BAYESIAN STATISTICAL PROCESS

ADJUSTMENT FOR UNKNOWN PARAMETER SYSTEMS

A Thesis in
Industrial Engineering and Operations Research

by
Zilong Lian

©2005 Zilong Lian

Submitted in Partial Fulfillment
of the Requirements
for the Degree of
Doctor of Philosophy

May 2005
The thesis of Zilong Lian has been reviewed and approved* by the following:

Enrique del Castillo
Professor of Industrial Engineering and Statistics
Thesis Adviser
Chair of Committee

M. Jeya Chandra
Professor of Industrial Engineering

Natarajan Gautam
Associate Professor of Industrial Engineering

Duncan Fong
Professor of Marketing and Statistics

Richard J. Koubek
Professor of Industrial Engineering
Head of the Department of Industrial & Manufacturing Engineering

*Signatures are on file in the Graduate School.
Abstract

A central problem in manufacturing is how to adjust a production process which is not performing as desired. Interventions or adjustments need to be performed to bring the process back to a desired or target performance. For many years, statistical techniques have been used in industry for this purpose, mainly to monitor and analyze process variation and maintain process consistency hence leading to better product quality. In the last decade, advancements in computer power and statistical methodology have taken place that can help manufacturers in facing new challenges with respect to product quality. In this dissertation, some recently developed Bayesian statistical methodologies are introduced for the first time in the process control field. The goal is to control or adjust a manufacturing process in order to improve the quality of the manufactured products.

Traditional statistical process adjustment methods used in manufacturing are based on the assumption that system parameters (e.g., the different variance components that make up the total variability of the measurements) are known. In practice, it is necessary to estimate these parameters before a statistical process adjustment policy can be put into place. If the sample size of the data used to get estimates is limited, parameter estimates
can be subject to substantial sampling error. Alternatively, biased estimates may also be obtained if the conditions are different between previous runs and the conditions over which the process will operate. This can lead to a poor performance of the adjustment method if the assumption of known parameters is made.

This research is focused on developing adjustment rules under unknown parameters for different process models and cost structures. The process models under study can all be characterized as univariate linear gaussian models. The process can be described by linear equations containing additive error terms that follow a normal distribution. The variances of these terms are assumed unknown. Such mathematical models are very useful in many discrete-part manufacturing industries, from metal-cutting to nano-scale manufacturing. The costs considered in this research include symmetric and asymmetric off-target costs and fixed adjustment costs. These different process models and cost structures have been studied in the literature on process adjustment under known parameters. The importance of considering unknown parameters has not been fully recognized and the related problems have not been studied.

In this dissertation, the mathematical derivation of the proposed adjustment methods is based on Bayesian statistics, an important branch of Statistics that has received renewed interest in recent years. The Bayesian approach is carried out through two computational techniques, Markov Chain Monte Carlo (MCMC) methods and Sequential Monte Carlo (SMC) methods.

This research provides fundamental new results in the area of statistical process control,
which are likely to impact the practice of this field.
# Contents

List of Figures xi  
List of Tables xvi  
List of Acronyms xviii  
Acknowledgements xix  

1 Introduction and Research Objectives 1  

1.1 Research Topics .............................................................. 3  
   1.1.1 Setup Adjustment Problems ........................................... 3  
   1.1.2 Process Control under a Fixed Adjustment Cost ..................... 3  
   1.1.3 Asymmetric Off-target Cost Functions .............................. 4  
   1.1.4 Bayesian Models .......................................................... 4  
   1.1.5 Computation of the Posterior Distribution .......................... 5  

1.2 Research Objectives .......................................................... 5
1.2.1 Setup Adjustment of Multiple Lots using a Markov Chain Monte Carlo Method ......................................................... 5
1.2.2 Setup Adjustment of Multiple Lots using a Sequential Monte Carlo Method .......................................................... 6
1.2.3 Setup Adjustment Under Fixed Adjustment Cost ......................... 6
1.2.4 Setup Adjustment for Asymmetric Off-target Cost ..................... 7
1.2.5 Adaptive Deadband Control of a Process with Random Drift in the Presence of a Fixed Adjustment Cost ..................... 7
1.3 Dissertation Outline ................................................................. 8

2 Literature Review ................................................................. 10
2.1 Setup adjustment problems ................................................. 10
2.2 Process Control under Fixed Adjustment Costs ....................... 14
2.3 Asymmetric Cost Functions .................................................. 15
2.4 Bayesian Models ................................................................. 16
2.5 Computation of the Posterior Distribution .............................. 18

3 Setup Adjustment of Multiple Lots using a Markov Chain Monte Carlo Method ................................................................. 20
3.1 Introduction ........................................................................... 20
3.2 Rules for adjusting an initial offset over a set of batches when parameters are unknown ............................................... 21
3.3 Evaluation of different adjustment rules with reference to different performance indexes .................................................. 25
3.4 Motivation for a modified rule .............................................. 32
3.5 Conditional First Adjustment Rule based on MCMC Approach .................................... 36
3.6 Conclusions ................................................................. 42
Appendix of Chapter 3: A Bayesian adjustment rule based on MCMC simulation 44

4 Setup Adjustment of Multiple Lots using a Sequential Monte Carlo Method 47
4.1 Process models and existing adjustment procedures ......................... 48
4.2 Sequential Monte Carlo methods ........................................ 51
4.3 Comparison with Grubbs’ Adjustment Rules .................................. 58
  4.3.1 Comparison with Grubbs’ Harmonic Rule ............................. 58
  4.3.2 Comparison with Grubbs’ Extended Rule ............................. 62
4.4 An Improved SMC method .................................................. 65
4.5 Conclusions ................................................................. 68
Appendix of Chapter 4: Markov Chain Monte Carlo And Sequential Monte Carlo
  methods ................................................................. 71

5 Setup Adjustment Under Fixed Adjustment Cost 79
5.1 Introduction ................................................................. 79
5.2 Process Model, Cost Criterion and Parameter Estimates ..................... 80
5.3 Dynamic programming solution to the setup adjustment problem ........ 83
5.3.1 Deadband structure of the optimal adjustment policy 86

5.4 Computer implementation of the solution 88
5.4.1 Generating the control table 88
5.4.2 Adjusting a Process with the Control Table 92

5.5 Sensitivity Analysis and Extremal Situations 95
5.5.1 System Parameters: $\theta_0, \sigma_\nu$ 96
5.5.2 Changes in initial prior state variables: $\mu_0, \sigma_0, \kappa_0, \nu_0$ 98
5.5.3 Fixed Adjustment Cost: $c$ 99

5.6 A simpler case: $\sigma_\nu^2$ known 100
5.6.1 Dynamic Programming 101
5.6.2 Computer Implementation 103

5.7 Comparison of known vs. unknown $\sigma_\nu^2$ approaches 105

Appendix of Chapter 5: Mathematical Derivation 108

6 Setup Adjustment for Asymmetric Off-target Cost 113
6.1 Introduction 113
6.2 Process model, Cost Function and Parameter Estimation 114
6.3 Optimal Adjustment 118
6.4 Simulation Study 122
6.5 Conclusions 126

Appendix of Chapter 6: Solution to Equation $\partial[E(C_i)]/\partial\mu = 0$ 128
7 Adaptive Deadband Control of a Process with Random Drift in the Presence of a Fixed Adjustment Cost

7.1 Process Model and Estimation ........................................... 131
7.2 Fixed Adjustment Cost and Adaptive Deadband Adjustment Method .... 138
7.3 Simulation Study .......................................................... 142
7.4 Conclusion ............................................................... 145

Appendix of Chapter 7: Algorithm for estimating the process parameters .... 146

8 Research Contributions and Further Research .......................... 148

8.1 Research Contributions .................................................. 148
8.1.1 Contributions to Statistical Process Control ........................... 148
8.1.2 Contributions to Applications of Bayesian Statistics .................... 152
8.2 Further Research ......................................................... 153

Bibliography .......................................................................... 156

Appendix A: List of Computer Programs used in Chapter 4 .................. 164

Appendix B: List of Computer Programs used in Chapter 5 .................. 168

Appendix C: List of Computer Programs used in Chapter 6 ................. 175

Appendix D: List of Computer Programs used in Chapter 7 .................. 178
List of Figures

2.1 Scheme of the air bending process and of springback phenomenon, where $\alpha$ represents the bent angle. .......................................................... 11

3.1 Boxplots of the percentage savings (using the quadratic cost function) induced by the MCMC adjustment rule over Grubbs’ harmonic rule and the two EWMA controllers ($\lambda = 0.1$ and $\lambda = 0.4$) as a function of different parameters: the number of lots ($I$); the number of parts in lots ($J$); the ratio of the mean offset to the within-lot standard deviation ($\mu/\sigma_v$); the ratio of the between to the within lot standard deviation ($\sigma_\theta/\sigma_v$). .................................. 27

3.2 Boxplots on the percentage savings (using the quadratic bias cost function) induced by the MCMC adjustment rule over Grubbs’ harmonic rule and the two EWMA controllers ($\lambda = 0.1$ and $\lambda = 0.4$) as a function of different parameters: the number of lots ($I$), the number of parts in lots ($J$), the ratio of the mean offset to the within-lot standard deviation ($\mu/\sigma_v$), and the ratio of the between to the within-lot standard deviation ($\sigma_\theta/\sigma_v$). ................. 31
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3</td>
<td>Quadratic costs in two simulation runs (a: case 16 and b: case 9) as a function of the number of parts processed in the lot.</td>
</tr>
<tr>
<td>3.4</td>
<td>Plot of the interval $\hat{\theta}<em>i' \pm \hat{\sigma}</em>{\theta_i'}$, as a function of lots processed.</td>
</tr>
<tr>
<td>3.5</td>
<td>The maximum savings vs. the minimum savings among the 16 cases on the Tables as a function of the adjustment limit $k$ when applying the Conditional First Adjustment MCMC rule.</td>
</tr>
<tr>
<td>3.6</td>
<td>Boxplots on the percentage savings (using the quadratic cost function) induced by the Conditional First Adjustment MCMC policy over Grubbs’ harmonic rule and the two EWMA controllers ($\lambda = 0.1$ and $\lambda = 0.4$) as a function of different parameters: the number of lots ($I$), the number of parts in lots ($J$), the ratio of the mean offset to the within-lot standard deviation ($\mu/\sigma_v$), and the ratio of the between to the within-lot standard deviations ($\sigma_\theta/\sigma_v$).</td>
</tr>
<tr>
<td>3.7</td>
<td>Boxplots on the percentage savings (using the quadratic bias cost function) induced by the Conditional First Adjustment MCMC policy over Grubbs’ harmonic rule and the two EWMA controllers ($\lambda = 0.1$ and $\lambda = 0.4$) as a function of different parameters: the number of lots ($I$), the number of parts in lots ($J$), the ratio of the mean offset to the within-lot standard deviation ($\mu/\sigma_v$), and the ratio of the between to the within-lot standard deviations ($\sigma_\theta/\sigma_v$).</td>
</tr>
<tr>
<td>4.1</td>
<td>Adjusted value of the true offset $\theta_{ij}$ under the SMC method and Grubbs’ harmonic rule for parts $j = 1, 2, 3, 4, 5$ and lots $i = 1, 3, 10$.</td>
</tr>
</tbody>
</table>
4.2 Plots of the posterior distributions of $\mu$, $\sigma^2_\theta$ and $\sigma^2_\upsilon$ obtained at the end of the process ................................................................. 57

4.3 Average saving rates induced by SMC method v.s. the harmonic rule for some selected process cases (Table 4.2 and priors (Table 4.3 combinations and their 95% confidence intervals) ............................................................... 61

4.4 Average saving rates induced by the original SMC method v.s. the extended rule using different ratios $k \times \sigma^2_\upsilon/\sigma^2_\theta$ for each process case, and their 95% confidence intervals ......................................................... 64

4.5 Average saving rates induced by the original SMC method (method 1), the modified SMC method without the second type of adjustments (method 2) and the conditional first-adjustment method (method 3) against the harmonic rule for each process case, and their 95% confidence intervals .............................................. 66

4.6 The changing of the maximum and the minimum average savings with different $\alpha$ among 16 process cases for prior setting 2 ......................................................... 67

5.1 Adjustment limits $\alpha_{i,\sigma_i}$ for all parts $i$ and $\sigma_i = 1, 4, 8$ in the numerical example. 91

5.2 Plot of the true means and observations ......................................................... 94

5.3 Plot of $\mu_i$ and corresponding $\alpha_{i,\sigma_i}$ ..................................................... 95

6.1 Plot of targets $\mu^*_i$ for Method I with $b = 0.5$, Method I with $b = 1$ (optimal targets) and Method II with $b = 0.5$. ......................................................... 121
6.2 Plot of targets $\mu^*_i$ for Method I with $b = 2$, Method I with $b = 1$ (optimal targets) and Method II with $b = 2$. .................................................. 122

6.3 The average saving in the total off-target costs from part 2 to part 5 induced by Method I and Method II for each combination of $b$ and $c_1$ and its 95% confidence interval. ................................................................. 124

6.4 The average saving in the total off-target costs from part 2 to part 10 induced by Method I and Method II for each combination of $b$ and $c_1$ and its 95% confidence interval. ................................................................. 125

6.5 The average saving in the total off-target costs from part 2 to part 20 induced by Method I and Method II for each combination of $b$ and $c_1$ and its 95% confidence interval. ................................................................. 126

7.1 Average savings induced by SMC method v.s. the KF method for each combination of a prior scenario and a process case, and their 95% confidence intervals ................................................................. 137

7.2 The control limits/half widths for the (1) Constant Deadband method with the true process parameters (2) Constant Deadband method with the wrong process parameters (3) Adaptive Deadband method with biased prior information ................................................................. 141

7.3 Average savings induced by the Adaptive Deadband method v.s. the Constant Deadband method for each combination of a prior scenario and a process case, and their 95% confidence intervals, when $c = 100$. .................................................. 143
7.4 Average savings induced by the Adaptive Deadband method v.s. the Constant Deadband method for each combination of a prior scenario and a process case, and their 95% confidence intervals, when $c = 16$. . . . . . . . . . . . . . . . . . . . . . . 144
# List of Tables

1.1 Problems considered in this research. ........................................ 8

3.1 Parameters characterizing different scenarios examined. ............. 26

3.2 Set of parameters characterizing cases studied in this chapter. ....... 26

3.3 Confidence intervals (at level 95%) on the median of the percentage savings in total quadratic costs induced by the MCMC method compared to each alternative adjustment rule. ........................................ 29

3.4 Confidence intervals (at level 95%) on the median of the percentage savings in quadratic bias costs induced by the MCMC method. .............. 32

3.5 Mean percentage savings in total quadratic costs induced by the MCMC method in simulated scenarios. .......................................... 33

3.6 Confidence intervals (at level 95%) on the median of the percentage savings in total quadratic costs induced by the Conditional First Adjustment MCMC method. .................................................. 42
3.7 Confidence intervals (at level 95%) on the median of the percentage savings in quadratic bias costs induced by the Conditional First Adjustment MCMC method. ................................................................. 42

4.1 Means and 95% credibility intervals for the posterior distributions of the process parameters after all parts are observed. ................................. 57

4.2 Parameters characterizing the true behavior of the process. ..................... 59

4.3 Scenarios for prior distributions. ......................................................... 60

5.1 Control table showing the adjustment limits \( \alpha_{i,\sigma_i} \) calculated for the example. 90

5.2 Savings made by the adjustment ............................................................. 96

5.3 Summary of the behavior of the adjustment rule ................................. 97

5.4 Savings made by the adjustment when \( \kappa_0 = 0.01 \) and \( \mu_0 = 0 \) ............. 99

5.5 Control table showing the adjustment limits \( \alpha_i \) calculated for the example. 104

5.6 Average savings compared to total losses under no adjustment ............... 106

6.1 Total off-target cost for each method and each value of \( b \) and the saving made by Method I v.s. Method II for each value of \( b \) ......................... 120

7.1 Scenarios with different prior information. ........................................... 136

7.2 Different processes investigated. ......................................................... 136
## List of Commonly Used Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCMC</td>
<td>Markov Chain Monte Carlo</td>
</tr>
<tr>
<td>SMC</td>
<td>Sequential Monte Carlo</td>
</tr>
<tr>
<td>DP</td>
<td>Dynamic Programming</td>
</tr>
<tr>
<td>IMA</td>
<td>Integrated Moving Average</td>
</tr>
<tr>
<td>EWMA</td>
<td>Exponentially Weighted Moving Average</td>
</tr>
<tr>
<td>SIR</td>
<td>Sampling/Importance Resampling</td>
</tr>
<tr>
<td>ESS</td>
<td>Effective Sample Size</td>
</tr>
<tr>
<td>IG</td>
<td>Inverse Gamma</td>
</tr>
<tr>
<td>CDF</td>
<td>Commutative Distribution Function</td>
</tr>
<tr>
<td>KF</td>
<td>Kalman Filter</td>
</tr>
<tr>
<td>SPC</td>
<td>Statistical Process Control</td>
</tr>
<tr>
<td>LSL</td>
<td>Lower Specific Limit</td>
</tr>
<tr>
<td>USL</td>
<td>Upper Specific Limit</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to take this opportunity to express my gratitude and respect to my advisor, Dr. Enrique del Castillo. Without his support or supervision, this thesis would not be accomplished. His commitment and passion in research have been, and will always be inspiring me.

I am deeply grateful to Dr. M. Jeya Chandra, Dr. Natarajan Gautam and Dr. Duncan Fong. I appreciate their time spent on serving in my committee. The advices that they gave on my research and the knowledge they imparted in classes have improved the quality of this thesis.

My special thanks to Dr. Bianca M. Colosimo from Politecnico di Milano in Italy for sharing research results and co-authoring two papers. I learned a lot from her during her short stay in the Penn State.

I thank the student fellows in Industrial Engineering Department, in particular, the Engineering Statistics Lab. I want to give my special thanks to Guillermo Miro Quesada, Ramkumar Rajagopal, Omer Arda Vanli and Eduardo Santiago Duran. I want to thank the staff and faculty in Industrial Engineering Department for the top-quality teaching and
student services. I also want to thank the Pennsylvania State University for the fine study and living environment she offered.

Finally, I want to say thank you to my parents. I will always be indebted to them. There will never be enough thanks comparable to the love, caring and support they give me.
Chapter 1

Introduction and Research Objectives

The behavior of a manufacturing process and the resulting product quality can be summarized by a mathematical model and characterized by a number of parameters. Some of these parameters are uncontrollable, such as the variances characterizing the random deviations from part to part or from time to time, which are usually caused by the intrinsic variation existing in any manufacturing and measurement system or caused by environmental disturbances. Some of them, for instance the process mean, can be controlled as the process operates. In a typical process adjustment method, the controllable parameters will be adjusted to improve the quality characteristics by utilizing the estimates of all the parameters. The improvements or performance of the adjustment method can be measured by a given cost function. One type of cost function is only related to the deviations between the quality characteristics and its target, the so-called off-target cost function. Other cost functions include the adjustment costs in addition to the off-target costs.
Statistical methods have been widely used in process adjustment. Traditionally, these methods are based on the assumption that the uncontrollable parameters are known. These known parameters can be used to obtain estimates of the controllable parameters, with which adjustments can be made. In practice, a pilot study, in which process data is obtained before actual production, is necessary for estimating the parameters so that a statistical process adjustment policy can be applied. If the sample size of the data from the pilot study is small, parameter estimates can be subject to substantial sampling error. Alternatively, biased estimates may also be obtained if the conditions are different between previous runs and the process that needs to be adjusted. The bias in the estimated parameters can lead to a poor performance of the adjustment method if the assumption of known parameters is made.

In modern manufacturing, companies are faced with the need for adjusting their processes without \textit{a priori} knowledge of system parameters. For instance, manufacturers are forced to shorten their process volumes by quickly changing technologies and increasing in this way competitiveness. Updates in production lines are frequently needed to accommodate customized features or newly developed products. The process parameters may also change with time. The cost of a pilot study becomes relatively expensive compared to the cost of a short-run process. In some high-technology industries where a unit of product is expensive, a pilot study is sometimes not feasible, or product is also required to be controlled during the test for economic purposes, i.e., control is needed from the first part produced.

This research focuses on developing adjustment rules for different systems with unknown
parameters considering different cost functions utilizing modern Bayesian statistical techniques. The next section describes the specific problems in process control and the topics in Statistics involved in this research.

1.1 Research Topics

The following is a summary of the adjustment problems addressed in this dissertation and the approaches utilized in their solutions. The research objectives are given in Section 1.2.

1.1.1 Setup Adjustment Problems

Adjusting a discrete-part manufacturing process is frequently necessary when the setup of the machine is improperly done. An incorrect setup can result in an offset in the quality characteristic (usually, some dimension) of the parts produced in the batch of product made subsequently to the setup. The offset cannot be observed directly, and sequential adjustments must take place after fabrication and measurement of each part in a series. This class of adjustment problems receives the name setup adjustment problem in the engineering statistics literature.

1.1.2 Process Control under a Fixed Adjustment Cost

Due to its simple analytical form, a quadratic loss function is frequently used to model the cost associated with the deviations from target in the quality characteristics. Another type of cost of interest in the literature and in practice is associated with each adjustment, the
so-called fixed adjustment cost. This is a cost incurred whenever an adjustment is made, irrespective of the magnitude of the adjustment. In the presence of such adjustment cost, an adjustment should be justified only when the resulting savings in the off-target cost is worth the cost spent on the adjustment.

1.1.3 Asymmetric Off-target Cost Functions

Besides symmetric off-target quadratic losses, asymmetric off-target cost functions are also often of interest in manufacturing practice. This type of cost function is more appropriate when the costs of oversized and undersized quality characteristics are different. The optimal targets of the process are usually located in the “less expensive” side, depending on the cost structures and the distributions of the quality characteristics.

1.1.4 Bayesian Models

In classical process control schemes, process parameters are estimated based on sampling. Parameter estimates are used in process control policies but the the precisions are not considered once the process starts. In this research, where the parameters are assumed unknown, a Bayesian model is used to make statistical inferences on the parameters. Instead of a single estimator, the posterior distribution of a parameter is obtained. This posterior distribution, containing information for both the estimate and its precision, is used to derive optimal process control rules that are sequentially updated as more observations are obtained.
1.1.5 Computation of the Posterior Distribution

In some problems considered in this research, *conjugate prior distributions* are used and hence the computations of the posterior distributions are convenient. However, in other cases, conjugate prior distributions are not available, and this requires a numerical method to carry out the Bayesian computations. Two simulation-based computational methods, *Markov Chain Monte Carlo* (MCMC) methods and *Sequential Monte Carlo* (SMC) methods, are studied and used in the adjustment rules proposed in this dissertation. SMC methods are better suited than MCMC methods for *on-line* process control when the time between observations is short.

1.2 Research Objectives

The overall goal of this research is to develop a series of adjustment methods for different control problems under the assumption that the system parameters are unknown. The specific objectives of this research are to study and solve the following process adjustment problems:

1.2.1 Setup Adjustment of Multiple Lots using a Markov Chain Monte Carlo Method

The first objective of this research is to investigate a setup adjustment problem for a multiple-batch process when no previous knowledge on the distribution of the setup offset or on the
process variability is available. The sensitivity analysis of a previously proposed on-line sequential adjustment procedure [11] will be conducted. This method is based on a hierarchical Bayesian model and uses Markov Chain Monte Carlo (MCMC) to derive estimates used to compute the adjustments at each observation. Based on this method, an improved procedure will be proposed for more robust performance. In this problem, no adjustment cost will be assumed, and only a symmetric quadratic off-target cost will be considered.

1.2.2 Setup Adjustment of Multiple Lots using a Sequential Monte Carlo Method

The same type of problem as in the first objective is again considered. A new adjustment method, which is based on the same Bayesian model yet uses a different posterior distribution computing technique (a SMC method) is proposed to achieve more time-efficient performance and to meet the time demands of on-line adjustment. Due to the reduction in time spent on computation, more simulation results can be obtained for sensitivity analysis.

1.2.3 Setup Adjustment Under Fixed Adjustment Cost

Assuming again no prior knowledge on the process parameters in the setup adjustment problem, the third objective of this research is to use a conjugate Bayesian model for statistical inferences on these parameters and give the optimal solution to a single-lot setup adjustment problem in the presence of the fixed adjustment cost. In this case, a symmetric quadratic off-target cost and a fixed adjustment cost will be joint considered. The optimal adjustment
rule will be proven to be of a deadband form. This implies that as long as the process response is predicted to lie inside some interval or “band”, no adjustment is made. The optimal width of the deadband is calculated through a dynamic programming (DP) formulation.

1.2.4 Setup Adjustment for Asymmetric Off-target Cost

This objective of the research is to derive a solution to the setup adjustment problem for asymmetric cost functions. The system model is identical to that in the previous objective. Therefore the same conjugate Bayesian model can be used to make inferences and predict the process mean. However, the off-target cost is asymmetric and no adjustment cost is considered. Comparisons with the known-parameters solution presented by Colosimo et al [10] will be carried out.

1.2.5 Adaptive Deadband Control of a Process with Random Drift in the Presence of a Fixed Adjustment Cost

The final objective of this research is to study the control of a process that drifts. A process model where the process mean can drift randomly will be considered. An integrated moving average model, or IMA(1,1) model will be assumed, given the popularity of this model in the practice of time series analysis. Under the assumption that the parameters are unknown, a Bayesian model will be used for estimating and forecasting. The computation of the posterior distributions are carried out by the SMC method. An adaptive deadband control scheme is proposed for solving this problem, which is a combination of the Bayesian estimates and
the traditional solutions assuming known parameters.

These five research objectives are summarized in Table 1.1.

<table>
<thead>
<tr>
<th>Objective</th>
<th>Setup error?</th>
<th>Drift?</th>
<th>Off-target cost</th>
<th>Adjustment cost</th>
<th>Posterior Computation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>yes</td>
<td>no</td>
<td>quadratic</td>
<td>none</td>
<td>MCMC</td>
</tr>
<tr>
<td>2</td>
<td>yes</td>
<td>no</td>
<td>quadratic</td>
<td>none</td>
<td>SMC</td>
</tr>
<tr>
<td>3</td>
<td>yes</td>
<td>no</td>
<td>quadratic</td>
<td>fixed</td>
<td>Conjugate</td>
</tr>
<tr>
<td>4</td>
<td>yes</td>
<td>no</td>
<td>asymmetric</td>
<td>none</td>
<td>Conjugate</td>
</tr>
<tr>
<td>5</td>
<td>no</td>
<td>yes</td>
<td>quadratic</td>
<td>fixed</td>
<td>SMC</td>
</tr>
</tbody>
</table>

Table 1.1: Problems considered in this research.

1.3 Dissertation Outline

Chapter 2 reviews the previous research in the topics covered in this research. Such work falls into two categories. The first category includes the traditional solutions to adjustment problems assuming known parameters, in contrast to the proposed methods in this dissertation that solve the same adjustment problems under the assumption that the parameters are known. The second body of work that is reviewed includes an introduction to the statistical methodologies development relevant to this research.

Chapter 3-7 presents the proposed work corresponding to each of the objectives shown in Table 1.1. In each chapter, new control methods are proposed and evaluated through the comparison with the existing methods. In particular:

Chapter 3 gives a sensitivity analysis of a previously proposed multiple-lot process ad-
justment method based on a Bayesian model and MCMC techniques. An improved method is proposed to provide more robust performance.

Chapter 4 introduces a new setup adjustment method of a multiple-lot process using an SMC method, under the assumption that the process parameters are unknown.

Chapter 5 provides a single-lot process setup adjustment method in the presence of a fixed adjustment cost. Again, the parameters are assumed to be unknown. A conjugate Bayesian model is used to give inferences on the parameters, based on which an optimal Dynamic Programming solution is given.

Chapter 6 presents an optimal solution to a single-lot process setup adjustment problem under an asymmetric off-target cost. The same conjugate Bayesian model as in Chapter 5 is utilized for statistical inferences, and the optimal target at each step is determined according to the posterior distributions at that step.

In Chapter 7, a process that drifts randomly is studied under fixed adjustment and quadratic off-target costs. An adaptive deadband adjustment method that utilizes both the Bayesian estimates and the traditional solution is proposed.

In Chapter 8, a summary of the research findings is given and possible directions of future research are discussed.
Chapter 2

Literature Review

This chapter reviews literature relevant to the problems presented in Section 1.2. Literature relevant to the statistical techniques utilized in their solutions is reviewed as well.

2.1 Setup adjustment problems

The setup adjustment problem was originally proposed and studied by F.E. Grubbs [27]. The problem relates to a machine that produces discrete-parts in lots or batches, with the possibility of an improper setup operation resulting in an error or offset in the quality characteristic of interest. This problem often arises in discrete part manufacturing. For instance, Majeske and Hammett [36] studied sources of variation in sheet metal stamping, identifying three possible components: within-lot variation (i.e., variability of the part processed around a stable or trending mean of a lot); lot-to-lot variation (i.e., variability of the batch mean between setups); and within-lot level variation (movement of the mean during a given lot, e.g.,
trend). Using as case study the stamping of automotive body panels, these authors showed that within-lot, lot-to-lot and within-lot level variations account respectively for 21 %, 79 % and 0 % of the whole variability observed in the output measurements. For processes like this, in which setup significantly contributes to the overall process variation, if a manipulable variable exists (or can be found) in the machine or process, the total observed variance can be reduced by selecting a sequence of adjustments.

Figure 2.1: Scheme of the air bending process and of springback phenomenon, where $\alpha$ represents the bent angle.

As a motivational example of a process in which the initial offset varies from lot to lot but can be adjusted, consider the sheet metal bending process schematically represented in Figure 2.1. In air bending, a blank sheet is supported by two shoulders of a stationary die and the depth of the punch stroke (which acts as controllable parameter) determines the bend angle obtained on the workpiece after unloading (this is the response variable). The actual angle obtained on the unloaded sheet depends on the “springback”, i.e., the elastic
recovery of the original geometry, as represented in Figure 2.1. Springback is mainly affected by the thickness and the mechanical properties of the blank sheet, which vary from batch to batch because of different suppliers or different conditions of the supplier’s process [20]. Therefore, bent angles obtained processing a batch will often be biased with respect to the target angle and the punch depth can be used to compensate for this offset. Tsz-Sin Siu [44] has recently studied the effect of controlling the punch depth in sheet metal bending (Figure 2.1). In this experimental study, offset compensation is achieved using an integral controller.

Similarly to the setup adjustment problem, the process is characterized by no dynamics and the natural variability of the bent angle can be modeled as white noise.

Grubbs proposed two solutions to this problem, depending on whether one considers a single lot of parts or multiple lots of parts with one setup operation before each lot is started. The objective, in either case, is to minimize the sum of squared deviations from target of the quality characteristic (i.e., the only relevant cost is a quadratic off-target cost). These two solutions are often referred to as Grubbs’ harmonic rule and Grubbs’ extended rule. Grubbs’ harmonic rule is an optimal solution for adjusting a single-lot process when the initial offset is not predictable. Grubbs’ extended rule can be considered to be optimal for multiple lots, only when both the distribution of the initial offsets and the distribution of process variability are known. Grubbs’ solutions to the setup adjustment problem have been studied and extended recently by Trietsch [51] and by Del Castillo, Pan and Colosimo [16, 17]. Del Castillo, Pan and Colosimo [17] showed how Grubbs’ extended rule has a Bayesian interpretation based on a Kalman filter where a priori knowledge of parameters characterizing both batch-to-
batch and within-batch distributions is required. Grubbs’ harmonic rule can be considered a special case of the extended rule when the prior distribution on the initial offset has infinite variance. Although derived for single lot, as indicated by Castillo, Pan and Colosimo [17], Grubbs’ harmonic rule and the simple integral controller (or EWMA controller [6]) can perform better than the extended rule, since optimality of the later rule is no longer guaranteed if the assumed distribution of the initial offset is not exactly equal to the true distribution.

Although the integral controller is not developed to solve the setup adjustment problem, it has a competitive performance in this case. Del Castillo, Pan and Colosimo [16] showed that a discrete integral controller [6] has a relationship with Grubbs’ harmonic adjustment rule when the EWMA weights are designed according to the approach proposed by Guo and Chen [28], who presented an EWMA feedback controller with a “time varying” weight.

The Grubbs’ extended rule and the integral controller require the knowledge of the process parameters. If such knowledge is not available, they can not be applied or the performance may be poor. Recently, a different adjustment rule was proposed [11] for situations in which no prior knowledge on parameters characterizing offset and process distributions is available. This rule is based on computing sequential Bayesian estimates of the unknown parameters using Markov Chain Monte Carlo (MCMC) methods. The performance of the MCMC rule is very good, but the computational cost is excessive for on-line control applications.
2.2 Process Control under Fixed Adjustment Costs

Process control problems under fixed adjustment costs have been studied by Box and Jenkins [5], Crowder [13], and Box and Luceño [6]. Solution to these problems is of the form of a deadband, where the process is adjusted only if the response is predicted to be far enough from target, with the deadband width denoting the action limits that depend on the costs involved. These authors consider a process that would drift off-target if let uncontrolled. Box and Jenkins [5] solved the infinite-horizon version of this problem by minimizing the long-run expected cost. The solution to the finite-horizon version of the problem was given by Crowder [13] through a dynamic programming formulation. Jensen and Vardeman [30] considered a similar problem in which additional deterministic drifts and random adjustment errors are considered.

In the literature on setup adjustment, a fixed adjustment cost is also considered. Following Grubbs’ work, Triestch [50] showed that if fixed adjustment costs are considered, adjusting is not needed after each observation. When certain cost or error is associated with each adjustment, some adjustment instants should be "skipped". Triestch [51] proposed a method to design an optimal “schedule” of adjustments in the presence of fixed adjustment cost and fixed measurement cost. That is, if there are fixed adjustment costs and all process parameters are known a priori, it is possible to determine the best time instants when adjustments should be made, assuming no further offsets or shifts occur after setup. Pan and Del Castillo [39] studied the same problem from a different perspective. They identified that the formulation for the optimal schedule of adjustments is analogous to the well-known
Wagner-Whitin (W-W) algorithm [52] in the inventory control literature. Hence a W-W algorithm can be applied to obtain an optimal adjustment schedule. In addition to that, a Silver-Meal(S-M) heuristic [46], which is simpler but has close to optimal performance in inventory control applications, was also suggested. The solutions proposed by Triestch and Pan and Del Castillo can be referred to as Scheduling Methods. The magnitudes of the adjustments are determined on-line according to the process estimates obtained from the observations, so in this way they are feedback control rules. However, the times to make these adjustments are scheduled before the process starts. The adjustment times are independent of the observations and of the estimates and are based on the prior knowledge on the process. Triestch assumed the distribution of the initial offset $\theta_0$ is known. Pan and Del Castillo considered $\theta_0$ as an unknown constant, to which a prior distribution is given.

All the methods discussed above are based on the assumption that the variance parameters of the process are known.

2.3 Asymmetric Cost Functions

The asymmetric cost function is considered in the case when the economic value of undersized and oversized parts is different. In such case, it may not be economical for an adjustment procedure to bring the process mean back to the target or nominal value. Research in tolerance design and process centering (e.g., [38, 37, 34]) focuses on finding the optimal target for the process mean prior to starting the process. For setup adjustment under an asymmetric cost function, it is intuitive to have the value of the quality characteristic converge,
as parts are produced, to the optimal setting from the lower cost side [39]. Solving a setup adjustment problem under an asymmetric cost function was first considered by Colosimo, Pan and Del Castillo [39]. They utilized a Kalman Filter, which is shown to have a Bayesian interpretation, for estimating and forecasting. With this technique, a predictive distribution of the quality characteristic of the next part can be obtained before each part is processed. The mean of the predictive distribution can be linearly changed by the adjustment. Subject to this distribution, the authors found the values of the controllable factor that optimize the cost function. This solution is obtained based on the assumption that the variance of the part-to-part variation is known, which is used in estimating the process mean and determining the optimal target.

2.4 Bayesian Models

Bayesian modelling has advantages in many cases where there exist difficulties with classical sampling-based estimation [7, 24]. Furthermore, Bayesian approaches can be easily extended to deal with a wide range of problems (i.e., unbalanced data, non homogeneous variances, missing data). Such situations also exist in the process control field [11], in which a Bayesian approach seems to be more appropriate to be chosen as a statistical tool to give inferences on the variance components.

Traditional (non-Bayesian) control policies require estimates of process parameters before the policy can start to adjust the process. The lack of such prior information can cause difficulties in applying these policies. Even in the presence of certain prior information,
the errors that may exist in the prior information are ignored. In Bayesian approaches, prior distributions are used to incorporate the prior knowledge, which can include more information such as the precision of the estimator. In the case that no prior information is available, non-informative priors (usually a distribution with large variance) can be used to represent the lack of confidence. The flexibility of a prior distribution, in contrast to a single estimator, allows various forms of prior information to be utilized. For instance, a designer or an expert can hardly provide a single estimate of an unknown parameter, but he/she may give a region or interval that the parameter is mostly like to be in. Such information can be easily included.

As the process runs under the adjustment of a policy based on a Bayesian model, new statistical inferences on these parameters will be given by the posterior distributions, which will be updated and be more accurate as more data is produced. The updated posterior distributions will have an effect immediately in the adjustment procedures, which leads to more accurate adjustments.

While adaptive control techniques are frequently based on a Bayesian approach [40, 54, 1, 9, 32, 14], the class of problems this proposal presents have not been studied in such literature. In particular, there is no reference to setup adjustment or to fixed adjustment costs in the large body of literature on adaptive and self-tuning systems. These two characteristics of the problem under study seem to occur more in discrete part manufacturing and less in electronic, chemical or mechanical systems over which fast control through adaptation is required.
2.5 Computation of the Posterior Distribution

In some cases when a Bayesian model is used, conjugate prior distributions are available. This means the posterior distributions will have a closed form and belong to the same distribution class as the priors do. Therefore the computation of the posterior distributions can be easily carried out by mathematically calculating the variables that define the posterior distributions. In most cases, such conjugate priors do not exist. The posterior distributions have to be computed numerically, which is one of the biggest challenges in Bayesian approaches. Profiting from the development of computer technologies, more reliable and efficient numerical computation methods have been built in the last 15 years and the applications of Bayesian methods have broaden.

The Markov Chain Monte Carlo (MCMC) method, specifically the Gibbs sampler, has been used to compute posterior distributions in a multiple-lot process setup adjustment method based on a Bayesian model [11]. Although the Gibbs sampler method is not the only approach available to compute the Bayesian analysis required, it has been selected because of its simplicity and flexibility. As Gelfand et al. [22] pointed out “the efficiency of other approaches is at the expense of detailed sophisticated applied analysis [7] or tailored “one-off” numerical tricks or sophisticated adaptive quadrature methodology, in conjunction with subtle sensitivity of parameterization issues”.

Despite the excellent performance of the MCMC adjustment method, it is a rule that requires substantial computational time at each point in time, an obvious disadvantage if on-line control is needed for a process where the time between parts is relatively short. The
reason of this is that the Markov Chain computations are restarted from scratch each time a new observation is obtained.

As an alternative to MCMC methods, SMC methods can also be used for computation in an adjustment rule. SMC methods also rely on Monte Carlo algorithms for the solution of Bayesian inference problems in which posterior distributions of the unknown parameters are created numerically from the generation of a large number of random variates or “particles”. In contrast to MCMC methods, which build these distributions from scratch every time a new observation becomes available (i.e., they start from the prior distributions every time and pass through all data up to the current time), SMC techniques if applied to Bayesian inference problems update sequentially the previous posterior distributions at the previous step and transform them into the new posteriors at the current step, from which inferences useful for adjustment purposes can be made. Thus, in SMC methods Bayes’ formula is used to update samples that approximate the underlying distributions. Clearly, the computational efficiency of SMC techniques is relevant in production environments where the time to produce parts is very short and fast on-line adjustments required. More detailed comparison between MCMC methods and SMC methods is shown in Chapter 4 and its appendix.

A relative new area of research in Statistics, SMC techniques have appeared in the last decade in the Statistical literature under names such as Sampling/Importance Resampling (SIR), bootstrapping filters, and particle filters, among other names. A detailed presentation of SMC methods can be found in [19].
Chapter 3

Setup Adjustment of Multiple Lots using a Markov Chain Monte Carlo Method

3.1 Introduction

In this chapter, an in-depth comparison of the performance of setup adjustments rules that can be applied when no parameter estimates are available, namely, Grubbs’ harmonic rule, the EWMA controller, and the MCMC policy, is provided. Performance will be evaluated considering different production scenarios characterized by the following set of parameters:

- the number of lots,
- the number of parts in each lot,
• the variability within and between-batches,

• the presence of a systematic error affecting the mean offset over a set of lots.

Not all parameters are equally uncertain in practice. In fact, the first two parameters, i.e., the number of lots and the lot size, can be assumed known even when a new product and/or a new process is considered. Given these two parameters, the better rule will be considered as the one which determines enhanced performance over all the alternative scenarios associated with a set of possible values of the remaining unknown parameters (i.e., the within-batch and between-batches variability and the mean offset over a set of lots).

It will be shown that the MCMC-based approach outperforms the other rules in almost all the cases. A closer study of situations in which performance of the MCMC policy is actually worse has motivated a modified version of this rule, which, as we will show, has an enhanced performance in all the cases examined.

3.2 Rules for adjusting an initial offset over a set of batches when parameters are unknown

When discrete parts are processed in batches and no adjustment rule is adopted, a common model used in the statistical literature to describe the quality characteristic observed after
each part is machined is the random effects model [7], given by:

$$Y_{ij} = \theta_i + v_{ij}$$  \hspace{1cm} (3.1)

$$\theta_i \sim N(\mu, \sigma_{\theta}^2)$$

$$v_{ij} \sim N(0, \sigma_v^2)$$;

where:

- $i = 1, \ldots, I$ is the batch index;
- $j = 1, \ldots, J$ is the part index;
- $Y_{ij}$ is the deviation from the nominal value for the quality characteristic observed at
  the $j^{th}$ part of the $i^{th}$ batch;
- $\theta_i$ is the (unknown) mean in each batch, here representing the initial offset due to setup
  errors. This is assumed to be normally distributed with mean $\mu$ and variance $\sigma_{\theta}^2$;
- $v_{ij}$ represents the random error due to the combined effect of the intrinsic variability
  in the machining process and the variability in the measurement system. It is assumed
  to be normally distributed with mean equal to zero and variance $\sigma_v^2$.

The mean offset over a set of batches $\mu$ can be interpreted as the systematic error in the
set-up operations, while the two variance components $\sigma_v^2$ and $\sigma_{\theta}^2$ represent the within and
between-batches variability, respectively.

To improve the quality of parts processed, adjustment rules should be adopted to elim-
inate the initial offset of parts processed in each lot by means of a compensatory variable
When the adjustment rule is applied, the mean of the quality characteristic of the \( j^{th} \) part processed in the \( i^{th} \) lot is given by:

\[
\theta_{ij} = \theta_{ij-1} + U_{ij-1}
\]  

where \( \theta_{i0} = \theta_i \), \( U_{ij-1} = u_{ij-1} - u_{ij-2} \) is the magnitude of the adjustment, and the quality characteristic can be thus expressed as:

\[
Y_{ij} = \theta_{ij} + v_{ij}
\]  

In view of equation (3.2), the “ideal” adjustment would be given by \( U_{ij-1} = -\theta_{ij-1} \), but the actual adjustment should be based on an estimate of \( \theta_{ij-1} \) since it is unknown. Solving recursively equation (3.2) and considering that \( U_{ij} = u_{ij} - u_{ij-1} \), the quality characteristic at the \( j^{th} \) part in the \( i^{th} \) lot can be rewritten as:

\[
Y_{ij} = \theta_i + u_{ij-1} + v_{ij}
\]  

and the adjustment problem can be reformulated as selecting \( u_{ij-1} \) (\( j = 2, \ldots, J \)) to be closer as possible to \(-\theta_i\).

If parameters \( \mu, \sigma_\theta, \sigma_v \) characterizing the distribution of the quality characteristic \( Y_{ij} \) (equation 3.1) are unknown, traditional adjustment rules that can be applied in this case are:

- **Grubbs’ harmonic rule** [27], where \( U_{ij} = -\frac{1}{j}Y_{ij} \) (\( U_{i0} = 0 \ \forall i \)). We note that Grubbs’ extended rule, characterized by \( U_{ij} = Y_{ij}/(j + \sigma_v^2/\sigma_\theta^2) \), is not applicable in practice as it requires knowledge of \( \sigma_v^2/\sigma_\theta^2 \).
• the integral or EWMA controller [6], where \( U_{ij} = -\lambda Y_{ij} \) (\( U_{i0} = 0 \) \( \forall i \)) and \( \lambda \) is some weight parameter (the integral constant) that needs to be chosen a priori;

• the sequential Bayesian adjustment rule [11], in which required parameters are sequentially predicted or estimated using Markov Chain Monte Carlo (MCMC) as detailed in the appendix of this chapter. In particular:

1. before processing a new lot \( i \) (where \( i \geq 3 \) because at least two lots must be processed to estimate the variance between lots \( \sigma^2 \)), the initial set-point \( u_{i0} \) that has to be set on the machine before processing lot \( i \) is based on the predictive distribution of the offset;

2. after observing at least one part in a lot, adjustments \( U_{ij} (j = 1, .., J) \) are based on the posterior distribution of the offset in the current lot, as seen from equation (3.4).

The appendix of this chapter describes the hierarchical normal means model used to estimate/predict required parameters and details on computation performed using the Gibbs Sampler method.
\subsection*{3.3 Evaluation of different adjustment rules with reference to different performance indexes}

Colosimo, Pan and del Castillo [11] compare the performance of the three adjustment rules mentioned above under the usual quadratic cost function, given by:

\begin{equation}
C = \sum_{i=1}^{I} \sum_{j=1}^{J} Y_{ij}^2.
\end{equation}

Two classical variance components examples presented in [7] were used for evaluation purposes. Percentage advantages in quadratic cost determined by the MCMC Bayesian procedure over Grubbs’ rule and EWMA controllers varied from 27\% to 63\% for the first example and from 18\% to 50\% for the second one [11].

To compare in more detail the adjustment rules applicable when no knowledge of parameters is \textit{a priori} available, we refer to a manufacturing process where parts are processed in lots. To represent a wide range of production situations, we considered as initial reference the second example presented in [7], further studied in [22] (this is characterized by $I = 6$, $J = 5$, $\sigma_v = 4$, $\mu = 4$, $\sigma_\theta = 2$) and we perturbed all the parameters to generate different scenarios, as reported in Table 3.1. The number of lots ($I$) was considered fixed and equal to 20, since performance for a number of lots $I < 20$ can be computed from partial results of the complete 20-lot simulations. For the other four parameters affecting the performance of the rules, the set of cases studied in this chapter was derived considering all the possible combinations of the parameters at two values, as reported in Table 3.2. Three replications were conducted for each scenario, thus the total number of simulations in this analysis was
equal to \(2^4 \times 3 = 48\) runs.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value in example 2 by Box and Tiao ([7] p. 247)</th>
<th>Perturbed value</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>J</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>(\mu)</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>(\sigma_\theta)</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>(\sigma_v)</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 3.1: Parameters characterizing different scenarios examined.

<table>
<thead>
<tr>
<th>Case</th>
<th>(J)</th>
<th>(\sigma_v)</th>
<th>(\mu)</th>
<th>(\sigma_\theta)</th>
<th>Case</th>
<th>(J)</th>
<th>(\sigma_v)</th>
<th>(\mu)</th>
<th>(\sigma_\theta)</th>
<th>Case</th>
<th>(J)</th>
<th>(\sigma_v)</th>
<th>(\mu)</th>
<th>(\sigma_\theta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>20</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>11</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>12</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 3.2: Set of parameters characterizing cases studied in this chapter.

For each scenario, percentage savings in quadratic costs (equation 3.5) obtained using the MCMC-based approach instead of Grubbs’ harmonic rule and two EWMA controllers (with \(\lambda = 0.4\) and \(\lambda = 0.1\)) were computed. Nonparametric statistical analysis was used to analyze results because of the lack of normality of percentage savings obtained in simulated scenarios. Figure 3.1 reports the interquartile (IQ) and the confidence interval (CI) boxplots of the percentage savings induced by the MCMC-based adjustment rules over competitor rules. In particular, each CI boxplot is plotted inside the IQ one, and refers to the 95% confidence interval on the median. IQ and CI boxplots are reported in Figure 3.1 as a function of: the number of lots \(I\); the number of parts in lots \(J\); the ratio of the mean offset to the within-lot standard deviation \((\mu/\sigma_v)\); the ratio of the between to the within-lot standard deviation \((\sigma_\theta/\sigma_v)\).
Figure 3.1: Boxplots of the percentage savings (using the quadratic cost function) induced by the MCMC adjustment rule over Grubbs’ harmonic rule and the two EWMA controllers ($\lambda = 0.1$ and $\lambda = 0.4$) as a function of different parameters: the number of lots ($I$); the number of parts in lots ($J$); the ratio of the mean offset to the within-lot standard deviation ($\mu/\sigma_v$); the ratio of the between to the within lot standard deviation ($\sigma_\theta/\sigma_v$).
As it can be observed, the MCMC approach outperforms the competitor adjustment rules in almost all the scenarios, inducing sometimes significant advantages. In particular, actual applications of the adjustment rules should be based on the behavior of the percentage savings with respect to the number of lots \((I)\) and the number of parts in each lot \((J)\). In fact, when a new product is considered or a new process is introduced \(I\) and \(J\) are the only parameters that can be assumed known. Table 3.3 reports the 95% confidence interval on the median of the percentage savings induced by the MCMC policy over the competitor rules, as a function of the number of lots \(I\) and the number of parts in each lot \(J\). These confidence intervals represent possible advantages induced by the MCMC policy if all the scenarios simulated for the other set of parameters \((\mu, \sigma_\theta \text{ and } \sigma_v)\) can be considered \textit{a priori} equally likely. We note that the lower confidence intervals on the percentage savings induced by the MCMC policy is almost always greater than zero, except for cases in which the number of lots processed \((I)\) and the lot size \((J)\) are small. Thus, for example, when \(I = 5\) there is no significant advantage in adopting the MCMC rule instead of Grubbs’ rule, because the 95% confidence interval on the median of percentage savings is given by \((-1.9\% , 5.8\%)\) and contains 0%. A similar comment applies to the case when the lot size is small, i.e. \(J = 5\), where there is no significant advantage compared to Grubbs’ rule and the EWMA controller with \(\lambda = 0.4\).

To better investigate the performance of the different methods, the quadratic cost function reported in equation (3.5) was expanded into its different components. The total sum of squares criterion contains the part-to-part errors \(v_{ij}\) that are not controllable. Therefore,
<table>
<thead>
<tr>
<th></th>
<th>Grubbs</th>
<th></th>
<th></th>
<th>EWMA (0.1)</th>
<th></th>
<th></th>
<th>EWMA (0.4)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>median</td>
<td>lower CI</td>
<td>upper CI</td>
<td>median</td>
<td>lower CI</td>
<td>upper CI</td>
<td>median</td>
<td>lower CI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
</tr>
<tr>
<td>I</td>
<td>5</td>
<td>2.7%</td>
<td>-1.9%</td>
<td>5.8%</td>
<td>18.1%</td>
<td>5.4%</td>
<td>29.3%</td>
<td>9.0%</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>3.7%</td>
<td>0.4%</td>
<td>6.7%</td>
<td>20.6%</td>
<td>13.9%</td>
<td>32.8%</td>
<td>9.7%</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>6.8%</td>
<td>3.6%</td>
<td>9.5%</td>
<td>24.4%</td>
<td>18.3%</td>
<td>33.8%</td>
<td>10.7%</td>
</tr>
<tr>
<td>J</td>
<td>5</td>
<td>5.1%</td>
<td>-1.6%</td>
<td>11.1%</td>
<td>28.5%</td>
<td>20.1%</td>
<td>34.0%</td>
<td>-0.4%</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>3.7%</td>
<td>2.6%</td>
<td>6.5%</td>
<td>16.7%</td>
<td>9.1%</td>
<td>25.5%</td>
<td>11.7%</td>
</tr>
</tbody>
</table>

Table 3.3: Confidence intervals (at level 95%) on the median of the percentage savings in total quadratic costs induced by the MCMC method compared to each alternative adjustment rule.

A more informative evaluation of the performance of any adjustment rule is to discount the variability induced by the \( v_{ij} \). This alternative criterion is evidently more preferable the larger \( \sigma_v^2 \) is. To do this, we partition the total quadratic cost function as follows:

\[
C = \sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij}^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} (\theta_i + u_{ij} + v_{ij})^2 = \\
= \sum_{i=1}^{I} \sum_{j=1}^{J} (\theta_i + u_{ij})^2 + \sum_{i=1}^{I} \sum_{j=1}^{J} 2(\theta_i + u_{ij})v_{ij} + \sum_{i=1}^{I} \sum_{j=1}^{J} v_{ij}^2
\]  

(3.6)

Since \( v_{ij} \sim \text{NID}(0, \sigma_v^2) \) and is independent of \((\theta_i + u_{ij})\), the second term in equation (3.6) \((\sum_i \sum_j 2(\theta_i + u_{ij})v_{ij})\) will be close to zero for relative large \( I \) and \( J \). The third term, which is independent of \( u_{ij} \), is not controllable. The first term \( \sum_i \sum_j (\theta_i + u_{ij})^2 \), the only part left in the total cost function that could be improved by an adjustment rule, will be referred to in what follows as the quadratic bias cost function.

The quadratic bias directly measures how well \( u_{ij} \) converges to \(-\theta_i\), where the gap between \( u_{ij} \) and \(-\theta_i\) is the bias. Notice that, for an ideal adjustment rule without any bias, i.e. when \( u_{ij} \) always equals \(-\theta_i\) (a situation impossible to achieve in practice), the corresponding saving rates would be 100%.
Considering this new performance index, IQ and CI boxplots of percentage savings induced by the MCMC policy are shown in Figure 3.2. Focusing just on parameters known before processing, Table 3.4 shows the 95% confidence interval on the median of the percentage savings induced by the MCMC policy as a function of the number of lots $I$ and the size of the lots $J$. Although basic considerations drawn for the quadratic cost function hold in this case too, the adoption of the quadratic bias cost function better outlines the advantages and disadvantages implied by the use of the MCMC policy. In particular, considering just the parameters known before processing lots ($I$ and $J$), the new performance index confirms that when few lots are processed (i.e., $I = 5$) there is no statistical evidence to assess that the MCMC policy performs better than Grubbs’ rule. Similarly, when lot size is small (i.e., $J = 5$) the performance of the MCMC policy is equivalent to that of an EWMA controller with $\lambda = 0.4$ (note that this conclusion was reached also using the quadratic cost function, but using such function the performance of Grubbs’ rule was also found to be equivalent). However, the relative performance of the EWMA controller is very sensible to a proper selection of the weight $\lambda$. In fact, when $\lambda = 0.1$ in the EWMA controller, savings induced by the MCMC policy are significantly greater than zero even for a small lot size ($J = 5$) since in this case the median of the percentage savings is equal to 45.7% and the 95% confidence interval is given by (37.5% , 55.8%). It is important to emphasize that in practice, there is no way to determine $\lambda$ appropriately in an EWMA controller if the process parameters are unknown.
Figure 3.2: Boxplots on the percentage savings (using the quadratic bias cost function) induced by the MCMC adjustment rule over Grubbs’ harmonic rule and the two EWMA controllers ($\lambda = 0.1$ and $\lambda = 0.4$) as a function of different parameters: the number of lots ($I$), the number of parts in lots ($J$), the ratio of the mean offset to the within-lot standard deviation ($\mu/\sigma_v$), and the ratio of the between to the within-lot standard deviation ($\sigma_\theta/\sigma_v$).
Table 3.4: Confidence intervals (at level 95%) on the median of the percentage savings in quadratic bias costs induced by the MCMC method.

<table>
<thead>
<tr>
<th></th>
<th>Grubbs</th>
<th>EWMA (0.1)</th>
<th>EWMA (0.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>median</td>
<td>lower CI</td>
<td>upper CI 95%</td>
</tr>
<tr>
<td>I</td>
<td>5</td>
<td>7.3%</td>
<td>-2.3%</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>15.2%</td>
<td>8.2%</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>25.4%</td>
<td>16.0%</td>
</tr>
<tr>
<td>J</td>
<td>5</td>
<td>14.0%</td>
<td>0.3%</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>20.2%</td>
<td>11.1%</td>
</tr>
</tbody>
</table>

3.4 Motivation for a modified rule

As shown in the previous section, the MCMC method can result in significant savings over the other three adjustment methods in most, but not all, cases. To better explore situations in which the MCMC adjustment policy induces worst performance, a particular study was conducted to better investigate the reasons of this behavior. The two cases where the MCMC policy results in worse performance compared to competitor rules were studied in more detail. Table 3.5 reports the mean of the percentage savings implied by using the MCMC policy for the three simulations in each scenario. From it, cases 16 (\( J = 5; \mu = 0, \sigma_v = 2, \sigma_\theta = 4 \)) and 9 (\( J = 5; \mu = 4, \sigma_v = 4, \sigma_\theta = 2 \)) were considered as representative of the worst and the best performance of the MCMC policy. Behavior of adjustment rules in one simulation from each of these cases was specifically analyzed.

Figure 3.3 shows the sum of quadratic costs induced over the set of lots \( I \) by all the competitor rules as a function of the number of parts in the lot, i.e., \( \sum_{i=1}^{20} Y_{ij}^2 \) for \( j = 1, 2, ..., 5 \). In particular, graphs (a) and (b) in Figure 3.3 refer to simulation of case 16 (worst performance case) and 9 (best performance case), respectively. As it can be observed in
Table 3.5: Mean percentage savings in total quadratic costs induced by the MCMC method in simulated scenarios.

Figure 3.3 (a), the worst performance of the MCMC policy is mainly due to the first part \((j = 1)\) processed in the lot. For parts other than the first one (i.e., \(j = 2, \ldots, 5\)), costs of the MCMC approach are almost identical to the one obtained with Grubbs’ rule, which is the best approach in this case. On the contrary, Figure 3.3(b) shows that the MCMC policy outperforms the other rule starting from the first part \((j = 1)\) processed in each lot.

This behavior could be better explained considering the two components of the MCMC policy:

1. the initial setpoint \(u_{i0}\) which is computed starting at the third lot \((i \geq 3)\) and is based on the posterior predictive distribution of a future offset \(\theta'_i\) given all the data collected from previously produced lots (equation 3.12 in the appendix);
Figure 3.3: Quadratic costs in two simulation runs (a: case 16 and b: case 9) as a function of the number of parts processed in the lot.

2. the adjustments $U_{ij}$ for $j = 1, \ldots, J$ based on the posterior distribution of the offset $\theta_i$ and which affect the quality characteristic of parts other than the first (equations 3.11 and 3.14 in the appendix).

Figure 3.3 clearly shows how the first component of the MCMC policy, the initial setpoint (and initial adjustment since $U_{i0} = u_{i0}$) does not perform well for some cases. From equation (3.12) in the appendix, the initial setpoint $u_{i0}$ set on the machine before processing the first
part in lot $i$, is given by:

$$u_{i0} = -E(\theta'_i | x^{i-1}, M_1) \quad \text{for } i = 3, ..., I,$$

(3.7)

where $x^{i-1}$ represents information (data and adjustments) available at this point in time and $M_1$ refers to the hierarchical normal means model used in the MCMC simulation (see the appendix of this chapter for further details). Figure 3.4 shows a plot of the estimated $\bar{\theta}'_i$ and of the accuracy of this estimate, measured by the interval $\bar{\theta}'_i \pm \hat{\sigma}_{\theta'_i}$, as a function of lots processed. Both the sample mean $\bar{\theta}'_i$ and the sample standard deviation $\hat{\sigma}_{\theta'_i}$ were obtained from the MCMC simulation.

As it can be seen, the sample mean converges to the true value of $\mu$ (equal to 0 and 4 for cases represented in Figure 3.4 (a) and (b), respectively) as the number of lots processed increases. Thus, the initial set-point $u_{i0}$ tends to be set to the proper value $-\mu$. That is, once the steady-state behavior is achieved, the MCMC policy allows to correct a possible systematic error before observing parts in a lot. However, there is some slowness in converging to the right value, as it is clear from Figure 3.4 (a). When $\mu = 0$, Grubbs’ rule and the EWMA controllers perform better because they do not include the predictive feature in the control rule, i.e., $u_{i0} = 0$ is set by default, and this happens to be the optimal selection for the initial set-point $u_{i0}$. Figure 3.4 (a) and (b) show that also the accuracy of the estimate depends on the specific scenario examined. Hence, a new modified version of the MCMC adjustment rule should take into account the disadvantage of using the initial adjustment when there is no systematic error (i.e., when $\mu = 0$) and when the uncertainty in the predicted lot mean
Figure 3.4: Plot of the interval $\bar{\theta}_i \pm \hat{\sigma}_{\theta_i}$, as a function of lots processed.

$\theta_i$ is large. This is discussed next.

3.5 Conditional First Adjustment Rule based on MCMC Approach

It is clear from the preceding discussion that applying the first adjustment in each lot $U_{i0} = u_{i0}$ can lead to an advantage or a disadvantage when using the MCMC policy, depending
on whether there is a systematic error ($\mu \neq 0$) or not ($\mu = 0$). However, the actual value of $\mu$ can not be known in advance, due to our assumption on unknown parameters. Therefore a different MCMC approach should be based on selecting "online" the best strategy for the $u_{i0}$’s. This can be done taking into account the accuracy of the estimate $\hat{\sigma}_{\theta_i}$. Figure 3.4 shows that the width of the interval $\bar{\theta}_i \pm \hat{\sigma}_{\theta_i}$ reduces as the number of lots processed increases. Hence, problems related to the slow convergence of $\bar{\theta}_i$ to $\mu$ can be overcome by including information on estimate accuracy in the adjustment rule.

A new “Conditional First Adjustment” MCMC method is thus proposed as follows:

1. At the beginning of each lot $i$ (with $i \geq 3$), substitute equation (3.12) of the original approach with:

$$
\begin{cases}
\text{adjust} & (u_{i0} = -\bar{\theta}_i) \quad \text{if} \quad |\bar{\theta}_i| > k\hat{\sigma}_{\theta_i} ; \\
\text{do not adjust} & (u_{i0} = 0) \quad \text{otherwise}.
\end{cases}
$$

(3.8)

2. Use the original MCMC method (equations 3.11 and 3.14) for all the other adjustments.

Here, $k$ is a tuning constant of the method that we shall discuss later and $\bar{\theta}_i$ and $\hat{\sigma}_{\theta_i}$ are respectively the mean and standard deviation of the posterior predictive distribution of the offset, conditional on the data observed before lot $i$. When the ratio $|\bar{\theta}_i|/\hat{\sigma}_{\theta_i}$ is relatively high, i.e., the estimate has a high precision compared with its magnitude, we have a strong belief of the estimate. This is similar to a frequentist test of significance for a normal population mean. While strictly speaking we should use a Bayesian factor for testing significance since the posterior distribution is not necessarily normal, the simpler conditional rule works very
well, as will be shown below. Therefore, we decided to keep its simplicity instead of the rigor of Bayesian factors.

As \( k \to 0 \), the conditional first adjustment rule becomes the same as the original MCMC method. As \( k \) increases, the percentage savings in the cases where the MCMC performs well start to drop, while those in the cases with poor performances become higher. When \( k \to \infty \), the conditional rule becomes the same as an MCMC rule that always omits the first adjustment in each lot.

Figure 3.5 shows how the maximum and minimum of the percentage savings among the 16 cases vary as \( k \) is changed in the Conditional First Adjustment rule. In particular, this Figure indicates that \( k \) should be chosen to have a proper trade-off between reduction of the maximum savings (i.e., the best performance of the MCMC policy in Table 3.5) and improvement of the minimum savings (i.e., the worst performance of the MCMC policy in Table 3.5). The figure indicates that the best performance deteriorates very little while the worst performance improves considerably for the first few values of \( k \) that were tried. From this it was concluded that the value of \( k = 0.64 \) showed the best performance and was used in what follows.

Figures 3.6 and 3.7, report IQ and CI boxplots of the percentage savings obtained with the Conditional First Adjustment MCMC rule (\( k = 0.64 \)) using the quadratic and the quadratic bias cost functions, respectively. Compared with previous boxplots obtained adopting the original MCMC approach (Figures 3.1 and 3.2, respectively), these graphs show that the new policy induces advantages over the competitor rules. In particular, when the quadratic
Figure 3.5: The maximum savings vs. the minimum savings among the 16 cases on the Tables as a function of the adjustment limit $k$ when applying the Conditional First Adjustment MCMC rule.

cost function is used as performance index, the 95% CI boxplots obtained with the new rule shifts all above the line at 0% (Figure 3.6).

Focusing only on the performance related to parameters that are always known ($I$ and $J$), Tables 3.6 and 3.7 report the 95% confidence interval on the median of savings induced by the Conditional First Adjustment MCMC method with $k = 0.64$ (using the quadratic and quadratic bias costs, respectively). Comparing these two tables with Tables 3.3 and 3.4 (obtained with the original MCMC method), we find that the new policy is preferable to other competitor rules for all the values of $I$ and $J$, since the left confidence interval limits are now always greater than 0%. In terms of the medians of the percentage savings induced
Figure 3.6: Boxplots on the percentage savings (using the quadratic cost function) induced by the Conditional First Adjustment MCMC policy over Grubbs’ harmonic rule and the two EWMA controllers ($\lambda = 0.1$ and $\lambda = 0.4$) as a function of different parameters: the number of lots ($I$), the number of parts in lots ($J$), the ratio of the mean offset to the within-lot standard deviation ($\mu/\sigma_v$), and the ratio of the between to the within-lot standard deviations ($\sigma_\theta/\sigma_v$).
Figure 3.7: Boxplots on the percentage savings (using the quadratic bias cost function) induced by the Conditional First Adjustment MCMC policy over Grubbs’ harmonic rule and the two EWMA controllers ($\lambda = 0.1$ and $\lambda = 0.4$) as a function of different parameters: the number of lots ($I$), the number of parts in lots ($J$), the ratio of the mean offset to the within-lot standard deviation ($\mu/\sigma_v$), and the ratio of the between to the within-lot standard deviations ($\sigma_\theta/\sigma_v$).
by this new rule, (Tables 3.6 and 3.7), it should be noticed that the median of advantages ranges from 2.8% to 29.0% when using the quadratic cost function and from 9.6% to 57.6% when using the quadratic bias cost function. Therefore, we can conclude that the Conditional First Adjustment MCMC rule should be preferred to competitor rules when new products or new processes need to be introduced.

<table>
<thead>
<tr>
<th></th>
<th>Grubbs</th>
<th>EWMA (0.1)</th>
<th>EWMA (0.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>median</td>
<td>lower CI 95%</td>
<td>upper CI 95%</td>
</tr>
<tr>
<td>I</td>
<td>5</td>
<td>2.8%</td>
<td>1.5%</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>3.8%</td>
<td>1.7%</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>5.8%</td>
<td>2.7%</td>
</tr>
<tr>
<td>J</td>
<td>5</td>
<td>4.2%</td>
<td>2.0%</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>4.1%</td>
<td>2.6%</td>
</tr>
</tbody>
</table>

Table 3.6: Confidence intervals (at level 95%) on the median of the percentage savings in total quadratic costs induced by the Conditional First Adjustment MCMC method.

<table>
<thead>
<tr>
<th></th>
<th>Grubbs</th>
<th>EWMA (0.1)</th>
<th>EWMA (0.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>median</td>
<td>lower CI 95%</td>
<td>upper CI 95%</td>
</tr>
<tr>
<td>I</td>
<td>5</td>
<td>9.6%</td>
<td>2.6%</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>19.1%</td>
<td>8.3%</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>21.6%</td>
<td>17.6%</td>
</tr>
<tr>
<td>J</td>
<td>5</td>
<td>18.7%</td>
<td>6.9%</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>19.4%</td>
<td>10.3%</td>
</tr>
</tbody>
</table>

Table 3.7: Confidence intervals (at level 95%) on the median of the percentage savings in quadratic bias costs induced by the Conditional First Adjustment MCMC method.

3.6 Conclusions

In this chapter, we dealt with the problem of adjusting initial offsets for quality characteristics of discrete-parts manufactured in batches. At the beginning of each batch, a set-up error
can cause the mean of the quality characteristic to be off-target and the adjustment procedure is designed to compensate for this initial off-set as the number of parts processed in the lot increases. In particular, we focused on the case in which off-target costs are quadratic, adjustment costs can be neglected and no previous knowledge on parameters characterizing the off-set distribution and the process intrinsic variability is available. This situation can properly model adjustment rules designed for new products or newly installed processes, in which there is no previous experience on set-up operations. We compared different adjustment rules that can be applied in this case: Grubbs’ rule, the integral or EWMA controller, and the adjustment rule based on a Bayesian sequential estimation of unknown parameters using MCMC simulation (MCMC adjustment rule). Considering different production scenarios, the last rule was shown to outperform other existing rules unless very few lots has to be machined or lots size is particularly small. A further study of performance of the MCMC adjustment rules in these last cases motivated a revised version of the rule, that has consistent advantages over Grubbs’ rule and the EWMA controller in all the cases examined.

The next chapter discusses the application of other Bayesian inference technique to the solution of the same problem as discussed in this chapter, that of setup adjustment for multiple lots.
Appendix of Chapter 3: A Bayesian adjustment rule based on MCMC simulation

When no prior knowledge on parameters is available, an adjustment rule can be based only on the history of data acquired which includes the response values and the previous adjustments. Under this assumption, a transformed variable $X_{ij}$ computed at the time $Y_{ij}$ is observed (since at that time $u_{ij-1}$ is known) is completely defined as:

$$X_{ij} = Y_{ij} - u_{ij-1} = \theta_i + v$$

(3.9)

and allows to derive a one-way random effects model for the adjusted process. Adopting a Bayesian perspective, the one-way random effects model is a special case of a hierarchical model, where the first stage models the distribution of observable data conditionally on unknown parameters, the second stage in the hierarchy specifies the prior distribution of these unknown parameters which can depend on some hyperparameters, which in turn have a prior at a third stage. The three-stage hierarchical model (that we will call $M_1$) is hence given by:

\begin{align*}
\text{first stage: } & \quad X_{kp} | \theta_k, \sigma_v^2 \sim N(\theta_k, \sigma_v^2) \quad (k = 1, \ldots, i \text{ and } p = 1, \ldots, J); \quad (3.10) \\
\text{second stage: } & \quad \theta_k | \mu, \sigma_\theta^2 \sim N(\mu, \sigma_\theta^2) \quad (k = 1, \ldots, i) , \\
& \quad \sigma_v^2 | a_1, b_1 \sim IG(a_1, b_1) ; \\
\text{third stage: } & \quad \mu | \mu_0, \sigma_\theta^2 \sim N(\mu_0, \sigma_\theta^2) , \\
& \quad \sigma_\theta^2 | a_2, b_2 \sim IG(a_2, b_2) ;
\end{align*}
where conjugacy has been used at each step of the hierarchical model (a common choice
for the random effects model [22], [23]) and where IG represents an Inverse-Gamma dis-
tribution. Parameters $\mu_0, \sigma_0^2, a_1, b_1, a_2, b_2$ are assumed known. In particular, they were
selected according to values suggested in the literature [47] to model "vague" prior dis-
tributions, i.e., $\mu_0 = 0, \sigma_0^2 = 1.0E + 10, a_1 = b_1 = a_2 = b_2 = 0.001$. Denoting with
$x^{ij} = \{x_{11}, x_{12}, ..., x_{1J}, ..., x_{i1}, ..., x_{ij}\}$ all (transformed) data observed after the $j^{th}$ part in
the $i^{th}$ lot has been just machined, the adjustment $U_{ij} = u_{ij} - u_{ij-1}$ ($j > 0$) can be com-
puted using the posterior distribution $(\theta_i|x^{ij}, M_1)$ when at least one part in the lot has been
processed. The predictive distribution $\theta'_{i+1}|x^{iJ}, M_1$ [25] can be used to select the initial set-
point $u_{i+10}$ that has to be set on the machine before processing the first part in the next lot
(lot $i + 1$), i.e.:

$$u_{ij} = -E(\theta_i|x^{ij}, M_1) \quad i = 2, ..., I \quad \text{and} \quad j = 1, ..., J,$$

(3.11)

$$u_{i+10} = -E(\theta'_{i+1}|x^{iJ}, M_1) \quad \text{for} \quad i = 2, ..., I$$

(3.12)

where the index $i$ is set to consider that adjusting can be performed only if $\sigma_0^2$ can be
estimated (i.e., when at least two parts from different lots have been already processed) and
$\theta'_{i+1}|x^{iJ}, M_1$ is the posterior predictive distribution [25] of a future offset given all the data
collected from previously produced lots.

Adjustments for parts in the first lot can be similarly derived considering a "reduced"
two-stage hierarchical model $M_2$ given by:

\begin{align*}
\text{first stage:} & \quad X_{1j}|\theta_1, \sigma_v^2 \sim N(\theta_1, \sigma_v^2) \\
\text{second stage:} & \quad \theta_1|\mu, \sigma^2 \sim N(\mu, \sigma^2) \\
& \quad \sigma_v^2 \sim IG(a_1, b_1).
\end{align*}

Since at least two parts have to be processed to estimate the within-lot variance $\sigma_v^2$, the adjustments $U_{1j}$ for $j = 2, \ldots, J$ can be computed using the hierarchical model $M_2$ while a trivial estimator (as in Grubbs' harmonic rule) can be used to start adjusting just after the first part in the first lot has been machined, i.e.:

\begin{align*}
u_{11} &= -x_{11}, \\
u_{1j} &= -E(\theta_1|x_{1j}, M_2) \quad j = 2, 3, \ldots, J.
\end{align*}

As each observation becomes available, a Gibbs Sampler is run to estimate the parameters of the hierarchical normal means model given the observations up to that point in time. The current lot mean estimate is then used for adjustment. The MCMC simulation coded in the Bugs (Bayesian inference Using Gibbs Sampling, [47]) language was used to perform Gibbs Sampling. Following the literature on convergence diagnostic ([8], [12], [24]) both the algorithms of Raftery and Lewis [42] and Gelman and Rubin [26] were used within the MCMC simulation for assessing convergence of the chains (for details on the integration of software packages used, the interested readers can refer to [11]).
Chapter 4

Setup Adjustment of Multiple Lots

using a Sequential Monte Carlo Method

In the previous chapter, a multiple-lot process setup adjustment method based on a Bayesian model was discussed. In that chapter, the MCMC method was used for posterior distribution computation. Although this method has excellent performance in reducing the total quadratic off-target cost, we noticed the disadvantage of time-inefficiency, also briefly discussed in Section 2.5. As an alternative to MCMC methods, in this chapter we consider the use of Sequential Monte Carlo (SMC) methods applied to the multiple lot, unknown parameters, setup adjustment problem.

The remainder of this chapter is organized as follows. Section 4.1 presents the statistical
model we consider for the multiple lot adjustment problem and techniques that have been previously proposed for the solution of its different versions. In Section 4.2 we present the proposed adjustment rule, and indicate the reasons that lead us to the particular choice of SMC techniques we recommend. This section also presents an illustrative example of the use of the proposed rule. Section 4.3 presents a study of the performance of the SMC adjustment rule compared to other rules. Then, Section 4.4 presents a modified SMC adjustment rule with improved performance. Finally, Section 4.5 gives a summary of the results of this chapter.

4.1 Process models and existing adjustment procedures

We consider a multiple-lot discrete-part manufacturing process, in which an initial offset can occur at the setup of each lot. It is assumed the process can be controlled or adjusted according to a linear control rule at each time instant before a part is processed. The statistical model can be expressed as follows:

\[ Y_{ij} = \theta_{ij} + \nu_{ij} \] (4.1)

\[ \theta_{ij} = \theta_{i(j-1)} + u_{ij} \] (4.2)

\[ \theta_{i0} \sim N(\mu, \sigma_{\theta}^2) \] (4.3)

\[ \nu_{ij} \sim N(0, \sigma_{\nu}^2) \] ; (4.4)

where:

- \( i = 1, ..., I \) is the lot or batch index;
• $j = 1, \ldots, J$ is the part index;

• $Y_{ij}$ is the deviation from the nominal value for the quality characteristic associated with the $j^{th}$ part in the $i^{th}$ batch (part $(i,j)$);

• $\theta_{ij}$ is the (unknown) mean for the $j^{th}$ part in the $i^{th}$ batch. $\theta_{i0}$ represents the initial offset due to setup errors, which is assumed to be normally distributed with mean $\mu$ and variance $\sigma^2_\theta$;

• $\nu_{ij}$ represents the random error due to the combined effect of the intrinsic variability in the machining process and the variability in the measurement system. It is assumed to be normally distributed with mean equal to zero and variance $\sigma^2_v$.

• $u_{ij}$ is the adjustment made before the $j^{th}$ part in the $i^{th}$ lot. It can be represented as the difference between the levels of two consecutive control variable setpoints: $U_{ij} = u_{ij} - u_{i(j-1)}$. Equation (4.2) can be rewritten as

$$\theta_{ij} = \theta_{i0} + u_{ij}$$

(4.5)

It will be assumed we can always set the desired quality characteristic to be 0. The performance of an adjustment rule will be evaluated by a quadratic symmetric loss function, given by:

$$C = \sum_{i=1}^{I} \sum_{j=1}^{J} Y_{ij}^2.$$  

(4.6)

Similarly to Chapter 3, there are several adjustment rules that can be applied to this problem based on different assumptions. For instance, one can apply:
• **Grubbs’ “Harmonic” rule** [27] where $U_{ij} = -\frac{1}{j} Y_{ij}$ ($U_{i0} = 0 \ \forall i$). Grubbs’s Harmonic rule is optimal if the initial offsets $\theta_{i0}$ are totally unpredictable (or $\mu = 0$ and $\sigma_\theta^2 \to \infty$ in (4.3)). But in practice, the initial offsets follow a certain distribution (assumed to be normal in (4.3)) which can be inferred and used to predict the offsets.

We can keep updating the estimates of the parameters characterizing the offset distribution while more and more parts are observed, which will improve the accuracy of the adjustments;

• **Grubbs “extended” rule** [27], characterized by $U_{ij} = -Y_{ij}/(j + \sigma_v^2/\sigma_\theta^2)$, requires knowing $\sigma_v^2/\sigma_\theta^2$ and assumes $\mu = 0$. To use this rule in practice, a pilot study needs to be conducted before the process begins from which parameter estimates are obtained. However, the estimates may be biased either due to the small samples used in the pilot study or due to the difference between the pilot study and the actual process to be controlled (the “scale up” phenomenon in manufacturing). Furthermore, leaving the pilot study uncontrolled may be unaffordable for some expensive-part manufacturers;

• **An integral or EWMA controller** [6, 15], where $U_{ij} = -\lambda Y_{ij}$ ($U_{i0} = 0 \ \forall i$) and $\lambda$ is some weight parameter (the integral constant) that needs to be chosen a priori. It is difficult to choose a proper $\lambda$ when the process parameters are not known. This controller, however, has the advantage that it remains “alert” in case further shifts or offsets occur within a lot.
• **The Bayesian adjustment rule of [11]**, in which required parameters are sequentially predicted or estimated through posterior distributions by using Markov Chain Monte Carlo (MCMC) and the adjustments are made accordingly. At each step, the computation of a posterior distribution using MCMC method is based on the original prior distribution and the whole data set up to that step. Each iteration of the MCMC updating always starts over from the beginning of the data set and does not utilize the posterior distributions obtained in the previous step. As the operation of the process goes on, the size of the data set grows and the time for the MCMC computations becomes substantial.

### 4.2 Sequential Monte Carlo methods

The adjustment rule proposed in this chapter is based on a similar Bayesian model as in the Bayesian adjustment rule of [11] (see also Chapter 3). Similarly to [11], the model described in equations (4.1-4.4) can be rewritten as a one-way random effects model, by substituting equation (4.5) in equation (4.1), thus obtaining:

\[
Y_{ij} = \theta_0 + u_{i(j-1)} + v_{ij}.
\]

Considering this last equation, the transformed variable:

\[
X_{ij} = Y_{ij} - u_{i(j-1)} = \theta_0 + v_{ij}
\]

(4.7)

can be computed each time a new part is processed (since the adjustment \(u_{i(j-1)}\) is known before observing deviation from target \(Y_{ij}\) characterizing the \(j^{th}\) part in the \(i^{th}\) batch).
Equation (4.7) together with equations (4.3-4.4) form a traditional one-way random effects model and adjustments can be made as in [11] at two levels:

1. before processing a new lot $i$ (where $i \geq 3$), the initial set-point $u_{i0}$ that has to be set on the machine before processing lot $i$ is based on the predictive distribution of the setup offset, that is:

$$U_{i0} = -\hat{\mu}_{|D_{(i-1)J}} \text{ for } i \geq 3,$$

(4.8)

or

$$u_{i0} = -\hat{\mu}_{|D_{(i-1)J}},$$

(4.9)

where $\hat{\mu}_{|D_{(i-1)J}}$ is the mean of the posterior distribution $p(\mu|D_{(i-1)J})$ and $D_{(i-1)J} = \{x_{11}, \ldots, x_{(i-1)1}, \ldots, x_{(i-1)J}\}$ represents all the (transformed) data observed before lot $i$ starts (i.e., after the last part in the previous lot $i - 1$ is observed);

2. after observing at least one part in a lot, adjustments are based on the posterior distribution of the offset in the current lot:

$$U_{ij} = -\hat{\theta}_{ij} | D_{ij}, \text{ for } j = 1, 2, \ldots, J - 1,$$

(4.10)

or

$$u_{ij} = -\hat{\theta}_{i0} | D_{ij},$$

(4.11)

where $\hat{\theta}_{i0}|D_{ij}$ is the mean of the posterior distribution $p(\theta_{i0}|D_{ij})$ and $D_{ij} = \{x_{11}, \ldots, x_{ij}\}$ represents all the (transformed) data observed before the $(j + 1)^{th}$ part in the $i^{th}$ lot is processed.
To complete the Bayesian model adopted, the prior distributions of process parameters have to be selected. In this chapter we use:

\[ \mu | \mu_0, \sigma^2_0 \sim N(\mu_0, \sigma^2_0), \]  
\[ \sigma^2_\theta | \mu_1, \sigma^2_1 \sim LN(\mu_1, \sigma^2_1), \]  
\[ \sigma^2_\upsilon | \mu_2, \sigma^2_2 \sim LN(\mu_2, \sigma^2_2), \]

where \( \mu_0, \sigma^2_0, \mu_1, \sigma^2_1, \mu_2, \sigma^2_2 \) are known constants and \( LN \) stands for a Log-Normal distribution. A Log-Normal prior for the variances is chosen because SMC does not require the use of conjugate priors. Furthermore, making a conjugate Inverse-Gamma "non-informative" is actually difficult, a fact that has not been recognized until recently [49]. In contrast, the degree of information of the prior can be easily changed using the Log-Normal. In particular, \( \mu_0 \) is usually 0 and \( \sigma^2_0 \) is made large to model a vague prior. Further discussion on how to set up the prior log-normal distributions for \( \sigma^2_\theta \) and \( \sigma^2_\upsilon \) will be presented in the next section.

To implement the adjustment rule, the posterior distributions of the process parameters need to be computed at each step when a new part is observed. In this case, the MCMC method (briefly described in the appendix of this chapter) becomes time-inefficient as the size of the data set grows. Sequential Monte Carlo methods, on the other hand, keep track of the posterior distributions previously obtained and update them into new posterior distributions only utilizing the new data collected since the previous step. In this way, the computational effort does not grow with the size of the total data set.

SMC consists in generating a set of draws or "particles" from the prior distributions of unknown parameters and associating a weight to each set of particle. These weights
are sequentially updated each time a new data is observed as described in the appendix (equations 4.23-4.23). Despite the benefits of sequential updating, this method poses some difficulties in practice. The problem arises if particles generated remain the same throughout all iterations and only the weights or the frequencies associated with the particles change. In this case, a phenomenon known as the “degeneracy” of the sample can arise. Degeneracy means that after some iterations of SMC, just few of the \( N \) original particles will have weights greater than zero. In other words, the sequential algorithm will be based on a reduced number of effective draws, thus inducing biased estimates of unknown parameters. Such degeneracy problem is particularly severe when the original prior distributions have large variances or high dimension, when the number of particles is small or when the observed data set is large (i.e. the number of iterations of the SMC method is large). The degree of this type of degeneracy due to an “impoverished” sample [2] can be monitored via the Effective Sample Size(ESS) [33, 43], described also in the appendix of this chapter.

Recently, Balakrishnany and Madigan [2] propose a One-pass Particle Filtering (1PFS) algorithm, in which a “rejuvenation” step is used to disperse the particles when ESS drops below a specific level and reduce the degeneracy. In this paper, we utilize the rejuvenation step from the 1PFS algorithm and apply it to multiple-lot setup adjustment problems for distribution computations. A detailed algorithmic description of the estimation steps in the SMC method is given in the appendix of this chapter. A computer program that implements the SMC method in the appendix is listed in Appendix A and can also be downloaded from http://www.ie.psu.edu/researchLabs/EngineeringStatistics/
Example 4.1. Consider a process described by expressions (4.1-4.4). There are $i = 20$ lots of $j = 5$ parts each. The true process parameters $\mu = 4$, $\sigma_\theta = 2$ and $\sigma_\upsilon = 4$ are used in simulating the process but they are assumed unknown for illustration of the adjustment rule. Suppose little knowledge about $\sigma^2_\theta$ and $\sigma^2_\upsilon$ is available, but it is thought unlikely that that either of these could be smaller than 1 or larger than 50. In this case, a priori 95% credibility intervals can be set using the log-normal distributions such that they equal the $(1, 50)$ interval. The prior distribution for $\mu$ is set to $N(0, 100)$.

The effect of changing the true offsets $\theta_{ij}$ under the adjustments of the SMC method (eqs. 4.8-4.11) for parts $j = 1, 2, 3, 4, 5$ in lots $i = 1, 3, 10$ can be seen in Figure 4.1. The corresponding results under the adjustments given by Grubbs’ Harmonic rule based on the same random numbers are also presented for comparison. In the first lot, the prior distributions used in the SMC method are vague and the adjustments for both rules are based on the same data set, so the process behavior under these two rules is very similar. In the third lot, the SMC method starts to adjust before the first part in that lot according to the estimate of $\mu$ based on the observations in the previous lots. This type of adjustment is the main difference between the two methods since it speeds up the convergence of the offsets to the target. The SMC adjustment rule “anticipates” the setup error. In lot 10, the offsets are not improved much by the harmonic rule since only the observations in this lot are utilized and the observation errors in this lot are large. For the SMC method, all observations in all previous lots contribute to the estimation of the offsets through the updated prior
Figure 4.1: Adjusted value of the true offset $\theta_{ij}$ under the SMC method and Grubbs’ harmonic rule for parts $j = 1, 2, 3, 4, 5$ and lots $i = 1, 3, 10$

distributions. Therefore, the quadratic costs for the five parts in this lot are significantly reduced compared to the harmonic rule. The total quadratic cost for the SMC method in this simulated run is 23% lower than the cost under the adjustments of the harmonic rule (1995 vs. 2583).

The posterior distributions on the process parameters $\mu$, $\sigma_\theta^2$ and $\sigma_\nu^2$ at the end of the simulation are plotted in Figure 4.2. The estimates of these parameters (means of the
posterior distributions) are also summarized in Table 4.1.

Figure 4.2: Plots of the posterior distributions of $\mu$, $\sigma_\theta^2$ and $\sigma_\upsilon^2$ obtained at the end of the process

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Posterior mean</th>
<th>95% Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>4</td>
<td>4.14</td>
<td>(3.13, 5.15)</td>
</tr>
<tr>
<td>$\sigma_\theta^2$</td>
<td>4</td>
<td>4.12</td>
<td>(1.40, 8.40)</td>
</tr>
<tr>
<td>$\sigma_\upsilon^2$</td>
<td>16</td>
<td>16.89</td>
<td>(13.10, 21.63)</td>
</tr>
</tbody>
</table>

Table 4.1: Means and 95% credibility intervals for the posterior distributions of the process parameters after all parts are observed.

The MCMC method proposed in [11] can also be used to adjust the simulated process, where the prior distributions are given by $\sigma_\theta^2 \sim \text{IG}(0.01, 0.01)$, $\sigma_\upsilon^2 \sim \text{IG}(0.01, 0.01)$ and $\mu \sim \text{N}(0, 10000)$ and IG represents Inverse Gamma distribution. For this specific example, the total cost under the MCMC method is 2011, which is 22% less than the cost under the Harmonic rule. 9 more replications of this process were simulated. The average saving induced by the SMC method versus the Harmonic rule is 29.8% and its frequentist 95% confidence interval based on all 10 replications is (26.6%, 33.0%). The corresponding average saving made by using MCMC method as opposed to the Harmonic rule is 25.9% and the
95% confidence interval is \((22.0\%, 29.8\%)\). The SMC method performs slightly better than the MCMC method mainly due to the prior distributions used in the SMC method are more informative and less biased. Discarding the effect of the prior distribution, the difference between the SMC method and the MCMC method is not large since they use the same Bayesian model and the difference in the computation method should not affect the result.

The main difference between MCMC and SMC, however, is the computational time. The total time the SMC method spent in adjusting a process in this example can vary from 1 to 3 minutes on a computer with a Pentium 4 CPU. In contrast, the MCMC requires a computational time in the order of hours, thus the SMC method is much more time-efficient.

### 4.3 Comparison with Grubbs’ Adjustment Rules

#### 4.3.1 Comparison with Grubbs’ Harmonic Rule

Grubbs’ Harmonic Rule is a very simple adjustment rule that does not require knowing any process parameters. Therefore, its performance will be used as a benchmark in this section. The relative savings in quadratic cost induced by the proposed SMC adjustment rule compared to Grubbs’ Harmonic rule are:

\[
S = \frac{C_h - C_s}{C_h} = 1 - \frac{C_s}{C_h},
\]

where \(C_h\) is the quadratic cost of a process adjusted by Grubbs’ Harmonic Rule and \(C_s\) is the quadratic cost of the process adjusted by using Sequential Monte Carlo methods.
To compare the proposed adjustment rule with Grubbs’ rule in a variety of production situations, 16 different cases or scenarios were simulated according to Table 4.2. The lot number $I$ is fixed at 20. Generally, the difference between the two rules are small when $I$ is small and the advantage of Bayesian models is significant when $I$ is large.

<table>
<thead>
<tr>
<th>Case</th>
<th>$J$</th>
<th>$\sigma_v$</th>
<th>$\mu$</th>
<th>$\sigma_\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>4</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>20</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>20</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 4.2: Parameters characterizing the true behavior of the process.

The prior distribution of $\mu$ is usually easy to choose. Here $\mu_0$ is set to be 0 and $\sigma_0^2$ is chosen to be 10000. To investigate the impacts of the prior distributions for $\sigma_v^2$ and $\sigma_\theta^2$ on the adjustment performance, 9 additional scenarios were created as in Table 4.3. A prior can be considered to be relatively accurate when its mode is close to the true parameter (Prior scenarios 1-3), or to be inaccurate otherwise (underestimated in scenarios 4-5, overestimated in scenarios 7-9). A higher variance gives a more vague prior (e.g. scenarios 3,6 and 9), and a smaller variance represents more confidence on the prior (e.g. scenario 1,4 and 7).
Prior mode($\sigma^2_\nu$)  mode($\sigma^2_\theta$) variance of $\sigma^2_\nu$ and $\sigma^2_\theta$
---
1  $\sigma^2_\nu$  $\sigma^2_\theta$  0.25 $\times$ mode$^2$
2  $\sigma^2_\nu$  $\sigma^2_\theta$  2 $\times$ mode$^2$
3  $\sigma^2_\nu$  $\sigma^2_\theta$  100 $\times$ mode$^2$
4  $0.5 \times \sigma^2_\nu$  $0.125 \times \sigma^2_\theta$  0.25 $\times$ mode$^2$
5  $0.5 \times \sigma^2_\nu$  $0.125 \times \sigma^2_\theta$  2 $\times$ mode$^2$
6  $0.5 \times \sigma^2_\nu$  $0.125 \times \sigma^2_\theta$  100 $\times$ mode$^2$
7  $8 \times \sigma^2_\nu$  $2 \times \sigma^2_\theta$  0.25 $\times$ mode$^2$
8  $8 \times \sigma^2_\nu$  $2 \times \sigma^2_\theta$  2 $\times$ mode$^2$
9  $8 \times \sigma^2_\nu$  $2 \times \sigma^2_\theta$  100 $\times$ mode$^2$

Table 4.3: Scenarios for prior distributions.

For each combination of the case (Table 4.2) and the prior scenario (Table 4.3, 10 replications were made. The averages of the saving rates and their (frequentist) 95% confidence intervals can be calculated. Among the 9 prior scenarios, scenario 4 is the only one that has a significant difference in average savings from the other scenarios. The prior settings on $\sigma^2_\theta$ and $\sigma^2_\nu$ may affect the convergence of the variance component estimators, but their effect on estimating the offsets is negligible. Scenario 4 is an exception where the mode of the prior distributions are smaller than the true values and the variances are extremely small. For log-normal distributions, the probability density decays very fast on the right side of the mode. So in case 4, the probability density around the true value is very low, and this requires longer time for the mean of the posterior distribution to converge to the true values.

On the other hand, the prior distribution of the initial offset at the beginning of each lot depends on the prior distributions of $\mu$ and $\sigma^2_\theta$, which is given by $\theta_0 | \mu, \sigma^2_\theta \sim N(\mu, \sigma^2_\theta)$. A prior that indicates a very small $\sigma^2_\theta$ gives extra confidence to the prior distribution of $\theta_0$ (i.e., its variance is small). This obviously can be dangerous if the difference between the
actual setup offset $\theta_0$ and the estimate of $\mu$ is large. It is hard to create definite rules on how to determine the prior distributions of the variances. One suggestion is to use a credibility interval approach as follows:

- When little prior process knowledge is available, a situation generally true for a real manufacturing process, a prior distribution can be set up such that a certain (e.g. 95%) credibility interval covers the region that the true parameter is mostly likely to be in. This was done in the example of Section 4.2. This interval can be widened if deemed necessary.

![Figure 4.3: Average saving rates induced by SMC method v.s. the harmonic rule for some selected process cases (Table 4.2 and priors (Table 4.3 combinations and their 95% confidence intervals)](image)

The results for prior scenario 1, 4 and 9 are shown in Figure 4.3. The results for the other prior scenarios are omitted since they are close to those of scenario 1 or 9. Several
conclusions on the comparison between the two adjustment rules can be made according to these results:

- In most cases when an appropriate prior is chosen (i.e., not scenario 4), the Sequential Monte Carlo method shows significant advantages against the harmonic rule, except for case 8 and 16. In these two cases, the means of the initial offsets are 0 and the large lot-to-lot variances $\sigma^2_\theta$ make the initial offsets more unpredictable, which is close to the assumption on which the harmonic rule is based, and for which it works best. On the other hand, extra errors are introduced in estimating $\mu$ by Sequential Monte Carlo method when there is a large variance $\sigma^2_\theta$, hence the total quadratic costs are increased. Even in such disadvantageous cases, the performance of the SMC rule is still close to that of the harmonic rule (mean is below 0 but the confidence intervals include 0).

- Increasing the number of parts per lot from $J = 5$ to $J = 20$ makes the savings drop. This is because the largest adjustments are made during the first few parts in a lot. However, when there are more parts in a lot, more observations can be utilized in estimating the process parameters, so the performance of SMC approach is more stable, reflected in the fact that the deviations in saving rates are smaller.

### 4.3.2 Comparison with Grubbs’ Extended Rule

In this section, a comparison between the SMC method and the Grubbs’ extended rule is conducted. For a process described by (4.1-4.4), Grubbs’ extended rule is the optimal
adjustment rule (for total quadratic off-target costs) if the following two assumptions hold true:

1. $\mu = 0$ (and we know this)

2. the ratio of two process parameters $\sigma^2_v/\sigma^2_\theta$ is known.

It was shown in Section 4.3.1 that the performance of the SMC method is not sensitive to the prior distributions of $\sigma^2_v$ and $\sigma^2_\theta$ except for a prior containing biased information with extremely high confidence. Therefore, instead of varying the priors as in Table 4.3, the SMC method was applied to the 16 process cases in Table 4.2 with the same prior distributions as used in the example in Section 4.2. The extended rule was also applied to the same process cases using common random numbers. The variance ratio was changed according to $k \times \sigma^2_v/\sigma^2_\theta$ are used. When $k = 1$, an accurate ratio is used in the extended rule; otherwise, the ratio used is biased. When the ratio is underestimated ($k < 1$), the performance of the extended rule is close to that of the harmonic rule (compare $U_{ij} = -\frac{1}{j}Y_{ij}$ v.s. $U_{ij} = -Y_{ij}/(j + \sigma^2_v/\sigma^2_\theta)$).

Therefore, in this dissertation only the results for $k \geq 1$ are shown. The comparison between the SMC method and the extended rule is illustrated in Figure 4.4, which shows the savings induced by the SMC method with respect to the extended rule using the ratio $k \times \sigma^2_v/\sigma^2_\theta$ where $k = 1, 2, 4, 8$.

We can make the following observations based on these results:

- In cases when $\mu = 4$, the first assumption that the extended rule is based on is violated.

Even if the prior knowledge on the ratio $\sigma^2_v/\sigma^2_\theta$ is correct ($k = 1$), there still exist
Figure 4.4: Average saving rates induced by the original SMC method v.s. the extended rule using different ratios $k \times \sigma^2_v/\sigma^2_\theta$ for each process case, and their 95% confidence intervals.

significant savings made by the SMC method. When the ratio used in the extended rule is incorrect, the savings are more substantial.

- In the cases that $\mu = 0$, the first assumption for the extended rule is met. If the second assumption is also met ($k = 1$), then Grubbs’ extended rule becomes optimal and the savings are negative as expected. But even in such situation, the performance of the SMC method is not far below that of the optimal rule. When the second assumption is violated, the advantage of the SMC method appears and the savings become more significant as the variance ratio used in the extended rule becomes more biased.

The SMC method provides significant savings compared to the extended rule when the prior knowledge on parameters $\mu$, $\sigma^2_\theta$ and $\sigma^2_v$ is unavailable or incorrect. But if the first assumption is true or $\mu$ is known, the extended rule is still a competitive rule even in the cases that the
ratio used for $\sigma^2_v/\sigma^2_\theta$ is slightly biased, because of its fast speed and easiness to use.

An extension of Grubbs’ extended rule to the case when $\mu \neq 0$ but $\mu$ and $\sigma^2_v/\sigma^2_\theta$ are both known, is easily derived:

- At the beginning of a lot, before the first part is processed, make the adjustment
  
  $$U_{i0} = -\mu.$$ 

- After the $i$th part is observed, adjust
  
  $$U_{ij} = -Y_{ij}/(j + \sigma^2_v/\sigma^2_\theta)$$

In Section 4.5 a procedure is suggested, in which the extended rule or its extension can be adopted following the execution of the SMC method.

### 4.4 An Improved SMC method

There are two types of process adjustments made by the SMC method. The first type of adjustment is made according to the posterior distribution of $\theta_{ij}$ right after a new part is observed. These adjustments are, similar to those suggested by the harmonic rule. The second type of adjustments are made at the beginning of a lot when no part in that lot is observed based on the estimates of $\mu$, and there is no such adjustment provided by the harmonic rule.

Therefore, in cases 8 and 16 (Table 4.2) where $\mu$ is actually 0, extra errors can be introduced by the second type of adjustments (i.e., the quadratic cost will be inflated because there is no need to adjust if the lot means are really zero). In the first investigation in this section, all the second type of adjustments were omitted. The effect of making this modifi-
cation is illustrated in Figure 4.5, where the original SMC method is denoted as method 1 and the modified method without the second type of adjustments is denoted as method 2. Once again, the cost for the harmonic rule is used as benchmark to calculate the saving rates for different methods. The same prior scenario (scenario 2 in Table 4.3) were used in the SMC-based methods compared here. The average savings and confidence intervals were obtained based on 10 replications. The advantages of the modified SMC method are significant in all cases since the extra errors made by the second type of adjustments are eliminated in the cases where $\mu = 0$. But these adjustments are beneficial if $\mu \neq 0$, so eliminating the adjustments in the cases where $\mu = 4$ makes the saving rates drop significantly.

![Figure 4.5: Average saving rates induced by the original SMC method (method 1), the modified SMC method without the second type of adjustments (method 2) and the conditional first-adjustment method (method 3) against the harmonic rule for each process case, and their 95% confidence intervals](image)

To remove the disadvantages of the second type of adjustments while keeping their ad-
vantages, a *conditional first-adjustment SMC method* is proposed. Under this method, at the beginning of lot \(i (i > 2)\), the first adjustment can only be executed when the \(1 - \alpha\) credibility region of the posterior distribution of \(\mu\) excludes 0, otherwise the first adjustment will be omitted. In this way the first adjustment in a lot will be made only when the estimate of \(\mu\) is significantly different from 0. \(\alpha\) is a significance level parameter specified before the process starts. Note that when \(\alpha = 1\), this method works identically to the original SMC method of Section 4.2. When \(\alpha = 0\) all the first adjustments will be omitted. In Figure 4.6, the maximum and the minimum average savings for different \(\alpha\) among the 16 process cases for prior scenario 2 are shown.

![Graph showing the changing of the maximum and the minimum average savings with different \(\alpha\) among 16 process cases for prior setting 2.](image)

Figure 4.6: The changing of the maximum and the minimum average savings with different \(\alpha\) among 16 process cases for prior setting 2.

By choosing a proper \(\alpha\), the saving rate can be improved in the case with the worst performance, as the saving rate in the case with the best performance only decreases by a small amount. The effect of changing \(\alpha\) may differ due to different priors and processes.
investigated. Choosing a proper $\alpha$ also depends on users' preference on the tradeoff between the maximum savings and the minimum savings. Based on our numerical experiments, we suggest using an $\alpha$ between 0.05 and 0.3. Figure 4.5 also shows the performance of the conditional first-adjustment method with $\alpha = 0.1$ (method 3) compared to the harmonic rule. Compared with original SMC method (method 1) and the modified SMC method without any second type adjustments (method 2), the improved conditional first-adjustment method has more robustness to different process cases. This method retains the substantial savings made by the original SMC method in the cases where $\mu = 4$ (e.g. process cases 13,14), while efficiently eliminating the redundant second type of adjustments in cases where $\mu$ is actually 0 (e.g. process cases 15,16).

4.5 Conclusions

When a process such as (4.1-4.4) needs to be adjusted and the prior knowledge on the process parameters is unavailable, Grubbs' harmonic rule and two adjustment methods based on Bayesian estimation (MCMC method and SMC method) are applicable. The harmonic rule does not require prior knowledge at all, but it does not take advantage of the on-line process data to estimate process parameters. Bayesian methods starting with vague priors have similar behavior in the first few lots as the harmonic rule, but they keep updating the posterior distributions of the parameters as more observations are available, from which increasingly better adjustments are made. The SMC method is more time-efficient than the MCMC method as the size of the data set grows. Grubbs’ extended rule can be an
alternative if the knowledge on parameter $\mu$ and ratio $\sigma^2_u/\sigma^2_\theta$ is available. Specifically, if a good estimate of $\mu$ is available, the extended rule is robust to a small bias in the estimate of $\sigma^2_u/\sigma^2_\theta$.

Based on the properties of these three adjustment methods, the following procedure is suggested to choose the proper method in practice:

1. If no prior knowledge of the parameters is available at all and it is very difficult to choose the prior distributions for the SMC method, then the Harmonic rule can be applied to the first 2 or 3 lots to obtain the minimum data necessary to obtain parameter estimates and set up the priors, from which the SMC method can then take over. Otherwise, the SMC method can be directly used from the beginning.

2. Adjust the process using the SMC method.

3. As the operation of the process proceeds under the SMC method, if the estimates based on the posterior distributions of $\mu$, $\sigma^2_u$ and $\sigma^2_\theta$ reach a certain level of precision, start using the extended rule (or its extension presented at the end of Section 4.3.2 to adjust the process using the estimates of $\mu$ and $\sigma^2_u/\sigma^2_\theta$ obtained in the SMC method.

The performance of such combined approach is a matter of further research.

In order to improve the SMC method and make it more robust, a conditional first-adjustment method was proposed. In this method, redundant adjustments are omitted when $\mu \simeq 0$ based on the posterior distribution of $\mu$. A computer program that implements the SMC method developed in this chapter is listed in Appendix A and can be downloaded from
http://www.ie.psu.edu/researchLabs/EngineeringStatistics/.
Appendix of Chapter 4: Markov Chain Monte Carlo

And Sequential Monte Carlo methods

Consider equations (4.8-4.11). The adjustments are selected based on the posterior distributions $p(\mu|D_{i-1}, t)$ and $p(\theta_{i0}|D_{ij})$. In this appendix we use a more general notation in which $\psi = \{\psi_1, ..., \psi_r, ..., \psi_R\}$ denotes the vector of $R$ unknown parameters in the model and $t$ represents the whole number of parts machined at a given time. In particular, if we consider the time instant in which the $j^{th}$ part of the $i^{th}$ lot has been completed, $t$ is given by $t = i \times J + j$. Consequently, $D_t = \{x_1, x_2, ..., x_t\}$ represents all the data observed at time $t$. At time $t = i \times J + j$, the vector $\psi$ is composed by all unknown parameters $\{\theta_{10}, \theta_{20}, ..., \theta_{i0}, \mu, \sigma_\theta^2, \sigma_\upsilon^2\}$.

The posterior distribution $p(\psi|D_t)$ can be computed using Bayes’ theorem:

$$p(\psi|D_t) = \frac{p(\psi)p(D_t|\psi)}{\int p(\psi)p(D_t|\psi)d\psi} = \frac{f_t(\psi)}{\int f_t(\psi)d\psi}, \quad (4.16)$$

where $p(D_t|\psi)$ is the likelihood of observing data in $D_t$ and $p(\psi)$ the prior distribution of unknown parameters. Closed form of this posterior distribution is available just under specific conditions (e.g., using conjugate analysis for simple, non-hierarchical problems).

In particular, the one-way random effect model we are dealing with, is a special type of hierarchical model. In this case no closed form is available for the posterior and estimates for unknown parameters can be computed using random draws generated from the posterior distribution.
Among well-known approaches that can be used to draw samples from posterior distributions, Markov Chain Monte Carlo (MCMC) and Sequential Monte Carlo (SMC) approaches will be briefly described in this Appendix. Markov Chain Monte Carlo [21] is a method for drawing samples by constructing a Markov Chain whose stationary distribution is the target posterior distribution we want to sample from. In particular, one of the most popular MCMC algorithm is Gibbs sampling [23], implemented in the software WinBugs [48] and used in [11] to solve the setup adjustment problem.

Gibbs sampling is a Markovian updating scheme that consists of $K$ iterations of $R$ draws, where $R$ is the number of unknown parameters. To briefly show how Gibbs sampling works, assume that we need to compute the posterior distribution after $t$ data $D_t = \{x_1, \ldots, x_t\}$ have been observed. Starting at $k = 0$ from arbitrary values of unknown parameters $\psi_1(0), \psi_2(0), \ldots, \psi_R(0)$, the first iteration ($k = 1$) in Gibbs sampling consists in sampling from the full conditional distributions as follows:

\[
\text{Draw } \psi_1(1) \text{ from } p(\psi_1|D_t, \psi_2(0), \ldots, \psi_R(0)) ; \\
\text{Draw } \psi_2(1) \text{ from } p(\psi_2|D_t, \psi_1(1), \ldots, \psi_R(0)) ; \\
\ldots \\
\text{Draw } \psi_R(1) \text{ from } p(\psi_R|D_t, \psi_1(1), \ldots, \psi_{R-1}(1)) ;
\]

thus obtaining as a result a set of points $\psi_1(1), \psi_2(1), \ldots, \psi_R(1)$ which represent the starting values for the next step.

If the number $K$ of cycles is large enough, i.e. the Markov chain reaches the steady state after $K$ steps, $N$ further iterations of the algorithm allow to compute a final set of $N$ draws.
from the distribution of each parameter of interest.

As it is clear from this brief description, MCMC can become cumbersome if data arise sequentially as in the adjustment problem described in this paper. In particular, when a new observation is obtained at time $t + 1$, all the $K + N$ draws generated to compute the posterior distributions of the parameters at time $t$ can not be re-used and the algorithm has to restart drawing $K + N$ new samples. Hence, “MCMC methods need to be restarted each time a new data is collected, making their implementation in real time impractical” (Berzuini in [21] p. 327).

An alternative solution for the Bayesian computation of posterior distributions of parameters is Sequential Monte Carlo (SMC) or Sequential Importance Sampling (SIS) [19]. This method is a sequential variant of the SIR (Sampling/Importance Resampling) procedure presented by Rubin and described by Smith and Gelfand [45]. It is also known in the literature as weighted bootstrap.

To describe how SIS works, let us start applying SIR to compute the posterior distribution of unknown parameters at time $t$, i.e., $p(\psi|D_t)$. The SIR approach consists of drawing $N$ samples, called "particles", of all unknown parameters $\psi_1, ..., \psi_R$ from a distribution $g(\psi)$ which is called the importance sampling distribution. A weight $w^{(n)} = q_t^{(n)} / \sum_{l=1}^N q_t^{(l)}$ is then computed for the $n^{th}$ draw $\psi^{(n)}$ using:

$$q_t^{(n)} = f_t(\psi^{(n)}) / g(\psi^{(n)}) ,$$

(4.21)

where $f_t(\psi^{(n)})$ is the numerator of expression (4.16) evaluated at the $n^{th}$ particle $\psi^{(n)}$. If we now sample a draw from the discrete distribution $\psi^{(1)}, ..., \psi^{(n)}, ..., \psi^{(N)}$ allocating mass
Consider now the special case in which the importance sampling function \( g(\psi) \) is the prior distribution \( p(\psi) \) in equation (4.16). Since \( f_t(\psi) \) is the product of the likelihood times the prior (equation 4.16) in this case, the weight \( q_t^{(n)} \) given by (4.21) can be rewritten as:

\[
q_t^{(n)} = \frac{f_t(\psi^{(n)})}{p(\psi^{(n)})} = \frac{p(D_t|\psi^{(n)})p(\psi^{(n)})}{p(\psi^{(n)})} = p(D_t|\psi^{(n)}) .
\]

As showed by [45], this resampling strategy simply means that more weight is given to prior samples that are more "likely" to happen.

The Sequential Importance Sampling (SIS) approach [19], used in this paper, is a sequential variant of the SIR algorithm. When data arise sequentially, SIS allows to use past particles generated before observing the last outcome from the process. To show how SIS works, consider the sequential nature of Bayes’ theorem when \( x_{t+1} \) has been just observed:

\[
p(\psi|D_{t+1}) = \frac{p(\psi|D_t)p(x_{t+1}|\psi)}{\int p(\psi|D_t)p(x_{t+1}|\psi) d\psi} ,
\]

(4.22)

where \( p(\psi|D_t) \) is the posterior computed before observing \( x_{t+1} \), and \( p(x_{t+1}|\psi) \) is the likelihood of the last observation \( x_{t+1} \). Therefore, if we use \( p(\psi|D_t) \) as the importance sampling distribution instead of the prior, the weights at time \( t + 1 \) can be computed as a function of weight at time \( t \) as follows:
\[ q_{t+1}^{(n)} = \frac{p(x_{t+1}|\psi^{(n)})w_t^{(n)}}{q_{t+1}}, \quad (4.23) \]

\[ w_{t+1}^{(n)} = q_{t+1}^{(n)}\sum_{l=1}^{N} q_l^{(l)} \quad (n = 1, \ldots, N), \quad (4.24) \]

where the weight associated with the \( n^{th} \) particle changes accordingly to the likelihood of observing \( x_{t+1} \) evaluated at each particle \( \psi^{(n)} \). This expression allows to sequentially update the weights each time a new data is observed and is thus applicable in situations in which data arise sequentially.

Problems associated with the described SMC approach concerns “degeneracy” of particles. When the \( N \) particles are generated from noninformative priors, most of these particles (the ones that are less likely, given data observed) will have weights equal to zero after few iterations. In this cases, particles associated with weights greater than zero will be very few, say \( N^* \ll N \). This “degeneration” of the initial number of \( N \) particle into a much smaller number \( N^* \) will deeply affect the effectiveness of the SMC approach, inducing biased estimates of the unknown parameters. Such problem is severe when the original prior distributions have large variances or high dimension, the number of particles is small or when the observed data set is large (i.e. the number of iterations of the SMC method is large).

The degree of this type of degeneracy due to an “impoverished” sample [2] can be monitored via the Effective Sample Size (ESS) [33, 43], that at time index \( t \), is given by:
\[ \text{ESS}_t = \frac{N}{1 + N^2 \text{var}(w_t^{(n)})}, \] 

(4.25)

where \( N \) is the original number of particles and \( w_t^{(n)} \) is the weight associated to particle \( n \) at time \( t \). In expression (4.25), the degree of “degeneracy” is described by the variance of weights \( w_t^{(n)} \) at time \( t \). In fact if at time \( t \) just few particles have associated weights greater than zero, the variance \( \text{var}(w_t^{(n)}) \) will be small and the \( \text{ESS}_t \) will tend to \( N \).

To overcome the “degeneracy” problem, a rejuvenation step of the \( N \) particles can be performed using the “1PFS” algorithm described in [43]. The original 1PFS algorithm can not handle the posterior distribution for the variance parameters, where the density is positive only in the positive region. The SMC algorithm used in this research is a modified algorithm based on the original version. It can compute the posterior distributions of the the positive parameters such as \( \sigma^2_\theta \) and \( \sigma^2_\upsilon \), in which the rejuvenation procedures take place in the logarithm space of the variance parameters.

The final SMC algorithm (including the test on degeneracy and possible rejuvenation step), adopted for the multiple-lot setup adjustment problem is summarized in the following.

**Algorithm for estimating the process parameters \( \mu, \theta_0, \sigma^2_\theta \) and \( \sigma^2_\upsilon \)**

At the beginning of the process:

draw \( N \) random numbers \( \mu^{(n)} \) from the prior distribution of \( \mu \);

draw \( N \) random numbers \( \sigma^2_\theta^{(n)} \) from the prior distribution of \( \sigma^2_\theta \);
draw N random numbers $\sigma_v^{2(n)}$ from the prior distribution of $\sigma^2_v$, $n = 1, 2, ..., N$;

Create an initial weight vector $(w_1, w_2, w_3, ..., w_N)$, where $w_n = 1/N$ for all $n$.

Iterations throughout all lots/parts processed:

For $i = 1, 2, 3, ..., I$

Generate 1 random number $\theta_{i0}^{(n)}$ from the distribution $N(\mu^{(n)}, \sigma^{2(n)}_\theta)$ for each $n$.

For $j = 1, 2, 3, ..., J$

Obtain the new observation $y_{ij}$ and calculate the new variable $x_{ij} = y_{ij} - u_{i(j-1)}$.

Calculate the likelihood of the $n^{th}$ particle (combination $(\mu^{(n)}, \theta_{i0}^{(n)}, \sigma^{2(n)}_\theta, \sigma_v^{2(n)})$): 

$$L_n = \frac{1}{\sqrt{\sigma_v^{2(n)}}} \exp\left\{ -\frac{(x_{ij} - \theta_{i0}^{(n)})^2}{2\sigma_v^{2(n)}} \right\}$$

Update the new weight vector and normalize it:

$$w_n \leftarrow w_n \times L_n$$

Then

$$w_n \leftarrow w_n / \sum_{m=1}^{N} w_m.$$ 

Obtain the new parameter estimators:

$$\hat{\theta}_{i0}|D_{ij} = \sum_{n=1}^{N} w_n \theta^{(n)}$$

and

$$\hat{\theta}_{ij}|D_{ij} = \hat{\theta}_{i0}|D_{ij} - u_{ij}.$$ 

$$\hat{\mu}|D_{ij} = \sum_{n=1}^{N} w_n \mu^{(n)}$$

$$\hat{\sigma}_v^{2}|D_{ij} = \sum_{n=1}^{N} w_n \sigma_v^{2(n)}$$

$$\hat{\sigma}_\theta^{2}|D_{ij} = \sum_{n=1}^{N} w_n \sigma_\theta^{2(n)}$$

Calculate the effective sample size factor:

$$ESS = \frac{N}{1+N^2\text{var}(w_n)}$$

If $ESS < pN$, where $p$ is a specified level between 0 and 1 ($p = 0.5$ was used in our computations),
rejuvenate $N$ particles using the “1PFS” algorithm (reference [2]).
Chapter 5

Setup Adjustment Under Fixed Adjustment Cost

5.1 Introduction

In the previous two chapters, the setup adjustment problem for a multiple-lot process was considered. In this chapter, we consider the setup adjustment problem for a single-lot process, and provide an adjustment method to minimize the expected value of the sum of quadratic off-target costs and fixed adjustment costs. Apart of the unknown initial offset, the process is supposed to be in a state of statistical control, so the process model is applicable to discrete-part production processes. The process variance is also assumed unknown.
5.2 Process Model, Cost Criterion and Parameter Estimates

Suppose $N$ parts are processed sequentially in time and suppose the machine setup results in an unknown initial offset $\theta_0$, which, if unadjusted, will affect all $N$ parts. Let us denote by $Y_i$ the deviation from target of the quality characteristic of interest for the $i^{th}$ part. We assume a controllable factor $u_i$ exists to adjust the process in a linear manner, and denote by $U_i = u_{i+1} - u_i$ the adjustment made before processing part $i+1$. The adjustment $U_i$ is then computed based on observations obtained up to and including part $i$. Then, the equations describing the observed quality characteristic can be expressed as follows:

$$Y_i = \theta_i + \upsilon_i \quad (5.1)$$
$$\theta_i = \theta_{i-1} + U_{i-1}, \quad i = 1, 2, ..., N \quad (5.2)$$

where $\upsilon \overset{iid}{\sim} N(0, \sigma^2_\upsilon)$ and $\sigma_\upsilon$ is unknown. We assume there exists a quadratic and symmetric off-target cost. Furthermore, we assume a fixed cost $c$ is incurred whenever an adjustment $U_i$ is non-zero. Under these assumptions, the expected loss function is

$$L = E\left\{ \sum_{i=1}^{N} (Y_i^2 + c\delta(U_{i-1})) \right\}, \quad (5.3)$$

where $\delta(x) = 1$ if $x \neq 0$ and $\delta(x) = 0$ otherwise. Following Crowder [13], the cost $c$ can be understood as the ratio of the adjustment cost to the per unit quadratic off-target cost. The objective is to find the optimal adjustments $\{U_i\}_{i=1}^{N}$ based on the observed deviations from target in order to minimize the cost criterion in (5.3).
We use a Bayesian model to make inferences about the two unknown parameters, \( \theta_0 \) and \( \sigma_v \), and be able to adjust the process based on these inferences. For this purpose, a useful model, discussed by Gelman et al. [25], is the two-parameter normal conjugate model with prior distributions given by:

\[
\theta_0 | \sigma_v \sim N(\mu_0, \sigma_v^2/\kappa_0) \quad (5.4)
\]

\[
\sigma_v^2 \sim Inv - \chi^2(\nu_0, \sigma_0^2) \quad (5.5)
\]

where the second distribution is a scaled inverse chi-square distribution with degrees of freedom \( \nu_0 \) and scale parameter \( \sigma_0 \). Notice that the \( Inv - \chi^2(\nu_0, \sigma_0^2) \) distribution is equivalent to a \( Inv - gamma(\alpha = \nu_0/2, \beta = \nu_0\sigma_0^2/2) \) distribution. We choose the mathematical expression of \( Inv - \chi^2 \) distribution in this chapter and chapter 6 due to the convenience in the regression relationship formulation as in equations (5.12)-(5.13). The joint model is frequently denoted \( (\theta_0, \sigma_v^2) \sim N - Inv - \chi^2(\mu_0, \sigma_0^2/\kappa_0; \nu_0, \sigma_0^2) \). Note that the parameters \( \theta_0 \) and \( \sigma_v^2 \) are assumed dependent in their joint prior distribution such that a large value of \( \sigma_v^2 \) induces a high-variance in the prior distribution of the initial mean \( \theta_0 \). As discussed by Gelman et al. [25], this is useful since prior beliefs about \( \theta_0 \) can be calibrated by the scale of measurement of \( Y \) so that they are equivalent to \( \kappa_0 \) prior measurements on this scale. This facilitates the assessment and incorporation of prior information, if any. Alternatively, these conjugate priors can also be set to be noninformative or “objective”.

The Bayesian model on the unknown parameters allows to make predictions on the quality characteristic \( Y \), which is directly observable in contrast to the unknowns \( \theta_0 \) and \( \sigma_v \). Our interest is to make inferences in future values of \( Y \) at time \( i \), for which we need to obtain an
expression for the posterior predictive density of $Y$. For this purpose, we use the conjugate prior distributions in (5.4-5.5) and equations (5.1-5.2) to get the posterior densities:

$$Y_i|\theta_i, \sigma^2 \sim N(\theta_i, \sigma^2)$$  (5.6)

or alternatively

$$Y_{i+1}|\theta_i, \sigma^2 \sim N(\theta_i + U_i, \sigma^2),$$  (5.7)

where

$$(\theta_i, \sigma^2)|Y^{(i)}, U^{(i)} \sim N - Inv - \chi^2(\mu_i, \sigma_i^2 / \kappa_i; \nu_i, \sigma_i^2).$$  (5.8)

Here, $Y^{(i)}$ and $U^{(i)}$ are sets containing all observations and adjustment values, respectively, through the time when part $i$ is finished and observed.

As shown in the appendix of this chapter, the posterior predictive density $f(Y_{i+1}| Y^{(i)}, U^{(i)})$ is obtained from integrating

$$\int_{0}^{\infty} \int_{-\infty}^{\infty} f(Y_{i+1}| \theta_i, \sigma^2) f(\theta_i, \sigma^2 | Y^{(i)}, U^{(i)}) d\theta_i d\sigma^2.$$  

This density is characterized by the four state variables $(\mu_i, \kappa_i, \nu_i, \sigma_i^2)$ and will be denoted by

$$f(Y_i| Y^{(i)}, U^{(i)}) = \Psi(\cdot | \mu_i, \kappa_i, \nu_i, \sigma_i^2)$$  (5.9)

where the mean of the distribution is adjusted by $U_i$. It is not difficult to show (see the appendix) that the posterior density $\Psi$ is a Student t distribution with $v_i$ degrees of freedom, mean equal to $\mu_i + U_i$, and scale parameter equal to $\sigma_i^2(k_i + 1)/k_i$. This is a valuable result we use in the next section to determine an optimal setup adjustment policy.
A nice feature of the two-parameter normal conjugate model is that the four state variables \((\mu_i, \kappa_i, \nu_i, \sigma_i^2)\) are easily updated recursively. From expressions in Gelman et al. [25] in conjunction with the posterior density (5.8) written at time \(i - 1\) (used as the prior density at time \(i\)), we obtain the recursive updating expressions:

\[
\begin{align*}
\mu_i &= \frac{\kappa_{i-1}}{\kappa_{i-1} + 1} (\mu_{i-1} + U_{i-1}) + \frac{1}{\kappa_{i-1} + 1} y_i \\
\kappa_i &= \kappa_{i-1} + 1 \\
\nu_i &= \nu_{i-1} + 1 \\
\nu_i \sigma_i^2 &= \nu_{i-1} \sigma_{i-1}^2 + \frac{\kappa_{i-1}}{\kappa_{i-1} + 1} (y_i - \mu_{i-1} - U_{i-1})^2,
\end{align*}
\]  

(5.10)

(5.11)

(5.12)

(5.13)

Where \(y_i\) is the observed value of \(Y_i\) at the end of period \(i\). In these equations it is assumed an adjustment of magnitude \(U_i\) modifies the updated mean after each observation \(y_i\) is obtained. The expressions will be used in Section 5.3 to build a Dynamic Programming formulation to solve the setup adjustment problem.

### 5.3 Dynamic programming solution to the setup adjustment problem

Suppose at stage \(i\), i.e. after part \(i\) is observed and before part \(i + 1\) is processed, the posterior predictive density in (5.9) is obtained by updating four state variables \(\kappa_i, \nu_i, \mu_i\) and \(\sigma_i^2\). Assuming \(\kappa_0\) and \(\nu_0\) are fixed when the prior information is given, the \(\kappa_i\)'s and \(\nu_i\)'s
are all constants determined by the stage index $i$, that is

$$\kappa_i = \kappa_0 + i, \ \nu_i = \nu_0 + i, \ \text{for} \ \forall \ i. \quad (5.14)$$

Therefore, only two state variables, $\mu_i$ and $\sigma_i^2$, are needed to describe the predictive density.

Two obvious properties of a Student t random variable $X \sim \Psi(\cdot|\mu, \kappa, \nu, \sigma^2)$, used in what follows, are that it is symmetric and has decreasing tails, namely:

$$\Psi(x|\mu, \kappa, \nu, \sigma^2) = \Psi(-x|\mu, \kappa, \nu, \sigma^2)$$

$$\Psi(x_1|\mu, \kappa, \nu, \sigma^2) > \Psi(x_2|\mu, \kappa, \nu, \sigma^2), \ \text{if} \ |x_1 - \mu| < |x_2 - \mu|. \quad (5.15)$$

In addition, its expectation and variance are given by

$$E\{X\} = \mu$$

$$Var\{X\} = \sigma^2 \frac{\nu}{\nu - 2} (1 + \frac{1}{\kappa}) \div V(\kappa, \nu, \sigma^2), \ \text{for} \ \nu > 2. \quad (5.16)$$

Since the state variables $\kappa$ and $\nu$ are uniquely defined by stage index $i$, we introduce the simpler notation

$$V(\kappa_i, \nu_i, \sigma^2) = V_i(\sigma^2)$$

and

$$\Psi(\cdot|\mu_i, \kappa_i, \nu_i, \sigma^2) = \Psi_i(\cdot|\mu, \sigma^2). \quad (5.17)$$

Define $R_i(\mu_i, \sigma_i^2)$ to be the minimum cost from parts $(i + 1)$ to $N$ given the current density $\Psi_i(\cdot|\mu_i, \sigma_i^2)$. Then $R_0(\mu_0, \sigma_0^2)$ is the expected cost of the optimal solution we seek to the problem.
At the boundary, i.e., at stage $N - 1$, we have that $Y_N | y_{N-1}, u_{N-1} \sim \Psi_{N-1}(|\mu_{N-1} + U_{N-1}, \sigma_{N-1}^2)$, so

$$R_{N-1}(\mu_{N-1}, \sigma_{N-1}^2) = \min_{U_{N-1}} E\{Y_N^2 + c\delta(U_{N-1})\}$$

$$= \min_{U_{N-1}} \{V(\kappa_{N-1}, \nu_{N-1}, \sigma_{N-1}^2) + (\mu_{N-1} + U_{N-1})^2 + c\delta(U_{N-1})\}$$

$$= V_{N-1}(\sigma_{N-1}^2) + \min\{\mu_{N-1}^2, c\}. \quad (5.18)$$

The optimal adjustment which minimizes this last expression is clearly

$$U_{N-1} = \begin{cases} -\mu_{N-1} & \text{if } |\mu_{N-1}| > c^{1/2} \\ 0 & \text{if } |\mu_{N-1}| \leq c^{1/2} \end{cases} \quad (5.19)$$

For stage $i < N - 1$, we recursively define by using backwards induction

$$R_i(\mu_i, \sigma_i^2) = \min_{U_i} \{V_i(\sigma_i^2) + (\mu_i + U_i)^2 + c\delta(U_i) + E\{R_{i+1}(\mu_{i+1}, \sigma_{i+1}^2)|U_i}\}, \quad (5.20)$$

where

$$E\{R_{i+1}(\mu_{i+1}, \sigma_{i+1}^2)|U_i\} = \int R_{i+1}(\mu_{i+1}, \sigma_{i+1}^2)f(Y_{i+1}|Y_i, U_i)dY_{i+1}$$

$$= \int R_{i+1}(\mu_{i+1}, \sigma_{i+1}^2)\Psi_i(\mu_i + U_i, \sigma_i^2)dY_{i+1}, \quad (5.21)$$

and from (5.10) and (5.13),

$$\mu_{i+1} = \frac{\kappa_i(\mu_i + U_i) + Y_{i+1}}{\kappa_i} \quad (5.22)$$

$$\sigma_{i+1}^2 = \frac{\nu_i \sigma_i^2 + \frac{\kappa_i}{\kappa_{i+1}}(Y_{i+1} - \mu_i - U_i)^2}{\nu_{i+1}}. \quad (5.23)$$
The optimal adjustment rule or controller $U^*_i$ depends on $\mu_i$ and $\sigma_i^2$, and is the optimal solution to (5.20). For each possible $U_i$, the expected cost can be calculated numerically using backwards induction and Monte Carlo integration. $U^*_i$ is chosen to be the $U_i$ yielding minimum cost, and this is a numerically intensive computation. The computational complexity can be reduced by narrowing the candidates $U_i(\mu_i, \sigma_i^2)$ we evaluate.

### 5.3.1 Deadband structure of the optimal adjustment policy

Let $R'_i(\mu_i, \sigma_i^2)$ be the minimum expected cost for parts $i + 1$ to $N$ given that no adjustment will be made for part $i + 1$. Since an adjustment will not affect $\sigma_i$, $R_i$ (minimum cost if an adjustment is made) can be rewritten in terms of $R'_i$:

$$R_i(\mu_i, \sigma_i^2) = \min_{U_i} \{ c\delta(U_i) + R'_i(\mu_i + U_i, \sigma_i^2) \}$$  

(5.24)

If we can find $\mu_i = \mu^*_{i,\sigma_i}$ that minimizes $R'_i(\mu_i, \sigma_i^2)$ for given $\sigma_i$, (5.24) can be rewritten as:

$$R_i(\mu_i, \sigma_i^2) = \min \{ R'_i(\mu_i, \sigma_i^2), c + R'_i(\mu^*_{i,\sigma_i}, \sigma_i^2) \}.$$  

(5.25)

The optimal controller is thus

$$U^*_i(\mu_i, \sigma_i^2) = \begin{cases} 
\mu^*_{i,\sigma_i} - \mu_i & \text{if } R'_i(\mu_i, \sigma_i^2) - R'_i(\mu^*_{i,\sigma_i}, \sigma_i^2) > c \\
0 & \text{o.w.}
\end{cases}$$  

(5.26)

Thus the decision is made based on a tradeoff between the adjustment cost and the cost savings gained by adjusting at the current stage. In such manner, we can find the optimal adjustment for each state $(\mu_i, \sigma_i)$ by numerically calculation.
However, the optimal controller (5.26) is not in a "deadband" form, which would be easier to implement. To get a deadband form for the optimal controller, consider two states at stage \(i\), \((\mu_i = x_1, \sigma_i^2)\) and \((\mu_i = x_2, \sigma_i^2)\), where the posterior densities are the same except for their means. It is easy to show that \(R'_i(x_1, \sigma_i^2) = R'_i(x_2, \sigma_i^2)\) when \(|x_1| = |x_2|\) due to the symmetries in the process and in the cost structure. Suppose \(|x_1| > |x_2|\), i.e., the first state represents a process with a larger expected offset while both processes have the same offset variance. The expected offset of the first process will either remain or be brought down at some cost, so it has a higher expected cost than the process at the second state. Hence \(R'_i(x_1, \sigma_i^2) > R'_i(x_2, \sigma_i^2)\), i.e., \(R'_i(\mu_i, \sigma_i^2)\) is symmetric around \(\mu_i = 0\) and increasing with \(|\mu_i|\). Following (5.25) and (5.26) we have

\[
R_i(\mu_i, \sigma_i^2) = \min \{ R'_i(\mu_i, \sigma_i^2), c + R'_i(0, \sigma_i^2) \} \tag{5.27}
\]

and the optimal controller is

\[
U^*_i(\mu_i, \sigma_i^2) = \begin{cases} 
-\mu_i & \text{if } |\mu_i| > \alpha_{i,\sigma_i} \\
0 & \text{if } |\mu_i| \leq \alpha_{i,\sigma_i}
\end{cases} \tag{5.28}
\]

where the adjustment or action limit \(\alpha_{i,\sigma_i}\) is a number depending on \(\sigma_i\) such that \(R'_i(\alpha_{i,\sigma_i}, \sigma_i^2) = c + R'_i(0, \sigma_i^2)\).

The adjustment rule is therefore of the deadband type [5, 6, 13]. This means that only for process states with a mean far enough from target an adjustment is justified. The action
limits defining the deadband are clearly a function of the adjustment cost \( c \).

According to the foregoing discussion, only two choices are compared for \( U^*_i \) in calculating each \( R_i(\mu_i, \sigma_i^2) \). During the backwards calculation, \( \alpha_{i, \sigma_i} \) can be obtained by finding the minimum \( |\mu_i| \) with \( |U^*_i(\mu_i, \sigma_i^2)| > 0 \). Once the limits \( \alpha_{i, \sigma_i} \) are obtained, the control policy in (5.28) can be applied to adjust the initial offset for the process described by (5.1)-(5.2).

### 5.4 Computer implementation of the solution

The complete procedure to adjust process (5.1-5.2) with the optimal controller (5.28) can be divided into two steps: i) generating the control table with the adjustment limits, a computation that can be done off-line, that is, prior to start producing parts, and ii) on-line adjustment of the process. We now explain each of these two steps in detail. An R program that implements this procedure (called chart.R) is listed in Appendix B and can be downloaded from


#### 5.4.1 Generating the control table

The state variables \((\mu_i, \sigma_i)\) belong to an unbounded continuous space \( R \oplus R^+ \) where \( \oplus \) denotes Cartesian set product. Such state space need to be mapped to a bounded discrete space \( D \) with finite elements so that for each \((\mu_i, \sigma_i) \in D, R_i(\mu_i, \sigma_i) \) can be approximated numerically by backwards induction. A typical form of \( D \) will be \( D = D_\mu \oplus D_\sigma \), where
\( D_\mu = \{ \mu | \mu = i \times d_\mu, i = -n_1, -n_1 + 1, \ldots, -1, 1 \} \) and \( D_\sigma = \{ \sigma | \sigma = i \times d_\sigma, i = 0, 1, \ldots, n_2 \} \).

d_\mu \) and \( d_\sigma \) are positive numbers and \( n_1, n_2 \) are positive integers. A state \((\mu, \sigma)\) is mapped to \((\mu', \sigma') \in D\) such that \( \mu' \) is the closest element to \( \mu \) in \( D_\mu \) and \( \sigma' \) is the closest element to \( \sigma \) in \( D_\sigma \). Increments \( d_\mu \) and \( d_\sigma \) are set to be small enough for accuracy. Upper bounds \( n_1 \times d_\mu \) and \( n_2 \times d_\sigma \) are set to be large enough in order to include most of the possible states that can occur.

The steps to generate the control table are summarized as follows:

- **Step 1.** Specify the value of \( c \), the relative adjustment cost. Set the prior state variables, \( \kappa_0, \nu_0, \mu_0 \) and \( \sigma_0 \), according to the prior information of the process. A proper selection of the prior distribution is discussed in the next section.

- **Step 2.** Determine the sample state space \( D \). The increments and upper bounds can vary for different processes, cost structure or accuracy requirements.

- **Step 3.** Calculate \( R_i(\mu_i, \sigma_i) \) for each \((\mu_i, \sigma_i) \in D\) backwards for each stage \( i \), by computing (5.20) using Monte Carlo integration.

During Step 3, for each \( \sigma_i \in D_\sigma \) and each \( i \), \( \alpha_{i,\sigma_i} \) is determined by setting it to be the smallest
positive $\mu_i \in D_\mu$ such that $R'_i(\mu_i, \sigma_i) \geq c + R'_i(0, \sigma_i)$. Thus a control table containing the $\alpha_{i, \sigma_i}$’s is obtained.

**Example 5.1.** To illustrate the procedure above, consider the following example. Suppose $N = 10$ parts are to be processed. The adjustment cost is $c = 9$. The initial prior state variables are set at $\kappa_0 = 1$, and $\nu_0 = 2.01$. The discretized state variable space $D$ is chosen to be $\{\mu, \sigma | \mu = 0.1a, \sigma = k\}$, where $a$ is an integer such that $|a| \leq 50$, and $k$ is a nonnegative integer such that $k \leq 10$. The control table generated from using the procedure described above is shown in Table 5.1.

<table>
<thead>
<tr>
<th>Part $i$</th>
<th>$\sigma_i$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>1.0</td>
<td>1.9</td>
<td>2.3</td>
<td>2.6</td>
<td>2.7</td>
<td>2.8</td>
<td>2.9</td>
<td>2.9</td>
<td>3.0</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>1.0</td>
<td>1.5</td>
<td>2.0</td>
<td>2.3</td>
<td>2.5</td>
<td>2.6</td>
<td>2.7</td>
<td>2.8</td>
<td>2.9</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>1.1</td>
<td>1.3</td>
<td>1.8</td>
<td>2.0</td>
<td>2.2</td>
<td>2.4</td>
<td>2.5</td>
<td>2.6</td>
<td>2.7</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>1.2</td>
<td>1.3</td>
<td>1.6</td>
<td>1.9</td>
<td>2.1</td>
<td>2.2</td>
<td>2.3</td>
<td>2.4</td>
<td>2.5</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>1.3</td>
<td>1.3</td>
<td>1.5</td>
<td>1.7</td>
<td>1.9</td>
<td>2.1</td>
<td>2.2</td>
<td>2.3</td>
<td>2.4</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>1.4</td>
<td>1.4</td>
<td>1.5</td>
<td>1.6</td>
<td>1.8</td>
<td>2.0</td>
<td>2.1</td>
<td>2.2</td>
<td>2.3</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>1.5</td>
<td>1.6</td>
<td>1.6</td>
<td>1.7</td>
<td>1.8</td>
<td>1.9</td>
<td>2.0</td>
<td>2.1</td>
<td>2.2</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.9</td>
<td>2.0</td>
<td>2.1</td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.3</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1: Control table showing the adjustment limits $\alpha_{i, \sigma_i}$ calculated for the example.

As shown in Table 5.1, the limits $\alpha_{i, \sigma_i}$ are generally increasing with $\sigma_i$, reflecting the fact that a higher $\sigma_i$ value provides less precision in the process mean or offset $\theta_i$. This can be better seen from the unconditional variance of the process mean, $Var(\theta_i)$, which equals to:

$$Var(\theta_i) = E[Var(\theta_i|\sigma_v)] + Var(E[\theta_i|\sigma_v]) = E[\sigma_v^2/\kappa_i] + Var(\mu_i) = \frac{\sigma_i^2}{\kappa_i} \frac{\nu_i}{\nu_i - 2}.$$

(5.29)
Figure 5.1: Adjustment limits $\alpha_{i,\sigma_i}$ for all parts $i$ and $\sigma_i = 1, 4, 8$ in the numerical example.

The “U” shape of the deadband implies that the optimal controller is more likely to suggest adjusting when the precision is higher. This fact is also reflected by another tendency observed from the table, that for a fixed $\sigma_i$, $\alpha_{i,\sigma_i}$ is “U” shaped when seen as a function of the part number $i$. This implies the adjustment limits are “U” shaped. This is explained as follows (see also Figure 5.1). The more observations, the higher the precision of the predictive density of the process mean. Since $\kappa_i$ increases linearly as a function of $i$, the variance $Var(\theta_i)$ is decreasing with $i$. However, the role of the precision becomes unimportant when the process approaches its end, when high precision is already obtained by the large number of observations and it can not be significantly increased further by simply increasing the
number of observations. At that point, all $\alpha_{i,\sigma_i}$ limits are close to each other for a given stage $i$ regardless of $\sigma_i$. This is so because as $\kappa_i$ gets larger it reduces the differences in precision obtained from the different $\sigma_i$'s. Near the end, if $\sigma_i$ is fixed, $\alpha_{i,\sigma_i}$ is increasing with $i$ because the future benefit of an adjusted process is not justified, given there are few parts to produce until the end of the batch of $N$ parts. A similar “funneling out” of the adjustment limits due to an end of horizon effect was reported by Crowder [13] for a different process model with known parameters.

5.4.2 Adjusting a Process with the Control Table

Once the control table is created, the control policy described in (5.28) can be applied to adjust the initial offset of a process. At each stage $i$ of the process, $i = 1, 2, ..., N - 1$, the following procedure is repeated.

- Step 1. Update the state variables $(\mu_i, \sigma_i)$ as in (5.10-5.13) given the new observation $y_i$.

- Step 2. Map $\sigma_i$ to $\sigma'_i \in D_\sigma$, i.e., find the $\sigma'_i$ on the control table closest to $\sigma_i$.

- Step 3. If $|\mu_i| > \alpha_{i,\sigma'_i}$, adjust the process by $U_i = -\mu_i$; otherwise make no adjustment $(U_i = 0)$.
Example 5.1 (cont.) An example is shown to illustrate the procedure to adjust a process with a generated table. The process is simulated such that it has 10 parts and the initial offset is $\theta_0 = 6$. Assume the adjustment cost is $c = 9$ and the prior state variables are $\mu_0 = 0$, $\sigma = 10$, $\kappa_0 = 1$, and $\nu_0 = 2.01$. The same discrete state space $D$ is chosen as in section 5.4.1, so table 5.1 is used to adjust the process at each stage.

At stage 0, i.e. at the beginning of the process, the prior state variable $\mu_0 = 0$ implies that no adjustment is necessary. Therefore, $U_0 = 0$.

At stage 1, the first part is made and observed, and suppose we get $Y_1 = 4.28$. The new current state variables are:

\[
\begin{align*}
\kappa_1 &= 2, \\
\nu_1 &= 3.01 \\
\mu_1 &= \frac{\kappa_0}{\kappa_0 + 1} \times 0 + \frac{1}{\kappa_0 + 1} Y_1 = 2.14 \\
\sigma_1 &= \sqrt{(\nu_0 \sigma_0^2 + \frac{\kappa_0}{\kappa_0 + 1} Y_1^2)/\nu_1} = 8.36
\end{align*}
\]

The closest element to $\sigma_1$ in $D_\sigma$ is $\sigma'_1 = 8$. Referring to table 5.1, $\alpha_{1,8} = 2.8$ is larger than $\mu_1$. Therefore, adjusting is not justified, so $U_1 = 0$.

At Stage 2, suppose we observe $Y_2 = 6.70$. The updated state variables are:

\[
\begin{align*}
\kappa_2 &= 3, \\
\nu_2 &= 4.01 \\
\mu_1 &= \frac{\kappa_1}{\kappa_1 + 1} \mu_1 + \frac{1}{\kappa_1 + 1} Y_2 = 3.66 \\
\sigma_2 &= \sqrt{(\nu_1 \sigma_1^2 + \frac{\kappa_1}{\kappa_1 + 1} Y_2 - \mu_0^2)/\nu_1} = 7.47
\end{align*}
\]

In this case, $\sigma'_2 = 7$ in $D_\sigma$ is the closest element to $\sigma_2$, and we find $\alpha_{2,7} = 2.6$ in the control table. Since $\mu_2 > \alpha_{2,7}$, we should adjust the process by $U_2 = -3.66$. 
Continuing in this form, at each stage from 3 to 9, we update the state variables and decide whether to adjust or not according to the control table. An illustration after ten hypothetical observation is shown in Figure 5.2. The figure shows all observations and how the true means $\theta_i$ change due to the adjustments.

![Figure 5.2: Plot of the true means and observations](image)

As we can see in the figure, for this specific sample process, only two adjustments are made at times 2 and 5. Figure 5.3 shows an “adjustment plot” which indicates when the process mean $\mu_i$ exceeds the adjustment limits $\alpha_{i,\sigma_i}$. As it can be seen, the adjustments efficiently drive the true mean closer to target.
5.5 Sensitivity Analysis and Extremal Situations

Thus far, only one example has shown how the adjustment rule works. In this section, more cases are investigated and for each case more replications are simulated to evaluate the average performance of the optimal policy. In each of the next subsections the sensitivity of the optimal policy with respect to different process and cost parameters is investigated.
5.5.1 System Parameters: $\theta_0, \sigma_v$

To show the impact of the proposed adjustment rule on a real processes, the control table for the example in Section 5.4.1 is applied to different types of processes with 10 parts and fixed adjustment cost $c = 9$, assuming the prior state variables are the same for all cases, $\mu_0 = 0$, $\sigma = 10$, $\kappa_0 = 1$, $\nu_0 = 2.01$. For each type of process, 1000 replications are made to estimate the expected total cost. The results are summarized in Table 5.2, where $\theta_0$ is the true initial offset and $\sigma_v$ is the standard deviation of the noise. $\bar{L}_A$ is the average loss when the adjustment rule was applied and $\bar{L}_N$ is the average loss without any adjustment. $\Delta = 1 - \bar{L}_A/\bar{L}_N$ is the saving rate made by the adjustment rule. The 99% confidence intervals of $\Delta$ are included in the table.

<table>
<thead>
<tr>
<th>Case</th>
<th>$\theta_0$</th>
<th>$\sigma_v$</th>
<th>$\bar{L}_A$</th>
<th>$\bar{L}_N$</th>
<th>$\Delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>40.2</td>
<td>39.4</td>
<td>-1.1%±0.8%</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>75.3</td>
<td>81.1</td>
<td>3.2%±1.6%</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2</td>
<td>99.2</td>
<td>201.3</td>
<td>49.2%±0.9%</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>3</td>
<td>100.1</td>
<td>89.7</td>
<td>-9.3%±2.0%</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>3</td>
<td>141.3</td>
<td>183.0</td>
<td>18.6%±1.9%</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>3</td>
<td>193.8</td>
<td>448.4</td>
<td>55.7%±0.8%</td>
</tr>
</tbody>
</table>

Table 5.2: Savings made by the adjustment

The behavior of the adjustment rule can also be seen in Table 5.3, where $|\bar{\theta}_i|$ is the average level of true offset for each part and $P_{Ai}$ is the proportion of times an adjustment is made right before part $i$. $P_{Ai}$ is obtained from the total number of adjustments made right before part $i$ divided by the total number of replications.
As one can see in Tables 5.2 and 5.3, the proposed adjustment rule reduces the costs significantly as $\theta_0/\sigma_v$ goes up. Limited number of adjustments are made as early as the offsets are estimated with high confidence, and the initial offsets are brought down. The worst performance occurs when the initial offset $\theta_0$ is actually 0 (cases 1 and 4). In such cases, any adjustment is redundant and will increase both the quadratic losses and the adjustment costs. However, if the rule is applied in these cases, very few adjustments are made and the cost increase is small. Therefore, the optimal adjustment rule is robust with respect to different process parameters without requiring the parameter estimates before processing.
5.5.2 Changes in initial prior state variables: $\mu_0, \sigma_0, \kappa_0, \nu_0$

According to (5.5), $\nu_0$ and $\sigma_0^2$ determine the precision of the initial prior density of $\sigma_v^2$. From (5.4) and (5.29), $\kappa_0$, $\nu_0$ and $\sigma_0$ together determine the precision of the initial prior density of $\theta_0$. Low precisions (i.e., large variances) should be utilized when no prior information about the process parameters $(\theta_0, \sigma_v^2)$ is available. In such case, $\sigma_0$ should be set to a reasonably large number and $\kappa_0$ and $\nu_0$ should be as small as possible. $\kappa_0$ can be a small positive number like 0.01, and $\nu_0$ may be a number slightly larger than 2, because $\nu_0 > 2$ is necessary for calculating the optimal expected loss $R_0(\mu_0, \sigma_0)$, where the variance of $Y_1$ described in equation (5.16) is needed. These choices for a noninformative prior are close to Jeffrey’s rule (see, e.g., [25]).

The parameter $\kappa_0$ determines (more than $\sigma_0^2$ and $\nu_0$) the precision of the initial prior density for $\theta_0$. A larger $\kappa_0$ indicates more confidence that the initial offset is close to $\mu_0$. The parameter $\kappa_0$ can also be considered a tuning variable used to trade-off the performance when $\theta_0/\sigma_v$ is small and when $\theta_0/\sigma_v$ is large.

**Example 5.1(cont.)** We illustrate the effect of changing $\kappa_0$ in the example of Section 5.4 by setting it equal to 0.01 (as opposed to $\kappa = 1$). The other parameters were not changed: $\mu_0 = 0$, $\sigma_v = 10$, and $\nu_0 = 2.01$. The new performance statistics are shown in Table 5.4 for the same six cases as before.

Comparing Table 5.4 with Table 5.2, although the smaller value of $\kappa_0$ slightly improves
the performance for processes with high $\theta_0/\sigma_0$, the loss for low $\theta_0/\sigma_0$ cases is dramatically increased. The reason lies in the forward updating of $\mu_i$ in equation (5.13). A smaller $\kappa_0$ gives more weight to new observations. When $\sigma_\nu$ is large, the fluctuation of the first several observations can drive $\mu_i$ high enough to justify adjustments no matter how small or nonexistent the true offset is. A slightly larger value of $\kappa_0$ (around 1.0) makes the adjustment rule more robust.

### 5.5.3 Fixed Adjustment Cost: $c$

When $c$ is very large, the limits $\alpha$ will also be large and hence no adjustment will be made. When $c \to 0$, the $\alpha$’s will converge to 0. In that case an adjustment will be made at every stage. It is easy to show that, when $c = 0$ and $\kappa_0 \to 0$, the proposed adjustment rule works exactly as Grubbs’ Harmonic Rule [27].

<table>
<thead>
<tr>
<th>Case</th>
<th>$\theta_0$</th>
<th>$\sigma_\nu$</th>
<th>$L_A$</th>
<th>$L_N$</th>
<th>$\Delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>53.0</td>
<td>40.3</td>
<td>-25.8%±4.7%</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>77.1</td>
<td>79.2</td>
<td>-4.3%±3.3%</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2</td>
<td>84.4</td>
<td>198.4</td>
<td>55.7%±1.5%</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>3</td>
<td>123.6</td>
<td>90.1</td>
<td>-35.0%±3.8%</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>3</td>
<td>146.9</td>
<td>180.9</td>
<td>13.2%±2.8%</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>3</td>
<td>172.6</td>
<td>446.0</td>
<td>59.9%±1.4%</td>
</tr>
</tbody>
</table>

Table 5.4: Savings made by the adjustment when $\kappa_0 = 0.01$ and $\mu_0 = 0$
5.6 A simpler case: $\sigma^2_u$ known.

Now we consider a simpler problem, in which the system equations and cost function are the same as in (5.1) - (5.3), but the system parameter $\sigma_u$ is assumed to be known. A different Bayesian model is employed to estimate the unknown parameter $\theta_0$, based on which we can adjust the process.

A conjugate prior distribution is given by

$$
\theta_0 \sim N(\mu_0, \tau_0^2) \tag{5.30}
$$

$$
Y_i|\theta_0 \sim N(\theta_0 + U_0, \sigma^2_u). \tag{5.31}
$$

At state $i$, when part $i$ is finished, we have a new observation $y_i$ of $Y_i$. A posterior distribution of the offset can thus be obtained through the Bayesian model,

$$
\theta_i|Y^{(i)}, U^{(i)} \sim N(\mu_i, \tau_i^2), \tag{5.32}
$$

where similarly as before $Y^{(i)}$ and $U^{(i)}$ are sets that contain all known observations and adjustment values, respectively, from the beginning of the process till part $i$ is finished and observed. The model can be used to predict the quality characteristic $Y$. The marginal density is

$$
Y_{i+1}|\theta_i \sim N(\theta_i + U_i, \sigma^2_u). \tag{5.33}
$$

The posterior predictive density of $Y_i$ is

$$
Y_{i+1}|Y^{(i)}, U^{(i)} \sim N(\mu_i + U_i, \sigma^2_u + \tau_i^2) \tag{5.34}
$$
This density is characterized by two state variables, \( \mu_i \) and \( \tau^2_i \). These two variables can be easily updated recursively. Using the prior density in stage \( i - 1 \) as the prior, we obtain the recursive updating equations:

\[
\frac{1}{\tau^2_i} = \frac{1}{\tau^2_{i-1}} + \frac{1}{\sigma^2_v} \quad (5.35)
\]

\[
\mu_i = \frac{\mu_{i-1} + U_{i-1} + y_i}{\tau^2_{i-1} + \frac{1}{\sigma^2_v}} = \frac{\mu_{i-1} + U_{i-1} + \frac{\tau^2_{i-1}}{\sigma^2_v} y_i}{1 + \frac{\tau^2_{i-1}}{\sigma^2_v}} \quad (5.36)
\]

5.6.1 Dynamic Programming

Once the prior state variable \( \tau_0 \) is specified, all other state variables \( \tau_i \) are constants determined by the stage index \( i \). Therefore, the predictive density at stage \( i \) is described by only one state variable, \( \mu_i \). The predictive density \( Y_{i+1} | Y^{(i)}, U^{(i)} \) can be denoted as \( \phi_i(Y_i - \mu_i) \), a normal distributed density with mean \( \mu_i \) and variance \( \sigma^2_v + \tau^2_i \).

Define \( R_i(\mu_i) \) to be the minimum cost from parts \((i + 1)\) to \( N \) given the current density \( \phi_i(Y_{i-1} - \mu_i) \). At stage \( N - 1 \), we only need to consider the cost associated with current adjustment and part \( N \):

\[
R_{N-1}(\mu_{N-1}) = \min_{U_{N-1}} E\{Y_N^2 + c\delta(U_{N-1})\}
\]

\[
= \min_{U_{N-1}} \{\sigma^2_v + \tau^2_{N-1} + (\mu_{N-1} + U_{N-1})^2 + c\delta(U_{N-1})\}
\]

\[
= \sigma^2_v + \tau^2_{N-1} + \min \{\mu^2_{N-1}, c\}. \quad (5.37)
\]
The optimal adjustment which minimizes this last expression is clearly

\[ U_{N-1} = \begin{cases} 
-\mu_{N-1} & \text{if } |\mu_{N-1}| > c^{1/2} \\
0 & \text{if } |\mu_{N-1}| \leq c^{1/2}
\end{cases} \]  
(5.38)

For stage \( i < N - 1 \), we recursively define by using backwards induction

\[ R_i(\mu_i) = \min_{U_i} \left\{ \sigma_i^2 + \tau_i^2 + (\mu_i + U_i)^2 + c\delta(U_i) + E\{R_{i+1}(\mu_{i+1})|U_i}\} \right\}, \]  
(5.39)

where

\[ E\{R_{i+1}(\mu_{i+1})|U_i\} = \int R_{i+1}(\mu_{i+1})\phi_i(Y_{i+1} - (\mu_i + U_i))dY_{i+1}, \]  
(5.40)

and

\[ \mu_{i+1} = \frac{\mu_i + U_i + \tau_i^2Y_{i+1}}{1 + \frac{\tau_i^2}{\sigma_i^2}} \]  
(5.41)

Now we introduce a theorem which is proved in the appendix of this chapter.

**Theorem 5.1.** For the problem described by equation (5.39), the following three statements are true:

\( i \) \( R_i(\mu_i) \) is a nonnegative function which is symmetric about \( \mu_i = 0 \) and nondecreasing in \( |\mu| \).

\( ii \) \( E\{R_{i+1}(\mu_{i+1})|U_i\} \) is a function of \( \mu_i + U_i \); it is symmetric about \( \mu_i + U_i = 0 \) and nondecreasing in \( |\mu_i + U_i| \)

\( iii \) If the optimal controller \( U^* \) is not zero, \( U^* = -\mu_i \).
According to the third statement in Theorem 5.1, (5.39) can be transformed to

\[ R_i(\mu_i) = \sigma_i^2 + \tau_i^2 + \min\{\mu_i^2 + E\{R_{i+1}(\mu_{i+1})|U_i = 0\}, c + E\{R_{i+1}(\mu_{i+1})|U_i = -\mu_i\}\} \]  

(5.42)

The corresponding optimal controller is

\[ U_i^*(\mu_i) = \begin{cases} 
-\mu_i & \text{if } \mu_i^2 + E\{R_{i+1}(\mu_{i+1})|U_i = 0\} > c + E\{R_{i+1}(\mu_{i+1})|U_i = -\mu_i\} \\
0 & \text{o.w.}
\end{cases} \]  

(5.43)

We notice that, in (5.43), \( \mu_i^2 + E\{R_{i+1}(\mu_{i+1})|U_i = 0\} \) is an increasing symmetric continuous function of \(|\mu_i|\), whose minimum value is at \( E\{R_{i+1}(\mu_{i+1})|U_i = 0\} \) when \( \mu_i = 0 \) and a maximum value of infinite. The right hand side of the inequality is a constant for all \( \mu_i \), which is between the minimum and maximum values of the left hand side. So there exist a nonnegative critical number \( \alpha_i \) such that both sides are equal when \(|\mu_i| = \alpha_i\). Thus the optimal controller can be rewritten as

\[ U_i^*(\mu_i) = \begin{cases} 
-\mu_i & \text{if } |\mu_i| > \alpha_i \\
0 & |\mu_i| \leq \alpha_i
\end{cases} \]  

(5.44)

This control policy is of the deadband form. We can also conclude that all deadband half widths \( \alpha_i \) are less or equal to \( \sqrt{c} \), because \( E\{R_{i+1}(\mu_{i+1})|U_i = 0\} > E\{R_{i+1}(\mu_{i+1})|U_i = -\mu_i\} \).

### 5.6.2 Computer Implementation

In order to find the half widths \( \alpha_i \) of the deadband, we need to calculate \( R_i(\mu_i) \) for all possible \( \mu_i \in R \), or alternatively, for all \( \mu_i \in [-\alpha_i, \alpha_i] \) since \( R_i(\mu_i) = R_i(\alpha_i) \) for all \(|\mu_i| > \alpha_i\). To make the calculation possible, we can map \( \mu_i \) to \( \mu'_i \in D \), where \( D \) is a discrete state space.
In this paper we let \( D_\mu = \{ \mu | \mu = i \times d_\mu, k = -n, -n+1, \ldots, n-1, n \} \), where the increment \( d_\mu \) is a positive number and \( n \) is a positive integer such that \( n \times d_\mu > \sqrt{c} \). This way, \( \alpha_i \) can be approximated by \( \alpha'_i \in D \).

**Example 5.2.** Suppose \( N = 10 \) parts are to be processed and the process parameter \( \sigma_v = 1 \) is known. The adjustment cost is \( c = 9 \). The initial prior parameter is set at \( \tau_0 = 1 \). The discrete state variable space \( D \) is chosen to be \( \{ \mu_i | \mu_i = 0.1k \} \), where \( k \) is an integer such that \( |a| \leq 50 \). The control limits \( \alpha_i \) are shown in Table 5.6.2.

<table>
<thead>
<tr>
<th>Stage ( i )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_i )</td>
<td>2.7</td>
<td>2.3</td>
<td>2.1</td>
<td>1.9</td>
<td>1.7</td>
<td>1.6</td>
<td>1.7</td>
<td>1.8</td>
<td>2.2</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 5.5: Control table showing the adjustment limits \( \alpha_i \) calculated for the example.

The usage of the control limits table is very similar to the procedure described in section 5.4.2. At each stage \( i, i = 0, \ldots, N-1 \), we compare the updated state variable \( \mu_i \) with the control limit \( \alpha_i \). If \( |\mu_i| > \alpha_i \), we make an adjustment \( U_i = -\mu_i \). Otherwise, no adjustment is needed at this stage.

An R computer program that implements the variance known case (chart.n.R) is also listed in Appendix B and can be downloaded from http://www.ie.psu.edu/researchlabs/EngineeringStatistics/software.htm.
5.7 Comparison of known vs. unknown $\sigma^2_v$ approaches

We have presented two methods for the adjustment problem described in (5.1) - (5.3): 1. The method derived in sections 5.2 to 5.4, in which the system parameter $\sigma_v$ is unknown; 2. The method described in section 5.6, where the system parameter $\sigma_v$ is assumed to be known. A comparison between each of these two methods and Grubbs’ Harmonic Rule is conducted next.

Suppose a process is going to be adjusted when $N = 10$, and no prior information about the system parameters $\theta_0$ and $\sigma_v$ is given. For method 2, a guess of $\sigma_v$ is required, and suppose this guess is $\sigma_v = 5$. The prior parameters are set to be $\mu_0 = 0$ and $\tau_0 = 5$. For method 1, the prior parameters are set as following: $\mu_0 = 0$, $\sigma_0 = 5$, $\kappa_0 = 1$, and $\nu_0 = 2.01$. Notice that from equations (5.10) and (5.36) if $\kappa_0 = \frac{\tau_0^2}{\sigma_v^2}$, the recursive updating of $\mu_i$ is essentially the same for methods 1 and 2. The only difference between the methods will be the estimate of true parameter $\sigma_v^*$ and the confidence of $\mu_i$ as a estimator of $\theta_i$. ($\sigma_v^*$ here denotes the true parameter to differ from the parameter $\sigma_v$ used in method 2.) Several cases were investigated to compare method 1, method 2 and Grubbs’ rule. The settings of the prior parameters were kept the same for all cases. The results are summarized in Table 5.6. Each saving rate in that table is the percentage savings induced by an adjustment rule, compared to the total loss incurred when there is no adjustment. The savings are average estimators based on 1000 replications. In each replication, the same random errors were used for method 1, method 2, Grubbs’ rule, and no adjustment case. The 99% confidence intervals are also shown in the
Some general conclusions can be reached from Table 5.6:

- As the process \( \{Y_t\} \) becomes more noisy (larger \( \sigma^*_v \)), it becomes harder to control. Thus in every case when adjustments may be needed (i.e., when \( \theta_0 \neq 0 \)), the advantages of the more complex methods 1 and 2 over Grubbs’ rule disappear as the noise \( (\sigma^*_v) \) increases.
• Only in a few cases method 1 is significantly better than method 2. This occurs when the offset is large and the $\sigma_v$ estimate used in method 2 is far from the true value $\sigma_v^*$. Method 1, although started with a quite non-informative prior for $\sigma_v^2$, has the ability to update its estimates of $\sigma_v$ as more observations become available, an ability that method 2 lacks. It is interesting to note that even when the $\sigma_v$ estimate of method 2 is correct (i.e., when $\sigma_v = \sigma_v^*$), the difference in performance between methods 1 and 2 is negligible. We can conclude that method 1 is the most robust of the three methods considered.

• Grubbs’ harmonic rule has a performance always dominated by the other two more complex policies. This is because it was not designed considering adjustment costs, thus it suggests adjustments for every part $i$. However, given its notorious simplicity it should be the preferred method when the adjustment cost is very low or for very noisy processes.
Appendix of Chapter 5: Mathematical Derivation

Derivation of the predictive density

We have that

\[ f(Y_{i+1}|Y^{(i)}, U^{(i)}) = \int_{0}^{\infty} \int_{-\infty}^{\infty} f(Y_{i+1} | \theta_i, \sigma^2_i) f(\theta_i, \sigma^2_i | Y^{(i)}, U^{(i)}) \, d\theta_i \, d\sigma^2_i. \]

where the two densities in the integral are as in (5.6) and (5.8), respectively. Therefore,

\[ f(Y_{i+1}|Y^{(i)}, U^{(i)}) \propto \int_{0}^{\infty} \int_{-\infty}^{\infty} (\sigma^2_i)^{-\nu_i/2-2} \exp \left\{ -\frac{1}{2\sigma^2_i} \left( k_i (\mu_i - \theta_i)^2 + (Y_{i+1} - \theta_i)^2 + \nu_i \sigma^2_i \right) \right\} d\theta_i \, d\sigma^2_i. \]

Integrating first with respect to \( \theta_i \) we get

\[ f(Y_{i+1}|Y^{(i)}, U^{(i)}) \propto \int_{0}^{\infty} \sigma^{-\nu_i/2-3/2} \exp \left\{ -\frac{1}{2} \left( \frac{k_i (Y_{i+1} - \mu_i)^2}{\kappa_i + 1} + \nu_i \sigma^2_i \right) \right\} d\sigma^2_i 
\]

\[ \propto \left[ \frac{k_i (Y_{i+1} - \mu_i)^2}{\kappa_i + 1} + \nu_i \sigma^2_i \right]^{-\left( \nu_i+1 \right)/2} \]

\[ \propto \left[ \frac{\kappa_i}{\kappa_i + 1} \left( \frac{Y_{i+1} - \mu_i}{\sigma_i} \right)^2 \frac{1}{\nu_i + 1} \right]^{-\left( \nu_i+1 \right)/2} \]

which is the kernel of a student t density \( t_{\nu_i}(\mu_i, \sigma_i^2(\kappa_i + 1)/\kappa_i) \). If prior to producing part \( i+1 \) we adjust the process mean by \( U_i \), then \( Y_{i+1}|Y^{(i)}, U^{(i)} \sim t_{\nu_i}(\mu_i + U_i, \sigma_i^2(\kappa_i + 1)/\kappa_i) \).

Proof of Theorem 5.1

Lemma 5.1 Let \( Z \) be a random variable with probability density \( \Psi(z) \), which is symmetric about \( z = 0 \) and nonincreasing in \( |z| \). Let \( H(z) \) be a nonnegative function that is
symmetric about $z = 0$ and nondecreasing in $|z|$. Then, the function

$$G(\mu) = E\{H(\mu + Z)\}$$

(5.45)

is nonnegative, symmetric about $\mu = 0$ and nondecreasing in $|\mu|$.

**Proof.** It is obvious that $G(\mu)$ is nonnegative since $H(\mu + z) \geq 0$ for $\forall z$. It is also easy to show the symmetry:

$$G(-\mu) = \int_{-\infty}^{\infty} H(-\mu + z)\Psi(z)dz = \int_{-\infty}^{\infty} H(-\mu - z)\Psi(-z)d(-z) = \int_{-\infty}^{\infty} H(\mu + z)\Psi(z)dz = G(\mu).$$

(5.46)

Now we are going to prove that $G(\mu)$ is nondecreasing in $|\mu|$.

Since $H(z)$ is symmetric and nondecreasing, there exits a nonnegative value $\beta_r$,

$$\beta_r = \inf z|H(z) > r, z \geq 0$$

(5.47)

such that

$$P\{H(z) > r\} = P\{|z| > \beta_r\}$$

(5.48)

Since $H(z)$ is nonnegative, we have

$$G(\mu) = E\{H(\mu + Z)\} = \int_{0}^{\infty} P\{H(\mu + Z) > r\}dr.$$

(5.49)

Without loss of generality, let $\mu_2 > \mu_1 \geq 0$, and $\delta = \mu_2 - \mu_1 > 0$.

$$G(\mu_2) - G(\mu_1) = \int_{0}^{\infty} \{P\{H(\mu_2 + Z) > r\} - P\{H(\mu_1 + Z) > r\}\}dr$$

(5.50)
Consider,

\[ P\{H(\mu_2 + Z) > r\} - P\{H(\mu_1 + Z) > r\} \]  \quad (5.51)

\[ = P\{|\mu_2 + Z| > \beta_r\} - P\{|\mu_1 + Z| > \beta_r\} \]

\[ = P\{\mu_2 + Z < -\beta_r\} + P\{\mu_2 + Z > \beta_r\} - P\{\mu_1 + Z < \beta_r\} - P\{\mu_1 + Z > \beta_r\} \]

\[ = P\{Z < -\beta_r - \mu_1 - \delta\} + P\{Z > \beta_r - \mu_1 - \delta\} - P\{Z < -\beta_r - \mu_1\} - P\{Z > \beta_r - \mu_1\} \]

\[ = P\{\beta_r - \mu_1 - \delta < Z < \beta_r - \mu_1\} - P\{-\beta_r - \mu_1 - \delta < Z < -\beta_r - \mu_1\} \]

\[ = \int_{\beta_r - \mu_1 - \delta}^{\beta_r - \mu_1} \Psi(z)dz - \int_{-\beta_r - \mu_1 - \delta}^{-\beta_r - \mu_1} \Psi(z)dz \]

\[ = \int_{0}^{\delta} \{\Psi(\beta_r - \mu_1 - \delta + s) - \Psi(-\beta_r - \mu_1 - \delta + s)\}ds. \quad (5.52) \]

Since \(\delta - s > 0\) for \(s \in [0, \delta]\), we have

\[ | - \beta_r - \mu_1 - \delta + s| = \beta_r + \mu_1 + \delta - s \geq |\beta_r - \mu_1 - \delta + s| \quad (5.53) \]

In addition, \(\Psi(z)\) is nonincreasing in \(|z|\), for any \(s \in [0, \delta]\),

\[ \Psi(\beta_r - \mu_1 - \delta + s) - \Psi(-\beta_r - \mu_1 - \delta + s) \geq 0. \quad (5.54) \]

Hence, \((5.52) \geq 0\), and therefore \((5.50) \geq 0\). So \(G(\mu)\) in nondecreasing in \(|\mu|\). Q.E.D.

**Theorem 5.1.** For the problem described by equation (5.39), the following three statements are true:

\(i)\) \(R_i(\mu_i)\) is a nonnegative function which is symmetric about \(\mu_i = 0\) and nondecreasing in \(|\mu|\).
ii) $E\{R_{i+1}(\mu_{i+1})|U_i\}$ is a function of $\mu_i + U_i$; it is symmetric about $\mu_i + U_i = 0$ and nondecreasing in $|\mu_i + U_i|$

iii) If the optimal controller $U^*$ is not zero, $U^* = -\mu_i$.

Proof. Obviously, statements i), ii), and iii) are true for stage $N - 1$, where $R_N$ is defined to be 0.

By induction, assume i), ii), and iii) hold true for stage $i + 1$. At stage $i$, first consider

$$E\{R_{i+1}(\mu_{i+1})|U_i\} = \int R_{i+1} \left( \frac{\mu_i + U_i + \frac{\tau^2_i}{\sigma_v^2} Y_{i+1}}{1 + \frac{\tau^2_i}{\sigma_v^2}} \phi_i(Y_{i+1} - (\mu_i + U_i)) \right) dY_{i+1}$$

$$= \int R_{i+1} \left( \mu_i + U_i + \frac{1}{\sigma_v^2/\tau^2_i + 1} x \right) \phi_i(x) dx$$

$$= \int R_{i+1}(\mu_i + U_i + z) \phi'_i(z) dz, \quad (5.55)$$

where,

$$x = Y_{i+1} - (\mu_i + U_i), \quad (5.56)$$

$$z = \frac{1}{\sigma_v^2/\tau^2_i + 1} x, \quad (5.57)$$

and $\phi'_i(z)$ is the density function of a normal distribution with mean 0 and variance $\frac{\sigma_v^2 + \tau^2_i}{(\sigma_v^2/\tau^2_i + 1)^2}$, which is symmetric about $z = 0$ and nonincreasing in $|z|$.

Lemma 5.1 implies that (5.55) is symmetric about $\mu_i + U_i = 0$ and nondecreasing in $|\mu_i + U_i|$. Therefore, ii) is proved.

Consider

$$R_i(\mu_i) = \min_{U_i} \{ \sigma_v^2 + \tau^2_i + (\mu_i + U_i)^2 + c\delta(U_i) + E\{R_{i+1}(\mu_{i+1})|U_i\} \}. \quad (5.58)$$
If the optimal control $U^*_i \neq 0$, obviously (5.58) is optimized when $U^*_i = -\mu_i$. Therefore, $iii)$ is proved.

Finally, note that (5.58) can be rewritten as

$$R_i(\mu_i) = \sigma^2_v + \tau^2_i + \min\{\mu_i^2 + E\{R_{i+1}(\mu_{i+1})|U_i = 0\}, c + E\{R_{i+1}(\mu_{i+1})|U_i = -\mu_i\}\}. (5.59)$$

Consider the terms in (5.59). $\mu_i^2$ and $E\{R_{i+1}(\mu_{i+1})|U_i = 0\}$ are both functions that are symmetric about $\mu_i = 0$, and nondecreasing in $|\mu_i|$. All the other terms are constants for $\forall \mu_i$. So $R_i(\mu_i)$ is indeed a function that is symmetric about $\mu_i = 0$ and nondecreasing in $|\mu_i|$, and $i)$ is proved to be true. Q.E.D.
Chapter 6

Setup Adjustment for Asymmetric Off-target Cost

6.1 Introduction

In this chapter, we consider the same single-lot process studied in the previous chapter. Similarly as before, the process is assumed to start off-target and the initial offset is unknown. The process variance is also assumed unknown. The Bayesian model of Section 5.2 can be used for inferring the unknown process parameters. However, the objective cost function considered in this chapter is an asymmetric quadratic off-target cost and the adjustment cost is not considered. A Bayesian-based setup adjustment method is proposed in this chapter to minimize the total asymmetric off-target cost. In this method, the process mean is estimated and the optimal target, which is different from the nominal value due to the asymmetry in
the cost function, is determined at each step according to the posterior distribution.

6.2 Process model, Cost Function and Parameter Estimation

Recall the model in Chapter 5 for a single-lot process with an unknown initial offset:

\[
Y_i = \theta_i + \upsilon_i \quad (6.1)
\]

\[
\theta_i = \theta_{i-1} + U_{i-1}, \ i = 1, 2, ..., N \quad (6.2)
\]

where \( \upsilon_i \sim N(0, \sigma^2_\upsilon) \) is the random part-to-part deviation and \( \sigma^2_\upsilon \) is assumed unknown; \( Y_i \) is the quality characteristic of the \( i^{th} \) part and it can be observed; \( \theta_i \) is the unobservable process mean while part \( i \) is processed.

Suppose the nominal value of these parts is 0. An off-target quadratic cost should be proportional to the square of the deviation between the quality characteristic and the nominal value, i.e. \( Y_i^2 \). In the cases considered in this chapter, the oversized parts \( (Y_i > 0) \) and undersized parts \( (Y_i < 0) \) have different economic value. Due to such asymmetry, the asymmetric quadratic cost of the \( i^{th} \) part is assumed to be:

\[
C_i(Y_i) = \begin{cases} 
  c_1 Y_i^2 & \text{if } Y_i < 0 \\
  c_2 Y_i^2 & \text{if } Y_i \geq 0.
\end{cases} \quad (6.3)
\]

Without loss of generality, we can set \( c_1 > c_2 > 0 \). The goal of an adjustment rule is thus to
minimize the total expected quadratic cost for all parts:

\[ E[C] = \sum_{i=1}^{N} E[C_i(Y_i)] \]  \hspace{1cm} (6.4)

through a sequence of adjustments \{U_i\} made to the process means \{\theta_i\}. Specifically, since the adjustment can be made at each step without introducing errors or costs, each adjustment \(U_{i-1}\) is made to modify \(\theta_i\) and minimize the expected off-target cost of the next part only, which is given by \(E[C_i(Y_i)]\).

Despite the different form of cost function, the two Bayesian models in Chapter 5, the model with unknown variances of Section 5.2 and the model with known variances of Section 5.6, can still be used to make inferences and estimate the process parameters, depending on the assumption on the knowledge of the variances \(\sigma^2_\upsilon\).

The process described by equations (6.1)-(6.2) under the asymmetric quadratic cost in (6.3) has been recently studied by Colosimo, Pan and del Castillo [10]. Their solution is based on the assumption that the variance parameter \(\sigma^2_\upsilon\) is known and can be interpreted according to the Bayesian model of Section 5.6. In this model, the conjugate prior distribution for the setup error is given by

\[ \theta_0 \sim N(\mu_0, \tau^2_0), \] \hspace{1cm} (6.5)

and the posterior predictive distribution of the response after the \(i^{th}\) part is observed and before the \((i + 1)^{th}\) part is processed is given by

\[ Y_{i+1}|Y^{(i)}, U^{(i)} \sim N(\mu_i + U_i, \sigma^2_\upsilon + \tau^2_i), \] \hspace{1cm} (6.6)

where \(Y^{(i)}\) and \(U^{(i)}\) are the sets of all observations and adjustments at current step. The
variables $\tau_i^2$ and $\mu_i$ can be updated recursively through the equations:

\[
\tau_i^2 = \frac{1}{\tau_{i-1}^2} + \frac{1}{\sigma_v^2}, \quad (6.7)
\]

\[
\mu_i = \frac{\mu_{i-1} + U_{i-1}}{\tau_{i-1} + 1} + \frac{y_i}{\sigma_v^2} = \frac{\mu_{i-1} + U_{i-1} + \tau_{i-1}^2 y_i}{1 + \tau_{i-1}^2}, \quad (6.8)
\]

where $y_i$ is the observed value of $Y_i$. In the solution given by Colosimo et al. [10], $\mu_0$ and $\tau_0^2$ are assumed to be 0 and $\infty$, respectively, to represent a noninformative prior. With this assumption, equation (6.7) can be simplified to be

\[
\tau_i^2 = \frac{i}{\sigma_v^2}, \quad (6.9)
\]

and the equation (6.8) becomes

\[
\mu_i = \mu_{i-1} + U_{i-1} + \frac{y_i}{1 + \frac{1}{i}}, \quad (6.10)
\]

which is an EWMA estimate of the process mean. With such results, the predictive distribution in (6.6) can be rewritten as

\[
Y_i | Y^{(i-1)}, U^{(i-1)} \sim N(\mu_{i-1} + U_{i-1}, (1 + 1/i)\sigma_v^2). \quad (6.11)
\]

It is easy to find the optimal target $\mu = \mu_i^*$ to minimize

\[
E[C_i(Y_i) | Y^{(i-1)}, U^{(i-1)}] = \int_{-\infty}^{\infty} C_i(y)\phi_i(y - \mu)dy, \quad (6.12)
\]

where $\phi(\cdot)$ is the probability density function of a $N(0, (1 + 1/i)\sigma_v^2)$ distribution. According to the optimal target, the adjustment can be made by $U_{i-1} = \mu_i^* - \mu_{i-1}$. This method is the optimal adjustment method in the cases that the variance $\sigma_v^2$ is known, and will be referred to as Method I in this chapter.
In this chapter, the case when the variance $\sigma^2_v$ is unknown is considered. The Bayesian model in Section 5.2 is used to make inferences about the two unknown parameters, $\theta_0$ and $\sigma^2_v$. Recall the conjugate prior distribution:

$$\theta_0|\sigma_v \sim N(\mu_0, \sigma^2_v/\kappa_0)$$  \hspace{1cm} (6.13)

$$\sigma^2_v \sim \text{Inv-Chi}^2(\nu_0, \sigma^2_0).$$  \hspace{1cm} (6.14)

Posterior densities are given by

$$Y_i|\theta_i, \sigma^2_v \sim N(\theta_i, \sigma^2_v),$$  \hspace{1cm} (6.15)

or alternatively

$$Y_{i+1}|\theta_i, \sigma^2_v \sim N(\theta_i + U_i, \sigma^2_v),$$  \hspace{1cm} (6.16)

where

$$(\theta_i, \sigma^2_v)|Y^{(i)}, U^{(i)} \sim N - \text{Inv-Chi}^2(\mu_i, \sigma^2_i/\kappa_i; \nu_i, \sigma^2_i).$$  \hspace{1cm} (6.17)

Again the recursive updating expressions are:

$$\mu_i = \frac{\kappa_{i-1}}{\kappa_{i-1} + 1}(\mu_{i-1} + U_{i-1}) + \frac{1}{\kappa_{i-1} + 1}y_i$$  \hspace{1cm} (6.18)

$$\kappa_i = \kappa_{i-1} + 1$$  \hspace{1cm} (6.19)

$$\nu_i = \nu_{i-1} + 1$$  \hspace{1cm} (6.20)

$$\nu_i \sigma^2_i = \nu_{i-1} \sigma^2_{i-1} + \frac{\kappa_{i-1}}{\kappa_{i-1} + 1}(y_i - \mu_{i-1} - U_{i-1})^2,$$  \hspace{1cm} (6.21)

where $y_i$ is the observed value of $Y_i$. Notice that, when $\mu_0 = 0$ and $\kappa_0 = 0$, i.e. when the prior distribution of the initial offset $\theta_0$ has mean at 0 and infinite variance, the updating
of the process mean estimator $\mu_i$ has exactly the same form as that in Method I (equation 6.10). This new method will be referred as Method II in this chapter.

### 6.3 Optimal Adjustment

In this section, an optimal setup adjustment rule is proposed for the cases when the variance is unknown, based on the model described in (6.15-6.21).

It has been shown in Chapter 5, that the predictive distribution at state $i-1$, i.e., after the $(i-1)^{th}$ part is observed and before the $i^{th}$ part is processed, is a scaled t distribution given by

$$Y_i | Y^{(i-1)}, U^{(i-1)} \sim t_\nu(\mu, \sigma^2), \quad (6.22)$$

where $\nu = \nu_i$, $\mu = \mu_{i-1} + U_{i-1}$, and $\sigma^2 = \sigma^2_{i-1}(\kappa_{i-1} + 1)/\kappa_{i-1}$). The expected cost at the current stage associated with the $i^{th}$ part is given by

$$E[C_i(Y_i)] = c_1 \int_{-\infty}^{0} Y_i^2 f(Y_i - \mu) dY_i + c_2 \int_{0}^{\infty} Y_i^2 f(Y_i - \mu) dY_i \quad (6.23)$$

$$= c_2 \int_{-\infty}^{0} Y_i^2 f(Y_i - \mu) dY_i + (c_1 - c_2) \int_{-\infty}^{0} Y_i^2 f(Y_i - \mu) dY_i \quad (6.24)$$

$$= c_2(\mu^2 + \sigma^2 \frac{\nu}{\nu - 2}) + (c_1 - c_2) \int_{-\infty}^{-\mu} (x + \mu)^2 f(x) dx, \quad (6.25)$$

where $f(\cdot)$ is the density function of the $t_\nu(0, \sigma^2)$ distribution and $x \equiv Y_i - \mu$. We can compute the first derivative of $E(C_i)$ with respect to $\mu$:

$$\frac{\partial E(C_i)}{\partial \mu} = 2c_2 \mu + (c_1 - c_2) \left[ -(x + \mu)^2 f(x)|_{x=-\mu} + \int_{-\infty}^{-\mu} 2(x + \mu) f(x) dx \right] \quad (6.26)$$

$$= 2c_2 \mu + 2(c_1 - c_2) \left[ \mu F(-\mu) + \int_{-\infty}^{-\mu} x f(x) dx \right], \quad (6.27)$$
where $F(\cdot)$ is the CDF of a $t_\nu(0, \sigma^2)$ distribution. The second derivative is given by

$$
\frac{\partial^2 [E(C_i)]}{\partial \mu^2} = 2c_2 + 2(c_1 - c_2) \left[ - (x + \mu) f(x) \Big|_{x=-\mu} + \int_{-\infty}^{-\mu} f(x) dx \right]
$$

(6.28)

$$
= 2c_2 + 2(c_1 - c_2) F(-\mu).
$$

(6.29)

Obviously the second derivative is positive, hence $E(C_i)$ is a strict convex function with respect to $\mu$. Thus, given $\nu$ and $\sigma^2$, there exists one and only one $\mu_i^*$ such that $E(C_i)$ is minimized at $\mu = \mu_i^*$. The optimal adjustment at stage $i - 1$ is hence $U_i^{*\mu}_{i-1} = \mu_i^* - \mu_i$. $\mu_i^*$ can be obtained by solving the following equation:

$$
\frac{\partial [E(C_i)]}{\partial \mu} = 0.
$$

(6.30)

An algorithm to numerically solve equation (6.30) is shown in the Appendix of Chapter 6.

An example will be shown to illustrate this adjustment rule, in comparison with Method I, the adjustment rule for the known-variance case of reference [10].

**Example 6.1.** Suppose a process that can be described by equations (6.1)-(6.2) needs to be adjusted. The process parameters are given by $N = 20$, $\theta_0 = 3$, $\sigma_v = 1$. The variables in the asymmetric cost are $c_1 = 4$ and $c_2 = 1$. Suppose certain prior information of $\sigma_v$ is available but it may be wrong. The prior estimator of $\sigma_v$ is assumed to be $b\sigma_v$, where the bias factor $b$ is a positive number and $\sigma_v$ is the true value. If Method I is used, the prior estimator will be accepted as the true value, despite the errors that may exist. In Method II, $\sigma_v$ is still assumed unknown. The prior information is only used to set up the prior distribution. In this example, we assume the mean of the prior distribution of $\sigma_v^2$ is $b^2\sigma_v^2$ and
the variance is 200 times of the square of the mean \((200b^4\sigma^4_0)\). This large variance represents lack of confidence in the prior knowledge. For both methods, the prior distributions of \(\theta_0\) have means at 0 and infinite variances.

The process was simulated with common random numbers, under the adjustment of these two rules, for the cases \(b = 0.5\) (underestimated), \(b = 1\) (accurate) and \(b = 2\) (overestimated). Method I with \(b = 1\), which is the best adjustment rule, is also applied for comparison. The total off-target costs of all \(N\) parts for different method with different value \(b\) are listed in Table 6.1. Percentage savings are created to compare these two method, which is given by the equation:

\[
S = 1 - \frac{C^{II}}{C^I},
\]

where \(C^I\) and \(C^{II}\) are the total costs under the adjustment of Method I and Method II respectively, in which the same prior knowledge is applied (same \(b\) for both methods).

<table>
<thead>
<tr>
<th>bias</th>
<th>(b = 0.5)</th>
<th>(b = 1)</th>
<th>(b = 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>method</td>
<td>I</td>
<td>II</td>
<td>I</td>
</tr>
<tr>
<td>cost</td>
<td>54.3</td>
<td>51.9</td>
<td>47.2</td>
</tr>
<tr>
<td>saving</td>
<td>4.4%</td>
<td>-4.5%</td>
<td>14.6%</td>
</tr>
</tbody>
</table>

Table 6.1: Total off-target cost for each method and each value of \(b\) and the saving made by Method I v.s. Method II for each value of \(b\).

In this specific example, Method II has better performance (positive savings) than Method I when the prior information is biased \((b = 0.5, 2)\). The only exception is the case \(b = 1\), when the prior estimator is accurate and Method I is the optimal method as expected. More explicit comparisons will be conducted in Section 6.4 based on more simulation results.

Since the estimates of the process means, \(\mu_i\), are essentially the same for both methods
due to the setup of the prior distribution of $\theta_0$, the main difference between these methods is the determination of the targets $\mu_i^\ast$. Figures 6.1 and 6.2 give the targets calculated with both methods when the prior information used is underestimated ($b = 0.5$) and overestimated ($b = 2$), respectively. The optimal targets given by Method I for $b = 1$ are plotted in both figures for reference. At the beginning of the process, due to the biased prior information, the targets given by both Method I and Method II are away from the optimal targets (as given by Method I with accurate prior information, i.e., $b = 1$). As the process runs, the targets given by Method II, in contrast to those given by Method I, gradually converge to
Figure 6.2: Plot of targets $\mu_i^*$ for Method I with $b = 2$, Method I with $b = 1$ (optimal targets) and Method II with $b = 2$.

the best values, and this results in the reduction in the total off-target cost.

6.4 Simulation Study

In the example in the previous section, one process was simulated to compare the performance of Method I and Method II. To evaluate the new method by comparing it with Method I in a more comprehensive manner, more cases were investigated and more replications were made in this section.
In the simulation study, we assume the prior distribution of $\theta_0$ is the same as in Example 6.1, so the estimates of the process means are the same for both methods. Hence, the magnitude of the initial offset will not affect the difference between these two methods. Suppose the ratios between the variance of the priors of $\sigma^2_\nu$ and the means are given, which is 200 in this chapter. For any given process length, the only factors that have effect on the relative savings are the relative bias in the prior information ($b$) and the ratio $\frac{c_1}{c_2}$. The part-to-part variance $\sigma^2_\nu$ can always be scaled to be 1 and $c_2$ can also be fixed to 1. We varied each of the two factors $b$ and $c_1$ for three levels, which are $b = 0.5, 1, 2$ and $c_1 = 2, 4, 6$. For each combination of $b$ and $c_1$, 5000 replications were simulated. In each replication, a process with 20 parts was simulated using common random numbers, under the adjustment of Method I and Method II, similarly to that in Example 6.1. The percentage saving for this replication was calculated according to equation (6.31). The average saving for each combination of $b$ and $c_1$ and its 95% confidence interval were obtained based on 5000 replications. The effect of length of the process ($N$) was also examined. Since the first part in a process is always left unadjusted, the savings were calculated with respect to the total costs incurred when producing parts 2 to 5, parts 2 to 10 and parts 2 to 20. The results are shown in Figures 6.3-6.5.

In most cases that the prior information is biased ($b = 0.5, 2$), Method II shows significant positive savings due to its adaptability. The only exception is the case when $b = 0.5$ for parts 2 to 5, a case which will be addressed later. In cases when $b = 1$, where Method II is being compared to the best method, the savings are significantly below 0.
Figure 6.3: The average saving in the total off-target costs from part 2 to part 5 induced by Method I and Method II for each combination of $b$ and $c_1$ and its 95% confidence interval.

It can be observed that the longer the process is (the larger $N$ is), the more significant the advantage of Method II over Method I. This is because it takes a certain transient time for the targets ($\mu_i^*$) to converge from values far away from the optimal values (as given by Method I with $b = 1$) to the values close to the best ones. When more data is available, better inferences of the process parameters are provided by Method II, and this results in better adjustments.

It is notable that the advantages of Method II in cases when $b = 0.5$ are not as significant as those in cases $b = 2$. If the length of the process is short, the average savings are even
Figure 6.4: The average saving in the total off-target costs from part 2 to part 10 induced by Method I and Method II for each combination of $b$ and $c_1$ and its 95% confidence interval.

below 0. This can be explained by the skewness of the prior $Inv - \chi^2$ distributions of $\sigma^2_\upsilon$.

In a $Inv - \chi^2$ distribution, the probability density is relatively lower on the right side of the mean than it is on the left side, i.e. such prior distribution weights the smaller values more heavily than the larger values. Hence it takes a longer time for the estimate of $\sigma^2_\upsilon$ to be tuned to a larger value than to a smaller value. Thus, in practice, a user should not be too conservative in selecting the prior mean of $\sigma^2_\upsilon$, as overestimating this parameter is clearly less problematic than understimating it.

The simulation results also show that the savings are more significant when $c_1$ is larger,
Figure 6.5: The average saving in the total off-target costs from part 2 to part 20 induced by Method I and Method II for each combination of $b$ and $c_1$ and its 95% confidence interval.

either for positive savings or negative savings. This implies that the determination of the optimal targets becomes more important as the cost function is more asymmetric.

6.5 Conclusions

Through the simulation study, it was shown that Method II, in which $\sigma_v^2$ is assumed unknown, has robust overall performance compared to Method I, in which $\sigma_v^2$ is assumed known. In particular, in the cases when the prior information is biased, Method II gives better results
due to its self-adaptation. So when there exits a probability that there exist errors in the
prior estimation, or when the prior knowledge is lacked, Method I is recommended.

The sensitivity analysis of the variances of the prior distributions of $\sigma^2_\nu$ was not included
in the simulation study. For a $Inv - \chi^2$ prior distribution as in (6.14), the ratio between
the variance and the square of mean is given by $2/(\nu_0 - 4)$ when $\nu_0 > 4$. When $\nu < 4$, the
variance is infinite, which is not recommended for this problem. Unlike other problems with
symmetric cost functions, where the effects of prior or posterior distributions of variance
parameters are reflected in the estimates of process means and the adjustment targets are
automatically the nominal values, in this problem with asymmetric costs the distributions
of $\sigma^2_\nu$ is directly used to determine the optimal targets. If the variance of the prior of $\sigma^2_\nu$
is infinite, it will drastically increase the variance of the predictive distribution in (6.16) or
in (6.22). If the variances of the predictive distributions are exaggerated, the corresponding
targets can be too far away from the best values. On the other hand, if the variance is
too small, it will reduce the adaptability of the Bayesian model in Method II, which will
also result in poor performance for short-run processes. Therefore, our suggestion for $\nu_0$
is a number slightly larger than 4 ($\nu_0 = 4.01$ was used in this chapter). The effect of the
prior distribution will decay gradually as the process runs, and hence its importance will be
relatively small for long-run process.

A computer program that implements the adjustment method developed in this chapter
is listed in Appendix C and can be downloaded from

http://www.ie.psu.edu/researchLabs/EngineeringStatistics/.
Appendix of Chapter 6: Solution to Equation 6.30

As shown in Section 6.3, the second derivative of $E(C_i)$ is positive, therefore the first derivative $\frac{\partial[E(C_i)]}{\partial \mu}$ is a monotonically increasing function of $\mu$. Recall equation 6.27:

$$\frac{\partial[E(C_i)]}{\partial \mu} = 2c_2\mu + 2(c_1 - c_2) \left[ \mu F(-\mu) + \int_{-\infty}^{-\mu} x f(x) dx \right].$$

(6.32)

Notice that $\frac{\partial[E(C_i)]}{\partial \mu}$ is less than 0 at $\mu = 0$ under the assumption $c_1 > c_2$ (or larger than 0 when $c_1 < c_2$). Therefore the solution $\mu^*_i$ to 6.30 should be larger than 0 (or less than 0 when $c_1 < c_2$). Integrate the integral part in (6.32), we obtain:

$$\frac{\partial[E(C_i)]}{\partial \mu} = 2c_2\mu + 2(c_1 - c_2) \left[ \mu F(-\mu) + \frac{\Gamma(1+\nu)}{\sqrt{\pi} \Gamma(\nu/2)} \frac{\sigma\sqrt{\nu}}{1 - \nu} \left( \frac{\mu^2}{\sigma^2\nu} + 1 \right)^{\frac{1-\nu}{2}} \right],$$

(6.33)

where the CDF $F(\cdot)$ for a t-distribution and the Gamma function $\Gamma(\cdot)$ can be calculated by functions in MATLAB. Hence it is easy to calculate the value of $\frac{\partial[E(C_i)]}{\partial \mu}$ for given $\mu$.

Considering above properties of $\frac{\partial[E(C_i)]}{\partial \mu}$, a searching algorithm is proposed to find the solution $\mu = \mu^*_i$ to $\frac{\partial[E(C_i)]}{\partial \mu} = 0$:

- Initialization:

  $$x_1 = 0; \quad x_2 = 1 \text{ if } c_1 > c_2; \quad x_2 = -1 \text{ if } c_1 < c_2.$$

- Find the lower or upper bound of the solution:

  iterate while $\frac{\partial[E(C_i)]}{\partial \mu}|_{\mu=x_1}$ and $\frac{\partial[E(C_i)]}{\partial \mu}|_{\mu=x_2}$ have the same sign:

  $$x_2 \leftarrow 2 \times x_2.$$

- Line search for $\mu^*_i$:
iterate while $|x_1 - x_2| > p$ ($p$ is the specified precision for $\mu_i^*$, 0.01 is used in this chapter).

\[ x_3 = (x_1 + x_2)/2, \]

if $\frac{\partial E(C_i)}{\partial \mu}|_{\mu=x_3}$ and $\frac{\partial E(C_i)}{\partial \mu}|_{\mu=x_1}$ have the same sign,

\[ x_1 = x_3; \]

else,

\[ x_2 = x_3. \]

- Final solution $\mu_i^* = (x_1 + x_2)/2$. 


Chapter 7

Adaptive Deadband Control of a Process with Random Drift in the Presence of a Fixed Adjustment Cost

Chapters 3-6 considered the setup adjustment problems with different cost structures. In those problems, the process mean did not change unless an adjustment was made. In this chapter, a different type of process is studied, one in which the process mean can drift randomly if no adjustment is applied. Such processes can appear not only in discrete-part manufacturing but also in continuous processes.

Similarly to chapter 3, both the symmetric off-target cost and the fixed adjustment cost are considered. The problem has been solved by Box and Jenkins [5] and Crowder [13] based on the assumption that the variance parameters are known. In this chapter, we give
a solution to this problem for the case that the parameters are unknown. The solution is based on a Bayesian model and utilizes a result from Box and Jenkins’ solution [5]. It will be shown later in this chapter that the proposed method provides better results either in terms of estimation or in terms of adjustments.

Due to the complexity of this problem, the Bayesian computation has to be carried out by numerical methods. The SMC method was chosen to accomplish this job. Because of the time-efficiency of the SMC method, the proposed adjustment method can be applied for on-line adjustment purpose.

7.1 Process Model and Estimation

Consider a process that can be modelled by the system and observation equations

\[
\begin{align*}
\theta_i &= \theta_{i-1} + U_{i-1} + \nu_i \\
Y_i &= \theta_i + \epsilon_i & i = 1, 2, ..., N,
\end{align*}
\]

where \( \nu_i \overset{iid}{\sim} N(0, \sigma^2_\nu) \), \( \epsilon_i \overset{iid}{\sim} N(0, \sigma^2_\epsilon) \) are two random sequences that are independent of each other. \( \theta_i \) is the system mean at stage \( i \), which will follow a random walk if the process is not adjusted (i.e., if \( U_i = 0 \) for \( i = 1, 2, 3, ..., N \)). The system mean can be linearly adjusted by \( U_i \) units at each stage to keep it close to its target, where the target can be assumed to be 0 without loss of generality. The process is assumed to start on target, i.e. \( \theta_0 = 0 \). We observe \( Y_i \), which includes the information given by the system mean and an observational (or measurement) error \( \epsilon_i \).
A symmetric quadratic loss function

\[ L^q = \sum_{i=1}^{N} \theta_i^2 \]  

(7.3)

is often used to evaluate the performance of a process under a given adjustment rule. To minimize the total loss in (7.3), an adjustment needs to be made at each stage by the amount

\[ U_i = -\hat{\theta}_i, \]  

(7.4)

where \( \hat{\theta}_i \) is an estimate of \( \theta_i \) obtained at stage \( i \) after \( Y_i \) is observed. When the adjustments are carried out at every stage, loss function in (7.3) not only evaluates the performance of an adjustment rule but also reflects the accuracy of the process mean estimate. Traditionally, the estimate is calculated through Kalman Filter. Given the observations from the beginning until stage \( i \), \( y^i = \{Y_1, Y_2, ..., Y_i\} \), the estimate of \( \theta_i \) at stage \( i \) is given by

\[
\hat{\theta}_i = \hat{\theta}_{i-1} + U_{i-1} + \omega [Y_i - (\hat{\theta}_{i-1} + U_{i-1})] \\
= (1 - \omega)(\hat{\theta}_{i-1} + U_{i-1}) + \omega Y_i,
\]

(7.5)

where \( \hat{\theta}_0 = 0 \) and

\[
\omega = \frac{1 + \sqrt{1 + 4\sigma^2/\sigma_\nu^2}}{1 + \sqrt{1 + 4\sigma^2/\sigma_\nu^2 + 2\sigma^2/\sigma_\nu^2}}.
\]

(7.6)

Notice that the Kalman Filter estimate [18] is essentially an EWMA estimate, specifically, a weighted average of the old estimate \( \theta_{i-1} + U_{i-1} \) and the new observation \( Y_i \). The weight \( \omega \) is uniquely determined by the ratio of two variances \( \sigma^2/\sigma_\nu^2 \). In the case that \( \sigma^2/\sigma_\nu^2 \) is large, which implies relatively large observation errors, a small weight \( \omega \) will be assigned to the new observation. When observations are more precise, i.e. when \( \sigma^2/\sigma_\nu^2 \) is small, the weight
for the new observation will be larger. Duncan and Horn [18] show how the Kalman filter estimate provides the minimum mean square error estimator of $\theta_i$. This assumes the error distributions are normal with the ratio $\sigma^2_\epsilon/\sigma^2_\nu$ known.

In the case when $\sigma^2_\epsilon$ and $\sigma^2_\nu$ are unknown, Kalman Filter estimates can not be applied. If biased estimates of $\sigma^2_\epsilon$ and $\sigma^2_\nu$ are used, Kalman Filter estimation may lead to a poor performance of the adjustment methods based on it. Notice that the Kalman Filter estimate can be interpreted as the mean of the posterior distribution of $\theta_i$ conditioning on the data set $y^i$ for a Bayesian model that assumes $\sigma^2_\epsilon$ and $\sigma^2_\nu$ are known. In the case that these two process parameters are unknown, a new adjustment method is proposed in this paper, which also has the form as in (7.4). However the estimates of $\theta_i$ are obtained based on a different Bayesian model in which the process parameters $\sigma^2_\epsilon$ and $\sigma^2_\nu$ are unknown. In this model, the prior distributions of $\sigma^2_\epsilon$ and $\sigma^2_\nu$ are required before the process starts. At each stage when a new observation is obtained, the posterior distribution on the process mean at that stage and the parameters $\sigma^2_\epsilon$ and $\sigma^2_\nu$ can be computed, according to which the adjustment can be made. The steps to update the posterior distributions and adjust the process can be summarized as follows:

- Start with prior distributions

\[
\sigma^2_\epsilon \sim LN(\mu_1, \sigma^2_1) \\
\sigma^2_\nu \sim LN(\mu_2, \sigma^2_2),
\]  

(7.7)

where $LN$ represents a log-normal distribution. (A justification of these priors is given
below). The predictive distribution of $\theta_1$ is

$$\theta_1|\sigma^2, \theta_0 \sim N(\theta_0, \sigma^2), \quad (7.8)$$

where $\theta_0 = 0$ is assumed. The joint distribution of $\sigma^2, \sigma^2$ and $\theta_1$ before the first observation in a process is denoted as $\pi(\sigma^2, \sigma^2, \theta_1|y^0)$, where $y^0$ is an empty data set.

- At stage $i$, $i = 1, 2, 3, ..., N$, when the new observation $Y_i$ is obtained,

  - Update the posterior distribution:

    $$p(\sigma^2, \sigma^2, \theta_i|y^i) \propto \pi(\sigma^2, \sigma^2, \theta_i|y^{(i-1)}) \times \text{likelihood}(\sigma^2, \sigma^2, \theta_i|Y_i)$$

    $$\propto \pi(\sigma^2, \sigma^2, \theta_i|y^{(i-1)}) \times \frac{1}{\sigma^2} \exp\left[-\frac{(Y_i - \hat{\theta}_i)^2}{2\sigma^2}\right]. \quad (7.9)$$

  - Adjust the process by

    $$U_i = -\hat{\theta}_i = -E(\theta_i|y^i). \quad (7.10)$$

  - Compute the new predictive distribution

    $$\pi(\sigma^2, \sigma^2, \theta_{(i+1)}|y^i) = \int p(\sigma^2, \sigma^2, \theta_i|y^i) \times p(\theta_{(i+1)}|\sigma^2, \theta_i) d\theta_i$$

    $$= \int p(\sigma^2, \sigma^2, \theta_i|y^i) \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\theta_{(i+1)} - \theta_i)^2}{2\sigma^2}\right], \quad (7.11)$$

    where $p(\bullet|y^i)$ is the posterior distribution computed in (7.9).

  - Iterate on $i$

The posterior distribution computation is carried out by a Sequential Monte Carlo (SMC) method, specifically, the One-Pass Particle Filter (1PFS) algorithm [2]. The “1PFS” algorithm was modified in this research for computing the posterior distributions of those
parameters that has bounded domain (variance parameters in this problem). The detail of the algorithm is shown in the appendix of this chapter. The SMC method does not require conjugate priors, so various forms of priors can be applied. The Log-Normal distribution is chosen in this paper because its convenience in incorporating the prior information. The parameters characterizing the prior distributions can be determined depending on the availability and precision of the prior information. When the prior distribution is not available or not much confidence can be had on prior information, larger variances can be assigned to the prior distributions to represent a more noninformative prior.

To compare numerically the proposed adjustment method (SMC method) and the adjustment method based on Kalman Filter (KF method), 9 scenarios with different prior information as shown in Table 7.1 are investigated. In each scenario, suppose the prior estimates of the parameters $\sigma_\epsilon$ and $\sigma_\nu$ are available, and they are $b_\epsilon \sigma_\epsilon$ and $b_\nu \sigma_\nu$ respectively. These estimates may be accurate ($b_\epsilon, b_\nu = 1$) or biased ($b_\epsilon, b_\nu = 0.5, 2$). The estimates on the ratio $\sigma_\epsilon^2 / \sigma_\nu^2$ are given by $b_\epsilon^2 / b_\nu^2 \times \sigma_\epsilon^2 / \sigma_\nu^2$, which may also be biased. If the KF method is applied, these estimates are plugged into equations (7.5-7.6) to estimate the process mean. In the SMC method, on the other hand, these estimates are only used as the means of the corresponding prior distributions. To represent lack of confidence on them, we set the variances to be equal to 4 times of the square of the distribution means ($4b_\epsilon^4 \sigma_\epsilon^4$ and $4b_\nu^4 \sigma_\nu^4$, respectively).

Three cases of process parameters were considered as shown in Table 7.2. Each of them had 500 stages. For each case in Table 7.2, common random numbers were used to simulate
<table>
<thead>
<tr>
<th>Scenario</th>
<th>( b_{\epsilon} )</th>
<th>( b_{\nu} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 7.1: Scenarios with different prior information.

<table>
<thead>
<tr>
<th>Case</th>
<th>( \sigma_{\epsilon} )</th>
<th>( \sigma_{\nu} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 7.2: Different processes investigated.

The adjusted process under each of the two adjustment methods (the KF method and the SMC method). Each case in Table 7.2 was tested with each scenario in Table 7.1. The total quadratic losses were calculated for both methods, and the performance was evaluated by the percentage savings induced by the SMC method over the KF method:

\[
S^q = \frac{L^q_{k} - L^q_{s}}{L^q_{k}},
\]

(7.12)

where \( L^q_{k} \) is the loss under the KF adjustments and \( L^q_{s} \) is the loss under the SMC adjustments.

20 replications were made, based on which the average savings and the frequentist 95% confidence interval were obtained for each combination of a case (Table 7.2) and a scenario (Table 7.1). The results are shown in Figure 7.1.

As it can be seen, in scenarios 1, 5 and 9, where the ratios used in the KF method happen
Figure 7.1: Average savings induced by SMC method v.s. the KF method for each combination of a prior scenario and a process case, and their 95% confidence intervals to be equal to the true values, the KF method provides the best estimates as expected. However, notice that the negative savings incurred by the SMC method are only slightly below 0. In other scenarios, where the wrong ratio was used in the KF method, the advantage of using the SMC method turn to be very significant. In particular, the ratios are underestimated in scenarios 2, 3, and 6, such that the KF method gives too much weight to new observations that have considerable observation error. In scenarios 4, 7 and 8, overestimated ratios are used in the KF method so that the errors in the new observations are exaggerated and too little weight is assigned to the new observations. Although the SMC method also starts with biased prior information, the effect of the prior distributions decays very soon as more data is available. Thus, in a situation where the prior information is likely to be biased, the SMC method can achieve a more robust estimation and hence better adjustment performance.
The robustness of the SMC method is also reflected by the fact that the performance the
SMC method is insensitive to the errors in prior distribution means when the correct variance
ratio is used \((b_e/b_v=1\), i.e. scenarios 1, 5 and 9).

### 7.2 Fixed Adjustment Cost and Adaptive Deadband

**Adjustment Method**

Sometimes in practice, the costs associated with the adjustments need to be taken into
account. Consider the following total loss function:

\[
L^t = \sum_{i=1}^{N} \theta_i^2 + c \sum_{i=1}^{N} \delta(U_i),
\]

where the function \(\delta(x) = 1\) if \(x \neq 0\) and \(\delta(x) = 0\) otherwise. This loss function includes
the total quadratic loss in (7.3) and the sum of fixed adjustment costs. Constant \(c\) is a
cost associated with one adjustment relative to a one unit quadratic loss. Such cost may
represent the time and labor spent on stopping machine and making an adjustment, which is
independent of the amount of the adjustment. The fixed adjustment cost generally implies
an adjustment method with a “deadband” form [5, 13, 30], i.e. instead of adjusting at each
step as in (7.4), the adjustment will only be carried when the magnitude of the process mean
is large enough such that the adjustment is worthy:

\[
U_i = \begin{cases} 
-\hat{\theta}_i & \text{if } |\hat{\theta}_i| > \alpha_i \\
0 & \text{if } |\hat{\theta}_i| \leq \alpha_i,
\end{cases}
\]
where \( \alpha_i \) is the action limit or the half width of the “deadband” at stage \( i \). Adjustment methods with such form have been proposed for the known parameter case, to control the process described in (7.1-7.2) under the loss function in (7.13) based on Kalman Filter estimations [5, 13]. Box and Jenkins [5] considered the case when the the length of the process is infinite \( (N = \infty) \). In the solution they provided, \( \alpha_i \) is constant for all stages and is uniquely determined by \( \sigma_\nu \) for a fixed adjustment cost \( c \), namely

\[
\alpha_i = \left[(6c/\sigma_{\nu}^2)^{1/4} - 0.63\right]\sigma_\nu.
\]

(7.15)

We refer to this method as the Constant Deadband Adjustment method. Crowder [13] gave an optimal solution to this known-parameters problem for the finite horizon case using Dynamic Programming, in which the optimal \( \alpha_i \) is nondecreasing with \( i \). Specifically, when \( (N - i) \to \infty \), \( \alpha_i \) will converge to the constant control limit given by (7.15). When the process is approaching the end, i.e. when \( i \to N - 1 \), the action limit \( \alpha_i \) will become larger until it reaches \( \sqrt{c} \) at the last stage \( N - 1 \). As shown by Crowder [13], this “funneling effect” is significant only for short-run process or an extremely large adjustment cost. For a moderately long process, the optimal control limits at most stages can be well approximated by equation (7.15). Hence, the Constant Deadband Adjustment method works well in most finite-horizon problems.

Both the finite and infinite horizon solution to this problem require the prior knowledge of \( \sigma^2_i \) and \( \sigma^2_\nu \), which is used in the Kalman Filter estimates and in determining the control limits. When such knowledge is unavailable, the SMC method described in Section 7.1 can be utilized to estimate the process mean. The only question that remains is how to determine
the control limits. When $\sigma_\epsilon^2$ and $\sigma_\nu^2$ are unknown, the optimal control limits should depend on the posterior distributions of $\theta_i$, $\sigma_\epsilon^2$ and $\sigma_\nu^2$, in both the infinite-horizon and finite-horizon problem. It is very difficult to find such optimal limits for this problem, in which the posterior distributions can only be numerically computed based on a Monte-Carlo method.

We propose an Adaptive Deadband Adjustment method, in which the control limits are approximated based on the sequential Bayesian estimate of the process parameter $\sigma_\nu^2$. In particular, at each stage the SMC method gives the estimates of $\theta_i$ and $\sigma_\nu^2$, which are the means of their posterior distributions. The estimate of $\sigma_\nu^2$ can be substituted into (7.15) to calculate the control limit for the current stage. This limit, together with the estimate of the process mean, can be used in (7.14) to determine if an adjustment is needed. When more observations are available, the limits calculated in this way are expected to converge to the optimal constant limit for the infinite-horizon problem. An example is shown next to illustrate the calculation of the adaptive deadband.

**Example 7.1.** Consider a process with 500 stages and process parameters given by $\sigma_\epsilon = 4$ and $\sigma_\nu = 2$. The fixed adjustment cost is assumed to be $c = 100$. Suppose the process can be adjusted under the following three methods: (1) Constant Deadband adjustment method with the true values of the process parameters known; (2) Constant Deadband adjustment method with biased process parameters $\sigma'_\epsilon = 1$ and $\sigma'_\nu = 4$; and (3) Adaptive Deadband adjustment method with prior information also biased such that the means of the prior distributions of $\sigma'_\epsilon^2$ and $\sigma'_\nu^2$ equal to $\sigma_\epsilon^2$ and $\sigma_\nu^2$ respectively. The variances of the prior distri-
butions of both parameters are 4 times of the square of the means. The advantages in terms of estimation using the SMC method, which is used in the Adaptive Deadband method, has been shown in the previous section. The calculated control limits/half widths are shown in Figure 7.2. The Adaptive Deadband method starts with biased prior distributions, so the calculated control limits at the beginning are close to those in the Constant Deadband method where the same biased information is used. As the process continues, better estimates of $\sigma^2$ are obtained through the SMC method and they are used in the Adaptive Deadband method. The control limits converge to the constant control limit using the true process parameters. The total costs are calculated for each method by simulating the process using common random numbers. The total cost for method (1), which should the best method in

Figure 7.2: The control limits/half widths for the (1) Constant Deadband method with the true process parameters (2) Constant Deadband method with the wrong process parameters (3) Adaptive Deadband method with biased prior information
this case, is 13347; the total cost for method (2) is 15127; and the total cost for method (3) is only 13605. The Adaptive Deadband adjustment method with biased prior information induced a cost only slightly larger than the cost caused by the best method. When it is compared to the Constant Deadband method with the same biased prior information, the savings are significant.

7.3 Simulation Study

To deepen the comparison, more cases were investigated in this section. For the cases of processes in Table 7.2, each prior information scenario was used both in the Constant Deadband method and the Adaptive Deadband method. The fixed adjustment cost was assumed to be $c = 100$ in all cases. The variances of the prior distributions of the process parameters were always equal to 4 times of the square of the means. The relative savings induced by the Adaptive Deadband method were calculated with respect to the Constant Deadband method with the same prior information scenario using common random numbers. 20 replications were made to calculate the average savings and their 95% confidence intervals for each combination of a process case and a prior scenario. The results are shown in Figure 7.3.

The only prior information scenario that has negative average savings is scenario 5, where the true values of the parameters was used in the Constant Deadband method and hence it is the best method either in estimation or in determining the control limits. However, the difference between the best adjustment method and the Adaptive Deadband method is not
Figure 7.3: Average savings induced by the Adaptive Deadband method v.s. the Constant Deadband method for each combination of a prior scenario and a process case, and their 95% confidence intervals, when $c = 100$.

significant. For scenarios 1 and 9, where the Kalman filter estimates are accurate since the ratios $\frac{\sigma_e^2}{\sigma_v^2}$ that were used happened to be equal to the true values, the positive savings are mainly due to the improving control limits obtained in the Adaptive Deadband method. For scenarios 2 and 8, where the constant control limits used in the Constant Deadband method are optimal since the true value of $\sigma_v^2$ was used to determine the limits, the savings made by the Adaptive Deadband method are mainly due to the better process mean estimates obtained in the SMC method. For other scenarios, the savings can be considered to be due to a mixed effect of the control limit determination and the process mean estimation. Generally, when the adjustment cost $c$ is larger, the control limit will be larger and there will be fewer adjustments. Hence the effect of the process mean estimation will be less
dominant. Contrarily, when \( c \) is smaller, the effect of the process mean estimation will be more dominant than that of the control limit determination.

Figure 7.4 shows the average savings and their 95% confidence intervals induced by the Adaptive Deadband method versus the Constant Deadband method for the same scenarios and cases when a smaller adjustment cost \( c = 16 \) was considered. The savings for \( c = 16 \) are closer to those in Figure 7.1, where the savings are purely due to the process mean estimation. Specifically, for the scenarios when the Kalman filter estimates are more accurate (e.g., scenarios 1 and 9), the savings for \( c = 16 \) drop compared to those for \( c = 100 \). For the scenarios when the Kalman filter estimates are not accurate (e.g., scenarios 3 and 7), the savings for \( c = 16 \) are more significant than those for \( c = 100 \).

Figure 7.4: Average savings induced by the Adaptive Deadband method v.s. the Constant Deadband method for each combination of a prior scenario and a process case, and their 95% confidence intervals, when \( c = 16 \).


7.4 Conclusion

As demonstrated in Section 7.1, the Bayesian method using a SMC technique provides better estimates of the process means than the Kalman filter estimates does, in the cases that the prior information of the variance ratio $\sigma_i^2/\sigma_\nu^2$ is not available or biased. Such advantage in estimation exists in the process adjustment problems regardless of whether or not the adjustment cost is considered.

An Adaptive Deadband adjustment method was proposed in Section 7.2 based on the Bayesian estimation method. In contrast to the Constant Deadband method, the Adaptive Deadband method calculates the deadband limits at each step based on the Bayesian estimate of $\sigma_\nu^2$ and hence it gives more appropriate limits when the prior information of $\sigma_\nu^2$ is biased.

The simulation study in Section 7.3 shows that the Adaptive Deadband method has more robust performance than the Constant Deadband method in adjusting a process with random IMA(1,1) drifts in the presence of a fixed adjustment cost. Such robustness comes from the advantage of Bayesian applications in estimation and in determining the control limits.

A computer program that implements the Adaptive Deadband adjustment method developed in this chapter is listed in Appendix D and can also be downloaded from http://www.ie.psu.edu/researchLabs/EngineeringStatistics/.
Appendix of Chapter 7: Algorithm for estimating the process parameters $\theta_i$, $\sigma^2_u$ and $\sigma^2_\epsilon$

At the beginning of the process:

- draw $M$ random numbers $\theta_0^{(m)}$ from the prior distribution of $\theta_0$ (In this dissertation, $\theta_0^{(m)} = 0$ for all $m$);
- draw $M$ random numbers $\sigma_\epsilon^{2(m)}$ from the prior distribution of $\sigma_\epsilon^2$;
- draw $M$ random numbers $\sigma_u^{2(m)}$ from the prior distribution of $\sigma_u^2$,
  for $m = 1, 2, \ldots, M$.

Create an initial weight vector $(w_1, w_2, w_3, \ldots, w_M)$, where $w_m = 1/M$ for all $m$.

Iterations throughout all stages:

For $n = 1, 2, 3, \ldots, N$

Observe $Y_n$

Generate 1 random number $\theta_n^{(m)}$ from the distribution $N(\theta_{n-1}^{(m)}, \sigma_u^{2(m)})$ for each $m = 1, \ldots, M$.

Calculate the likelihood of the $m^{th}$ particle (combination $(\theta_i^{(m)}, \sigma_\epsilon^{2(m)}, \sigma_u^{2(m)})$):

$$L_m = \frac{1}{\sqrt{\sigma_\epsilon^{2(m)}}} \exp\left\{-\frac{(Y_i - \theta_i^{(m)})^2}{2\sigma_\epsilon^{2(m)}}\right\}$$

Update the new weight vector and normalize it:

$$w_m \leftarrow w_m \times L_m \text{ then } w_m \leftarrow w_m / \sum_{m=1}^{M} w_m.$$ 

Obtain the new parameter estimators:
\[ \hat{\theta}_n|D_n = \sum_{m=1}^{M} w_m \theta_n^{(m)} \]
\[ \hat{\sigma}_\epsilon^2|D_{ij} = \sum_{m=1}^{M} w_m \sigma^{2(m)}_\epsilon \]
\[ \hat{\sigma}_\nu^2|D_{ij} = \sum_{m=1}^{M} w_m \sigma^{2(m)}_\nu \]

Calculate the effective sample size factor:

\[ ESS = \frac{M}{1 + M \text{var}(w_m)} \]

If \( ESS < pM \), where \( p \) is a specified level between 0 and 1 (\( p = 0.5 \) was used in our computations),

rejuvenate \( M \) particles using the "1PFS" algorithm (reference [2]).
Chapter 8

Research Contributions and Further Research

This chapter summarizes the contributions of this research. Some further research directions are also suggested.

8.1 Research Contributions

This research mainly contributes to two major fields: Statistical Process Control (SPC) and applications of Bayesian statistics.

8.1.1 Contributions to Statistical Process Control

Several fundamental problems in the field of SPC are studied in this research. These problems cover three types of manufacturing basic processes:
• Single-lot processes with initial offsets,

• Multiple-lot processes with initial offsets,

• Processes with random drift.

Three types of cost functions were considered:

• Symmetric quadratic costs,

• Asymmetric quadratic costs,

• Fixed adjustment costs.

These problems have been noticed in literature and solved based on the assumption that the variance parameters are known. In this research we considered the situation when these parameters are not known and provided new solutions that can work properly in this situation. Bayesian models were introduced to give inference to the unknown parameters, based on which the adjustment methods were created. The major advantages of the new adjustment methods over the traditional approaches are summarized as follows:

• When no prior information on the process parameters is available, the traditional approaches that depend on such information can not be applied. The proposed Bayesian methods can still function by assigning vague prior distributions to the unknown parameters.
• When prior information on some process parameters is not available and any source of information is valuable, the new methods based on Bayesian models have the ability to take advantage of various forms of prior information.

• In the case that the prior information/estimate is likely to contain errors, the new adjustment methods have more robust performance compared to corresponding traditional approaches, mainly due to the adaptability of the Bayesian models.

• Even in the case that accurate prior information can be obtained for traditional methods, new methods still have competitive performance.

• The new methods estimate all process parameters while adjusting the controllable parameters. They produce better estimates in the case that classical sample-based estimation methods are facing difficulties, particularly in estimating variance parameters.

• Statistical inferences to unknown parameters are sequentially updated while the new methods are controlling processes, which benefits the latter adjustments in the processes. Moreover, the posterior distributions obtained in the end of a process can be utilized in future operations.

Above advantages are not limited to the problems considered in this dissertation. They can be generalized when solutions based on Bayesian models are obtained for other process control problems. More specific findings in the problems considered in this thesis are summarized as follows:
In the multiple-lot setup adjustment problems considered in Chapters 3 and 4, it was demonstrated that in most cases the proposed Bayesian methods have better performance in comparison with the traditional methods that require precise prior knowledge. The only case that the Bayesian methods have relatively poor performance happened when \( \mu \) (the mean of the distribution of the initial offsets) is extremely small. For this reason, improved Bayesian methods were created, in which the first adjustment in a lot is conditional on the statistical significance of the Bayesian estimate of \( \mu \), so that positive average savings can be gained in all cases studied.

In Chapter 5, we proved that the optimal adjustments have a deadband form for the setup adjustment problems with fixed adjustment costs. We also showed and explained the “U” shape of the deadband limits, which is related to the lack of the estimation precision at the beginning of a process and the horizon effect [13] at the end.

The asymmetric cost off-target function was considered for the setup adjustment problem in Chapter 6. We showed that the targets of the process means are different from the nominal value, and they can be found according to the posterior predictive distributions of the quality characteristics. The deviation of a target from the nominal value can be increased by the asymmetry of the cost function or the variance of the predictive distribution at the current stage.

A process with random IMA(1,1) drifts was studied in Chapter 7. It was shown that the Bayesian methods give better process mean estimates than a Kalman filter in the
cases that the prior knowledge of the variance parameters $\sigma^2/\epsilon \sigma_\nu^2$ is not available. In the presence of the fixed adjustment cost, the adjustments also have the deadband form. The deaband limits can be determined at each step through Box and Jenkins’ solution [5], a method that assumes infinite process length and known variances, by substituting the Bayesian estimates. Due to the adaptability of the Bayesian estimates, better limits can be obtained and hence the Bayesian methods provide better adjustments.

8.1.2 Contributions to Applications of Bayesian Statistics

The applications of Bayesian Statistics have increased dramatically in the past two decades. However, in the field of SPC, which is one of the most important applications of statistics in manufacture, the advantages of Bayesian methods have not been fully utilized.

Such situation may be due to the computational complexity of Bayesian models, which is obviously a restriction for Bayesian methods to be used for on-line adjustments. In this dissertation, we introduced the SMC method for Bayesian computation as an alterative to the MCMC method used previously. It was demonstrated that SMC methods can be used for on-line process control, and that at least for the problems under study, they work equally well than MCMC at a fraction of the computational cost. The speed and accuracy of SMC methods can be improved with the assistant of better computers and better algorithms. We believe there is a great potential for more applications of SMC methods in SPC, and this tendency will remain in the near future.

The development of fundamental science is always motivated by its applications. When
we applied Bayesian models to solve these new problems, some questions were still not answered. For instance, how to set up the prior distributions properly or how to find a vague prior distributions is still a challenge in Bayesian Statistics applications. In our research, we analyzed the sensitivity of various adjustment methods to the prior distributions and gave suggestions on how to determine the priors for specific problems. We expect this work can help the fundamental research in Bayesian statistics and further benefit applications in other fields.

This research also helps improving existing posterior distribution computation algorithms. When applying the SMC method, we modified the 1PFS algorithm of [2] such that the rejuvenation steps can work for half-space posterior distributions (e.g. in the case of the density for $\sigma^2$, which is positive only for the positive region).

### 8.2 Further Research

In this research, some fundamental problems in process control were tackled. Based on the results and following the same philosophy, we believe other similar process control problems with unknown parameters can also be solved by using Bayesian methods, specifically:

- In the class of asymmetric off-target costs, besides the quadratic cost, the asymmetric constant off-target cost is also often used to represent the different cost associated with violating LSL or USL. This type of cost is considered in [10] for setup adjustment problems under the assumption of known variance. Similarly to the derivation of the
solution of the asymmetric quadratic cost in Chapter 6, we could find an optimal setup adjustment problem under the asymmetric constant cost based on the assumption of unknown variance.

- Among the three types of cost functions, the symmetric quadratic cost and the fixed adjustment cost has been considered for both setup adjustment problems and processes with random drifts. However the asymmetric off-target cost function is only considered in the setup adjustment problems. To adjust a process with random drifts under asymmetric off-target cost is also of interest.

- In some cases, both of the asymmetric off-target cost and the fixed adjustment cost need to be taken into account together. The solution to a process control problem under these two costs could be numerically solved through dynamic programming formulation, similarly to that in Chapter 5. Intuitively, the solution should also have a deadband form with different control limits on different side of the nominal value and a sequence of optimal targets for different stages. However, it may be hard to mathematically prove the existence of such form and the computation complexity may be large.

- Another important adjustment cost, the quadratic adjustment cost, was not studied in this research. Such cost is proportional to the square of the adjustment magnitude, which is usually associated with energy. Process control problems with unknown parameters under this type of adjustment cost may be important to some industries.
• In Chapter 7, we considered a process with random drifts, which can be characterized by IMA(1,1) series without adjustments. In more general cases, a process may follow other type of time series when no adjustment is made. It is also interesting to adjust such process under a quadratic cost function and a fixed adjustment cost based on the assumption of unknown variances. The SMC method can again be used to estimate the process mean, which is relatively easy to implement. The only step left before finding a solution is to determine the optimal control limits.

• In this research, we assume that the adjustment is precise, i.e. there will not be errors introduced by the adjustments. Sometimes in practice, such errors are not negligible. Jensen and Vardeman [30] considered the adjustment errors in their solution to the control of a process with random IMA(1,1) drifts and deterministic drifts based on the known variances. In assumption of unknown variances, an adaptive control method to this process control problem may be created, similarly to the adaptive deadband method in Chapter 7.
Bibliography


Appendix A: List of Computer Programs used in Chapter 4

In this appendix, three MATLAB codes used in Chapter 4 are listed, in particular:

**SMCadjust.m**: the main code that allows a user to input the observations and recommends the adjustments. The size of the process and the prior distributions need to be specified at the beginning of the code prior to the execution.

**log_kitagawa_resample_move.m**: a MATLAB function for conducting the “rejuvenation” of the distributions.

**randomnorm.m**: a function for generating multi-variate normal random numbers.

**SMCadjust.m**

```
%Multi-lot, true parameters
I=3;   %Number of lots
J=3;    %Number of parts in a lot
mu=4;   %mean of the initial offsets

% prior distribution
%the prior dist of sigma^2_upsilon ~ lognorm(a1,b1), i.e. exp(norm(a1,b1))
a1=1.956011503;
b1=0.997965052;
%the prior dist of sigma^2_upsilon ~ lognorm(a2,b2)
a2=1.956011503;
b2=0.997965052;
%the prior dist of mu ~ norm(mu0,sigma0)
mu0=0;
sigma0=100;

% preference for the SMC step
N=10000;    %number of samples
```
% specific level to determine when to rejuvenate
% starting time

tol=0.5;    % specific level to determine when to rejuvenate
t0=cputime; % starting time

% Samples (particles) from prior
sample_s2_upsilon=exp(normrnd(a1,b1,1,N));
sample_s2_theta=exp(normrnd(a2,b2,1,N));
sample_mu=normrnd(mu0,sigma0,1,N);
weights1=ones(1,N)/N;  % initial weights
u=0; % level of controller

for i=1:I
    disp(’**************************************************’)
    disp([’Begin to adjust lot i=’ num2str(i)]);
    if i>2
        smean_mu=sample_mu*weights1’;
        disp([’Please set the controller to be u=’ num2str(-smean_mu) ’ before the 1st part is processed’]);
        u=-smean_mu;
    else
        disp([’Please reset the controller to be u=0 before the 1st part is processed’]);
    end
    % sample from the prior distribution of the initial offset theta_i
    sample_theta=normrnd(sample_mu+u,sqrt(sample_s2_theta));
    for j=1:J
        disp([’Please process part ’ num2str(j) ’ in lot ’ num2str(i)’]);
        disp([’Please process part ’ num2str(j) ’ in lot ’ num2str(i)’]);
        y=input([’y(‘ num2str(i) ’,’ num2str(j) ’)=’]);
        % importance weights
        like1=1./sqrt(sample_s2_upsilon).*exp(-((y-sample_theta).^2)./2./sample_s2_upsilon);
        weights1=weights1.*like1;
        weights1=weights1./sum(weights1);
        ESS1(i,j)=1./((1+N*N*var(weights1)));
        % calculate the mean and stdev of the samples.
        if j<J
            smean_theta=weights1*sample_theta’;  % estimate of theta (mean of the posterior distribution of theta_i)
            disp([’Please make an adjustment by amount U(’ num2str(i) ’,’ num2str(j) ’)=’ num2str(-smean_theta) ’ before the next part is processed’]);
            sample_theta=sample_theta-smean_theta;
        end
        % rejuvenation
        if ESS1(i,j)<tol %if ESS/N drop below a specified level, call rejuvenation steps
            newsample=log_kitagawa_resample_move(weights1,[sample_theta;sample_s2_upsilon;sample_mu;sample_s2_theta]);
            weights1=ones(1,N)/N;
            flag_distr=true;
            sample_theta=newsample(:,1)’;
        end
    end
end
log_kitagawa_resample_move.m

function [new_betas]=log_kitagawa_resample_move(wts, betas)
% Rejuvenation steps
% Requires randnorm() from Lightspeed

dims2=size(betas); %getting size of the number of particles
p=dims2(2);
M=dims2(1);
W=cumsum(wts);
W=W./W(M);
u=rand(1,M+1);
T= cumsum(u);
% update posterior distribution by resampling
k=1;
m=1;
while m<M+1
    if (W(m)*T(M+1)) > T(k)
        res_betas(k,:)=betas(m,:);
        k = k+1;
    else
        m = m+1;
    end
end
%the second and the forth parameters are defined on (0,Infty)
res_betas(:,2)=log(res_betas(:,2));
res_betas(:,4)=log(res_betas(:,4));
theta_bar=mean(res_betas,1);
V=cov(res_betas);
%From Stavropoulos and Titterington
bm_sq=(4/((p+2)*M))^(2/(p+4));
a=sqrt(1-bm_sq);
sig = bm_sq * V;
m = a * res_betas + repmat((1 - a) * thetabar, M, 1); % Shift the kernel means
%m = res_betas;
new_betas = randnorm(M, m, [], sig);
new_betas(:, 2) = exp(new_betas(:, 2));
new_betas(:, 4) = exp(new_betas(:, 4));

randomnorm.m

function x = randnorm(n, m, S, V)
% RANDNORM      Sample from multivariate normal.
% RANDNORM(n,m) returns a matrix of n columns where each column is a sample
% from a multivariate normal with mean m and unit variance.
% RANDNORM(n,m,S) specifies the standard deviation, or more generally an
% upper triangular Cholesky factor of the covariance matrix.
% This is the most efficient option.
% RANDNORM(n,m,[],V) specifies the covariance matrix.

if nargin == 1
  x = randn(1, n);
  return;
end
[d, nm] = size(m);
x = randn(d, n);
if nargin > 2
  if nargin == 4
    if d == 1
      S = sqrt(V);
    else
      S = chol(V);
    end
  end
  if d == 1
    x = S .* x;
  else
    x = S' .* x;
  end
end
if nm == 1
  x = x + repmat(m, 1, n);
else
  x = x + m;
end
Appendix B: List of Computer Programs used in Chapter 5

In this appendix, two R source codes used in Chapter 5 are listed, which are:

**chart.R**: R code that computes the control table for unknown-variance case,

**chart_n.R**: R code that computes the control table for unknown-variance case.

---

**chart.R**

```r
#Purpose: generate control table containing \( \alpha_i_{\sigma_i} \) for unknown-variance model in Chapter 5.
#Input: input parameters should be provided by a row of numbers in file "input1.dat" separated by spaces.
#   From left to right respectively, they are
#   N: number of parts to produce
#   C: fixed adjustment cost
#   \( \nu_0 \): prior parameter, \( \nu_0 \) in Chapter 5
#   \( \kappa_0 \): prior parameter, \( \kappa_0 \) in Chapter 5
#   range.mu: the upper limit of \( \mu_i' \), i.e. \( n_1*d_\mu \) in Chapter 5
#   incr.mu: the increment of discrete \( \mu_i' \), i.e. \( d_\mu \) in Chapter 5
#   range.sigma: the upper limit of \( \sigma_i' \), i.e. \( n_2*d_\sigma \) in Chapter 5
#   incr.sigma: the increment of discrete \( \sigma_i' \), i.e. \( d_\sigma \) in Chapter 5
#   Rep: number of replications used in one monte carlo integration
#   Example: for example 1 in Chapter 5, the input file "input1.dat" should contain following content:
#       10 9 2.01 1 5 0.1 10 1 10000
#Output: The output, a row of number \( \alpha_i_{\sigma_i} \) will be in file "control1.dat" by the end of the program.

#functions for transfering a real number to integer
mu.code <- function(x)
{
  i <- pmin(ceiling((abs(x)+incr.mu/2)/incr.mu),N.mu)
  return(i)
}
sigma.code <- function(x)
{
  i <- pmin(ceiling(x+incr.sigma/2)/incr.sigma,N.sigma)
}
```

---
#functions for transferring an integer to a real number
mu.decode <- function(i)
{
  x <- (i-1)*incre.mu
  return(x)
}

sigma.decode <- function(i)
{
  x <- (i-1)*incre.sigma
  return(x)
}

#read the input parameters
input <- scan("input1.dat")
N <- input[1]
C <- input[2]
nu0 <- input[3]
kappa0 <- input[4]
rangemu <- input[5]
incre.mu <- input[6]
rangesigma <- input[7]
incre.sigma <- input[8]
Rep <- input[9]

N.mu <- rangemu/incre.mu+1
N.sigma <- rangesigma/incre.sigma+1
loss <- array(0,dim=c(N.mu,N.sigma))
band <- array(0,dim=c(N,N.sigma))
t.start<-Sys.time()

# at final stage N
nu <- nu0+(N-1)
kappa <- kappa0+(N-1)

for(j in 1:N.sigma)
{
  is.control <- FALSE
  loss.upper <- 0
  sigma <- sigma.decode(j)
  #var <- sigma^2*nu/(nu-2)*(1+1/kappa)
  for(i in 1:N.mu)
  {
    if(is.control)
    {
      loss[i,j] <- loss.upper
    }
else{
    mu <- mu.decode(i)
    mean <- mu

    if(mean^2>=C)
    {
        loss[i,j] <- C+var
        loss.upper <- loss[i,j]
        band[N,j] <- mean
        is.control <- TRUE
    }
    else{
        loss[i,j] <- mean^2+var
    }
}

value.file <- paste("value1",as.character(N),"dat",sep=".")
write.table(loss,value.file,quote=FALSE)

# Other Stages
for (n in (N-1):1)
{
    print(n)
    loss.next <- loss
    nu <- nu0+(n-1)
    kappa <- kappa0+(n-1)
    for(j in 1:N.sigma)
    {
        is.control <- FALSE
        loss.upper <- 0
        sigma <- sigma.decode(j)
        #var <- sigma^2*nu/(nu-2)*(1+1/kappa)
        for(i in 1:N.mu)
        {
            if(is.control)
            {
                loss[i,j]=loss.upper
            }
        }
    }

    mu <- mu.decode(i)
    mean <- mu
    if(sigma!=0)
    {
        sum.n <- 0
        sum.y <- 0
        for (r in 1:Rep)
        {
y1 <- rt(1,nu)*sigma+mu
mu.new.n <- (kappa*mu+y1)/(1+kappa)
mu.new.y <- mu.new.n-mean
sigma.new <- sqrt((nu*sigma^2+kappa/(kappa+1)*(y1-mu)^2)/(nu+1))
mu.i.new.n <- mu.code(mu.new.n)
mu.i.new.y <- mu.code(mu.new.y)
sigma.j.new <- sigma.code(sigma.new)
loss.sim.n <- loss.next[mu.i.new.n,sigma.j.new]
loss.sim.y <- loss.next[mu.i.new.y,sigma.j.new]
sum.n <- sum.n+loss.sim.n
sum.y <- sum.y +loss.sim.y
loss.mean.n <- sum.n/Rep
loss.mean.y <- sum.y/Rep
}
else
{
mean <- mu
#var <- 0
mu.i.n <- i
mu.i.y <- mu.code(0)
sigma.j <- j
loss.mean.n <- loss.next[mu.i.n,j]
loss.mean.y <- loss.next[mu.i.y,j]
}
if(loss.mean.y+C<=loss.mean.n+mean^2)
{
loss[i,j] <- loss.mean.y+C+var
is.control <- TRUE
loss.upper <- loss[i,j]
band[n,j] <- mu
}
else
{
loss[i,j] <- loss.mean.n+mean^2+var
}
}
value.file <- paste("value1",as.character(n),"dat",sep=".")
write.table(loss,value.file,quote=FALSE)
}
#Record the control table.
control.file <- paste("control1","dat",sep=".")
write.table(band,control.file,quote=FALSE)
t.total<-Sys.time()-t.start
chart_n.R

#Purpose: generate control table containing \( \alpha_i \) for known-variance model in Chapter 5
#Input: input parameters should be provided by a row of numbers in file "input2.dat" separated by spaces.
# From left to right respectively, they are
# N: number of parts to produce
# C: fixed adjustment cost
# sigma: the revealed standard deviation of part-to-part error, \( \sigma_\upsilon \) in Chapter 5.
# prec.0: the initial precision, i.e. \( 1/\tau_0^2 \) in Chapter 5
# range.mu: the upper limit of \( \mu_i' \), i.e. \( n*d_\mu \) in Chapter 5
# incre.mu: the increment of discrete \( \mu_i' \), i.e. \( d_\mu \) in Chapter 5
# Rep: number of replications used in one monte carlo integration
# Example: for example 2 in Chapter 5, the input file "input2.dat" should contain following content:
# 10 9 1 5 0.1 10000
#Output: The output, a row of number \( \alpha_i, i=1,...,N \) will be in file "control2.dat" by the end of the program.

#functions for transfering a real number to integer
mu.code <- function(x)
{
  i <- pmin(ceiling((abs(x)+incre.mu/2)/incre.mu),N.mu)
  return(i)
}
#functions for transfering an integer to real
mu.decode <- function(i)
{
  x <- (i-1)*incre.mu
  return(x)
}
#Read the input parameters
input <- scan("input2.dat")
N <- input[1]
C <- input[2]
sigma <- input[3]
prec.0 <- input[4]
range.mu <- input[5]
incre.mu <- input[6]
Rep <- input[7]

N.mu <- range.mu/incre.mu+1 #calculate the number of levels of mu
loss <- array(0, dim=c(N, mu, N)) # matrix loss, the expected loss for each stage, each mu
band <- array(0, dim=c(N)) # half-width of deadband for each stage

t.start <- Sys.time()

# at final stage N
prec = prec.0 + (N-1)/sigma^2
is.control <- FALSE
loss.upper <- 0
for(i in 1:N, mu)
{
  if(is.control)
  {
    loss[i, N] <- loss.upper
  }
  else{
    mu <- mu.decode(i)
    if(mu^2 >= C)
    {
      loss[i, N] <- C
      loss.upper <- loss[i, N]
      band[N] <- mu
      is.control <- TRUE
    }
    else{
      loss[i, N] <- mu^2
    }
  }
}

# Other Stages
for(n in (N-1):1)
{
  print(n)
  prec = prec.0 + (n-1)/sigma^2
  is.control <- FALSE
  loss.upper <- 0
  for(i in 1:N, mu)
  {
    if(is.control)
    {
      loss[i, n] <- loss.upper
    }
    else{
      mu <- mu.decode(i)
    
      sum.n <- 0
      sum.y <- 0
    }
  }
}
for (r in 1:Rep)
{
  y.r <- rnorm(1,mu,sqrt(sigma^2+1/prec))
  mu.new.n <- (mu+1/prec/sigma^2*y.r)/(1+1/prec/sigma^2)
  mu.new.n <- mu.new.n-mu
  mu.i.new.n <- mu.code(mu.new.n)
  mu.i.new.y <- mu.code(mu.new.y)
  sum.n <- sum.n+loss[mu.i.new.n,n+1]
  sum.y <- sum.y +loss[mu.i.new.y,n+1]
  }
  loss.mean.n <- sum.n/Rep
  loss.mean.y <- sum.y/Rep

if(loss.mean.y+C<=loss.mean.n+mu^2)
{
  loss[i,n] <- loss.mean.y+C
  is.control <- TRUE
  loss.upper <- loss[i,n]
  band[n] <- mu
}
else
{
  loss[i,n] <- loss.mean.n+mu^2
}
}

value.file <- paste("value2","dat",sep=".")
write.table(loss,value.file,quote=FALSE)
#Record the control table.
control.file <- paste("control2","dat",sep=".")
write(band,control.file,ncolumns=N)
t.total<-Sys.time()-t.start
print(t.total)
Appendix C: List of Computer Programs used in Chapter 6

In this appendix, three MATLAB source codes used in Chapter 6 are listed, which are:

**AsymmetricAdjust.m**: the main MATLAB code that provides an interface for a user to enter observations of the quality characteristics, based on which the program gives the recommendations for adjustments. Users need to specify the number of parts in the process, the asymmetric cost function and the prior distributions in the code.

**Optimal_t.m**: a MATLAB function to search for the optimal target given a posterior predictive t-distribution.

**D1EC_t.m**: a function calculating the first derivative of the $E(C_i)$ for given value of $\mu$.

AsymmetricAdjust.m

N=5; %Number of parts in the process

%Specification of the cost function
c1=3;
c2=1;

%Specification of the prior distribution
kappa(1)=0; %kappa_0 in Chapter 6
mu(1)=0; %mu_0 Chapter 6
ratio=200; %the ratio between the prior variance and the square of the prior mean
mean=2; %the prior mean

%Calculate nu_0 and sigma_0 in Chapter 6
nu(1)=4+2/ratio;
sigma(1)=sqrt(mean)*sqrt((nu(1)-2)/nu(1));
%Initialization: the first adjust before the first part is always 0
U(1)=0;

disp('Start the process');
for i=1:(N-1)
    disp('************************************');
    disp(['please enter the observation of part ' num2str(i)]);
    Y(i)=input(['y(' num2str(i) ')=']);
    kappa(i+1)=kappa(i)+1;
    nu(i+1)=nu(i)+1;
    mu(i+1)=(kappa(i)*(mu(i)+U(i))+Y(i))/kappa(i+1);
    sigma(i+1)=sqrt((nu(i)*sigma(i)^2+kappa(i)/kappa(i+1)*(Y(i)-mu(i)-U(i))^2)/nu(i+1));

    %finding the optimal target
    mu_o(i)=Optimal_t(c1,c2,nu(i+1),sigma(i+1)*sqrt((kappa(i+1)+1)/kappa(i+1)));
    %Make the adjustment
    U(i+1)=mu_o(i)-mu(i+1);
    disp(['Please make an adjustment by amount U(' num2str(i) ')=' num2str(U(i+1)) ', then start to process the next part.']);
end

Optimal_t.m

function u0=Optimal_t(c1,c2,v,sigma)
% Search for the optimal target for given predictive t-distribution

% Intialization
% One of the initial lower/upper bounds of the target is always 0
u1=0;
sign1=sign(D1EC_t(c1,c2,v,sigma,u1));

% Find the other initial lower or upper bound of the target
% first trial
if c1==c2
    u0=0;
    return
elseif c1>c2
    u2=1;
else
u2=-1;
end

% find the initial lower/upper bound
while sign(D1EC_t(c1,c2,v,sigma,u2))==sign1;
    u2=2*u2;
end

% line search for the target
while abs(u2-u1)>0.001;
    u_temp=(u2+u1)/2;
    if sign(D1EC_t(c1,c2,v,sigma,u_temp))==sign1
        u1=u_temp;
    else
        u2=u_temp;
    end
end
% solution
u0=(u2+u1)/2;

D1EC_t.m

function D1EC=D1EC_t(c1,c2,v,sigma,u)
% Calculate the first derivative of E(C) for given mu
D1EC=2*c2*u+2*(c1-c2)*(u*tcdf(-u/sigma,v)+Gamma((1+v)/2)/sqrt(pi)/Gamma(v/2)*sigma*sqrt(v)/(1-v)*(u^2/sigma^2/v+1)^(1/2-v/2));
Appendix D: List of Computer Programs used in Chapter 7

In this appendix, two MATLAB codes used in Chapter 7 are listed, which are:

AdaptiveAdjust.m: The main MATLAB code that provides an interface for a user to enter observations of the process means, based on which the program gives the recommendations for adjustments. The specifications of the number of stages, the adjustment cost and the prior distributions need to be given in the code.

log_kitagawa_resample_move.m: a MATLAB function for conducting the “rejuvenation” of the distributions. This code is slight different from the code log_kitagawa_resample_move.m in Appendix A due to the difference in the process models. In addition to above two programs, the MATLAB function randomnorm.m in Appendix A is also required.

AdaptiveAdjust.m

%Number of stages in the process
N=100;

% the fixed adjustment cost
c=16;

% specification of the prior distributions
% prior for sigma^2_epsilon ~ exp(N(a1,b1^2))
ratio1=4; %variance/mean^2
mean1=4; %the prior mean

% prior for sigma^2_nu ~ exp(N(a2,b2^2))
ratio2 = 4; % variance/mean^2
mean2 = 4; % the prior mean

% Calculate the variables characterizing the priors
b1 = sqrt(log(ratio1 + 1));
a1 = log(mean1) - 0.5 * b1^2;
b2 = sqrt(log(ratio2 + 1));
a2 = log(mean2) - 0.5 * b2^2;

% Number of particles for computing the posteriors
n = 10000;
% Specified number for justifying rejuvenation
tol = 0.5;

% Initialization
Y(1) = NaN;
smean_theta(1) = 0; % estimate by Bayesian method
U(1) = 0; % adjustment by Bayesian

% Samples from prior
sample_s2_epsilon = exp(normrnd(a1, b1, 1, n));
sample_s2_nu = exp(normrnd(a2, b2, 1, n));
sample_theta = zeros(1, n);
weights1 = ones(1, n) / n;

% Start Process
disp('Start the process');
for i = 2:N
    disp('************************************');
    disp(['please enter the observation of part ' num2str(i - 1)]);
    Y(i) = input(['y(' num2str(i - 1) ')=']);
    sample_theta = normrnd(sample_theta + U(i - 1), sqrt(sample_s2_nu));
    like1 = 1 ./ sqrt(sample_s2_epsilon) .* exp(- (Y(i) - sample_theta).^2 ./ sample_s2_epsilon);
    weights1 = weights1 .* like1;
    weights1 = weights1 ./ sum(weights1);

    % Compute ESS/n
    ESS1(i) = 1 / (1 + n * n * var(weights1));
    if (ESS1(i) < tol)
        newsample = log_kitagawa_resample_move(weights1, [sample_theta; sample_s2_epsilon; sample_s2_nu]);
        weights1 = ones(1, n) / n;
        sample_theta = newsample(:, 1);
        sample_s2_epsilon = newsample(:, 2);
        sample_s2_nu = newsample(:, 3);
    end
\%calculate the mean of the sample theta.
smmean_theta(i)=weights1*sample_theta';
smmean_s2_nu(i)=weights1*sample_s2_nu';
smmean_s2_epsilon(i)=weights1*sample_s2_epsilon';
k_b(i)=((6*c/smmean_s2_nu(i))^(1/4)-0.63)*sqrt(smmean_s2_nu(i));
if abs(smmean_theta(i))>k_b(i)
    U(i)=-smmean_theta(i);
else
    U(i)=0;
end
disp(['Please make an adjustment by amount U(' num2str(i) ')=' num2str(U(i)) ', then resume the process till the next stage.']);
end

log_kitagawa_resample_move.m

function [new_betas]=log_kitagawa_resample_move(wts, betas)

    %function [new_betas]=log_kitagawa_resample_move(wts, betas)
    % Based on Liu and West Kernel density methods.
    %
    % betas = D1&part(D2) based Particles
    % matrix size = Mx10
    % wts = Wts after incorporating observations
    % vector size = Mx1
    % new_betas = Resampled and moved 'New' particles
    % matrix size = Mx10
    % Requires randnorm() from Lightspeed

dims2=size(betas); %getting size of the number of particles
p=dims2(2);
M=dims2(1);

W=cumsum(wts);
W=W./W(M);
u=randn(1,M+1);
T= cumsum(u);
% update posterior distribution by resampling
    k=1;
    m=1;
    while m<M+1
        if (W(m)*T(M+1)) > T(k)
            res_betas(k,:)=betas(m,:);
        end
        m=m+1;
    end

new_betas=res_betas(1:M,:);
end
\begin{verbatim}
    k = k+1;
    else
        m = m+1;
    end
end

res_betas(:,2)=log(res_betas(:,2));
res_betas(:,3)=log(res_betas(:,3));
theabar=mean(res_betas,1);

V=cov(res_betas);

%From Stavropoulos and Titterington
bm_sq=(4/((p+2)*M))^(2/(p+4));
a=sqrt(1-bm_sq);
sig=bm_sq*V;

m=a*res_betas+repmat((1-a)*thetabar,M,1); %Shift the kernel means
%m=res_betas;
new_betas=randnorm(M,m,[],sig);
new_betas(:,2)=exp(new_betas(:,2));
new_betas(:,3)=exp(new_betas(:,3));
\end{verbatim}
Vita

Zilong Lian was born in Weihai, China. He received a Bachelor of Science degree in Physics from Peking University in 1999. He studied in Columbia University and obtained a degree of Master of Science in Operations Research. Since January 2002, he has been a graduate student in the Department of Industrial & Manufacturing Engineering at the Pennsylvania State University. During the time, he worked as a research assistant in the Engineering Statistics Lab under Professor Enrique del Castillo, focused on developing new statistical process control methods. At a more general level, his research interests involve developing and improving statistical methods that are used in all aspects of the manufacturing industry. From the research presented in this dissertation, he has submitted five papers for publication in refereed journals. He is expected to receive a dual degree of Doctor of Philosophy in Industrial Engineering and Operations Research with a minor degree in Statistics in May 2005.