Leveraged Plans for Measurement System Assessment

by

Ryan Browne

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2009

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In manufacturing, measurement systems are used to control processes and inspect parts with the goal of producing high quality product for the customer. Modern Quality Systems require the periodic assessment of key measurement systems to ensure that they are functioning as expected. Estimating the proportion of the process variation due to the measurement system is an important part of these assessments.

The measurement system may be simple, for example, with one gauge automatically measuring a single characteristic on every part or complex with multiple characteristics, gauges, operators etc. Traditional assessment plans involve selecting a random sample of parts and then repeatedly measuring each part under a variety of conditions that depend on the complexity of the measurement system.

In this thesis, we propose new plans for assessing the measurement system variation based on the concept of leveraging. In a leveraged plan, we select parts (non-randomly) with extreme initial values to measure repeatedly. Depending on the context, parts with initial measurements may be available from regular production or from a specially conducted baseline study. We use the term leveraging because of the re-use of parts with extreme values.

The term leverage has been used by the proponents of the problem solving system initially proposed by Dorian Shainin. Parts with relatively large and small values of the response are compared to identify the major causes of the variation. There is no discussion of the theory of leveraging in the literature or its application to measurement system assessment. In this thesis, we provide motivation for why leveraging is valuable and apply it to measurement system assessments.

We consider three common contexts in the thesis:

- Simple measurement systems with one gauge, no operator effects and no external information about the process performance;
- Measurement systems, as stated above, where we have external information, as would be the case, for example, if the measurement system was used for 100% inspection;
- Measurement systems with multiple operators.

For each of these contexts, we develop new leveraged assessment plans and show that these plans are substantially more efficient than traditional plans in estimating the proportion of the process variation due to the measurement system. In each case, we also provide methodology for planning the leveraged study and for analysing the data generated.

We then develop another new application of leveraging in the assessment of a measurement system used for 100% inspection. A common practice is to re-measure all parts with a first measurement outside of inspection limits. We propose using these repeated measurements to assess the variation in the measurement system. Here the system itself does the leveraging since we have repeated measurements only on relatively large or small parts. We recommend using maximum likelihood estimation but we show that the ANOVA estimator, although biased, is comparable to the MLE when the measurement system is reliable. We also provide guidelines on how to schedule such assessments.

To outline the thesis, in the first two chapters, we review the contexts described above. For each context, we discuss how to characterize the measurement system performance, the common assessment plans and their analysis. In Chapter 3, we introduce the concept of leveraging and provide motivation for why it is effective. Chapters 4 to 7 contain the bulk of the new results in the thesis. In Chapters 4, 5 and 6, which correspond to the three contexts described above, we provide

new leveraged plans, show their superiority to the standard plans and provide a methodology to help design leveraged plans. In Chapter 7, we show how to assess an inspection system using repeated measurements on initially rejected parts. In the final chapter, we discuss other potential applications of leveraging to other measurement system assessment problems and to a problem in genetics.

Acknowledgements

I would like to express my sincerest gratitude to my supervisors Dr. Jock MacKay and Dr. Stefan Steiner for their guidance and patience throughout my study. Much appreciation to Riley Metzger for his assistance and suggestions in formulating my teaching style. Many thanks to my fellow grad students Oana Danila, Audrey Boruvka, Dagmar Mariaca and Andrew Rae. Also, thanks to my family for their encouragement and support throughout my study. Finally, I would like to say, to my best friend by default, Kim, thanks.

Dedication

To Zoey Tigerlily.

Contents

List of Tables **xvii**

List of Figures **xxiv**

1 Introduction **1**

1.1 A Measurement System with No Operators and a Single Gauge . . 2

1.1.1 Characterizing Measurement System Quality 6

1.1.1.1 Metrics based on Measurement Standard Deviation 7

1.1.1.2 Metrics based on Comparing Measurement Variation to the Specification Limits 9

1.1.1.3 Metrics based on Measurement and Process Variation 11

1.1.1.4 Metrics based on the Process Mean and Variation, Measurement Variation and the Specification Limits 16

1.1.1.5 Numerical Example 21

1.2 A Measurement System with 100% Inspection 23

1.3 A Measurement System with Operators 24

1.3.1 Characterizing Measurement System Quality 27

1.4 Goal and Outline 28

2 Standard Measurement Assessment Plan **31**

2.1 A Measurement System 32

2.1.1 Estimation 33

2.2 A Measurement System with 100% Inspection 37

2.2.1 Analysis 37

2.2.2 The Value of the Baseline Information 41

2.3 A Measurement System with Operators 42

2.3.1 Analysis 43

3 Leveraged Plan and Why it Works **47**

3.1 Leveraging and Regression 48

3.2 Known True Value Versus an Extreme Initial Measurement 52

4 Leveraged Plan for a Measurement System **55**

4.1 Context 55

4.2 Estimation 57

4.2.1 Maximum Likelihood 57

4.2.2 Regression Estimator 63

4.2.3 ANOVA Estimator 65

4.2.4 Combined Estimator 66

4.2.5 Numerical Example of Various Estimates for ρ Based on Leveraged Plan 69

4.2.6 Comparison of the Various Estimators for ρ Based on the Leveraged Plan 73

4.2.7 The Effect of Varying the Number of Repeated Measurements on Each Part 74

4.3 Leveraged Plan Design 77

4.3.1 Optimal Designs 79

4.4 Model Assumptions 83

4.5 The Leveraged versus Standard Plan 85

4.6 Conclusions 94

5 Leveraged Plan for a Measurement System with 100% Inspection 99

5.1 Context 99

5.2 Design 100

5.3 Estimation 100

5.3.1 Maximum Likelihood 101

5.3.2 Regression Estimator 102

5.3.3 ANOVA Estimator 103

5.3.4 Combined Estimator 104

5.3.5 Comparison of Leveraged Plan Analysis Methods 105

5.4 Leveraged Plan Design 106

5.5 Leveraged versus Standard Plan 109

5.6 Conclusions and Discussion 113

6 Leveraged Plan for a Measurement System with Operators 115

6.1 Context 115

6.2 Leveraged Plan A 118

6.2.1 Plan . . . 118
6.2.2 Estimation . . . 119
6.2.2.1 Fisher Information . . . 122
6.2.3 Numerical Example . . . 124
6.2.4 The Effect of Direction in a Leveraged Plan . . . 126
6.2.5 Leveraged Plan A Design . . . 132
6.2.6 Leveraged Plan A versus Standard Plan . . . 135
6.3 Leveraged Plan B . . . 137
6.3.1 Plan B . . . 137
6.3.2 Estimation for Leveraged Plan B . . . 138
6.3.2.1 Fisher information . . . 142
6.3.3 Leveraged Plan B Design . . . 143
6.3.4 Leveraged Plan B versus Standard Plan . . . 144
6.4 Leveraged Plan A versus B . . . 146
6.5 Discussion and Conclusions . . . 147

7 Assessment of an Inspection System with Production Data 149

7.1 Context . . . 149
7.2 Inspection System Analysis . . . 153
7.2.1 ANOVA . . . 154
7.2.2 Regression Estimator . . . 156
7.2.3 Maximum Likelihood . . . 158
7.2.4 Comparison . . . 160

7.3 Model Assessment 161

7.4 Assessment Intervals 162

7.5 Discussion and Conclusions 163

8 Conclusions and Extensions **165**

8.1 Summary 165

8.2 Extensions 166

8.2.1 Nonparametric Measurement System Assessments 167

8.2.2 Genetic Variation Studies 167

8.2.3 Random Effects for Operators 168

8.2.4 Assessing a Bivariate Measurement System 171

Appendices **174**

A Two Stage Leveraged Plan **175**

A.1 Fisher Information 175

B Leveraged Plan with Operators as Fixed Effects **177**

B.1 Direction 177

B.2 Tables with optimal designs for $m = 3$ 183

B.3 Tables with optimal designs for $m = 3$ leveraged plan B 187

C Extended Rank 1 Update **193**

C.1 Rank 1 Update 193

C.2 Properties of Fisher information 194

D Assessment of a Measurement System with Production Data **197**

D.1 Truncated Normal 197

D.2 Normal Moments 198

D.3 Expectation of S_m^2 199

D.4 Variance of S_m^2 200

References **201**

List of Tables

1.1 Cutoffs for Acceptable and Unacceptable Measurement Systems in Terms of ρ 14

1.2 Measures of Customer and Producer Risk 17

1.3 Numerical Example of Metrics for Consumer and Producer Risk (given in terms of %) 23

2.1 Analysis of Variance for Gauge R&R Study 33

2.2 ANOVA for Fixed Effects Model (2.18) 43

2.3 Statistics for Standard Plan with Fixed Effects 44

2.4 Estimators for Standard Plan with Fixed Effects 44

4.1 Baseline Data of 100 Camshaft Journal Diameters 70

4.2 Example of a Stage 2 Sample with 2 Extreme Parts Repeatedly Measured 18 Times Each 71

4.3 Estimates of ρ for the Camshaft Journal Diameters Example 72

4.4 The LP Designs that have the Smallest, Middle and Largest stdev($\tilde{\rho}_c$) when $b + nk = 60$ 78

4.5 The Five Plans with the Lowest stdev($\tilde{\rho}_c$) for Different Values of $N = b + nk$ 80

4.6 Values of N for Estimating ρ with a Particular Standard Deviation when $b = N - 5\lfloor N/10 \rfloor$, $n = 5$ and $k = \lfloor N/10 \rfloor$ 81

4.7 The Asymptotic Optimal Leveraged Designs (b, k, n) for a Given ρ When $N = 30$, 60 and 100 82

4.8 Optimal Leveraged Designs (b, k, n) for Estimating ρ with a Particular Standard Deviation 83

6.1 Example Baseline (Stage 1) 124

6.2 Example repeats (Stage 2) 125

6.3 The LP designs (b,k,n) with the smallest asymptotic standard deviation of η given λ, η when $m = 3$. 133

6.4 The LP B designs (b,k,n) with the smallest asymptotic standard deviation of η given λ, η when $m = 3$. 144

7.1 100 Production Test Observations or 1^{st} Measurements 151

7.2 17 2^{nd} Measurements 152

B.1 The LP designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.05, 0.05, 0.15, 0.25, 0.35, 0.45)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 60$. 183

B.2 The LP designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.55, 0.65, 0.75, 0.85, 0.95)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 60$. 184

B.3 The LP designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.05, 0.05, 0.15, 0.25, 0.35, 0.45)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 90$. 185

B.4 The LP designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.05, 0.05, 0.15, 0.25, 0.35, 0.45)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 90$. 186

B.5 The LP B designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.05, 0.05, 0.15, 0.25, 0.35, 0.45)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 60$. 187

B.6 The LP designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.55, 0.65, 0.75, 0.85, 0.95)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 60$. 188

B.7 The LP B designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.05, 0.05, 0.15, 0.25, 0.35, 0.45)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 90$. 189

B.8 The LP B designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.55, 0.65, 0.75, 0.85, 0.95)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 90$. 190

B.9 The LP B designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.05, 0.05, 0.15, 0.25, 0.35, 0.45)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 150$. 191

B.10 The LP B designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.55, 0.65, 0.75, 0.85, 0.95)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 150$. 192

List of Figures

1.1 Scatterplot of Two Measurements on 100 Parts. Left panel has $\rho =$ 0.2 and the right panel $\rho = 0.9$. 4

1.2 Two Measurement Systems: Zero Bias (left), Positive Bias (right) . 5

1.3 Non-Linearity of Measurement Variability; left panel linearity and right panel non-linearity 5

1.4 Comparison of Three Common Measures of Measurement System Reliability 13

2.1 Power Curves for Testing (1.19) when $\rho_0 = 0.80$ and $\rho_0 = 0.91$, . . 41

3.1 Left panel; 2 measurements on 10 parts chosen at random. Right panel; 10 parts the most extreme initial measurements chosen non-randomly. 49

3.2 The ratio Var $\left(\widetilde{\beta}_1\right)$ /Var $\left(\widetilde{\beta}_2\right)$, when both plans have the same sample size. 51

3.3 Plots of $\text{Var}(\widetilde{\rho}_2)/\text{Var}(\widetilde{\rho}_1) = [(1-\rho^2)/z^2] \,/\, [2(1-\rho)^2]$, the left panel is a contour plot by z and ρ and the right panel displays the quantity when $z = 2$. 53

4.1 Comparison of the Bias and Standard Deviation for Estimators of ρ with a Leveraged Plan 74

4.2 The approximate ratio of $\text{Var}(\widetilde{\rho}_r)$ as given by (4.34) 76

4.3 The standard deviation and bias of the LP estimators of ρ for different distributional assumptions 86

4.4 Comparison of Standard Deviation and Bias for a Leveraged and Standard Plan 87

4.5 The Sample Size N Required for a Leveraged Plan to Have the Same Standard Deviation as the Standard Plan ($k = 10$, $n = 6$) Across the Range of ρ 88

4.6 Comparison of Standard Deviation and Bias of $\tilde{\rho}$ for an LP and SP 89

4.7 Comparison of Optimal LP and SP. The left panels show the ratio of standard deviations for the optimal LP and SP. The right panels show the sample size of an LP that has the same standard deviation as the optimal SP that has N=60 or 100 measurements. 91

4.8 Left panel shows the ratio of standard deviations for the optimal LP and SP when using ML to estimate ρ. Right panel shows the difference $N-N_o$ where N_o is defined as the number of measurements required for an optimal LP to have the same standard deviation as an optimal SP with N measurements and a given ρ. 92

4.9 Left panel shows the ratio of standard deviations for the recommended LP and optimal SP when using ML to estimate ρ. Right panel shows contours of $N-N_r$ by ρ and N where N_r is defined as the number of measurements required for the recommended LP to have the same standard deviation as an optimal SP with N measurements and a given ρ. 93

4.10 The MSE of the MLE's for the SP ($k = 10, n = 6$) and LP ($b = 30, k = 6, n = 5.$) under different distributional assumptions; 1. both the part and measurement error distribution are normal, 2. the part distribution is t_5 and the measurement error distribution is normal, 3. the part distribution is normal and the measurement error distribution is t_5, 4. both the part and measurement error distributions are t_5. 95

4.11 The MSE of the MLE's for the SP ($k = 50, n = 2$) and LP ($b = 52, k = 16, n = 3.$) under different distributional assumptions; 1. both the part and measurement error distribution are normal, 2. the part distribution is t_5 and the measurement error distribution is normal, 3. the part distribution is normal and the measurement error distribution is t_5, 4. both the part and measurement error distributions are t_5. 96

5.1 Power Curves for Testing the Hypothesis (1.19) when $\rho_0 = 0.80$ and $\rho_0 = 0.91$ at $\alpha = 0.05$, with six repeated measurements and ten parts having SSS = 10, 40 and 90. 102

5.2 Power Curves for Testing the Hypothesis (1.19) when $\rho_0 = 0.80$ (left), $\rho_0 = 0.91$ (right), ten parts having $\sum z_{i,0}^2 = 40$ and $k = 10$, $n = 6$. 106

5.3 Contours of ρ for Testing (1.19) when $\rho_0 = 0.80$ and $\rho_0 = 0.91$ at $\alpha = 0.05$, power = 0.80, with $\sum_{i=1}^{k} z_{i,0}^2 = 4k$. 107

5.4 Contours of ρ for Testing (1.19) when $\rho_0 = 0, .2, .4$ and $\rho_0 = 0.6$ at $\alpha = 0.05$, power = 0.80, with $\sum_{i=1}^{k} z_{i,0}^2 = 4k$. 108

5.5 The ratio of the square root of the MSE for the LP ($n = 6$ and ten part with standardized initial measurements $z_0 = 2$) and two SPs with $k = 10, n = 6$ (left panel) and $k = 10, n = 7$ (right panel). . . . 110

5.6 Power Curves for the LP and SP from Testing (1.19) when $\rho_0 = 0.80$, $\rho_0 = 0.91$, ten parts having $\sum z_{i,0}^2 = 40$ and 6 repeated measurements. 111

5.7 Power Curves for the LP and SP from Testing (1.19) when $\rho_0 = 0, .2, .4$ and $\rho_0 = 0.6$, ten parts having $\sum z_{i,0}^2 = 40$ and 6 repeated measurements. 112

6.1 Checking model fit. The left panel shows a QQ plot of the baseline residuals $y_{ij0} - \overline{y}_{.j0}$ and the right panel shows these residuals by operator. 125

6.2 Checking model fit. The left panel shows a QQ plot of the 2^{nd} stage residuals $y_{ijlh} - \overline{y}_{ijl.}$ and the right panel shows these residuals by operator. 126

6.3 The asymptotic standard deviation for the MLE of η when $b = 15, k = 2, n = 2, m = 2$ and the initial measurements u_1 and u_2 are from operator 1 and 2, respectively. Each plot is scaled by its minimum. 129

6.4 The ratio of asymptotic standard deviations for the MLEs of η from $\mathbf{u}^t = (2, -2)$ compare to $\mathbf{u}^t = (-2, 2)$ when $b = 15, k = 2, n = 2, m = 2$ and the initial measurements u_1 and u_2 are from operator 1 and 2, respectively. 130

6.5 The ratio of asymptotic standard deviations for the MLEs of η from the balanced (left panel) and worst (right panel) selection methods compared to the optimal selection method. All three selection methods have the same design ($b = 18, k = 6, n = 2$ and $m = 3$). 131

6.6 The standard deviation of the LP estimator from two designs $b = 11, k = 3, n = 3$ (left panel) and $b = 18, k = 6, n = 2$ (right panel) divided by the standard deviation of the optimal plan given $N = 60$ (left panel), 90 (right panel), η, λ and $m = 3$. 134

6.7 The MSE ratios for estimators of η (left panel) and λ (right panel) from the leveraged plan ($b = 11$, $k = 3$, $n = 3$) and standard plan ($k = 10$, $n = 2$) across η and λ when the total sample size is 60. . . 136

6.8 The MSE ratios for estimators of η (left panel) and λ (right panel) from the leveraged plan ($b = 18$, $k = 6$ and $n = 2$) and standard plan ($k = 10$ and $n = 3$) across η and λ when the total sample size is 90. 137

6.9 The mean squared error (MSE) ratios for estimating η (left panel) and λ (right panel) from the leveraged plan ($b = 16, k = 2, n = 2$) and standard plan ($k = 10, n = 2$) across η and λ when the total sample size is 60. 145

6.10 The mean squared error (MSE) ratios for estimating η (left panel) and λ (right panel) from the leveraged plan (b = 26, k = 2 and n = 2) and standard plan (k = 10 and n = 3) across η and λ when the total sample size is 90. 146

6.11 The mean squared error (MSE) ratios for estimating η (left panel) and λ (right panel) from the leveraged plan A ($b = 11, k = 3, n = 3$) and leveraged plan B ($b = 16, k = 2n = 2$) across η and λ when the total sample size is 60. 147

6.12 The mean squared error (MSE) ratios for estimating η (left panel) and λ (right panel) from the leveraged plan A ($b = 18, k = 6, n = 2$) and leveraged plan B ($b = 26, k = 2, n = 2$) across η and λ when the total sample size is 90. 148

7.1 Typical Inspection Protocol 150

7.2 The left panel gives β_1 for different values of the standardized inspection limits. The right panel is the bias of $\tilde{\rho}_a$ as a function of ρ and β_1. 156

7.3 The asymptotic standard deviation of the MLE by ρ and β_1 160

7.4 $\sqrt{\text{MSE}(\widetilde{\rho}_a)}$ and $\sqrt{\text{MSE}(\widetilde{\rho}_r)}$ divided by $\sqrt{\text{MSE}(\widetilde{\rho}_{MLE})}$ by ρ and β_1 161

7.5 Normal quantile plots of the production data (left panel) and the residuals from regressing y_{i2} onto y_{i1} (right panel). 161

Chapter 1

Introduction

The quality of the measurement system is crucial to any manufacturing process, scientific investigation, or clinical trial because all measurements are subject to error. As discussed in Shrout and Fleiss [1979], "measurement error can seriously affect statistical analysis and interpretation; it therefore is important to quantify the statistical properties of measurement systems such as bias, variation, etc."

This thesis focuses on the context of a manufacturing process but quantifying measurement variation is also important in other contexts. For instance, in the medical field we measure people rather than parts for health measures such as blood cholesterol levels.

In a manufacturing setting, measurements of critical characteristics are taken to determine if a part is being built to specification. For simplicity of language, we assume that the characteristic of interest is a physical dimension but this assumption is not critical. We also assume each part characteristic has a true value which is unknown but fixed. That is, we assume that measuring the characteristic does not affect its true value, so for example, we do not consider destructive measurement systems. Another critical assumption is that we can make independent measurements on the same part.

In this thesis, we consider three contexts:

- a measurement system with no operators and a single gauge,
- a measurement system used with 100% inspection where the overall process characteristics are known or estimated with negligible error,
- a measurement system with operators and a single gauge.

In the following sections we introduce and describe each context. In addition, in each context, we review the metrics that characterize measurement system quality. In the final section of the chapter, we describe the goal and outline of the thesis.

1.1 A Measurement System with No Operators and a Single Gauge

The observations from a measurement system have two sources of variability; the true part dimensions and the measurement errors. Measurement variation is the variation observed among measurements on the same part. Part variation is the additional variation observed among measurements on different parts. The part variation is also referred to as the process variation.

Using manufacturing terminology, a commonly used statistical model for the outcomes from a measurement system is the random effects model

$$Y = X + E \tag{1.1}$$

where X is the random variable representing the possible true values for the dimension of a part and E is a random variable representing the measurement error. It is

commonly assumed that the part effects (X) are independent and identically distributed normal random variables with mean μ and variance σ_p^2, the measurement errors (E) are independent and identically distributed normal random variables with mean zero and variance σ_m^2, and X and E are mutually independent. The variance of Y, called the total variation, is $\sigma_t^2 = \sigma_p^2 + \sigma_m^2$. We refer to σ_m^2 as the measurement variation and σ_p^2 as the process variation.

Studies that quantify the measurement variation are used in several applications: assessing a new measurement system, a routine assessment of an existing system, and assessing a system as one of the first steps in process problem solving.

One way to summarize the relative sizes of the measurement and process variation is to use the proportion of variances. One such proportion is called the intraclass correlation coefficient, $\rho = \frac{\sigma_p^2}{\sigma_p^2+\sigma_m^2}$. It is the correlation between two measurements on the same part, since two measurements on the same part, (Y_1 & Y_2) using model (1.1), can be written as $Y_1 = X + E_1$ and $Y_2 = X + E_2$. The covariance between these two measurement is

$$\begin{aligned} Cov\left(Y_1, Y_2\right) = Cov\left(X + E_1, X + E_2\right) &= Cov\left(X, X\right) = Var\left(X\right) = \sigma_p^2 \\ \text{and} \qquad Cor(Y_1, Y_2) = \frac{Cov\left(Y_1, Y_2\right)}{\sqrt{Var\left(Y_1\right) Var\left(Y_2\right)}} &= \frac{\sigma_p^2}{\sigma_t^2} = \rho. \end{aligned}$$

Also, note that correlation between the true part dimension X and a single measurement Y_1 on that part is

$$Cor(Y_1, X) = \frac{Cov\left(Y_1, X\right)}{\sqrt{Var\left(Y_1\right) Var\left(X\right)}} = \frac{\sigma_p^2}{\sigma_p \sigma_t} = \sqrt{\rho}.$$

The metric ρ is bounded between $[0, 1]$ and values near 1 indicate that the part variation is large relative to the measurement variation. Conversely, if ρ is near 0 then the measurement variation is large relative to the part variation.

Figure 1.1, shows the relationship between two measurements on 100 different parts when the intraclass correlation coefficient, ρ, is 0.2 and 0.9. When the process variation is large relative to the measurement system, e.g. $\rho = 0.9$, the relationship between the two measurements on the same part is strong. When the measurement variation is small, we expect two measurements on the same part to be similar. On the other hand, when the measurement system is the larger source of variation, e.g. $\rho = 0.2$, there is little correlation between repeated measurements on the same part.

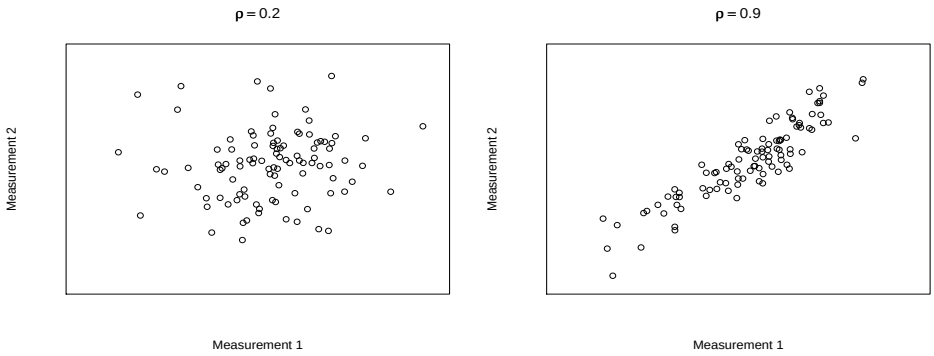

Figure 1.1: Scatterplot of Two Measurements on 100 Parts. Left panel has $\rho = 0.2$ and the right panel $\rho = 0.9$.

An important property of measurement systems is bias. We can include bias in model (1.1) if the measurement error E has a non-zero mean, μ_m, instead of zero. We demonstrate a bias in a measurement system using Figure 1.2. In each panel, the line corresponds to equality of the measured and true dimensions. The left panel in Figure 1.2, shows a measurement system which is unbiased, i.e. $\mu_m = 0$. A simple form of bias is a positive or negative constant adjustment across the true dimensions. A constant positive bias is shown the right panel of Figure 1.2. In general, to check measurement bias, we require parts with known dimensions. Later, we will see that for measurement systems involving multiple operators it is possible to check for relative bias of each operator with a part of unknown true

dimension.

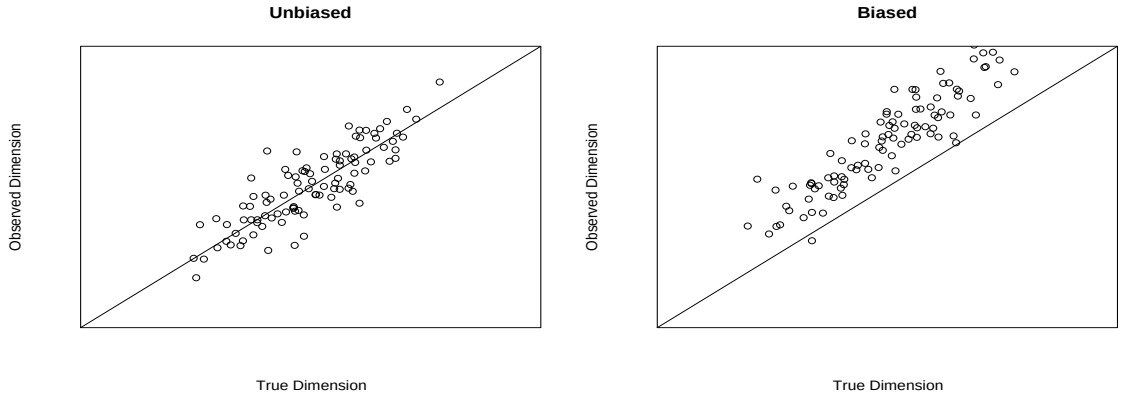

Figure 1.2: Two Measurement Systems: Zero Bias (left), Positive Bias (right)

In a widely used manual on measurement system assessment [Automotive Industry Action Group, 2002], linearity is a term used to describe how the bias or measurement variability changes over the range of true dimensions. The measurement bias and variability are considered "linear" if μ_m and σ_m do not depend on the true dimension (as in either panel of Figure 1.2). A measurement system showing signs of non-linearity in the measurement variance is shown in the right panel of Figure 1.3. In the plot, the measurement variance increases with the true dimension.

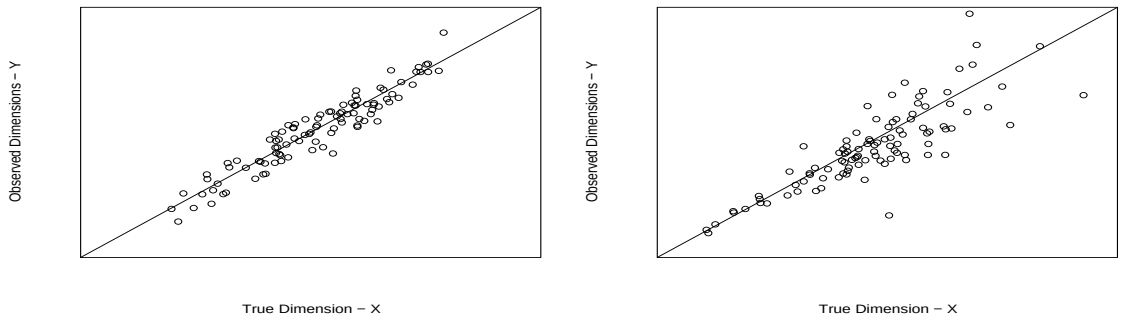

Figure 1.3: Non-Linearity of Measurement Variability; left panel linearity and right panel non-linearity

Stability is a property similar to linearity. A measurement system is unstable if its measurement bias or variability depend on time. An example of an unstable measurement system is drift in the bias, perhaps due to degradation of the measurement system over time.

Linearity and stability are important quantities, but by adopting model (1.1) we implicitly assume that the measurement system is linear and stable. That is, we assume the measurement bias and variation are constant over both time and the true dimension. However, when planning and analyzing a measurement system study, we should always think about checking these assumptions and we provide ways for making such checks in the recommended new study plans in Chapters 4 to 7.

1.1.1 Characterizing Measurement System Quality

A variety of performance metrics are used to quantify the precision or quality of the measurement system. The appropriate choice in any application depends on the goal and custom. The four common categories of metrics are functions of

- measurement variation (σ_m) only,
- measurement variation (σ_m) and the lower and upper specification limits (LSL & USL),
- measurement and process variation (σ_m & σ_p), and
- measurement and process variation (σ_m & σ_p), process mean (μ), LSL and USL.

The LSL and USL for the measured characteristic are given by the product/process design. To calculate any metric defined in terms of the parameters σ_m, σ_p and μ, these need to be estimated with an assessment study.

In most cases, the following metrics summarize the performance of the measurement system as a single number. This is clearly a loss of information. These metrics must be interpretable because otherwise one should report the original parameter(s) instead.

1.1.1.1 Metrics based on Measurement Standard Deviation

How can we assess the quality of a measurement system with only the variation? One answer is to report the standard deviation σ_m, but does a more interpretable measure of performance exist? A measurement system is used to estimate the true dimension of a part. Thus a natural metric is to quantify how well a measurement system can estimate the true dimension of a part with a single observation. This idea is called the effective resolution of a measurement system in Wheeler and Lyday [1984].

One possibility is the width of a $100\,(1-\alpha)\,\%$ confidence interval for the true dimension with a single observation. For example, if the measurement error has a normal distribution with standard deviation σ_m the metric is $2\sigma_m Z_{1-\alpha/2}$, where $Z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of a standard normal distribution. The width of a 95% ($\alpha = 0.05$) confidence interval for the true dimension using a single measurement is roughly $4\sigma_m$.

Wheeler and Lyday [1984] use the same idea, but their metric is based on the half width of a 50% confidence interval. If we again assume normality, this quantity is $0.67\sigma_m$ and known as the "Median Uncertainty" or the "Probable Error" of a measurement.

Another argument which leads to a similar result is based on the question: What is the required distance between two true dimensions for the measurement system to reliably distinguish between them? We can quantify this question by examining

the power of a statistical test of hypothesis to detect differences among the true values. The power of a statistical test is the probability that the test will reject a false null hypothesis. As the difference between the true dimensions gets larger the power will increase. How large a difference in the true dimensions will result in an acceptably high power?

To set up the appropriate test of hypothesis, we assume δ_1 and δ_2 are the true values of the two parts. Then, the measured values Y_1, Y_2 from each part have a normal distribution with means δ_1 and δ_2, and variance σ_m^2. To test the hypothesis $H_0 : \delta_1 = \delta_2$ vs. $H_A : \delta_1 \neq \delta_2$ we use the statistic

$$\frac{y_1 - y_2}{\sqrt{2}\sigma_m}.$$

where the y_1, y_2 are the observed values. The test of hypothesis with size α will be rejected if

$$\left| \frac{y_1 - y_2}{\sqrt{2}\sigma_m} \right| > Z_{1-\alpha/2} \tag{1.2}$$

The size of a test α is the probability of rejecting the null hypothesis H_0 when it is true. In the above test, there is an α chance of rejecting $\delta_1 = \delta_2$ when it is true.

The question of interest becomes how large does the difference between δ_1 and δ_2 have to be for the test to have a power equal to $1 - \beta$. The power of the test is

$$1 - \beta = 1 - \Phi\left(Z_{1-\alpha/2} - \frac{\delta_0}{\sqrt{2}\sigma_m}\right) + \Phi\left(-Z_{1-\alpha/2} - \frac{\delta_0}{\sqrt{2}\sigma_m}\right) \tag{1.3}$$

where $\delta_0 = \delta_1 - \delta_2$. This means the difference between the two parts has to be

$$\delta_0 \geq \sqrt{2}\sigma_m \left| Z_\beta + Z_\alpha \right| \tag{1.4}$$

for the test to have size α and power $1-\beta$. Thus, another metric for a measurement

system is $\sqrt{2}\,|Z_\beta + Z_\alpha|\,\sigma_m$.

To illustrate this idea, consider the following numerical example. If the measurement standard deviation is $\sigma_m = 1$ and the test of hypothesis is required to have $\alpha = 0.05$ and power $1 - \beta = 0.80$, then the true dimensions have to be at least 3.96 units away from each other to be reliably distinguished by the measurement system. The term "reliably" is used to represent the size and power that were chosen and represents one's own personal risk level. I have chosen two particular levels of the size and power but another person may choose different values.

All of these metrics define a form of effective resolution for the measurement system and are a constant times σ_m. They simplify the interpretation of the standard deviation but since they are simply scalar multiples of σ_m, I suggest reporting the standard deviation only. Then, one can obtain any chosen metric by multiplying the standard deviation by the appropriate constant. From a statistical perspective, any of these metrics can be estimated once we have an estimate of σ_m.

1.1.1.2 Metrics based on Comparing Measurement Variation to the Specification Limits

Burdick et al. [2005] and Automotive Industry Action Group [2002] use the precision-to-tolerance ratio (PTR) to assess a measurement system. PTR compares the width of the measurement error distribution to the width of the specification limits (tolerance). It is calculated as

$$\text{PTR} = \frac{k\sigma_m}{USL - LSL}$$

where k is either 5.15 or 6. The values $k = 6$ and 5.15 correspond to the width, centered at the mean, of an interval containing 99.73% and 99.00% percent of the standard normal distribution, respectively. The guidelines in Automotive Industry

Action Group [2002] suggest using k=5.15.

The precision-to-tolerance ratio applies when the key characteristic has a two-sided specification or tolerance. The supplier or manufacturer will compare PTR to some standard for the purpose of approving the measurement system. The Automotive Industry Action Group [2002] suggests classifying measurement systems using PTR as

- PTR < 0.1: the measurement system is capable.
- PTR > 0.3: the measurement system is not capable.
- $0.1 <$ PTR < 0.3: the measurement system may be capable.

What does this metric quantify? It does not assess the measurement system relative to a process because the PTR does not take into account the distribution of the true part dimensions. Montgomery and Runger [1993a] and Larsen et al. [1999] noted that PTR does not necessarily indicate how well a measurement system performs for a particular process. This can occur because a process which always produces parts well within specification can tolerate a measurement system that is not capable.

It is hard to state a good reason why a measurement system should be assessed only relative to the specification limits. A measurement system is intended to be used along with a production process and not just the specification limits. One justification for comparing the measurement variation and the specification limits would be for a process that has not started producing yet. In this situation a measure of PTR would be useful because the process variation is unknown and there would be a requirement for the measurement system to be able to distinguish parts inside and outside the specification limits.

Wheeler and Lyday [1984] give an example of why measurement variation should not be compared only to the specifications limits. In their example, a statistician determined a measurement system was not capable relative to the tolerance because σ_m was 8.0 units and the tolerance was 20.0 which gave a $PTR = 2.06$. Based on these results, the statistician recommended a measurement system upgrade costing \$1.6 million. The plant manger, after seeing this price tag, brought the statistician back to reassess the conclusion. It was then discovered that the standard deviation due to the true dimensions (σ_p) was 36.0 units. Relative to the process, the measurement system was good enough because upgrading the measurement system would only help them better quantify how bad the process was. So, rather than spending \$1.6 million on an upgrade to the measurement system they invested these resources into improving their process.

This example shows why comparing measurement system to the tolerance can be misleading and should be avoided. Metrics based on the measurement variation and the specification limits do not fully characterize the measurement system which can lead to a misallocation of resources.

1.1.1.3 Metrics based on Measurement and Process Variation

There are two classes of metrics that compare the measurement and process variation. One class simply compares the relative size of the two. The other tries to answer the question, "how many distinct categories of parts from the process can the measurement system identify?"

Relative Variance

Comparing the size of measurement and process variation is valuable for process improvement. Often the goal of process improvement is to reduce the total variation

$\sigma_t^2 = \sigma_m^2 + \sigma_p^2$. If $\sigma_m > \sigma_p$ then reducing the variation due to the measurement system will have a bigger impact on σ_t^2 than reducing the process variation. Four common ways to quantify the relative size of measurement and process variation are:

1. Gauge Repeatability (GR) defined by the proportion of measurement variation relative to the total variation on the standard deviation scale [Automotive Industry Action Group, 2002]

$$GR = \frac{\sigma_m}{\sqrt{\sigma_m^2 + \sigma_p^2}} = \frac{\sigma_m}{\sigma_t}.$$

2. The intraclass correlation coefficient, ρ, defined as the proportion of process variation relative to the total variation [Donner and Eliasziw, 1987],

$$\rho = \frac{\sigma_p^2}{\sigma_m^2 + \sigma_p^2} = \frac{\sigma_p^2}{\sigma_t^2}. \tag{1.5}$$

3. The discrimination ratio, D, defined as the process standard deviation divided by the measurement standard deviation [Steiner and Mackay, 2005]

$$D = \frac{\sigma_p}{\sigma_m}. \tag{1.6}$$

   The metric D has also been called the signal-to-noise ratio (SNR) (See Burdick et al. [2003], Automotive Industry Action Group [2002], and Larsen [2002]).

4. The classification ratio, D_R, defined as the eccentricity of the bivariate probability ellipse from two measurements on the same part. Eccentricity is the deviation of a curve or orbit from circularity. In the literature D_R is called the discrimination ratio but to avoid confusion with D we call D_R the classification ratio. It is defined in terms of ρ, by Wheeler and Lyday [1984]

as

$$D_R = \sqrt{\frac{1+\rho}{1-\rho}}.$$

Clearly, these four measures are one-to-one functions of each other, since as shown in Figure 1.4, $\rho = 1 - GR^2$, $\rho = \frac{D^2}{1+D^2}$ and $\rho = \frac{D_R^2-1}{D_R^2+1}$. This means they all convey the same information. It also means that a test of hypothesis phrased in terms of one measure can be re-phrased in terms of another. Table 1.1 shows common cutoff values of the GR, discrimination ratio and classification ratio used to determine if the measurement system is acceptable. Note that the GR guideline is the strictest because it requires the measurement variation to be less than one percent of the total variation or ten percent on the standard deviation scale.

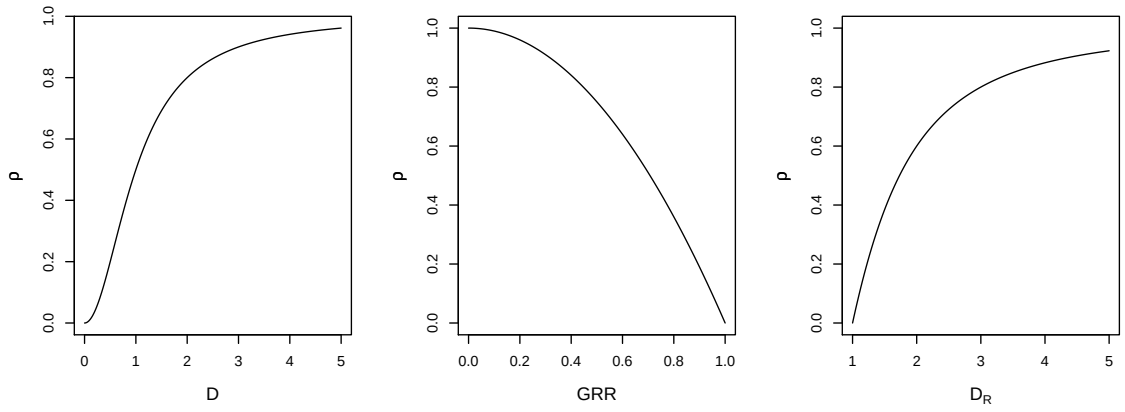

Figure 1.4: Comparison of Three Common Measures of Measurement System Reliability

Any of these metrics are equivalent when comparing the relative variances of σ_p^2 and σ_m^2. Generally, in the medical industry ρ is preferred whereas in the manufacturing industry using GR% is more common. Some favour using the intraclass correlation coefficient (1.5) because it can be interpreted as the correlation between two measurements on the same part. For this thesis, we focus on the intraclass correlation coefficient, ρ.

Table 1.1: Cutoffs for Acceptable and Unacceptable Measurement Systems in Terms of ρ

	Mesurement System		
	acceptable	needs improvement	not acceptable
GR	$< 10\%$	$.99 > \rho > .91$	$> 30\%$
Discrimination Ratio D	≥ 3	$.90 > \rho > .80$	< 2
Classification Ratio D_R	≥ 4	$.88 > \rho > .60$	< 2

Number of Distinct Categories

Another metric of interest is the number of distinct categories that a measurement system can readily identify. Here we define several metrics that claim to report this quantity. The metrics are useful but the notion of the number of distinct categories is not well defined.

One metric, recommended by Automotive Industry Action Group [2002][pg. 117] is the number of distinct categories (ndc). It is defined as the number of non-overlapping 97% confidence intervals for the true value of the measured characteristic that will span the expected product variation. The ndc is also referred to as the number of distinct categories that can be reliably distinguished by the measurement system. Equation (1.7) shows how to calculate the ndc.

$$ndc = \lfloor 1.41\frac{\sigma_p}{\sigma_m} \rfloor \approx \lfloor \frac{6.12\sigma_p}{2 \times (2.17) \times \sigma_m} \rfloor \tag{1.7}$$

where $\lfloor \rfloor$ means to truncate to a whole number. The Automotive Industry Action Group [2002] suggests ndc should be at least 5 for the measurement system to be acceptable. The corresponding value of ρ is 0.926.

From Woodall and Borror [2007], ndc can be derived using two facts. Both use properties of the normal distribution. The denominator of (1.7) is based on the

fact that $2 \times (2.17)\sigma_m$ is the width of a 98.5% confidence interval based on a single observation. The value 2.17 corresponds to the quantile of the 98.5^{th} percentile of the standard normal distribution. The numerator of (1.7) uses the fact that for a normal distribution 99.8% of the process variation is captured by an interval of width $6.12\sigma_p$. Then, we round $\frac{6.12}{2\times(2.17)}$ to the second decimal place to get 1.41.

Woodall and Borror [2007] state that on pg. 113 of Automotive Industry Action Group [2002], "The number of data categories is often referred to as the discrimination ratio since it describes how many classifications can be reliably distinguished given the observed process variation." A similar statement is made by Wheeler and Lyday [1984], who state, "The ratio of the major axis to the minor axis defines the number of distinct product categories which could be established with the measurements while making allowance for measurement error, and this ratio is estimated by the Discrimination Ratio."

Wheeler and Lyday [1984] suggest truncating D_R to give the number of distinct categories that a measurement system can readily identify. Woodall and Borror [2007] and Wheeler and Lyday [1984] showed there is a close relationship between the ndc (before truncation) and D_R. If one uses 1.41 as an approximation to $\sqrt{2}$ then

$$D_R = \sqrt{ndc^2 + 1}.$$

Measures of the number of distinct categories represent a summary of the performance of a measurement system. However, this class of metrics is not easy to understand because the term "readily identify" is not clearly defined. Another clear disadvantage is that to derive the number of distinct categories, we truncate a continuous measure of performance with the resulting loss of information.

1.1.1.4 Metrics based on the Process Mean and Variation, Measurement Variation and the Specification Limits

Another way to assess the performance of a measurement system is by relating measurement and process variation to the specification limits. Specification limits denote the interval of true dimensions that are acceptable. These limits are defined by the producer and/or customer. Due to measurement error, a part may have a measured value within specifications while its true value is outside the specifications limits or vice versa. When true part dimensions are close to either the LSL or USL, it is easy to imagine that some parts could be labeled FAIL (measured value outside of specifications) when the true dimension is GOOD (true dimension within specifications).

The quality of a measurement system can be defined by how well it discriminates between good and bad parts. Problems occur when parts are misclassified. The missed failures (MF) and false failures (FF) are probabilities that can summarize these problems. Doganaksoy [2000] prefers the Customer and Producer Risk which are different probabilities although they are related to the MF and FF. These probabilities [Burdick et al., 2003, Doganaksoy, 2000, Larsen et al., 1999] involve the following events for individual parts:

- BAD - the true dimension is outside specification,
- GOOD - the true dimension is within specification,
- FAIL - the measured dimension is outside specifications, and
- PASS - the measured dimension is within specifications.

Two definitions of customer and producer risk are typically considered in the literature. One is the manufacturing version of sensitivity and specificity called

the conditional MF & FF, (See Burdick et al. [2003] and Larsen et al. [1999]). Alternatively, Doganaksoy [2000] uses the joint probabilities of MF & FF to define customer and producer risk. Two other useful measures not considered by the literature are the escaped failures and detained quality probabilities. All three measures of risk are defined in Table 1.2.

Table 1.2: Measures of Customer and Producer Risk

	Risk	
	Consumer	Producer
Conditional MF & FF	Pr(PASS | BAD)	Pr(FAIL | GOOD)
Joint MF & FF	Pr(BAD and PASS)	Pr(GOOD and FAIL)
Escaped Failures & Detained Quality	Pr(BAD | PASS)	Pr(GOOD | FAIL)

The different measures of consumer and producer risk are not one-to-one functions of each other. However, if we include the pass rate, Pr(PASS), to any group of risks we can calculate the other measures of risk. Often Pr(PASS) is known or well estimated from production records since it represents the current yield of the process. Note that Pr(PASS) is a function of both the performance of the process and the measurement system.

The conditional MF & FF are equivalent to type I and type II errors in hypothesis testing. Doganaksoy [2000] considers the joint probabilities to be more representative of the process because "the magnitudes of these risks help determine the (economic) impact of the measurement error and whether or not the measurement system requires a capability upgrade."

While all three sets of misclassification probabilities are called consumer and

producer risks, it is the escaped failures and detained quality metrics that the consumers and producers use to measure performance. The consumer is typically interested in how many bad parts are being shipped to them. This quantity is calculated using the probability that a part is bad (out of specifications) given that it was shipped which is the escaped failures probability. The supplier and the customer are interested in this probability or risk because this is usually related to contractual agreements. On the other hand, while the supplier is interested in the escaped failures they are also curious to know how many good parts are being scrapped. These parts cost the company since they should be shipped.

One problem with all these measures of risk is that there are no established cutoff values or acceptable levels. Burdick et al. [2005] suggest comparing the conditional MF & FF rates to the the misclassification rates that one would get by employing a chance measurement system. The chance measurement system classifies parts not by taking measurements, but by instead labeling a part PASS with probability π and FAIL with probability $1-\pi$, where the probability π is the assumed known value of the proportion of GOOD parts. The joint MF & FF probabilities for the chance measurement system are both $\pi\,(1-\pi)$. The conditional MF & FF probabilities are π and $(1-\pi)$, respectively and escaped failures and detained quality probabilities are $(1-\pi)$ and π, respectively. Note that the chance measurement system depends on knowing the probability of producing a good part which needs to be estimated from the process.

Currently, there are no guidelines on how to reduce an unacceptable misclassification probability. For example, we do not know how changing the process mean affects these probabilities. The escaped failures and detained quality measures can be reduced by remeasuring parts that failed or passed. For example, if we want to reduce the number of bad parts that are being shipped, we can simply remeasure parts that have passed. Assuming the additional measurements are independent,

more bad parts will be found. Conversely, if we remeasure the failed parts and ship ones that pass the second time we will be decreasing the detained quality and increasing the escaped failures. Another possible solution is to measure a part once and then examine the probability of a bad part given the measured value. Then make a decision based on this probability.

To calculate the consumer and producer risks we use the joint probability density function (pdf) for Y and X from model (1.1). We derive results using the bivariate normal probability density function (pdf) given by

$$g(x,y) = \frac{1}{2\pi |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} \begin{bmatrix} x-\mu_p \\ y-\mu_p \end{bmatrix} \Sigma^{-1} \begin{bmatrix} x-\mu_p \\ y-\mu_p \end{bmatrix} \right)$$

where Σ is the covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_p^2 & \sigma_p^2 \\ \sigma_p^2 & \sigma_p^2 + \sigma_m^2 \end{bmatrix} = \sigma_t^2 \begin{bmatrix} \rho & \rho \\ \rho & 1 \end{bmatrix}$$

The marginal pdfs of X and Y are both normal with means μ_p, μ_p and variances σ_p^2 and $\sigma_p^2 + \sigma_m^2$ respectively. The pdf for a normal random variable Z with mean μ and variance σ^2 is

$$f(z;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2} \right)$$

To calculate the consumer and producer risks we simply integrate $f(z)$ and $g(x,y)$ over the appropriate regions.

- **Process yield:**

$$\Pr(\text{PASS}) = \Pr(Y \in [LSL, USL]) = \int_{LSL}^{USL} f\left(z;\mu_p,\sqrt{\sigma_p^2+\sigma_m^2}\right) dz$$

- **True process yield:**

$$\Pr(\text{GOOD}) = \Pr\left(X \in [LSL, USL]\right) = \int_{LSL}^{USL} f\left(z; \mu_p, \sigma_p\right) dz$$

- **Passing a good part**

$$\begin{aligned} \Pr(\text{GOOD and PASS}) &= \Pr\left(X \in [LSL, USL] \cap Y \in [LSL, USL]\right) \\ &= \int_{LSL}^{USL} \int_{LSL}^{USL} g\left(x, y\right) dxdy \end{aligned}$$

- **Joint Consumer Risk**

$$\begin{aligned} \Pr(\text{BAD and PASS}) &= \Pr\left(X \notin [LSL, USL] \cap Y \in [LSL, USL]\right) \\ &= \int_{LSL}^{USL} \int_{-\infty}^{LSL} g\left(x, y\right) dxdy + \int_{LSL}^{USL} \int_{USL}^{\infty} g\left(x, y\right) dxdy \end{aligned}$$

- **Joint Producer Risk**

$$\begin{aligned} \Pr(\text{GOOD and FAIL}) &= \Pr\left(X \in [LSL \cap USL], Y \notin [LSL, USL]\right) \\ &= \int_{-\infty}^{LSL} \int_{LSL}^{USL} g\left(x, y\right) dxdy + \int_{USL}^{\infty} \int_{LSL}^{USL} g\left(x, y\right) dxdy \end{aligned}$$

- **Failing a bad part**

$$\begin{aligned} \Pr(\text{BAD and FAIL}) &= \Pr\left(X \notin [LSL, USL] \cap Y \notin [LSL, USL]\right) \\ &= \int_{-\infty}^{LSL} \int_{-\infty}^{LSL} g\left(x, y\right) dxdy + \int_{-\infty}^{LSL} \int_{USL}^{\infty} g\left(x, y\right) dxdy \\ &\quad + \int_{USL}^{\infty} \int_{-\infty}^{LSL} g\left(x, y\right) dxdy + \int_{USL}^{\infty} \int_{USL}^{\infty} g\left(x, y\right) dxdy \end{aligned}$$

- **Conditional Consumer Risk**

$$\Pr\left(\text{PASS}|\,\text{BAD}\right) = \frac{\Pr\left(\text{BAD and PASS}\right)}{\Pr\left(\text{BAD}\right)}$$

- **Conditional Producer Risk**

$$\Pr\left(\text{FAIL}|\,\text{GOOD}\right) = \frac{\Pr\left(\text{GOOD and FAIL}\right)}{\Pr\left(\text{GOOD}\right)}$$

- **Escaped Failures**

$$\Pr\left(\text{BAD}|\,\text{PASS}\right) = \frac{\Pr\left(\text{BAD and PASS}\right)}{\Pr\left(\text{PASS}\right)}$$

- **Detained Quality**

$$\Pr\left(\text{GOOD}|\,\text{FAIL}\right) = \frac{\Pr\left(\text{GOOD and FAIL}\right)}{\Pr\left(\text{FAIL}\right)}$$

1.1.1.5 Numerical Example

This numerical example relates to the quality of a lamp used to illuminate the target for an optical scanning device and is taken from Larsen et al. [1999]. In an industrial process 150,000 lamps are produced per month and 100% inspection is utilized. An important test parameter for lamp performance is the luminance in units of candelas per square meter (cd/m^2). A measurement assessment study was conducted. Using the study results, we estimate μ, σ_p^2 and σ_t^2 as $35.2, 16.81$, and 17.41 respectively. By using the relation $\sigma_p^2 + \sigma_m^2 = \sigma_t^2$ we estimate $\sigma_m^2 = 0.60$. The LSL and USL are 30 and 42. All the previously described metrics for the performance of the measurement system are estimated below.

1. Measurement variability

- 95% Confidence Interval: $2\sigma_m Z_{1-0.05/2} = 3.063$
- Median Uncertainty or Probable Error: $0.67\sigma_m = 0.52$
- Minimum distance between 2 dimensions to reject when size 0.05 and power 0.20: $\sqrt{2}\sigma_m \left|Z_\beta + Z_\alpha\right| = 0.80$

2. Measurement variability to specification limits
   - $PTR = \frac{k\sigma_m}{USL-LSL}$ with k = 5.15: PTR = 0.33, gauge not capable
3. Metrics based on Measurement and Process Variation
   - $GRR = \frac{\sigma_m}{\sigma_t} = 0.186 \Rightarrow$ gauge needs improvement
   - Intraclass Correlation Coefficient $\rho = \frac{\sigma_p^2}{\sigma_t^2} = 0.966$
   - Discrimination Ratio $D = \frac{\sigma_p}{\sigma_m} = 5.29 \Rightarrow$ gauge acceptable for process improvement
   - Classification Ratio $D_R = \sqrt{\frac{1+\rho}{1-\rho}} = 7.55 \Rightarrow$ gauge acceptable
   - $ndc = \lfloor 1.41\frac{\sigma_p}{\sigma_m} \rfloor = \lfloor 7.46 \rfloor = 7 \Rightarrow$ gauge acceptable
4. Metrics based on Measurement, Process Variation and Specification Limits
   - Conditional Consumer and Producer Risk: 0.1180 & 0.0292.
   - Joint Consumer and Producer Risk: 0.0178 & 0.0248.
   - Escaped Failure Rate: 0.0211 and Detained Quality Rate: 0.1571

Similarly, a table of the estimated conditional and joint probabilities was formed and it is replicated in Table 1.3. Also shown are the escaped failure and detained quality rate measures of risk.

In Larsen et al. [1999] on page 572, they state "the customer can expect to receive 1.78% nonconforming lamps or 2670 nonconforming lamps per month from

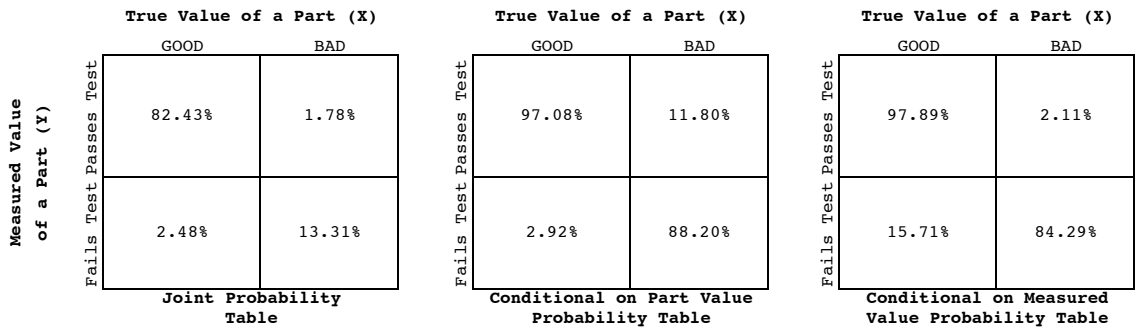

Table 1.3: Numerical Example of Metrics for Consumer and Producer Risk (given in terms of %)

the supplier." Although the stated number of 2670 nonconforming lamps is correct, the stated probability is false. The customer does not receive all the lamps, they only receive lamps that have passed inspection. The customer can expect that 2.11% of their lamps are bad which is the probability of a bad part given that it has been shipped, i.e. $\Pr(\text{BAD}|\,\text{PASS})$.

We have included all the metrics of measurement system quality for completeness but for this thesis, we focus on the intraclass correlation coefficient, ρ, given in (1.5).

1.2 A Measurement System with 100% Inspection

Many manufacturing processes use 100% inspection to ensure parts meet specifications. In an inspection system, every part produced is measured at least once. If the system stores the measured values, we can estimate the current process mean, μ and standard deviation σ_t with negligible error. Therefore in this context, the only unknown parameter is the relative size of σ_p and σ_m. Also, because of the high volume, we have parts available with values spread across the whole distribution.

One such measurement system is used to inspect journal diameters and several other characteristics of finished crankshafts in an engine assembly plant. Another involves the measurement of many functions of a circuit board assembled for use in a hand-held electronic device. In both of these examples and many others, Quality Systems (such as ISO/TC 16949 or QS 9000) adopted by the manufacturer require periodic assessments of the measurement system to ensure that the current measurement variability is relatively small compared to the underlying process variation.

If a manufacturing process does not use 100% inspection, estimates of the process mean μ and standard deviation σ_t might be available from baseline studies. In a baseline study we measure many parts once. Baseline studies are used for performance evaluation and Steiner and Mackay [2005] suggest doing a baseline study as the first stage in a variance reduction project. Note that a baseline study alone cannot assess the performance of a measurement system. Typically, baseline studies have relatively large sample sizes, so estimation of the baseline parameters, μ and σ_t, is very precise.

1.3 A Measurement System with Operators

In measurement systems, operators are often thought to be a substantial source of variability. Each operator is assumed to have a different mean effect on measurements so that there are relative biases among the operators. That is, two operators measuring the same part will obtain different mean values. This is an additional source of measurement variability. We will use multiple operator terminology, but the context would be the same, if there are multiple automated gauges with the same variability.

The most common measurement assessment plan for a system which includes

operators is called a Gauge Reproducibility and Repeatability (R&R) as described in Automotive Industry Action Group [2002]. A Gauge R&R labels the two sources of measurement variability as repeatability and reproducibility. Repeatability is the variation associated with one operator repeatedly measuring the same part. Reproducibility is the variation associated with different operators when measuring the same part. The gauge R&R plan has each operator measure each part (typically 10 parts) 2 or 3 times.

The effect from operators can be fixed or random depending on the situation. Assuming fixed effects means there are a finite (usually a small) number of operators. A random effects model characterizes the effects from operators (when chosen at random) as being from some distribution (usually assumed to be normal). If operator effects are assumed to be random, the model for the standard plan (SP) will be a two-way random effects model which is the common model assumed by Burdick et al. [2005], Montgomery and Runger [1993a] and Wheeler and Lyday [1984]. For the remainder of this thesis we treat the effect from operators as fixed. In section 8.2.3, we discuss treating the effects from operators as random.

We recommend assuming the effects from operators are fixed because in practice, only a few operators are trained to measure parts and/or actually measure parts in regular production. If a site has multiple gauges, again a fixed effect model should be assumed as there are rarely large numbers of gauges.

We extend model (1.1) to include operators, using a mixed effect model given by

$$Y = \mu_j + X + E \tag{1.8}$$

where X is a random effect of the true part dimensions, μ_j is the mean effect from operator j, $j = 1, \ldots, m$ and E is the random effect from the repeatability or a single operator repeatedly measuring the same part. X and E are assumed to

be independent normals with zero means and standard deviations σ_p, σ_g, respectively. The mean of the true part dimensions is included in the mean effect for each operator.

The following characterizes the total variation when assuming fixed effects for operators when each operator is equally likely to measure a part. We define

$$\begin{aligned}
\underline{\mu} &= (\mu_1, \ldots, \mu_m)^t, && (1.9)\\
\overline{\mu} &= \frac{1}{m}\sum_{j=1}^{m} \mu_j && (1.10)\\
\sigma^2_{pg} &= \sigma^2_p + \sigma^2_g && (1.11)\\
\sigma^2_m &= \sigma^2_o + \sigma^2_g && (1.12)\\
\text{and} \qquad \sigma^2_t &= \sigma^2_o + \sigma^2_p + \sigma^2_g && (1.13)\\
\text{where} \qquad \sigma^2_o &= \frac{1}{m}\sum_{j=1}^{m} \left(\mu_j - \overline{\mu}\right)^2, && (1.14)
\end{aligned}$$

m is the number of operators and μ_j is the mean for j^{th} operator. The parameter σ^2_{pg} is the variation seen in measurements made by any single operator on a sample of parts from the process. The parameter σ^2_t represents the total variation seen in the process if each operator measured the same proportion of parts in regular production. The parameter σ^2_o captures the variation due to differences among the m operator means (i.e. the effects of relative bias), but is not a variance in the usual sense. Finally, the parameter σ^2_m represents the total variation seen in the measurement of any particular part if each operator is used in the system with the same intensity. If each operator has the same mean, then $\sigma_0 = 0$ and we can interpret σ_m and σ_t as the standard deviations defined in model (1.1). Using manufacturing jargon, σ_m represents the overall measurement variability, σ_g the repeatability and σ_o the reproducibility. We assume σ_g is the same for each operator and part.

1.3.1 Characterizing Measurement System Quality

In the mixed effects model, we can partition the variance into three components; parts, reproducibility, and repeatability. Quantities of interest for these studies are split into two groups.

1. Metrics that use the total variation are

   - the intraclass correlation coefficient, ie. the process variation divided by the total variation
   $$\eta = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_o^2 + \sigma_g^2} = \frac{\sigma_p^2}{\sigma_t^2}, \tag{1.15}$$
   - the gauge repeatability and reproducibility, i.e. the measurement standard deviation divided by the total standard deviation.
   $$\text{GRR} = \frac{\sigma_m}{\sigma_t} = \sqrt{\frac{\sigma_m^2}{\sigma_t^2}} = \sqrt{\frac{\sigma_o^2 + \sigma_g^2}{\sigma_t^2}}, \tag{1.16}$$

2. Metrics that compare the components of the measurement variation are

   - proportion of the variation due to operator bias
   $$\lambda = \frac{\sigma_o^2}{\sigma_g^2 + \sigma_o^2} \tag{1.17}$$
   - reproducibility over the repeatability
   $$R/R = \frac{\sigma_g^2}{\sigma_o^2} \tag{1.18}$$

We have shown that the intraclass correlation coefficient and the gauge repeatability and reproducibility are equivalent metrics of measurement system quality. When trying to improve a measurement system, metrics like the reproducibility over the repeatability and the proportion of the variation due to operator bias can identify

which measurement system component contributes the majority of the variation. In this thesis, we focus on the estimation of the two quantities η and λ. We will treat η as the primary parameter of interest and λ as of secondary interest.

1.4 Goal and Outline

The goal of this thesis is to show that new measurement assessment plans, using the concept of leveraging, are more efficient than the traditional measurement assessment plans, called the standard plan (SP), in the three measurement system contexts described in Sections 1.1, 1.2 and 1.3. We also provide detailed recommendations on the design of these new plans and also indicate how to analyse the data that are collected.

We compare the estimators derived from the leveraged plan (LP) and the standard plan SP using mean squared error (MSE) which is the variance plus the bias squared. When the bias is not a significant contribution to the MSE, we will use the standard deviation. Additionally, in some contexts we contrast the two classes of plans using the hypothesis test

$$H_0 : \rho \leq \rho_0 \text{ versus } H_A : \rho > \rho_0, \tag{1.19}$$

and/or the width of confidence intervals while holding the coverage probability fixed.

To achieve our goal we begin with Chapter 2, where we describe the plan, design and analysis for the standard approach for each context. In Chapter 3, we introduce leveraging and give general guidelines for its implementation. Then, in the following three chapters we describe the new LP designs and analyses for each context and compare them to the SPs. Next, we present an application for leveraging in the

assessment of inspection systems with production data. Finally we suggest some other problems where we may exploit leveraging.

Chapter 2

Standard Measurement Assessment Plan

This chapter introduces the standard plan (SP) design and analysis used to estimate the parameters of interest in the three contexts of interest as described in chapter 1. That is, we present the details on how to perform the study and analyze the data obtained from the study. We focus on assessing the parameter ρ in the first and second context and η in the third context. To improve estimation for these ratios, we show how to incorporate data from another study called a baseline investigation.

Note throughout this thesis, we use a circumflex ($\hat{}$) to overscore a parameter to denote the estimate (a number) and an overscore tilde ($\tilde{}$) to denote the corresponding estimator (a random variable). We feel it is important to distinguish between an estimate and estimator because an estimate is single number and the estimator has distributional properties.

2.1 A Measurement System

The measurement system has two components of variation, the true part values and the measurement error. Using the standard measurement system study, we can estimate all the parameters (σ_p, σ_m and μ). In this study, we choose k parts randomly from the process and then repeatedly measure each part n times (Donner and Eliasziw [1987], and Burdick et al. [2003]). A typical design for the standard plan is when $k = 10$ and $n = 6$ [Automotive Industry Action Group, 2002]. A commonly adopted model for the data from this standard plan is the one-way random effects model

$$Y_{ij} = X_i + E_{ij} \tag{2.1}$$

where $i = 1, 2, \ldots, k, j = 1, 2, \ldots, n$, X_i is a random variable representing the possible true values for the dimension of part i, and E_{ij} is a random variable representing the measurement error, n is the number of repeated measurements on each part, and k is the number of parts. We assume that the part effects $\{X_i\}$ are independent and identically distributed normal random variables with mean μ_p and variance σ_p^2, the measurements errors $\{E_{ij}\}$ are independent and identically distributed normal random variables with mean zero and variance σ_m^2, and the $\{X_i\}$ and $\{E_{ij}\}$ are mutually independent. The variance of Y_{ij}, called the total variation, is $\sigma_t^2 = \sigma_p^2 + \sigma_m^2$. Also we assume that σ_p, σ_m and μ do not depend on the true part dimension and do not change over time; that is, the measurement system is linear and stable. Note that model (2.1) is equivalent to model (1.1).

Burdick et al. [2003] and Automotive Industry Action Group [2002] call this type of study a gauge repeatability and reproducibility, although there are no operators in this context. The acronym GR&R is widely used to describe this study plan. Another name used by Wheeler and Lyday [1984] for these studies is evaluating the measurement process (EMP) study. In this thesis, we call GR&R the standard

plan (SP).

2.1.1 Estimation

Traditionally, the Range method (Automotive Industry Action Group [2002], Burdick et al. [2003]) is used for estimation of σ_p and σ_m. However, an analysis of variance (ANOVA) and/or maximum likelihood (ML) are more efficient, can easily be carried out with easily available software and can be adapted to analyze more complex studies. Table 2.1 is from Donner and Eliasziw [1987] and shows the ANOVA table for the SP.

Table 2.1: Analysis of Variance for Gauge R&R Study

Analysis of variance

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Among parts	$k-1$	SSA	$\text{MSA} = \text{SSA}/(k-1)$	MSA/MSW
Within parts	$k(n-1)$	SSW	$\text{MSW} = \text{SSW}/[k(n-1)]$	

$$SSA = \sum_{i=1}^{k} n\left(\overline{y}_{i.} - \overline{y}_{..}\right)^2 \qquad SSW = \sum_{i=1}^{k}\sum_{j=1}^{n}\left(y_{ij} - \overline{y}_{i.}\right)^2$$

$$\overline{y}_{i.} = \sum_{j=1}^{n} y_{ij}/n \qquad \overline{y}_{..} = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}/nk$$

From Burdick et al. [2003], to estimate ρ, we use

$$\widehat{\rho} = \frac{\text{MSA} - \text{MSW}}{\text{MSA} - (n-1)\text{MSW}} \tag{2.2}$$

which is the sample intraclass correlation coefficient. MSA is the mean squared error across parts and MSW is the mean squared error within parts. See Table 2.1

for details on how to calculate these quantities. An estimate of ρ larger than 1/2 suggests the part-to-part variation is greater than the measurement variation.

The test of hypothesis (1.19) can be used to determine whether the desired level of precision, ρ_0, is achieved by the measurement system. The size of a test, α, is defined as the probability of rejecting the null hypotheses H_0 when the true value of ρ is ρ_0. The power of a test is the probability of rejecting the null hypotheses H_0 when the true value of ρ is $\rho_1 \neq \rho_0$.

Donner and Eliasziw [1987] carry out the test of hypothesis in (1.19), by comparing the F value obtained from Table 2.1 to the quantity $C_0 F_\alpha(v_1, v_2)$, where $C_0 = 1 + [n\rho_0/(1-\rho_0)]$, and $F_\alpha(v_1, v_2)$ is the value of the F distribution with v_1, v_2 degrees of freedom at the α level of significance, $v_1 = k - 1$ and $v_2 = k(n-1)$. To calculate the power of (1.19) when $\rho = \rho_1$, we use

$$\beta = 1 - P\left(F \leq C\ F_\alpha; v_1, v_2\right) \tag{2.3}$$

where

$$C\ =\ \frac{(1 + n\frac{\rho_0}{1-\rho_0})}{(1 + n\frac{\rho_1}{1-\rho_1})}.$$

Details on how to derive this formula can be found in Donner and Eliasziw [1987].

Donner and Eliasziw [1987] use (2.3) to determine appropriate sample size requirements for the test of hypothesis in (1.19). They accomplish this by rearranging (2.3), expressing ρ_1 as a function of n and k when testing (1.19) with $\alpha = 0.05$ and $\beta = 0.80$. They plot contours of ρ_1 as a function of n and k.

Burdick et al. [2003] and Burdick et al. [2005] derive confidence intervals for the SP using modified large sample methods and generalized intervals.

When using ML to analyze data, we are required to specify the distribution of the data. Assuming (2.1) and normality, n measurements on a randomly selected

part with mean μ have the joint distribution given in (2.4).

$$\begin{pmatrix} Y_{i,1} \\ Y_{i,2} \\ \vdots \\ Y_{i,n} \end{pmatrix} \backsim N \left(\mu \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \Sigma_n = \sigma_t^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & & \\ \vdots & & \ddots & \vdots \\ \rho & & \dots & 1 \end{bmatrix} \right) \tag{2.4}$$

Observations from the k different parts are independent of each other. Thus, combining the likelihoods across k parts, we obtain the log-likelihood

$$\begin{aligned} l_{s1}(\mu, \sigma_t^2, \rho) &= -\frac{k}{2} \left\{ n \log \sigma_t^2 + (n-1) \log(1-\rho) + \log\left[\rho(n-1)+1\right] \right\} \\ &\quad - \frac{\left\{ (1+(n-1)\rho) SSW + (1-\rho)\left(SSA + nk(\overline{y}_{\cdot\cdot} - \mu)^2\right) \right\}}{2\sigma_t^2 (1-\rho)\left[\rho(n-1)+1\right]} \end{aligned} \tag{2.5}$$

where SSW, SSA and $\overline{y}_{\cdot\cdot}$ are defined in Table 2.1. The maximum likelihood estimates for μ, σ_t^2 and ρ are

$$\begin{aligned} \widehat{\mu} &= \overline{y}_{\cdot\cdot}, \\ \widehat{\sigma}_t^2 &= \frac{SSW + SSA}{nk}, \qquad \text{and} \\ \widehat{\rho}_s &= \frac{MSA(k-1) - MSWk}{MSWk(n-1) + MSA(k-1)} \\ &= \frac{MSA\frac{(k-1)}{k} - MSW}{MSW(n-1) + MSA\frac{(k-1)}{k}} \\ &\approx \frac{MSA - MSW}{MSW(n-1) + MSA} \end{aligned}$$

Thus for large k, ML and ANOVA, shown in (2.2), yield the same estimates for ρ.

If we use ML, we can obtain the asymptotic variance-covariance matrix of the estimators using the Fisher information. The Fisher information is the negative of the expectation of the second derivative of the log-likelihood with respect to the

parameters. For the SP, the Fisher information, $J_{s1}\left(\mu, \sigma_t^2, \rho\right)$ is

$$\begin{pmatrix} \frac{nk}{\sigma_t^2[1+\rho(n-1)]} & 0 & 0 \\ 0 & \frac{kn}{2\sigma_t^4} & -\frac{nk\rho(n-1)}{2\sigma_t^2(1-\rho)[1+\rho(n-1)]} \\ 0 & -\frac{nk\rho(n-1)}{2\sigma_t^2(1-\rho)[1+\rho(n-1)]} & \frac{1}{2}\frac{nk(n-1)\left[\rho^2(n-1)+1\right]}{(1-\rho)^2[1+\rho(n-1)]^2} \end{pmatrix} \tag{2.6}$$

The asymptotic variances of the maximum likelihood estimators (MLEs) are the diagonal elements of the inverse of the Fisher information matrix. For the MLE of ρ, this is

$$Var\left(\tilde{\rho}_s\right) = \frac{2(1-\rho)^2\left[1+\rho(n-1)\right]^2}{kn(n-1)}. \tag{2.7}$$

Now, suppose we are planning an SP and are given the total number of measurements, N and a hypothesized value of ρ. In this situation we have choice on how to design the SP; that is we get to choose n and k. We call a design optimal if it yields a MLE with the smallest variance. To minimize (2.7) with respect to n and k, we set $n = N/k$ and then optimize. The plan with the lowest asymptotic variance has

$$k = \frac{\rho}{1+\rho}N\ . \tag{2.8}$$

Inserting these optimal k and n into the asymptotic variance (2.7), we have for the optimal plan

$$Var\left(\tilde{\rho}\right) = \frac{4(1-\rho)^2(1+\rho)^4}{\rho N^3}.$$

Note however that this optimal variance may not be achievable because the design parameters k and n must be positive integers.

2.2 A Measurement System with 100% Inspection

We examine two methods to incorporate baseline information into measurement system assessments. Both methods assume σ_t and μ are known or estimated with neglible error from the available inspection data.

2.2.1 Analysis

We use ANOVA and ML for estimation when the baseline parameters are known. Practitioners often do not recognize that there is substantial value in making use of the known process characteristics μ and σ_t. In the next section, we demonstrate the considerable value of this information.

In the standard plan, as described in Section 2.1, we estimated both σ_p and σ_m from the measurement investigation. To incorporate the known parameters μ and σ_t into the ANOVA method, we utilize the equations $\rho = \frac{\sigma_p^2}{\sigma_p^2+\sigma_m^2}$ and $\sigma_t^2 = \sigma_p^2 + \sigma_m^2$. Since σ_t is now known we only need to estimate σ_m in the measurement investigation. By estimating σ_m^2 using MSW from Table 2.1 and rearranging the equations, we obtain the estimate

$$\widehat{\rho}_a = 1 - \frac{\sum_{i=1}^{k}\sum_{j=1}^{n}\left(\overline{y}_{ij} - \overline{y}_{i.}\right)^2}{k(n-1)\sigma_t^2} = 1 - \frac{\text{MSW}}{\sigma_t^2}. \tag{2.9}$$

The corresponding estimator denoted $\tilde{\rho}_a$, is unbiased and has variance

$$\frac{2(1-\rho)^2}{k(n-1)}. \tag{2.10}$$

A benefit of this estimator is that it does not depend on how the parts were selected. This means the parts selected for the measurement study do not have to be

representative of the process.

Since the error terms $\{E_{ij}\}$ in model (2.1) are assumed to be normally distributed with mean zero and variance σ_m^2, the statistic

$$\frac{k(n-1)\text{MSW}}{\sigma_t^2(1-\rho)} = W \sim \chi^2_{k(n-1)}. \tag{2.11}$$

The critical region for testing if ρ is less than or equal to ρ_0 is

$$MSW \geq (1-\rho_0)\,\sigma_t^2 \chi^2_{\alpha,k(n-1)}.$$

where $\chi^2_{\alpha,k(n-1)}$ is the α quantile from a chi-square distribution with $k(n-1)$ degrees of freedom. The critical region, R, is the region such that if $\hat{\rho} \in R$ then the null hypothesis is rejected. The size of a test is defined as $P\left(\tilde{\rho}_a \in R\right) = \alpha$ when $\rho = \rho_0$. To find the critical region of (1.19) with size α we need to determine c in

$$\begin{aligned} P\left(\tilde{\rho}_a \ > \ c\right) &= \alpha \\ P\left(W \leq \frac{(1-c)k(n-1)}{(1-\rho_0)}\right) &= \alpha \end{aligned}$$

Solving for c we have $c = 1 - \frac{1-\rho_0}{k(n-1)}\chi^2_{\alpha,k(n-1)}$

The power of this test is the probability the estimate is in the critical region when the parameter ρ is different than the null value of ρ_0. The power when $\rho = \rho_1$ is

$$\begin{aligned} & P\left(\tilde{\rho}_a \ > \ 1 - \frac{1-\rho_0}{k(n-1)}\chi^2_{k(n-1)}\right) \\ = \ & P\left(1 - \frac{1-\rho_1}{k(n-1)}W \ > \ 1 - \frac{1-\rho_0}{k(n-1)}\chi^2_{k(n-1)}\right) \\ = \ & P\left(W \ \leq \ \frac{1-\rho_0}{1-\rho_1}\chi^2_{\alpha,k(n-1)}\right) \end{aligned} \tag{2.12}$$

Above, we modified the ANOVA-based analysis of the SP to include known values of μ and σ_t^2. The estimator $\widetilde{\rho}_a$ does not depend on how we sampled the parts in measurement system study. However if we assume the parts were randomly sampled, then we can use ML to estimate the parameters. To apply ML we need to determine the distribution of the repeated measurements.

Assuming (2.1) and normality, n measurements on the same randomly selected part have the joint distribution given in (2.4). Observations from the k different parts are independent of each other. Thus, combining the likelihoods across k parts, we obtain the log-likelihood

$$\begin{aligned} l_{s2}(\rho) &= -\frac{k}{2}\left\{n\log\sigma_t^2 + (n-1)\log(1-\rho) + \log\left[\rho(n-1)+1\right]\right\} \\ &\quad -\frac{1}{2\sigma_t^2(1-\rho)\left[\rho(n-1)+1\right]}\left(\left[\rho(n-1)+1\right]SSW^* - \rho n^2 SSA^*\right) \quad (2.13) \end{aligned}$$

where $SSW^* = \sum_{i=1}^{k}\sum_{j=1}^{n}\left(y_{ij}-\mu\right)^2$ and $SSA^* = \sum_{i=1}^{k}\left(\overline{y}_{i.}-\mu\right)^2$. The "*" signifies that we are using the known parameter μ and not $\overline{y}_{i.}$ and $\overline{y}_{..}$ as in SSW and SSA.

There is no closed form solution for the maximum likelihood estimate of ρ. We can however examine the asymptotic variance of the estimator with the Fisher information of the parameter. The Fisher information is

$$J_{s2}(\rho) = E\left[-\frac{\partial^2}{\partial\rho^2}\, l_1\left(\rho\right)\right] = \frac{1}{s2}\frac{nk(n-1)(\rho^2(n-1)+1)}{(1-\rho)^2(1+n\rho-\rho)^2}. \quad (2.14)$$

The asymptotic variance of the MLE is the inverse of the Fisher information. This means a larger value for the Fisher information is better.

Maximum likelihood large sample theory provides three ways to test the hypothesis in (1.19), the likelihood ratio test, the score test and the Wald test. The score test is based on the score function (2.15) which is the first derivative of the

log-likelihood function.

$$S_{s2}(\rho; Y) = \frac{\partial}{\partial \rho}\, l_{s2}\,(\rho) \tag{2.15}$$

The score test has critical region

$$R = \{\ y;\ S_{s2}(\rho_0 :\ y)\,[J_{s2}(\rho_0)]^{-1/2}\ >\ Z_{1-\alpha}\ \}$$

where α is the size of the test and $Z_{1-\alpha}$ is the $1-\alpha$ quantile from the standard normal distribution. The critical region is determined by the asymptotic distribution of the score function in maximum likelihood theory. That is, as $n \to \infty$

$$S_{s2}(\rho_0; Y)\,[J_{s2}(\rho_0)]^{-1/2}\ \to_D\ \ Z \sim N(0,1). \tag{2.16}$$

To carry out the score test, we calculate $S_{s2}(\rho_0; Y)\,[J_{s2}(\rho_0)]^{-1/2}$ using the data and reject the null hypothesis if the observed value is greater than $Z_{1-\alpha}$.

To apply the Wald test, we suppose that $\tilde{\rho}$ is the MLE of ρ. Then, we have

$$W_{s2}(\rho_0; Y) = (\tilde{\rho} - \rho_0)\,[J_{s2}(\rho_0)]^{1/2} \to_D Z \sim N(0,1).$$

To perform the Wald test, we calculate the maximum likelihood estimate, calculate $\widehat{W}_{s2}(\rho_0; Y)$ and then reject the hypothesis $\rho \leq \rho_0$, if the calculated value of $\widehat{W}_{s2}(\rho_0; Y)$ is greater than $Z_{1-\alpha}$.

The approximate power of the Wald test when $\rho = \rho_1$ is

$$P\left\{Z > \left(Z_{1-\alpha}\,[J_{s2}(\rho_0)]^{-1/2} + \rho_0 - \rho_1\right)[J_{s2}(\rho_1)]^{1/2}\right\}. \tag{2.17}$$

We use the Wald test when comparing maximum likelihood estimation to the other methods.

2.2.2 The Value of the Baseline Information

We use statistical power to compare the three methods of testing the hypothesis (1.19): maximum likelihood, standard ANOVA and ANOVA incorporating baseline information. Figure 2.1 shows power curves for testing (1.19), when the plans consist of ten parts and six repeated measurements on each part, the default choice in Gauge R&R studies [see Automotive Industry Action Group, 2002].

When comparing power curves from different tests, we ensure they all have the same size $\alpha = 0.05$, which is the power when $\rho = \rho_0$.

The formulas (2.3), (2.12) and (2.17) were used to create Figure 2.1. These formulas determine the power for each of the following analyzes: ANOVA, ANOVA incorporating baseline information and MLE, respectively.

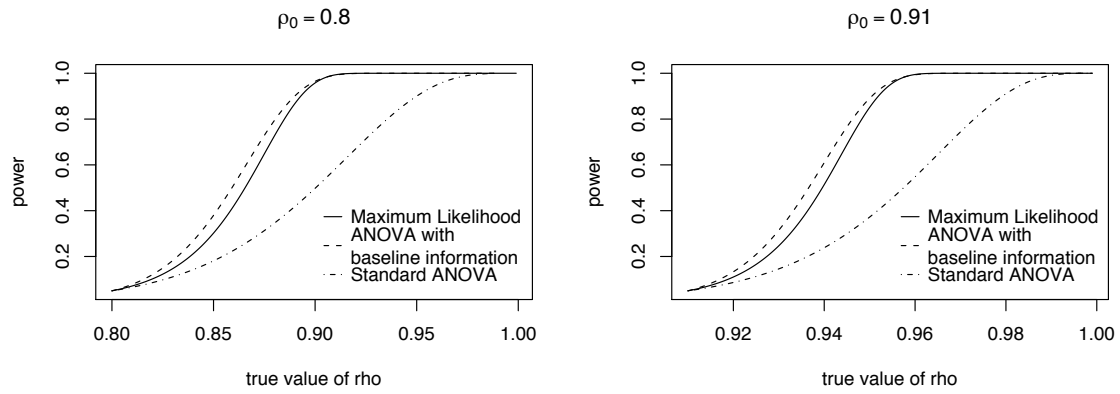

Figure 2.1: Power Curves for Testing (1.19) when $\rho_0 = 0.80$ and $\rho_0 = 0.91$,

Figure 2.1 indicates that Maximum Likelihood and ANOVA with baseline information are significantly more powerful than the standard ANOVA analysis when ρ_0 is 0.80 or 0.91. Clearly, if μ and σ_t^2 are known, this information should be used in these situations. The values of $\rho_0 = 0.80$ and 0.91 are important because they correspond to cutoff values for unacceptable measurement systems as suggested by the GRR and Discrimination Ratio in Table 1.1.

Under the assumption that the k parts are sampled randomly from the process, ML should be the most efficient method of estimation. However, Figure 2.1 shows that ANOVA with baseline information has a higher power curve than the MLE. This likely happened only because the Wald test used in Figure 2.1 is based on the asymptotic distribution of the MLE. For large sample sizes, the order would be reversed. We verified, using simulation that for $\rho \geq 0.80$ and finite sample sizes, maximum likelihood and ANOVA with baseline information are virtually equivalent in terms of power. We believe these two estimators are similar when ρ is close to one because most of the information about ρ is contained in MSW. We can infer this because the variances of MSW and MSA are proportional to $(1-\rho)^2$ and ρ^2 respectively.

In summary, we recommend the analysis that estimates ρ using ANOVA with baseline information. The approach has a closed form estimate for ρ and performs as well as Maximum Likelihood when ρ is larger than 0.80, which are the typical values of interest. In addition, the method does not require the parts selected for the measurement system study to be representative of the process.

2.3 A Measurement System with Operators

The standard measurement assessment plan (SP) in this context is to sample k parts selected at random from the process and then have each of the m operators measure each part n times for a total of $N = kmn$ measurements. The plan commonly uses two or three operators ($m = 2, 3$), each of whom measure the same $k = 10$ parts two or three ($n = 2, 3$) times for a total of 40 to 90 measurements [See Automotive Industry Action Group, 2002, Burdick et al., 2003]. .

The data from this SP is described using a mixed effect model given by

$$Y_{ijl} = \mu_j + X_i + E_{ijl} \tag{2.18}$$

where $i = 1, \ldots, k$, $j = 1, \ldots, m$, $l = 1, \ldots, n$, X_i is a random effect of the true part dimensions, μ_j is the mean effect from operator j, and E_{ijk} is the random effect from the repeatability or a single operator repeatedly measuring the same part. X and E are assumed to be independent normals with zero means and standard deviations σ_p, σ_g, respectively. The mean of the true part dimensions is included into the mean effect for each operator.

2.3.1 Analysis

The analysis for the (SP) in this context is based on the ANOVA which can found in Burdick et al. [2005]. The ANOVA is reproduced in Table 2.2.

Table 2.2: ANOVA for Fixed Effects Model (2.18)

Source of variation	Degrees of freedom	Mean square	Expected mean square
Parts (P)	$k-1$	S_P^2	$\theta_P = \sigma_g^2 + mn\sigma_p^2$
Operators (O)	$m-1$	S_O^2	$\theta_O = \sigma_g^2 + kn\left(\frac{m\sigma_O^2}{m-1}\right)$
Replicates	$kmn-k-m+1$	S_E^2	$\theta_O = \sigma_g^2$

The statistics in Table 2.2 are defined in Table 2.3. The ANOVA estimators for each parameter are defined in Table 2.4.

Burdick et al. [2005] focus on confidence intervals for parameters. We consider standard errors which are directly related to the length of the approximate confidence intervals.

To apply maximum likelihood, we form the likelihood using the parameter-

Table 2.3: Statistics for Standard Plan with Fixed Effects

Statistic	Definition
S_P^2	$\frac{mn\sum_i(\overline{Y}_{i**}-\overline{Y}_{***})^2}{k-1}$
S_O^2	$\frac{kn\sum_j(\overline{Y}_{*j*}-\overline{Y}_{***})^2}{m-1}$
S_E^2	$\frac{kn\sum_i\sum_j\sum_l(Y_{ijl}-\overline{Y}_{ij*})^2}{km(n-1)}$
$\overline{Y}_{i**}$	$\frac{\sum_j\sum_l Y_{ijl}}{mn}$
$\overline{Y}_{*j*}$	$\frac{\sum_i\sum_l Y_{ijl}}{kn}$
$\overline{Y}_{ij*}$	$\frac{\sum_l Y_{ijl}}{n}$
$\overline{Y}_{***}$	$\frac{\sum_i\sum_j\sum_l Y_{ijl}}{kmn}$

Table 2.4: Estimators for Standard Plan with Fixed Effects

Parameter	Estimator
$\overline{\mu}$	$\overline{Y}_{***}$
μ_j	$\overline{Y}_{*j*}$
σ_p^2	$\widetilde{\sigma_p}^2 = \frac{S_P^2-S_E^2}{mn}$
σ_g^2	$\widetilde{\sigma_g}^2 = S_E^2$
σ_o^2	$\widetilde{\sigma_o}^2 = (S_O^2 - S_E^2)\frac{1}{kn}\frac{m-1}{m}$
σ_t^2	$\widetilde{\sigma}_p^2 + \widetilde{\sigma}_o^2 + \widetilde{\sigma}_g^2$
σ_m^2	$\widetilde{\sigma}_o^2 + \widetilde{\sigma}_g^2$
η	$\widetilde{\sigma}_p^2/\widetilde{\sigma}_t^2$
λ	$\widetilde{\sigma}_g^2/\widetilde{\sigma}_m^2$

ization $\left\{\underline{\mu} = (\mu_1, \ldots, \mu_m), \sigma^2_{pg} = \sigma^2_p + \sigma^2_g, \rho = \frac{\sigma^2_p}{\sigma^2_p+\sigma^2_g}\right\}$ because it simplifies the likelihood. Note, $\underline{\mu}$ is defined in (1.9) and represents the vector of operator means. The maximum likelihood estimates are found by maximizing the likelihood,

$$\begin{aligned} l_{s3}(\mu, \sigma^2_{pg}, \rho) &= -\frac{1}{2}k\Big[nm\log(\sigma^2_{pg}) + (nm-1)\log(1-\rho) + \log(1+(nm-1)\rho)\Big] \\ &\quad -\frac{1}{2}\frac{(1+(nm-1)\rho)SSW' - \rho(nm)^2 SSA'}{\sigma^2_{pg}(1-\rho)(1+(nm-1)\rho)} \end{aligned} \tag{2.19}$$

$$\begin{aligned} \text{where} \quad SSW' &= \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{k=1}^{n}(y_{ijk} - \mu_j)^2 \\ SSA' &= \sum_{i=1}^{k}(\overline{y}_{i..} - \overline{\mu})^2 \\ \text{and} \quad \overline{y}_{i..} &= \sum_{j=1}^{m}\sum_{k=1}^{n} y_{ijk}/mn \end{aligned} \tag{2.20}$$

Now to get the MLE's for η and λ, we apply the appropriate transformations, given by

$$\eta = \frac{\rho\sigma^2_{pg}}{\sigma^2_o + \sigma^2_{pg}} \qquad \text{and} \qquad \lambda = \frac{\sigma^2_o}{\sigma^2_o + (1-\rho)\sigma^2_{pg}} \tag{2.21}$$

where σ^2_o is $\frac{1}{m}\sum_{j=1}^{m}(\mu_j - \overline{\mu})^2$ and $\overline{\mu} = \frac{1}{m}\sum_{j=1}^{m}\mu_j$.

The Fisher Information for the MLEs is

$$J_{s3}\left(\underline{\mu}, \sigma^2_{pg}, \rho\right) = \begin{bmatrix} \frac{kn(1-\rho+\rho nm-\rho n)}{\sigma^2_{pg}(1-\rho)(1+\rho(nm-1))} & \cdots & -\frac{\rho n^2 k}{\sigma^2_{pg}(1-\rho)(1+\rho(nm-1))} & 0 & 0 \\ & \ddots & \vdots & \vdots & \vdots \\ & & \frac{kn(1-\rho+\rho nm-\rho n)}{\sigma^2_{pg}(1-\rho)(1+\rho(nm-1))} & 0 & 0 \\ & & & \frac{knm}{2\sigma^2_{pg}} & \frac{-kmn(nm-1)\rho}{2\sigma^2_{pg}(1-\rho)(1+\rho nm-\rho)} \\ & & & & \frac{k(nm-1)nm(\rho^2(nm-1)+1)}{2(1-\rho)^2(1+\rho(nm-1))^2} \end{bmatrix}.$$

We have the Fisher information for $\left(\underline{\mu}, \sigma^2_{pg}, \rho\right)$ but we want the Fisher information for $(\underline{\mu}, \lambda, \eta)$. Fortunately, the Fisher information for $(\underline{\mu}, \lambda, \eta)$ can be written as

the matrix multiplication of $J_{s3}\left(\mu_1,\ldots,\mu_m,\sigma_{pg}^2,\rho\right)$ and a matrix

$$D=\begin{bmatrix} I_m & -\frac{2}{m}\frac{\lambda\eta+1-\lambda}{\lambda(\eta-1)}(\underline{\mu}-\overline{\mu}\underline{1}_m) & 0_m \\ 0_m^t & -\frac{\sigma_o^2}{\lambda^2(1-\eta)} & \frac{\eta(1-\eta)}{(\lambda\eta+1-\lambda)^2} \\ 0_m^t & \frac{\sigma_o^2}{\lambda(1-\eta)^2} & \frac{1-\lambda}{(\lambda\eta+1-\lambda)^2} \end{bmatrix} \tag{2.22}$$

where $\overline{\mu}$ and σ_o^2 are defined in (1.10) and (1.14), respectively. The matrix D is formed by column vectors which are partial derivatives of the vector function

$$(\underline{\mu},\sigma_{pg}^2,\rho)=\underline{h}(\underline{\mu},\lambda,\eta)=\underline{h}(\mu_1,\ldots,\mu_m,\frac{\sigma_o^2(\lambda\eta+1-\lambda)}{\lambda(1-\eta)},\frac{\eta}{(\lambda\eta+1-\lambda)}) \tag{2.23}$$

with respect to $(\underline{\mu},\lambda,\eta)$. For example, the first column of D is the gradient vector $\frac{\partial\underline{h}(\underline{\mu},\lambda,\eta)}{\partial\mu_1}$. Note, the vector function $\underline{h}$ is mapping from $(\underline{\mu},\lambda,\eta)$ to $\left(\underline{\mu},\sigma_{pg}^2,\rho\right)$ and that σ_o^2 is a function of $(\mu_1,\mu_2,\ldots,\mu_m)$.

Then, the Fisher information for $(\underline{\mu},\lambda,\eta)$ is

$$J_{s3}\left(\underline{\mu},\lambda,\eta\right)=DJ_{s3}\left(\mu_1,\ldots,\mu_m,\sigma_{pg}^2,\rho\right)D^t. \tag{2.24}$$

Finally, to obtain asymptotic variance-covariance matrix for $(\underline{\mu},\lambda,\eta)$ we take the inverse of $J_{s3}\left(\underline{\mu},\lambda,\eta\right)$. We use these results when we compare the SP to the proposed leveraged plan in Chapter 6.

Chapter 3

Leveraged Plan and Why it Works

In the first context, a leveraged plan is conducted in two stages: In Stage 1, we sample b parts at random from the process and measure each part once to obtain a baseline. In Stage 2, from the baseline sample, we select k extreme parts using the observed measured values. Then the k selected parts are repeatedly measured n times each.

We use the term leveraging because of the re-use of units with extreme values. In short, leveraging is a method of non-random sampling where we select extreme units using a measured response.

In a different context, the term leverage is sometimes used by the proponents of the problem solving system initially proposed by Dorian Shainin. Units with relatively large and small values of the response are compared to identify the major causes of the variation. See Steiner et al. [2008] for a more complete description. However, there is no theoretical discussion in the literature regarding leveraging methodology or its application to measurement system assessment. In this chapter, we provide motivation for why leveraging is valuable.

3.1 Leveraging and Regression

The measurement problem can be translated to a regression context if we let the covariate x be the first measurement on a part and the response y be the second measurement. The slope of the regression line is the correlation ρ, defined in (1.5), because each measurement has the same variability.

To mathematically demonstrate that a measurement system assessment can be translated into a regression context, we use the properties of the bivariate normal distribution. Using model (1.1) the distribution of two measurements on the same part Y_1, Y_2 is

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \backsim N\left(\mu \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \Sigma = \sigma_t^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

and then by conditioning on Y_1 we obtain

$$Y_2 | \left(Y_1 = y_1\right) \backsim N\left(\mu + \rho\left(y_1 - \mu\right), \quad \sigma_t^2\left(1+\rho\right)\left(1-\rho\right)\right).$$

Since, Y_2 depends on ρ linearly through the mean we can use regression to estimate this parameter.

Suppose that μ and σ_t are known and we are interested in estimating ρ using the slope of a regression line between the 1^{st} measurement and the 2^{st} measurement. In addition suppose we have already measured 100 different parts once but now we can only afford to remeasure 10 parts. We consider two methods of selecting the 10 parts;

- select 10 parts at random or
- choose the 10 parts associated with the five smallest and five largest initial measurements.

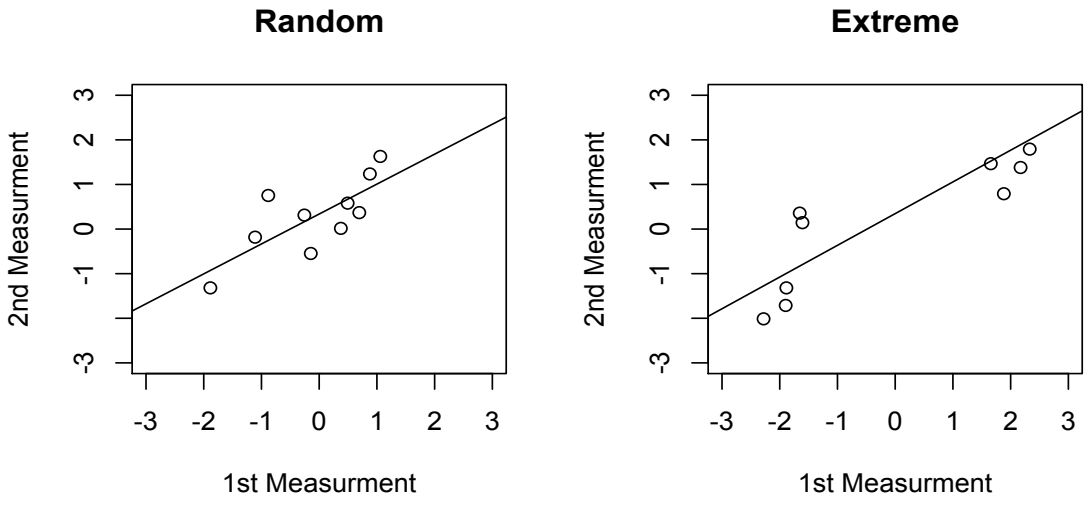

Figure 3.1: Left panel; 2 measurements on 10 parts chosen at random. Right panel; 10 parts the most extreme initial measurements chosen non-randomly.

These two sampling methods were simulated and the measured values are shown in Figure 3.1. It is known that the second sampling method provides an estimator with a smaller standard error because we chose extreme initial measurements or x values [Montgomery et al., 2001]. To minimize the standard error of the regression estimator for β we would remeasure parts that had the most extreme initial measurements relative to their average.

This example illustrates that if given the opportunity, we should sample parts non-randomly instead of randomly. But what if we do not have 100 parts already measured? Suppose instead we had the resources for $2N$ measurements and assuming 1^{st} and 2^{nd} measurements cost the same, then we consider two sampling plans

Plan 1. randomly sample and measure N parts twice $(y_{i1}, y_{i2}, i = 1, \ldots, N)$,

Plan 2. randomly sample b parts and measure each part once $(y_{i1}, i = 1, \ldots, b)$.

Then from those b parts, non-randomly sample, by choosing k parts with

extreme measurements and remeasure the selected parts. Here b and k are chosen such that $b \geq k$ and $b + k = 2N$. We specify the k selected parts with the set S which is a subset of $\{1, \ldots, b\}$. The additional data is $(y_{i2}, i \in S)$.

A simple linear regression model is

$$Y = \beta X + E$$

where $X \sim N(0, \sigma_x^2)$ and $E \sim N(0, \sigma_e^2)$. In general, if we have m observations of (y_i, x_i) and estimate β using least squares regression, the conditional variance for the estimator $\widetilde{\beta}$ is

$$\text{Var}\left(\widetilde{\beta} \mid x_i, i \in S\right) = \frac{\sigma_e^2}{\sum_{i \in S} (x_i - \overline{x})^2}.$$

Noting that the 1^{st} measurements are the covariate x in a regression context, both sampling plans have the above conditional variance with $\sigma_e^2 = \sigma_t^2(1+\rho)(1-\rho)$ and $\sigma_x^2 = \sigma_t^2$.

When we consider the first sampling plan, the unconditional variance is

$$\text{Var}\left(\widetilde{\beta}_1\right) = \sigma_e^2 E\left[\frac{1}{\sum_{i=1}^{N} \left(Y_{i1} - \overline{Y}_{\cdot 1}\right)^2}\right] = \frac{\sigma_e^2}{\sigma_x^2} \frac{1}{N-2}$$

because the parts where randomly sampled and measured twice. For the second sampling plan the unconditional variance is more complicated because we chose parts with the most extreme 1^{st} measurements. We have

$$\text{Var}\left(\widetilde{\beta}_2\right) = \sigma_e^2 E\left[\frac{1}{\sum_{i \in S} \left(Y_{i1} - \overline{Y}_{i1}\right)^2}\right] = \frac{\sigma_e^2}{\sigma_x^2} E\left[\frac{\sigma_x^2}{\sum_{i \in S} \left(Y_{i1} - \overline{Y}_{i1}\right)^2}\right]$$

where $\overline{Y}_{i1} = \sum_{i \in S} Y_{i1}/k$ is the average of the 1^{st} measurements in S.

If we fix $2N = 100$ and set $k = 100 - b$ then $\text{Var}\left(\widetilde{\beta}_2\right)$ depends only on b. Note, b is bounded between N and $2N$ and when $b = N$ then $\text{Var}\left(\widetilde{\beta}_2\right) = \text{Var}\left(\widetilde{\beta}_1\right)$. The

ratio, Var $\left(\widetilde{\beta}_1\right)$ /Var $\left(\widetilde{\beta}_2\right)$, is shown as a function of b in the left panel of Figure 3.2. The right panel shows a contour plot of this ratio while varying b and k and setting $N = (b+k)/2$ in Var $\left(\widetilde{\beta}_1\right)$ to ensure both plans have the same total number of measurements. For example at $b = 100$ and $k = 50$, the Var $\left(\widetilde{\beta}_1\right)$ is calculated with $N = 75$. Note $k \leq b$ because we cannot select more than b parts. Figure 3.2 shows that when $\frac{b}{b+k} \approx 0.70$, the second sampling plan yields estimators with smaller standard errors.

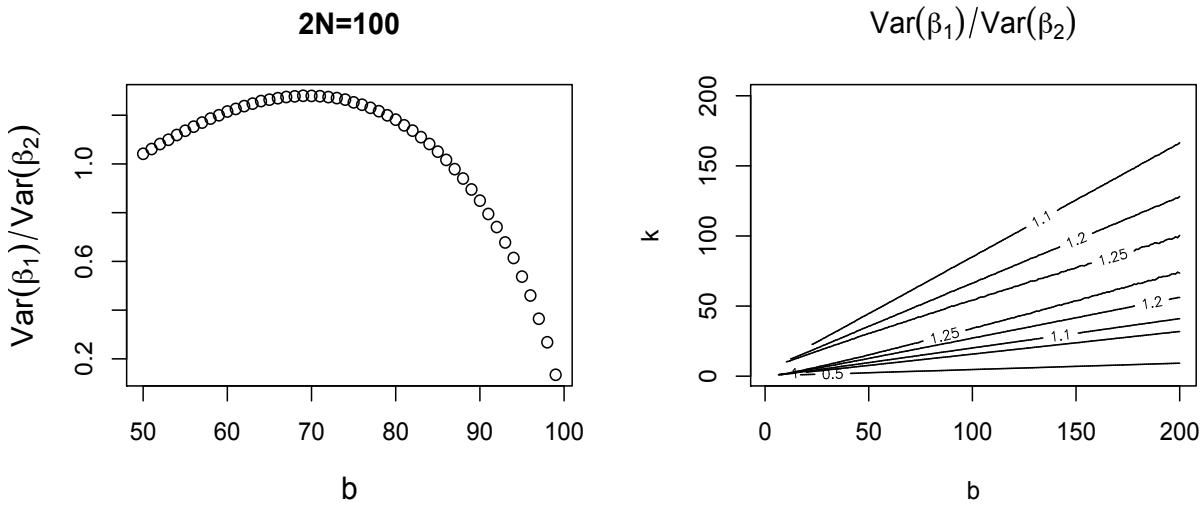

Figure 3.2: The ratio Var $\left(\widetilde{\beta}_1\right)$ /Var $\left(\widetilde{\beta}_2\right)$, when both plans have the same sample size.

In summary, if we sample parts with extreme first measurements (i.e. use leveraging) rather than sampling randomly, we can reduce the standard error for the regression estimator of ρ.

3.2 Known True Value Versus an Extreme Initial Measurement

When we submitted our first paper on Leveraging, we got a quick response from the Associate Editor. It read: "On its face, the proposed method [leveraging] cannot work as well as claimed. Two extreme cases are (1) where the true value for each of the parts is known, and (2) where no prior information about these true values is available. The proposed method is one in which some prior information is available. In assessing the measurement variance, the difference between the extreme cases is one degree of freedom for each of the parts. Thus, the proposed method cannot save more than one measurement per part. "

To refute the Associate Editor's erroneous comments, we consider estimating ρ with a single measurement. We assume model (2.1) holds and that we know σ_t and μ. Now we propose two options. We can choose

1. a part with known true dimension x or
2. a part with initial measurement y.

If we pick option 1, the distribution of the single measurement M_1 on a part with known value x is

$$M_1 | \left(X = x\right) \backsim N\left(x, \quad \sigma_t^2 \left(1 - \rho\right)\right).$$

Here, a natural estimator for ρ is

$$\widetilde{\rho}_1 = \left[1 - \frac{1}{\sigma_t^2} \left(M_1 - \mu\right)^2\right] \sim \left(1 - \rho\right) \chi_1^2.$$

If we pick option 2, the distribution of the second measurement M_2 on a part

that has an initial measurement, y, is

$$M_2|\,(Y = y) \backsim N\left(\mu + \rho\,(y - \mu)\,, \quad \sigma_t^2\,(1 + \rho)\,(1 - \rho)\right).$$

Since, σ_t^2 and μ are known we can standardize the initial measurement y with $y = \sigma_t z + \mu$. A regression estimator for the second option is

$$\widetilde{\rho}_2 = \frac{(M_2 - \mu)}{(y - \mu)} \sim N\left(\rho, \quad \frac{(1 + \rho)\,(1 - \rho)}{z^2}\right)$$

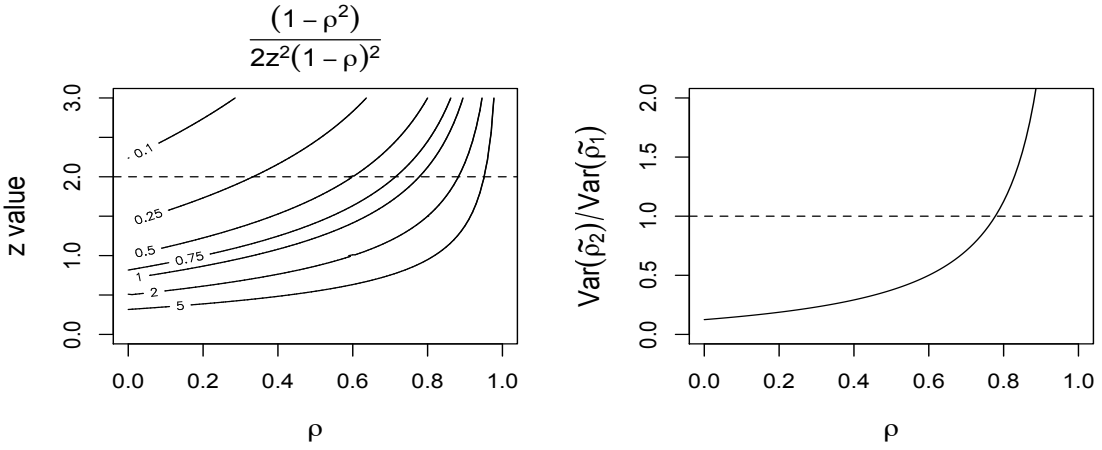

Figure 3.3: Plots of $\mathrm{Var}(\widetilde{\rho}_2)/\mathrm{Var}(\widetilde{\rho}_1) = [(1 - \rho^2)/z^2] \,/\, [2(1 - \rho)^2]$, the left panel is a contour plot by z and ρ and the right panel displays the quantity when $z = 2$.

We display the ratio of variances $Var(\widetilde{\rho}_2)/Var(\widetilde{\rho}_1)$ in Figure 3.3. The left panel is a contour plot of the ratio by z and ρ. In this plot, a value less than one means that having a part with a standardized initial measurement, z, is better than knowing the true dimension x. This demonstrates that the AE's intuition was not true for all ρ. The value 0.25 means that the variance of option 2 is four times smaller than option 1. The right panel displays the contour line along $z = 2$. When $\rho \geq 0.8$, option 1 is more efficient, but in the next chapter, we will show that by making repeated measurements on the same part, leveraging (i.e. option 2) also works well

for large values of ρ.

Leveraging, in the measurement system context, is defined as purposely selecting parts with extreme initial measurement to repeatably measure. In this demonstration of the value of leveraging we have assumed μ and σ_t were known. In cases where we do not know these parameters, the leveraged plan will include a baseline. A baseline involves randomly sampling parts and measuring them once. This will allow us to estimate the overall parameters σ_t^2 and μ. We present a leveraged plan in the following chapters for each of the contexts of interest. These plans, with increased efficiency over the plans currently in use, are all new to the literature.

Chapter 4

Leveraged Plan for a Measurement System

4.1 Context

In this chapter we compare the standard plan (SP) (as described in Section 2.1) to a leveraged plan (LP) when the parameters μ, σ_t^2 and ρ are unknown. This context occurs when assessing a new measurement system and when we assess an existing measurement system ignoring any information about the process parameters. In practice this is the most common situation.

The leveraged measurement system assessment plan is conducted in two stages:

Stage 1: Sample b parts from the process randomly (chosen to be representative of the process) to obtain a baseline. We denote the observed values for these measurements as $\{y_{10}, y_{20}, \ldots, y_{b0}\}$ and the baseline average and sample variance by $\overline{y}_b = \frac{1}{b}\sum_{i=1}^{b} y_{i0}$ and $s_b^2 = \frac{1}{b-1}\sum_{i=1}^{b}(y_{i0} - \overline{y}_b)^2$.

Stage 2: From the baseline sample, we select k parts (non randomly) using the observed measured values. In particular, to improve our estimate for ρ, we

sample k parts that are extreme relative to the baseline average, $\overline{y}_b$. We denote the k selected parts using the set S. These k parts are then repeatedly measured n times each to give the additional data $\{y_{ij}, i \in S \quad \& \quad j = 1, \ldots, n\}$. The total number of measurements for the leveraged plan is $N = b + nk$.

For example, for a leveraged plan with $k = 2$ we may pick the parts with the minimum and maximum initial measurement. If the largest and smallest parts from the baseline indexed as parts 5 and 11 we have $S = \{5, 11\}$.

We recommend repeatedly measuring the parts in Stage 2 over the range of conditions (operators/time, environment, etc.) expected to capture the major sources of measurement variation. Note that this recommendation matches the requirements for a standard plan.

As stated above, we assume the total variation σ_t^2 and the process mean μ are unknown but interest lies in estimating ρ. The parameters μ and σ_t are viewed as nuisance parameters.

This chapter has the following structure. In the next section, we describe the analysis for an LP, including properties of the MLE for ρ and other simpler estimators. In Section 4.3, we compare different designs (i.e. different values of b, k and n) for leveraged plans when the total sample size is fixed. Based on empirical evidence, we recommend specific plans for any total sample size. In Section 4.5, we compare the derived estimators of ρ from standard and leveraged plans using the bias and standard deviation.

4.2 Estimation

We present four approaches for estimation of μ, σ_t^2 and ρ. The first method uses Maximum Likelihood. The MLEs have no closed form and must be found numerically. The other three methods estimate μ and σ_t^2 with the baseline information only and then estimate ρ using the repeated measurements, conditional on the baseline observations. The second estimate uses a regression approach since the conditional mean of the repeated measurements depends on ρ. The third uses the variation within the repeated measurements to estimate ρ. Finally, the fourth estimate is a combination of the second and third estimates.

4.2.1 Maximum Likelihood

We decompose the LP likelihood into two pieces by conditioning on the baseline measurements. These two pieces are the baseline log-likelihood, denoted by

$$l_{b1}\left(\mu, \sigma_t^2; y_{10}, \ldots, y_{b0}\right)$$

and the log-likelihood of the repeated measurements conditional on the baseline measurement denoted by

$$l_{r1}\left(\mu, \sigma_t^2, \rho;\ y_{ij}, i \in S\ \ \&\ \ j = 1, \ldots, n \,\middle|\, y_{10}, \ldots, y_{b0}\right).$$

The b parts in the baseline sample are selected at random from the process, so the baseline log-likelihood is

$$l_{b1}(\mu, \sigma_t^2) = -\frac{b}{2}\log\sigma_t^2 - \frac{1}{2\sigma_t^2}\left\{(b-1)s_b^2 + b\left(\overline{y}_b - \mu\right)^2\right\}.$$

Now, to obtain l_{r1} we start with a single part. For a single part (selected to be

repeatedly measured), the joint distribution of the initial measurement Y_0 and the n repeated measurements $\{Y_1, \ldots, Y_n\}$ is

$$\begin{pmatrix} Y_0 \\ Y_1 \\ \vdots \\ Y_n \end{pmatrix} \backsim N \left(\mu \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \Sigma_{n+1} = \sigma_t^2 \begin{bmatrix} 1 & \rho & \ldots & \rho \\ \rho & 1 & & \\ \vdots & & \ddots & \vdots \\ \rho & & \ldots & 1 \end{bmatrix} \right). \tag{4.1}$$

where $\sigma_t^2 = \sigma_p^2 + \sigma_m^2$ is the total variation.

Using the properties of the multivariate normal, the distribution of the repeated measurements $\{Y_1, \ldots, Y_n\}$ on a single part given the initial measurement $Y_0 = y_0$ is

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \Bigg| Y_0 = y_0 \backsim N \left([\mu + \rho(y_0 - \mu)] \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \ \Sigma_c \right) \tag{4.2}$$

where the covariance matrix,

$$\Sigma_c = \sigma_t^2 \begin{bmatrix} 1-\rho^2 & & \rho(1-\rho) \\ & \ddots & \\ \rho(1-\rho) & & 1-\rho^2 \end{bmatrix}, \tag{4.3}$$

has a special form which allows us to obtain the following well known properties [Dillon and Goldstein, 1984, reprinted in Appendix C]:

$$\Sigma_c^{-1} = \frac{1}{\sigma_t^2(1-\rho)(1+n\rho)} \begin{bmatrix} 1+\rho(n-1) & & -\rho \\ & \ddots & \\ -\rho & & 1+\rho(n-1) \end{bmatrix}$$

$$|\Sigma_c| = \sigma_t^{2n}(1-\rho)^n(1+n\rho)$$

Using the properties of Σ_c, we can write down the conditional likelihood (conditional on y_0) for the repeated measurements on a single part. The measurements for one part are independent of the measurements from another part, so the conditional likelihood for k parts, each with n measurements, is the product of their likelihoods. Thus, the conditional log-likelihood for n repeated measurements on k parts is

$$\begin{aligned} l_{r1}\left(\mu, \sigma_t^2, \rho \,|y_{10}, \ldots, y_{k0}\right) &= -\frac{nk}{2}\log\sigma_t^2 - \frac{nk}{2}\log(1-\rho) - \frac{k}{2}\log(1+n\rho) \\ &\quad -\frac{1}{2}\frac{1}{\sigma_t^2(1-\rho)(1+n\rho)} \times \\ &\quad \left\{(1+n\rho)SSW + n\sum_{i=1}^{k}\left[\overline{y}_{i.} - \mu - \rho(y_{i0}-\mu)\right]^2\right\} \end{aligned} \tag{4.4}$$

where y_{i0} is the baseline measurement for the i^{th} part, $\overline{y}_{i.} = \frac{1}{n}\sum_{i=1}^{n} y_{ij}$ is the average of the repeated measurements for the i^{th} part and $SSW = \sum_{i\in S}\sum_{j=1}^{n}(y_{ij} - \overline{y}_{i.})^2$.

Finally, we add (4.1) and (4.4) to get the (unconditional) log-likelihood for the LP:

$$l_{L1}(\mu, \sigma_t^2, \rho) = l_{b1}\left(\mu, \sigma_t^2\right) + l_{r1}\left(\mu, \sigma_t^2, \rho \big| y_{10}, \ldots, y_{b0}\right) \tag{4.5}$$

To get the MLEs of μ, σ_t^2 and ρ, we can numerically maximize (4.5). In theorem 1, we prove that this likelihood is appropriate regardless of the part selection method in Stage 2.

Theorem 1. *If $Y_{ij} = P_i + E_{ij}$ where $P_i \sim N\left(0, \sigma_p^2\right)$, $E_{ij} \sim N\left(0, \sigma_m^2\right)$ $i = 1, 2, \ldots, b$ and $j = 0, 1, 2, \ldots, n$, then if we sample $\{Y_{10}, \ldots, Y_{b0}\}$ and order them such that $\left\{Y_{(1)0} \leq \ldots \leq Y_{(b)0}\right\}$ then the conditional distribution of $\left\{Y_{(i)1}, \ldots, Y_{(i)n} \big| Y_{(i)0} = y_{(i)0}\right\}$ where $i \in S$ and S is a subset of $\{1, \ldots, b\}$ is given by (4.2).*

Proof: We begin with a joint distribution assuming we had repeated measure-

ments on every part in the baseline.

$$
\begin{aligned}
& f\left(y_{10}, y_{i1}, \ldots, y_{1n}, y_{20}, \ldots, y_{2n}, \ldots, y_{b0}, \ldots, y_{bn}\right) \\
= & \prod_{i=1}^{b} f\left(y_{i0}, y_{i1}, \ldots, y_{in}\right) \quad \left(\text{ the distribution defined in (4.1) }\right) \\
= & \prod_{i=1}^{b} f\left(y_{i1}, \ldots, y_{in} \middle| y_{i0}\right) f\left(y_{i0}\right) \quad \left(\text{ the distribution defined in (4.2) }\right) \\
= & \prod_{i=1}^{b} f\left(y_{i1}, \ldots, y_{in} \middle| y_{i0}\right) \prod_{i=1}^{b} f\left(y_{i0}\right) \\
& \text{do a change of variables such that } \; y_{(1)0} \leq y_{(2)0} \leq \ldots \leq y_{(b)0} \\
= & \left[\prod_{k=1}^{b} f\left(y_{k1}, \ldots, y_{kn} \middle| y_{(k)0}\right)\right] \; n! \; \prod_{k=1}^{b} f\left(y_{(k)0}\right) \\
= & \; n!\left[\prod_{k=1}^{b} f\left(y_{k1}, \ldots, y_{kn} \middle| y_{(k)0}\right) f\left(y_{(k)0}\right)\right] \\
= & \; n!\left[\prod_{k=1}^{b} f\left(y_{k1}, \ldots, y_{kn}, y_{(k)0}\right)\right] \\
& \text{if } k \notin S \text{ integrate } \; \left\{y_{k1}, \ldots, y_{kn}\right\} \\
= & \; n!\left[\prod_{k \notin S} f\left(y_{(k)0}\right)\right]\left[\prod_{k \in S} f\left(y_{k1}, \ldots, y_{kn}, y_{(k)0}\right)\right] \\
= & \; n!\left[\prod_{k \notin S} f\left(y_{(k)0}\right)\right]\left[\prod_{k \in S} f\left(y_{k1}, \ldots, y_{kn} \middle| y_{(k)0}\right) f\left(y_{(k)0}\right)\right] \\
= & \; n!\left[\prod_{k=1}^{b} f\left(y_{(k)0}\right)\right]\left[\prod_{k \in S} f\left(y_{k1}, \ldots, y_{kn} \middle| y_{(k)0}\right)\right] \\
& \text{do a change of variables back to the original order} \\
= & \left[\prod_{i=1}^{b} f\left(y_{i0}\right)\right]\left[\prod_{i \in S} f\left(y_{i1}, \ldots, y_{in} \middle| y_{i0}\right)\right]
\end{aligned}
$$

We can see that this is the joint distribution of $\left\{Y_{(i)0}, Y_{(i)1} \ldots, Y_{(i)n}\right\}$. Thus, the conditional distribution of $\left\{Y_{(i)1}, \ldots, Y_{(i)n} \middle| Y_{(i)0}\right\}$ where $i \in S$ is (4.2). Note, the conditional distribution given a baseline measurement does not depend on the rank of the baseline measurement from a sample.

Theorem 1 can be explained in words by imagining that we had repeated measurements on every part in the baseline so the conditional distribution of repeats depends only on the initial measurement. Then selection has no effect through the value of the initial measurement. In other words, the conditional distribution of the repeats are independent of the baseline measurements given the initial measurement. So, conditional on the initial measurement, the rank of that measurement in the baseline does not matter. The part could have come from a baseline of size 100 or 1000.

Score Function

The score function for ρ, the partial derivative of l_{r1} with respect to ρ, can be written as a linear combination of three estimating functions,

$$S_\rho = \frac{\partial l_{r1}}{\partial \rho} = \Psi_1(\rho)c_1(\rho) + \Psi_2(\rho)c_2(\rho) + \Psi_3(\rho)c_3(\rho)$$

where

$$\begin{aligned}
\Psi_1(\rho) &= \left[(1-\rho)\sigma_t^2 k(n-1) - SSW\right] && (4.6)\\
\Psi_2(\rho) &= \left[\sum_{i=1}^{k} (\overline{y}_{i.} - \mu)(y_{i0} - \mu) - \rho \sum_{i=1}^{k} (y_{i0} - \mu)^2\right] && (4.7)\\
\Psi_3(\rho) &= \left[k\sigma_t^2 (1-\rho)\left(\rho + \frac{1}{n}\right) - \sum_{i=1}^{k} [\overline{y}_{i.} - \mu - \rho(y_{i0} - \mu)]^2\right] && (4.8)
\end{aligned}$$

and

$$\begin{aligned}
c_1(\rho) &= \frac{1}{2}\frac{1}{(1-\rho)^2\sigma_t^2},\\
c_2(\rho) &= \frac{n}{\sigma_t^2(1-\rho)(1+n\rho)},\\
c_3(\rho) &= \frac{1}{2}\frac{n(1+2n\rho-n)}{\sigma_t^2(1-\rho)^2(1+n\rho)^2}.
\end{aligned}$$

These three estimating functions all have expectation zero and are important because later we see that each provides an alternative way to estimate ρ.

Fisher Information

We recommend choosing parts with extreme initial measurements because this decreases the asymptotic variance of the MLE of ρ. The asymptotic variance-covariance matrix of the maximum likelihood estimator is the inverse of the Fisher information matrix, which is

$$J_{L1}\left(\mu,\sigma_t^2,\rho\right)=\begin{pmatrix} \frac{(1-\rho)nk+b(n\rho+1)}{\sigma_t^2(n\rho+1)} & 0 & \frac{nE[SC]}{\sigma_t(n\rho+1)} \\ 0 & \frac{1}{2}\frac{b+nk}{\sigma_t^4} & -\frac{1}{2}\frac{nk\rho(n+1)}{\sigma_t^2(n\rho+1)(1-\rho)} \\ \frac{nE[SC]}{\sigma_t(n\rho+1)} & -\frac{1}{2}\frac{nk\rho(n+1)}{\sigma_t^2(n\rho+1)(1-\rho)} & E\left[-\frac{\partial^2}{\partial\rho^2}\, l_{L1}\left(\mu,\sigma_t^2,\rho\right)\right] \end{pmatrix} \tag{4.9}$$

$$\text{where}\quad E\left[-\frac{\partial^2}{\partial\rho^2}\, l_{L1}\left(\mu,\sigma_t^2,\rho\right)\right]=\frac{1}{2}\frac{kn(n+1)(n\rho^2+1)}{(1+n\rho)^2(1-\rho)^2}+\frac{n\left(E\left[SSC\right]-k\right)}{(1+n\rho)(1-\rho)}, \tag{4.10}$$

$$SSC=\sum_{i\in S}\left[\frac{Y_{i0}-\mu}{\sigma_t}\right]^2 \qquad \text{and} \qquad SC=\sum_{i\in S}\left[\frac{Y_{i0}-\mu}{\sigma_t}\right]. \tag{4.11}$$

It is shown in Appendix A.1 that if a sampling plan is chosen such that $E\left[SC\right]=0$, then the asymptotic variance of the MLE for ρ is reduced. Also, we show that the variance of the MLE is reduced by choosing a sampling plan where $E\left[SSC\right]$ is large. A plan with both these properties is to choose an equal number of parts with extreme initial measurements on either side of the baseline average.

When using maximum likelihood, standard errors for the estimates can be obtained from the inverted information matrix (see Appendix A.1) with the parameters replaced by their estimates.

4.2.2 Regression Estimator

Maximum Likelihood is an efficient method of estimation but when explicit expressions of the MLE cannot be found, it is useful to search for estimators with a closed form. The distribution of the average of the repeated measurements on a single part, given the initial measurement y_{i0}, is

$$\overline{Y}_{i.}\,|(Y_{i0} = y_{i0}) \backsim N\left(\mu + \rho(y_{i0} - \mu),\ \ \sigma_t^2(1-\rho)\left(\rho + \frac{1}{n}\right)\right) \tag{4.12}$$

The averages of the repeated measurements on different parts are mutually independent. We can use regression to estimate ρ because in (4.12), the mean depends on ρ linearly and the variance is the same for each part. The conditional mean of $\overline{Y}_{i.}$ also depends on μ but we use the baseline average $\overline{y}_b$ to estimate this unknown.

The regression estimate of ρ [Montgomery, Peck, and Vining, 2001] is

$$\widehat{\rho}_r = \frac{\sum_{i\in S}\left(\overline{y}_{i.} - \overline{y}_b\right)\left(y_{i0} - \overline{y}_b\right)}{\sum_{i\in S}\left(y_{i0} - \overline{y}_b\right)^2} \tag{4.13}$$

If we standardize each quantity on the right hand side in (4.13), the marginal distribution of $\tilde{\rho}_r$, the corresponding estimator of the regression estimate, depends only on ρ. The distribution of $\tilde{\rho}_r$, conditional on the baseline sample, is normal with mean

$$\begin{aligned}
& E\left[\widetilde{\rho}_r \middle| y_{10}, \ldots, y_{b0}\right] \\
= \ & \frac{\sum_{i\in S}\left(E\left[\overline{Y}_{i.}\middle| y_{10}, \ldots, y_{b0}\right] - \overline{y}_b\right)\left(y_{i0} - \overline{y}_b\right)}{\sum_{i\in S}\left(y_{i0} - \overline{y}_b\right)^2} \\
= \ & \rho\,\frac{\sum_{i\in S}\left(y_{i0} - \mu\right)\left(y_{i0} - \overline{y}_b\right)}{\sum_{i\in S}\left(y_{i0} - \overline{y}_b\right)^2} + \left(\mu - \overline{y}_b\right)\frac{\sum_{i\in S}\left(y_{i0} - \overline{y}_b\right)}{\sum_{i\in S}\left(y_{i0} - \overline{y}_b\right)^2} \\
= \ & \rho\,\frac{\sum_{i\in S}\left(y_{i0} - \overline{y}_b + \overline{y}_b - \mu\right)\left(y_{i0} - \overline{y}_b\right)}{\sum_{i\in S}\left(y_{i0} - \overline{y}_b\right)^2} + \left(\mu - \overline{y}_b\right)\frac{\sum_{i=1}^{k}\left(y_{i0} - \overline{y}_b\right)}{\sum_{i\in S}\left(y_{i0} - \overline{y}_b\right)^2}
\end{aligned}$$

$$
\begin{aligned}
&= \rho \frac{\sum_{i\in S}(y_{i0}-\overline{y}_b)^2 + (\overline{y}_b-\mu)\sum_{i\in S}(y_{i0}-\overline{y}_b)}{\sum_{i\in S}(y_{i0}-\overline{y}_b)^2} + (\mu-\overline{y}_b)\frac{\widehat{SC}}{\widehat{SSC}} \\
&= \rho\left(1+(\mu-\overline{y}_b)\frac{\widehat{SC}}{\widehat{SSC}}\right) + (\mu-\overline{y}_b)\frac{\widehat{SC}}{\widehat{SSC}} \\
&= \rho + \left[(\mu-\overline{y}_b)\frac{\widehat{SC}}{\widehat{SSC}}\right](1+\rho) \qquad (4.14)
\end{aligned}
$$

and variance

$$
Var\left[\tilde{\rho}_r \middle| y_{10},\ldots,y_{b0}\right] = \frac{\sigma_t^2(1-\rho)(1/n+\rho)}{\sum_{i\in S}(y_{i0}-\overline{y}_b)^2} = \frac{\sigma_t^2(1-\rho)(1/n+\rho)}{s_b^2\widehat{SSC}} \qquad (4.15)
$$

$$
\text{where} \quad \widehat{SC} = \sum_{i\in S}\left[\frac{y_{i0}-\overline{y}_b}{s_b}\right] \quad \text{and} \quad \widehat{SSC} = \sum_{i\in S}\left[\frac{y_{i0}-\overline{y}_b}{s_b}\right]^2 \qquad (4.16)
$$

are the baseline estimates of SC and SSC as defined in (4.11).

The estimator $\tilde{\rho}_r$ has a small bias (conditionally) if, as we recommend, we choose parts so that $\widehat{SC} \approx 0$ and $\widehat{SSC}$ is large. We also expect that $\overline{y}_b$ will be close to μ since the baseline sample is selected at random from the process. Unconditionally, $\tilde{\rho}_r$ is unbiased because $\overline{Y}_b$ is independent of the random variables corresponding to $\widehat{SC}$ and $\widehat{SSC}$.

The unconditional variance of $\tilde{\rho}_r$ is

$$
\sigma_r^2 = Var(\tilde{\rho}_r) \approx (1-\rho)\left(\rho+\frac{1}{n}\right)\ E\left[\frac{\sigma_t^2}{\sum_{i\in S}(Y_{i0}-\overline{y}_b)^2}\right] \qquad (4.17)
$$

because $\tilde{\rho}_r$ is unbiased. We estimate $E\left[\frac{\sigma_t^2}{\sum_{i\in S}(Y_{i0}-\overline{y}_b)^2}\right]$ from the baseline observations with the inverse of $\widehat{SSC}$ as given by (4.16). Similar to the MLE, choosing parts to re-measure with extreme baseline measurements relative to the baseline average reduces the conditional variance of this estimator. Since the estimator is unbiased, selecting extreme parts will also reduce the unconditional variance.

Note that the regression based estimator uses the average of the repeated mea-

surements to estimate ρ. It does not, however, use the variability of the repeated measurements within each part unlike the next estimator.

4.2.3 ANOVA Estimator

We can use the variation within the repeated measurements to get an ANOVA-like estimate of ρ. For each part, the variation within the repeated measurements $\sum_{j=1}^{n}\left(Y_{ij}-\overline{Y}_{i.}\right)^2$ is independent of Y_{i0} and

$$MSW = \frac{\sum_{i\in S}\sum_{j=1}^{n}\left(y_{ij}-\overline{y}_{i.}\right)^2}{k(n-1)} \tag{4.18}$$

is an estimate of σ_m^2. Note that the average $\overline{y}_{i.}$ in the above expression does not include the baseline measurements. Since the baseline variation is an estimate of σ_t^2 and $\rho = \frac{\sigma_p^2}{\sigma_p^2+\sigma_m^2}$, by rearrangement, we obtain the estimate

$$\widehat{\rho}_a = 1 - \frac{\text{MSW}}{s_b^2}. \tag{4.19}$$

Transforming the ANOVA estimator, we see that $(1-\tilde{\rho}_a)/(1-\rho)$ has an F-distribution with $k(n-1)$ and $b-1$ degrees of freedom and so the distribution of the ANOVA estimator depends only on ρ and not the other unknown parameters μ and σ_t. We have

$$E(\widetilde{\rho}_a) = 1-(1-\rho)\,\frac{b-1}{b-3} \quad = \rho\,\left(\frac{b-1}{b-3}\right) - \frac{2}{b-3} \tag{4.20}$$

$$\text{and} \qquad \sigma_a^2 = Var(\widetilde{\rho}_a) = (1-\rho)^2 v_F \tag{4.21}$$

$$\text{where} \qquad v_F = Var(F_{k(n-1),b-1}) = \frac{2\,(b-1)^2(k(n-1)+(b-1)-2)}{k(n-1)((b-1)-2)^2((b-1)-4)} \tag{4.22}$$

Note that both the regression and ANOVA estimates do not require that the

parts selected to be re-measured be representative of the process. They do however require that the measurement errors be representative.

4.2.4 Combined Estimator

An estimator which has a closed form and turns out to have similar properties to the MLE is a combination of the two estimators $\widetilde{\rho}_r$ and $\widetilde{\rho}_a$ as described in Sections 4.2.3 and 4.2.2 respectively. We combine these two estimators because they are uncorrelated.

To show that the covariance between $\widetilde{\rho}_r$ and $\widetilde{\rho}_a$ is zero, we use the conditional covariance formula

$$\mathrm{Cov}\left(\widetilde{\rho}_r, \widetilde{\rho}_a\right) = E\left[\mathrm{Cov}\left(\widetilde{\rho}_r, \widetilde{\rho}_a\right) \middle| \{y_{i0}\}_{i=1}^{b}\right] + \mathrm{Cov}\left(E\left[\widetilde{\rho}_r \middle| \{y_{i0}\}_{i=1}^{b}\right], E\left[\widetilde{\rho}_a \middle| \{y_{i0}\}_{i=1}^{b}\right]\right),$$

where we used $\{y_{i0}\}_{i=1}^{b}$ to represent $\{y_{10}, \ldots, y_{b0}\}$ to save space. We show that each term is equal to zero.

The first term,

$$E\left[\mathrm{Cov}\left(\widetilde{\rho}_r, \widetilde{\rho}_a\right) | y_{10}, \ldots, y_{b0}\right] = 0 \tag{4.23}$$

because from repeated measurements, $\widetilde{\rho}_r$ is a function of the means and $\widetilde{\rho}_a$ is a function of the variance. Since, the measurement error is assumed to have normal distribution and the sample means and variances from observations with a normal distribution are independent, their covariance is zero.

For the second term,

$$\begin{aligned} & \mathrm{Cov}\left(E\left[\widetilde{\rho}_r \middle| y_{10}, \ldots, y_{b0}\right], E\left[\widetilde{\rho}_a \middle| y_{10}, \ldots, y_{b0}\right]\right) \\ = \; & \mathrm{Cov}\left(\rho + \left[(\mu - \overline{y}_b)\frac{\widehat{SC}}{\widehat{SSC}}\right](1+\rho), 1 - \frac{\sigma_t^2(1-\rho)}{s_b^2}\right) \end{aligned}$$

$$
\begin{aligned}
&= \text{Cov}\left(\left[(\mu-\overline{y}_b)\,\frac{\widehat{SC}}{\widehat{SSC}}\right](1+\rho), -\frac{\sigma_t^2(1-\rho)}{s_b^2}\right)\\
&= -\sigma_t^2(1-\rho)(1+\rho)\ \text{Cov}\left(\left[(\mu-\overline{y}_b)\,\frac{\widehat{SC}}{\widehat{SSC}}\right], \frac{1}{s_b^2}\right)\\
&= -\sigma_t^2(1-\rho)(1+\rho)\ \left(E\left[(\mu-\overline{y}_b)\,\frac{\widehat{SC}}{\widehat{SSC}}\frac{1}{s_b^2}\right] - E\left[(\mu-\overline{y}_b)\,\frac{\widehat{SC}}{\widehat{SSC}}\right]E\left[\frac{1}{s_b^2}\right]\right)\\
&= -\sigma_t^2(1-\rho)(1+\rho)\ \left(E\Big[\mu-\overline{y}_b\Big]E\left[\frac{\widehat{SC}}{\widehat{SSC}}\frac{1}{s_b^2}\right] - E\Big[\mu-\overline{y}_b\Big]E\left[\frac{\widehat{SC}}{\widehat{SSC}}\right]E\left[\frac{1}{s_b^2}\right]\right)\\
&\quad \text{because the mean}, (\mu-\overline{y}_b)\,, \text{ is independent of the residuals, } \frac{\widehat{SC}}{\widehat{SSC}}\frac{1}{s_b^2}.\\
&= -\sigma_t^2(1-\rho)(1+\rho)\ \left(0\times E\left[\frac{\widehat{SC}}{\widehat{SSC}}\frac{1}{s_b^2}\right] - 0\times E\left[\frac{\widehat{SC}}{\widehat{SSC}}\right]E\left[\frac{1}{s_b^2}\right]\right)\\
&= 0
\end{aligned}
$$

Therefore, the covariance between $\widetilde{\rho}_r$ and $\widetilde{\rho}_a$ is zero.

If we suppose the two uncorrelated estimators of ρ, $\widetilde{\rho}_r$ and $\widetilde{\rho}_a$, had known variances σ_r^2 and σ_a^2 then the minimum variance linear combination is

$$w\ \widetilde{\rho}_r + (1-w)\ \widetilde{\rho}_a = \frac{\sigma_a^2}{\sigma_r^2+\sigma_a^2}\ \widetilde{\rho}_r + \frac{\sigma_r^2}{\sigma_r^2+\sigma_a^2}\ \widetilde{\rho}_a. \tag{4.24}$$

This combined estimator is unbiased because it is a weighted sum of two unbiased estimators, $\tilde{\rho}_a$ and $\tilde{\rho}_r$. The linear combination has

$$E\left[w\widetilde{\rho}_r + (1-w)\widetilde{\rho}_a\right] = \frac{\sigma_a^2}{\sigma_r^2+\sigma_a^2}\ E\left[\widetilde{\rho}_r\right] + \frac{\sigma_r^2}{\sigma_r^2+\sigma_a^2}\ E\left[\widetilde{\rho}_a\right] = \frac{\sigma_a^2}{\sigma_r^2+\sigma_a^2}\ \rho + \frac{\sigma_r^2}{\sigma_r^2+\sigma_a^2}\ \rho = \rho.$$

An estimating function can be created (from 4.24) by subtracting its expectation ρ. Multiplying by $\sigma_r^2+\sigma_a^2$, we get

$$\Psi_c(\rho) = \sigma_a^2\ \tilde{\rho}_r + \sigma_r^2\ \tilde{\rho}_a - (\sigma_a^2+\sigma_r^2)\rho. \tag{4.25}$$

An estimating function has expectation zero. Parameters are estimated by setting the estimating function to zero and solving. Substituting the quantities in (4.21) for σ_a^2 and (4.17) for σ_r^2, we obtain the combined estimate ρ_c, as a root of the quadratic equation (4.26).

$$\left(v_F - E\left[\frac{1}{SSC}\right]\right)\rho_c^2 + \left(E\left[\frac{1}{SSC}\right]\left[\hat{\rho}_a - \frac{1}{n}\right] - v_F\left[1+\hat{\rho}_r\right]\right)\rho_c + \left(v_F\hat{\rho}_r + E\left[\frac{1}{SSC}\right]\frac{\hat{\rho}_a}{n}\right) = 0 \qquad (4.26)$$

where $v_F = Var\left(F_{k(n-1),b-1}\right)$. As with the regression estimator, we estimate $E\left[\frac{1}{SSC}\right]$ from the baseline observations with the inverse of $\widehat{SSC} = \sum_{i\in S}\left[\frac{y_{i0}-\hat{\mu}}{s_b}\right]^2$.

In this case based on simulations, the appropriate estimator is the smaller root because the larger root gives estimates of ρ which are greater than one and ρ is bounded between zero and one. Note that $\tilde{\rho}_c$ is not just a simple weighted average of the two previous estimators because the variances σ_r^2 and σ_a^2 depend on ρ.

To find the asymptotic distribution, we use general results from Jorgensen and Knudsen [2004] that establish the asymptotic normality of the estimates corresponding to the estimating functions. The asymptotic variance of the combined estimator is approximately

$$\mathrm{Var}(\tilde{\rho}_c) \approx \frac{Var\left[\Psi_c(\rho)\right]}{\left\{E\left[\frac{\partial}{\partial\rho}\Psi_c(\rho)\right]\right\}^2} = \frac{\sigma_a^2\sigma_r^2}{(\sigma_a^2+\sigma_r^2)}. \qquad (4.27)$$

The asymptotic variance-covariance matrix [see Jorgensen and Knudsen, 2004] of $\tilde{\mu}, \tilde{\sigma}_t^2$ as estimated from the baseline and $\tilde{\rho}_c$, as given by solving (4.26), is

$$\mathrm{Cov}\begin{pmatrix}\tilde{\mu}\\ \tilde{\sigma}_t^2\\ \tilde{\rho}_c\end{pmatrix} \approx \begin{pmatrix} \frac{\sigma_t^2}{b} & 0 & 0 \\ 0 & \frac{2\sigma_t^4}{b-1} & \frac{2\sigma_t^2(1-\rho)}{b-3}\frac{\sigma_r^2}{\sigma_r^2+\sigma_a^2} \\ 0 & \frac{2\sigma_t^2(1-\rho)}{b-3}\frac{\sigma_r^2}{\sigma_r^2+\sigma_a^2} & \frac{\sigma_a^2\sigma_r^2}{(\sigma_a^2+\sigma_r^2)} \end{pmatrix}. \qquad (4.28)$$

The variance of the combined estimator depends on ρ through σ_a^2 and σ_r^2. Throughout our simulations, we noticed that the distribution of $\tilde{\rho}_c$ is skewed towards zero when ρ is close to one. Qualitatively, this occurs because $\tilde{\rho}_a$ (which has a skewed distribution) has more weight in this situation.

To construct confidence intervals for parameters with skewed distribution it is common to work on a transformed scale. A transformation that seems to perform well is Fisher's z-transformation. We let

$$\theta = \frac{1}{2}\log\frac{1+\rho}{1-\rho} \qquad \text{and} \qquad \frac{\partial\theta}{\partial\rho} = \frac{1}{1-\rho^2} \tag{4.29}$$

then its approximate variance is

$$\text{Var}\left(\widetilde{\theta}\right) \approx \frac{Var\left[\Psi_c(\rho)\right]}{E\left\{\left[\frac{\partial}{\partial\rho}\Psi_c(\rho)\right]\right\}^2 \left[\frac{\partial\rho}{\partial\theta}\right]^2}. \tag{4.30}$$

To construct approximate confidence interval for ρ, we suggest first constructing a confidence interval on the θ scale based on asymptotic normality on the transformed scale. Then transform the confidence interval limits to create a confidence interval for ρ. See the example in the next section.

To perform the test of hypothesis in (1.19) we use the same approach as for a one-sided confidence interval. First, we make a one-sided confidence interval on the transform scale under normality and using the variance given in (4.30). Then, we transform the limit to determine the critical value on the ρ scale.

4.2.5 Numerical Example of Various Estimates for ρ Based on Leveraged Plan

Steiner and Mackay [2005] present an example of a leveraged measurement as-

sessment study on an automated crankshaft main journal gauge. Although they calculate only the ANOVA estimator, we can apply all four methods of estimation for illustration. In their example, three parts, a large, small and medium sized part, were selected from the baseline study to be re-measured. To more closely match the suggestions in this thesis to select an equal number of extreme parts on each side of the baseline average, we proceed by assuming only the large and small parts were selected. In the example, a baseline of 100 parts was randomly selected from the process. The baseline data, given as a difference from a target value, are shown in Table 4.1.

Table 4.1: Baseline Data of 100 Camshaft Journal Diameters

5.3	0.0	-4.1	-6.4	-5.7	7.1	-0.5	-1.7	-2.7	2.1
0.9	-1.5	-5.4	3.3	6.0	2.4	-1.2	3.4	-2.9	-6.4
-12.8	-7.3	1.5	1.9	5.6	-5.2	2.4	0.9	-2.5	-0.8
4.6	4.1	-7.8	10.3	0.0	-0.9	-3.3	5.7	8.2	1.5
-5.3	4.2	4.6	10.5	-3.4	0.5	1.4	9.1	-1.1	**12.8**
-2.7	-3.2	4.4	1.0	1.2	-4.0	-1.6	-2.5	-6.9	1.2
-2.2	-0.6	-5.4	-6.0	-1.1	0.1	-3.5	2.5	1.4	**-12.2**
-1.5	-6.0	9.7	5.2	10.4	2.2	9.2	3.6	1.8	1.7
-2.0	-0.8	-4.1	-4.5	4.2	7.8	-3.2	1.9	-0.4	0.5
4.3	2.3	6.1	5.0	4.6	8.4	6.1	-7.1	4.7	-7.4

The baseline average $\overline{y}_b$ is 0.540 and baseline variance is s_b^2 is 25.865. The parts chosen to be repeatedly measured were parts 50 and 70 (i.e. S=$\{50, 70\}$), with baseline measurements 12.8 and -12.2 respectively. These two parts were measured an additional 18 times each. Note that the average of the two baseline measurements for the selected parts is very close to the baseline average. Part 21 with baseline value -12.8 that is more extreme than part 70 could have been selected instead but for reasons unknown to us, this was not done.

The individual measurements for each re-measured part are shown in Table 4.2 with the average and standard deviation for the repeated measurements within each

part. We see regression toward the mean for the measurements on part 70 but not on part 50. That is, all the repeated measurement for part 70 are below the initial measurement but for part 50 the repeated are balanced by the initial measurement. In addition, we see that the measurement system is easily able to distinguish the two selected parts and that the measurement variation for the two parts is roughly the same.

Table 4.2: Example of a Stage 2 Sample with 2 Extreme Parts Repeatedly Measured 18 Times Each

Part 50 $y_{50,0} = 12.8$							Part 70 $y_{70,0} = -12.2$					
10.9	13.2	12.8	12.6	12.7	14.1		-10.3	-11.1	-10.0	-12.2	-11.0	-11.1
12.9	13.1	12.0	13.3	12.6	13.4		-10.9	-10.0	-10.6	-11.4	-11.5	-11.1
12.0	12.9	11.7	11.8	12.2	14.1		-11.4	-10.7	-10.3	-11.4	-9.8	-11.5
$\overline{y}_{50.} = 12.7$ $s^2_{50} = 0.68029$							$\overline{y}_{70.} = -10.9$ $s^2_{70} = 0.40997$					

To check the normality assumptions, we constructed QQ-plots (not shown here) of the baseline data and the residuals of the repeated measurements excluding the baseline measurements. There is no evidence to contradict the model assumptions in this example. It is interesting to note that all of the re-measured values for part 70 are larger than the baseline value. This suggests that there was a large measurement error in the baseline value of this part. We would expect such behaviour if the measurement variation is large relative to the process variation.

Using (4.18), the ANOVA estimate is

$$\widehat{\rho}_a = 1 - \frac{MSW}{s_b^2} = 1 - \frac{(s_{50}^2 + s_{70}^2)/2}{s_b^2} = 1 - \frac{(0.40997 + 0.68029)/2}{25.865} = 0.97892$$

The estimates of SC and SSC using the two selected parts and the baseline

summary statistics are

$$\widehat{SC} = -2.51 + 2.41 = -0.10 \quad \text{and} \quad \widehat{SSC} = 6.275 + 5.811 = 12.086$$

The maximum likelihood estimates for (μ, σ_t^2, ρ) are $(0.551, 25.392, 0.97809)$.

Using (4.13), the regression estimate is

$$\widehat{\rho}_r = \frac{\sum_{i \in S} (\overline{y}_{i.} - \overline{y}_b)(y_{i0} - \overline{y}_b)}{\sum_{i \in S} (y_{i0} - \overline{y}_b)^2} = \frac{145.8 + 148.9}{162.3 + 150.3} = \frac{294.7}{312.6} = 0.94267$$

We need v_F and the baseline estimate of SSC to determine the coefficients of the quadratic equation (4.26) used for the combined estimator. Using $v_F = Var(F_{34,99}) = 0.0845$, the combined estimate of ρ is the smaller root of the quadratic equation

$$0.001755011\rho_c^2 - 0.0877455\rho_c + 0.08414984 = 0$$

The two roots of this equation are 0.97816 and 49.019. Therefore $\widehat{\rho}_c = 0.97816$ is the combined estimate of ρ. Table 4.3 summarizes the four estimates and their corresponding standard errors. Since ρ appears to be large for this measurement system, there is little difference in the estimates and their standard errors with the exception of the regression estimate which has much higher standard error.

Table 4.3: Estimates of ρ for the Camshaft Journal Diameters Example

	Estimate	Standard Error
ρ_a	0.97892	0.00613
ρ_r	0.94267	0.06881
ρ_c	0.97816	0.00628
ρ_{mle}	0.97809	0.00597

We illustrate the calculations for confidence intervals, as given in (4.29), using

the combined estimate. The transformed estimate, using (4.29) is $\widehat{\theta} = \frac{1}{2}\log\left(\frac{1+0.97816}{1-0.97816}\right)$ = 2.2531. Using the variance of this estimate given in (4.30) and the standard error of the combined estimate (found in Table 4.3), the standard error of $\widehat{\theta}$ is $\sqrt{\frac{0.00628^2}{(1-0.97816^2)^2}} = 0.14535$. Thus, a 95% confidence interval for θ is $2.25 \pm 1.96(0.145) = (1.968, 2.538)$ and the approximate 95% confidence interval in terms of ρ is $(0.962, 0.988)$.

4.2.6 Comparison of the Various Estimators for ρ Based on the Leveraged Plan

We consider a sampling plan with $b = 30$, $k = 6$ and $n = 5$ because this is the plan that will be recommended in Section 4.3 when the total number of measurements is 60. We suggest choosing the six parts corresponding to the three largest and three smallest measurements from the baseline study of 30 parts. Figure 4.1 shows the bias and standard deviation for the MLE, regression estimator ($\tilde{\rho}_r$), ANOVA estimator ($\tilde{\rho}_a$) and the Combined Estimator ($\tilde{\rho}_c$). The figure was created by simulating ten thousand samples for each value of ρ. The results of the simulation are based on ten thousand samples for 23 values of ρ spread over the interval (0.01,0.99) with higher density where the bias and standard deviation are changing rapidly. We used the same set of values for ρ in all simulations in this chapter.

The two individual components of the combined estimator, the regression and ANOVA estimators are efficient for different values of ρ. The standard deviation of the ANOVA estimator is much larger than the regression estimator when $\rho = 0.2$ but it performs well when ρ is larger than 0.9. All estimators are slightly biased except for the regression estimator.

Notice that the combined and the MLE estimators perform similarly when $\rho \geq 0.3$. Since usually measurement systems are reasonably good, i.e. ρ is larger than 0.5, we can use the combined estimator without loss of efficiency. We see similar

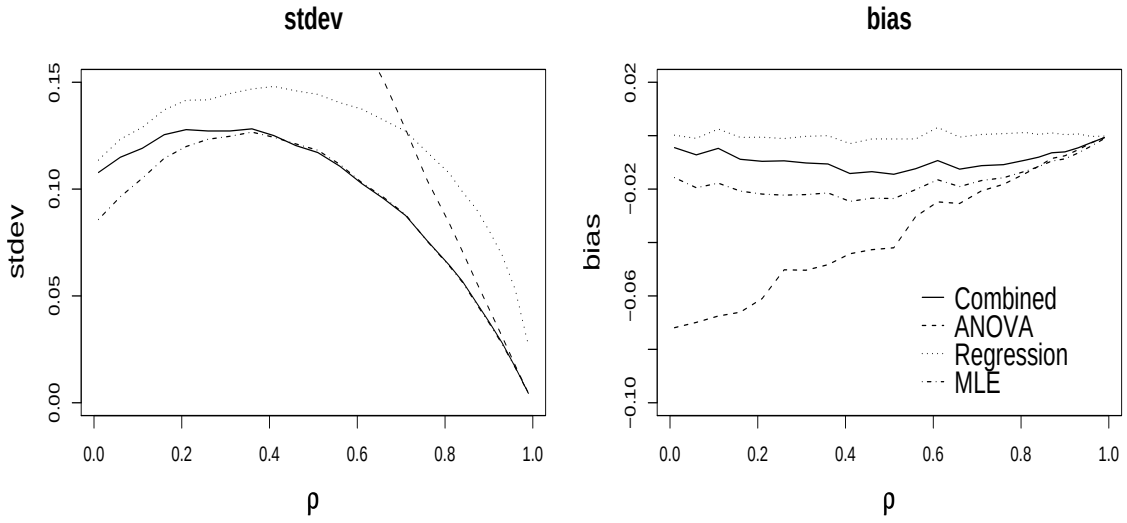

Figure 4.1: Comparison of the Bias and Standard Deviation for Estimators of ρ with a Leveraged Plan

results for leveraged sampling plans with other values of b, k and n.

4.2.7 The Effect of Varying the Number of Repeated Measurements on Each Part

In Stage 2, the proposed LP recommends choosing k parts and remeasuring them n times each. We may wonder if there is any benefit to varying the number of repeated measurements and if so, does this benefit outweigh the cost of a more complex plan. For example, we could increase the number of measurements for parts with more extreme initial measurements.

Letting the number of repeated measurements on part i be n_i, the degrees of freedom for the ANOVA will be the same as long as each part has at least two measurements. The variance for the regression estimator (4.15), now with n_i

measurements on each part, is

$$Var\left[\tilde{\rho}_r \middle| y_{10}, \ldots, y_{b0}\right] = \frac{\sigma_t^2\left(1-\rho\right)}{\sum_{i\in S}\left(y_{i0}-\overline{y}_b\right)^2}\left\{\rho + \frac{\sum_{i\in S}\left[\left(y_{i0}-\overline{y}_b\right)^2/n_i\right]}{\sum_{i\in S}\left(y_{i0}-\overline{y}_b\right)^2}\right\}.$$

We can use Lagrange multipliers to determine the optimal n_i for each part given the baseline measurements. If we assume b and k are fixed then $\sum_{i\in S} n_i = N - b$. To find the optimal n_i we minimize

$$L(n_i, \lambda) = Var\left[\tilde{\rho}_r \middle| y_{10}, \ldots, y_{b0}\right] + \lambda\left(\sum_i^k n_i - (N-b)\right)$$

$$\frac{\partial L(\lambda)}{\partial n_i} = -\frac{\sigma_t^2\left(1-\rho\right)}{\left[\sum_{i\in S}\left(y_{i0}-\overline{y}_b\right)^2\right]^2}\sum_{i\in S}\frac{\left(y_{i0}-\overline{y}_b\right)^2}{n_i^2} + \lambda$$

$$\Rightarrow \quad n_i \propto \left|y_{i0}-\overline{y}_b\right| \quad \Rightarrow \quad n_i = (N-b)\frac{\left|y_{i0}-\overline{y}_b\right|}{\sum_{i\in S}\left|y_{i0}-\overline{y}_b\right|} \tag{4.31}$$

$$\text{but for any application we set } \; n_i = \text{round}\left[(N-b)\frac{\left|y_{i0}-\overline{y}_b\right|}{\sum_{i\in S}\left|y_{i0}-\overline{y}_b\right|}\right]$$

Plugging in the optimal n_i into the variance for the regression estimator (4.15) we get the conditional variance

$$Var\left[\tilde{\rho}_r \middle| y_{10}, \ldots, y_{b0}\right] = \frac{\sigma_t^2\left(1-\rho\right)}{\sum_{i\in S}\left(y_{i0}-\overline{y}_b\right)^2}\left\{\rho + \frac{1}{(N-b)}\frac{\left[\sum_{i\in S}\left|y_{i0}-\overline{y}_b\right|\right]^2}{\sum_{i\in S}\left(y_{i0}-\overline{y}_b\right)^2}\right\} \tag{4.32}$$

We compared the variance in (4.32) to the variance of the design where each part gets the same number of repeated measurements given by (4.15). To reduce the number of parameters, we will also assume that $b \approx N/2$ which means $\sum_{i\in S} n_i = b$. This assumption is reasonable because further ahead in Section 4.3 we show that for the best leveraged plans, b should be roughly half of the sample size N. This

means we have that $n = \frac{k}{b}$ because $b = N/2$. The ratio is

$$\frac{Var\left[\widetilde{\rho}_r \middle| y_{10}, \ldots, y_{b0}, n_i\right]}{Var\left[\widetilde{\rho}_r \middle| y_{10}, \ldots, y_{b0}, n_i = n = \frac{k}{b}\right]} = \frac{\left\{\rho + \frac{1}{b}\frac{\left[\sum_{i\in S}|y_{i0}-\overline{y}_b|\right]^2}{\sum_{i\in S}(y_{i0}-\overline{y}_b)^2}\right\}}{\left\{\rho + \frac{k}{b}\right\}} \tag{4.33}$$

For large b this ratio can be approximated by

$$1 + \frac{q(b,k)}{\rho} + O\left(\frac{1}{b^2}\right) \quad \text{where} \quad q(b,k) = \frac{1}{b}\left\{\frac{\left[\sum_{i\in S}|y_{i0} - \overline{y}_b|\right]^2}{\sum_{i\in S}(y_{i0} - \overline{y}_b)^2} - k\right\} \tag{4.34}$$

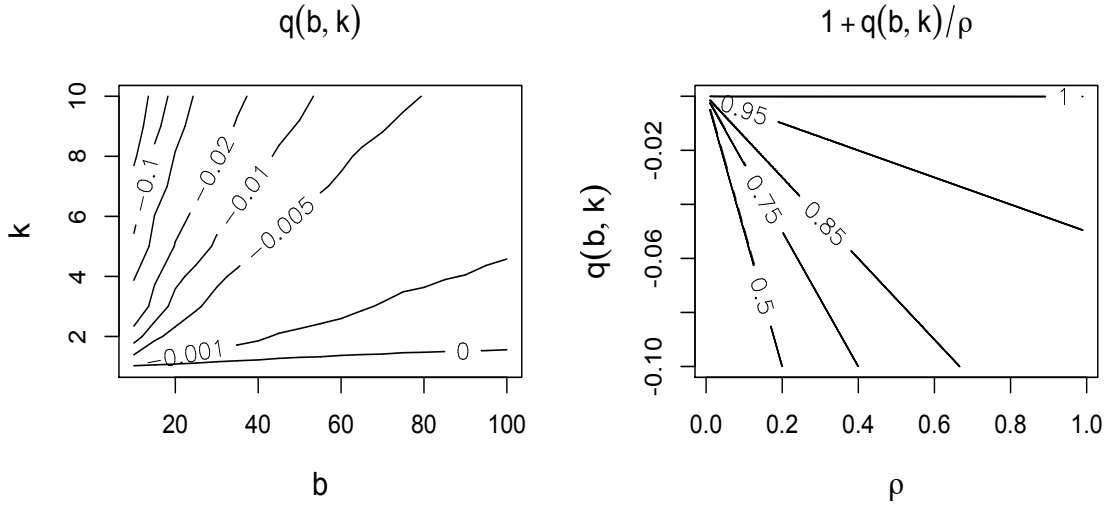

Figure 4.2: The approximate ratio of Var$(\widetilde{\rho}_r)$ as given by (4.34)

Figure 4.2, shows the approximated ratio in two panels. The left panel is a contour plot of $q(b,k)$, which is given in (4.34), by k and b and was generated by simulation. The right shows the ratio by $q(b,k)$ and ρ. For example, if $b = 40$ and $k = 6$ then from the left panel $q(b,k)$ is approximately -0.01 and from the right panel we see that the ratio in (4.33) is between 0.95 and 1 for $\rho \geq 0.2$. That is, when $b = 40$ and $k = 6$ the plan with optimal n_i is at most 0.95 more efficient that the plan with $n_i = n$ when $\rho \geq 0.2$. Using the two panels we see that varying n_i does reduce the variance of the regression estimator but for typical large values of

ρ the difference is small and would likely not compensate for the added complexity.

4.3 Leveraged Plan Design

In this section, we determine how to choose a leveraged plan (i.e. values for b, k and n) when the total number of measurements is N and the precision desired for the estimate of ρ using the combined estimator is specified. As with most sample size calculations, we must also specify a value of ρ to select the plan. We consider two specific values of ρ, namely 0.80 and 0.91. These values of ρ are chosen because in terms of measurement repeatability they are equivalent to 0.45 and 0.30, respectively. Thus $\rho = 0.91$ corresponds to the minimum acceptable value in Automotive Industry Action Group [2002]. We also include $\rho = 0.80$ in Tables 4.4 and 4.5 to show how the standard deviation of $\tilde{\rho}_c$ behaves with a relatively poor measurement system. The goal of this section is to understand, for fixed N, how the standard deviation of the estimator $\widetilde{\rho}_c$ depends on b,k and n. In particular, we want to find values of b,k and n that minimize the standard deviation.

When calculating the asymptotic variance (4.27) for the combined estimator, we need to replace $\frac{1}{\widehat{SSC}}$ by its expected value because σ_r^2, as shown in (4.17), depends on $E\left[\frac{1}{\widehat{SSC}}\right]$. This quantity corresponds to the sum of the standardized squares of the k observations chosen for Stage 2 from the baseline. In an LP, we choose parts to be remeasured based on their extreme initial values which typically correspond to (assuming k is even) the $k/2$ smallest and $k/2$ largest observed values in the baseline. This implies these extreme initial values can be represented as order statistics from the standard normal distribution. We can write

$$E\left[\frac{1}{SSC}\right] = E\left(\frac{1}{Z^2_{[1:b]} + \cdots + Z^2_{[k/2:b]} + Z^2_{[b-k/2+1:b]} + \cdots + Z^2_{[b:b]}}\right) \tag{4.35}$$

where $Z_{[i:b]}$ is the i^{th} order statistic from a sample of b standard normal random variables. We estimate (4.35) by simulating ten thousand samples of b observations.

To start, we consider $N = 60$. The first step is an exhaustive search looking at all possible values b,k and n with $b + kn = 60$. There are approximately 200 designs. Table 4.4 displays the approximate standard deviation, obtained using the asymptotic variance (4.27), of the combined estimator for ρ. We show the five plans with the smallest standard deviation, five plans with the largest standard deviation and the five plans with close to the median standard deviation. To check the asymptotic approximations we compared with simulation results. These results are very close to those obtained through simulation. Poor plans with large standard deviations tend to have a small number of observations allocated to the baseline. The designs with the lowest standard deviation have $b \backsimeq nk$, i.e. $b \backsimeq N/2$.

Table 4.4: The LP Designs that have the Smallest, Middle and Largest stdev($\tilde{\rho}_c$) when $b + nk = 60$

	$\rho = 0.80$					$\rho = 0.91$			
	b	k	n	stdev($\tilde{\rho}_c$)		b	k	n	stdev($\tilde{\rho}_c$)
smallest 5	32	7	4	0.0684		32	4	7	0.0350
	30	6	5	0.0688		33	3	9	0.0351
	33	9	3	0.0688		30	5	6	0.0351
	30	10	3	0.0689		30	6	5	0.0352
	35	5	5	0.0690		30	3	10	0.0352
middle 5	18	14	3	0.0785		38	11	2	0.0394
	42	1	18	0.0785		45	3	5	0.0396
	22	2	19	0.0788		45	1	15	0.0397
	25	1	35	0.0789		21	1	39	0.0401
	18	7	6	0.0792		20	2	20	0.0401
largest 5	7	1	53	0.1766		6	3	18	0.1017
	6	2	27	0.1831		6	2	27	0.1058
	5	5	11	0.1870		6	1	54	0.1138
	6	1	54	0.2053		5	5	11	0.1203
	5	1	55	0.2475		5	1	55	0.1496

In Table 4.5, we show the plans corresponding to the lowest stdev($\tilde{\rho}_c$) for differ-

ent values of $N = b+nk$ when ρ equals 0.80 and 0.91. The differences in performance among the plans with the same N is small but notice that the baseline size b is close to $N/2$ for all the best plans. Using this empirical evidence, we suggest using $b \backsimeq N/2$, $n \backsimeq 5$ and then $k \backsimeq N/10$. This plan is in every set of the top 5 in Table 4.5. Since the LP design parameters must be integers, given a total sample size N, we recommend the plan with $k = \lfloor N/10 \rfloor$, $n = 5$ and $b = N - 5\lfloor N/10 \rfloor$, where $\lfloor \rfloor$ is the floor function that rounds down to nearest integer.

In Table 4.6, assuming that we use the recommended plan, we give the value of N required to achieve a specified standard error of the transformed variable (4.29) when given a value of ρ. We used the transformed scale because the distribution of the estimator is skewed when ρ is close to 1. To illustrate use of Table 4.6, suppose historical data suggests ρ is approximately 0.91 and we want to estimate ρ with a standard deviation of at most 0.025. Then using the variance from (4.30), we obtain the standard deviation on the transformed scale to be $\frac{0.025}{1-0.91^2} \approx 0.145$. Now in Table 4.6, we look down the column with $\rho = 0.91$ and row corresponding to $\text{stdev}(\tilde{\theta}) = 0.15$ to get the total sample size of 101. Using the suggested plan, we require 51 parts for the baseline study. Then, from the baseline sample, we select 10 extreme parts to repeatedly measure 5 times each.

4.3.1 Optimal Designs

The previous subsection used simulation to obtain the approximate standard deviation to find a recommended plan. We choose only two values of ρ because this task was computationally intensive. To further explore the optimal designs at different values of ρ and N, we now use the Fisher information. This means that optimal designs in this subsection will have the the smallest asymptotic standard deviation in contrast to the previous analysis which found the plans with the smallest ap-

Table 4.5: The Five Plans with the Lowest stdev($\tilde{\rho}_c$) for Different Values of $N = b + nk$

$\rho = 0.80$						$\rho = 0.91$				
N	b	k	n	stdev($\tilde{\rho}_c$)		N	b	k	n	stdev($\tilde{\rho}_c$)
30	18	3	4	0.1065		30	18	2	6	0.0552
30	18	4	3	0.1068		30	16	2	7	0.0555
30	15	5	3	0.1076		30	18	3	4	0.0556
30	18	2	6	0.1078		30	18	1	12	0.0557
30	15	3	5	0.1081		30	17	1	13	0.0558
50	26	6	4	0.0766		50	26	4	6	0.0393
50	26	8	3	0.0769		50	26	3	8	0.0393
50	25	5	5	0.0770		50	29	3	7	0.0393
50	30	5	4	0.0771		50	28	2	11	0.0394
50	29	7	3	0.0771		50	25	5	5	0.0395
75	39	9	4	0.0599		75	40	5	7	0.0306
75	40	7	5	0.0601		75	39	6	6	0.0306
75	43	8	4	0.0602		75	39	4	9	0.0307
75	39	12	3	0.0603		75	40	7	5	0.0307
75	35	10	4	0.0603		75	43	4	8	0.0308
100	52	12	4	0.0507		100	51	7	7	0.0259
100	48	13	4	0.0509		100	52	6	8	0.0259
100	50	10	5	0.0509		100	52	8	6	0.0259
100	56	11	4	0.0509		100	50	10	5	0.0260
100	55	9	5	0.0510		100	50	5	10	0.0260
125	65	15	4	0.0447		125	62	9	7	0.0228
125	61	16	4	0.0448		125	65	10	6	0.0229
125	65	12	5	0.0448		125	61	8	8	0.0229
125	69	14	4	0.0449		125	62	7	9	0.0229
125	60	13	5	0.0449		125	65	6	10	0.0229
200	100	25	4	0.0346		200	102	14	7	0.0177
200	104	24	4	0.0347		200	98	17	6	0.0177
200	100	20	5	0.0347		200	95	15	7	0.0177
200	96	26	4	0.0347		200	96	13	8	0.0177
200	105	19	5	0.0347		200	104	12	8	0.0177

proximate standard deviation. Using the Fisher information allows us to examine the optimal designs over all ρ, as seen in Table 4.7. It presents the asymptotic optimal leveraged designs given ρ when $N = 30$, 60 and 100 by showing the range of ρ for which a design is optimal. For example, if we wanted the optimal design

Table 4.6: Values of N for Estimating ρ with a Particular Standard Deviation when $b = N - 5\lfloor N/10 \rfloor$, $n = 5$ and $k = \lfloor N/10 \rfloor$

	Assumed value of ρ					
stdev($\tilde{z}$)	0.2	0.4	0.6	0.8	0.91	0.99
0.25	22	27	32	39	44	49
0.20	31	38	45	55	62	69
0.15	48	60	73	89	101	113
0.14	54	68	82	101	115	127
0.13	62	77	94	115	131	146
0.12	71	89	109	133	152	168
0.11	83	105	128	157	178	198
0.10	98	125	153	188	213	236
0.09	120	154	188	231	261	289
0.08	151	194	238	292	329	362
0.07	197	256	314	383	429	469
0.06	273	356	436	528	586	633
0.05	409	538	657	780	852	908

for $\rho = 0.80$ when $N = 60$ then we look down the eighth column. Since 0.8 falls into the range 0.52 to 0.99 the optimal plan has $b = 30, k = 10$ and $n = 3$.

Table 4.7 has two interesting features; the optimal design for each N is stationary when $\rho \geq 0.54$ and these optimal designs all have $n = 3$. This suggested small number of repeated measurements is contrary to the conclusions from Table 4.5, which suggested a larger n. However, Table 4.5 was created using simulation on the combined estimator and not the maximum likelihood estimator. Also, we have noticed that the optimal design space is fairly flat as long as b, the baseline size, is roughly half N, the total sample size. This feature is shown in both Tables 4.5 and 4.7.

We can use the asymptotic variance to determine the optimal plans for estimating ρ given a constraint on the precision. Table 4.8 presents this on a transformed scale, named q. The scale, $q =$stdev$(\widetilde{\rho})/\sqrt{(1-\rho^2)\rho(1-\rho)}$, was required to make the table compact and scales the precision requirement so that we require higher

Table 4.7: The Asymptotic Optimal Leveraged Designs (b, k, n) for a Given ρ When $N = 30$, 60 and 100

$N = 30$				$N = 60$				$N = 100$			
b	k	n	ρ	b	k	n	ρ	b	k	n	ρ
2	1	28	0.01 to 0.03	2	1	58	0.01	4	1	96	0.01
4	1	26	0.04	5	1	55	0.02	10	2	45	0.02
6	2	12	0.05 to 0.07	8	2	26	0.03	13	3	29	0.03
8	2	11	0.08 to 0.10	10	2	25	0.04	16	4	21	0.04
10	2	10	0.11 to 0.14	12	2	24	0.05	20	4	20	0.05
12	2	9	0.15 to 0.16	12	3	16	0.06	24	4	19	0.06
12	3	6	0.17 to 0.23	15	3	15	0.07	25	5	15	0.07
14	4	4	0.24 to 0.53	16	4	11	0.08 to 0.10	28	6	12	0.08 to 0.10
15	5	3	0.54 to 0.99	20	4	10	0.11 to 0.13	30	7	10	0.11 to 0.12
				20	5	8	0.14 to 0.15	36	8	8	0.13 to 0.17
				24	6	6	0.16 to 0.26	37	9	7	0.18
				25	7	5	0.27 to 0.29	40	10	6	0.19 to 0.25
				28	8	4	0.30 to 0.51	45	11	5	0.26 to 0.34
				30	10	3	0.52 to 0.99	48	13	4	0.35 to 0.50
								52	12	4	0.51 to 0.52
								52	16	3	0.53 to 0.99

precision for large ρ values. For example, suppose we thought ρ was 0.6 (i.e. the measurement system was poor) then we might only require a standard deviation of 0.05 whereas if ρ was 0.91 we would require a smaller standard deviation. To illustrate the use of Table 4.8, suppose we thought ρ was about 0.6 and we required a standard deviation equal to 0.05, then we find the q value to be approximately 0.12. To find the optimal plan that achieves this q value we search for the entry that corresponds to $\rho = 0.60$ and $q = 0.12$ in Table 4.8. We get the LP with $b = 139, k = 42$ and $n = 3$. In summary, this means we would require a total of at least 265 measurements to achieve a standard deviation of 0.05 when $\rho = 0.60$.

Table 4.8: Optimal Leveraged Designs (b, k, n) for Estimating ρ with a Particular Standard Deviation

	$q = \text{stdev}(\widetilde{\rho})/\sqrt{(1-\rho^2)\rho(1-\rho)}$						
ρ	0.32	0.28	0.24	0.2	0.16	0.12	0.08
0.1	(16,4,11)	(25,5,11)	(33,6,12)	(46,9,11)	(71,14,11)	(120,25,11)	(275,55,11)
0.2	(20,5,6)	(23,6,7)	(34,8,7)	(53,12,6)	(78,16,7)	(132,29,7)	(297,64,7)
0.3	(21,6,4)	(28,8,4)	(35,9,5)	(50,12,5)	(80,18,5)	(135,33,5)	(305,73,5)
0.4	(21,6,4)	(27,7,4)	(35,10,4)	(49,14,4)	(80,20,4)	(136,36,4)	(314,79,4)
0.5	(22,6,3)	(27,7,4)	(34,9,4)	(48,13,4)	(79,19,4)	(134,34,4)	(296,76,4)
0.6	(22,6,3)	(26,8,3)	(37,11,3)	(52,16,3)	(78,24,3)	(139,42,3)	(311,93,3)
0.7	(22,6,3)	(26,8,3)	(37,11,3)	(50,15,3)	(78,24,3)	(137,41,3)	(309,92,3)
0.8	(22,6,3)	(26,8,3)	(37,11,3)	(50,15,3)	(78,24,3)	(139,42,3)	(305,95,3)
0.9	(19,7,3)	(26,8,3)	(37,11,3)	(52,16,3)	(80,25,3)	(138,44,3)	(311,98,3)
0.99	(20,5,4)	(28,9,3)	(34,12,3)	(49,17,3)	(79,27,3)	(139,47,3)	(313,104,3)

4.4 Model Assumptions

To use an LP, we recommend selecting a number of extreme parts in Stage 2. Questions then arise about the sensitivity of the leveraged plan (relative to a standard plan) to the model assumptions and to methods for detecting departures from the assumed model.

A key assumption of model (1.1) is that the properties of the measurement system are independent of those of the underlying process. This assumption is sometimes called linearity of the measurement system [Automotive Industry Action Group, 2002]. In particular, we are assuming that the bias (if any) and the standard deviation σ_m do not depend on the part size P. Many measurement systems are non-linear in that σ_m increases as the part size increases.

If the bias is constant across part size, this bias gets subsumed into the process mean and has no effect on the estimation of ρ for either an LP or SP. If the bias varies across part size so that, given $P_i = p_i$, the mean of E_{ij} in (1.1) depends on

p_i, then we can re-write the model so that the part effect is redefined to include this bias. With either plan, we can estimate ρ but its value is inflated because the varying bias is included in the process variation. Neither the SP or LP can detect varying bias.

If the measurement variability is a function of part size, then, with either plan, the meaning of ρ is not clear and we are unsure what it is we are estimating. Since we repeatedly measure parts a number of times with either an SP or an LP, we have some power to detect if σ_m is varying over part size. With a leveraged plan, we typically have fewer repeated measurements. However, if the measurement variability is increasing with part size, by using extreme parts we have a greater chance to detect the difference. It is unclear which plan has the advantage to detect this type of non-linearity.

For an LP, we can check the normality assumption for the baseline measurements using a QQ plot. We can also construct a QQ plot for the residuals of the repeated measurements, ignoring the baseline measurement to check the normality of the measurement errors. In practice, if an outlier is observed in the baseline measurements, we do not recommend the use of the corresponding part in the second stage. Such an outlier may be due to either the process or the measurement system. We would advise a separate study of this part, because, if the extreme value is due to the measurement system, finding such an outlier in a small baseline study suggests that there may be a larger problem with the measurement system.

Another question is the issue of robustness of the leveraged plan to departures from normality to underlying distributions with heavier tails. One might expect problems because of the use of parts with extreme baseline measurements. For the standard plan we could not find any results on robustness of the estimators. To investigate this issue briefly, we conducted a small simulation with four cases:

1. both the part and measurement error distribution are normal

2. the part distribution is t_5 and the measurement error distribution is normal

3. the part distribution is normal and the measurement error distribution is t_5,

4. both the part and measurement error distributions are t_5.

Although this distributional exchange does not confirm the robustness of leveraging, it does give some evidence to that effect. Since the variance of a t_5 is 5/3, the random variables simulated from this distribution were scaled by $\sqrt{3/5}$. Note that in the simulation all the estimates of ρ were truncated to fall between zero and one to match common practice and to reduce the unrealistic variation produced by very large or small values. The results of the simulation, based on 10,000 replicates for 23 values of ρ spread over the interval (0.01,0.99), are shown in Figure 4.3.

When comparing the other cases to Case 1 in Figure 4.3, we see that changing the part or measurement error distribution has surprisingly little effect other than a small increase in the standard deviation for all estimators and a change in the bias of the ANOVA estimate.

In summary, we can assess the assumptions of model (1.1) as easily as for the LP as for the SP. Departures from the model affect both plans, but there is no evidence, based on our very cursory study, that the LP has greater sensitivity to these departures than does the SP. To be convincing, a much larger study is required.

4.5 The Leveraged versus Standard Plan

To demonstrate the value of leveraging, we resort to simulation. We consider five situations. We will compare the LP to the SP where we chose $k = 10$ parts and

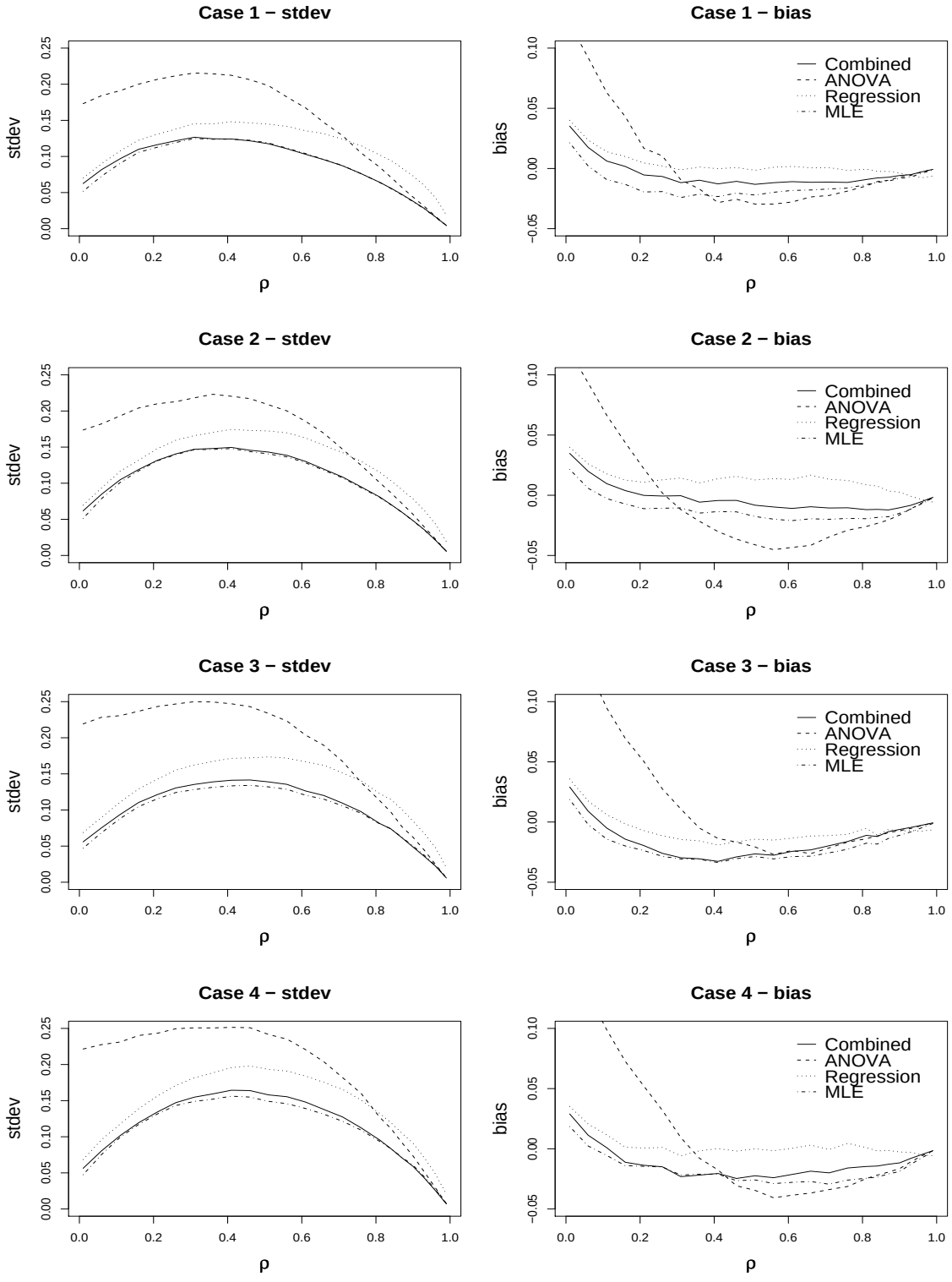

Figure 4.3: The standard deviation and bias of the LP estimators of ρ for different distributional assumptions

make $n = 6$ repeated measurements for each part as recommended in Automotive Industry Action Group [2002]. We used maximum likelihood estimation in all cases to make the comparisons fair. We quantify the difference between the plans using bias and standard deviation calculated from 10,000 simulations at each value of ρ spread over the interval (0.01,0.99) with higher density where the bias and standard deviation are changing rapidly. Also, recall that in most realistic situations ρ is fairly large.

Comparison 1: $N = 60$

In this comparison, we assume that no parts have been previously measured. We compare the following two plans each with a total of 60 measurements

- SP with k =10 and n=6
- LP with $b = 30$, $k = 6$ and $n = 5$.

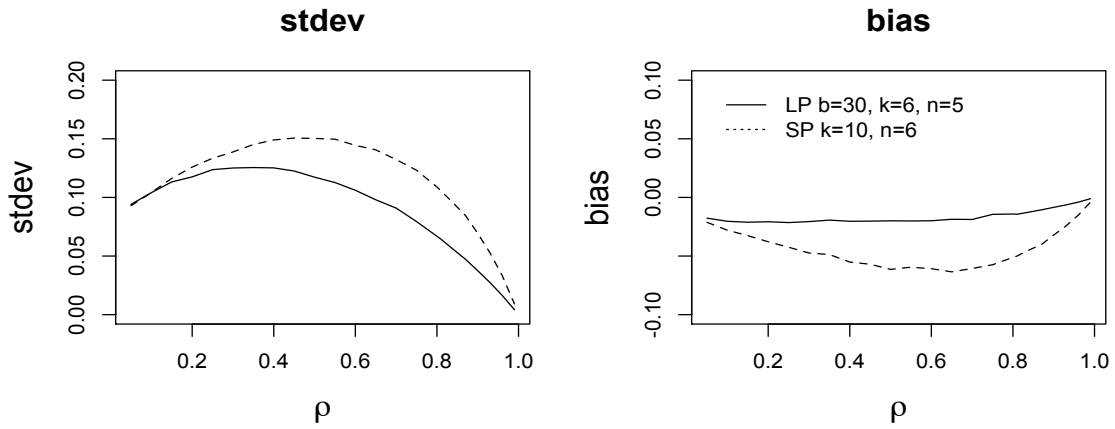

Figure 4.4: Comparison of Standard Deviation and Bias for a Leveraged and Standard Plan

We quantify the difference between the plans using bias and standard deviation calculated from simulation. We see from Figure 4.4 that the LP is substantially better than the SP with smaller standard deviation and bias for all values of ρ.

We can also compare the LP and SP by looking at the total number of measurements required to give a desired precision in the estimation of ρ. Figure 4.5 shows the total number of measurements required for an LP to have the same precision (standard deviation) as the SP ($k = 10, n = 6$) for different values of ρ. In Figure 4.5, the selected LP corresponds to the suggested plan from Section 4.3. For example, at $\rho = 0.91$, when the SP has a standard deviation of 0.060 (see Figure 4.4) the LP with the same standard deviation for estimating ρ has a total sample size of 34, where $k = 3$, $n = 5$ and $b = 19$.

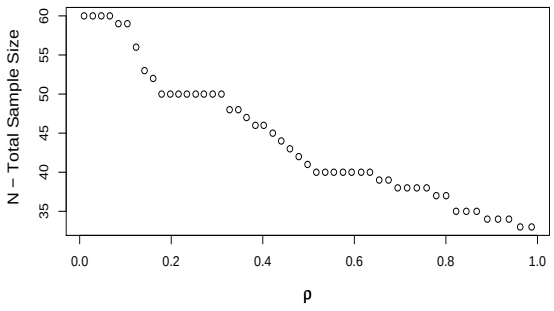

Figure 4.5: The Sample Size N Required for a Leveraged Plan to Have the Same Standard Deviation as the Standard Plan ($k = 10$, $n = 6$) Across the Range of ρ

Comparison 2: a free baseline

Suppose we have a baseline sample with $b = 100$ from a previous investigation. We compare the following four sampling plans.

- SP with $k = 10$ parts selected at random from the baseline and $n = 6$ (does not include baseline measurements).
- LP with $k = 10$ and $n = 6$ (selecting 10 parts at random). Note this is equivalent to having an SP with $k = 10$, $n = 6$ and including the baseline measurements into the analysis.

- LP with $k = 2$ and $n = 2$ (selecting the largest & smallest parts from the baseline).
- LP with $k = 10$ and $n = 6$ (selecting the 5 largest & 5 smallest parts from the baseline).

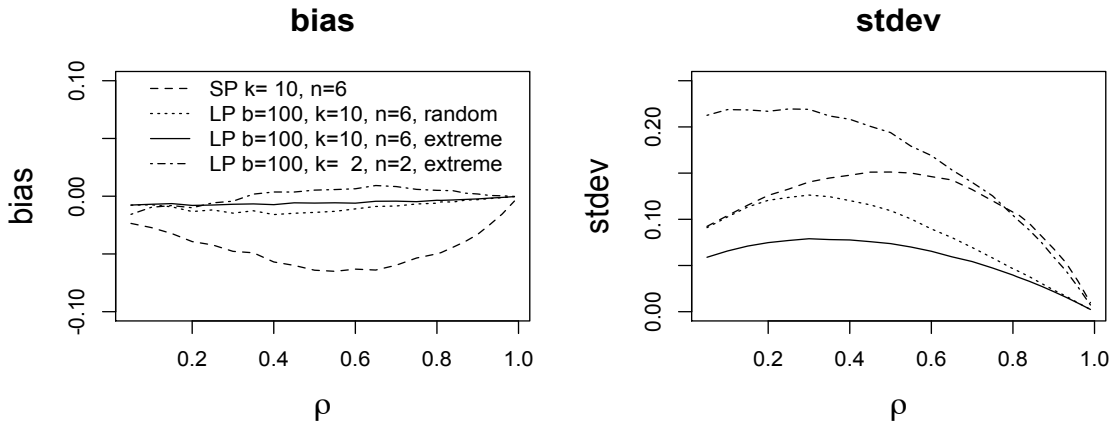

Figure 4.6: Comparison of Standard Deviation and Bias of $\tilde{\rho}$ for an LP and SP

For the SP, it is common practice to estimate ρ using only the information from the 10 repeatedly measured parts, but the estimation procedure can be modified to include the baseline sample information. We include this second SP as the LP with $k = 10$, $n = 6$ and selecting 10 parts at random because this is equivalent to having an SP with $k = 10$, $n = 6$ and including the baseline measurements into the analysis. Figure 4.6 shows that the LP is more efficient than the SP in this situation. Moreover, an LP with $b = 100$, $k = 2$, $n = 2$, is as efficient as the SP when, in the analysis of the SP, the baseline information is ignored.

Not all of these four plans/analyses have the same number of total measurements, so in some ways this is an unfair comparison. However, Figure 4.6 illustrates that the baseline information should not be ignored when planning a measurement assessment. See also Browne et al. [2007]. We show this comparison to promote the use of baseline information in the analysis. Also, we suggest that planning a

measurement assessment study after a baseline study is an effective use of resources in terms of the number of the measurements. Baseline data are often readily and cheaply available from regular production.

Comparison 3: Optimal Designs

In the previous two comparisons we fixed the standard plan design to be $k = 10$ and $n = 6$. In this subsection we compare the optimal design for both the LP and SP given ρ and the total sample size N. For each plan the optimal design has the (b, k, n) or (k, n), for the LP or SP respectively, which minimizes the asymptotic standard deviation of the MLE for ρ given the total sample N and ρ. To find the optimal designs we used an exhaustive search. That is, for a fixed N we calculate the asymptotic standard deviation for all combinations of b, k and n such that $b + kn = N$.

Figure 4.7 displays an optimal plan comparison when N is set at either 60 or 100. For each setting of N, two plots are shown. The left plot is the ratio of standard deviations for the optimal LP and SP when using MLE to estimate ρ and fixing N. The right plot shows the sample size of an LP that has the same standard deviation as the optimal SP for a given ρ and fixed N. For example in the top right plot, when $\rho = 0.4$ the curve indicates that there is an LP with 45 measurements that has the same standard deviation as the optimal SP with 60 measurements.

When N equals 60 or 100, the LP has maximum standard deviation that is at most roughly 15% more efficient than the optimal SP. This occurs when ρ is around 0.4. Since the SP is a subset of the LP the ratio of standard deviations is less 1. From the right panel of Figure 4.7, the largest savings by using the optimal LP instead of the optimal SP is about 25% of the original sample size.

Figure 4.8 expands on Figure 4.7 by including the total sample size as a variable.

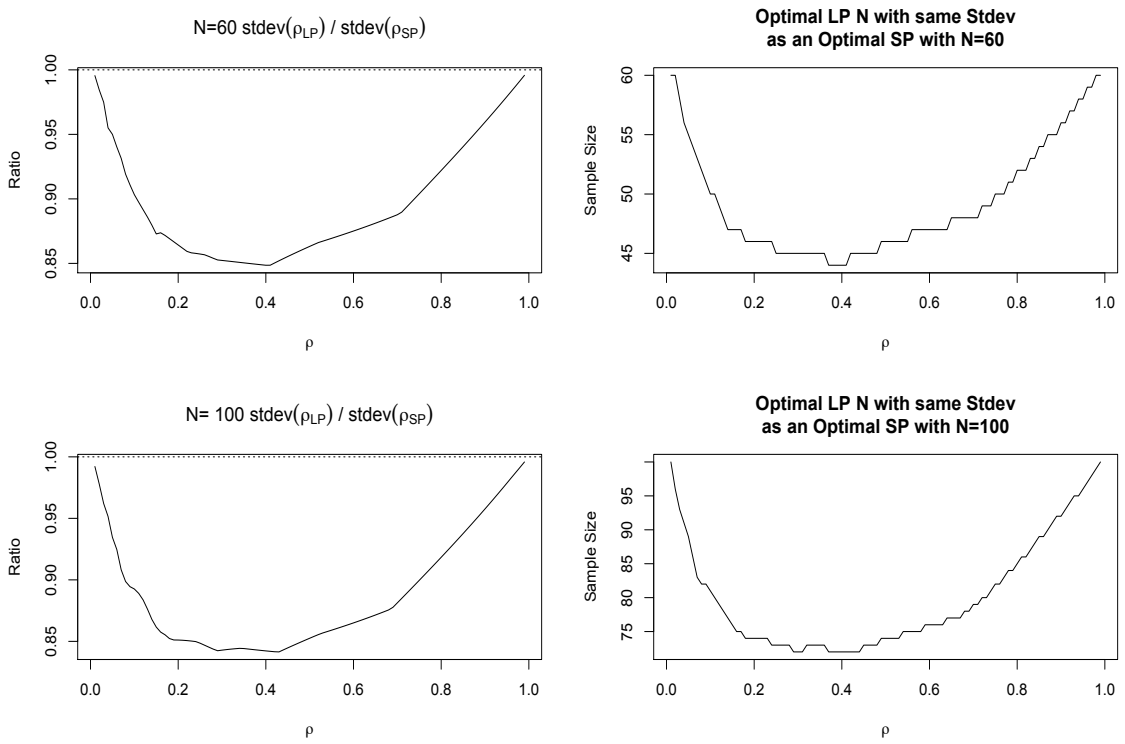

Figure 4.7: Comparison of Optimal LP and SP. The left panels show the ratio of standard deviations for the optimal LP and SP. The right panels show the sample size of an LP that has the same standard deviation as the optimal SP that has N=60 or 100 measurements.

The left panel of Figure 4.8 uses contours to demonstrate that the optimal LP has a smaller standard deviation than the SP when using maximum likelihood to estimate ρ. Although, the region explored in Figure 4.8 does not include values of N larger than 250, we can modestly assume that the contours hold for larger N. In fact, the contours depend on ρ more than N. The most common region for ρ is near one. Here we see modest gains from using the LP.

The right panel 4.8 shows contours of $N - N_o$ by ρ and N where N_o is defined as the number of measurements required for an optimal LP to have the same standard deviation as an optimal SP with N measurements and a given ρ. For example, if we were planning an SP with resources for 150 (N) measurements and we thought ρ was 0.8, then we could save approximately 20 ($N - N_o$) measurements by using

an LP with 130 (N_o) measurements and have no loss of precision.

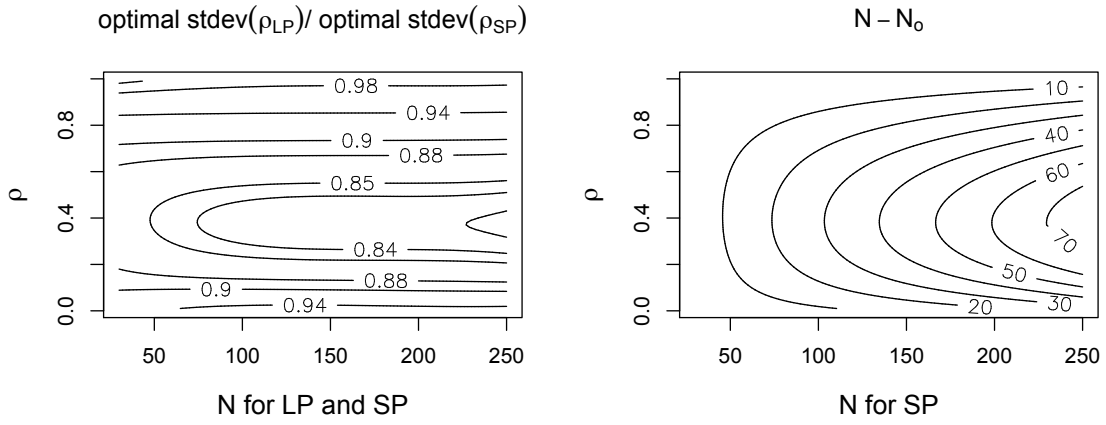

Figure 4.8: Left panel shows the ratio of standard deviations for the optimal LP and SP when using ML to estimate ρ. Right panel shows the difference $N - N_o$ where N_o is defined as the number of measurements required for an optimal LP to have the same standard deviation as an optimal SP with N measurements and a given ρ.

Comparison 4: Optimal SP versus the Recommended LP

An optimal plan is dependent on knowing the value of ρ. When we do not have this information we would have to guess it. In contrast, the recommended LP design given by $k = \lfloor N/10 \rfloor$, $n = 5$ and $b = N - 5\lfloor N/10 \rfloor$ does not depend on ρ. We compare the optimal SP and recommended LP in Figure 4.9. Both panels show that the recommended LP is more efficient than the SP when we restrict ourselves to the region $\rho \geq 0.1$. For a typical measurement system we would expect a small measurement error, thus ρ should be near 1. Although, when $\rho < 0.1$, the optimal SP quickly becomes more efficient than the LP but again, in measurement system studies, ρ near zero rarely occurs.

Restricting to the region $\rho \geq 0.1$, both panels of Figure 4.9 quantify the advantage of using the recommended LP instead of the SP. Similar, to Figure 4.8, the left

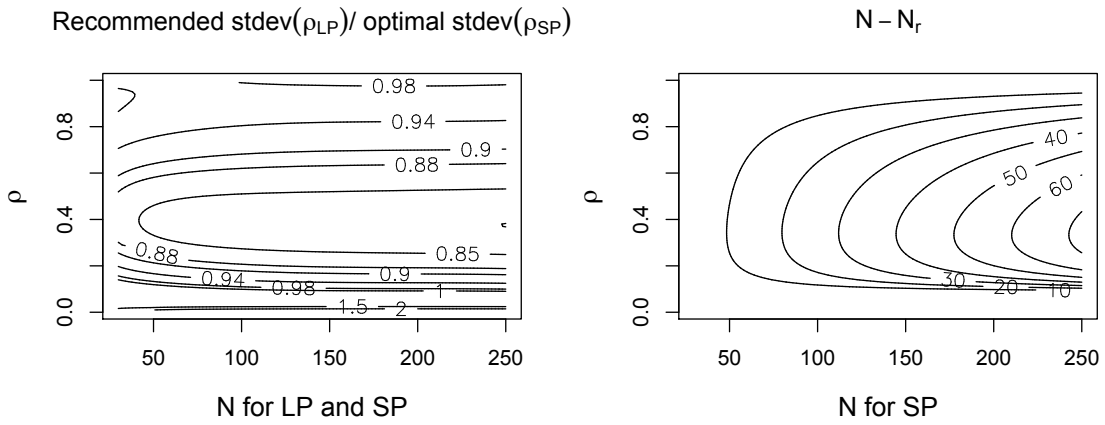

Figure 4.9: Left panel shows the ratio of standard deviations for the recommended LP and optimal SP when using ML to estimate ρ. Right panel shows contours of $N - N_r$ by ρ and N where N_r is defined as the number of measurements required for the recommended LP to have the same standard deviation as an optimal SP with N measurements and a given ρ.

panel of Figure 4.9 shows contours of the ratio of standard deviations from the recommended LP and SP given N and ρ. The right panel of Figure 4.9 shows contours of $N - N_r$ by ρ and N where N_r is defined as the number of measurements required for the recommended LP to have the same standard deviation as an optimal SP with N measurements and a given ρ. For example, if we were planning an SP with resources for 100 (N) and we thought ρ was 0.8, then we could save approximately 10 ($N - N_r$) measurements by using the recommended LP with 90 (N_r) measurements and have the same standard error. This means we can obtain modest gains, over the optimal SP, from using the recommended LP when we restrict ourselves to the region $\rho \geq 0.1$.

Comparison 5: Non-normality

This subsection examines the robustness of the Leveraged and Standard plans when the part and measurement distributions have longer tails. We repeat the experiment in Section 4.4 which is a full factorial experiment, where the measurement and

part distributions which are usually normal are exchanged with t_5 distributions. The results shown in this subsection do not constitute a complete study on the robustness of the LP or SP but they give some insight as to how they perform under different distributional assumptions.

An important comparison is against the recommended SP design, $k = 10$ and $n = 6$. We choose to compare this SP against the recommended LP design, $b = 30$, $k = 6$ and $n = 5$. The results for both the LP and SP are shown in Figure 4.10. When $\rho \geq 0.2$ the LP has a lower MSE in all four cases. In addition, when the part distribution is t_5, i.e. cases 2 and 4, the SP's MSE becomes considerably worse than the LP. However, both the LP's and SP's MSE increase when we stray from normality.

Figure 4.11 compares the optimal SP and LP under the various distributional assumptions. The optimal designs for both plans were chosen based on minimizing the Fisher information when $\rho = 0.80$ and $N = 100$. The SP design was $k = 50, n = 2$ and the LP was $b = 52, k = 16, n = 3$. Figure 4.11 suggests that under different distributional assumptions these two plans are comparable when $\rho \geq 0.8$. When we consider $\rho \leq 0.80$, the LP has a lower MSE in all four cases.

We used a limited robustness study to see how the recommended SP and LP act under different distributional assumptions. If the sample size were larger, we would be able to detect changes in the part and measurement distribution and then modify our analysis to compensate for the change.

4.6 Conclusions

In this chapter, we have presented a new leveraged plan for estimating the intraclass correlation coefficient of a measurement system. We defined leverage to be the

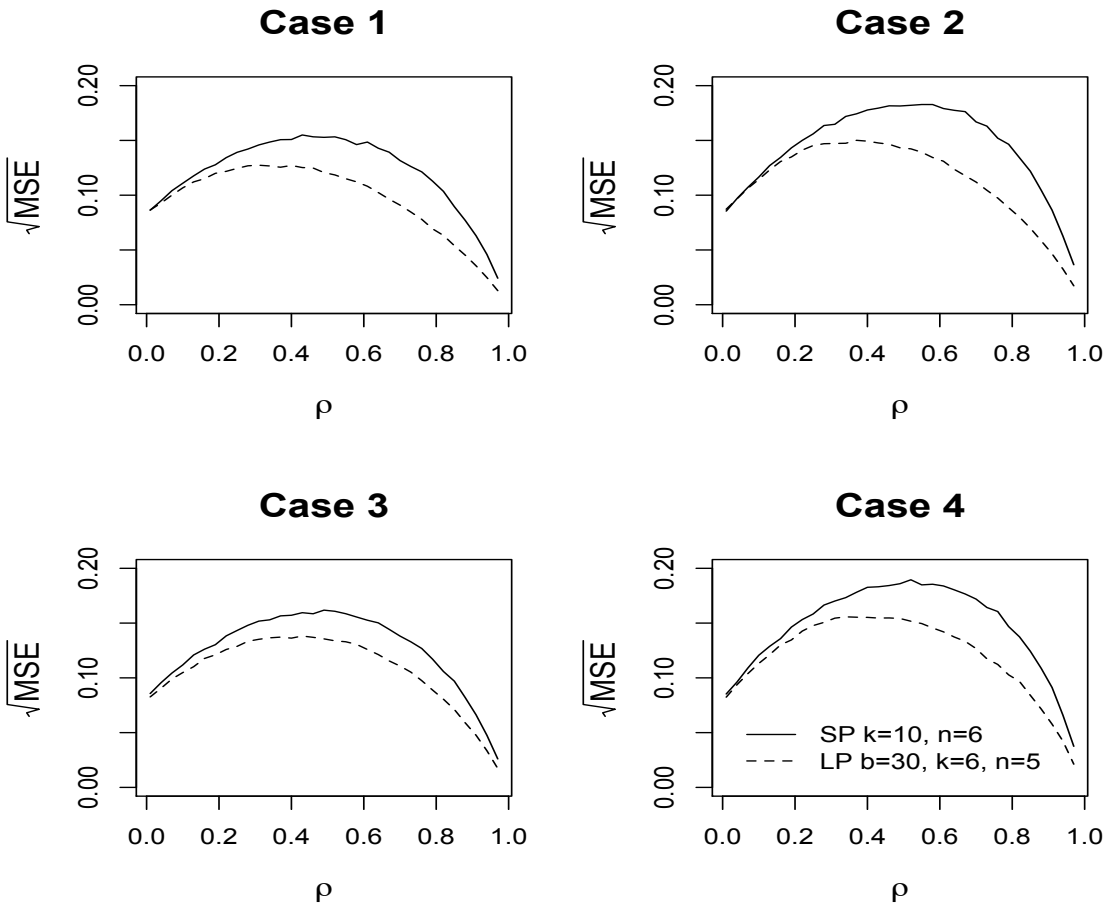

Figure 4.10: The MSE of the MLE's for the SP ($k = 10, n = 6$) and LP ($b = 30, k = 6, n = 5.$) under different distributional assumptions; 1. both the part and measurement error distribution are normal, 2. the part distribution is t_5 and the measurement error distribution is normal, 3. the part distribution is normal and the measurement error distribution is t_5, 4. both the part and measurement error distributions are t_5.

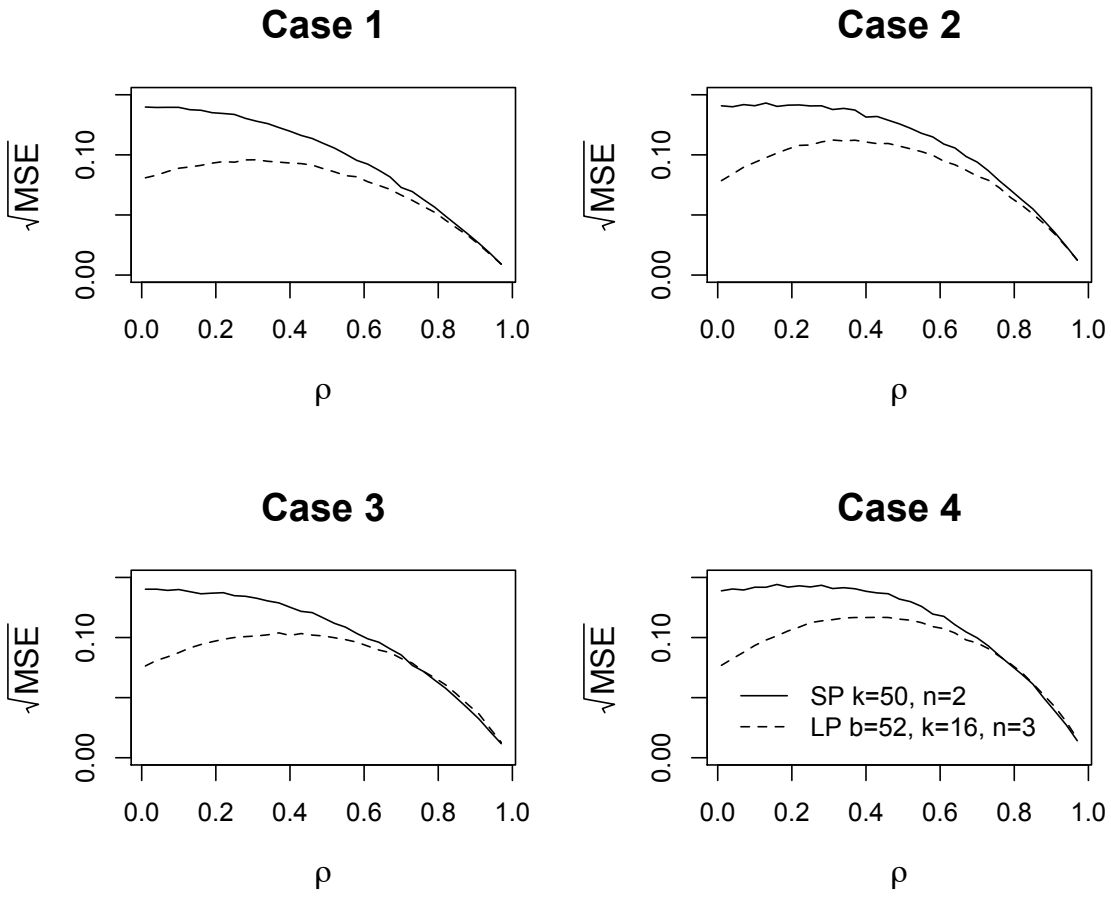

Figure 4.11: The MSE of the MLE's for the SP ($k = 50, n = 2$) and LP ($b = 52, k = 16, n = 3.$) under different distributional assumptions; 1. both the part and measurement error distribution are normal, 2. the part distribution is t_5 and the measurement error distribution is normal, 3. the part distribution is normal and the measurement error distribution is t_5, 4. both the part and measurement error distributions are t_5.

purposeful selection of parts with extreme initial measured values to be remeasured. We showed that the leveraged plan with the same number of total measurements is more efficient than the standard plan. We provided a closed form estimator for ρ that performs as well as the maximum likelihood estimator. We recommend an LP for a fixed number of total measurements N that has a baseline sample of size $b = N - 5\lfloor N/10 \rfloor$, in Stage 2 we select a sample of the $k = \lfloor N/10 \rfloor$ most extreme parts from the baseline and we repeatedly measure each selected part $n = 5$ times.

We demonstrated that the LP is superior to the SP under a variety of conditions. If we restrict ourselves to reasonable values of ρ, i.e. $\rho \geq 0.1$, the recommended LP has a lower MSE than even the optimal SP that depends on the unknown ρ.

Chapter 5

Leveraged Plan for a Measurement System with 100% Inspection

5.1 Context

In this chapter, we compare the leveraged plan with the standard plan, in a manufacturing setting which has a high volume in-production measurement system and measured values are stored. For this chapter, we assume that there is no operator effect (as is the case with many automated measurement systems) and that the process involves 100% inspection. The consequences of this context are

- there are a large number of once-measured parts available,
- the total process variation σ_t^2 and mean μ are known (or estimated with negligible error). We call μ and σ_t the baseline information.

We first compare the LP and SP based on the mean squared error (MSE) of the estimates for ρ. Then, we compare the power of the LP and SP when testing

the hypothesis (1.19) to determine whether the measurement system is acceptable. To compare study plans, we compare the power of these tests over all values of ρ when each test has equivalent size and the same number of total measurements.

5.2 Design

The LP plan is similar to the SP but instead of using a random sample, we sample parts based on their initial measurements as recorded. In particular, we sample parts that are extreme relative to the process known mean μ.

Since a large number of parts are measured in the regular process, we assume that for a measurement study we can select a part with any observed initial measurement (as long as we do not ask for a highly extreme part). To obtain these parts, we can wait until a part is measured close to the desired measurement or find the part associated with the desired measurement from the part inventory.

Recall that in the SP, we measure a sample of $k = 10$ parts $n = 6$ times each. In this context, with the LP, k parts are selected based on their initial measurement from a large sample of measured parts and each selected part is then re-measured n times.

5.3 Estimation

We present three approaches for estimating ρ and testing the hypothesis of interest,

$$H_0 : \rho \leq \rho_0 \text{ versus } H_A : \rho > \rho_0.$$

The first method of estimation is Maximum Likelihood. The second is based on a regression estimate. A third is based on a weighted average of the regression

estimate and the ANOVA estimate using baseline information (from Section 2.2.1).

5.3.1 Maximum Likelihood

The log-likelihood for n repeated measurements on k parts is the same as (4.4). To estimate ρ, we can maximize (4.4) numerically. The Wald test used to test the hypothesis in (1.19) is found as in Section 2.2.1. However, the Fisher information is a special case of (4.10) because the chosen initial measurements are no longer random and the baseline parameters are known. The Fisher information is

$$J_{L2}(\rho) = \frac{1}{2}\frac{kn(n+1)(n\rho^2+1)}{(1+n\rho)^2(1-\rho)^2} + \frac{n\,(SSS-k)}{(1+n\rho)(1-\rho)}, \tag{5.1}$$

The quantity SSC in (4.11) is replaced with the quantity SSS to signify that it is no longer random in this context because we choose parts with specific initial measurements. We define SSS as sum of the squared standardized values of the initial measurements.

$$SSS = \sum_{i=1}^{k} z_{i0}^2 = \sum_{i=1}^{k} \left[\frac{y_{i0}-\mu}{\sigma_t}\right]^2. \tag{5.2}$$

From (5.1) we see that on the standardized scale, increasing the z_{i0}^2's will increase the Fisher information. Figure 5.1 shows the effect on the power of selecting 10 parts, with specified SSS value. It is based on the asymptotic power of the Wald test as shown in (2.17) with $J_{s2}(\rho)$ replaced by $J_{L2}(\rho, y_0)$, as given in (5.1). SSS $= 40$ corresponds to 10 parts with initial values equal $\mu \pm 2\sigma_t$ or any other set of $\{y_{10}, \ldots, y_{k0}\}$ such that $\sum_{i=1}^{k}\left[\frac{y_{i0}-\mu}{\sigma_t}\right]^2 = \sum_{i=1}^{k} z_{i0}^2 = 40$. If we select 10 parts at random, the expected value of $SSS = 10$. This means that Figure (5.1) shows the difference between random and non-random (choosing the most extreme) sampling.

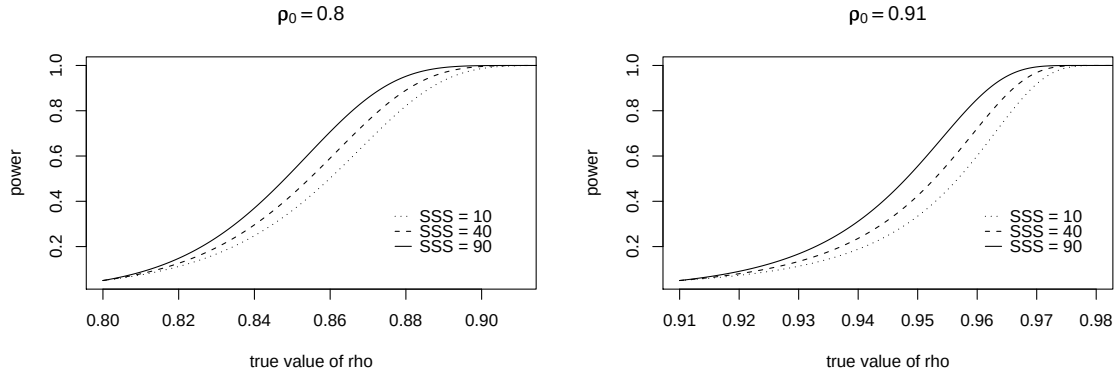

Figure 5.1: Power Curves for Testing the Hypothesis (1.19) when $\rho_0 = 0.80$ and $\rho_0 = 0.91$ at $\alpha = 0.05$, with six repeated measurements and ten parts having SSS = 10, 40 and 90.

5.3.2 Regression Estimator

When μ and σ_t are known the regression estimate (4.13) becomes

$$\widehat{\rho}_r = \frac{\sum_{i \in S} (\overline{y}_{i.} - \mu)(y_{i0} - \mu)}{\sum_{i \in S} (y_{i0} - \mu)^2}$$

By defining $r_i = (\overline{y}_{i.} - \mu)/\sigma_t$ and $z_{i,0} = (y_{i,0} - \mu)/\sigma_t$ the regression estimate can be written as

$$\hat{\rho}_r = \frac{\sum_{i=1}^{k} r_i z_{i0}}{\sum_{i=1}^{k} z_{i0}^2} \tag{5.3}$$

The estimator is unbiased and has variance

$$Var\,(\widetilde{\rho}_r; \rho) = \frac{(1-\rho)(\rho + 1/n)}{\sum_{i=1}^{k} z_{i0}^2} = \frac{(1-\rho)(\rho + 1/n)}{SSS}. \tag{5.4}$$

where SSS is defined in (5.2). We see from the denominator of (5.4) that this estimator has smaller variance when we choose parts that increase the sum of squares of the standardized initial measurements.

Using $\widehat{\rho}_r$, we reject the hypothesis $\rho \leq \rho_0$ if

$$\frac{\widehat{\rho}_r - \rho_0}{Var(\widetilde{\rho}_r; \rho_0)^{1/2}} > Z_{1-\alpha} \tag{5.5}$$

The power of the test when $\rho = \rho_1$ is given by

$$P\left\{Z \geq \left(Z_{1-\alpha}\left[Var(\widetilde{\rho}_r; \rho_0)\right]^{1/2} + \rho_0 - \rho_1\right)\left[Var(\widetilde{\rho}_r; \rho_1)\right]^{-1/2}\right\} \tag{5.6}$$

Note that this test based on the regression estimator does not use the information from the variability of the repeated measurements to help to estimate ρ.

5.3.3 ANOVA Estimator

When μ and σ_t are known the ANOVA estimate (4.19) reduces to

$$\widehat{\rho}_a = 1 - \frac{\text{MSW}}{\sigma_t^2}. \tag{5.7}$$

Transforming the ANOVA estimator, we see that $(1 - \tilde{\rho}_a)/(1 - \rho)$ has a χ^2-distribution with $k(n-1)$ degrees of freedom. Using its distributional properties we have

$$E(\widetilde{\rho}_a) = 1 - (1 - \rho) = \rho\ \left(\frac{b-1}{b-3}\right) - \frac{2}{b-3}$$

$$Var(\widetilde{\rho}_a) = (1-\rho)^2 Var\left(\chi^2_{k(n-1)}\right) = \frac{2(1-\rho)^2}{k(n-1)} \tag{5.8}$$

This estimator performs poorly unless ρ is near one but when combined with the regression estimator the result has some good properties.

5.3.4 Combined Estimator

In the next subsection, we show that the test based on the regression estimator does not perform well when compared to the MLE. To improve the power, we propose to estimate ρ using a weighted average of the regression and ANOVA estimators when ρ is large.

The variances of the estimators (5.8) and (5.4) depend on ρ which makes finding optimal weights impossible since ρ is unknown. We can, however, find optimal weights for a given value of $\rho = \rho_0$, the hypothesized value in (1.19).

If we have two unbiased independent estimators of ρ, $\tilde{\rho}_1$ and $\tilde{\rho}_2$ with known variances σ_1^2 and σ_2^2, the optimal linear combination is

$$\tilde{\rho} = w_1\,\tilde{\rho}_1 + w_2\,\tilde{\rho}_2 = \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2}\,\tilde{\rho}_1 + \frac{1/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2}\,\tilde{\rho}_2 \tag{5.9}$$

and thus the combined estimator has variance

$$Var\,(\tilde{\rho}) = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \tag{5.10}$$

Now, if we obtain $\widehat{\rho}_1$ using (5.3) and $\widehat{\rho}_2$ with (5.7), it can be shown that the two estimators are independent. Applying (5.9), we define the combined leveraged estimate as

$$\begin{aligned} \hat{\rho}_c &= w_1\widehat{\rho}_r + w_2\widehat{\rho}_a \\ \hat{\rho}_c &= w_1\left(\frac{\sum_{i=1}^k r_i z_{i,0}}{\sum_{i=1}^k z_{i,0}^2}\right) + w_2\left(1 - \frac{\mathrm{MSW}}{\sigma_t^2}\right) \end{aligned} \tag{5.11}$$

$$\text{with} \quad w_1 = \frac{2n\sum_{i=1}^k z_{i,0}^2(1-\rho_0)}{2n(1-\rho_0)\sum_{i=1}^k z_{i,0}^2 + k(n-1)(\rho_0 n + 1)}$$

$$\text{and} \quad w_2 \quad = \frac{k(n-1)(\rho_0 n+1)}{2n(1-\rho_0)\sum_{i=1}^{k} z_{i,0}^2 + k(n-1)(\rho_0 n+1)}$$

Using (5.10), the variance of $\tilde{\rho}_c$ when $\rho = \rho_0$ is

$$Var\left(\tilde{\rho}_c; \rho = \rho_0\right) = \frac{2(1-\rho_0)^2(\rho_0 n+1)}{2n(1-\rho_0)\sum_{i=1}^{k} z_{i,0}^2 + k(n-1)(\rho_0 n+1)}. \tag{5.12}$$

When $\rho = \rho_1$, the variance of $\tilde{\rho}_c$ is

$$Var\left(\tilde{\rho}_c; \rho = \rho_1\right) = 2(1-\rho_1)\frac{\left[2n(1-\rho_0)^2(\rho_1 n+1)\sum_{i=1}^{k} z_{i,0}^2 + k(n-1)(\rho_0 n+1)^2(1-\rho_1)\right]}{\left[2n(1-\rho_0)\sum_{i=1}^{k} z_{i,0}^2 + k(n-1)(\rho_0 n+1)\right]^2}. \tag{5.13}$$

We construct the test of the hypothesis (1.19) using the normal approximation for the estimator. The approximate power of the test is

$$P\left\{Z > \left(Z_{1-\alpha}\left[Var(\tilde{\rho}_c; \rho = \rho_0)\right]^{1/2} + \rho_0 - \rho_1\right)\left[Var(\tilde{\rho}_c; \rho = \rho_1)\right]^{-1/2}\right\}. \tag{5.14}$$

Here we use a normal approximation of the distribution of MSW. Actually, MSW follows a chi-square distribution with $k(n-1)$ degrees of freedom. A normal approximation of a chi-square distribution is reasonable if $k(n-1)$ is larger than say, 30. The default choice for the SP is $k = 10$ and $n = 6$ which means $k(n-1) = 50$.

To create a confidence interval for ρ we assume normality of the estimator and use the variance of the combined estimator as given in (5.12).

5.3.5 Comparison of Leveraged Plan Analysis Methods

We compare the three estimation methods for an LP using the power of the tests because the combined estimator requires the specification of ρ_0 for estimation. We

use the power from the Wald tests from (5.6) and (5.14) to generate the power curve for the Leveraged MLE. In all the comparisons, we use a sample of ten parts with $SSS = \sum_{i=1}^{k} z_{i,0}^2 = 40$ and six repeated measurements on each part.

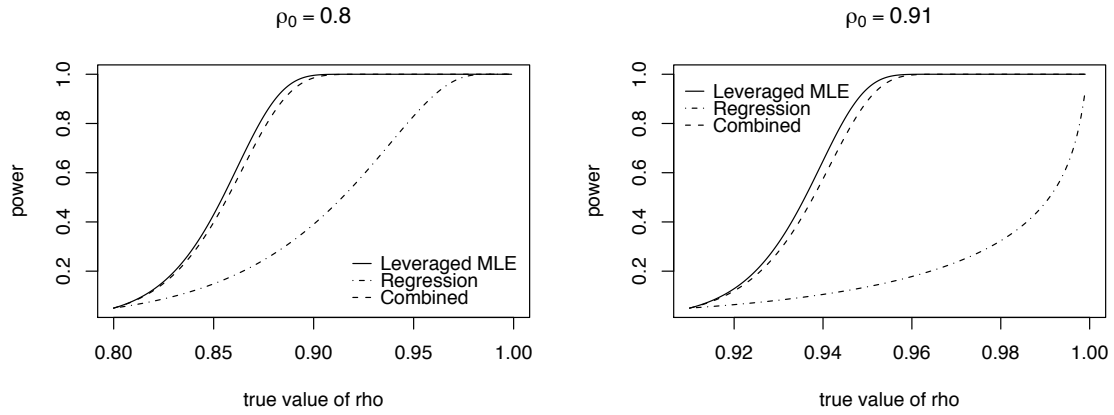

Figure 5.2: Power Curves for Testing the Hypothesis (1.19) when $\rho_0 = 0.80$ (left), $\rho_0 = 0.91$ (right), ten parts having $\sum z_{i,0}^2 = 40$ and $k = 10$, $n = 6$.

Figure 5.2 shows that the test based on the regression estimator performs poorly and that the test based on the combined estimator performs almost as well as the Wald test based on the MLEs, with the advantage of having a closed form solution.

5.4 Leveraged Plan Design

The goal of a measurement investigation is to demonstrate that the measurement system is reliable. In terms of model parameters, the goal can be translated to stating (with some uncertainty) that ρ is greater than some predetermined value denoted as ρ_0. This goal is achieved by rejecting the test of hypothesis in (1.19). Figure 5.3 shows the true values of ρ, denoted as ρ_1, that have 0.80 power when testing ρ_0 at various values of k (number of parts) and n (number of repeated measurements). Thus, these contours show the (n, k) combinations that satisfy the required size and power for the assumed values of ρ.

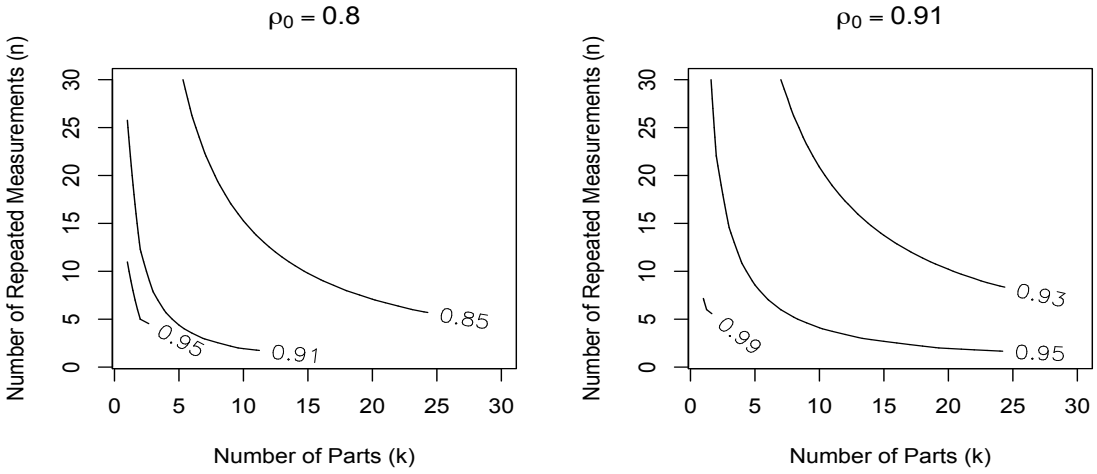

Figure 5.3: Contours of ρ for Testing (1.19) when $\rho_0 = 0.80$ and $\rho_0 = 0.91$ at $\alpha = 0.05$, power $= 0.80$, with $\sum_{i=1}^{k} z_{i,0}^2 = 4k$.

To obtain the contours in Figure 5.3, we determined, for all discrete points (n, k), the values of ρ that have power 0.80 using the asymptotic Wald test when testing ρ_0 in (1.19). To determine the power of the Wald test for Leveraged MSA, we use (2.17) but instead of $J_{s2}(\rho)$, we substituted $J_{L2}(\rho)$ as given in (5.1).

Figure 5.3 can be used to determine the necessary sample sizes for testing (1.19) when $\rho_0 = 0.80$ and $\rho_0 = 0.91$ with size 0.05 and power 0.80. For example, suppose we wish to determine if the process variability is 80% or more of the total variation (i.e. $\rho_0 = 0.80$) and it is currently thought that ρ is around 0.95. The point on Figure 5.3, with $\rho_0 = 0.80$, that corresponds to five parts with $\sum_{i=1}^{5} z_{i,0}^2 = 20$ and five repeated measurements on each part satisfies the needs of this investigation because it is above the $\rho = 0.95$ contour.

For another example, suppose we are planning an investigation to determine if the discrimination ratio, as defined in (1.6), is greater than 2 and we want to detect a ratio greater than 3 with 0.80 power. As shown in Table 1.1, these discrimination ratio values correspond to $\rho = 0.80$ and 0.90, respectively. Any combination of a number of parts and repeated measurements which is above and to the right of the

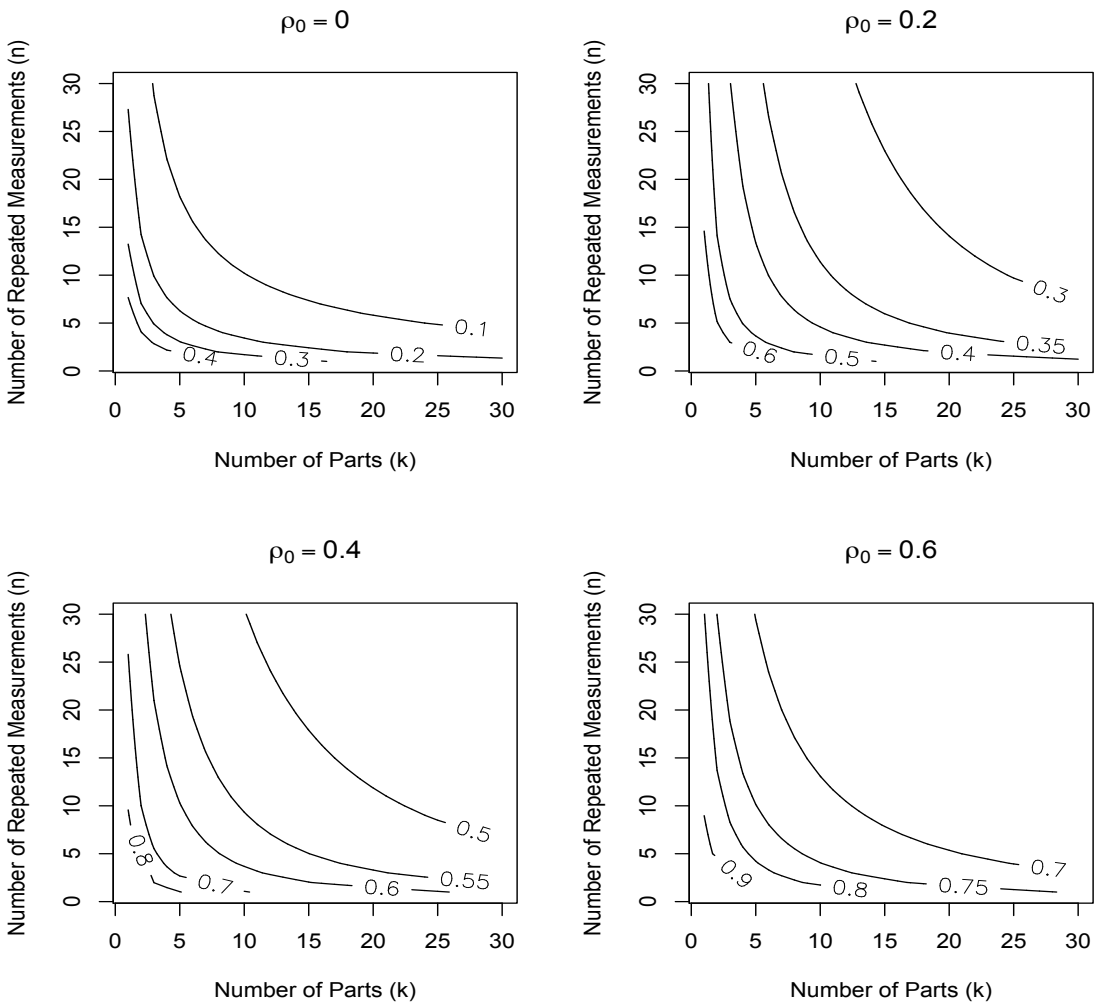

Figure 5.4: Contours of ρ for Testing (1.19) when $\rho_0 = 0, .2, .4$ and $\rho_0 = 0.6$ at $\alpha = 0.05$, power $= 0.80$, with $\sum_{i=1}^{k} z_{i,0}^2 = 4k$.

0.90 contour on Figure 5.3, with $\rho_0 = 0.80$, will be satisfactory.

Examining Figure 5.3 shows that theoretically a design with $k \leq 2$ can satisfy the power requirements. This type of design, though possible, is not recommended. A more desirable plan is to use several extremes on either side of the mean and to choose a couple of initial measurements close to the mean. This type of design will also help to check if the measurement variance is constant across the range of true part dimensions.

Figure 5.4 can be used in the same same way as Figure 5.3 for testing (1.19) when $\rho_0 = 0, 0.2, 0.4$ and 0.6 with size 0.05 and power 0.80.

In practice one will not be able to find k parts with initial measures of exactly some predetermined value, for example, $\mu \pm 2\sigma_t$. Fortunately, the analysis is conditional on the set of initial measurements $\{y_{01}, y_{02}, \ldots, y_{0k}\}$ because the distribution of these measurements does not depend on ρ. The conditionality principle is regularly used in regression where the analysis is conditional on the covariates. See Cox and Hinkley [1974] for a discussion of the conditionality principle. As seen in (5.12) and (5.13) we obtain a reduction in the variance as long as $\sum_{i=1}^{k} z_{i,0}^2$ is large. The results presented in Figure 5.6, given in section 5.5, illustrate the effectiveness of leveraging and Figure 5.3 is useful to assist in planning an LP.

5.5 Leveraged versus Standard Plan

In this subsection, we show that leveraging is beneficial whenever there is a supply of parts with measured (and recorded) values so that we can select extremes. Such a supply will be available if the measurement system is used routinely in production or if the measurement system assessment is preceded by a baseline study to assess the performance of the process. There would be a small cost to find the extreme

parts. What is the gain? To address this question we first compare the MSE and the power curves for leveraged and standard plans with the same number of parts and repeated measurements.

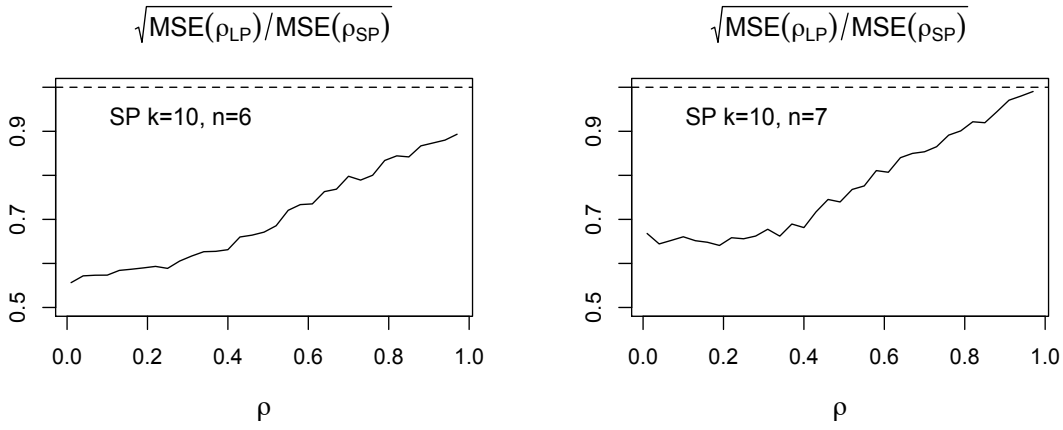

Figure 5.5: The ratio of the square root of the MSE for the LP ($n = 6$ and ten part with standardized initial measurements $z_0 = 2$) and two SPs with $k = 10, n = 6$ (left panel) and $k = 10, n = 7$ (right panel).

The ratios of the MSE for LP and SP are shown in Figure 5.5. Two SP designs are compared against one LP design. The LP design has $n = 6$ repeated measurements and ten parts with standardized initial measurements $z_0 = 2$, i.e. $SSS = 40$. This LP design was chosen to match the common practice SP design, $k = 10$ and $n = 6$. The left panel shows the comparison with the most common SP that has 10 parts measured six times each. For illustration, the right panel shows the comparison with the SP with $k = 10$ and $n = 7$. We include the 1 additional measurement per part since the LP has $n + 1$ measurements on each part, if we include the extreme initial measurement.

Next we compare the power for the same LP and SP discussed above. We generated the power curves for testing the hypothesis in (1.19) using the Maximum Likelihood Wald test for the LP and SP. The power of the Wald test for the SP is given by (2.17). The power of the Wald test for Leveraged MSA uses the same

formula but instead of $J_{s2}(\rho)$ it uses $J_{L2}(\rho)$ which is given in (5.1). We used Maximum Likelihood for both power curves because this way, the two designs can be evaluated fairly.

Figures 5.6 and 5.7 show that the LP is substantially more powerful than the SP when the 10 parts selected have $SSS = 40$ and six repeated measurements on each part.

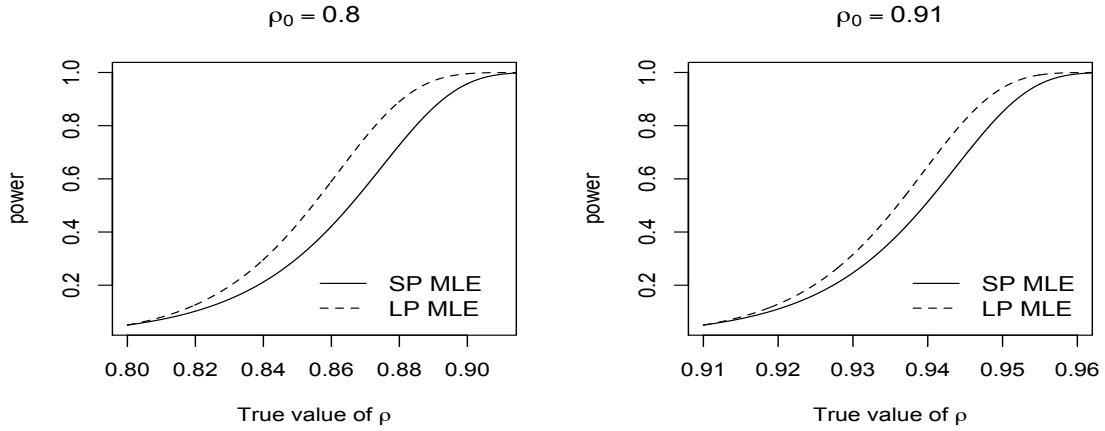

Figure 5.6: Power Curves for the LP and SP from Testing (1.19) when $\rho_0 = 0.80$, $\rho_0 = 0.91$, ten parts having $\sum z_{i,0}^2 = 40$ and 6 repeated measurements.

We can also quantify the effects of leveraging by comparing sample sizes. Suppose we select five parts with initial measured values that are two standard deviations from the mean and then measure each part five times. Using the test based on the combined estimate, the power of the test for $\rho_0 = 0.8$ at $\rho = 0.9$ is about 0.80. To get the same power with the unleveraged plan as the standard plan and the test based on the ANOVA estimate, we would need to measure each of five randomly selected parts seven times.

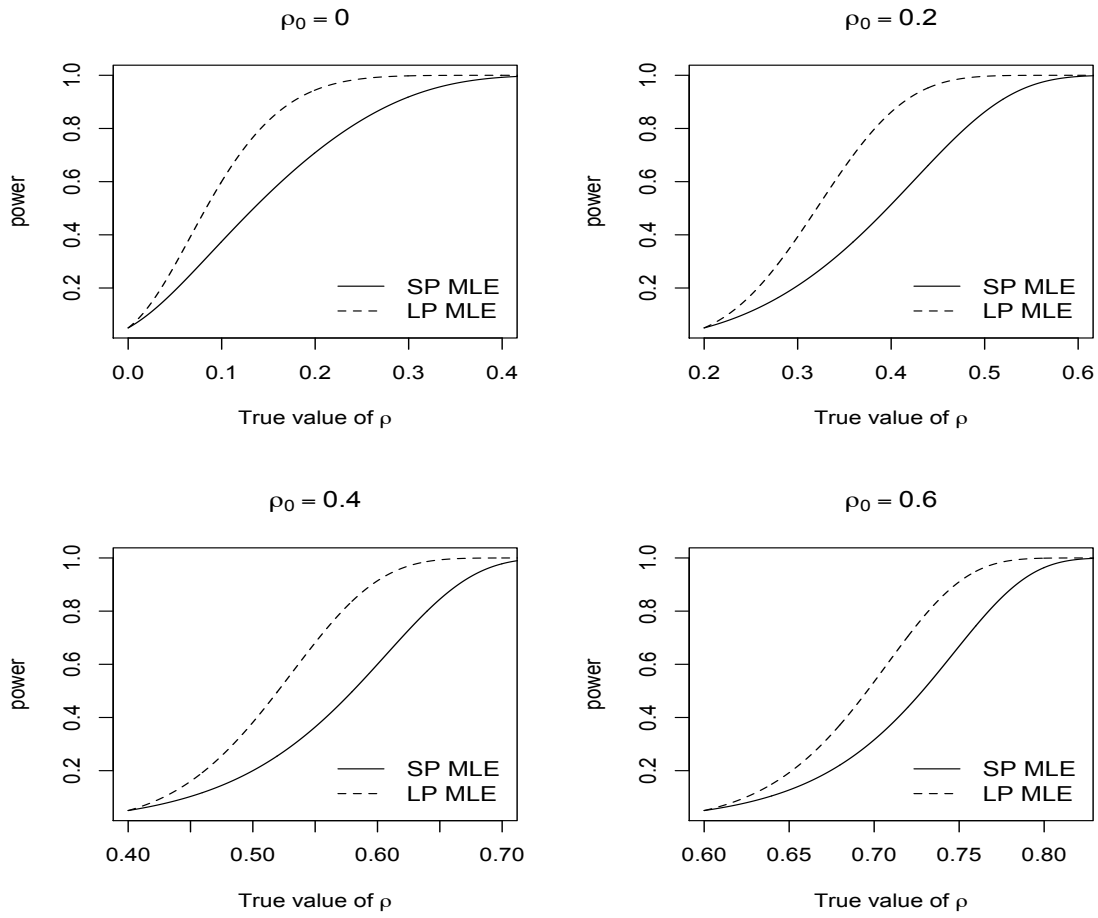

Figure 5.7: Power Curves for the LP and SP from Testing (1.19) when $\rho_0 = 0, .2, .4$ and $\rho_0 = 0.6$, ten parts having $\sum z_{i,0}^2 = 40$ and 6 repeated measurements.

5.6 Conclusions and Discussion

In this chapter we presented a new leveraged assessment plan when the process mean μ and overall process variance σ_t^2 are known or estimated with neglible error. This would be the case for a process with 100% inspection, since then every part produced is measured at least once. Under these conditions, we showed that leveraging increases the precision of estimation of ρ, the intraclass correlation coefficient and increases the power of the hypothesis test in (1.19) compared to the standard plan. Complicated analysis is not required to realize this gain in power. Section 5.3.4 shows that the estimator (5.11) which combines the regression (5.3) and ANOVA using baseline information (4.19) estimators achieves almost the same power as the MLE. The benefit of leveraging increases when more extreme parts are chosen in the LP.

We also provided a methodology for designing an LP to meet pre-specified power values in (1.19).

In conclusion, when we have 100% inspection, leveraging can increase power or correspondingly, the precision of the estimate of ρ, with no increase in cost.

Chapter 6

Leveraged Plan for a Measurement System with Operators

6.1 Context

In many measurement systems, operators are thought to be a substantial source of variability. Each operator is assumed to have a different mean effect on measurements so that there are relative biases among the operators. In this chapter, we extend our previous work on leveraging in Chapters 4 and 5 to allow multiple operators. Each operator is treated as a fixed effect denoted by μ_i. That is, there is a relative bias between two measurements on the same part by different operators. Recall from Section 1.3 that we extended model (1.1) and (2.1) to include operators, using a mixed effect model given by

$$Y_{ijk} = \mu_j + X_i + E_{ijk} \tag{6.1}$$

where X_i is a random effect of the true part dimension, μ_j is the mean effect from operator j, and E_{ijk} is the random effect from repeatedly measuring the same part with the same operator. The random variables X and E are assumed to be independent normals with zero means and standard deviations σ_p, σ_g, respectively. The mean of the true part dimension is included into the mean effect for each operator. Also, in Section 1.3 we defined

$$\underline{\mu} = (\mu_1, \ldots, \mu_m)^t, \tag{6.2}$$

$$\overline{\mu} = \frac{1}{m}\sum_{j=1}^{m} \mu_j \tag{6.3}$$

$$\sigma_{pg}^2 = \sigma_p^2 + \sigma_g^2 \tag{6.4}$$

$$\sigma_m^2 = \sigma_o^2 + \sigma_g^2 \tag{6.5}$$

$$\text{and} \qquad \sigma_t^2 = \sigma_o^2 + \sigma_p^2 + \sigma_g^2 \tag{6.6}$$

$$\text{where} \qquad \sigma_o^2 = \frac{1}{m}\sum_{j=1}^{m} (\mu_j - \overline{\mu})^2, \tag{6.7}$$

m is the number of operators and μ_j is the mean for j^{th} operator. The parameter σ_{pg} is the variation seen in measurements made by any single operator on a sample of parts from the process. The parameter σ_t represents the total variation seen in the process if each operator measured the same proportion of parts in regular production. The parameter σ_o captures the variation due to differences among the m operator means (i.e. the effects of relative bias), but is not a standard deviation in the usual sense. Finally, the parameter σ_m represents the total variation seen in the measurement of any particular part if each operator is used in the system with the same intensity. If each operator has the same mean, then $\sigma_0 = 0$ and we can interpret σ_m and σ_t as standard deviations as defined in model (1.1). Using manufacturing jargon, σ_m represents the overall measurement variability, σ_g the repeatability and σ_o the reproducibility. We assume σ_g is the same for each operator

and part.

We shall compare the LP and SP using the two metrics:

- the intraclass correlation coefficient or the process variation divided by the total variation
$$\eta = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_o^2 + \sigma_g^2} = \frac{\sigma_p^2}{\sigma_t^2}, \tag{6.8}$$
- and the proportion of measurement variation attributable to operator bias,
$$\lambda = \frac{\sigma_o^2}{\sigma_g^2 + \sigma_o^2}. \tag{6.9}$$

We will treat η as the primary parameter of interest while λ is of secondary interest. Both quantities η and λ are bounded between zero and one. If η is larger than $1/2$ the process variation is the dominant source of the overall variation. Similarly, if λ is larger than $1/2$ the variation due to operator differences is the dominant source of variation in the measurement system. Note that the overall quality or reliability of a measurement system depends on η and not λ. A good measurement system will have a value of η near one. If we have a poor measurement system, examining λ can help us determine how to improve the measurement system. For example, if $\lambda < 1/2$, reducing the gauge variation will have a larger impact than addressing the operator relative biases.

In this chapter, we introduce two new Leveraged plans (LPs), denoted A and B, where we deliberately select extreme parts to re-measure from an initial baseline sample. The two LPs have the same Stage 2 but different Stage 1. For both LPs, in Stage 2 every operator measures each of the k selected parts n times. The baseline or Stage 1 for LP A has m operators each measuring b different parts. The baseline for LP B has m operators each measuring the same b parts. Both LPs require $b \times m$ measurements in Stage 1 but LP A has $b \times m$ parts whereas LP B has b parts.

We demonstrate the advantages of the LPs over the standard plan (SP) by comparing the standard deviations of the estimators for η and λ defined in (6.8) and (6.9), respectively.

6.2 Leveraged Plan A

6.2.1 Plan

This plan is conducted in two stages:

Stage 1: Sample $(b \times m)$ parts at random from the process to obtain a baseline. Then allocate b parts to each of the m operators. Each operator measures their b allocated parts once. In this stage, no part is measured twice. We denote the observed values from part i and operator j as $\{y_{ij0}\}$ for $i = 1, \ldots, b$ and $j = 1, \ldots, m$. We label the Stage 1 parts and operators as $\{(1,1), \ldots, (b,m)\}$

Stage 2: From the baseline sample, select k parts using the observed measured values. In particular, to improve the estimation for η, sample k parts such that both

- the initial measurements are extreme relative to their operator average, defined as $\left|y_{ij0} - \overline{y}_{.j0}\right|$ where $\overline{y}_{.j0} = \frac{1}{b}\sum_{i=1}^{b} y_{ij0}$, and
- the average of the initial measurements of the selected parts is close to the baseline average.

Next each operator measures each of the k parts n times to give the additional data $\{y_{ijlh},\ \ (i,j) \in S,\ \ l = 1, \ldots, m \ \text{ and } \ h = 1, \ldots, n\}$ where S is a subset of $\{(1,1), \ldots, (b,m)\}$ with k elements. The total number of measurements in the leveraged plan is $N = m(b + nk)$.

For example, in a leveraged plan A with $b = 10, m = 3, k = 6$ and $n = 2$, we first sample 30 parts at random from the process and allocate 10 parts to each of the 3 operators. The operators measure each of their 10 assigned parts once. Then, we pick the parts with the minimum and maximum initial measurement from each operator. These $k = 6$ parts are then used in the standard plan where each operator measures each part twice. This plan has a total of $10 \times 3 + 6 \times 3 \times 2 = 66$ measurements. Alternatively, we might pick the six parts which are most extreme relative to the operator averages, so the number of parts used in Stage 2 is not necessarily balanced over operators in Stage 1. Note that the standard plan is not a special case or subset of a leveraged plan A.

6.2.2 Estimation

We form the likelihood using the parameterization $\left\{\underline{\mu} = (\mu_1, \ldots, \mu_m), \sigma^2_{pg} = \sigma^2_p + \sigma^2_g, \rho = \frac{\sigma^2_p}{\sigma^2_p + \sigma^2_g}\right\}$ because it simplifies the likelihood and we apply the likelihood (4.4) from Section 4.2.1. Note, $\underline{\mu}$ is defined in (1.9) and represents the vector of operator means.

To obtain the likelihood for the LP, we (as in Section 4.2.1) decompose the likelihood into two pieces by conditioning on the baseline measurements. The log-likelihood for the LP is

$$l_{L3}(\mu, \sigma^2_{pg}, \rho) = l_{b3}(\underline{\mu}, \sigma^2_{pg}) + l_{r3}\left(\underline{\mu}, \sigma^2_{pg}, \rho \;\middle|\; y_{ij0}\; i = 1, \ldots, b\; j = 1, \ldots, m\right) \quad (6.10)$$

To get the complete likelihood, we begin with the likelihood for Stage 1. In the baseline, each of the m operators measure b parts for total of $(b \times m)$ measurements. Assuming the $(b \times m)$ parts in the baseline sample are selected at random from the

process, the marginal log-likelihood of the baseline is

$$l_{b3}(\mu, \sigma^2_{pg}) = -\frac{bm}{2}\log\sigma^2_{pg} - \frac{1}{2\sigma^2_{pg}}\left\{\sum_{j=1}^{m}\sum_{i=1}^{b}\left(y_{ij0} - \overline{y}_{.j0}\right)^2 + b\sum_{j=1}^{m}\left(\overline{y}_{.j0} - \mu_j\right)^2\right\}. \tag{6.11}$$

where $\overline{y}_{.j0} = \frac{1}{b}\sum_{i=1}^{b} y_{ij0}$ is the average of the baseline measurements for operator j.

To obtain the likelihood for Stage 2, we first consider a single part i (selected to be repeatedly measured) which is initially measured by operator j. The joint distribution of this initial measurement Y_{ij0} and the n repeated measurements from each of the m operators is

$$\begin{pmatrix} Y_{ij0} \\ \underline{Y}_{ij1n} \\ \vdots \\ \underline{Y}_{ijmn} \end{pmatrix} \backsim N\left(\begin{bmatrix} \mu_j \\ \underline{\mu} \otimes \mathbf{1}_n \end{bmatrix}, \quad \sigma^2_{pg}\left[(1-\rho)\,\mathbf{I}_{mn+1} + \rho\mathbf{J}_{mn+1}\right]\right). \tag{6.12}$$

where $\rho = \sigma^2_p/(\sigma^2_p + \sigma^2_g)$, $\underline{\mu} = (\mu_1, \mu_2, \ldots, \mu_m)^t$, $\otimes$ denotes the kronecker product, $\underline{Y}_{ijln} = (Y_{ijl1}, Y_{ijl2}, \ldots, Y_{ijln})^t$ for $l = 1, \ldots, m$, $\mathbf{I}_q$ is an identity matrix with dimension q, $\mathbf{J}_q$ is a square matrix of ones with dimension q and $\mathbf{1}_q$ is column vector of ones with q rows. $\underline{Y}_{ijkn}$ is vector of n measurements from operator k on a part i which was initially measured by the j^{th} operator.

The distribution of the repeated measurements $\left\{\underline{Y}_{ij1n}, \ldots, \underline{Y}_{ijmn}\right\}$ on a single part given the initial measurement $Y_{ij0} = y_{ij0}$ is

$$\left(\begin{array}{c} \underline{Y}_{ij1n} \\ \vdots \\ \underline{Y}_{ijmn} \end{array}\middle|\, Y_{ij0} = y_{ij0}\right) \backsim N\left(\left[\underline{\mu} + \rho(y_{ij0} - \mu_j)\mathbf{1}_m\right] \otimes \mathbf{1}_n, \;\; \Sigma_f\right).$$

where the covariance matrix

$$\Sigma_f = \sigma_{pg}^2(1-\rho)\Big[\mathbf{I}_{mn} + \rho\mathbf{J}_{mn}\Big] \tag{6.13}$$

has a special form which allows us to obtain the following well known properties [Dillon and Goldstein, 1984]:

$$\begin{aligned}
\Sigma_f^{-1} &= \frac{1}{\sigma_{pg}^2(1-\rho)(1+mn\rho)}\Big\{\,(1+mn\rho)\,\mathbf{I}_{mn} - \rho\mathbf{J}_{mn}\Big\} \\
|\Sigma_f| &= \sigma_{pg}^{2mn}(1-\rho)^{mn}(1+mn\rho).
\end{aligned}$$

Using these properties of Σ_f, we can write the conditional likelihood (conditional on y_{ij0}) for the repeated measurements on a single part. The measurements for one part are independent of the measurements from another part. Thus the conditional likelihood for k parts, each with n measurements, is the product of the individual likelihoods. By carefully defining new variables, $z_{ij0} = (y_{ij0} - \mu_j)$, we can use (4.4) to obtain the conditional log-likelihood for all repeated measurements conditional on the baseline. We get

$$\begin{aligned}
&l_{r3}\left(\underline{\mu}, \sigma_{pg}^2 = \sigma_p^2 + \sigma_g^2, \rho = \frac{\sigma_p^2}{\sigma_p^2+\sigma_g^2}\,\middle|\;\; y_{ij0}\right) \\
&= \left[-\frac{mnk}{2}\log\sigma_{pg}^2 - \frac{mnk}{2}\log(1-\rho) - \frac{k}{2}\log(1+mn\rho)\right] \\
&\quad -\frac{1}{2}\frac{1}{\sigma_{pg}^2(1-\rho)(1+mn\rho)}\left\{(1+mn\rho)SSO + mn\sum_{(i,j)\in S}\left[\overline{z}_{ij..} - \rho z_{ij0}\right]^2\right\}
\end{aligned} \tag{6.14}$$

where $z_{ij0} = (y_{ij0} - \mu_j)$ is the baseline measurement centred by the mean for operator j, $SSO = \sum_{(i,j)\in S}\sum_{l=1}^{m}\sum_{h=1}^{n}(z_{ijlh} - \overline{z}_{ij..})^2$, $\overline{z}_{ij..} = \frac{1}{mn}\sum_{l=1}^{m}\sum_{h=1}^{n} z_{ijlh}$ and $z_{ijlh} = (y_{ijlh} - \mu_l)$. The parameterization $\left\{\underline{\mu}, \sigma_{pg}^2 = \sigma_p^2 + \sigma_g^2, \rho = \frac{\sigma_p^2}{\sigma_p^2+\sigma_g^2}\right\}$ simplifies the likelihood but it can also easily be expressed in terms of $\left\{\underline{\mu}, \sigma_p^2, \sigma_g^2\right\}$.

To get the MLEs of $\underline{\mu}, \sigma^2_{pg}$ and ρ, we maximize (6.10). Solutions can be found numerically. Then to get the MLE's for η and λ, we apply the appropriate transformations,

$$\eta = \frac{\rho\sigma^2_{pg}}{\sigma^2_o + \sigma^2_{pg}} \qquad \text{and} \qquad \lambda = \frac{\sigma^2_o}{\sigma^2_o + (1-\rho)\sigma^2_{pg}} \tag{6.15}$$

where σ^2_o is defined in (6.7).

6.2.2.1 Fisher Information

The asymptotic variance-covariance matrix of the maximum likelihood estimators is the inverse of the Fisher information matrix. We obtain the Fisher information matrix by summing several pieces. The first piece comes from the baseline likelihood

$$J_{b3}\left(\underline{\mu}, \sigma^2_{pg}, \rho\right) = \begin{bmatrix} \frac{b}{\sigma^2_{pg}}\mathbf{I}_m & \underline{0}_m & \underline{0}_m \\ & \frac{1}{2}\frac{mb}{\sigma^4_{pg}} & 0 \\ & & 0 \end{bmatrix}, \tag{6.16}$$

where $\mathbf{I}_m$ is the identity matrix with dimension m and $\underline{0}_m$ is a column vector of m zeros. Since Fisher information matrices are symmetric, we do not show the lower triangles in these matrices. In addition to the baseline Fisher information matrix, we have one matrix for every part selected to be repeatedly measured. Each matrix has the following form. Suppose, we have a part i with initial measurement Y_{ij0} which was measured by operator j. Then the Fisher information is

$$J_{r3ij}(\underline{\mu}, \sigma^2_{pg}, \rho) = \begin{bmatrix} \mathbf{M}_0 + \mathbf{M}_{(j)} & \underline{0}_m & \underline{v}_0 + \underline{v}_{(j)} \\ & \frac{1}{2}\frac{mn}{\sigma^4_{pg}} & -\frac{1}{2}\frac{mn\rho(mn+1)}{\sigma^2_{pg}(1-\rho)(1+mn\rho)} \\ & & \frac{1}{2}\frac{nm(nm+1)(nm\rho^2+1)}{(1+mn\rho)^2(1-\rho)^2} + \frac{mn\left(E\left[Z^2_{i10}\right]-1\right)}{(1+mn\rho)(1-\rho)} \end{bmatrix} \tag{6.17}$$

$$\begin{array}{cc} \text{where} \quad \mathbf{M}_0 = n(1+mn\rho)\mathbf{I}_m - n^2\rho\mathbf{J}_m, & \mathbf{M}_{(j)} = n\rho(m\rho-1)\mathbf{I}_{(j)} - n\rho\mathbf{J}_{(m)}, \\ \mathbf{I}_{(j)} = \underline{e}_j\underline{e}_j^t, & \mathbf{J}_{(j)} = \underline{1}_m\underline{e}_j^t + \underline{e}_j\underline{1}_m^t - \underline{e}_j\underline{e}_j^t, \\ \underline{v}_0 = \frac{nE[Z_{i10}]}{\sigma_{pg}(1-\rho)(1+mn\rho)}\underline{1}_m, & \underline{v}_{(i,j)} = \frac{-nm\rho E[Z_{ij0}]}{\sigma_{pg}(1-\rho)(1+mn\rho)}\underline{e}_j, \\ \text{and} & Z_{ij0} = (Y_{ij0}-\mu_j)/\sigma_{pg} \end{array} \tag{6.18}$$

Note, $\underline{e}_j$ is a column vector of zeros of length m except for a 1 at the j^{th} position and $\underline{1}_m$ is a column vector of m 1s. We require special matrices and vectors $\mathbf{M}_{(j)}$ and $\underline{v}_{(j)}$ to represent the extra information obtained about μ_j when we select a part from that operator. Also, $\mathbf{M}_0$ and $\underline{v}_0$ represent the base amount of information obtained about the operator means. To estimate the quantities $E[Z_{ij0}]$ and $E\left[Z_{ij0}^2\right]$ we suggest using the observed value y_{ij0} and the maximum likelihood estimates for μ_j and σ_{pg}. To get the complete Fisher information we sum over all parts used in the 2^{nd} Stage and the baseline Fisher information.

$$J_{L3}\left(\underline{\mu}, \sigma_{pg}^2, \rho\right) = J_{b3}\left(\underline{\mu}, \sigma_{pg}^2, \rho\right) + \sum_{(i,j)\in S} J_{r3ij}(\underline{\mu}, \sigma_{pg}^2, \rho) \tag{6.19}$$

To obtain the Fisher information for a leveraged plan in terms of $(\underline{\mu}, \lambda, \eta)$ we apply the transformation given in (2.24). Specifically, we apply a matrix, D, given in (2.22) to $J_{L31}\left(\underline{\mu}, \sigma_{pg}^2, \rho\right)$. Then the Fisher information for $(\underline{\mu}, \lambda, \eta)$ is

$$J_{L3}\left(\underline{\mu}, \lambda, \eta\right) = DJ_{L3}\left(\mu_1, \ldots, \mu_m, \sigma_{pg}^2, \rho\right) D^t. \tag{6.20}$$

Finally, to obtain the asymptotic variance-covariance matrix for $(\underline{\mu}, \lambda, \eta)$ we take the inverse of $J_{L3}\left(\underline{\mu}, \lambda, \eta\right)$.

6.2.3 Numerical Example

To illustrate how to get estimates, standard errors and check model assumptions for a leveraged plan we simulated data with $m = 3$ operators suggested by the design section with $b = 11$, $k = 3$ and $n = 3$. This means in Stage 1, 33 parts were sampled at random and distributed equally to the three operators. The baseline data are presented in Table 6.1. The three extreme parts chosen for Stage 2 are

Table 6.1: Example Baseline (Stage 1)

	Baseline by Operator		
(i,j)	1	2	3
1	-1.37	**2.16**	0.35
2	-0.05	-0.72	-0.18
3	-0.95	-0.07	0.14
4	-0.22	-1.34	-1.12
5	**1.13**	-1.84	1.20
6	0.22	0.61	-0.98
7	-0.61	-0.93	-0.49
8	0.23	-1.10	1.60
9	-1.24	1.03	-1.20
10	1.11	-0.05	**1.89**
11	-0.96	0.10	0.97
$y_{.j0}$	-0.25	-0.20	0.20
$s^2_{.j0}$	0.76	1.34	1.22

highlighted in bold. These three parts were then measured by each operator $n = 3$ times to yield the data in Table 6.2.

Using the observed data, the maximum likelihood estimates for $\left(\mu_1, \mu_2, \mu_3, \sigma^2_{pg}, \rho\right)$ are (-0.187, -0.095, 0.046, 1.026, 0.997). Using the transformations (6.15), the maximum likelihood estimates for (λ, γ) are $(0.765, 0.108)$. We can also obtain standard errors by applying (6.20). In this example, the standard errors for λ and γ are 0.0609 and 0.0159, respectively. We can then construct approximate confidence intervals using these standard errors.

Table 6.2: Example repeats (Stage 2)

Repeated Measurements for Part (i, j)											
$(5, 1)$				$(1, 2)$				$(10, 3)$			
	Repeats				Repeats				Repeats		
Operator	1	2	3		1	2	3		1	2	3
1	1.16	1.08	1.27	1	2.05	1.96	2.03	1	1.82	1.65	1.73
2	1.23	1.29	1.17	2	2.15	2.15	2.08	2	1.79	1.87	1.83
3	1.35	1.36	1.37	3	2.26	2.37	2.35	3	1.93	1.96	1.94

To check the model fit, we created a QQ plot of the baseline residuals defined as $y_{ij0} - \overline{y}_{.j0}$. This plot is shown in the left panel of Figure 6.1. The right panel shows the baseline residuals by operator which can be used to verify the assumption of constant variance σ_g across the three operators. These two plots show no evidence to dispute the model.

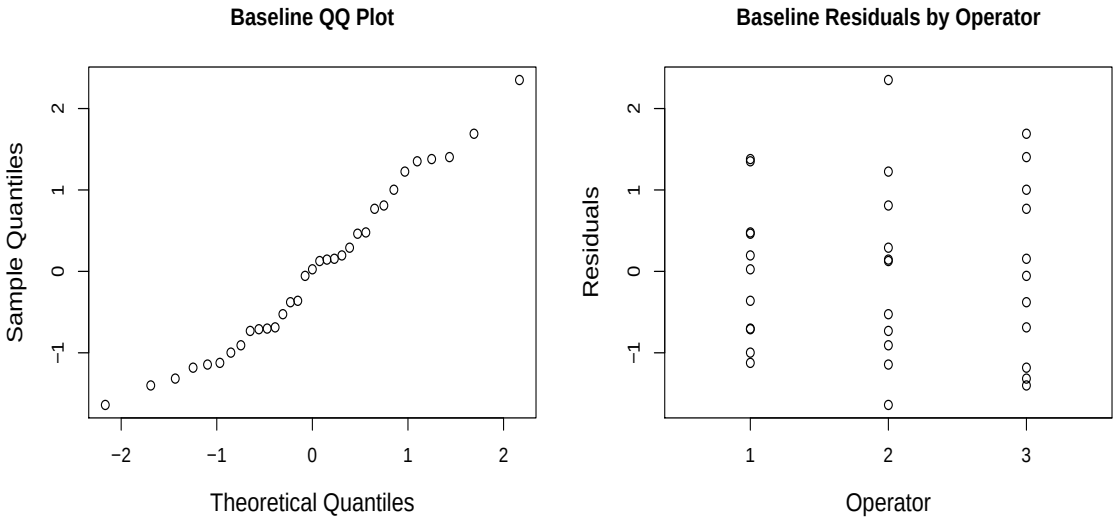

Figure 6.1: Checking model fit. The left panel shows a QQ plot of the baseline residuals $y_{ij0} - \overline{y}_{.j0}$ and the right panel shows these residuals by operator.

To further check the model fit, we created a QQ plot of the Stage 2 residuals defined as $y_{ijlh} - \overline{y}_{ijl.}$ where $\overline{y}_{ijl.} = \frac{1}{n}\sum_{h=1}^{n} y_{ijlh}$. This is shown in the left panel of Figure 6.2. The right panel shows the Stage 2 residuals by operator which can be

used as a further check of constant variance across the three operators. These two plots also show no evidence to reject the model for this example.

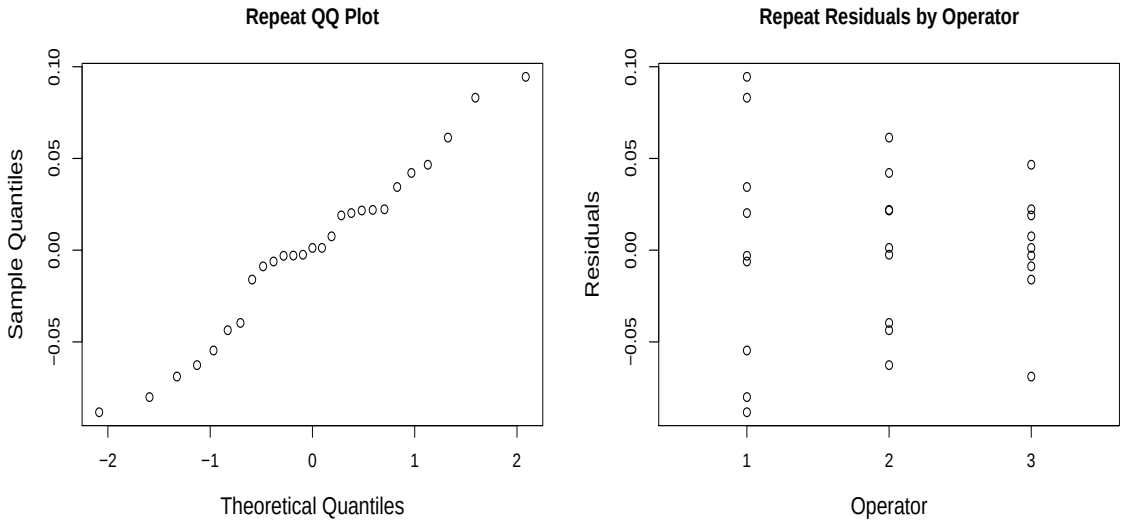

Figure 6.2: Checking model fit. The left panel shows a QQ plot of the 2^{nd} stage residuals $y_{ijlh} - \overline{y}_{ijl.}$ and the right panel shows these residuals by operator.

6.2.4 The Effect of Direction in a Leveraged Plan

This subsection was motivated by the following problem. Suppose in Stage 2, we select all large parts (or all small) from each operator. For example, in one sampling plan we could choose large parts only for each operator. Then, is there a sampling plan that minimizes the asymptotic variance for η? The answer is yes but we show that the benefit of this plan is likely not worth the extra complexity. That is, there is little benefit from selecting all large parts (or all small) from each operator.

Specifying if we select large or small parts for remeasurement affects what we call direction of an LP. The term direction refers to the direction of the vector $\mathbf{u}$, defined in (6.21). Informally, $\mathbf{u}$ is the vector of sums by operator of the initial measurements associated with the parts selected for Stage 2. In Appendix B.1,

we show that asymptotic variance of η can be written as a function of the inner product of

$$\mathbf{r} = (\underline{\mu} - \overline{\mu}\underline{1}_m) \qquad \text{and} \qquad \mathbf{u} = \sum_{(i,j)\in S} E\left[Z_{ij0}\right] \underline{e}_j = \begin{bmatrix} \sum_{(i,1)\in S} E\left[Z_{ij0}\right] \\ \vdots \\ \sum_{(i,m)\in S} E\left[Z_{ij0}\right] \end{bmatrix} \tag{6.21}$$

where $\underline{e}_j$ a column vector of zeros except for the j^{th} position which contains an one. Note $\mathbf{r}$ is a vector of differences between each operator's mean and the overall mean so that $\sigma_o^2 = \frac{1}{m}\mathbf{r}^t\mathbf{r}$.

In Appendix B.1, we show that the asymptotic variance of η is smaller when the inner product $(\mathbf{u}^t\mathbf{r}) > 0$ or when $\mathbf{u}$ and $\mathbf{r}$ tend to point in the same direction. To control the inner product, $(\mathbf{u}^t\mathbf{r})$, we modify the selection method for Stage 2. We use the notation (a, b) to indicate that we chose parts with initial measurements a and b from an operator. For example, if $m = 3$, $\underline{\mu} = (-1, 0, 1)$, and we choose parts with values $(-3, -3)$, $(-2, +2)$ and $(+3, +3)$ from operator one, two and three respectively. Then the vector $\mathbf{u}^t = (-4, 0, 4)$ and the inner product $(\mathbf{u}^t\mathbf{r}) = 4 + 0 + 4 = 8$. In contrast, if we choose parts with values $(+3, +3)$, $(+2, -2)$ and $(-3, -3)$ from operators one, two and three respectively and then the inner product $(\mathbf{u}^t\mathbf{r}) = -4 + 0 - 4 = -8$.

The asymptotic variance of η depends on the Stage 2 sampling plan through the vector $\mathbf{u}$ and $\sum_{(i,j)\in S}\left[Z_{ij0}^2\right]$. This means we can summarize the behaviour of the asymptotic variance of η with two quantities. If we select one part for remeasurement from each operator then $\sum_{(i,j)\in S}\left[Z_{ij0}^2\right]$ reduces to $\mathbf{u}^t\mathbf{u}$. But if we have more than one part from each operator the relationship between $\sum_{(i,j)\in S} E\left[Z_{ij0}^2\right]$ and $\mathbf{u}$ is more complicated.

To explore how changing the Stage 2 selection method affects the asymptotic

standard deviation of the MLE for η, we let $m = 2, b = 15, n = 2$, fix the parameters $(\underline{\mu}, \lambda, \eta)$ by setting $\sigma_t^2 = 1$, $\mu_2 \geq 0$, $\mu_1 = -\mu_2$, $\mu_1 + \mu_2 = 0$ and $(\mu_1^2 + \mu_2^2)/2 = \sigma_o^2$, and select one part from each operator with initial measurements y_{10} and y_{20}. Note, that by applying the above constraints we have $\mathbf{r} = \sigma_o^2(-1, 1)$. Next, to ensure the initial measurements are reasonable, we standardize them with $u_1 = (y_{10} - \mu_1)/\sigma_{pg}$ and $u_2 = (y_{20} - \mu_1)/\sigma_{pg}$ so that $\mathbf{u}^t = (u_1, u_2)$ and $E\left[Z_{ij0}^2\right] = \sigma_{pg}^2(u_1^2 + u_2^2)$. Figure 6.3 displays the contours of the asymptotic standard deviation for the MLE of η when $b = 15$, $k = 2$, $n = 2$ and $m = 2$ versus the initial measurements u_1 and u_2 taken from operator 1 and 2, respectively. In each plot we divide the standard deviation by its minimum to give a comparable scale across the plots. For each panel, the minimum standard deviation occurs when $\mathbf{u}^t = (-2, +2)$. This point has two features; it maximizes the length, $\mathbf{u}^t\mathbf{u}$ and is parallel to or has the same direction as the vector $\mathbf{r} = \underline{\mu} - \overline{\mu}\underline{1}_m$. The other corners also maximize the length. The worst corner $(2, -2)$ still performs well relative to the best corner $(-2, +2)$ because in the top row of plots, the ratio between the minimum and maximum standard deviation is small and in the other plots, maximizing the length $(\mathbf{u}^t\mathbf{u})$ is far more important than the direction $(\mathbf{u}^t\mathbf{r})$. From this interpretation, Figure 6.3 suggests that if we select parts with extreme initial measurements, we can ignore the direction.

We continue to compare the direction and length in Figure 6.4 by comparing two points from Figure 6.3; $(-2, 2)$ the corner which yielded the smallest standard deviation within each plot and $(2, -2)$ the point with the same length that had the largest standard deviation. These corners yield the smallest and largest standard deviations because $\mathbf{r} = \sigma_o^2(-1, 1)$ and these two corners maximize and minimize the inner product $\mathbf{r}^t\mathbf{u}$ while fixing the length. Figure 6.4 demonstrates that by choosing extreme parts, even in the worst direction, we can achieve results comparable to the best direction.

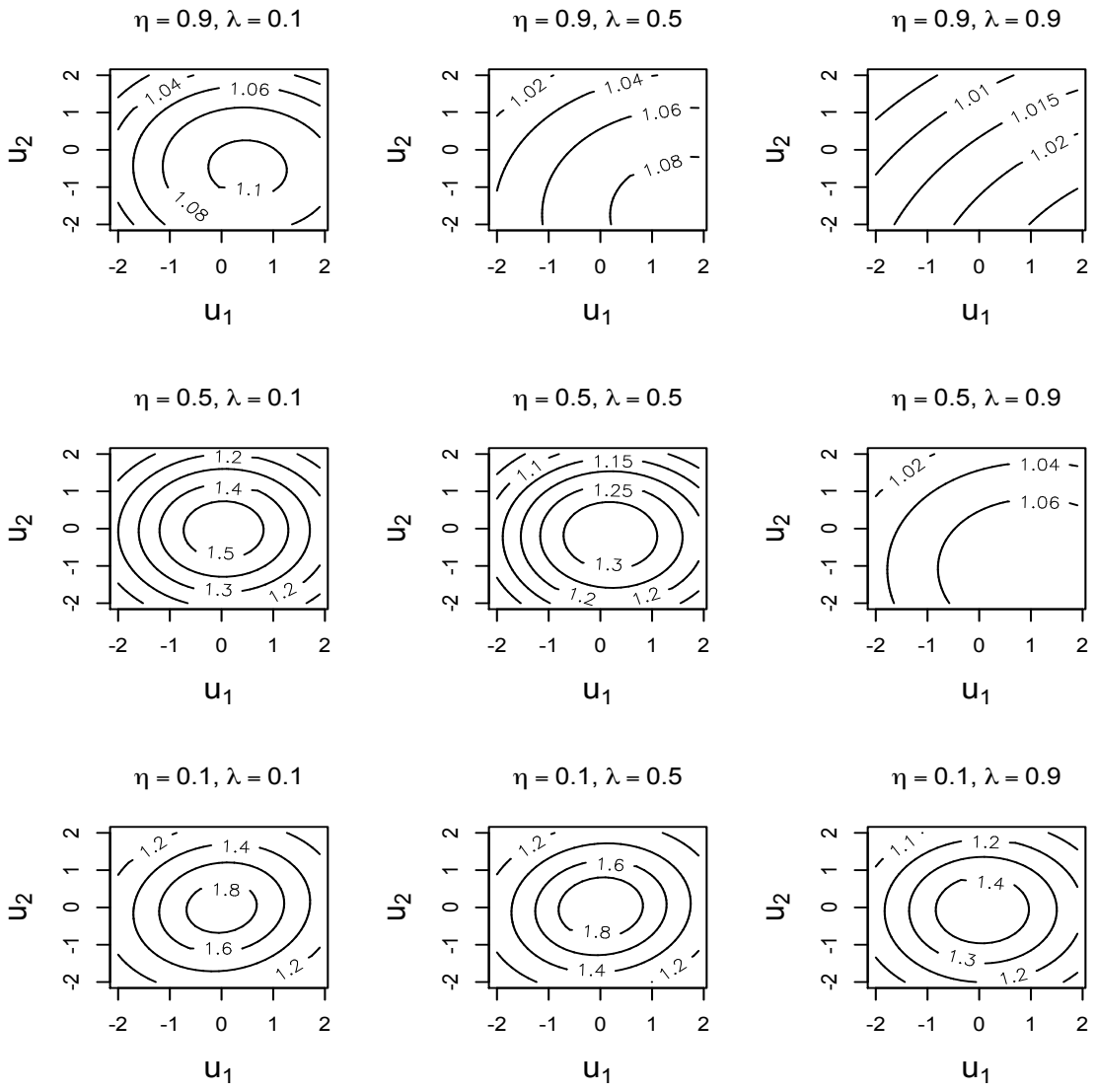

Figure 6.3: The asymptotic standard deviation for the MLE of η when $b = 15, k = 2, n = 2, m = 2$ and the initial measurements u_1 and u_2 are from operator 1 and 2, respectively. Each plot is scaled by its minimum.

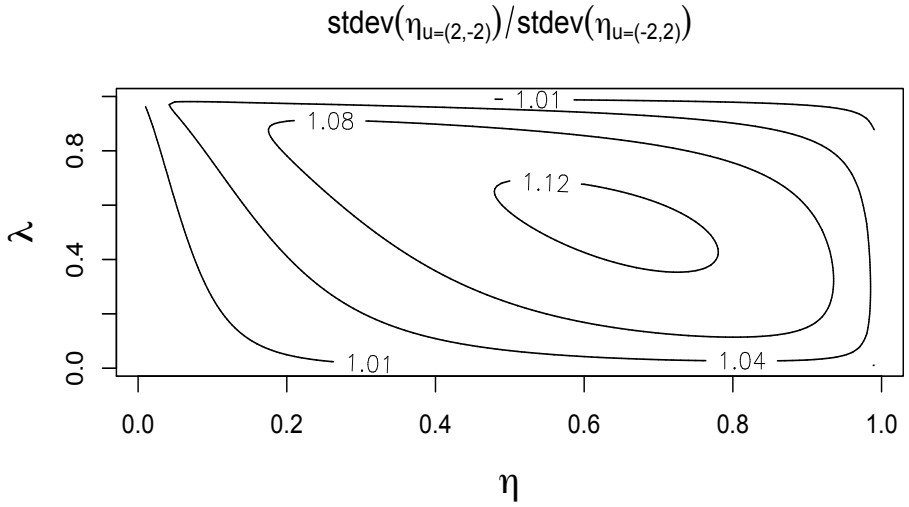

Figure 6.4: The ratio of asymptotic standard deviations for the MLEs of η from $\mathbf{u}^t = (2, -2)$ compare to $\mathbf{u}^t = (-2, 2)$ when $b = 15, k = 2, n = 2, m = 2$ and the initial measurements u_1 and u_2 are from operator 1 and 2, respectively.

To end this section, we explore selection methods for Stage 2 where two parts are selected from each operator. Similar to the previous comparison we allow the standardized initial measurements to take on values between -2 and $+2$ but now we assume there are $m = 3$ operators such that $\underline{\mu} = (-a, 0, a)$ and $a = \sqrt{\frac{3}{2}\sigma_o^2}$. Other compositions of $\underline{\mu}$ are more realistic but since we restricted the standard initial measurements to be between -2 and 2 the chosen composition allows a corner, which maximizes the length, to be parallel to the optimal direction. For this comparison we consider three different selection methods with the same design $b = 18, k = 6, n = 2$. The three selection methods are

1. The optimal selection method selects parts along the direction of $(\underline{\mu} - \overline{\mu}\underline{1}_m)$ which means we select parts with standardized initial measurements $(-2, -2)$ from operator 1, $(-2, +2)$ from operator 2 and $(+2, +2)$ from operator 3.

2. The balanced selection method selects two parts, one with a large and one with a small initial measurement, which means we select parts with standardized

initial measurements $(-2, +2)$ from each of the three operators.

3. The worst selection method selects parts in the opposite direction of $(\underline{\mu} - \overline{\mu}\underline{1}_m)$ which means we select parts with standardized initial measurements $(+2, +2)$ from operator 1, $(-2, +2)$ from operator 2 and $(-2, -2)$ from operator 3.

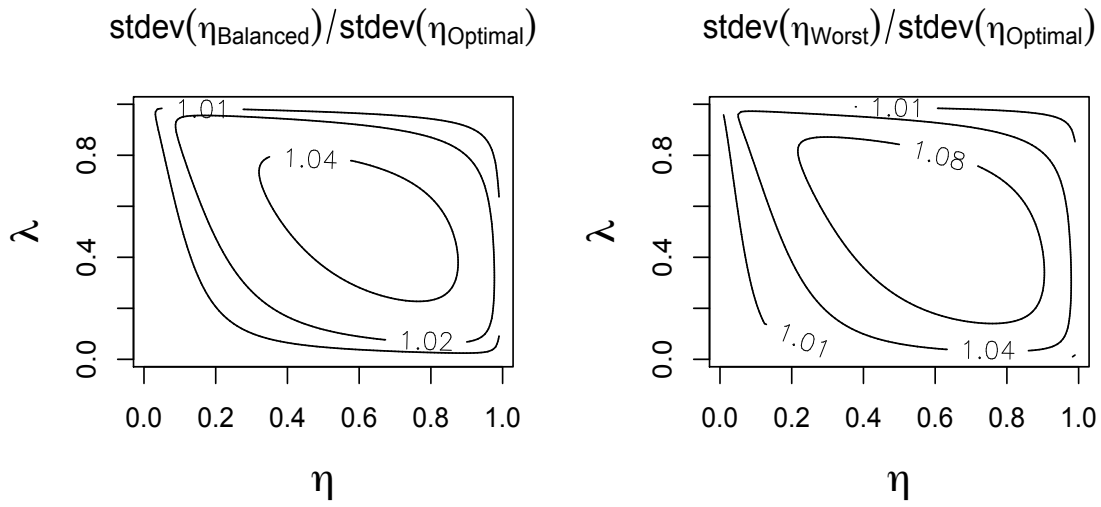

Figure 6.5: The ratio of asymptotic standard deviations for the MLEs of η from the balanced (left panel) and worst (right panel) selection methods compared to the optimal selection method. All three selection methods have the same design ($b = 18, k = 6, n = 2$ and $m = 3$).

Figure 6.5 shows similar results as Figure 6.4 in that the non-optimal selection methods are comparable to the optimal selection method. In practice, it is possible to adopt an approximately optimal selection method because we can estimate $(\underline{\mu} - \overline{\mu}\underline{1}_m)$ from the baseline. Based on the results seen for this particular example, the gains will be small (smaller than what is presented here) and it is probably not worth the additional complexity. In addition, here we have assumed that μ were known and in practice they would have to be estimated. We suggest using a balanced selection method. That is, select a roughly equal number of parts with large and small initial measurements for each operator because these parts will

typically have the most extreme measurements. Also, a balanced selection method will help detect if the measurement system has constant variance across the part values.

6.2.5 Leveraged Plan A Design

In this section, we present some general guidelines for choosing a type A leveraged plan (i.e. choosing values for b, k and n) when the total number of measurements is N=60, 90 or 150 and there are $m = 3$ operators. We based our guidelines on designs that have the smallest asymptotic standard deviation of η calculated using the Fisher information. To find the preferred plans with N and m fixed, we pick a point in the region $\eta \in (0.01, 0.99)$ and $\lambda \in (0, 1)$ and calculate the asymptotic standard deviations for all possible designs. Again we assume that the operator effects are equally spaced. That is, if $\sigma_o^2 = 1$ and $\overline{\mu} = 0$ then $\underline{\mu} = (-1.225, 0, 1.225)$. When k is not a multiple of $2m = 6$, the set S of parts selected for Stage 2 is not balanced across operators. To select the parts, we use a balanced selection method as suggested in the last paragraph of subsection 6.2.4.

In Table 6.3, we present a cross-section of the full results of the search (given in Appendix B). The optimal designs in Table 6.3 change slowly with η. Choosing an optimal design in practice is not possible because of the dependence on the unknown parameters. We might suggest a criterion that summarizes the performance of a design over all values of η and λ. However, any such criterion would be difficult to justify. In Table 6.3, the baseline size ($b \times m$) is around half the total number of measurements N and the number of repeated measurements n is two or three. We thus consider only two leveraged plans ($b = 11, k = 3, n = 3$) and ($b = 18, k = 6, n = 2$) because they match designs featured in Table 6.3 and have k equal to an integer multiple of the number of operators. This property ensures that an equal

Table 6.3: The LP designs (b,k,n) with the smallest asymptotic standard deviation of η given λ, η when $m = 3$.

η	N=60	$\lambda \in$	N=90	$\lambda \in$	N=150	$\lambda \in$
0.5	(8,6,2)	(0.05,0.08)	(14,8,2)	(0.01,0.14)	(24,13,2)	(0.01,0.05)
	(10,5,2)	(0.09,0.29)	(16,7,2)	(0.15,0.31)	(26,12,2)	(0.06,0.35)
	(12,4,2)	(0.30,0.87)	(18,6,2)	(0.32,0.67)	(28,11,2)	(0.36,0.62)
	(14,3,2)	(0.88,0.95)	(20,5,2)	(0.68,0.86)	(30,10,2)	(0.63,0.83)
			(22,4,2)	(0.87,0.95)	(32,9,2)	(0.84,0.86)
					(38,6,2)	(0.87,0.91)
0.75	(10,5,2)	(0.05,0.26)	(14,8,2)	(0.05, 0.05)	(26,12,2)	(0.05,0.23)
	(12,4,2)	(0.27,0.84)	(16,7,2)	(0.06, 0.31)	(28,11,2)	(0.24,0.56)
	(14,3,2)	(0.85,0.94)	(18,6,2)	(0.32, 0.65)	(30,10,2)	(0.57,0.8)
	(16,2,2)	(0.95,0.95)	(20,5,2)	(0.66, 0.85)	(32,9,2)	(0.81,0.85)
			(22,4,2)	(0.86, 0.95)	(35,5,3)	(0.86,0.88)
					(36,7,2)	(0.89,0.9)
					(38,6,2)	(0.91,0.94)
					(40,5,2)	(0.95,0.96)
0.9	(10,5,2)	(0.05,0.35)	(16,7,2)	(0.05,0.34)	(26,12,2)	(0.05,0.26)
	(12,4,2)	(0.36,0.84)	(18,6,2)	(0.35,0.67)	(28,11,2)	(0.27,0.57)
	(14,3,2)	(0.85,0.94)	(20,5,2)	(0.68,0.85)	(30,10,2)	(0.58,0.79)
	(16,1,4)	(0.95,0.95)	(22,4,2)	(0.86,0.95)	(32,9,2)	(0.80,0.86)
					(34,8,2)	(0.87,0.87)
					(36,7,2)	(0.88,0.92)
					(38,6,2)	(0.93,0.95)

number of parts measured by each operator in the baseline are represented in the 2^{nd} Stage. Figure 6.6 shows the asymptotic standard deviation of η for these two plans versus the optimal designs given in Table 6.3, over the parameter space. For either plan, there is at most a 15% increase in the standard deviation of η from using the suggested LP design as opposed to an optimal design which depends on the unknown parameters. The benefit of choosing the optimal over the recommended plan diminishes as we increase the total number of measurements (i.e. move from the left panel to the right panel in Figure 6.6).

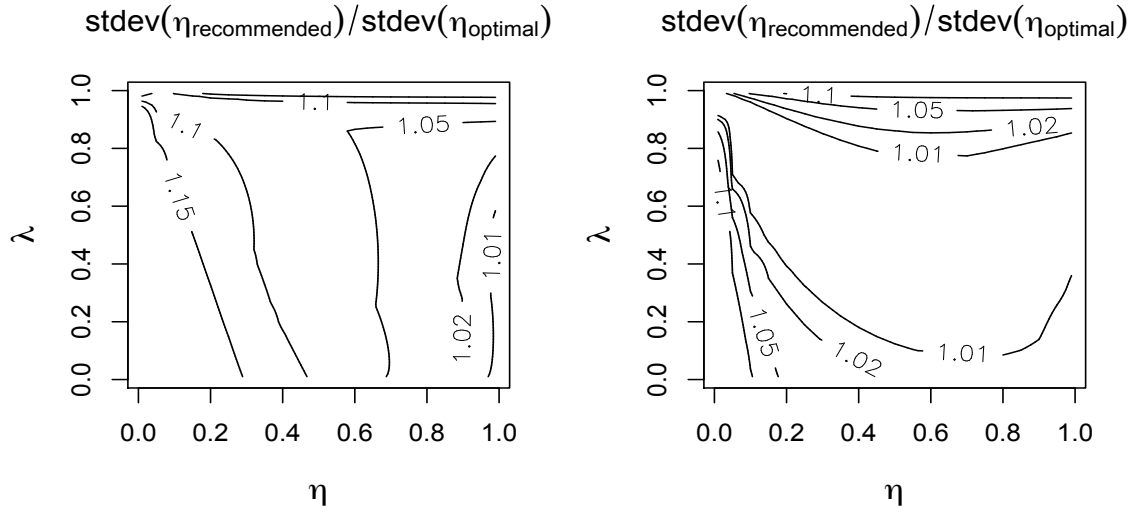

Figure 6.6: The standard deviation of the LP estimator from two designs $b = 11, k = 3, n = 3$ (left panel) and $b = 18, k = 6, n = 2$ (right panel) divided by the standard deviation of the optimal plan given $N = 60$ (left panel), 90 (right panel), η, λ and $m = 3$.

We also investigated how these two plans behave when the operator effects are not equally spaced. Because we select the same number of parts for each operator from the baseline, we see in (B.7) that the estimators from these plans do not depend on how the operator effects are distributed to make up σ_o.

To summarize, we have the following guidelines for selecting a leveraged plan when only a few operators make regular use of the measurement system and the

total number of measurements N available is roughly specified:

- involve all m operators
- in Stage 1, select bm parts at random from the process where bm is close to $N/2$. Each operator measures b parts.
- in Stage 2, select an equal number of extreme parts from each operator based on the initial measurements to give a total of k parts. Since each of these parts will be measured two or three times by each operator, select k so that $2mk$ or $3mk$ is approximately $N/2$.

6.2.6 Leveraged Plan A versus Standard Plan

To demonstrate the value of the leverage plan we resort to simulation. We compare the following two plans when there are $m = 3$ operators and the total number of measurements is $N = 60$ or 90:

1. $N = 60$, with $m = 3$
    - SP with $k = 10$ and $n = 2$ (a commonly used plan in practice)
    - LP with $b = 11$, $k = 3$ and $n = 3$ (as recommended in Section 6.2.5)
2. $N = 90$, with $m = 3$
    - SP with $k = 10$ and $n = 3$ (a commonly used plan in practice)
    - LP with $b = 18$, $k = 6$ and $n = 2$ (as recommended in Section 6.2.5)

We use MLE for both the leveraged plan and standard plan. We quantify the difference between the plans using the ratio of square root of the mean squared error (MSE) of the estimators for η and λ from the LP and SP. To calculate the

ratio we simulated 1,000 repeats at a grid of values of η and λ spread over the region $(0,1) \times (0,1)$. Then to present the data in Figure 6.9 we applied a smoother across the parameter space. We used fitted values from a non-linear regression model because of simulation error.

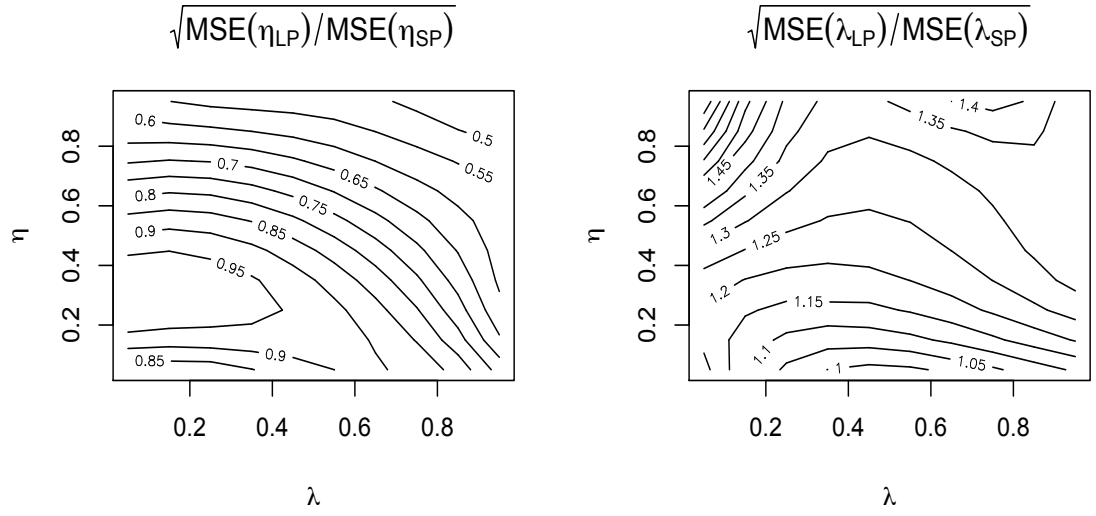

Figure 6.7: The MSE ratios for estimators of η (left panel) and λ (right panel) from the leveraged plan ($b = 11$, $k = 3$, $n = 3$) and standard plan ($k = 10$, $n = 2$) across η and λ when the total sample size is 60.

Figure 6.9 shows that the LP has a lower MSE for η than does SP. Conversely, the SP has a lower MSE for λ than the LP. We think estimation of η is more important than λ because the λ value does not matter if we have a good measurement system. A good measurement system [as recommended by Automotive Industry Action Group, 2002] should have $\eta \geq .91$. In this range the SP is 1.3 to 1.4 times more efficient in estimating λ while the LP is 1.6 to 2 times more efficient in estimating η.

Figure 6.10 shows the results when the total sample size is $N = 90$. For this case we get similar qualitative results but now estimation of η with the LP is 2 times more efficient than the SP for $\eta \geq 0.91$.

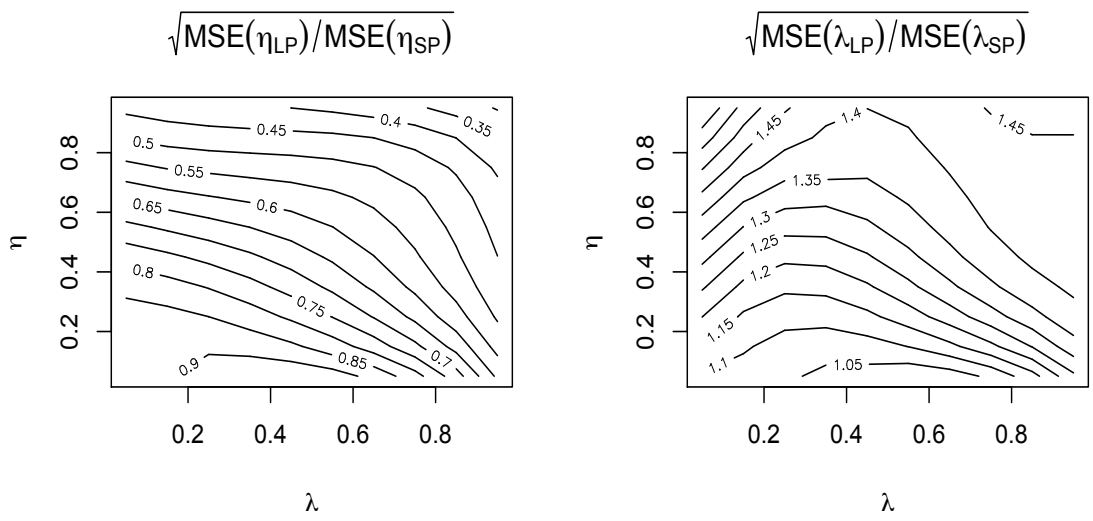

Figure 6.8: The MSE ratios for estimators of η (left panel) and λ (right panel) from the leveraged plan ($b = 18$, $k = 6$ and $n = 2$) and standard plan ($k = 10$ and $n = 3$) across η and λ when the total sample size is 90.

6.3 Leveraged Plan B

We consider a second version of the leveraged plan where in both Stages 1 and 2 any selected part is measured by all operators.

6.3.1 Plan B

This second LP is conducted in two stages:

Stage 1: Sample b parts at random from the process and have the m operators measure each part once for a total $(b \times m)$ measurements. In this stage, every part is measured m times. We denote the observed values from part i as a vector, $\underline{y}_{i0} = (y_{i10}, \ldots, y_{im0})$. Individually, $\{y_{ij0}\}$ is the measured value from operator j on part i.

Stage 2: From the baseline sample, select k parts using the observed measured

values. In particular, to improve the estimation for η, sample k parts such that both

- the part average of initial measurements are extreme relative to the overall average, defined as $|\overline{y}_{i.0} - \overline{y}_{..0}|$ where $\overline{y}_{i.0} = \frac{1}{m}\sum_{j=1}^{m} y_{ij0}$ and $\overline{y}_{..0} = \frac{1}{bm}\sum_{i=1}^{b}\sum_{j=1}^{m} y_{ij0}$, and
- the average of the initial measurements of the selected parts is close to the baseline average.

These k parts are then used in a standard plan. That is each operator measures each of the k parts n times to give the additional data $\{y_{ijlh},\ \ (i,j) \in S,\ \ l = 1, \ldots, m \ \text{ and } \ h = 1, \ldots, n\}$ where the S is a subset of $\{(1,1), \ldots, (b,m)\}$ with k elements. The total number of measurements in the leveraged plan is $N = m(b + nk)$.

For example, for a type B leveraged plan with $b = 10$, $m = 3$, $k = 2$ and $n = 2$, we sample 10 parts at random from the process and then have each of the three operators measured every part once. Then, we pick the parts with the minimum and maximum average initial measurements. These $k = 2$ parts are then measured by each operator twice more. This plan has a total of $3(10 + 2 \times 2) = 42$ measurements. Note, that the standard plan is a subset of a leveraged plan where $k = b$. If, in this leveraged plan, we repeatedly measure all the parts from the baseline we have the standard plan.

6.3.2 Estimation for Leveraged Plan B

We form the likelihood using the parameterization $\left\{\underline{\mu}, \sigma^2_{pg} = \sigma^2_p + \sigma^2_g, \rho = \frac{\sigma^2_p}{\sigma^2_p + \sigma^2_g}\right\}$ because it simplifies the likelihood and we apply the likelihood (4.4) from Section 4.2.1. Note, $\underline{\mu}$ is defined in (1.9) and represents the vector of operator means.

To obtain the likelihood for the LP, we again decompose the likelihood into two pieces by conditioning on the baseline measurements. The log-likelihood for the LP is

$$l_{L4}(\mu, \sigma_t^2, \rho) = l_{b4}(\underline{\mu}, \sigma_{pg}^2) + l_{r4}\left(\underline{\mu}, \sigma_{pg}^2, \rho \;\middle|\; y_{ij0}\right). \tag{6.22}$$

To get the complete likelihood, we begin with the likelihood for Stage 1. In the baseline, each of the m operators measure b parts for total of $(b \times m)$ measurements. This is the same as the standard plan. Thus we can use the standard plan likelihood given in (2.5) but instead we have $k = b$ and $n = m$. The baseline likelihood is

$$\begin{aligned} l_{b4}(\underline{\mu}, \sigma_{pg}^2) &= -\frac{1}{2} b \Big[m \log(\sigma_{pg}^2) + (m-1)\log(1-\rho) + \log(1+\rho(m-1)) \Big] \\ &\quad -\frac{1}{2} \frac{[1+\rho(m-1)]\, SSW' - \rho m^2 SSA'}{\sigma_{pg}^2 (1-\rho)\,[1+(m-1)\rho]} \end{aligned} \tag{6.23}$$

where $SSA' = \sum_{i=1}^{b} \overline{z}_{i.0}^2$, $SSW' = \sum_{i=1}^{b} \sum_{j=1}^{m} z_{ij0}^2$, $\overline{z}_{i.0} = \sum_{j=1}^{m} z_{ij0}/m$ and $z_{ij0} = y_{ij0} - \mu_j$.

To obtain the likelihood for Stage 2, we first consider a single part i (selected to be repeatedly measured). Note that in LP B each part was already measured once by each operator in Stage 1. The joint distribution of these initial measurements $\underline{Y}_{i0} = (Y_{i10}, \ldots, Y_{im0})$ and the n repeated measurements $\{\underline{Y}_{10}, \ldots, \underline{Y}_{m0}\}$ is

$$\begin{pmatrix} \underline{Y}_{i0} \\ \underline{Y}_{i1} \\ \vdots \\ \underline{Y}_{in} \end{pmatrix} \backsim N\left(\begin{bmatrix} \underline{\mu} \\ \underline{\mu} \otimes \mathbf{1}_n \end{bmatrix}, \quad \Sigma_{m(n+1)} = \begin{matrix} \Sigma_m & \rho\sigma_t^2 \mathbf{J}_{m,mn} \\ \rho\sigma_t^2 \mathbf{J}_{mn,m} & \Sigma_{mn} \end{matrix} \right), \tag{6.24}$$

where Σ_n is defined in (4.1), $\rho = \sigma_p^2/(\sigma_p^2 + \sigma_m^2)$, $\underline{Y}_{il} = (Y_{i1l}, Y_{i2l}, \ldots, Y_{iml})^t$ represents a vector of the l^{th} measurement by the m operators on part i for $l = 1, \ldots, n$ and Y_{ijl} is a random variable representing the k^{th} measurement by operator j on part i.

Since, the random variables $\{\underline{Y}_{i0}, \underline{Y}_{i1}, \ldots, \underline{Y}_{in}\}$ have a multivariate normal distribution, if we condition on the initial measurements $\{\underline{Y}_{i0} = \underline{y}_{i0}\}$, the distribution of the repeated measurements $\{\underline{Y}_{i1}, \ldots, \underline{Y}_{in}\}$ is

$$\left(\begin{array}{c} \underline{Y}_{i1} \\ \vdots \\ \underline{Y}_{in} \end{array} \middle| \underline{Y}_{i0} = \underline{y}_{i0} \right) \backsim N \left(\left[\underline{\mu} + \frac{\rho m}{1-\rho+\rho m} \sum_{j=1}^{m} \left(\underline{y}_{ij0} - \underline{\mu} \right) / m \right] \otimes \underline{1}_{mn}, \quad \Sigma_v \right).$$

This distribution has a similar form as seen in (4.2). Specifically, we notice that the initial measurements affect the mean but not the variance. Additionally, the mean is only affected by the average of the initial measurements.

We derive the covariance matrix Σ_v, using the properties of matrices

$$\begin{aligned}
\Sigma_v &= \Sigma_{mn} - \left(\rho\sigma_t^2 \mathbf{J}_{mn,m}\right) \Sigma_m^{-1} \left(\rho\sigma_t^2 \mathbf{J}_{m,mn}\right) \\
&= \Sigma_{mn} - \rho^2\sigma_t^4 \mathbf{J}_{mn,m} \left(\frac{1}{\sigma_t^2} \left\{ \frac{1}{1-\rho} \mathbf{I}_m - \frac{\rho}{(1-\rho)(1-\rho+\rho m)} \mathbf{J}_m \right\} \right) \mathbf{J}_{m,mn} \\
&= \Sigma_{mn} - \rho^2\sigma_t^2 \left\{ \frac{1}{1-\rho} \mathbf{J}_{mn,m}\mathbf{I}_m\mathbf{J}_{m,mn} - \frac{\rho}{(1-\rho)(1-\rho+\rho m)} \mathbf{J}_{mn,m}\mathbf{J}_m\mathbf{J}_{m,mn} \right\} \\
&= \Sigma_{mn} - \rho^2\sigma_t^2 \left\{ \frac{1}{1-\rho} m\mathbf{J}_{mn,mn} - \frac{\rho}{(1-\rho)(1-\rho+\rho m)} m^2\mathbf{J}_{mn} \right\} \\
&= \Sigma_{mn} - \rho^2\sigma_t^2 \left\{ \frac{m}{(1-\rho+\rho m)} \right\} \mathbf{J}_{mn} \\
&= \sigma_t^2 \left[(1-\rho)\mathbf{I}_{mm} + \rho\mathbf{J}_{mn} \right] - \rho^2\sigma_t^2 \left\{ \frac{m}{(1-\rho+\rho m)} \right\} \mathbf{J}_{mn} \\
&= \sigma_t^2 \left((1-\rho)\mathbf{I}_{mm} + \frac{\rho(1-\rho)}{(1-\rho+\rho m)} \mathbf{J}_{mn} \right) \\
&= \sigma_t^2 (1-\rho) \left(\mathbf{I}_{mm} + \frac{\rho}{(1-\rho+\rho m)} \mathbf{J}_{mn} \right).
\end{aligned}$$

Σ_v has a special form which allows us to obtain the following well known properties [Dillon and Goldstein, 1984]:

$$\Sigma_v^{-1} = \frac{1}{\sigma_t^2(1-\rho)} \left(\mathbf{I}_{mn} - \frac{\rho}{1-\rho+\rho m+\rho nm} \mathbf{J}_{mn} \right)$$

$$|\Sigma_v| = \left(\sigma_t^2(1-\rho)\right)^{mn}\left(\frac{1-\rho+\rho m+\rho nm}{1-\rho+\rho m}\right)$$

Using these properties of Σ_v, we can write the conditional likelihood (conditional on the vector $\underline{y}_{i0}$) for the repeated measurements on a part. The measurements on one part are independent of the measurements from another part. Thus the conditional likelihood for the k parts is the product of the individual likelihoods. This conditional likelihood has the same form as (4.4) but is slighly different because here we are conditioning on more than one measurement. The conditional log-likelihood is

$$\begin{aligned} l_{r4}(\,\underline{\mu}, \sigma_{pg}^2, \rho \;\Big|\; \underline{y}_{i0}\,) = \\ -\frac{k}{2}\Big\{ mn\log\left(\sigma_{pg}^2(1-\rho)\right) - \log\left(1-\rho+\rho m\right) + \log\left(1-\rho+\rho m+\rho nm\right)\Big\} \\ -\frac{1}{2}\frac{1}{\sigma_{pg}^2(1-\rho)(1-\rho+\rho m+\rho nm)}\Big\{ (1-\rho+\rho m+\rho nm)SSV \\ +nm(1-\rho+\rho m)\sum_{i\in S}\left(\overline{v}_{i..} - \frac{\rho m}{1-\rho+\rho m}\overline{v}_{i.0}\right)^2\Big\} \end{aligned} \tag{6.25}$$

where $v_{ij0} = (y_{ij0} - \mu_j)$ is the baseline measurement for the i^{th} part measured by operator j, $SSV = \sum_{i\in S}\sum_{j=1}^{m}\sum_{k=1}^{n}(v_{ijk} - \overline{v}_{i..})^2$, $\overline{v}_{i..} = \frac{1}{mn}\sum_{j=1}^{m}\sum_{k=1}^{n} v_{ijk}$, $v_{ijk} = (y_{ijk} - \mu_j)$ and $\overline{v}_{i.0} = \sum_{j=1}^{m}(y_{ij0} - \mu_j)/m$. Surprisingly, even though we condition on a vector of measurements the conditional likelihood depends only on the average of the initial measurements, denoted by $\overline{v}_{i.0}$.

To find the MLEs of $\underline{\mu}, \sigma_{pg}^2$ and ρ, we maximize (6.22). Solutions can be found numerically. Then to get the MLE's for η and λ, we apply the appropriate transformations given in (6.15).

6.3.2.1 Fisher information

We obtain the Fisher information matrix by summing two pieces. The first piece comes from the baseline likelihood,

$$J_{b4}\left(\underline{\mu}, \sigma_{pg}^2, \rho\right) = \begin{bmatrix} \frac{b}{\sigma_{pg}^2(1-\rho)(1+\rho(m-1))}\left[(1-\rho+\rho m)\mathbf{I}_m - \rho\mathbf{J}_m\right] & \underline{0}_m & \underline{0}_m \\ & \frac{bm}{2\sigma_{pg}^2} & \frac{-bm(m-1)\rho}{2\sigma_{pg}^2(1-\rho)(1+\rho(m-1))} \\ & & \frac{b(m-1)m(\rho^2(m-1)+1)}{2(1-\rho)^2(1+\rho(m-1))^2} \end{bmatrix} \quad (6.26)$$

and the second piece is from the Stage 2 likelihood. The Stage 2 Fisher information is shown element by element,

$$\begin{aligned}
E\left[-\frac{\partial^2 l_{r4}}{\partial \mu_j^2}\right] &= \frac{kn\left[(1-\rho)^2 + (1-\rho+m\rho\left[n+1\right])(1-\rho+m\rho)(m-1)\right]}{\sigma_{pg}^2(1-\rho)(1-\rho+m\rho\left[n+1\right])(1-\rho+m\rho)m} \\
E\left[-\frac{\partial^2 l_{r4}}{\partial \mu_j \partial \mu_l}\right] &= \frac{kn\left[(1-\rho)^2 - (1-\rho+m\rho\left[n+1\right])(1-\rho+m\rho)\right]}{\sigma_{pg}^2(1-\rho)(1-\rho+m\rho\left[n+1\right])(1-\rho+m\rho)m} \quad (6.27) \\
E\left[-\frac{\partial^2 l_{r4}}{\partial \mu_j \partial \sigma_{pg}^2}\right] &= 0 \quad (6.28) \\
E\left[-\frac{\partial^2 l_{r4}}{\partial \mu_j \partial \rho}\right] &= -\frac{nm\sum_{i\in S}\overline{v}_{i.0}}{\sigma_{pg}^2(1-\rho+\rho m)^2(1-\rho+\rho m+mn\rho)} \quad (6.29) \\
E\left[-\frac{\partial^2 l_{r4}}{\partial (\sigma_{pg}^2)^2}\right] &= \frac{1}{2}\frac{kmn}{\sigma_{pg}^2} \quad (6.30) \\
E\left[-\frac{\partial^2 l_{r4}}{\partial \sigma_{pg}^2 \partial \rho}\right] &= -\frac{1}{2}\frac{mnk\rho(\rho(n+1)m^2 + (1-\rho)(n+2)m - 1 + \rho)}{(1-\rho)(1-\rho+\rho m+\rho nm)(1-\rho+\rho m)\sigma_{pg}^2} \quad (6.31) \\
E\left[-\frac{\partial^2 l_{r4}}{\partial \rho^2}\right] &= \left\{\frac{2nm^3(1-\rho)(1-\rho+\rho m+\rho nm)}{k(1-\rho+\rho m)}\sum_{i\in S}\overline{v}_{i.0}^2/\sigma_{pg}^2\right. \\
&\quad -(2\rho-1)(mn-1) \\
&\quad +\rho^2 m(4(mn+m-1)+n(mn-1)) \\
&\quad +\rho^3 2(m-1)(mn+2m-1)(mn+m-1) \\
&\quad \left.+\rho^4(m-1)^2(mn+m-1)^2\right\} \times
\end{aligned}$$

$$\frac{kmn}{2(1-\rho)^2(1-\rho+\rho m)^2(1-\rho+\rho m+\rho nm)^2} \tag{6.32}$$

The first thing to notice about this Fisher information is that (6.32) increases when we select parts associated with initial measurements that are far from the overall mean $\overline{\mu}$. In contrast to the LP A, the Fisher information is the same for all the means $\underline{\mu}$ because this plan is symmetric with respect to the operators.

To get the complete Fisher information we add the baseline and the 2^{nd} Stage Fisher information matrices.

$$J_{L4}\left(\underline{\mu}, \sigma^2_{pg}, \rho\right) = J_{b4}\left(\underline{\mu}, \sigma^2_{pg}, \rho\right) + J_{r4}(\underline{\mu}, \sigma^2_{pg}, \rho) \tag{6.33}$$

To obtain the Fisher information for η, λ we use the method given in (2.24).

6.3.3 Leveraged Plan B Design

In this section, we present some general guidelines for choosing a leveraged plan (i.e. choosing values for b, k and n) when the total number of measurements is N=60, 90 or 120 and there are $m = 3$ operators. We based our guidelines on designs that have the smallest asymptotic standard deviation of η calculated using the Fisher information. To find the preferred plans with N and m fixed, we pick a point in the region $\eta \in (0.01, 0.99)$ and $\lambda \in (0, 1)$ and calculate the asymptotic standard deviations for all possible designs. Again, as in section 6.2.5, we assume that the operator effects are equally spaced. To select the parts, we take a roughly balanced number of large and small parts.

In Table 6.4, we present a cross-section of the full results of the search (given in Appendix B). The optimal designs in Table 6.4 change slowly with η. Choosing an optimal design is not possible because of the dependence on the unknown

Table 6.4: The LP B designs (b,k,n) with the smallest asymptotic standard deviation of η given λ, η when $m = 3$.

η	N=60	$\lambda \in$	N=90	$\lambda \in$	N=150	$\lambda \in$
0.25	(14,3,2)	(0.01,0.17)	(22,4,2)	(0.01,0.41)	(34,8,2)	(0.01,0.05)
	(16,2,2)	(0.18,0.64)	(24,3,2)	(0.42,0.48)	(36,7,2)	(0.06,0.16)
	(18,1,2)	(0.65,0.99)	(26,2,2)	(0.49,0.71)	(38,6,2)	(0.17,0.41)
			(28,1,2)	(0.72,0.99)	(40,5,2)	(0.42,0.45)
					(42,4,2)	(0.46,0.62)
					(44,3,2)	(0.63,0.64)
					(46,2,2)	(0.65,0.76)
					(48,1,2)	(0.77,0.99)
0.5	(16,2,2)	(0.01,0.19)	(26,2,2)	(0.01, 0.33)	(42,4,2)	(0.05,0.17)
	(18,1,2)	(0.20,0.99)	(28,1,2)	(0.34, 0.99)	(44,3,2)	(0.18,0.21)
					(46,2,2)	(0.22,0.44)
					(48,1,2)	(0.45,0.99)
0.75 to 0.99	(18,1,2)	(0.05,0.35)	(28,1,2)	(0.01,0.99)	(48,1,2)	(0.01,0.99)

parameters. In Table 6.4, n, the number of repeated measurements is 2 and k is small.

6.3.4 Leveraged Plan B versus Standard Plan

To demonstrate the value of the leveraged plan, we resort to simulation. We compare the following two plans when there are $m = 3$ operators and the total sample size is $N = 60$ and 90 measurements:

1. $N = 60$, with $m = 3$
   - SP with $k = 10$ and $n = 2$ (a commonly used plan in practice)
   - LP with $b = 16$, $k = 2$ and $n = 2$ (as recommended in Section 6.3.3)
2. $N = 90$, with $m = 3$
   - SP with $k = 10$ and $n = 3$ (a commonly used plan in practice)

- LP with $b = 26$, $k = 2$ and $n = 2$ (as recommended in Section 6.3.3)

We use MLE for both the leveraged plan and standard plan. We quantify the difference between the plans using the ratio of the square root of the mean squared error (MSE) of the estimators for η and λ from the LP and SP. To calculate the ratio we simulated 1,000 repeats at a grid of values of η and λ spread over the region $(0, 1) \times (0, 1)$. Then to present the data in Figure 6.9 we applied a smoother across the parameter space.

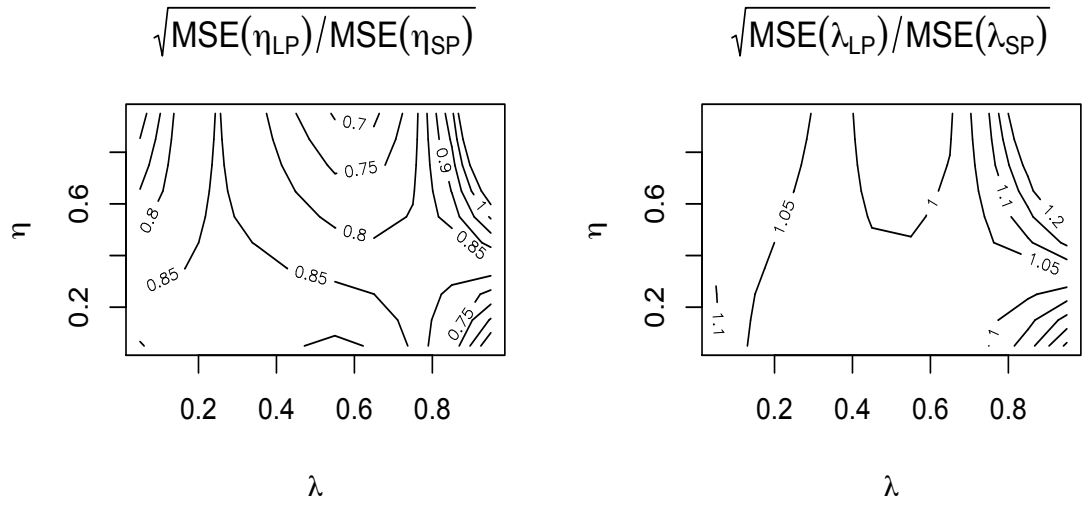

Figure 6.9: The mean squared error (MSE) ratios for estimating η (left panel) and λ (right panel) from the leveraged plan ($b = 16, k = 2, n = 2$) and standard plan ($k = 10, n = 2$) across η and λ when the total sample size is 60.

Figure 6.9 shows that the LP B has a lower MSE for η than does SP except when λ is near 1. Unlike the LP A, LP B is comparable to the SP when it comes to estimating λ, although the SP still has a lower MSE for λ. Figure 6.10 shows similar results when the total sample size is $N = 90$.

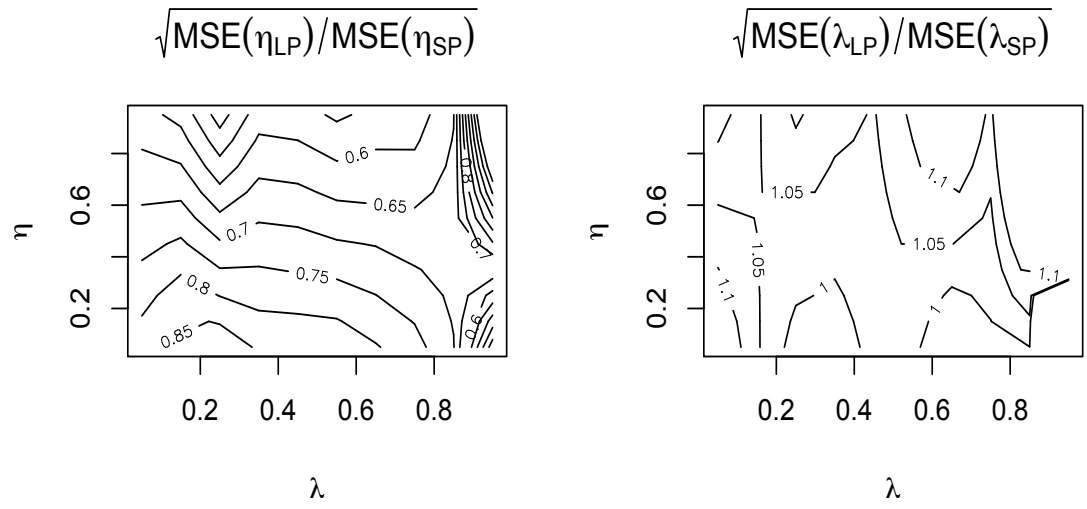

Figure 6.10: The mean squared error (MSE) ratios for estimating η (left panel) and λ (right panel) from the leveraged plan (b = 26, k = 2 and n = 2) and standard plan (k = 10 and n = 3) across η and λ when the total sample size is 90.

6.4 Leveraged Plan A versus B

To compare the two types of leveraged plans, we use simulation. We compare the following two plans when there are $m = 3$ operators and the total sample size is $N = 60$ and 90 measurements:

1. $N = 60$, with $m = 3$

   LP A with $b = 11$, $k = 3$ and $n = 3$

   LP B with $b = 16$, $k = 2$ and $n = 2$

2. $N = 90$, with $m = 3$

   LP A with $b = 18$, $k = 6$ and $n = 2$

   LP B with $b = 26$, $k = 2$ and $n = 2$

We use MLE for these two leveraged plans. Figures 6.11 and 6.12 suggest that the LP A is more efficient than LP B for estimating η when the true value of

$\eta \geq 0.4$. Conversely, as shown in the right panels of the two Figures LP B is more efficient than LP A for estimating λ except when λ is near 1. We recommended LP A because of its efficiency when estimating the primary parameter η.

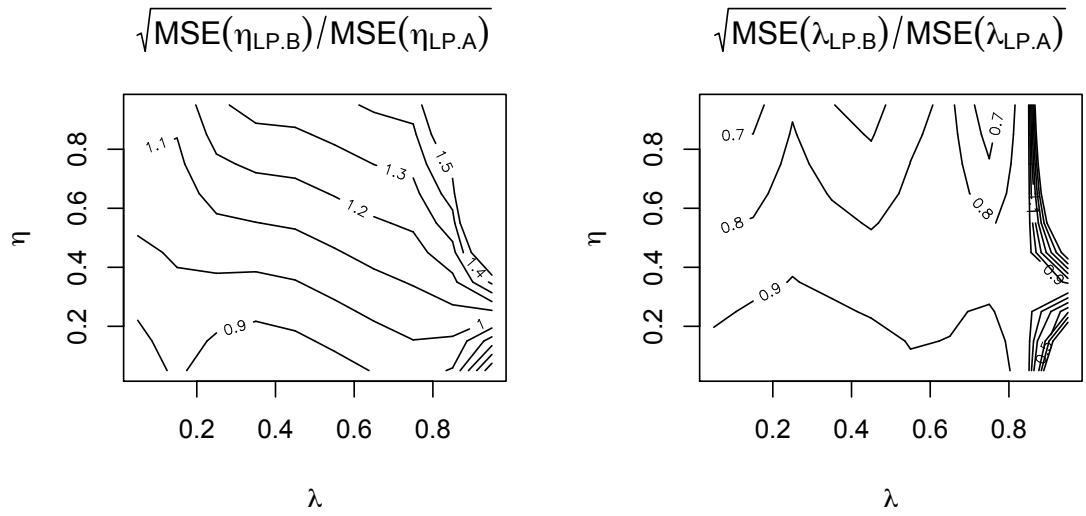

Figure 6.11: The mean squared error (MSE) ratios for estimating η (left panel) and λ (right panel) from the leveraged plan A ($b = 11, k = 3, n = 3$) and leveraged plan B ($b = 16, k = 2n = 2$) across η and λ when the total sample size is 60.

6.5 Discussion and Conclusions

In this chapter, we introduced two new leveraged plans for assessing measurement systems with operators. We compared these two plans to the standard measurement system assessment under specific conditions. We treat η, the part variation divided by the total variation, as the primary parameter of interest while λ, the proportion of the measurement variation due to operator bias, is of secondary interest, since the main purpose of the assessment is to validate the measurement system as a whole. The comparisons were based on the MSE of the ML estimators from each plan.

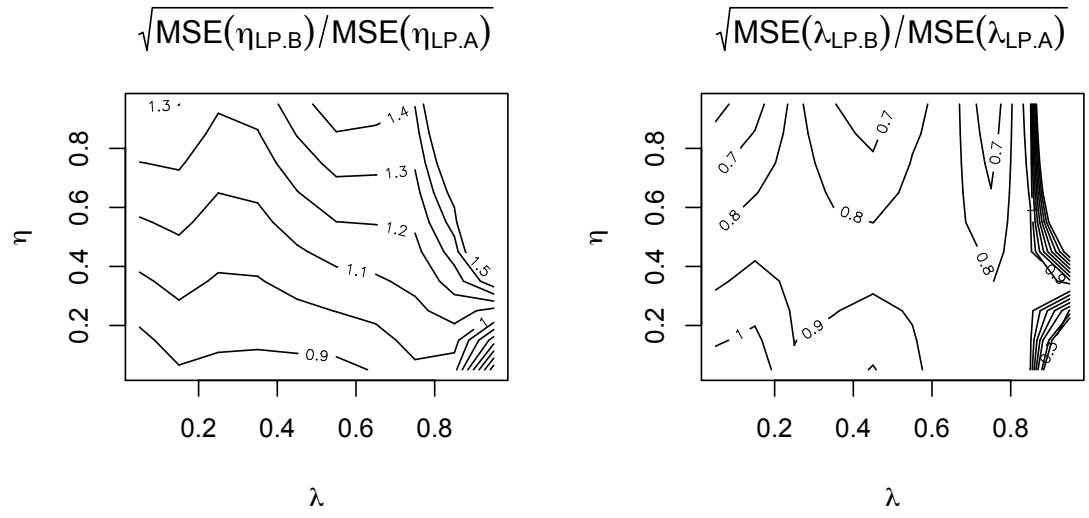

Figure 6.12: The mean squared error (MSE) ratios for estimating η (left panel) and λ (right panel) from the leveraged plan A ($b = 18, k = 6, n = 2$) and leveraged plan B ($b = 26, k = 2, n = 2$) across η and λ when the total sample size is 90.

When comparing the MSEs for estimators of η we found that the LP A was most efficient followed by the LP B and then the SP. Conversely in the comparison based on estimating λ, SP and LP B were more efficient than LP A. Relative to the SP, the LP B was more efficient in estimating η and was comparable when estimating λ whereas LP A was more efficient in estimating η and less efficient when estimating λ. We recommend LP A because of its efficiency when estimating the primary parameter η.

We also provided a methodology for practicioners to aid in the design of an LP A.

Chapter 7

Assessment of an Inspection System with Production Data

7.1 Context

Many manufacturers require parts to pass an inspection system before being shipped. The purpose of the inspection is to prevent customers from receiving poor quality parts. The ideal system rejects each part with a true value outside of inspection limits but due to measurement error, the actual system rejects parts with observed or measured characteristics outside of these limits. Thus, an inspection system will reject some good parts and accept some bad parts. Accepting and rejecting the wrong parts can be costly, making it essential to verify or quantify the performance of the inspection system. Measurement variability explains why inspection limits are often tighter than the specification limits.

In general, an inspection system has two parts, a measurement system and an inspection protocol. The measurement system is the method or device used to measure the characteristic of interest. The inspection protocol is the set of decision

rules for the inspection system. Figure 7.1 gives an example of a commonly used protocol but there are many possibilities.

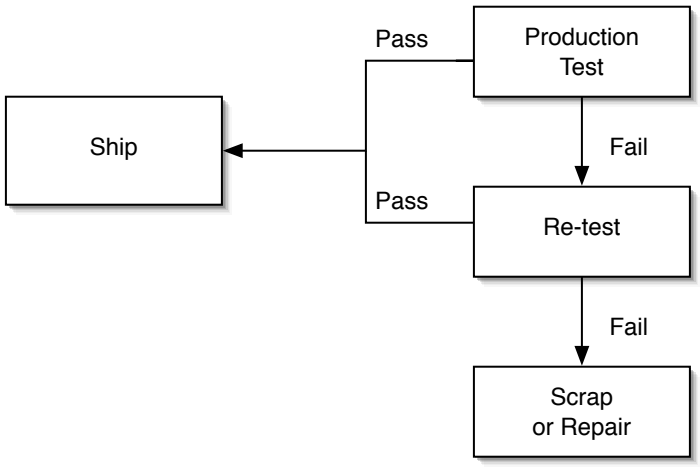

Figure 7.1: Typical Inspection Protocol

The performance of any inspection system is highly reliant on the measurement system used to measure the characteristic of interest. To assess a continuous measurement system, a Gauge Repeatability and Reproducibility study is typically performed. See Automotive Industry Action Group [2002] and Burdick et al. [2003]. Since these studies are conducted off-line, they can be costly and may not reflect the properties of the measurement system during production.

In some industries, such as integrated circuit or electronic device manufacturing it is common to use the protocol given in Figure 7.1. The inspection protocol starts with each part being measured. This first measurement is called the production test. The test result can be either pass or fail based on whether the measured value lies within the inspection limits or not. A pass allows the part to be shipped whereas a failure means the part is retested. Commonly, this retest is carried out immediately and if the part passes the second test it is shipped. Otherwise it is sent to be repaired or scrapped.

Table 7.1: 100 Production Test Observations or 1^{st} Measurements

103.6	100.2	107.6	97.4	92.4	96.1	97.3	102.1	95.2	101.6
96.8	105.8	100.9	101.6	105.5	107.6	112.9	104.2	104.3	91.9
105.5	96.0	92.9	101.1	92.6	94.9	97.7	98.8	105.0	104.2
105.3	104.4	99.5	103.1	101.5	93.8	101.6	99.4	101.2	98.9
100.6	105.9	103.9	98.3	99.5	98.0	98.1	97.3	100.9	93.9
96.5	97.8	98.8	100.3	99.1	93.6	107.1	85.7	107.2	101.5
100.1	97.9	107.8	99.8	104.0	99.3	96.8	95.8	103.1	100.4
112.2	97.8	95.3	97.5	101.5	99.1	107.9	111.5	89.5	91.9
93.8	101.6	99.2	98.1	99.8	103.9	101.2	103.1	102.4	93.3
95.6	96.9	97.3	94.5	104.1	98.6	104.4	98.3	105.8	100.6

Using the inspection protocol shown in Figure 7.1, some parts are measured twice. As a result it is possible to carry out an assessment of the measurement system using data from the inspection system alone and avoid off-line studies such as a standard Gauge R&R. The data from the inspection system has a special form because a part is measured a second time if and only if the first measurement falls outside the "pass" region, denoted B = (LIL, UIL) where (LIL) and (UIL) are the lower and upper inspection limits, respectively. The possible outcomes for any part are (PASS), (FAIL, PASS) and (FAIL, FAIL).

Suppose, for example, we have the results of an inspection system with limits (95, 110) for 100 parts. Of these, 17 have second measurements. The data are shown in two tables. Table 7.1 gives the production measurements and Table 7.2 gives the repeated measurements. The 1^{st} measurements or production data average and standard deviation are 100.1 and 4.86, respectively.

Generally, any existing process operates well so we assume that the majority of the observed measurements are within the inspection limits. Thus, there are typically a large number of first measurements and a relatively small number of second measurements.

We use the following notation. The production data, the first measurement from each of n_1 parts is denoted by $\{y_{11}, y_{21}, \ldots, y_{n_1 1}\}$. For the retest data, we use

Table 7.2: 17 2^{nd} Measurements

Part #	1^{st}	2^{nd}	Part #	1^{st}	2^{nd}
5	92.4	91.3	71	112.2	111.8
17	112.9	111.1	78	111.5	110.8
20	91.9	92.2	79	89.5	88.8
23	92.9	93.3	80	91.9	91.1
25	92.6	94.1	81	93.8	95.4
26	94.9	94.2	90	93.3	90.8
36	93.8	92.4	94	94.5	93.6
50	93.9	92.9			
56	93.6	92.2			
58	85.7	84.6			

S, a subset of $\{1, 2, \ldots, n_1\}$, to indicate all the parts that have failed the production test so that $y_{i1} \notin B$. Suppose there are n_2 such parts. The retest data, the second measurements, are denoted by $\{y_{i2}, i \in S\}$.

To model an inspection system, we follow Burdick et al. [2003] and Doganaksoy [2000] by assuming that a normal random effects model (7.1) describes the observed characteristics. The model is

$$Y_{ij} = P_i + E_{ij} \tag{7.1}$$

where P_i is a random variable representing the possible values for the true dimension of part i $(i = 1, \ldots, n)$ and E_{ij} is a random variable representing the error on each measurement $(j = 1, 2)$ for part i. We assume that the part effects P_i are independent and identically distributed normal random variables with mean μ and variance σ_p^2, the measurement errors E are independent and identically distributed normal random variables with mean zero and variance σ_m^2, and P and E are mutually independent. The variance of Y_{i1}, called the total variation, is $\sigma_t^2 = \sigma_p^2 + \sigma_m^2$. By adopting model (7.1), we assume that μ, σ_p and σ_m are constant over the time needed to conduct the investigation and that σ_m is constant across true part dimensions.

We also assume the measurement system has no material operator effects. This

contrasts with what is typically assumed in the literature [See Burdick et al., 2003], but examples with no or little operator error are common in practice. For instance, in one example, piston diameters were inspected by an inline gauge with automated part handling. Using manufacturing jargon, with no operator effects, σ_m captures measurement repeatability but not reproducibility.

Burdick et al. [2003] describe a variety of metrics used to quantify measurement system quality or reliability. The metric we use for this chapter is again the intraclass correlation coefficient ρ.

This chapter provides justification for using the ANOVA estimator when assessing the measurement system data from regular production. Note our purpose is not to assess or try to optimize the inspection protocol. We assume the inspection protocol is described by Figure 7.1.

The ANOVA estimator is based on the pooled variance of the measured values for each part that has two measurements. We derive the bias and variance of the estimator using properties of the truncated normal distribution. We compare the root mean squared errors of the ANOVA and MLE. We show that the ANOVA estimator has a relatively small bias and high efficiency when compared the maximum likelihood estimator for most common values of ρ. Finally, we consider some other applications for this assessment method.

7.2 Inspection System Analysis

Three methods of analysis are presented and compared: analysis of variance (ANOVA), regression and maximum likelihood (ML). Although interest lies in estimating ρ, two other parameters σ_t^2 and μ are unknown and need to be estimated. To estimate these additional parameters, the ANOVA procedure uses the first measurements, i.e. the production data, only. In contrast, the ML procedure uses all the data.

7.2.1 ANOVA

A natural estimate of the measurement variation σ_m^2 is the average within part variance from those parts with two measurements. We estimate σ_t^2 by the sample variance of all the first measurements. The ANOVA estimate, denoted as $\widehat{\rho}_a$ is

$$\widehat{\rho}_a = 1 - \frac{s_m^2}{s_1^2} \tag{7.2}$$

where $s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} \left(y_{i1} - \overline{y}_{.1}\right)^2$ is the production data variance, $s_m^2 = \sum_{i \in S} \sum_{j=1}^{2} (y_{ij} - \overline{y}_{i.})^2 / n_2$ is the average variation within parts with two measurements, $\overline{y}_{i.} = (y_{i1} + y_{i2})/2$ is the average for any part i with two measurements and $\overline{y}_{.1} = \sum_{i}^{n_1} y_{i1}/n_1$ is the production data average. In the corresponding estimator $\widetilde{\rho}_a$, each y_{ij} is replaced with the corresponding random variable Y_{ij}.

As one would predict, this estimator is biased because the second measurement is not independent of the first measurement. To find the expectation and variance of the estimator $\widetilde{\rho}_a$, we note that the distribution of the second measurement conditional on the first measurement is given by

$$Y_{i2} \,|(Y_{i1} = y_{i1}) \backsim N\left(\mu + \rho(y_{i1} - \mu), \;\; \sigma_t^2 \left(1 + \rho\right)\left(1 - \rho\right)\right). \tag{7.3}$$

and when a second measurement occurs, the first measurement is outside the inspection limits. This means, that if there is a second measurement, the distribution of the first measurement, Y_{i1}, is a truncated $N(\mu, \sigma_t^2)$ such that $Y_{i1} \notin B$. By conditioning on the first measurements, we can determine the expectation of S_m^2 as

$$E\left[S_m^2\right] = \sigma_t^2 \left(1 - \rho\right)\left[1 - \left(1 - \rho\right)\frac{\beta_1}{2}\right] \tag{7.4}$$

and the variance of S_m^2 is

$$\begin{aligned} Var\left(S_m^2\right) &= \frac{2\sigma_t^4\left(1-\rho\right)^2}{n_2}\left\{1-\left(1-\rho\right)\beta_1+\frac{\left(1-\rho\right)^2}{8}\left(3\beta_1-\beta_1^2-\beta_3\right)\right\} \\ &\approx \frac{2\sigma_t^4\left(1-\rho\right)^2}{n_2}\left\{1-\left(1-\rho\right)\beta_1\right\} \end{aligned} \tag{7.5}$$

where β_i is defined in equation (D.3). The derivation of both variance and expectation are shown in Appendix (D). The expression for $Var\left(S_m^2\right)$ can be simplified as given in (7.5) because the contribution of the term $\frac{1}{8}\left(1-\rho\right)^2\left(3\beta_1-\beta_1^2-\beta_3\right)$ is small. This simplification shows that variance of S_m^2 is inflated by a multiplicative factor, $\left(1-\left(1-\rho\right)\beta_1\right)$, relative to measuring a randomly chosen part twice. Since, the covariance between S_m^2 and S_1^2 is near 0, the approximate expectation and variance of $\widetilde{\rho}_a$ are

$$E\left[\widetilde{\rho}_a\right] \approx \rho+\frac{1}{2}\beta_1\left(1-\rho\right)^2, \qquad \text{and} \tag{7.6}$$

$$Var\left(\widetilde{\rho}_a\right) \approx 2\left(1-\rho\right)^2\left\{\frac{1}{n_1-1}+\frac{1}{n_2}\left[1-\left(1-\rho\right)\beta_1\right]\right\}. \tag{7.7}$$

Equation (7.6) shows that $\widetilde{\rho}_a$ is biased. Figure 7.2, shows the bias, $\frac{1}{2}\beta_1\left(1-\rho\right)^2$, as a function of the standardized inspection limits ($\alpha_1=\frac{LIL-\mu}{\sigma_t}, \alpha_2=\frac{UIL-\mu}{\sigma_t}$) and ρ. Note that the vertical axis is $\alpha_2-\alpha_1$ in the left panel of Figure 7.2. For example, if ρ is 0.8 and $(\alpha_1,\alpha_2)=(-2,2)$, then we find the point (-2, 4) on the left panel of Figure 7.2 to obtain $\beta_1=-5$ and then on the right panel, we find the bias to be ≈ -0.1.

For one-sided inspection limits, β_1, can be determined from the left panel of Figure 7.2 because β_1 is the same for limits of the form $(-k,k)$, $(-\infty,k)$ and $(-k,\infty)$. For example, to find β_1 when the standardized inspection limits are $(-\infty,1.5)$, we look up the point $\alpha_1=-1.5$ and $\alpha_2-\alpha_1=2\times 1.5=3$ on the left

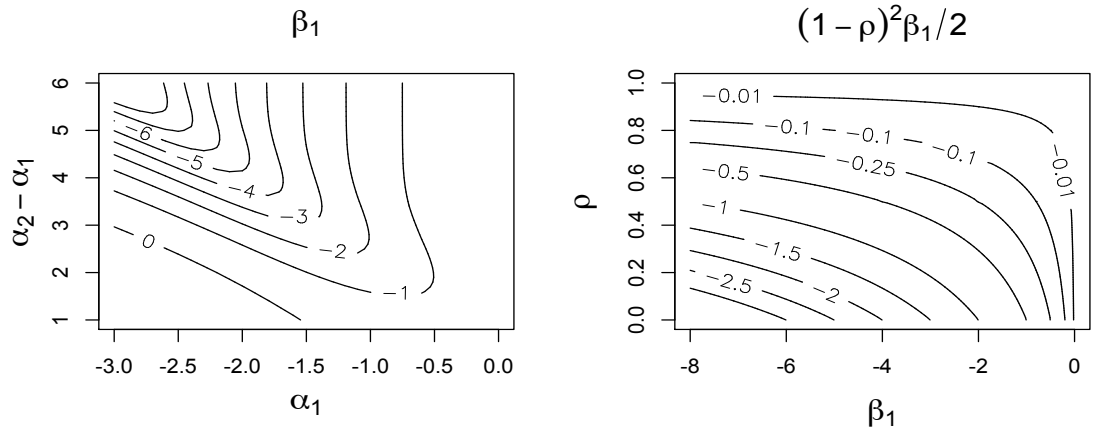

Figure 7.2: The left panel gives β_1 for different values of the standardized inspection limits. The right panel is the bias of $\tilde{\rho}_a$ as a function of ρ and β_1.

panel of Figure 7.2.

We analyze the example presented in section 7.1 using the above method. From Table 7.2, we calculate $s_m^2 = 0.725$. Thus the ANOVA estimate, $\widehat{\rho}_a$, from (7.2) is 0.969. Using Table 7.1, we can estimate β_1 from our example to be -2.05. This estimate is useful in determining the approximate bias of the ANOVA estimator. In Figure 7.2, viewing the line along $\beta_1 = -2.05$ we can see how the bias for this estimator depends on ρ. Additionally, from (7.7) the standard error of $\widehat{\rho}$ can be approximated as 0.0117.

7.2.2 Regression Estimator

The distribution of the second measurement Y_{i2} given the first measurement y_{i1}, is given in (7.3). Since the mean in (7.3) depends on ρ linearly, the variance is the same for each part and the measurements on different parts are mutually independent, we can use regression to estimate ρ. The conditional mean of Y_{i2} also depends on μ but we use the 1^{st} pass average $\overline{y}_{.1}$ to estimate this unknown. Applying the

regression estimate of ρ from (4.13) we obtain

$$\widehat{\rho}_r = \frac{\sum_{i \in S_1} (y_{i2} - \overline{y}_{.1})(y_{i1} - \overline{y}_{.1})}{\sum_{i \in S_1} (y_{i1} - \overline{y}_{.1})^2}$$

From Section 4.2.2, the distribution of $\tilde{\rho}_r$, conditional on the 1^{st} pass data, is normal with mean

$$E[\hat{\rho}_r | \, y_{i1}, i \in S_1] = \rho + \left[(\mu - \overline{y}_{.1}) \frac{\widehat{SC}}{\widehat{SSC}} \right] (1 + \rho) \tag{7.8}$$

$$\text{where} \quad \widehat{SC} = \sum_{i \in S} \left[\frac{y_{i1} - \overline{y}_{.1}}{s_1} \right] \quad \text{and} \quad \widehat{SSC} = \sum_{i \in S} \left[\frac{y_{i1} - \overline{y}_{.1}}{s_1} \right]^2. \tag{7.9}$$

Typically, n_1 will be large and thus the estimator $\tilde{\rho}_r$ will have only a small bias (conditionally) because $\overline{y}_{.1}$ will be close to μ. Unconditionally, the estimate is unbiased up to $O\left(\frac{1}{\sqrt{n_1}}\right)$ because we can replace $\overline{y}_{.1}$ with $\mu + O\left(\frac{1}{\sqrt{n_1}}\right)$.

We use the methodology in Jorgensen and Knudsen [2004] and given in (4.27) to derive two expressions for the asymptotic variance of $\tilde{\rho}_r$. The first variance is based on a particular dataset where we obtained the 1^{st} pass data. The second covariance is appropriate when we repeatedly or regularly use this method of estimation. We suggest using expression (7.11) for analysis and expression (7.10) for planning.

- variance conditional on the data

$$Var(\tilde{\rho}_r) \approx \frac{(1-\rho)^2}{n_1} \left[\frac{\widehat{SC}}{\widehat{SSC}} \right]^2 + \frac{(1-\rho)(1+\rho)}{\widehat{SSC}} \tag{7.10}$$

- variance

$$Var(\tilde{\rho}_r) \approx \frac{(1-\rho)^2}{n_1} \left[\frac{\beta_0}{1-\beta_1} \right]^2 + \frac{(1-\rho)(1+\rho)}{n_2(1-\beta_1)} \tag{7.11}$$

The standard error for the regression estimator can be estimated in two ways using either (7.10) or (7.11). The second expression (7.11) uses the properties of the truncated normal distribution estimated from the 1^{st} pass data. The first estimate

(7.10) is more robust than (7.11) because it uses $\widehat{SC}$ and $\widehat{SSC}$ defined in (7.9).

We analyze the example presented in section 7.1 using the regression method. The regression estimate is $\widehat{\rho}_r = 1.02$ which is above the upper bound for ρ. When we obtain estimates outside the range (0,1) we suggest setting the estimate to the value of the closest boundary. In this case we would set $\widehat{\rho}_r$ to 1.

To illustrate how to calculate the standard errors for the regression estimator we use $\widehat{\rho}_a = 0.969$ to estimate ρ. Using Table 7.1, we can estimate σ_t, β_0 and β_1 to be 4.86, 1.072 and -2.05, respectively. Then using Table 7.1, we calculate $\widehat{SC} = -14.50$ and $\widehat{SSC} = 56.15$. Finally, the estimate of the standard error conditional on the data is 0.0330 and the unconditional standard error is 0.0343

7.2.3 Maximum Likelihood

The log-likelihood for the inspection system data is the sum of two log-likelihoods: $l_p(\mu, \sigma_t^2)$, the likelihood of the production data and $l_{rp}\left(\mu, \sigma_t^2, \rho \,|y_{i1}, i \in S_1\right)$, the likelihood of the retest data given the production data. The distribution of the production data is independent $N(\mu, \sigma_t^2)$. The distribution of Y_{i2} given $Y_{i1} = y_{i1}$ is given in (7.3). Thus the two log-likelihoods are

$$l_p(\mu, \sigma_t^2) = -\frac{n_1}{2}\log\sigma_t^2 - \frac{1}{2\sigma_t^2}\left\{n_1 s_1^2 + n_1\left(\overline{y}_{.1} - \mu\right)^2\right\}, \quad \text{and} \tag{7.12}$$

$$l_{rp}\left(\rho, \mu, \sigma_t^2 \,|y_{i1}, i \in S_1\right) = -\frac{n_2}{2}\log\left[\sigma_t^2(1+\rho)(1-\rho)\right] - \frac{1}{2}\frac{\sum_{i\in S}\left[\overline{y}_{i2} - \mu - \rho(y_{i1} - \mu)\right]^2}{\sigma_t^2(1+\rho)(1-\rho)}. \tag{7.13}$$

The complete log-likelihood for the inspection process is the sum of the two log-likelihoods (7.12) and (7.13). To get the MLEs of μ, σ_t^2 and ρ, we numerically maximize the complete log-likelihood.

When using maximum likelihood, asymptotic standard errors for the estimators

can be obtained from the inverted Fisher information matrix with the parameters replaced by their estimates.

The Fisher information is obtained by taking the expectation of negative one times the log-likelihood second derivatives. The expectations depend on the distribution of the observations. For retest data, the distribution of the second measurements Y_{i2} is given in (7.3) and the first measurements Y_{i1} have a truncated $N(\mu, \sigma_t^2)$ with Y_{i1} outside the inspection limits. Corresponding to the two components of the log-likelihood, the Fisher information is the sum of two matrices, J_p and J_{rp}. Taking derivatives and applying expectations, we obtain

$$J_p\left(\mu, \sigma_t^2, \rho\right) = n_1 \begin{pmatrix} \frac{1}{\sigma_t^2} & 0 & 0 \\ 0 & \frac{1}{2\sigma_t^4} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{7.14}$$

$$\text{and} \qquad J_{rp}\left(\mu, \sigma_t^2, \rho\right) = n_2 \begin{pmatrix} \frac{(1-\rho)}{\sigma_t^2(1+\rho)} & 0 & -\frac{\beta_0}{\sigma_t(1+\rho)} \\ 0 & \frac{1}{2}\frac{1}{\sigma_t^4} & -\frac{\rho}{\sigma_t^2(1-\rho^2)} \\ -\frac{\beta_0}{\sigma_t(1+\rho)} & -\frac{\rho}{\sigma_t^2(1-\rho^2)} & \frac{(1+\rho^2)}{(1-\rho^2)^2} - \frac{\beta_1}{(1-\rho^2)} \end{pmatrix} \tag{7.15}$$

Details of the calculations are given in Appendix D.

The asymptotic variance of the MLE for ρ can be obtained by inverting the matrix $J_p + J_{rp}$. In general, to get a reasonable number of retests, we need a large number of production tests n_1. We can simplify the calculations if we let n_1 tend to infinity, then the asymptotic variance of the maximum likelihood estimator for ρ becomes

$$Var(\tilde{\rho}_{mle}) = \frac{(1-\rho^2)^2}{n_2\left[1 - \beta_1(1-\rho^2)\right]} \tag{7.16}$$

This is the same as the variance we get by assuming μ and σ_t are known. Figure 7.3 displays the standard error of the MLE (square root of (7.16)) as a function of ρ and β_1 assuming n_1 is large. It shows that as β_1 and ρ increase, the standard

error decreases.

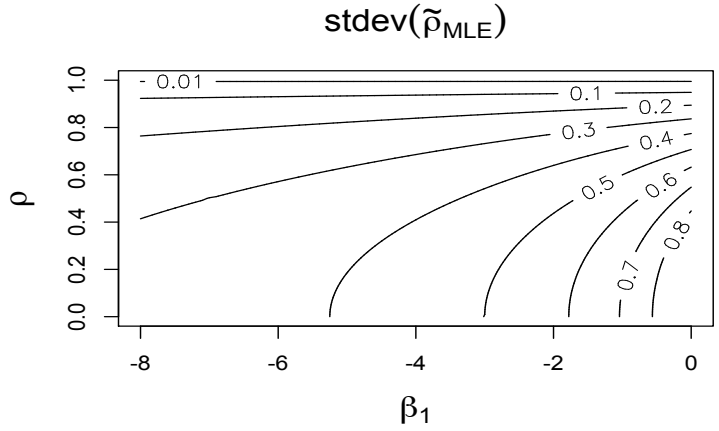

Figure 7.3: The asymptotic standard deviation of the MLE by ρ and β_1

Applying maximum likelihood to the example, we get estimates for μ, σ_t^2 and ρ of $(100.0, 24.03, 0.971)$, respectively. Using the MLEs, we estimate β_1 to be -1.98 and the standard error for the maximum likelihood estimator is 0.0105.

7.2.4 Comparison

In Figure 7.4, we compare the MSE of the ANOVA and regression estimators to the MLE. These ratios are a function of ρ and β_1. The figure shows, as expected, that the MLE is more efficient. When $\rho \geq 0.9$ the ANOVA and MLE estimators are almost equivalent but when $\rho < 0.9$ the ANOVA estimator is very inefficient. The regression estimator on the other hand is almost equivalent to the MLE for smaller ρ but does much worse when $\rho \geq 0.9$.

Inspection systems with good measurement systems will have $\rho > 0.90$. Restricting ourselves to this interval and given the cost and complexity of finding the MLE, we recommend using the ANOVA estimate.

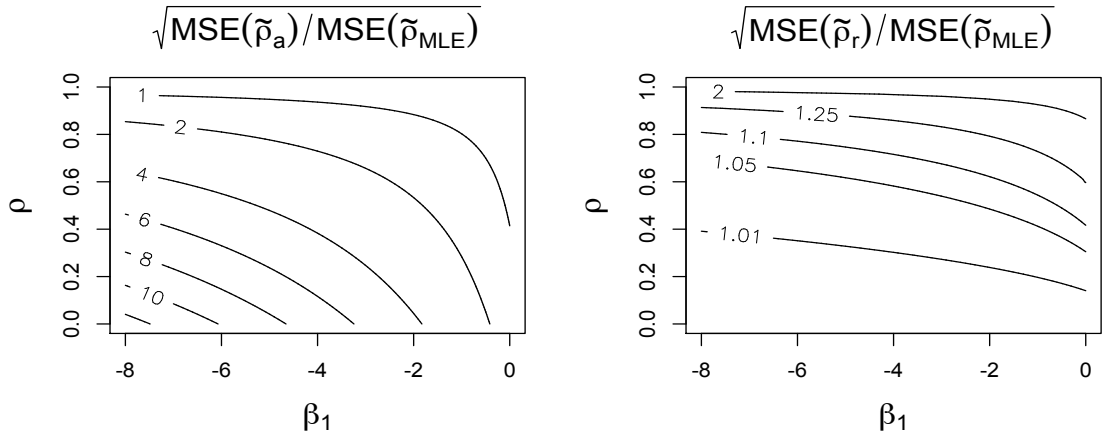

Figure 7.4: $\sqrt{\text{MSE}(\widetilde{\rho}_a)}$ and $\sqrt{\text{MSE}(\widetilde{\rho}_r)}$ divided by $\sqrt{\text{MSE}(\widetilde{\rho}_{MLE})}$ by ρ and β_1

7.3 Model Assessment

This section gives suggestions for assessing model (7.1) and applies them to the example. Normality is a key assumption in this proposal. We can assess the overall normality (part plus measurement error) from the production data. A normal quantile plot of the production data from Table 7.1 is plotted in the left panel of Figure 7.5 and shows no evidence to reject normality.

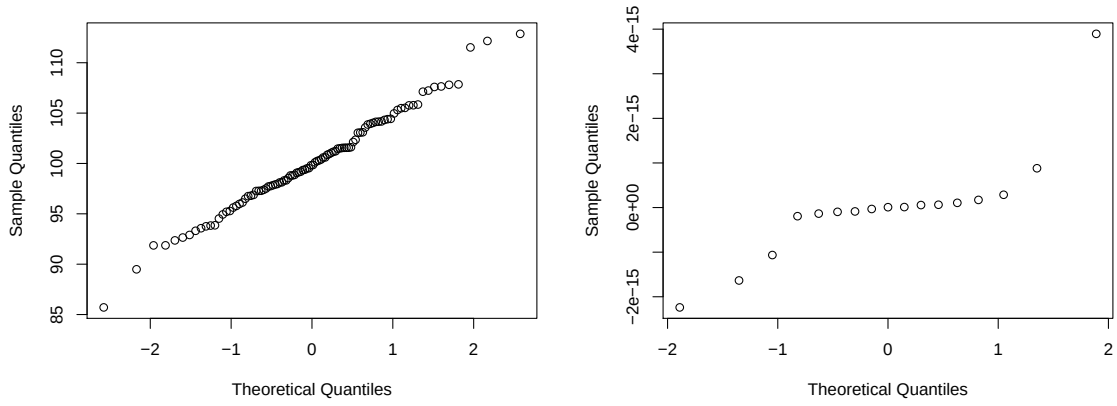

Figure 7.5: Normal quantile plots of the production data (left panel) and the residuals from regressing y_{i2} onto y_{i1} (right panel).

Another check on normality comes from equation (7.3). Since, Y_{i2} depends on Y_{i1} conditionally through the mean with the equal variance for each pair of measurements, we can regress y_{i2} onto y_{i1} and check the residuals for departures from normality. A normal quantile plot of the residuals for the data in Table 7.2 is shown from the right panel of Figure 7.5. There is no evidence to reject the model in this example.

7.4 Assessment Intervals

Using this approach we can assess the measurement system without conducting a separate off-line study but we need to specify how often we summarize the inspection data. Then we could conduct analysis on a regular schedule say, weekly/monthly, or when some desired precision is achieved. The precision can be achieved by specifying the number of parts n_1 or the number of second measurements n_2 to be included in the study.

For example, suppose we specify the desired precision for the ANOVA estimator. Using the ANOVA estimate from the example, namely 0.969, as the true value, equation (7.7) and assuming n_1 is large, we can obtain the number of second measurements, n_2, required to achieve the desired precision. If the desired standard error for $\hat{\rho}$ is 0.01 and the standardized inspection limits are $A = (-1.05, 2.04)$ and $\rho = 0.969$, then the number of second measurements needs to be at least 21.

To obtain standard errors for any inspection system use Figure 7.2 or equation (D.3) to calculate β_1. Then along with the approximate value of ρ, input these values into (7.7).

7.5 Discussion and Conclusions

In this chapter we looked at the case where there is a single response. These methods can be extended to inspection systems with k independent characteristics, although the independence assumption is likely unreasonable. If all characteristics from a part are remeasured when a single characteristic fails, then, because of independence, some of the repeated measurements are equivalent to randomly selecting a part and measuring twice. For example, suppose there are two independent characteristics X and Y and during the production test, Y fails resulting in a second measurement for both X and Y. Then the two repeated measurements for X are equivalent to taking a random part and measuring X twice. Other parts will have second measurements because the production test for X failed. So, in this situation, the ANOVA estimator will be composed of two types of repeated measurements.

The assumption that an inspection system has k independent characteristics is very restrictive but removing this assumption complicates things considerably. First we need to incorporate the dependency structure into the model (7.1). Second, we note that with multiple measurements, we are apt to get continuous, ordinal and binary characteristics determined by the same inspection system.

Using the analysis method presented in this chapter, we can also assess the measurement variation in situations where multiple gauges are used in parallel and we are trying to detect differences in variation among the gauges. This means σ_m now becomes σ_{mg} where g denotes the gauge. We assume the parts are randomly allocated to the gauges to ensure that over the long term the part variation is the same for each gauge. Here we assumed there are only two gauges but the methodology can be generalized.

Without the knowledge of Section 7.2, to assess the differences in the parallel

gauges we might use the production data or the first measurements. This method detects differences in σ_{m1} and σ_{m2} through the total variation $\sigma_{t1}^2 = \sigma_p^2 + \sigma_{m1}^2$ and $\sigma_{t2}^2 = \sigma_p^2 + \sigma_{m2}^2$. Detection of differences will be difficult if σ_p^2 is the dominant component of the variation. This is the typical situation. Thus, adding repeated measurements into this type of analysis will greatly improve the power to detect differences in the two measurement variation components when they are both small relative to the part variation. The analysis in Section 7.2 suggests that we can use the ratio of the ANOVA estimates for each gauge to compare σ_{m1} to σ_{m2g}, if $\rho \geq 0.9$ for each gauge.

Modifying the inspection protocol (see Figure 7.1) will change the results given in this thesis. For example, allowing two retests will inflate the variance of the ANOVA estimator beyond what is tolerable.

In summary we presented three ways to analyze a measurement system from inspection data. They enable to us to avoid off-line studies such as a standard Gauge R&R. We recommend the MLE but we showed that the ANOVA estimator, although biased, is comparable to the MLE if $\rho \geq 0.9$.

Chapter 8

Conclusions and Extensions

8.1 Summary

The goal of this thesis was to apply leveraging in designing new measurement system assessment plans in three contexts that commonly occur in manufacturing. In each context, we were able to show that the proposed LPs were more efficient for estimating the metric of interest than are the plans currently used in practice. We also showed, in each case, how to analyse the data from the LP and we also provided a methodology useful to practicioners for planning an LP. All of these LPs, their design and analysis, are new to the mesurement system asessment literature.

In general a leveraged plan is conducted in two stages. In the first stage, called a baseline study, we measure many randomly selected parts once. Then based on these initial measurements, we select a subsample of parts to be remeasured n times. In particular, to improve estimation, we sample parts that have initial measurements which are extreme relative to the baseline average. In addition we sample parts such that the average of the initial measurements of the selected parts is close to the baseline average. We showed, in general, that a good LP uses about half of its resources in the baseline study. This conclusion forms the basis for the

recommended LPs.

When the measurement system is used for 100% inspection, we do not need a baseline study and the LP requires only the second stage.

In addition, in Chapter 7, we presented a new approach for assessing a measurement system used as part of an inspection system with operational data. In this situation, the inspection protocol performs leveraging automatically by remeasuring parts that have a first measurement outside of the inspection limits. Using the methodology provided, practicioners can avoid costly off-line assessment studies and essentially assess the system for free.

8.2 Extensions

Further research exploiting leveraging can go in many directions. One application is determining how leveraging can be applied when comparing two measurement systems. Another is with systems that exhibit non-constant measurement variation. Here the use of leveraging would be valuable because we purposely select extreme parts. Four other topics that involve leveraging are discussed in further detail below:

- non-parametric measurement system assessments
- a problem in genetics,
- treating operator effects as random instead of fixed,
- multivariate measurement systems.

8.2.1 Nonparametric Measurement System Assessments

In a non-parametric measurement system assessment, we allow the part and/or measurement distributions to be nonparametric. Two modifications to the standard assessment are suggested. The first is robust estimation of the parameters as defined in the normal context [See Burdick et al., 2005]. For example, we could use the trimmed mean to estimate the expectation and a robust statistic to estimate the variance. The second suggestion is to use alternative parameters that are more robust [See Lai and Chew, 2000]. For example, instead of estimating variances, the parameters of interest might be defined in terms of the inter-quarantile ranges. Also, side information such as knowing the overall distribution of the measured values has never been included in non-parametric measurement system assessments but such information is readily available in high volume manufacturing settings.

8.2.2 Genetic Variation Studies

In genetic variation studies [Li et al., 2004], the goal is to determine if genetically related people (families) have more or less variability in a continuous characteristic than individuals in a population. This problem can be translated to the measurement system problem by labelling the family effect as the part effect and the individual effect as the measurement effect [See Amos, 1994, Falconer and Mackay, 1994]. There are two major differences between the measurement and the genetic problems. In the measurement problem, we can measure a part as many times as we desire whereas the number of individuals in a family is varying outside of our control. Second, the correlation between any two measurements on a part is constant whereas the correlation can vary between members within a family. To implement a leveraged genetic study, we first sample and measure b individuals (from different families) to obtain a baseline. Then from the baseline, we select k

individuals with large or small observed values and we measure each member of their families.

To overcome these two major differences we might have to select additional subjects in the second stage. Also, when planning an LP we will need to replace the number of repeated measurements with the expected family size. The analysis will be more difficult because the correlation between any two family members can vary. Some work will be required to find a closed form for the conditional mean and variance-covariance matrix. A simple solution is to numerically invert the variance-covariance although this will dramatically increase the computational effort.

8.2.3 Random Effects for Operators

In Section 1.3 and Chapter 6, we considered the addition of operators as fixed effects to the measurement system. Burdick et al. [2003] suggest that the majority of studies which treat operator effects as random do not have enough operators to properly estimate the variation due to operators. That is, when treating operator effects as random, we require a large number of operators to estimate the variance of the distribution.

In this section, we consider the effect from the operators to be random instead of fixed. For the effects to be random, we mean that the effects are realizations from a distribution. We follow the literature from Burdick et al. [2005] by assuming the operator effects have a normal distribution with mean 0 and variance σ_o^2. This extended model is written as

$$Y_{ijk} = \mu + X_i + R_j + E_{ijk} \tag{8.1}$$

where μ is the mean of the true part dimensions, X_i is a random effect of the true

part dimensions, R_j is the random effect from operator j, and E_{ijk} is the random effect from the same operator repeatedly measuring the same part on the same gauge. X, R and E are assumed to be independent normals with means equal to zero and standard deviations σ_p, σ_p and σ_m, respectively.

When we assume operators are random effects, the quantities of interest are the same as in subsection 1.3.1. We will treat η as the primary parameter of interest and λ as of secondary interest.

In this context, we again conduct a leveraged measurement system assessment in two stages:

Stage 1: Sample b parts at random from the process and have b operators measure a different part once each to obtain a baseline of b measurements. We denote the observed values $\{y_{10}, y_{20}, \ldots, y_{b0}\}$ and the baseline average and sample variance by $\overline{y}_b = \frac{1}{b}\sum_{i=1}^{b} y_{i0}$ and $s_b^2 = \frac{1}{b-1}\sum_{i=1}^{b}(y_{i0} - \overline{y}_b)^2$.

Stage 2: From the baseline sample, select k parts and/or operators using the observed measured values. In particular, to improve the estimation for η, sample k parts and/or operators such that

- the initial measurements are extreme relative to the baseline average, and
- the average of the initial measurements of the selected parts is close to the baseline average.

We denote the k selected parts using the set S. These k parts and/or operators are then repeatedly measured n times each to give the additional data $\{y_{ij}, i \in S \ \ and \ \ j = 1, \ldots, n\}$. The total number of measurements in the leveraged plan is $N = b + nk$.

For example, for a leveraged plan with $k = 2$, we might pick the parts and/or operators with the minimum and maximum initial measurement in the baseline sample.

There are at least three options for leveraging. We may leveraged by part, by operator or by part and operator. These options correspond to the type of sampling used for Stage 2. All three options are based on the idea that each measurement within the baseline has three components: a measured value, a part number and an operator number. We use the measured value to select either the part and/or operator to be used in the second stage.

1. When leveraging by part we select the parts associated with the extreme measurements in the baseline. Then in Stage 2 we repeatedly measure the chosen parts with m new operators n times each. The total sample size is $N = b + kmn$.

   For example, when leveraging by part with $k = 2, m = 3, n = 2$, we might pick the parts with the minimum and maximum initial measurements in the baseline sample. Then we recruit three new operators to measure these parts twice each.

2. Similarly, when leveraging by operator we select the operators associated with the extreme measurements in the baseline. Then in Stage 2 we repeatedly measure k new parts n times each with the m choosen operators. The total sample size is $N = b + kmn$.

   For example, when leveraging by operator with $k = 2, m = 3, n = 2$, we might pick the operators associated with the minimum, maximum and 2^{nd} largest initial measurements in the baseline sample. Then we obtain two new parts for these three operators to measure twice each.

3. In Leveraging by part and operator, we select l operators and parts associated with the extreme measurements in the baseline. Then in Stage 2 the l chosen operators measure each of the l parts n times each. Note that with this version of leveraging no new parts or operators are introduced in Stage 2. The total sample size is $N = b + l^2 n$.

   For example, when leveraging by part and operator with $l = 2, n = 2$, we might pick the parts and operators associated with the minimum and maximum initial measurements in the baseline sample. Then the two operators measure the two parts twice more each.

The three options for leveraging will behave very differently. We believe leveraging by part will be most efficient for estimating η and the least efficient for estimating λ. As for leveraging by operator we think it the least efficient for estimating η and the most efficient for estimating λ. Finally, we believe leveraging by operator and part will not be the most or least efficient but the balance between the two plans. Unfortunately, we think the optimal plan will depend on our location in the parameter space.

8.2.4 Assessing a Bivariate Measurement System

Another possible extension is to consider a bivariate measurement system where two characteristics are measured on each part. An example of two characteristics might be length and width or if we have people instead of parts, the two characteristics could be height and weight.

If these observed characteristics are independent (both the true dimensions and measurement errors are independent) then we can analyze them individually. However when correlation exists between them, how should we assess the quality of the bivariate measurement system? Sweeney [2007] shows how to properly analyze data

from a standard plan when it is used to assess a bivariate measurement system. The question of interest is how to carry out a leveraged plan and will this plan yield more efficient estimators than the standard plan in the multi-dimensional case?

The model assumed in (8.2) is the multivariate version of model (1.1). Similar to (1.1) we assume the random effects in (8.2) are multivariate normal (MVN).

$$Y = \mu + P + M \tag{8.2}$$

where

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \quad P = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}, \quad M = \begin{pmatrix} M_1 \\ M_2 \end{pmatrix},$$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad P \sim \text{MVN}\left(0, \Sigma_x\right), \quad M \sim \text{MVN}\left(0, \Sigma_e\right),$$

$$\Sigma_p = \begin{bmatrix} \sigma_{p1}^2 & \lambda_p \\ \lambda_p & \sigma_{p2}^2 \end{bmatrix}, \quad \Sigma_m = \begin{bmatrix} \sigma_{m2}^2 & \lambda_m \\ \lambda_m & \sigma_{m2}^2 \end{bmatrix} \quad \Rightarrow \quad \Sigma_t = \Sigma_p + \Sigma_m = \begin{bmatrix} \sigma_{t2}^2 & \lambda_t \\ \lambda_t & \sigma_{t2}^2 \end{bmatrix},$$

and

- σ_{ti}^2 is the total variation of characteristic $i, i = 1, 2$;
- σ_{pi}^2 is the part variation of characteristic $i, i = 1, 2$;
- σ_{mi}^2 is the measurement variation of characteristic $i, i = 1, 2$;
- λ_m is the covariance between the measurements of the two characteristics;
- λ_p is the covariance between the true values of the two characteristics;
- λ_t is the total covariance between the characteristic 1 and 2;
- μ_i is the mean of characteristic $i, i = 1, 2$.

The goal of a measurement system study is to assess $\rho_i = \sigma_{pi}^2/\sigma_{ti}^2, i = 1, 2$. Another goal used in Sweeney [2007] and Voelkel [2003], is to assess the diameter of a circle that captures 99% of repeated readings made on one part, where the center of the circle corresponds to the mean of the distribution. Voelkel [2003] suggests that this measure be used when the engineering tolerance is a diameter.

As in the one-dimensional case, the leveraged measurement assessment plan can be conducted in two stages. From a baseline study we will be able to estimate any of the parameters that involve the means, variances and covariance $(\mu_1, \mu_2, \sigma_{t1}^2, \sigma_{t2}^2, \lambda_t)$. Then a sample of k parts that are extreme (a definition of extreme in the bivariate case is to be determined) will be repeatedly measured n times each.

An extreme part in the one-dimensional case has a measured dimension which is extreme relative to the baseline average but it was the part's squared standardized value that reduced the asymptotic standard errors. In the bivariate case, we believe the squared standardized value will again play an important role. Although here the standardized value is slightly different because we need to account for the covariances. In the multivariate case, the squared standard initial measurement is

$$(\mathbf{y}_{i0} - \mu)^t \, \Sigma_t^{-1} \, (\mathbf{y}_{i0} - \mu) \tag{8.3}$$

Note that the squared standard initial measurement reduces to the Euclidean distance when the variance-covariance matrix is diagonal. i.e. when the characteristics are independent.

Using the baseline data we can estimate the value in (8.3) using the Mahalanobis distance squared, denoted D_M^2. The Mahalanobis distance is basically (8.3) with Σ_t and μ replaced by the baseline estimates $\widehat{\Sigma}_t$ and $\widehat{\mu}$, respectively.

$$\left[D_M \left(\mathbf{y}_{i0}, \overline{\mathbf{y}}_{.0}; \widehat{\Sigma}_t \right) \right]^2 = (\mathbf{y}_{i0} - \overline{\mathbf{y}}_{.0})^t \, \widehat{\Sigma}_t^{-1} \, (\mathbf{y}_{i0} - \overline{\mathbf{y}}_{.0}) \tag{8.4}$$

We believe this distance can be useful to define a leveraged part because when the two characteristics are not independent it will weight an initial measurement by the covariances and standard deviations.

One concern with using the Mahalanobis distance is that it may suggest parts which are extreme only in a single dimension. On average, this method may be most efficient but in particular cases one estimator of ρ_i might be good while the other is poor. In this context, the effect of different sampling methods needs to be explored. A possible alternate sampling approach with bivariate data is to select half the parts for remeasurement based solely on the first dimension and select the other half based solely on the second dimension.

In the bivariate case, we propose to consider two situations. The first occurs when interest lies in quantifying the measurement variation marginally in the two correlated characteristics. In this situation, two questions should be answered. Does an LP yield more efficient estimators than an SP and what is gained from using a two-dimensional method instead of doing the analyses marginally?

The other situation involves quantifying the variation of the distance (radius) from the true value and not the marginal components. This situation is considered in Sweeney [2007]. We will also consider if the LP is more efficient than the SP. Another question here is whether a two-dimensional analysis is even required. Can the same efficiency be achieved by performing the analysis marginally on the distance? The use of baseline information (known μ and Σ_t) can also be considered in the same manner as in Chapter 5.

Appendix A

Two Stage Leveraged Plan

A.1 Fisher Information

Why does SC=0 and SSC≫0 reduce the asymptotic variance of the MLE? Let

$$J\left(\mu, \sigma_t^2, \rho\right) = \begin{pmatrix} x & 0 & t \\ 0 & y & v \\ t & v & z \end{pmatrix}$$

where $x, y, z \geq 0$. Using the principal minors, the determinant and inverse of J are

$$det(J) = x \begin{vmatrix} y & v \\ v & z \end{vmatrix} - 0 + t \begin{vmatrix} 0 & t \\ y & v \end{vmatrix} = x\left(yz - v^2\right) - t^2 y$$

$$J^{-1} = \frac{1}{det(J)} \begin{pmatrix} yz - v^2 & vt & -yt \\ vt & xz - t^2 & -xv \\ -yt & -xv & xy \end{pmatrix}.$$

This means the asymptotic variance of maximum likelihood estimator of ρ is

$$Asvar\left(\tilde{\rho}\right) = \frac{xy}{x(yz - v^2) - t^2 y} = \frac{1}{z - v^2/y - t^2/x}$$

Ideally, $Asvar\left(\tilde{\rho}\right)$ is close to zero. From (4.10) we see that selecting parts to repeatedly remeasured affects t and z. The $Asvar\left(\tilde{\rho}\right)$ is reduced when z is large and $t = 0$. Since, $x, y, z \geq 0$, we can reduce $Asvar\left(\tilde{\rho}\right)$ by decreasing v^2 or t^2. We cannot change v, but we can set $t = 0$ by selecting parts with initial measurements such that $E\left[SC\right] = 0$. If we choose parts with large and small extreme measurements we can get $E\left[SC\right] = 0$, which means $t = 0$ and $E\left[SSC\right]$ is large which increases z.

Appendix B

Leveraged Plan with Operators as Fixed Effects

B.1 Direction

In this section we derive a detailed expression for the asymptotic variance of η by simplifying the result given in (6.20) which is product of three matrices. Then we show that asymptotic variance of η can be written as a function of the inner product of two vectors $\mathbf{u}^t\mathbf{r}$.

We start with the following two results about block matrix inversion established by Banachiewicz [1937]. For a general block matrix the inverse is

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix} \tag{B.1}$$

and the determinant is

$$\det\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \det(\mathbf{A})\ \det\left(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right).$$

From (6.19), we notice that the Fisher information matrix for $(\underline{\mu}, \sigma^2_{pg}, \rho)$ has the form

$$F_{A1} = \begin{pmatrix} \mathbf{M} & \mathbf{B} \\ \mathbf{B}^t & \mathbf{P} \end{pmatrix} \quad \text{where} \quad B = \begin{pmatrix} \mathbf{0}_m & \mathbf{v} \end{pmatrix}.$$

Applying (B.1) we obtain

$$F_{A1}^{-1} = \begin{bmatrix} \mathbf{M}^{-1} + \mathbf{K} & -\mathbf{M}^{-1}\mathbf{B}\mathbf{Q}^{-1} \\ -\mathbf{Q}^{-1}\mathbf{B}^{\mathbf{t}}\mathbf{M}^{-1} & \mathbf{Q}^{-1} \end{bmatrix}$$

$$\text{where} \quad \mathbf{Q} = \mathbf{P} - \mathbf{B}^{\mathbf{t}}\mathbf{M}^{-1}\mathbf{B} \quad \text{and} \quad \mathbf{K} = \mathbf{M}^{-1}\mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^{\mathbf{t}}\mathbf{M}^{-1}. \tag{B.2}$$

Additionally, the matrix $\mathbf{D}$ from (2.22) can be written as

$$\mathbf{D} = \begin{pmatrix} \mathbf{I}_m & \mathbf{E} \\ \mathbf{0}^t_{m\times 2} & \mathbf{T} \end{pmatrix} \quad \text{where} \quad E = \begin{pmatrix} \mathbf{m} & \mathbf{0}_m \end{pmatrix}.$$

Applying (B.1) to $\mathbf{D}$ we obtain

$$\mathbf{D}^{-1} = \begin{bmatrix} \mathbf{I} & -\mathbf{E}\mathbf{T}^{-1} \\ \mathbf{0}^t_{m\times 2} & \mathbf{T}^{-1} \end{bmatrix}$$

Now, we evaluate the product $\mathbf{D}^{-t}\mathbf{F}_{A1}^{-1}\mathbf{D}^{-1}$ to obtain the asymptotic variance of η and λ,

$$\mathbf{D}^{-t}{\mathbf{F}_{\mathbf{1}}}^{-1}\mathbf{D}^{-1}$$

$$
\begin{aligned}
&= \begin{bmatrix} \mathbf{I} & \mathbf{0}_{m\times 2} \\ -\mathbf{T}^{-1}\mathbf{E}^t & \mathbf{T}^{-t} \end{bmatrix} \begin{bmatrix} \mathbf{M}^{-1}+\mathbf{K} & -\mathbf{M}^{-1}\mathbf{B}\mathbf{Q}^{-1} \\ -\mathbf{Q}^{-1}\mathbf{B}^{\mathbf{t}}\mathbf{M}^{-1} & \mathbf{Q}^{-1} \end{bmatrix} \mathbf{D}^{-1} \\
&= \begin{bmatrix} \mathbf{M}^{-1}+\mathbf{K} & -\mathbf{M}^{-1}\mathbf{B}\mathbf{Q}^{-1} \\ -\mathbf{T}^{-1}\left(\mathbf{E}^t\left(\mathbf{M}^{-1}+\mathbf{K}\right)+\mathbf{Q}^{-1}\mathbf{B}^{\mathbf{t}}\mathbf{M}^{-1}\right) & \mathbf{T}^{-t}\left(\mathbf{E}^t\mathbf{M}^{-1}\mathbf{B}\mathbf{Q}^{-1}+\mathbf{Q}^{-1}\right) \end{bmatrix} \mathbf{D}^{-1} \\
&= \begin{bmatrix} \mathbf{M}^{-1}+\mathbf{K} & -\left(\left(\mathbf{M}^{-1}+\mathbf{K}\right)\mathbf{E}+\mathbf{M}^{-1}\mathbf{B}\mathbf{Q}^{-1}\right)\mathbf{T}^{-1} \\ -\mathbf{T}^{-1}\left(\mathbf{E}^t\left(\mathbf{M}^{-1}+\mathbf{K}\right)+\mathbf{Q}^{-1}\mathbf{B}^{\mathbf{t}}\mathbf{M}^{-1}\right) & \mathbf{Z} \end{bmatrix}
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{Z} &= \mathbf{T}^{-t}\left(\mathbf{E}^t\left(\mathbf{M}^{-1}+\mathbf{K}\right)+\mathbf{Q}^{-1}\mathbf{B}^{\mathbf{t}}\mathbf{M}^{-1}\right)\mathbf{E}\mathbf{T}^{-1}+\mathbf{T}^{-1}\left(\mathbf{E}^t\mathbf{M}^{-1}\mathbf{B}\mathbf{Q}^{-1}+\mathbf{Q}^{-1}\right)\mathbf{T}^{-1} \\
&= \mathbf{T}^{-t}\left(\left(\mathbf{E}^t\left(\mathbf{M}^{-1}+\mathbf{K}\right)+\mathbf{Q}^{-1}\mathbf{B}^{\mathbf{t}}\mathbf{M}^{-1}\right)\mathbf{E}+\left(\mathbf{E}^t\mathbf{M}^{-1}\mathbf{B}\mathbf{Q}^{-1}+\mathbf{Q}^{-1}\right)\right)\mathbf{T}^{-1} \\
&= \mathbf{T}^{-t}\left(\mathbf{E}^t\left(\mathbf{M}^{-1}+\mathbf{K}\right)\mathbf{E}+\mathbf{Q}^{-1}\mathbf{B}^{\mathbf{t}}\mathbf{M}^{-1}\mathbf{E}+\mathbf{E}^t\mathbf{M}^{-1}\mathbf{B}\mathbf{Q}^{-1}+\mathbf{Q}^{-1}\right)\mathbf{T}^{-1}.
\end{aligned}
$$

We are only interested in the matrix $\mathbf{Z}$ because it contains the asymptotic variance for λ and η. We can simplify $\mathbf{Z}$ further by letting $\mathbf{X} = \mathbf{B}^{\mathbf{t}}\mathbf{M}^{-1}\mathbf{E}$ and substituting $\mathbf{K} = \mathbf{M}^{-1}\mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^{\mathbf{t}}\mathbf{M}^{-1}$ then

$$
\mathbf{Z} = \mathbf{T}^{-t}\left[\mathbf{E}^t\mathbf{M}^{-1}\mathbf{E}+\left(\mathbf{I}_2+\mathbf{X}\right)^t\mathbf{Q}^{-1}\left(\mathbf{I}_2+\mathbf{X}\right)\right]\mathbf{T}^{-1}. \tag{B.3}
$$

The first component of $\mathbf{Z}$ can be determined from

$$
\mathbf{E}^t\mathbf{M}^{-1}\mathbf{E} = \begin{pmatrix} \mathbf{m}^t\mathbf{M}^{-1}\mathbf{m} & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and setting} \quad \mathbf{T}^{-1} = \frac{1}{\det\left(\mathbf{T}\right)}\begin{pmatrix} t_{2,2} & -t_{1,2} \\ -t_{2,1} & t_{1,1} \end{pmatrix}
$$

$$
\text{then} \quad \mathbf{T}^{-t}\left(\mathbf{E}^t\mathbf{M}^{-1}\mathbf{E}\right)\mathbf{T}^{-1} = \frac{\left(\mathbf{m}^t\mathbf{M}^{-1}\mathbf{m}\right)}{\det\left(\mathbf{T}\right)^2}\begin{pmatrix} t_{2,2}^2 & -t_{2,2}t_{1,2} \\ -t_{2,2}t_{1,2} & t_{1,2}^2 \end{pmatrix}. \tag{B.4}
$$

For the second component we find that

$$\mathbf{I}_2 + \mathbf{X} = \mathbf{I}_2 + \mathbf{B^t M^{-1} E} = \begin{pmatrix} 1 & 0 \\ \mathbf{v}^t\mathbf{M}^{-1}\mathbf{m} & 1 \end{pmatrix}$$

$$\text{and if we let} \qquad \mathbf{Q}^{-1} = \frac{1}{\det(\mathbf{Q})} \begin{pmatrix} q_{2,2} & -q_{1,2} \\ -q_{1,2} & q_{1,1} \end{pmatrix}$$

$$\text{and let} \qquad \alpha = \mathbf{v}^t\mathbf{M}^{-1}\mathbf{m} \tag{B.5}$$

$$\text{then} \qquad (\mathbf{I}_2 + \mathbf{X})^t \mathbf{Q}^{-1} (\mathbf{I}_2 + \mathbf{X}) = \frac{1}{\det(\mathbf{Q})} \begin{pmatrix} q_{2,2} - 2\alpha q_{1,2} + \alpha^2 q_{1,1} & -q_{1,2} + \alpha q_{1,1} \\ -q_{1,2} + \alpha q_{1,1} & q_{1,1} \end{pmatrix}$$

Finally, we can determine the $[2,2]$ element of $\mathbf{T}^{-t}(\mathbf{I}_2 + \mathbf{X})^t \mathbf{Q}^{-1}(\mathbf{I}_2 + \mathbf{X})\mathbf{T}^{-1}$ to be

$$\frac{\left[\left(t_{1,2}^2 q_{1,1}\right)\alpha^2 - 2\left(t_{1,2}t_{1,1}q_{1,1} + t_{1,2}^2 q_{1,2}\right)\alpha + \left(t_{1,2}^2 q_{2,2} + t_{1,1}^2 q_{1,1} + 2t_{1,2}t_{1,1}q_{1,2}\right)\right]}{\det(\mathbf{Q})\det(\mathbf{T})^2}.$$

If we complete the square on α, we get

$$\frac{1}{\det(\mathbf{Q})\det(\mathbf{T})^2} \frac{t_{1,2}^2}{q_{1,1}} \left[q_{1,1}^2 \left(\frac{t_{1,1}}{t_{1,2}} + \frac{q_{1,2}}{q_{1,1}} - \alpha \right)^2 + \det(\mathbf{Q}) \right] \tag{B.6}$$

Combining the first (B.4) and second (B.6) components we get the asymptotic variance of η as

$$\frac{1}{\det(\mathbf{T})^2} \frac{t_{1,2}^2}{q_{1,1}} \left[\frac{q_{1,1}^2}{\det(\mathbf{Q})} \left(\frac{t_{1,1}}{t_{1,2}} + \frac{q_{1,2}}{q_{1,1}} - \alpha \right)^2 + 1 + q_{1,1}\left(\mathbf{m}^t\mathbf{M}^{-1}\mathbf{m}\right) \right]. \tag{B.7}$$

To further explore how the asymptotic variance depends on the selection of parts for Stage 2, we look at the term that involves α. We first recall that $\mathbf{T}$ is the matrix

formed from the last two components of D, as defined in (2.22),

$$\mathbf{T} = \begin{bmatrix} -\frac{\sigma_o^2}{\lambda^2(1-\eta)} & \frac{\eta(1-\eta)}{(\lambda\eta+1-\lambda)^2} \\ \frac{\sigma_o^2}{\lambda(1-\eta)^2} & \frac{1-\lambda}{(\lambda\eta+1-\lambda)^2} \end{bmatrix}.$$

Then we can write down the two terms that are squared along with α in (B.7)

$$\frac{t_{1,1}}{t_{1,2}} = -\frac{\sigma_o^2(\lambda\eta+1-\lambda)^2}{\lambda^2\eta(1-\eta)^2} \tag{B.8}$$

and

$$\frac{q_{1,2}}{q_{1,1}} = -\frac{1}{\sigma_{pg}^2}\frac{(1-\rho)(1+mn\rho)}{\rho(mn+1)} \approx -\frac{1}{\sigma_{pg}^2}(1-\rho) = -\frac{\lambda(1-\lambda)(1-\eta)^2}{(\lambda\eta+1-\lambda)^2\sigma_o^2}. \tag{B.9}$$

Since, the two quantities (B.8) and (B.9) are both negative, to minimize the asymptotic variance of η in (B.7) we need to set $-\alpha$ to be positive. Unfortunately, we cannot we can minimize the asymptotic variance. To deduce how to use the selection method to get $\alpha = \mathbf{v}^t\mathbf{M}^{-1}\mathbf{m}$ to be negative, we begin with

$$\mathbf{m} = c_0\mathbf{r} \quad \text{where} \quad \mathbf{r} = (\underline{\mu} - \overline{\mu}\underline{1}_m) \quad \text{and} \quad c_0 = \frac{2}{m}\frac{(\lambda\eta+1-\lambda)}{\lambda(1-\eta)}. \tag{B.10}$$

Note, $\mathbf{r}$ is a vector of differences between each operator's mean and the overall mean and $\sigma_o^2 = \frac{1}{m}\mathbf{r}^t\mathbf{r}$. From the Fisher information (6.17), the vector $\mathbf{v}$ equals

$$\sum_{(i,j)\in S}\left[\mathbf{v}_0 + \mathbf{v}_{(i,j)}\right] = \frac{n}{\sigma_{pg}(1-\rho)(1+mn\rho)}\left[\sum_{(i,j)\in S} E\left[Z_{ij0}\right]\underline{1}_m - m\rho\sum_{(i,j)\in S} E\left[Z_{ij0}\right]\underline{e}_j\right].$$

If we define

$$\mathbf{u} = \sum_{(i,j)\in S} E\left[Z_{ij0}\right]\underline{e}_j = \begin{bmatrix} \sum_{(i,1)\in S} E\left[Z_{ij0}\right] \\ \vdots \\ \sum_{(i,m)\in S} E\left[Z_{ij0}\right] \end{bmatrix}, \tag{B.11}$$

$\mathbf{v}$ simplifies to

$$\mathbf{v} = \frac{n}{\sigma_{pg}(1-\rho)(1+mn\rho)}\left[\left(\mathbf{u}^t\underline{1}_m\right)\underline{1}_m - m\rho\mathbf{u}\right] = c_v\left(\mathbf{u}^t\underline{1}_m\right)\underline{1}_m - c_v m\rho\mathbf{u}.$$

The next term in α is $\mathbf{M}$ which is the Fisher information for $\underline{\mu}$ found in (6.17),

$$\mathbf{M} = kn(1+mn\rho)\mathbf{I}_m - kn^2\rho\mathbf{J}_m + \sum_{(i,j)\in S}\left[n\rho(m\rho-1)\mathbf{I}_{(j)} - n\rho\mathbf{J}_{(m)}\right].$$

If we select k parts and allocate them evenly to each operator then $\sum_{(i,j)\in S}\mathbf{I}_{(j)} \approx \frac{k}{m}\mathbf{I}_m$ and $\sum_{(i,j)\in S}\mathbf{J}_{(j)} \approx \frac{k}{m}\mathbf{J}_m$. If each operator is allocated the same number of parts then these approximations will be equalities and $\frac{k}{m}$ will be a positive integer. Under these approximations, we have

$$\mathbf{M} = \frac{kn}{m}\left[(m + m^2n\rho + m\rho^2 - \rho)\mathbf{I}_m - \rho(nm+1)\mathbf{J}_m\right]$$

This matrix has the usual special form, making the inverse

$$\mathbf{M}^{-1} = \frac{m}{kn}\frac{1}{(m+m^2n\rho+m\rho^2-\rho)}\left[\mathbf{I}_m + \frac{\rho(nm+1)}{[m(1+\rho^2)-\rho(m+1)]}\mathbf{J}_m\right] = c_1\mathbf{I}_m + c_2\mathbf{J}_m.$$

Finally, we can write an expression for $\alpha = \mathbf{v}^t\mathbf{M}^{-1}\mathbf{m}$ as

$$\begin{aligned}
& \mathbf{v}^t\mathbf{M}^{-1}\mathbf{m} \\
= \;& \left[c_v\left(\mathbf{u}^t\underline{1}_m\right)\underline{1}_m - c_v m\rho\mathbf{u}\right]^t\left[c_1\mathbf{I}_m + c_2\mathbf{J}_m\right]\left[c_0\mathbf{r}\right] \\
= \;& c_v c_0\left[\left(\mathbf{u}^t\underline{1}_m\right)\underline{1}_m - m\rho\mathbf{u}\right]^t\left[c_1\mathbf{I}_m + c_2\mathbf{J}_m\right]\mathbf{r} \\
= \;& c_v c_0\left[\left(\mathbf{u}^t\underline{1}_m\right)\underline{1}_m - m\rho\mathbf{u}\right]^t\left[c_1\mathbf{r} + c_2\underline{1}_m(\underline{1}_m^t\mathbf{r})\right] \\
= \;& c_v c_0\left[c_1\left(\mathbf{u}^t\underline{1}_m\right)\left(\underline{1}_m^t\mathbf{r}\right) - c_1 m\rho\left(\mathbf{u}^t\mathbf{r}\right) + mc_2\left(\mathbf{u}^t\underline{1}_m\right)\left(\underline{1}_m^t\mathbf{r}\right) - c_2 m\rho(\underline{1}_m^t\mathbf{r})\left(\mathbf{u}^t\underline{1}_m\right)\right] \\
= \;& c_v c_0\left[-c_1 m\rho\left(\mathbf{u}^t\mathbf{r}\right) + (c_1 + mc_2 - c_2 m\rho)\left(\mathbf{u}^t\underline{1}_m\right)\left(\underline{1}_m^t\mathbf{r}\right)\right]
\end{aligned}$$

We have that

$$\left(\underline{1}_m^t \mathbf{r}\right) = \underline{1}_m^t(\underline{\mu} - \overline{\mu}\underline{1}_m) = m\overline{\mu} - \sum_{j=1}^{m} \mu_j = 0,$$

$$\Rightarrow \qquad \alpha = -c_v c_0 m \rho c_1 \left(\mathbf{u}^t \mathbf{r}\right) \tag{B.12}$$

From this equation we can see that the value of α depends on the inner product $(\mathbf{u}^t\mathbf{r})$.

B.2 Tables with optimal designs for $m = 3$

Table B.1: The LP designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.05, 0.05, 0.15, 0.25, 0.35, 0.45)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 60$.

	0.05	0.15	0.25	0.35	0.45
0.01	(2,1,18)	(2,1,18)	(2,1,18)	(2,1,18)	(2,1,18)
0.05	(4,4,4)	(5,5,3)	(5,5,3)	(8,6,2)	(8,6,2)
0.10	(8,6,2)	(8,6,2)	(8,6,2)	(8,6,2)	(8,6,2)
0.15	(8,6,2)	(8,6,2)	(8,6,2)	(8,6,2)	(8,6,2)
0.20	(8,6,2)	(8,6,2)	(8,6,2)	(8,6,2)	(8,6,2)
0.25	(8,6,2)	(8,6,2)	(8,6,2)	(8,6,2)	(10,5,2)
0.30	(8,6,2)	(8,6,2)	(8,6,2)	(10,5,2)	(12,4,2)
0.35	(8,6,2)	(8,6,2)	(8,6,2)	(10,5,2)	(12,4,2)
0.40	(8,6,2)	(8,6,2)	(10,5,2)	(10,5,2)	(12,4,2)
0.45	(8,6,2)	(10,5,2)	(10,5,2)	(12,4,2)	(12,4,2)
0.50	(8,6,2)	(10,5,2)	(10,5,2)	(12,4,2)	(12,4,2)
0.55	(10,5,2)	(10,5,2)	(10,5,2)	(12,4,2)	(12,4,2)
0.60	(10,5,2)	(10,5,2)	(10,5,2)	(12,4,2)	(12,4,2)
0.65	(10,5,2)	(10,5,2)	(10,5,2)	(12,4,2)	(12,4,2)
0.70	(10,5,2)	(10,5,2)	(10,5,2)	(12,4,2)	(12,4,2)
0.75	(10,5,2)	(10,5,2)	(10,5,2)	(12,4,2)	(12,4,2)
0.80	(10,5,2)	(10,5,2)	(10,5,2)	(12,4,2)	(12,4,2)
0.85	(10,5,2)	(10,5,2)	(10,5,2)	(12,4,2)	(12,4,2)
0.90	(10,5,2)	(10,5,2)	(10,5,2)	(10,5,2)	(12,4,2)
0.95	(10,5,2)	(10,5,2)	(10,5,2)	(10,5,2)	(12,4,2)
0.99	(10,5,2)	(10,5,2)	(10,5,2)	(10,5,2)	(12,4,2)

Table B.2: The LP designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.55, 0.65, 0.75, 0.85, 0.95)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 60$.

	0.55	0.65	0.75	0.85	0.95
0.01	(2,1,18)	(5,5,3)	(8,6,2)	(8,6,2)	(8,6,2)
0.05	(8,6,2)	(8,6,2)	(8,6,2)	(10,5,2)	(12,4,2)
0.1	(8,6,2)	(8,6,2)	(12,4,2)	(12,4,2)	(12,4,2)
0.15	(8,6,2)	(10,5,2)	(12,4,2)	(12,4,2)	(12,4,2)
0.2	(10,5,2)	(12,4,2)	(12,4,2)	(12,4,2)	(14,3,2)
0.25	(12,4,2)	(12,4,2)	(12,4,2)	(12,4,2)	(14,3,2)
0.3	(12,4,2)	(12,4,2)	(12,4,2)	(12,4,2)	(14,3,2)
0.35	(12,4,2)	(12,4,2)	(12,4,2)	(12,4,2)	(14,3,2)
0.4	(12,4,2)	(12,4,2)	(12,4,2)	(12,4,2)	(14,3,2)
0.45	(12,4,2)	(12,4,2)	(12,4,2)	(12,4,2)	(14,3,2)
0.5	(12,4,2)	(12,4,2)	(12,4,2)	(12,4,2)	(14,3,2)
0.55	(12,4,2)	(12,4,2)	(12,4,2)	(12,4,2)	(14,3,2)
0.6	(12,4,2)	(12,4,2)	(12,4,2)	(12,4,2)	(14,3,2)
0.65	(12,4,2)	(12,4,2)	(12,4,2)	(12,4,2)	(14,3,2)
0.7	(12,4,2)	(12,4,2)	(12,4,2)	(12,4,2)	(16,2,2)
0.75	(12,4,2)	(12,4,2)	(12,4,2)	(14,3,2)	(16,2,2)
0.8	(12,4,2)	(12,4,2)	(12,4,2)	(14,3,2)	(16,2,2)
0.85	(12,4,2)	(12,4,2)	(12,4,2)	(14,3,2)	(16,1,4)
0.9	(12,4,2)	(12,4,2)	(12,4,2)	(14,3,2)	(16,1,4)
0.95	(12,4,2)	(12,4,2)	(12,4,2)	(14,3,2)	(16,1,4)
0.99	(12,4,2)	(12,4,2)	(12,4,2)	(14,3,2)	(16,1,4)

Table B.3: The LP designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.05, 0.05, 0.15, 0.25, 0.35, 0.45)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 90$.

	0.05	0.15	0.25	0.35	0.45
0.01	(2,1,28)	(2,1,28)	(2,1,28)	(2,1,28)	(2,2,14)
0.05	(6,6,4)	(12,6,3)	(12,6,3)	(12,6,3)	(12,6,3)
0.1	(12,6,3)	(12,6,3)	(12,6,3)	(12,6,3)	(14,8,2)
0.15	(12,6,3)	(12,6,3)	(14,8,2)	(14,8,2)	(14,8,2)
0.2	(14,8,2)	(14,8,2)	(14,8,2)	(14,8,2)	(16,7,2)
0.25	(14,8,2)	(14,8,2)	(14,8,2)	(14,8,2)	(16,7,2)
0.3	(14,8,2)	(14,8,2)	(14,8,2)	(16,7,2)	(16,7,2)
0.35	(14,8,2)	(14,8,2)	(14,8,2)	(16,7,2)	(18,6,2)
0.4	(14,8,2)	(14,8,2)	(16,7,2)	(16,7,2)	(18,6,2)
0.45	(14,8,2)	(14,8,2)	(16,7,2)	(18,6,2)	(18,6,2)
0.5	(14,8,2)	(16,7,2)	(16,7,2)	(18,6,2)	(18,6,2)
0.55	(14,8,2)	(16,7,2)	(16,7,2)	(18,6,2)	(18,6,2)
0.6	(14,8,2)	(16,7,2)	(16,7,2)	(18,6,2)	(18,6,2)
0.65	(14,8,2)	(16,7,2)	(16,7,2)	(18,6,2)	(18,6,2)
0.7	(14,8,2)	(16,7,2)	(16,7,2)	(18,6,2)	(18,6,2)
0.75	(14,8,2)	(16,7,2)	(16,7,2)	(18,6,2)	(18,6,2)
0.8	(16,7,2)	(16,7,2)	(16,7,2)	(18,6,2)	(18,6,2)
0.85	(16,7,2)	(16,7,2)	(16,7,2)	(16,7,2)	(18,6,2)
0.9	(16,7,2)	(16,7,2)	(16,7,2)	(16,7,2)	(16,7,2)
0.95	(15,5,3)	(15,5,3)	(15,5,3)	(16,7,2)	(16,7,2)
0.99	(15,5,3)	(15,5,3)	(16,7,2)	(16,7,2)	(16,7,2)

Table B.4: The LP designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.05, 0.05, 0.15, 0.25, 0.35, 0.45)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 90$.

	0.55	0.65	0.75	0.85	0.95
0.01	(6,4,6)	(12,6,3)	(12,6,3)	(12,6,3)	(16,7,2)
0.05	(12,6,3)	(12,6,3)	(16,7,2)	(18,6,2)	(18,6,2)
0.10	(14,8,2)	(16,7,2)	(18,6,2)	(18,6,2)	(20,5,2)
0.15	(16,7,2)	(18,6,2)	(18,6,2)	(18,6,2)	(22,4,2)
0.20	(16,7,2)	(18,6,2)	(18,6,2)	(20,5,2)	(22,4,2)
0.25	(18,6,2)	(18,6,2)	(18,6,2)	(20,5,2)	(22,4,2)
0.30	(18,6,2)	(18,6,2)	(18,6,2)	(20,5,2)	(22,4,2)
0.35	(18,6,2)	(18,6,2)	(20,5,2)	(20,5,2)	(22,4,2)
0.40	(18,6,2)	(18,6,2)	(20,5,2)	(20,5,2)	(22,4,2)
0.45	(18,6,2)	(18,6,2)	(20,5,2)	(20,5,2)	(22,4,2)
0.50	(18,6,2)	(18,6,2)	(20,5,2)	(20,5,2)	(22,4,2)
0.55	(18,6,2)	(18,6,2)	(20,5,2)	(20,5,2)	(22,4,2)
0.60	(18,6,2)	(20,5,2)	(20,5,2)	(20,5,2)	(22,4,2)
0.65	(18,6,2)	(20,5,2)	(20,5,2)	(20,5,2)	(22,4,2)
0.70	(18,6,2)	(20,5,2)	(20,5,2)	(20,5,2)	(22,4,2)
0.75	(18,6,2)	(18,6,2)	(20,5,2)	(20,5,2)	(22,4,2)
0.80	(18,6,2)	(18,6,2)	(20,5,2)	(20,5,2)	(22,4,2)
0.85	(18,6,2)	(18,6,2)	(20,5,2)	(20,5,2)	(22,4,2)
0.90	(18,6,2)	(18,6,2)	(20,5,2)	(20,5,2)	(22,4,2)
0.95	(18,6,2)	(18,6,2)	(18,6,2)	(20,5,2)	(22,4,2)
0.99	(16,7,2)	(18,6,2)	(18,6,2)	(20,5,2)	(24,3,2)

B.3 Tables with optimal designs for $m = 3$ leveraged plan B

Table B.5: The LP B designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.05, 0.05, 0.15, 0.25, 0.35, 0.45)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 60$.

	0.05	0.15	0.25	0.35	0.45
0.01	(3,1,17)	(3,1,17)	(3,1,17)	(4,1,16)	(4,1,16)
0.05	(8,2,6)	(8,2,6)	(10,2,5)	(10,2,5)	(10,2,5)
0.10	(12,2,4)	(12,2,4)	(12,2,4)	(12,4,2)	(12,4,2)
0.15	(12,4,2)	(12,4,2)	(14,3,2)	(14,3,2)	(14,3,2)
0.20	(14,3,2)	(14,3,2)	(14,3,2)	(16,2,2)	(16,2,2)
0.25	(14,3,2)	(14,3,2)	(16,2,2)	(16,2,2)	(16,2,2)
0.30	(16,2,2)	(16,2,2)	(16,2,2)	(16,2,2)	(16,2,2)
0.35	(16,2,2)	(16,2,2)	(16,2,2)	(16,2,2)	(16,2,2)
0.40	(16,2,2)	(16,2,2)	(16,2,2)	(16,2,2)	(18,1,2)
0.45	(16,2,2)	(16,2,2)	(16,2,2)	(18,1,2)	(18,1,2)
0.50	(16,2,2)	(16,2,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.55	(16,2,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.60	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.65	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.70	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.75	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.80	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.85	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.90	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.95	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.99	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)

Table B.6: The LP designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.55, 0.65, 0.75, 0.85, 0.95)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 60$.

	0.55	0.65	0.75	0.85	0.95
0.01	(4,1,16)	(4,1,16)	(4,2,8)	(10,2,5)	(14,3,2)
0.05	(12,2,4)	(12,2,4)	(14,3,2)	(16,2,2)	(18,1,2)
0.10	(14,3,2)	(14,3,2)	(16,2,2)	(16,2,2)	(18,1,2)
0.15	(16,2,2)	(16,2,2)	(16,2,2)	(18,1,2)	(18,1,2)
0.20	(16,2,2)	(16,2,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.25	(16,2,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.30	(16,2,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.35	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.40	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.45	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.50	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.55	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.60	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.65	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.70	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.75	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.80	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.85	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.90	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.95	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)
0.99	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)	(18,1,2)

Table B.7: The LP B designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.05, 0.05, 0.15, 0.25, 0.35, 0.45)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 90$.

	0.05	0.15	0.25	0.35	0.45
0.01	(4,1,26)	(4,1,26)	(4,1,26)	(4,1,26)	(4,1,26)
0.05	(12,3,6)	(12,3,6)	(14,4,4)	(14,4,4)	(14,4,4)
0.10	(14,4,4)	(18,4,3)	(18,4,3)	(18,4,3)	(18,4,3)
0.15	(18,4,3)	(18,4,3)	(20,5,2)	(22,4,2)	(22,4,2)
0.20	(20,5,2)	(22,4,2)	(22,4,2)	(22,4,2)	(22,4,2)
0.25	(22,4,2)	(22,4,2)	(22,4,2)	(22,4,2)	(24,3,2)
0.30	(22,4,2)	(22,4,2)	(22,4,2)	(24,3,2)	(26,2,2)
0.35	(22,4,2)	(22,4,2)	(24,3,2)	(26,2,2)	(26,2,2)
0.40	(22,4,2)	(24,3,2)	(26,2,2)	(26,2,2)	(26,2,2)
0.45	(24,3,2)	(26,2,2)	(26,2,2)	(26,2,2)	(28,1,2)
0.50	(26,2,2)	(26,2,2)	(26,2,2)	(28,1,2)	(28,1,2)
0.55	(26,2,2)	(26,2,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.60	(26,2,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.65	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.70	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.75	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.80	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.85	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.90	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.95	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.99	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)

Table B.8: The LP B designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.55, 0.65, 0.75, 0.85, 0.95)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 90$.

	0.55	0.65	0.75	0.85	0.95
0.01	(4,2,13)	(8,2,11)	(10,2,10)	(14,4,4)	(22,4,2)
0.05	(18,4,3)	(18,4,3)	(20,5,2)	(22,4,2)	(28,1,2)
0.10	(22,4,2)	(22,4,2)	(22,4,2)	(26,2,2)	(28,1,2)
0.15	(22,4,2)	(24,3,2)	(26,2,2)	(28,1,2)	(28,1,2)
0.20	(24,3,2)	(26,2,2)	(26,2,2)	(28,1,2)	(28,1,2)
0.25	(26,2,2)	(26,2,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.30	(26,2,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.35	(26,2,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.40	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.45	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.50	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.55	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.60	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.65	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.70	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.75	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.80	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.85	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.90	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.95	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)
0.99	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)	(28,1,2)

Table B.9: The LP B designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.05, 0.05, 0.15, 0.25, 0.35, 0.45)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 150$.

	0.05	0.15	0.25	0.35	0.45
0.01	(4,1,46)	(4,2,23)	(8,2,21)	(10,2,20)	(10,2,20)
0.05	(22,4,7)	(22,4,7)	(22,4,7)	(25,5,5)	(26,6,4)
0.10	(26,6,4)	(26,6,4)	(29,7,3)	(32,6,3)	(32,6,3)
0.15	(32,6,3)	(32,6,3)	(34,8,2)	(34,8,2)	(36,7,2)
0.20	(34,8,2)	(34,8,2)	(36,7,2)	(38,6,2)	(38,6,2)
0.25	(34,8,2)	(36,7,2)	(38,6,2)	(38,6,2)	(40,5,2)
0.30	(38,6,2)	(38,6,2)	(38,6,2)	(42,4,2)	(42,4,2)
0.35	(38,6,2)	(38,6,2)	(42,4,2)	(42,4,2)	(42,4,2)
0.40	(38,6,2)	(42,4,2)	(42,4,2)	(42,4,2)	(46,2,2)
0.45	(42,4,2)	(42,4,2)	(42,4,2)	(46,2,2)	(46,2,2)
0.50	(42,4,2)	(42,4,2)	(46,2,2)	(46,2,2)	(48,1,2)
0.55	(42,4,2)	(46,2,2)	(46,2,2)	(46,2,2)	(48,1,2)
0.60	(46,2,2)	(46,2,2)	(46,2,2)	(48,1,2)	(48,1,2)
0.65	(46,2,2)	(46,2,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.70	(46,2,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.75	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.80	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.85	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.90	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.95	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.99	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)

Table B.10: The LP B designs (b,k,n) with the smallest asymptotic standard deviation of γ given $\lambda = (0.55, 0.65, 0.75, 0.85, 0.95)$, $\eta \in (0, 1)$ when $m = 3$ and $N = 150$.

	0.55	0.65	0.75	0.85	0.95
0.01	(14,2,18)	(14,3,12)	(18,4,8)	(25,5,5)	(34,8,2)
0.05	(26,6,4)	(32,6,3)	(34,8,2)	(38,6,2)	(46,2,2)
0.10	(34,8,2)	(38,6,2)	(38,6,2)	(42,4,2)	(48,1,2)
0.15	(38,6,2)	(40,5,2)	(42,4,2)	(46,2,2)	(48,1,2)
0.20	(40,5,2)	(42,4,2)	(46,2,2)	(48,1,2)	(48,1,2)
0.25	(42,4,2)	(46,2,2)	(46,2,2)	(48,1,2)	(48,1,2)
0.30	(44,3,2)	(46,2,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.35	(46,2,2)	(46,2,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.40	(46,2,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.45	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.50	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.55	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.60	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.65	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.70	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.75	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.80	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.85	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.90	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.95	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)
0.99	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)	(48,1,2)

Appendix C

Extended Rank 1 Update

C.1 Rank 1 Update

Theorem 2. *If $W = C + \underline{w}\underline{w}^t$ where C is non-singular then*

1. $det(W) = det(C)\left[1 + \underline{w}^t C^{-1} \underline{w}\right]$
2. $W^{-1} = \left[I - \frac{1}{1+\underline{w}^t C^{-1} \underline{w}} C^{-1} \underline{w}\underline{w}^t\right] C^{-1}$

Corollary 3. *If $W = I + \underline{u}\underline{u}^t$ then*

1. $det(W) = I + \underline{u}^t \underline{u}$
2. $W^{-1} = I - \frac{1}{1+\underline{u}^t \underline{u}} \underline{u}\underline{u}^t$

Corollary 4. *If $W = aI_n + bJ_n$ then*

1. $det(W) = a^{n-1}\left[a + bn\right]$
2. $W^{-1} = \frac{1}{a} I_n - \frac{b}{a(a+bn)} J_n$

C.2 Properties of Fisher information

We will need to use the following two results about block matrix inversion established by Banachiewicz [1937].

Banachiewicz [1937] established the following two results about block matrix inversion.

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix}^{-1}$$

$$\det(\mathbf{A}) = \det \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \det(\mathbf{A}) \ \det\left(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)$$

We define

$$F_0 = \begin{pmatrix} \mathbf{M} & \mathbf{0}_{m\times 2} \\ \mathbf{0}_{m\times 2}^t & P \end{pmatrix} \quad \text{and} \quad F_1 = \begin{pmatrix} \mathbf{M} & s\mathbf{v} \\ s\mathbf{v}^t & r \end{pmatrix}$$

$$F_0^{-1} = \begin{pmatrix} \mathbf{M}^{-1} & \mathbf{0}_{m\times 2} \\ \mathbf{0}_{m\times 2}^t & \mathbf{P}^{-1} \end{pmatrix}$$

and

$$F_1^{-1} = \begin{pmatrix} \mathbf{M}^{-1} + \mathbf{M}^{-1}\mathbf{v}(r - \mathbf{v}^t\mathbf{A}^{-1}\mathbf{v})^{-1}\mathbf{v}^t\mathbf{M}^{-1} & -\mathbf{M}^{-1}\mathbf{v}(\mathbf{r} - \mathbf{v}^t\mathbf{M}^{-1}\mathbf{v})^{-1} \\ -(\mathbf{r} - \mathbf{v}^t\mathbf{M}^{-1}\mathbf{v})^{-1}\mathbf{v}^t\mathbf{M}^{-1} & (r - \mathbf{v}^t\mathbf{M}^{-1}\mathbf{v})^{-1} \end{pmatrix}$$

$\mathbf{M}$ is positive definite which implies $\mathbf{M}^{-1}$. For a positive definite matrix P of dimension q has the property that $\mathbf{x}^t\mathbf{P}\mathbf{x} > 0$ for any non-vector $\mathbf{x} \in \mathbb{R}^q$. Thus, $\mathbf{v}^t\mathbf{M}^{-1}\mathbf{v} > 0$. The asymptotic variance will be decreased if we we can set $\mathbf{v} = 0$. If

$\mathbf{v} \neq 0$ we can quantify the loss of efficiency, which is the ratio F_0/F_1, by

$$\frac{1}{1 - \frac{1}{r}\mathbf{v}^t\mathbf{M}^{-1}\mathbf{v}}$$

Appendix D

Assessment of a Measurement System with Production Data

D.1 Truncated Normal

The moment generating function (MGF) for a $X \sim N(\mu, \sigma^2)$ [See Fisher, 1931] that is truncated such that $X \in [b_1, b_2]$, where $b_2 > b_1$ (denoted as truncated $N(\mu, \sigma^2; b_1, b_2)$) is

$$M(t; b_1, b_2) = e^{\mu t + \sigma^2 t^2/2} \frac{\left[\Phi\left(\frac{b_2 - \mu - \sigma^2 t}{\sigma}\right) - \Phi\left(\frac{b_1 - \mu - \sigma^2 t}{\sigma}\right)\right]}{\left[\Phi\left(\frac{b_2 - \mu}{\sigma}\right) - \Phi\left(\frac{b_1 - \mu}{\sigma}\right)\right]} \tag{D.1}$$

where $\Phi(x)$ is the standard normal cumulative distribution function. Using the MGF the first four moments are

$$\begin{aligned}
E(X) &= \mu - \lambda_0\sigma \\
E(X^2) &= \mu^2 - 2\lambda_0\sigma\mu + (1 - \lambda_1)\sigma^2 \\
E(X^3) &= \mu^3 - 3\lambda_0\sigma\mu^2 + (3 - 3\lambda_1)\sigma^2\mu + (-2\lambda_0 - \lambda_2)\sigma^3 \\
E(X^4) &= \mu^4 - 4\lambda_0\sigma\mu^3 + (-6\lambda_1 + 6)\sigma^2\mu^2 + (-8\lambda_0 - 4\lambda_2)\sigma^3\mu + (-3\lambda_1 + 3 - \lambda_3)\sigma^4
\end{aligned}$$

where $\alpha_1 = \frac{b_1-\mu}{\sigma_t}$, $\alpha_2 = \frac{b_2-\mu}{\sigma_t}$, for $i = 0, 1, 2, 3$, $\lambda_i = \frac{\alpha_2^i\phi(\alpha_2)-\alpha_1^i\phi(\alpha_1)}{\Phi(\alpha_2)-\Phi(\alpha_1)}$ and $\phi(x)$ is the standard normal probability density function.

If Y is truncated $N(\mu, \sigma^2)$, such that $Y \notin A = (a_1, a_2)$, where $a_2 > a_1$ then we can write $Y = uX_1 + (1-u)X_2$ where $X_1 \sim$ truncated $N(\mu, \sigma^2; -\infty, a_1)$, $X_2 \sim$ truncated $N(\mu, \sigma^2; a_2, \infty)$ and $u = \left[\Phi\left(\frac{a_1-\mu}{\sigma}\right)\right] / \left[1 - \Phi\left(\frac{a_2-\mu}{\sigma}\right) + \Phi\left(\frac{a_1-\mu}{\sigma}\right)\right]$ then the moment generating function for Y is

$$M_Y(t) = uM(t; -\infty, a_1) + (1-u)M(t; a_2, \infty) \tag{D.2}$$

where $M(t; b_1, b_2)$ is given in (D.1). Thus Y has the same moments as X with the exception that for $i = 0, 1, 2, 3$ λ_i is replaced with

$$\beta_i = u\lambda_i \left(-\infty, \frac{a_1-\mu}{\sigma}\right) + (1-u)\ \lambda_i \left(\frac{a_2-\mu}{\sigma}, \infty\right) \tag{D.3}$$

$$\text{where} \quad \lambda_i(z_1, z_2) = \frac{z_2^i\phi(z_2) - z_1^i\phi(z_1)}{\Phi(z_2) - \Phi(z_1)} \tag{D.4}$$

D.2 Normal Moments

If $X \sim N(\mu, \sigma^2)$ then the first four moments [See Johnson and Kotz, 1970] are

$$\begin{aligned} E(X) &= \mu \\ E(X^2) &= \mu^2 + \sigma^2 \\ E(X^3) &= \mu^3 + 3\mu\sigma^2 \\ E(X^4) &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 \end{aligned}$$

D.3 Expectation of S_m^2

We define

$$S_m^2 = \frac{1}{n_2}\sum_{i\in S}\sum_{j=1}^{2}\left(Y_{ij}-\overline{Y}_{i.}\right)^2 = \frac{1}{n_2}\sum_{i\in S}\frac{1}{2}\left(Y_{i1}-Y_{i2}\right)^2 = \frac{1}{n_2}\sum_{i\in S}S_{im}^2. \tag{D.5}$$

Each pair of measurements from different parts are independent so we need to determine the properties of $\frac{1}{2}\left(Y_{i1}-Y_{i2}\right)^2$. To simplify the calculations, we notice that we can define $Y_{ij} = X_{ij} + \mu$ where X_{ij} has the same distribution as Y_{ij} but with paramter $\mu = 0$. Then

$$\begin{aligned}
E\left[S_{im}\right] &= \frac{1}{2}E\left[\left(Y_{i1}-Y_{i2}\right)^2\right] \\
&= \frac{1}{2}E\left[\left(X_{i1}+\mu-X_{i2}-\mu\right)^2\right] \\
&= \frac{1}{2}E\left[\left(X_{i1}-X_{i2}\right)^2\right] \\
&= \frac{1}{2}E\left[E\left[X_{i1}^2-2X_{i1}X_{i2}+X_{i2}^2\middle|X_{i1}\right]\right] \\
&= \frac{1}{2}E\left[X_{i1}^2-2X_{i1}E\left[X_{i2}\middle|X_{i1}\right]+E\left[X_{i2}^2\middle|X_{i1}\right]\right] \\
&= \frac{1}{2}E\left[X_{i1}^2-2X_{i1}\left[\rho X_{i1}\right]+\left[\rho X_{i1}\right]^2+\sigma_t^2\left(1-\rho\right)\left(1+\rho\right)\right] \\
&= \left[-\frac{1}{2}+1-\rho+\frac{1}{2}\rho^2\right]E\left[X_{i1}^2\right]+\frac{1}{2}\sigma^2\left(1-\rho\right)\left(1+\rho\right) \\
&= \frac{1}{2}\left[1-\rho\right]^2\left[\left(1-\beta_1\right)\sigma^2\right]+\frac{1}{2}\sigma^2\left(1-\rho\right)\left(1+\rho\right) \\
&= \sigma^2\left(1-\rho\right)\left[1-\left(1-\rho\right)\beta_1/2\right]
\end{aligned}$$

Thus, the expectation of S_{im}^2 is $\sigma^2\left(1-\rho\right)\left[1-\left(1-\rho\right)\beta_1/2\right]$.

D.4 Variance of S_m^2

To calculate the variance of S_m we use S_{im}, defined in (D.5), to get

$$\mathrm{Var}(S_m^2) = \left(\frac{1}{n_2}\right)^2 \sum_{i \in S} \mathrm{Var}(S_{im}^2) = \frac{1}{n_2}\mathrm{Var}(S_{im}^2) \tag{D.6}$$

because the measurements on different parts are independent. If we use the identity

$$\mathrm{Var}(S_{im}^2) = E\left[S_{im}^4\right] - E\left[S_{im}^2\right]^2 \tag{D.7}$$

we only need to calculate $E\left[S_{im}^4\right]$ to determine the variance.

$$\begin{aligned}
E\left[S_{im}^4\right] &= E\left[\left(\frac{1}{2}\left(Y_{i1} - Y_{i2}\right)^2\right)^2\right] \\
&= \frac{1}{4}E\left[X_{i1}^4 - 4X_{i1}^3X_{i2} + 6X_{i1}^2X_{i2}^2 - 4X_{i1}X_{i2}^3 + X_{i2}^4\right] \\
&= \frac{1}{4}E\left\{E\left[X_{i1}^4 - 4X_{i1}^3X_{i2} + 6X_{i1}^2X_{i2}^2 - 4X_{i1}X_{i2}^3 + X_{i2}^4 \,|X_{i1}\right]\right\} \\
&= \frac{1}{4}E\left\{X_{i1}^4 - 4X_{i1}^3E\left[X_{i2}\,|X_{i1}\right] + 6X_{i1}^2E\left[X_{i2}^2\,|X_{i1}\right]\right. \\
&\qquad\qquad \left. -4X_{i1}E\left[X_{i2}^3\,|X_{i1}\right] + E\left[X_{i2}^4\,|X_{i1}\right]\right\} \\
&\qquad \vdots \\
&= \frac{1}{4}\left[-3 + 4(1-\rho) - 4\rho^3 + \rho^4 + 6\rho^2\right]E\left[X_{i1}^4\right] \\
&\quad + \frac{1}{4}\left[6\sigma^2 + 6\sigma^2\rho^2 - 12\sigma^2\rho\right](1-\rho)(1+\rho)E\left[X_{i1}^2\right] \\
&\quad + \frac{1}{4}3\sigma^4(1-\rho)^2(1+\rho)^2 \\
&\qquad \vdots \\
&= \frac{1}{4}\sigma^4(1-\rho)^2\left(3(1-\rho)^2\beta_1 - (1-\rho)^2\beta_3 - 12\beta_1(1-\rho) + 12\right) \\
&= \frac{1}{4}\sigma^4(1-\rho)^2\left[(1-\rho)^2\left(3\beta_1 - \beta_3\right) - 12\beta_1(1-\rho) + 12\right]
\end{aligned}$$

Now, combing this result with the previous formula we have

$$\begin{aligned}
& \mathrm{Var}(S_{im}^2) \\
= \; & E\left[S_{im}^4\right] - E\left[S_{im}^2\right]^2 \\
= \; & \frac{1}{4}\sigma^4(1-\rho)^2\left[(1-\rho)^2(3\beta_1-\beta_3) - 12\beta_1(1-\rho) + 12\right] \\
& \qquad\qquad - \left[\sigma^2(1-\rho)\left[1-(1-\rho)\frac{\beta_1}{2}\right]\right]^2 \\
& \quad \vdots \\
= \; & \frac{1}{4}\sigma^4(1-\rho)^2\left[3(1-\rho)^2\beta_1 - (1-\rho)^2\beta_3 - 8\beta_1(1-\rho) - (1-\rho)^2\beta_1^2\right] \\
= \; & 2\sigma^4(1-\rho)^2\left(1-\beta_1(1-\rho) - \frac{(1-\rho)^2}{8}\left(\beta_3 - 3\beta_1 + \beta_1^2\right)\right)
\end{aligned}$$

and thus,

$$\mathrm{Var}(S_m^2) = \frac{2\sigma^4(1-\rho)^2}{n_2}\left(1-\beta_1(1-\rho) - \frac{(1-\rho)^2}{8}\left(\beta_3 - 3\beta_1 + \beta_1^2\right)\right) \qquad \text{(D.8)}$$

References

Christopher I. Amos. Robust variance-components approach for assessing genetic linkage in pedigrees. *American Journal of Human Genetics*, 54:535–543, 1994. 167

Automotive Industry Action Group. *Measurement Systems Analysis*. Southfield MI., third edition, 2002. 5, 9, 10, 12, 14, 15, 25, 32, 33, 41, 42, 77, 83, 87, 136, 150

T. Banachiewicz. Zur berechnung der determinanten, wie auch der inversen, und zur darauf basierten auflosung der systeme linearer gleichungen. *Acta Astronomica*, Serie C:41–67, 1937. 177, 194

R. P. Browne, R. J. Mackay, and S. H. Steiner. Leveraged measurement system assessment. *Submitted to Technometrics*, 2007. 89

R. K. Burdick, C. M. Borror, and D. C. Montgomery. A review of methods for measurement systems capability analysis. *J. Qual. Technol.*, 35:342–354, 2003. 12, 16, 17, 32, 33, 34, 42, 150, 152, 153, 168

R. K. Burdick, C. M. Borror, and D. C. Montgomery. *Design and Analysis of Gauge R&R Studies: Making Decisions with Confidence Intervals in Random and Mixed ANOVA*. ASA-SIAM Series on Statistics and Applied Probability, 2005. 9, 18, 25, 34, 43, 167, 168

D.R. Cox and D.V. Hinkley. *Theoretical Statistics*. Chapman and Hall, 1974. 109

W.R. Dillon and M. Goldstein. *Multivariate analysis methods and applications*. John Wiley & Sons, 1984. 58, 121, 140

N. Doganaksoy. Assessment of impact of measurement variability in the presence of multiple sources of product variability. *Quality Engineering*, 13:83–89, 2000. 16, 17, 152

A. Donner and M. Eliasziw. Sample size requirements for reliability studies. *Statistics in Medicine*, 6:441–448, 1987. 12, 32, 33, 34

Douglas S. Falconer and Trudy F.C. Mackay. *Introduction to Quantitative Genetics*. Benjamin Cummings; 4 edition, 1994. 167

R. A. Fisher. The truncated normal distribution. *British Association for Advancement of Science*, 5:xxxiii–xxxiv, 1931. 197

Norman Johnson and Samuel Kotz. *Distributions in Statistics: Continuous Univariate Distributions*. Houghton Mifflin, 1970. 198

B. Jorgensen and S. J. Knudsen. Parameter orthogonality and bias adjustment for estimating functions. *Scandinavian Journal of Statistics*, 31:93–114, 2004. 68, 157

Yock Wah Lai and Ek Peng Chew. Gauge capability assessment for high-yeild manufacturing process with truncated distribution. *Quality Engineering*, 13:203–210, 2000. 167

D.P. Larsen, J. Prins, and R.E. Lampe. The economic impact of measurement error. *Quality Engineering*, 11:563–574, 1999. 10, 16, 17, 21, 22

G.A. Larsen. Measurement system analysis - the usual metric can be noninformative. *Quality Engineering*, 15:293–298, 2002. 12

Miao-Xin Li, Peng-Yuan Liu, Yu-Mei Li, Yue-Juan Qin, Yao-Zhong Liu, and Hong-Wen Deng. A major gene model of adult height is suggested in chinese. *Journal of Human Genetics*, 49:148–153, March 2004. 167

D. C. Montgomery and G.C. Runger. Gauge capabilty analysis and designed experinments. part i: Basic methods. *Quality Engineering*, 6:115–135, 1993a. 10, 25

Douglas C. Montgomery, Elizabeth A. Peck, and Geoffrey G. Vining. *Introduction to Linear Regression Analysis*. John Wiley & Sons, 2001. 49, 63

P. E. Shrout and J. L. Fleiss. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86:420–428, 1979. 1

S. H. Steiner and R. J. Mackay. *Statistical Engineering: An Algorithm for Reducing Variation in Manufacturing Processes*. ASQ Quality Press, 2005. 12, 24, 69

S. H. Steiner, R. J. MacKay, and J. S. Ramberg. An overview of the shainin system for quality improvement. *Quality Engineering (with discussion)*, 20:6–45, 2008. 47

Shannon Sweeney. Analysis of two-dimensional gage repeatability and reproducibility. *Quality Engineering*, 19:29–37, 2007. 171, 173, 174

J. G. Voelkel. Gauge r&r analysis for two-dimensional data with circular tolerances. *Journal of Quality Technology*, 35:153–167, 2003. 173

D. J. Wheeler and R.W. Lyday. *Evaluating the Measurement Process*. SPC Press, Inc., 1984. 7, 10, 12, 15, 25, 32

W. H. Woodall and C. M. Borror. A review of methods for measurement systems capability analysis. *Quality and Reliability Engineering International*, 2007. 14, 15