 1-2, pp. 90-95.

Chan, W.M., Ibrahim, R.N., and Lochert, P.B. (2005) "Quality evaluation model using loss function for multiple S-type quality characteristics", *International Journal of Advanced Manufacturing Technology*, Vol. 26, Nos. 1-2, pp. 98-101.

Charnes, A., Cooper, W.W. (1961). *Management models and industrial applications of linear programmin*g. Wiley, New York, NY.

Chase, K.W., Parkinson, A.R. (1991) "A survey of research in the application of tolerance analysis to the design of mechanical assemblies", *Engineering Design*, Vol. 3, No. 1, pp. 23-27.

Chen, C.H. (2004) "Determining the optimum process mean based on asymmetric quality loss function and rectifying inspection plan", *IEEE: Industrial Engineering Management Conference*, pp. 1080-1084.

Chen, C.H., Chou, C.Y. (2003) "Determining the optimum process mean under the bivariate quality characteristics", *International Journal of Advanced Manufacturing Technology*, Vol. 21, No. 5, pp. 313-316.

Chen, C.H., Huang, K.W. (2011) "The determination of the optimal process mean for a one-sided specification limit product with manufacturing cost and linear quality loss". *Journal of Quality*, Vol. 18, No. 1, pp. 19-33.

Chen, S.L., Chou, C.Y., Huang, K.W. (2002) "Determining the optimum process mean for a poor process", *International Journal of Advanced Manufacturing Technology*, Vol. 20, No. 10, pp. 747-757.

Chen, W., Sahai, A., Messac, A., Sundararaj, G. (2000) "Exploration of the effectiveness of physical programming in robust design," *Journal of Mechanical Design*, Vol. 122, No. 6, pp. 155-163.

Chen, W., Wiecek, M.M., Zhang, J. (1999), "Quality Utility – A Compromise programming approach to robust design," *Journal of Mechanical Design*, Vol. 121, No. 6, 179-187.

Chen, Y., Ye,, K. (2009), "Bayesian hierarchical modeling on dual response surfaces in partially replicated designs", *Quality Technology and Quantitative Management*, Vol. 6, No. 4, pp. 371-389.

Cho, B.R. (1994), "Optimization issues in quality engineering", Ph.D. thesis, School of Industrial Engineering, University of Oklahoma.

Cho, B.R., Leonard, M.S. (1997) "Identification and extension of quasi-convex quality loss functions", *International Journal of Reliability, Quality, and Safety Engineering*, Vol. 4, No. 2, pp. 191-204.

Cho, B.R., Park, C. (2005), "Robust design modeling and optimization with unbalanced data", *Computers and Industrial Engineering*, Vol. 48, pp. 173-180.

Choi, S.H., Buckley, J.J. (2007), "Fuzzy regression using least absolute deviation estimators", *Statistics and Computing*, Vol. 12, pp. 257-263.

Clayton, E., Weber, W., Taylor III, B. (1982), “A goal programming approach to the optimization of multiresponse simulation models”, *IIE Transactions*, Vol. 14, 282-287.

Committee on Testing of Body Armor Materials for Use by the U.S. Army – Phase III, Board on Army Science and Technology, Division on Engineering and Physical Sciences, National Research Council (2012), Testing of Body Armor Materials: Phase III. National Academies Press; url=http://www.nap.edu/openbook.php?record_id=13390.

Copeland, K.A., Nelson, P.R. (1996), “Dual Response Optimization via Direct Function Minimization”, *Journal of Quality Technology*, Vol. 28, No. 3, pp. 331-336.

Costa, N.R.P. (2010), “Simultaneous Optimization of Mean and Standard Deviation”, *Quality Engineering*, Vol. 22, pp. 140-149.

Counsell, N., Cortina-Borja, M., Lehtonen, A., Stein, A. (2011), “Modeling psychiatric measures using skew-normal distributions”, *European Psychiatry*, Vol. 26, No. 2, pp. 112-114.

Das, M.N. (1963) "On the Construction of Second Order Rotatable Designs through Balanced Incomplete Block Designs with Blocks of Unequal Size." *Calcutta Statistical Association Bulletin*, Vol. 12, pp. 31-46.

Das, R.N., Lee, Y. (2010), “Analysis Strategies for Multiple Responses in Quality Improvement Experiments”, *International Journal of Quality Engineering Technology*, Vol. 1, No. 4, pp. 395-409.

Dasgupta, K., Sen, D., Mazumder, S., Basak, C.B., Joshi, J.B. and Banerjee, S. (2010), “Optimization of parameters by Taguchi method for controlling purity of carbon nanotubes in chemical vapour deposition technique”, *Journal of Nanoscience and Nanotechnology*, Vol. 10, No. 6, pp. 4030-7.

DeCarlo L. (1997), “On the meaning and use of kurtosis”, *Psychological Methods*; Vol. 2, pp. 292-307.

DeGroot, M. (1970), *Optimal Statistical Decisions.* McGraw-Hill, New York, NY.

DeGroot, M.H., Schervish, M.J. (2002), *Probability and Statistics (3rd ed.)*, Addison Wesley, Boston, MA.

Dehnad, K. (1989), *Quality Control, Robust Design, and the Taguchi Method*. Wadsworth and Brooks/Cole, Pacific Grove, CA.

Del Castillo, E., Fan, S.-K., Semple, J. (1997) “The Computation of Global Optima in Dual Response Systems”, *Journal of Quality Technology*, Vol. 29, No. 3, pp. 347-353.

Del Castillo, E., Montgomery, D.C. (1993), “A Nonlinear Programming Solution to the Dual Response Problem”, *Journal of Quality Technology*, Vol. 25, No. 3, pp. 199-204.

Ding, R., Lin, D.K.J., and Wei, D. (2004). “Dual response surface optimization: a weighted MSE approach”, *Quality Engineering*, Vol. 16, No. 3, pp. 377-385.

Dobrzanski, L.A., Domagala, J., Silva, J.F. (2007), “Application of Taguchi method in the optimisation of filament winding of thermoplastic composites”, *Archives of Materials Science and Engineering*, Vol. 28, No. 3, pp. 133-140.

Elsayed, E.A., Chen, A. (1993) “Optimal levels of process parameters for products with multiple quality characteristics”, *International Journal of Production Research*, Vol. 31, No. 5, pp. 1117-1132.

Fan, S.-K. (2000) “A Generalized Global Optimization Algorithm for Dual Response Systems”, edited by J.R. Simpson, *Journal of Quality Technology*, Vol. 32, No. 4, pp. 444-456.

Fang, H.F., Wu, Z.T. (2000) “Concurrent tolerance design and methods of technology economy assessment in process route”, *Chinese Journal of Mechanical Engineering*, Vol. 36, No. 4, pp. 74-77.

Feng, Q., Kapur, K.C. (2009) “Selection of optimal precision levels and specifications considering measurement error”, *International Journal of Advanced Manufacturing Technology*, Vol. 40, Nos. 9-10, pp. 960-970.

Fernandez, M.M. (2004), “Fundamentals on Estimation Theory,” Unpublished course notes obtained at http://lmi.bwh.harvard.edu/papers/pdfs/2004/martin-fernandezCOURSE04b.pdf, accessed 29 July, 2011.

Filliben J. (1975), “The probability plot correlation coefficient test for normality”, *Technometrics*, Vol. 17, pp. 111-117.

Fisher, R.A. (1922), “On the Mathematical Foundations of Theoretical Statistics", *Philosophical Transactions of the Royal Society*, *Series A*, Vol. 222, pp. 309-368.

Goethals, P.J., Aragon, L., Cho, B.R. (2009), “Experimental investigations of estimated response surface functions with different variability measures”, *International Journal of Experimental Design and Process Optimisation*, Vol. 1, No. 2-3, pp. 123-163.

Goethals, P.J., Boylan, G.L. (2011), “Analysing the effects of variability measure selection on process and product optimization”, *International Journal of Quality Engineering and Technology*, Vol. 2, No. 3, pp. 254-276.

Goethals, P.L., Cho, B.R. (2011a), “Solving the optimal process target problem using response durface designs in heteroscedastic conditions,” *International Journal of Production Research*, Vol. 49, No. 12, pp. 3455-3478.

Goethals, P.L., Cho, B.R. (2011b), “The development of a robust design methodology for time-oriented dynamic quality characteristics with a target profile,” *Quality and Reliability Engineering International*, Vol. 27, No. 4, pp. 403-414.

Goethals, P.L., and Cho, B.R. (2011c), "Using Higher Precision-Based Response Surface Designs to Determine the Optimal Process Target," *International Journal of Advanced Manufacturing Technology*, Vol. 56, pp. 13-30.

Goethals, P.L., and Cho, B.R. (2011d), "The development of quality models for military protective armor manufacturing," Proceedings of the 2011 Industrial Engineering Research Conference, May 21-25, Reno, NV.

Goethals, P.L., and Cho, B.R. (2012), "Designing the optimal process mean vector for mixed multiple quality characteristics," *IIE Transactions*, Vol. 44, pp. 1002-1021.

Golan, A., Judge, G.G., Miller, D. (1996). *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. Wiley, New York, NY.

Golhar, D.Y., Pollock, S.M. (1988) “Determination of the best mean and the upper limit for a canning problem”, *Journal of Quality Technology*, Vol. 20, No. 3, pp. 188-192.

Golhar, D.Y., Pollock, S.M. (1992) “Cost saving due to variance reduction in a canning problem”, *IIE Transactions*, Vol. 24, No. 1, pp. 89-92.

Gruber, M.H.J. (1990), *Regression Estimators: A comparative study*. Academic Press, Boston, MA.

Hampel, F. R. (1978). “Optimally bounding the gross-error sensitivity and the influence of position in factor space”, *Proceedings of the Statistical Computing Section of the American Statistical Association*, Washington, DC, 59–64.

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

Hardin, R. H., Sloane, N. J. A. (1993) “A new approach to the construction of optimal designs”, *Journal of Statistical Planning and Inference*, Vol. 37, pp. 339-369.

Harrington, E. (1965) "The Desirability Function", *Industrial Quality Control*, Vol. 21, No. 10, pp. 494-498.

Hasenkamp, T., Arvidsson, M., Gremyr, I. (2009), "A review of practices for robust design methodology", *Journal of Engineering Design*, Vol. 20, No. 6, pp. 645-657.

Hayter, A.J. (2006). *Probability and Statistics for Engineers and Scientists (3rd edn)*. Duxbury Press, Australia.

Hazen A. (1914), "Storage to be provided in the impounding reservoirs for municipal water supply", *Transactions of the American Society of Civil Engineers*; Vol. 77, pp. 1547-1550.

Hill, I.D., Hill, R., Holder, R.L. (1976), "Fitting Johnson curves by moments", *Applied Statistics*, Vol. 25, No. 2, pp. 180-184.

Hill, R. W. (1977). "Robust Regression When There Are Outliers in the Carriers". Unpublished Ph.D. dissertation, Harvard University, Boston, MA.

Hoaglin, D.C., Mosteller, F., Tukey, J.C. *Understanding Robust and Exploratory Data Analysis*. Wiley, New York, NY.

Hodges, J.L., Lehmann, E.L. (1963), "Estimates of location based on rank tests", *Annals of Mathematical Statistics*, Vol. 34, pp. 598-611.

Holmstrom, K., Petersson, J. (2002), "A review of the parameter estimation problem of fitting positive exponential sums to empirical data", *Applied Mathematics and Computation*, Vol. 126, pp. 31-61.

Huber, P. J. (1964), "Robust Estimation of a Location Parameter." *Annals of Mathematical Statistics*, Vol. 35, pp.73-101.

Huber, P. J. (2009). *Robust Statistics (2nd ed.)*. Wiley, New York, NY.

Huber, P.J. (1973), "Robust regression: asymptotics, conjectures, and Monte Carlo", *Annals of Statistics*, Vol. 1, No. 5, p. 799-821.

Hunter, W.C., Kartha, C.P. (1977) "Determining the most profitable target value for a production process", *Journal of Quality Technology*, Vol. 9, No. 4, pp. 176-181.

Jeong, I.-J., Kim, K.-J., and Chang, S.T. (2005), "Optimal weighting of bias and variance in dual response surface optimization, *Journal of Quality Technology*, Vol. 37, No. 3, pp. 236-247.

Johnson R, Wichern D. (1982), *Applied Multivariate Statistical Analysis*. Prentice-Hall, Englewood Cliffs, NJ.

Kapur, K.C., Cho, B.R. (1996), "Economic design of the specification region for multiple quality characteristics", *IIE Transactions*, Vol. 28, No. 3, pp. 237-248.

Karsak, E.E., Sozer, S., Alptekin, S.E. (2002), "Product planning in quality function deployment using a combined analytic network process and goal programming approach", *Computers and Industrial Engineering*, Vol. 44, pp. 171-190.

Karst, O.J. (1958), "Linear Curve Fitting Using Least Deviations," *Journal of the American Statistical Association*, Vol. 53, pp. 118-32.

Kazemzadeh, R.B., Bashiri, M., Atkinson, A.C., and Noorossana, R. (2008), "A General Framework for Multiresponse Optimization Problems Based on Goal Programming", *European Journal of Operational Research*, Vol. 189, pp. 421-429.

Kiefer, J. (1959), "Optimum Experimental Designs", *Journal of the Royal Statistical Society, Series B (Methodologies)*, Vol. 21, pp. 272-319.

Kiefer, J. (1961), "Optimum Designs in Regression Problems, II", *Annals of Mathematical Statistics*, Vol. 32, No. 1, pp. 298-325.

Kiefer, J., Wolfowitz, J. (1959), "Optimum Designs in Regression Problems", *Annals of Mathematical Statistics*, Vol. 30, No. 1, pp. 271-294.

Kim, K.-J., Lin, D. (2006), "Optimization of multiple responses considering both location and dispersion effects", *European Journal of Operations Research*, Vol. 169, No. 1, pp. 133-145.

Kim, K.-J., Lin, D.K.J. (1998), "Dual response surface optimization: a fuzzy modeling approach", *Journal of Quality Technology*, Vol. 30, No. 1, pp. 1-10.

Kim, Y.J., Cho, B.R. (2002), "Development of priority-based robust design", *Quality Engineering*, Vol, 14, No. 3, pp. 355-363.

Kim, Y.J., Cho, B.R., Phillips, M.D. (2000) "Determination of the optimal process means with the consideration of variance reduction and process capability", *Quality Engineering*, Vol. 13, No. 3, pp. 251-260.

Koksoy, O., Yalcinoz, T. (2008), "Robust Design Using Pareto Type Optimization: A Genetic Algorithm with Arithmetic Crossover", *Computers and Industrial Engineering*, Vol. 55, No. 1, pp. 208-218.

Koutrouvelis, I.A., Canavos G.C. (2000), “A comparison of moment-based methods of estimation for the log Pearson type 3 distribution”, *Journal of Hydrology*, Vol. 234, pp. 71-81.

Kovach, J., Cho, B.R. (2007a), “Constrained Robust Design Experiments and Optimization with the Consideration of Uncontrollable Factors”, *International Journal of Advance Manufacturing Technology*, Vol. 38, pp. 7-18.

Kovach, J., Cho, B.R. (2007b), “Development of a *D*-Optimal Robust Design Model for Restricted Experiments”, *International Journal of Industrial Engineering*, Vol. 14, No. 2, pp. 117-128.

Kovach, J., Cho, B.R. (2008), “Development of a Multidisciplinary-Multiresponse Robust Design Optimization Model”, *Engineering Optimization*, Vol. 40, No. 9, pp. 805-819.

Kovach, J., Cho, B.R., and Antony, J. (2009), “Development of a Variance Prioritized Multiresponse Robust Design Optimization Framework”, *International Journal of Quality and Reliability Management*, Vol. 26, No. 4, pp. 380-396.

Krasker,W. S. (1980), “Estimation in Linear Regression Models with Disparate Data Points,” *Econometrica*, Vol. 48, pp. 1333–1346.

Kuk, A.Y.C., Cheng, Y.W. (1999), “Pointwise and functional approximation in Monte Carlo maximum likelihood estimation”, *Statistics and Computing*, Vol. 9, pp. 91-99.

Kummar, J., Shunmugam, M.S. (2006), “Fitting of robust reference surface based on least absolute deviations”, Precision Engineering, Vol. 31, pp. 102-113.

Kutner M, Nachtsheim C, Neter J, Li W. (2005), *Applied Linear Statistical Models* (5th edn). McGraw-Hill Irwin, Boston, MA.

Lee, S.B., Park, C., Cho, B.R. (2007), “Development of a highly efficient and resistant robust design”, *International Journal of Production Research*, Vol. 45, No. 1, 157-167.

Lee, Y., Nelder, J.A. (1996), “Hierarchical generalized linear models (with discussion)”, *Journal of the Royal Statistical Society*, Vol. B58, pp. 619-656.

Lee, Y., Nelder, J.A. (2003) Robust design via generalized linear models, *Journal of Quality Technology*, Vol. 35, No. 1, pp. 2-12.

Lee, Y., Nelder, J.A. (2006), “Double hierarchical generalized mixed linear models (with discussion)”, *Applied Statistics*, Vol. 55, pp. 139-167.

Lee, Y., Nelder, J.A. (2007), "H-likelihood: problems and solutions", *Statistics and Computing*, Vol. 17, No. 1, pp. 49-55.

Lee, Y., Nelder, J.A., Park, H. (2011), "HGLMs for quality improvement", *Applied Stochastic Models in Business and Industry*, Vol. 27, No. 3, pp. 315-328.

Legendre, A. (1805), *Nouvelles m´ethodes pour la d´etermination des orbites des com`etes.*

Li, Y., Arce, G.R. (2004), "A maximum likelihood approach to least absolute deviation regression", *EURASIP Journal on Applied Signal Processing*, Vol. 12, pp. 1762-1769.

Liao, M.-Y. (2010) 'Economic tolerance design for folded normal data', *International Journal of Production Research*, Vol. 48, No. 14, pp. 4123-4137.

Lim, K. (2007), "Maximum likelihood estimation for linear regression models involving missing covariate observations", unpublished honors thesis, accessed at http://www.ms.unimelb.edu.au/publications/LimKen.pdf on 28 Jul 2011.

Lin, D.K.J., Tu, W. (1995), "Dual Response Surface Optimization", *Journal of Quality Technology*, Vol. 27, No. 1, pp. 34-39.

Looney S., Gulledge Jr., T. (1985), "Use of the correlation coefficient with normal probability plots", *The American Statistician*, Vol. 39, pp. 75-79.

Lumley, T. (2009), leaps. *Leaps*. The R project for statistical computing, http://www.r-project.org/.

Luo X, Mayer M, Heck B. (2010), "On the probability distribution of GNSS carrier phase observations", *GPS Solutions*, 1-11. Doi: 10.1007/s10291-010-0196-2.

Mallows, C. L. (1975). "On Some Topics in Robustness." Unpublished memorandum, Bell Telephone Laboratories, Murrray Hill, NJ.

Marazzi, A. (1993), *Algorithms, Routines, and S Functions for Robust Statistics*. Wadsworth & Brooks-Cole, Pacific Grove, CA.

Maronna, R.A. (2011), "Robust ridge regression for high-dimensional data", *Technometrics*, Vol. 53, No. 1, pp. 44-53.

Maronna, R.A., Zamar, R.H. (2002), "Robust estimates of location and dispersion for high-dimensional datasets", *Technometrics*, Vol. 44, pp. 307-317.

Messac, A. (1996), “Physical programming: effective optimization for computational design”, *American Institute of Aeronautics and Astronautics Journal*, Vol. 34, No. 1, pp. 149-158.

Michael, W., and Siddall, J.N. (1981) “The optimization problem with optimal tolerance assignment with full acceptance”, *Journal of Mechanical Design*, Vol. 103, No. 4, pp. 842-848.

Montgomery, D.C., Loredo E.N., Jearkpaporn D., Testik M.C. (2002) Experimental Designs for Constrained Regions. *Quality Engineering*, Vol. 14, pp. 587-601.

Moorhead, P.R., Wu, C.F.J. (1998) “Cost driven parameter design”, *Technometrics*, Vol. 40, No. 2, pp. 111-119.

Motley, B. (2005), “Introduction to Variability and variation reduction”, *Defense Acquisition, Technology, & Logistics*, May-June, pp. 53-55.

Muhlbauer, A., Spichtinger, P., Lohmann, U. (2009), “Application and comparison of robust linear regression methods for trend estimation”, *Journal of Applied Meteorology and Climatology*, Vol. 48, pp. 1961-1970.

Myers, R.H., Carter, W.H. (1973) "Response Surface Techniques for Dual Response Systems." *Technometrics* 15(2): 301-317.

Myers, R.H., Montgomery, D.C., and Vining, G.G. (2002) *Generalized Linear Models: With Applications in Engineering and the Sciences*, Wiley, New York, NY.

Myers, R.H., Montgomery, D.C. (1997) “A tutorial on generalized linear models”, *Journal of Quality Technology*, Vol. 29, pp. 274-291.

Myung, I.J. (2003), “Tutorial on maximum likelihood estimation”, *Journal of Mathematical Psychology*, Vol. 47, pp. 90-100.

Nair, V.N. (1992), “Taguchi’s Parameter Design: A Panel Discussion”, *Technometrics*, Vol. 34, pp. 127-161.

Nelder, J.A., Wedderburn, R.W.M. (1972), “Generalized linear models”, Journal of the Royal Statistical Society, Vol. 135, No. 3, pp. 370-384.

Nelson W. (1982), *Applied Life Data Analysis*. Wiley, New York, NY.

Noh, M., Lee, Y. (2007), “Restricted maximum likelihood estimation for binary data in generalised linear mixed models”, *Journal of Multivariate Analysis*, Vol. 98, pp. 896-915.

O'Hagan, X., Leonard (1976), "Bayes estimation subject to uncertainty about parameter constraints", *Biometrika*, Vol. 63, pp. 201-202.

Park, C., Cho, B.R. (2003), "Development of robust design under contaminated and non-normal data", *Quality Engineering*, Vol. 15, No. 3, pp. 463-469.

Park, G.-J., Lee, T.-H., Lee, K.H., Hwang, K.-H. (2006), "Robust Design: an overview", *American Institute of Aeronautics and Astronautics Journal*, Vol. 44, 181-191.

Peng, H.P., Jiang, X.Q. and Liu, X.J. (2008), "Concurrent optimal allocation of design and process tolerances for mechanical assemblies with interrelated dimension chains", *International Journal of Production Research*, Vol. 24, No. 15, pp. 6963-6979.

Pignatiello, J.J. (1993), "Strategies for robust multiresponse quality engineering", *IIE Transactions*, Vol. 25, pp. 5-25.

Pisarenko, V.F., Sornette, D. (2004), "Statistical methods of parameter estimation for deterministically chaotic time series", *Physical Review E*, Vol. 69, pp. 1-12.

Pukelsheim, F. (1995), *Optimum Design of Experiments*. Chapman & Hall, New York, NY.

Pyzdek, T. (1995), "Why normal distributions aren't [all that normal]", *Quality Engineering*, Vol. 7, No. 4, pp. 419-440.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rahim, M.A., Shaibu, A.B. (2000), "Economic selection of optimal target value", *Process Control Quality*, Vol. 11, No. 5, pp. 369-381.

Ranneby, B. (1984), "The maximum spacing method: an estimation method related to the maximum likelihood method", *Scandinavian Journal of Statistics*, Vol. 11, No. 2, pp. 93–112.

Ripley, B.D. (2004), "Robust Statistics", published course notes, accessed via www.stats.ox.ac.uk/pub/StatMeth/Robust.pdf on 28 July, 2011.

Robinson, T.J., Borror, C.M., Myers, R.H. (2004), "Robust parameter design: A review", *Quality and Reliability Engineering International*, Vol. 20, pp. 81-101.

Robinson, T.J., Wulff, S.S., Montgomery, D.C., Khuri, A.I. (2006), "Robust parameter design using generalized linear mixed models", *Journal of Quality Technology*, Vol. 38, No. 1, pp. 65-75.

Rocke, D.M. (1996), "Robustness properties of S-estimators of multivariate location and shape in high dimension," *The Annals of Statistics*, Vol. 24, No. 3, pp. 1327-1345.

Rousseeuw, P.J. (1984), "Least median squares regression", *Journal of the American Statistical Association*, Vol. 79, No. 388, pp. 871-880.

Rousseeuw, P.J., Croux, C. (1993) Alternatives to the median absolute deviation, *Journal of the American Statistical Association*, Vol. 88, pp. 1273-1283.

Rousseeuw, P.J., Leroy, A.M. (1987), *Robust Regression and Outlier Detection*. Wiley, New York, NY.

Rousseeuw, P.J., Yohai, V. (1984), "Robust regression by means of S-estimators." *Robust and Nonlinear Time Series Analysis*, edited by J. Franke, W. Hardle, and D. Martin, *Lecture Notes in Statistics No. 26*, pp. 256-272. Springer-Verlag, Berlin.

Ryan, T.P. (2009), *Modern Regression Methods*, Wiley, Hoboken, NJ.

Schlossmacher, E.J. (1973), "An iterative technique for absolute deviations curve fitting", *Journal of the American Statistical Association*, Vol. 68, No. 344, pp. 857-859.

Schmidt, R.L., Pfeifer, P.E., 1989. An economic evaluation of improvements in process capability for a single-level canning problem. *Journal of Quality Technology*, Vol. 21, pp. 16-19.

Serfling, R. (2011), "Asymptotic relative efficiency in estimation", to be published in the *International Encyclopedia of Statistical Sciences (1st ed.)*, Springer, Berlin.

Shaibu, A.B., Cho, B.R.(2009), "Another view of dual response surface modeling and optimization in robust parameter design", *International Journal of Advanced Manufacturing Technology*, Vol. 41, Nos. 7-8, pp. 631-641.

Shaibu, A.B., Cho, B.R., and Kovach, J., (2009), "Development of Censored Robust Design for Time-Oriented Quality Characteristics", *Quality and Reliability Engineering International*, Vol. 25, pp. 181-197.

Shao, J., Tu, D. (1995), *The Jackknife and Bootstrap*. Springer-Verlag, New York, NY.

Shapiro S, Brain C. (1981), "A review of distributional testing procedures and development of a censored sample distributional test", in *Statistical Distributions in Scientific Work*, eds. C. Taillie, G.P. Patil, and B.A. Baldessari, D. Reidel, Holland, 1-24.

Shapiro, S, Francia R. (1972), "An approximate analysis of variance test for normality. *Journal of the American Statistical Association*; Vol. 67, pp. 215-216.

Shapiro, S, Wilk M. (1965), "An analysis of variance test for normality (complete samples)", *Biometrika*; Vol. 52, pp. 591-611.

Sheskin, D.J. (2007), *Handbook of Parametric and Nonparametric Statistical Procedures (4th ed),* Chapman & Hall, Boca Raton, FL.

Shin, S., Cho, B.R. (2007), "Integrating a bi-objective paradigm to tolerance optimization", *International Journal of Production Research*, Vol. 45, No. 23, pp. 5509-5525.

Shin, S., Cho, B.R. (2008), "Development of a sequential optimization procedure for robust design and tolerance design within a bi-objective paradigm", *Engineering Optimization*, Vol. 40, No. 11, pp. 989-1009.

Shin, S., Cho, B.R. (2009), "Studies on a bi-objective robust design optimization problem", *IIE Transactions*, Vol. 41, pp. 957-968.

Shin, S., Govindaluri, M.S. and Cho, B.R. (2005), "Integrating the Lambert W function to a tolerance optimization problem", *Quality and Reliability Engineering International*, Vol. 21, No. 8, pp. 795-808.

Shin, S., Samanlioglu, F., Cho, B.R., Wiecek, M.M. (2011), "Computing Trade-offs in Robust Design: Perspectives of the Mean Squared Error", *Computers and Industrial Engineering*, Vol. 60, pp. 248-255.

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

Simpson, J. R., Montgomery, D. C. (1998), "A performance-based assessment of robust regression methods. *Communications in Statistics-Simulation and Computation*, Vol. 27, No. 4, pp. 1031–1049.

Smith, K. (1918) "On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations", *Biometrika, Vol.* 12, pp. 1-85.

Snedecor G, Cochran W. (1980), *Statistical Methods (7th ed).* Iowa State University Press, Ames, IA.

Spiring, F.A. (1994), "The reflected normal loss function", *La Revue Canadienne de Statistique*, Vol. 21, No. 3, pp. 321-330.

Springer, C.H. (1951), "A method for determining the most economic position of a process mean", *Industrial Quality Control*, Vol. 8, No. 1, pp. 36-39.

St. John, R.C., Draper, N.R. (1975) "D-Optimality for Regression Designs: A Review." *Technometrics*, Vol. 17, pp. 15-23.

Steuer, R.E. (1986), *Multiple Criteria Optimization: Theory, Computation, and Application*. Wiley, New York, NY.

Sundaram, R.M. (1978), "An application of a goal programming technique in metal cutting", *International Journal of Production Research*, Vol. 16, pp. 375-382.

Tabucanon, M. (19880, *Multiple Criteria Decision Making in Industry*. Elsevier, Amsterdam.

Taguchi, G. (1984), "Quality evaluation for quality assurance", *American Supplier Institute*, Romulus, MI.

Taguchi, G. (1986), *Introduction to Quality Engineering: Designing Quality into Products and Processes*. Kraus, White Plains, NY.

Taguchi, G. (1987), *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost*, UNIPUB/Kraus International, White Plains NY.

Tahera, K., Ibrahim, R.N., and Lochert, P.B. (2008) "A Fuzzy Logic Approach for Dealing with Qualitative Quality Characteristics of a Process." *Expert Systems with Applications* 34(4): 2630-2638.

Tang, L.C., Xu, K. (2002), "A Unified Approach for Dual Response Surface Optimization", *Journal of Quality Technology*, Vol. 34, No. 4, pp. 437-447.

Teeravaraprug, J., Cho, B.R. (2002), "Designing the optimal process target levels for multiple quality characteristics", *International Journal of Production Research*, Vol. 40, No. 1, pp. 37-54.

Truong, N.K.V., Shin, S. (2012), "Development of a new robust design methodology based on Bayesian perspectives", *International Journal of Quality Engineering and Technology*, Vol. 3, No. 1, pp. 50-78.

Tsai, R.-C., Bockenholt, U. (2001), "Maximum likelihood estimation of factor and ideal point models ofr paired comparison data", *Journal of Mathematical Psychology*, Vol. 45, pp. 795-811.

Tsui, K.L. (1992), "An overview of Taguchi method and newly developed statistical methods for robust design", *IIE Transactions*; Vol. 24, pp. 44-57.

Tsui, K.L. (1999), "Robust design optimization for multiple characteristic problems", *International Journal of Production Research*, Vol. 37, No. 2, pp. 433-445.

Vannman, K., Albing, M. (2007) 'Process capability indices for one-sided specification intervals and skewed distributions', *Quality and Reliability Engineering International*, Vol. 23, No. 6, pp. 755-765.

Vining G. (2010), "Technical Advice: Quantile plots to check assumptions", *Quality Engineering*; Vol. 2, pp. 364-367.

Vining, G.G., Myers, R.H. (1990), "Combining Taguchi and response surface philosophies: A Dual Response Approach", *Journal of Quality Technology*, Vol. 22, No.1, pp.38-45.

Waddington, D., Thompson, R. (2004), "Using a correlated probit model approximation to estimate the variance for binary matched pairs", *Statistics and Computing*, Vol. 14, pp. 83-90.

Weibull W. (1939), "The phenomenon of rupture in solids", *Ingeniors Vetenskaps Akademien Handingar*; Vol. 153, p.17.

Weins, D.P., Wu, E.K.H. (2010), "A comparative study of robust designs for M-estimated regression models", *Computational Statistics and Data Analysis*, Vol. 54, pp. 1683-1695.

Wheeler, R.E. (2004). optFederov. *AlgDesign*. The R project for statistical computing, http://www.r-project.org/.

Wilk M, Gnanadesikan R. (1968), "Probability plotting methods for the analysis of data", *Biometrika*; Vol. 55, pp. 1-19.

Willems, G., Van Aelst, S. (2005), "Fast and robust bootstrap for LTS", *Computational Statistics and Data Analysis*, Vol. 48, No. 4, pp. 703-715.

Willinger, W., Alderson, D., Lun, L. A pragmatic approach to dealing with high-variability in network measurements, *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*, 25-27 October, 2004, Taormina, Sicily, Italy.

Wu, L.L. (1985). "Robust M-estimation of location and regression", *Sociological Methodology*, Vol. 15, pp. 316-388.

Yang D, Qi H. (2010), "An effective nonparametric quickest detection procedure based on Q-Q distance", *IEEE Int Conf on Acoustics, Speech, and Signal Processing – Proceedings*, 3786-3789.

Yohai, V.J. (1987) "High breakdown-point and high efficiency robust estimates for regression," *The Annals of Statistics*, Vol. 15, No. 2, pp. 642-656.

Yu, P.L. (1973), “A class of solutions for group decision problems”, *Management Science*, Vol. 19, pp. 936-946.

Zeleny, M. (1973), *Compromise Programming* in: *Multiple Criteria Decision Making*. eds. J.L. Cochrane and M. Zeleny. University of South Carolina Press, Columbia, SC, pp. 262-301.