The Pennsylvania State University

The Graduate School

Department of Industrial and Manufacturing Engineering

STATISTICAL PROCESS ADJUSTMENT METHODS FOR

QUALITY CONTROL IN SHORT-RUN MANUFACTURING

A Thesis in

Industrial Engineering

by

Rong Pan

© 2002 Rong Pan

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2002

We approve the thesis of Rong Pan.

Date of Signature

_____   _____

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

_____   _____

Richard A. Wysk
Professor of Industrial and Manufacturing Engineering
Leonhard Chair, Department of Industrial and Manufacturing Engineering

_____   _____

M. Jeya Chandra
Professor of Industrial and Manufacturing Engineering

_____   _____

Bruce G. Lindsay
Professor of Statistics
Distinguished Professor, Department of Statistics

_____   _____

Richard J. Koubek
Professor of Industrial and Manufacturing Engineering
Head of the Harold and Inge Marcus Department of
Industrial and Manufacturing Engineering

Abstract

Process adjustment techniques based on the feedback control principle have become popular among quality control researchers and practitioners, due to the recent interest on integrating Statistical Process Control (SPC) and Engineering Process Control (EPC) techniques. Traditionally, quality engineers, who are more familiar with SPC methods, avoid using process adjustment methods because of process tampering concerns. This has resulted in very few systematic studies on how to apply process adjustment strategies for continuous quality improvement. Most of the work in this area concentrates on chemical processes which typically have long production runs. This thesis focuses on studying sequential adjustment methods, closely related to well-known Stochastic Approximation procedures, for the purpose of quality control of a short-run manufacturing process.

First, the problem of adjusting a machine that starts production after a defective setup operation is considered. A general solution based on a Kalman Filter estimator is presented. This solution unifies some well-known process adjustment rules, and is a particular case of Linear Quadratic (LQ) control methods. In essence, this solution calls for a sequential adjustment strategy which recursively calculates the value of an adjustable variable according to the prior knowledge of this variable and the most recent observation from the process.

Next, the integration of sequential adjustments with SPC control charts are investigated for controlling an abrupt step-type process disturbance on a manufacturing

process. The performance of this type of integrated methods depends on the sensitivity of the control chart to detect shifts in the process mean, on the accuracy of the initial estimate of shift size, and on the number of sequential adjustments that are made. It is found that sequential adjustments are superior to single adjustment strategies for almost all types of process shifts and shift sizes considered. A combined CUSUM chart plus sequential adjustments approach has better performance than other methods when the shift size is not very large.

If there are different costs associated with a higher-than-target quality characteristic compared to a lower-than-target quality characteristic, that is, an asymmetric cost function, the adjustment rule needs to be modified to avoid the quality characteristic falling into the higher cost side. For this case, a sequential adjustment rule with an additional bias term is proposed. A method to determine these bias terms is developed. Furthermore, the effect of process measurement and adjustment costs on the decision of whether or not to apply adjustment actions at each sampling instant is investigated. A modified Silver-Meal scheduling algorithm is found to be good at providing robust and close-to-optimal adjustment schedules for this problem.

Finally, methods for identifying and fine-tuning a manufacturing system operating in closed-loop are studied. When a process is operated under a linear feedback control rule, the cross-correlation function between the process input and output has no information on the process transfer function, and open-loop system identification techniques cannot be used. In this research, it is shown that under certain general assumptions on the controller and process disturbance structure, it is possible to identify the process disturbance models from data obtained under closed-loop operation. After

identification, it is proposed to tune the controller to a near-optimal setting according to a performance criterion that considers both the variance of the output and the variance of the adjustments.

In summary, a collection of mathematical models for short-run manufacturing processes are proposed and studied systematically in this thesis. It is demonstrated that by implementing proper adjustment strategies the stability of the process can be better maintained; thus, significant economic benefits obtained from the consistent quality of products will be achieved. This research contributes directly to the quality improvement program of the manufacturing industry and to the field of applied statistics.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to express my deep gratitude to my advisor, Dr. Enrique del Castillo. I would not be able to finish my graduate study without his ceaseless support. His advice and encouragement also guide me to my future career in academia. Gracias.

I want to send my special thanks to Dr. Richard A. Wysk for his advising in my first one-and-half-year study at Penn State and for encouraging me to pursue the research topic that I like. He was also in my thesis committee. I would like to thank the other committee members, Dr. Bruce G. Lindsay and Dr. M. Jeya Chandra, for their invaluable feedbacks and suggestions throughout the course of this thesis development.

There are many other individuals that I need to say thanks to: my lab mates and friends - Suntara, Guillermo, Ramkumar, and other friends - Hong, Yingzi, Celick, Ronn, Sree, Son, Lee, Qing, Jerry, and Kaizhi. They made my experience at Penn State more memorable. In particular, I must send my sincere appreciations to one colleague and friend, Dr. Bianca Colosimo, for her terrific enthusiasm and productivity on leading our collaborations of several research papers.

Finally, I want to thank my parents, my sister, Bin, and my brother, Kai, for their spiritual and financial supports at my good time and bad time throughout my life.

Thank you all.

To my parents

Chapter 1

Introduction

1.1   Motivation

With growing competition in consumer markets, today's manufacturers over-whelmingly rely on the quality of their products and service for survive and success. This research is motivated by the characteristics of modern manufacturing environments, where the life cycle of products has decreased rapidly and customized short-run manufacturing processes have become quite common for achieving customer satisfactions. The objective of this thesis is to develop and study some statistical process adjustment strategies used to maintain a manufacturing process on its stable and desirable level of performance.

An industrial process is any production activity that comes in physical contact with hardware or software that will be delivered to an external customer (Harrington [47]). Although it is commonly accepted that in order to produce high-quality products it is critical to maintain the stability of the manufacturing process and to make it robust to external disturbances that may drive the process off-target, a traditionally-trained quality engineer or quality manager does not consider process adjustments as potential tools for continuous quality improvement. On the other hand, the techniques of Automatic Process Control (APC), or Engineering Process Control (EPC), have been long adopted by process engineers, who mostly concentrate on on-line process adjustments

according to the type of disturbance on the process. The barrier between these two groups is mainly caused by their different interpretations of process models instead of their ultimate goals.

In this thesis, a typical manufacturing process is viewed as a stochastic process, in which, due to the inherent process variation and measurement variation, some process parameters cannot be observed directly. Traditional Statistical Process Control (SPC) methods assume that as long as the process is stable and within statistical control, the remaining observed variability is uncontrollable; therefore, no further process adjustment actions are able to reduce the process variance. Even when the process is deemed to be out-of-control, no adjustment strategies are explicitly specified in the SPC literature. However, in a modern manufacturing environment, many assumptions used by SPC are no longer valid. First, new manufacturing methods and sensor technologies often imply that the values of the quality characteristic of a process are serially correlated in time. Therefore, by applying a properly designed adjustment scheme, process variance revealed by the quality data could be reduced. Second, in order to satisfy more and more specific customer demands, customized short-run manufacturing operations have been widely adopted by many companies. Short-run manufacturing processes imply the high frequency of process setups, and thus a higher possibility of systematic process errors, which, in turn, could drastically deteriorate product quality if the process is not adjusted in time.

The technologies for implementing on-line adjustments on a manufacturing line are available and usually are inexpensive. It is reasonable for a quality engineer to be aware of their importance instead of ignoring them.

1.2   Identification of the Problems under Study

1.2.1   Process monitoring

Traditional SPC methods provide a group of statistical tests of a general hypothesis – that the mean value of the quality characteristic of a process, or process mean for short, is consistently on its target level. A variety of graphical tools have been developed for monitoring a process mean, e.g., Shewhart charts, CUSUM charts, and EWMA charts.

Shewhart charts, first introduced by W.A. Shewhart [87] in 1930's, plot either the individual process measure or the average value of a small sample group (usually not more than five samples) along with the target level and control limits. Under the assumption that the plotted data are normally distributed around the process target when the process is within statistical control, the possibility of observing a point that is out of the three-sigma control limits is less than 2.7 in a thousand. Therefore, when an out-of-control-limit point is indeed found, it signals an out-of-control alarm and calls for an investigation on the process. In principle, a Shewhart chart can be used for detecting any type of process abnormality; however, it is not the most effective tool for some common process errors, such as a small shift in the process mean.

The process mean is desired to be maintained at its target level consistently; however, random process errors, or random "shocks", could shift the process mean to an unknown level. A control chart is required to detect this shift as soon as possible. At the same time, it should not signal too many false alarms when the process mean is on the target. These criteria are usually defined in terms of the Average Run Length of the

control chart for the in-control operation and out-of-control operation of the process, i.e., in-control ARL and out-of-control ARL, respectively. When a shift in the process mean is to be detected and the size of the shift is known, then a cumulative sum (CUSUM) control chart is the most efficient method according to its ARL properties (Basseville and Nikiforov [9]).

CUSUM charts were first introduced by Page [77] and its procedure can be seen as equivalent to applying a sequential likelihood ratio test for a shift in the process mean (see, e.g., Barnard [8], Ewan and Kemp [38]). A CUSUM chart monitors the accumulated process deviation after the process is determined to be in the in-control state. The parameters of a CUSUM chart can be assigned such that this chart is the optimal likelihood ratio test on a particular shift size at each time.

The Exponentially Weighted Moving Average (EWMA) chart can be seen as a variation of the CUSUM control scheme. Based on the assumption that the most recent observed process deviation can have more information on process errors than the previous deviations, we may assign different weights to data according to their recorded time. An EWMA scheme lets the weights decrease exponentially with the age of each point, while a CUSUM scheme keeps the same value for the weights. A Shewhart scheme, in contrast, assigns the total weight to the most recent observation and zero to others. Usually, an EWMA chart can be designed to have similar ARL properties as a CUSUM chart through simulation studies (Crowder [25]).

For the quality control of a manufacturing process, one essential task is to detect and remove any possible abnormal change in the process mean. However, a control chart alone does not explicitly provide a process adjustment scheme even when the process

mean is deemed to be off-target. The lack of an adjustment scheme in SPC applications may cause a large process off-target cost, a problem of particular concern in short-run manufacturing. Therefore, it is of importance to explore some on-line process adjustment methods that are able to remove various types of process errors quickly and keep the process mean right on a desirable level with relatively small effort.

1.2.2   Process adjustment

Suppose a shift in the mean of a process has been detected using a SPC scheme. The next task is to adjust the process. There are two generic steps in process adjustment: one is to estimate the shift size in the process mean and the other one is to take an action on the manufacturing process to remove the error. Since the shift size is an unknown parameter due to the inherent process variation and measurement variation, it has to be estimated based on the data collected on-line. Therefore, a good process adjustment strategy must take the uncertainty of the parameter estimate into consideration. It seems reasonable to apply a sequence of adjustments, which refine the estimate of the shift size over a period of time when an initial parameter estimate is imprecise.

In this thesis, we consider a sequential process adjustment strategy based on the Stochastic Approximation (SA) technique. Stochastic approximation was first introduced by Robbins and Monro [79], who studied the following problem. Suppose x is a controllable factor taking values on the real line and y is a measurable process output that is desired to be on target T. Assume that the following relation holds:

$$y_t = M(x_{t-1}) + \varepsilon_t \tag{1.1}$$

where $M(\cdot)$ is a general regression function of y on x. A value θ corresponds to a value of x such that $M(x) = T$, and ε_t is a random variable with mean value of zero. The question is how to design a sequence of input variables x_t so that y_t will converge to its target in mean square. After further assuming that the regression function is such that $M(x) \leq 0$ if $x < \theta$ and $M(x) \geq 0$ if $x > \theta$, and that the distribution function of ε has finite tails, Robbins and Monro designed the controllable factor x in the following recursive fashion (this is the so-called R-M process):

$$x_{t+1} = x_t - a_t(y_t - T), \tag{1.2}$$

where the coefficient sequence $\{a_t\}$ must satisfy the following conditions:

$$a_t \to 0, \quad \sum_{t=1}^{\infty} a_t = \infty, \text{ and } \sum_{t=1}^{\infty} a_t^2 < \infty.$$

These conditions guarantee the convergence of x_t to the root θ in mean square, i.e., $lim_{n\to\infty} E[(x_t - \theta)^2] = 0$. More SA procedures and asymptotic convergence properties of SA were developed in subsequent research (see, e.g., Blum [11], Chung [23], Dvoretzky [36], Fabian [39], Lai and Robbins [59, 60]).

For the problem of applying quality control on a process having mean shifts, we assume that there exists a controllable factor on the process which has a direct effect on the process mean. Then, a sequence of adjustments on this factor which take the form of equation (1.2) can be used to remove the mean shift eventually. As will be shown in

Chapter 3, this process adjustment method can be derived from the well-known Kalman Filter technique and several existing adjustment methods can be unified.

The topic of process adjustment is typically in the reign of the EPC literature. In contrast to SPC, EPC promotes active regulation actions on a manufacturing process, because it assumes that without active control the process will be either unstable or highly correlated. However, it is well-known that when a process is in the state of in-control, frequent adjustments will inflate the process variance and thus increase the process off-target quality cost. Ironically, it has been reported in some manufacturing case studies that switching down the automatic process controller is a way to improve overall product quality (Kramer [56]).

Although both EPC and SPC are aimed at enhancing the consistency of the process by reducing the variation of the quality of the products, they take different approaches based on their different assumptions on the process model. For the problem of quality control of a process which experiences random mean shifts, an integration method of control charts and sequential adjustment schemes will activate control actions only after the process has been statistically tested to be in the state of out-of-control. Therefore, this method is expected to have a better performance in the sense of minimizing the process off-target cost than applying an automatic control alone.

1.2.3 An example: a modified Deming's funnel

In this section, we demonstrate the significant effect of process adjustments on process quality control through a modified version of Deming's funnel experiment (this experiment is also discussed in Del Castillo [31]). The original version of the funnel

Fig. 1.1. Funnel experiment setup (Adapted from Davis [28])

experiment was described by Deming [34] and has become a classic example for Deming's principles of quality improvement. The experiment was conducted by mounting a funnel over a target bull's eye placed on a flat surface. Marbles were consecutively dropped through the funnel, and their position with respect to the target was measured (see Figure 1.1). The goal of this experiment is to minimize the deviations of the marble's resting positions to the target.

However, in this modified version of Deming's funnel experiment, we suppose the funnel is initially off-target and the direction and magnitude of this offset are unknown. We drop 50 marbles through the funnel. Three adjustment strategies are considered in this computer experiment: i) no adjustments, ii) at most one adjustment, and iii) a few sequential adjustments. If the funnel is not adjusted, it is obvious that the dropped marbles will congregate in a small region around the initially biased position away from the true target as shown in Figure 1.2a. Figure 1.2b shows the result after one full adjustment on the funnel is made according to the first observation, i.e., the funnel was relocated onto the opposite position of the first dropped marble (such an adjustment

Fig. 1.2. Off-target Deming's funnel experiment. a. no adjustment; b. one adjustment; c. five sequential adjustments

scheme was advocated by Taguchi [97] and criticized by Luceño [68]). In Figure 1.2c, five sequential adjustments on the funnel position are conducted, one for each of the first five marbles (i.e., the funnel is not adjusted for the remaining 45 marbles). Let $\mathbf{p}_t$ be the position of the tth marble, measured relative to the target. The five adjustments used in the simulation follow the harmonic sequence $\{-\mathbf{p}_1, -\mathbf{p}_2/2, -\mathbf{p}_3/3, -\mathbf{p}_4/4, -\mathbf{p}_5/5\}$. This is an adjustment strategy proposed by Grubbs [44] which will be studied in more detail in Chapter 3. By comparing Figures 1.2b and c, it is clear that one adjustment is inadequate while five sequential adjustments are able to bring the funnel very close to its real target.

In practice, SPC control charts can be used to detect the shift in process mean and to estimate the initial value of the shift size, and sequential adjustments can then be

conducted to reset the off-target process back on-target. It is evident that the in-control and out-of-control ARL performance of the control chart will affect the overall performance of the integrated scheme. The performance of this type of integrated schemes need to be studied under a variety of shift sizes in order to apply it on general cases. In Chapter 4, the performance of several combinations of control charts and sequential adjustment methods are evaluated by simulation.

1.2.4　Cost considerations

In previous sections, we related the quality cost to the squared deviation of process measure from its target value because customer satisfaction usually deteriorates very rapidly when the quality characteristic of a product is away from its specification. This relation was originally advocated by Taguchi [98]; thus, it is often called the "Taguchi's quality loss function".

For a Taguchi-type cost function, the quality cost associated with over-adjusting a process and the cost associated with under-adjusting are assumed to be symmetric around the process target. This assumption is frequently violated in some real manufacturing settings. For instance, consider a drilling process on a pre-machined part. An under-size hole may be repaired by another refining machining process, but an over-size hole is cannot be recovered and may result in scrapping the whole part. Therefore, there are different cost implications associated with different types of process deviation from its target. It is necessary to modify the sequential adjustment schemes for a process with an asymmetric quality cost function.

Other costs besides the process quality cost need to be included at evaluating the benefit of any adjustment scheme. When either process measurement cost or process adjustment cost is significant in nature, frequent adjustments may not be desirable. Under such cost structures, problems such as determining the optimal number of adjustments or whether or not to skip an adjustment are worth of investigation. Some modified sequential adjustment schemes according to a variety of cost functions will be addressed in Chapters 5 and 6 of this thesis.

1.2.5   Closed-loop processes

In many industrial instances, manufacturing processes are operated under the adjustments of a pre-defined controller, but the controller works in an *ad hoc* manner, i.e., it is designed to stabilize the process but it is not an optimal control rule in any sense. In this closed-loop system, the process output is usually autocorrelated. Properties of the closed-loop system need to be examined in order to improve the performance of any on-line process controller.

However, it is well-known that the closed-loop identification is very difficult because of the collinearity between the input and output data. Most of the existing techniques for process identification require the process to be operated in an open-loop manner when the process data are collected, but this could be too expensive, particularly when the process is not stable on its own. Therefore, identifying the stochastic model of a process operating in closed-loop and then optimizing the controller's performance is a practical issue for process quality control.

1.3 Research Objectives and Contributions

The overall goal of this research is to develop advanced statistical process adjustment methods for application in short-run manufacturing processes. More specifically, the following research objectives will be addressed:

1. **Sequential process adjustment methods using Stochastic Approximation**

   A unifying view of several process adjustment methods found in the literature will be established based on the Kalman Filter technique. The performance of these methods will be compared according to their small-sample properties (see Chapter 3).

2. **Integration of sequential adjustments and SPC techniques for process monitoring**

   The success of an integrated quality control scheme relies on properly applying techniques from both EPC and SPC fields. The proposed sequential adjustment methods will be integrated with control charts so as to provide both detection and correction of process abnormalities. Properties of this integrated scheme will be studied (see Chapter 4).

3. **A sequential control strategy for an asymmetric off-target cost function**

   Asymmetric cost structures are frequently met in various manufacturing processes. This implies that an adjustment rule based on a symmetric cost function around

the target may be inappropriate. A modified adjustment rule for asymmetric off-target cost functions will be proposed in this thesis and its performance will be compared with the conventional rules (see Chapter 5).

4. **Modifying adjustment rules when measurement and adjustment costs are considered**

When the costs of making adjustments or taking measurements is significant, adjusting at every time instance may not be the most economical strategy to achieve process correction. Some cost-saving opportunities, such as skipping diminutive adjustments and determining optimal stopping rules, will be investigated in this thesis (see Chapter 6).

5. **Closed-loop process identification and fine-tuning**

When a manufacturing process is already being regulated by a controller with non-optimal parameter settings, it is beneficial to identify the process model on-line and then to tune the controller parameters so as to drive the process to reach a better controlled state. Methods for the identification and fine tuning of a closed-loop process will be discussed in this research (see Chapter 7).

This research focuses on studying sequential adjustment methods based on Stochastic Approximation (SA) techniques for the purpose of process quality control. In the Statistics literature, SA is studied mostly from the estimation perspective; but its importance in process control should not be ignored. An important difference between the estimation problem and the control problem is that in the former the asymptotic behavior of the SA procedure is of interest, whereas in the latter the transient effect is

of interest. A short-run manufacturing process as discussed in this proposal will require analytical or simulation results on the small-sample properties of SA procedure. These have not been investigated in the Statistics literature with exception of Ruppert [84] and Frees and Ruppert [40]. Furthermore, when the cost functions of some manufacturing process settings are deviated from common assumptions, modified sequential adjustment schemes are proposed and their performance are compared with that of the conventional scheme.

This research also considers the integration of SPC and sequential adjustment methods. A considerable amount of work on the integration of SPC and Engineering Process Control (EPC) techniques has taken place during the recent years, but no previous work has combined SPC with sequential process adjustments. Therefore, this research provides a new approach to process quality control and this approach is established on Stochastic Approximation techniques.

In summary, a collection of mathematical models for the adjustment of short-run manufacturing processes is proposed and studied systematically in this thesis. It is demonstrated that by implementing proper adjustment strategies the stability of the process can be better maintained; thus, significant economic benefits obtained from the consistent quality of products will be achieved. This research contributes directly to the quality improvement program of the manufacturing industry and to the field of applied statistics.

1.4 Overview of Thesis

In the next chapter, a literature review of process monitoring, process adjustment methods and integration of SPC and EPC is given. Then, the main body of this thesis is organized as follows.

In Chapter 3, we consider the problem of adjusting a machine that starts production after a defective setup operation. A general solution based on a Kalman Filter estimator is presented and this solution calls for a sequential adjustment strategy.

Next, in Chapter 4, the integration of sequential adjustments with SPC control charts is investigated for controlling an abrupt step-type process disturbance on both i.i.d. processes and autocorrelated processes. The performance of this integration method is evaluated and compared to using either control charts or automatic adjustments alone.

In Chapter 5, we suppose that there are different cost levels associated with a higher-than-target process quality measure than with a lower-than-target quality measure. Under such an asymmetric cost structure the adjustment rules discussed in Chapter 3 need to be modified so as to avoid the quality characteristic falling into the higher cost side. Adding a bias term into the linear feedback control rule is proposed and the formulae for calculating this bias term is developed.

Furthermore, in Chapter 6, we discuss the influence of process measurement and adjustment costs on the decision of whether to apply adjustment actions at each point in time.

In Chapter 7, a method for identifying and fine tuning an EPC controller from closed-loop data is studied. We show that under certain general assumptions on the controller structure and process disturbance, it is possible to identify an ARMA (autoregressive moving average) models for the process output. After identification, one is able to to tune the controller to a near-optimal setting according to some performance criteria.

Finally, Chapter 8 summarize the main contribution of this thesis and discusses the further work.

Chapter 2

Literature Review

Given the vast literature on the topics related to this research, it is important to summarize them and present their connection with this research. In this chapter, we first review several types of SPC control charts for both independent and identical distributed (i.i.d.) processes and for autocorrelated processes. Then, some process adjustment techniques are described. We also survey the literature on integration of process adjustments and statistical process control. Finally, previous work on closed-loop system identification is discussed in the last section.

2.1 Control Charts for IID Processes and Autocorrelated Processes

2.1.1 Performance of control charts on an i.i.d. process

Two types of control charts – attribute control charts and variables control charts – are usually employed for monitoring a manufacturing process. In this study, we are mostly interested in step-type disturbances on process means, which implies that the process mean is subject to shifting to unknown levels from its target and stays at that level if no adjustments on the process are made. Thus, a control chart for variables, which is used to detect a change on a quality characteristic that is measured on a continuous scale, is suitable for our purposes. The most frequently used variables control charts for monitoring process means include the Shewhart chart, the EWMA chart and the CUSUM

chart. We briefly describe them in this section. More illustrations and applications of them can be found in many quality control textbooks, for example, see Montgomery [73].

Shewhart control charts have been widely applied on industrial practice due to its simplicity and easy of interpretation. On a Shewhart chart, sub-group (sample) means or individual observations of a process quality characteristic, y_t, are plotted in a time sequence. Suppose these data are independently normally distributed and their standard deviation, σ, is known. Then, if the process mean is right on its target, T, there is a 99.73% probability of observing a plotted point that is inside of the $T \pm 3\sigma$ control limits. Any point that is out of the control limits indicates a strong evidence that a shift in the process mean has occurred. However, since a Shewhart chart makes inferences on the process mean based on one observed data, it is only efficient for detecting large shifts on the process mean and it is insensitive to small or moderate mean shifts. In order to detect small shifts quickly, EWMA and CUSUM charts are recommended instead.

EWMA charts use the EWMA smoothing method to predict the process mean. This utilizes not only the current measurement but the discounted historical data as well. The EWMA statistic is defined as

$$z_t = \lambda y_t + (1 - \lambda)z_{t-1}, \quad 0 < \lambda < 1, \tag{2.1}$$

where λ is the smoothing parameter or weight, which accounts for how much of the past data should be discounted out at computing the current EWMA statistic. The EWMA chart control limits are $\pm L \, \sigma \sqrt{\frac{\lambda}{(2-\lambda)}[1 - (1 - \lambda)^{2t}]}$, which varies with time and converges to $\pm L\sigma \sqrt{\frac{\lambda}{(2-\lambda)}}$ in the long run (Crowder [25]).

CUSUM charts also use previous observations, but unlike EWMA charts, CUSUM charts use undiscounted information and are executed through two steps. First, the process is judged whether it is deemed to be in-control. If it is in-control, the CUSUM statistic would be set to its initial value (it is usually 0); otherwise, the CUSUM statistic would be accumulated by adding the current observed data from the process and the next judgement on whether the process is out-of-control could be made. If the process data are independently normally distributed, the CUSUM scheme can be shown to be the generalized likelihood ratio test for the hypothesis $H_0 : \mu = 0$ versus $H_1 : \mu = \mu_0$ where μ_0 is a predetermined out-of-control process mean (Lorden [63]). To establish a two-sided tabular CUSUM chart, two CUSUM statistics are needed:

$$c_t^+ = max\{0, y_t - K + c_{t-1}^+\}$$

and

$$c_t^- = max\{0, -y_t - K + c_{t-1}^-\} \tag{2.2}$$

where $K = \frac{s}{2}\sigma$ and s is the shift size that one wishes to detect (Woodall and Adams [113]). The control limit of the CUSUM statistics is defined as $H = h\sigma$. Whenever c^+ or c^- exceeds H, an out-of-control alarm is signaled. Usually, CUSUM charts are designed to detect small shifts in the process mean. For monitoring a process with a broad range of shift sizes, a combined CUSUM-Shewhart chart is recommended (Lucas [66]).

The performance of a control chart is characterized by its run-length distribution. As we can see from the above formulations, for each chart several parameters need to be determined and they greatly affect on the run-length properties of the charts. The

Shift in mean	Shewhart chart for individuals	EWMA chart for individuals	CUSUM chart for individuals
0	370	500	465
0.25	281	150	139
0.50	155	41.8	38.0
0.75	81.2	18.2	17.0
1.00	43.9	10.5	10.4
1.50	15.0	5.5	5.75
2.00	6.30	3.7	4.01
2.50	3.24	2.9	3.11
3.00	2.00	2.4	2.57
4.00	1.19	1.9	2.01

Table 2.1. ARL values for a Shewhart chart (control limit $L = 3\sigma$), EWMA chart ($\lambda = 0.2$, $L = 2.962\sigma$), CUSUM chart ($k = 0.5$, $h = 5$) and CUSUM-Shewhart chart (Shewhart limits at 3.5σ) (Adapted from Montgomery [73] and Lucas and Saccucci [67])

average run lengths (ARLs) of some commonly used Shewhart chart, EWMA chart and CUSUM chart are given in Table 2.1.

2.1.2 Change-point detection in an autocorrelated process

Previous discussions on control charts are applicable to an i.i.d. process. It is well-known that the presence of autocorrelation on process observations will deteriorate the performance of control charts substantially. The detrimental effects of autocorrelation on the performance of control charts have been actively researched in recent years (see, e.g., Alwan [4]). The problem associated with autocorrelated data mainly lies on two facets: 1) autocorrelations on data can induce a large amount of false alarms on a control chart, which, in turn, damages the usability of the control chart; 2) it is hard to distinguish a process mean shift from positive autocorrelations and the signal of the shift might be delayed.

Fig. 2.1. Plot of 50 simulated data from an i.i.d. process and from an AR(1) process, $y_t = 0.5y_t + \varepsilon_t$, on a Shewhart individual chart. No alarm is generated from the i.i.d. process, while two alarms are generated from the autocorrelated process.

Stamboulis [93] showed that for an AR(1) model (AutoRegressive model of order one), $y_t = \phi y_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim^{iid} N(0, \sigma^2)$, the ratio of $Var(\bar{y})^{iid}/Var(\bar{y})^{AR(1)}$ is proportional to $(1 - \phi)/(1 + \phi)$ when the sample group is large. Therefore, applying $\pm 3\sigma$ control limits on an autocorrelated process will increase the probability of false alarms (see Figure 2.1). Vasilopoulos and Stamboulis [106] investigated the effect of autocorrelation on Shewhart charts for an AR(2) process. Alwan [3] provided a measure of the detrimental effect of autocorrelation given by a general ARMA(p,q) process on Shewhart individual charts. He demonstrated that the moving-range based control limits and run rules might suggest numerous searches for nonexist special causes of out-of-control and distract from understanding the true process.

2.1.2.1   Control charts on process residuals

Assume the process observations $\{y_t\}$ are generated from a general stable ARMA (AutoRegressive Moving Average) time series model, $A(\mathcal{B})y_t = B(\mathcal{B})\varepsilon_t$, where $A(\mathcal{B})$ and $B(\mathcal{B})$ are some polynomial functions of the backshift operator $\mathcal{B}$ defined by $\mathcal{B}y_t = y_{t-1}$. If the process is invertible, an i.i.d. process of the residuals, i.e., $\varepsilon_t = A(\mathcal{B})y_t/B(\mathcal{B})$, can be obtained. Straightforwardly, a conventional control chart can be applied on such residuals.

Filtering out the autocorrelation on the original data by an ARMA model and plotting the process residuals on a Shewhart chart was first proposed by Alwan and Roberts [4]. They called this chart the Special Cause Chart because it is used to detect any out-of-control signal triggered by special causes that needs to be removed. They suggested to use this chart along with a time series plot of the original data.

Montgomery and Mastrangelo [74] proposed to use an IMA(1,1)(Integral Moving Average) process to model a general nonstationary process so that the EWMA statistic of the process observations is the next step prediction of the process mean. This prediction and its 95% confidence intervals serve as the target and control limits for the next process observation and this chart is called a Moving Center Line EWMA Control Chart. However, MacGregor [70] and Ryan [85] pointed out that this chart is equivalent to a residual Shewhart chart after imposing an IMA(1,1) model as the underlining model of the observations; therefore, it has good performance only when the true model is close to IMA(1,1).

The run-length distributions of residual Shewhart charts were investigated in detail by Wardell, Moskowitz and Plante [110, 109]. They mentioned that when a shift in the mean occurrs, the probability of observing a signal at the first sample is much higher than the probability of the following samples.

The performance of EWMA charts on residuals and CUSUM charts on residuals for ARMA(1,1) processes was reported by Lu and Reynolds [64, 65]. They also compared the ARLs of these charts with those of EWMA and CUSUM charts on the original (i.e., not filtered) observations. They found that an EWMA chart of the residuals is able to detect shifts in the process mean more quickly than EWMA charts on the original observations, but when the shift size is small or the autocorrelation is strong, an EWMA chart on the original observations seems to have better run-length performance. The performance of EWMA charts on residuals and EWMA charts on observations is very sensitive to the parameter estimates of the ARMA model. They recommended that more than 1,000 observations are needed for model estimation. For CUSUM charts, similar conclusions were drawn.

English *et al.* [37] did a comparison of the performance of Shewhart charts and EWMA charts on the residuals of a general autoregressive process, AR(p). It was found that the EWMA chart was generally preferred.

It has long been observed that residual charts do not have the same capability of detecting shifts in the process mean as the traditional control charts applied to an i.i.d. process (Harris and Ross [49], Adams, Woodall and Superville [2]). The purpose of using an ARMA filter to transform the observations to residuals is to obtain i.i.d. data, on which traditional SPC can apply. However, the jump-type process mean shift

will no longer exist in the residuals; instead, the shift will change to some dynamic pattern depending on the filtering model. Hu and Roan [50] demonstrated such dynamic patterns for AR(1), ARMA(1,1) and ARMA(2,1) models using z-transform formulae. For example, when a "step shift" in the process mean is filtered through an AR(1) model with high positive autocorrelation, the shift pattern changes to a high peak at the first time and then settling down on a small steady value; if the autocorrelation is negative, the shift pattern gradually reaches a large steady value.

Apley and Shi [6] called these patterns "process fault signatures". Therefore, to detect a shift in the process mean we need to detect any potential fault signature from the residual data. This can be done by a generalized linear ratio test, which can also be seen as a generalized CUSUM test. Correlations between the process residuals and the predefined fault signatures are used to reveal when a mean shift has occurred. The robustness of this method for misidentified process models was also discussed in their paper. Similar ideas also appeared in Atienza, Tang and Ang [7], and Wright, Booth and Hu [114].

2.1.2.2 Conventional control charts with modified control limits

Modifying the control chart limits is another approach to reduce the detrimental effect of autocorrelation on chart performance when a control chart is applied on the original autocorrelated observations. Traditional control charts tend to generate frequent false alarms especially for positive autocorrelation. Therefore, it is desirable to relax their control limits to avoid so many alarms. This approach has the advantages of avoiding using ARMA filters for data transformation; thus the process mean shift pattern would

not be altered and there is no risk of misidentification of ARMA models as the previous approach has.

Stamboulis [93] and Vasilopoulos and Stamboulis [106] discussed the design of a Shewhart $\bar{x}$ chart for AR(1) and AR(2) processes. The key issue is to find a suitable control limit instead of 3σ. The property of CUSUM charts on an AR(1) process was analyzed by Timmer, Pignatiello and Longnecker [99]. They modified the CUSUM statistic and its control limits according to the sequential probability ratio test on the joint distribution of autocorrelated data.

Yashchin [116, 118] has also discussed the design of a CUSUM chart for autocorrelated data. He showed that using transformed data is better than using the original data, but the advantage is too small to justify using it over original data, especially when the autocorrelation is not very strong. To obtain the desired run-length property, he suggested to replace the autocorrelated data by some i.i.d. series for which the charts has similar run-length characteristics. That is, i.i.d. data $\{y_t^*\}$ are used to replace autocorrelated $\{y_t\}$ and $\sum_{j=1}^{r} y_j^*$ and $\sum_{j=1}^{r} y_j$ have the same distribution function (or the same first and second moments), where r is the expected run length before the first false alarm when the CUSUM statistics of autocorrelation data are plotted. Some analytical results of in-control ARLs with different CUSUM design parameters (h and k) for ARMA models were provided in his paper [118].

Zhang [120] proposed an EWMAST chart, which is an EWMA chart with wider control limits applied on a stationary time series. His simulation results showed that for the AR(1) model with mild positive autocorrelation ($\rho < 0.5$), the EWMAST chart outperformed other charts.

The EWMA statistic studied in Zhang was generalized by Jiang, Tsui and Woodall [52]. They proposed a new test statistic, ARMA statistic. The ARMA(1,1) statistic was tested and compared with EWMAST, but it did not give a better performance on detecting process mean shifts in general.

Two general approaches to detect a mean shift in an autocorrelated process have been surveyed so far. We notice that control charts on process residuals, although they are mostly discussed in literature, are not generally better than control charts on original data. Therefore, the second approach will be followed in this thesis. More specifically, the CUSUM chart designed by Yashchin [118] will be applied on autocorrelated observations in Chapter 4 when SPC/EPC integration schemes for non-i.i.d. processes are investigated.

2.2 Process Adjustment Techniques and Stochastic Approximation

It can be said that, in general, there are two types of industrial controllers: one is a feedforward controller and the other one is a feedback controller. Feedforward control adjusts some controllable process variables before the process starts based on the measurement of process inputs. For example, raw material properties may vary from batch to batch, but by adjusting some process variables accordingly it might be possible to obtain consistent process outputs. Feedback control, in contrast, adjusts controllable process variables directly according to process output measures. The most widely used feedback controllers include discrete integral controllers, also called the EWMA controller

in the semiconductor manufacturing literature, and proportional integral (PI) controllers. Feedback-type control is the focus of this research.

In industrial practice, some simple yet powerful feedback control schemes are becoming a relevant part of the toolkit of a quality engineer. This is especially true in the semiconductor industry. For example, the Run-to-Run (R2R) controller, with its embedded EWMA or double EWMA control algorithm, has been successfully implemented in the semiconductor industry for batch-to-batch product quality control (Sachs *et al.* [86], del Castillo and Hurwitz [33], Moyne *et. al* [76]).

The optimal feedback controller, especially the celebrated Kalman filter controller (Kalman [53]), has a long successful history in the field of automatic control (see, e.g., Åström and Wittenmark [95], Ljung and Söderström [62], Lewis [61]). But since automatic control mainly focuses on adjustments at the equipment level where there is usually a very large number of measurements, the short-run performance and the discontinuous adjustments often required for adjustment of the quality of the products are largely ignored.

In the statistical literature, there are many stochastic processes that can be studied from a control point of view. The Robbins-Monro process (Stochastic Approximation) as illustrated in Chapter 1 is one of the most famous one.

Let x be the controllable variable on a process. If the process output y is a regression function on x, $y = M(x) + \varepsilon$, then a sequence $\{x_t\}$ that is recursively computed via (1.2) makes y_t converge to its target T. Robbins and Siegmund [80] provided an elegant proof of the consistency of x_t. Chung [23] and Fabian [39] established the asymptotic distribution of x_t by making some further assumptions on the sequence $\{a_t\}$. Loosely

speaking, when a_t takes a "quasi-linear" form such as $a_t = c/t$, then we have that, asymptotically, x_t is normally distributed, i.e.,

$$\sqrt{t}(x_t - \theta) \to^D N(0, \sigma^2 c^2/(2\beta c - 1)) \tag{2.3}$$

where θ is a value such that $M(\theta) = T$ and β is the slope of regression function at θ. By choosing $c = 1/\beta$, the asymptotic variance in (2.3) is minimized. Since β is usually unknown to process operators, it needs to be estimated. SA procedures with a consistent estimate of β are called adaptive.

Venter [107] proposed taking observations in pairs at $x_t \pm c_t$ for a suitable positive sequence $\{c_t\}$. Let y_t^1 and y_t^2 be the corresponding responses, and define

$$z_t = \frac{y_t^2 - y_t^1}{2c_t},$$

then β can be estimated by

$$b_t = \frac{\sum_{i=1}^{t} w_i z_i}{\sum_{i=1}^{t} w_i}$$

where $\{w_i\}$ are suitable positive weights.

Lai and Robbins [60] showed that b_t can be estimated by the usual least square estimation, i.e.,

$$b_t = \frac{\sum_{i=1}^{t}(x_i - \bar{x}_t)(y_i - \bar{y}_t)}{\sum_{i=1}^{t}(x_i - \bar{x}_t)^2},$$

then the optimal asymptotic distribution of $(x_t - \theta)$ still holds.

It is well known that the conventional SA procedure has a slow convergence rate. Kesten [55] derived a procedure which permits to bring the process closer to target more quickly than the conventional scheme. He considered that if the measured variable, y_t, is near zero in the neighborhood of the target, it would have its sign greatly obscured by the process noise. Therefore, the number of reversals of sign indicates whether the search of x_t is near or far from the root, θ. The sequence $\{a_t\}$ can be set as

$$a_t = \frac{1}{1 + \sum_{i=2}^{t} sign(y_i y_{i-1})}$$

where

$$sign(x) = \begin{cases} 1 & \text{if } x \leq 0 \\ 0 & \text{if } x > 0 \end{cases}.$$

Although in the literature SA procedures have been mentioned to be useful to solve control problems (see, e.g., Comer [24], Lai and Robbins [59], Ruppert [82, 83]), they have not been broadly perceived by quality engineers as a powerful tool for process quality improvement. A reason for this lack of interest is that the small-sample properties of SA have often been neglected.

2.2.1 Grubbs' harmonic rule

One classical application of using process adjustments for product quality control is the Grubbs' harmonic rule for adjusting a machine after a bad setup operation.

To give some context to this problem, suppose it is of interest to adjust a machine tool that manufactures discrete metal parts because of concerns about a machine offset

due to a bad setup. Measurements $\{y_t\}$ correspond to the deviations from target of a predefined quality characteristic of parts as they are produced at sequential points in time t. The machine adjustment at time t ($\nabla x_t = x_t - x_{t-1}$) has a direct impact on the quality level of the subsequent part. In some machining processes, an incorrect setup operation can result in drastic consequences in the quality of the parts produced thereafter. However, the observation from the process is subject to both the process inherent variation and the measurement error. The precision of the offset estimate can be improved if more measurements are available. Therefore, a good adjustment strategy will sequentially estimate the offset and suggest adjustments to the process accordingly.

Grubbs [44] proved that by using equation (1.2) and choosing $\{a_t\}$ to be a harmonic sequence, i.e.,

$$a_t = 1/t, \quad t = 1, 2, ... \tag{2.4}$$

the expected value of the process characteristic of every next sample will be the same as the target value. Furthermore the variance of the quality characteristic is minimal comparing to using other linear adjustment methods.

Recently, Trietsch [100] and del Castillo [29] made detailed analysis and extension of the Grubbs' rule for the problem of adjusting machine setup errors. It is evident that unifying these control schemes by using SA and evaluating their small-sample performance are two prerequisite steps before applying them on the quality control of a general manufacturing process. This topic will be presented in detail in Chapter 3.

2.3 Integration of SPC and EPC

Statistical Process Control and Engineering Process Control are two sets of toolk-its used by Quality Engineers and Process Engineers, respectively. Although they have the common goal – to maintain the process outcome as close to its target as possible – these two types of techniques take different approaches, which seems to contradict each other in some sense. SPC emphasizes process monitoring and fault detection. It opposes frequent process adjustment actions because when the process is in the statis-tically in-control state process tampering can only inflate the process variance instead of decreasing it. In contrast, EPC advocates various process adjustment schemes for guarding the process from drifting away from its target. It should be pointed out that the different approaches of SPC and EPC come from their fundamental assumptions on a manufacturing process and in practice they are not rivals of each other at all. Recently, the integration of these two techniques have been discussed by many researchers (see, for example, Box and Kramer [14], MacGregor [69], Vander Wiel, *et. al.* [105] and Tucker, *et. al* [104])

If a manufacturing process is initially in the statistically in-control state, as as-sumed by SPC, control charts are employed to detect any random abnormal change on the process. When this abnormal change is indeed found, process adjustment actions that can compensate for this change are necessary. A second way in which EPC and SPC scheme can be integrated is as follows. If a manufacturing process exhibits drifting behavior or autocorrelation in its output measures, as assumed by EPC, an automatic controller can be designed to regulate this process. However, it is still wise to have a

Fig. 2.2.   Two types of SPC/EPC integration. a. SPC chart acting as a deadband; b. SPC monitoring EPC output.

control chart to monitor the quality characteristic of the process and signal alarms due to other types of process changes that the controller cannot guard against. However, if an EPC scheme is continuously functioning, the process output is usually autocorrelated in time; therefore, a control chart designed for autocorrelated data should be used. These two types of SPC/EPC integration are illustrated in Figure 2.2.

In this thesis, we mainly discuss the first type of integration. The basic procedure of applying both monitoring and adjustment techniques for quality control is through three steps: 1) employ a control chart on-line to detect any possible process change, such as shifts in the process mean; 2) estimate the parameters of the process change; and 3) adjust the process based on a control scheme.

Taguchi [97] emphasized the importance of adjustments and recommended adjusting by the opposite deviation $(-y_t)$ whenever y_t exceeds the control limit of a Shewhart chart. This means that the process mean at the time of the out-of-control is estimated

by the last observed data point. This estimate always gives a large shift size, thus is significantly biased when the actual shift size is small. Adams and Woodall [1] also showed that the optimal control parameters and loss functions given by Taguchi are severely misleading in many situations.

An alternative feedback adjustment method recommended by Luceño [68] is to use an EWMA statistic of the past data collected from the process. It has been shown that if the disturbance is an integrated moving average (IMA(1,1)) process with parameter θ, the EWMA statistic is optimal in the sense of mean square error when its smoothing parameter λ is equal to $(1 - \theta)$. If the disturbance is not an IMA(1,1), this adjustment scheme still contains integral action and is quite robust (Box and Luceño [18], Del Castillo [31]).

Kelton *et al.* [54] suggested that continuously observing (without adjustment) several data after receiving the first "out-of-control" alarm will improve the process mean estimates. For instance, they suggest that the average of 10 deviations y_t after an alarm occurs is a good estimate of a shift size of 1.5σ. Delaying the mean estimation was also recommended by Ruhhal, Runger and Dumitrescu [81], although they dealt with a generalized IMA process model. Evidently, this method is only acceptable for a manufacturing process with high process capabilities and long production runs. For a short-run process or a process with tight specification, this approach may produce a high scrap rate.

Wiklund [111, 112] proposed a maximum likelihood estimation (MLE) method based on a truncated normal probability density function. His argument relies on the fact that the estimation of the process mean is made on the condition that one finds a

μ_s	Taguchi's method	Wiklund's MLE method	CUSUM ($h = 5$, $k = 0.5$)	EWMA ($\lambda = 0.25$, $L = 3$)
0	0 (3.30)	0 (1.38)	0 (2.39)	0 (1.22)
0.5σ	3.1 (1.28)	1.1 (1.11)	1.0 (0.67)	1.24 (0.14)
1σ	3.3 (0.71)	1.3 (1.10)	1.3 (0.54)	1.27 (0.14)
1.5σ	3.4 (0.54)	1.5 (1.22)	1.6 (0.55)	1.31 (0.14)
2σ	3.5 (0.50)	1.8 (1.33)	1.9 (0.67)	1.36 (0.17)
3σ	3.8 (0.60)	2.5 (1.50)	2.6 (0.77)	1.44 (0.26)
4σ	4.3 (0.78)	3.5 (1.60)	3.2 (0.82)	1.55 (0.32)

Table 2.2.   Shift size estimates (and their standard errors) obtained using different methods [adapted from Wiklund [111]]

point exceeding the control limit of Shewhart chart. He also discussed other estimation methods based on using CUSUM and EWMA control charts. Table 2.2 provides results of the estimated shift size by different methods from simulation study, where the standard errors of the estimates appear in parenthesis. One can see that Taguchi's method is very misleading on small to moderate shifts, that the EWMA is not a sensitive estimator of the shift sizes, and that the MLE and CUSUM perform comparatively better, but they are still inefficient when the shift size becomes large.

More accurate estimation methods of the process mean has appeared in recent research. Chen and Elsayed [22] provide a Bayesian estimation method for detecting the shift size and estimating the time of the shift. Crowder and Eshleman [27] applied a Bayesian method to the short-run process control. They assumed that the process mean is subject to small frequent changes that result in serial autocorrelation, so the hypothesis test of whether a major shift in the mean occurred is not relevant. Yashchin [117] proposed an adaptive EWMA estimator of the process mean, and he showed that his estimator is good at detecting abrupt mean changes. However, this method require

extensive computational effort for estimating the process mean at each step and is unsuitable for on-line process control.

Most SPC/EPC integration methods found in the literature use only a single step adjustment; therefore, the accuracy and the precision of the process mean estimate determine the effectiveness of this adjustment. In contrast, sequential adjustment methods have the substantial benefit of reducing the negative effect of an inaccurate initial estimation. Guo, Chen and Chen [45] presented an EWMA process mean estimator with the capability of dynamic parameter tuning. They used two EWMA charts to detect moderate and large process shifts, and then developed an EWMA controller with a sequentially adjusted parameter. In Chapter 3 we show that their sequential adjustment procedure is actually the same as Grubbs' harmonic rule.

Integration of sequential adjustment strategy and control charts will be fully studied in Chapter 4. We will apply this method on both i.i.d and autocorrelated processes.

2.4 Modified Adjustment Rules under Some Special Cost Considerations

Two cost issues often arise from real manufacturing settings: 1) asymmetric cost functions and 2) significant measurement and adjustment costs. The sequential adjustment scheme, such as Grubbs' harmonic rule, must be modified for these cases.

It is well-known that in many practical process, like, for instance, in hole-finishing or milling operations, asymmetric off-target quality cost functions are appropriate because the cost of oversized and undersized quality characteristics are different. The impact of asymmetric cost functions has been studied from several perspectives. Wu

and Tang [115] and Maghsoodloo and Li [71] considered tolerance design with asymmetric cost functions, while Moorhead and Wu [75] analyzed the effect of this type of cost function on parameter design. Ladany [58] presented a solution to the problem of setting the optimal target of a production process *prior* to starting the process under a constant asymmetric cost function. Harris [48] discussed the design of minimum-variance controllers with asymmetric cost functions for a process characterized by a linear dynamic model and ARIMA (AutoRegressive Integrated Moving Average) noise. Despite of the generality of this model, a possible process start-up error has not been included into consideration, probably because his emphasis was on chemical processes, not discrete part manufacturing.

We once again look at the machine setup error problem. Former research on the setup adjustment procedure only dealt with the case of symmetric cost functions. When the setup error exists under an asymmetric off-target quality cost function, it is intuitive to have the value of the quality characteristic converge to the optimal setting from the lower cost side. This is related to certain stochastic approximation techniques in which a bias term is added to allow for one-side convergence, as discussed by Anbar [5] and Krasulina [57]. However, their approaches are oriented to asymptotic or long-term performance and the conditions they impose on the control rule parameters are too complicated for practical engineering application. In Chapter 5, a biased sequential adjustment rule for asymmetric off-target quality cost will be derived and its small sample properties will be studied.

In Chapter 6, a modified sequential adjustment strategy will be illustrated when the process measurement and process adjustment costs are significant. A variety of process adjustment and process measurement cost models have been discussed in literature. Adams and Woodall [1] and Box and Kramer [14] investigated several types of adjustment and measurement cost models for the "machine tool problem", which was first proposed in Box and Jenkins [12]. Crowder [26] derived the control limits for adjustments to minimize the total cost for a finite horizon (short-run) manufacturing process. Single-step adjustment methods (i.e., adjusting only once when the process is deemed to be out-of-control) were used in these papers. An optimal adjustment strategy for the setup error problem was discussed by Sullo and Vandeven [96]. They considered a single adjustment method with a 0-1 type quality cost for conforming or nonconforming manufactured items.

For the machine setup error problem, a sequential process adjustment scheme has two purposes – to move the process close to its target quickly and to collect the process information for the next adjustment. Sometimes, especially when there exists a significant process disturbance, these two purposes conflict with each other. The harmonic adjustment rule is an elegant procedure that achieves the process adjustment goal along with the information collection goal. It is also interpreted as a special version of mean square error regression (Lai and Robbins [59]). This procedure can be re-examined from another viewpoint, that is, the maximum likelihood estimation of the unknown process parameters, as illustrated in what follows.

Fig. 2.3.   A different view of the machine setup problem

Suppose the model for this problem is

$$y_t = d + x_{t-1} + \varepsilon_t \qquad (2.5)$$

where d is an unknown start-up offset and where we assume $\varepsilon_t \sim N(0, \sigma^2)$. In this case, the regression function of y on x is $M(x) = d + x$. Without loss of generality, we let the target of y be equal to zero. Using the harmonic adjustment rule, the process adjustment scheme is

$$x_t = x_{t-1} - \frac{1}{t} y_t. \qquad (2.6)$$

Graphically, the model is a linear regression line with a slope of 45° and an intercept at d, as shown in Figure 2.3. Of course, the regression line is buried in random

noise. The first measurement y_1 is taken before any adjustment action, i.e., at $x_0 = 0$. After y_1 is obtained, x_1 can be calculated by $x_1 = x_0 - y_1 = -d - \varepsilon_1$. The variable x_1 will likely fall into a $\pm 3\sigma$ interval around θ, where θ is the root of the regression function $M(x)$ and it is equal to $-d$ in this problem. Let y_2 be the second measurement taken after setting the machine at x_1, i.e., $y_2 = d + x_1 + \varepsilon_2$. If the noise ε_2 is coincidentally equal to ε_1, then $y_2 = d + x_1 + \varepsilon_1 = 0$. Therefore, two measurements could be obtained at the setpoint x_1. An unbiased maximum likelihood estimate (MLE) of y is $\frac{1}{2}(y_2 + 0)$, which is identical to the second term of equation (2.6) when $t = 2$. Following this argument, it can be seen that any further adjustment quantity is the same as the MLE of y at that point in time. Therefore, the adjustment formula can be rewritten as

$$x_t = x_{t-1} - \hat{y}(x_{t-1}). \tag{2.7}$$

In this way, it is clear that x_t is likely to fall inside a $\pm \frac{3\sigma}{t}$ range around θ.

The above explanation implies that the intermediate adjustments between two arbitrary adjustment steps are not necessary for maintaining the optimal property of the last adjustment. Skipping some unnecessary process adjustments (but keep measuring the process) will not alter the optimality of the subsequent adjustment if the harmonic rule is slightly modified. This is specially important when the process measurement cost and adjustment cost need to be included into consideration.

Trietsch [101] showed that after some slight modifications the Grubbs' harmonic adjustment rule can be carried out in a series of arbitrary discrete points of time, i.e.,

the adjustments need not be conducted one following another successively, and the optimality of this rule would not change. He proposed an approximated scheduling method for discontinued adjustments. Chapter 6 presents an even simpler and more powerful scheduling method for the adjustments.

2.5 Closed-loop Systems

In previous sections, the importance of selecting an adjustment scheme to regulate an out-of-control manufacturing process was discussed. Now, suppose instead that an automatic feedback controller has already been installed on a process. We intend to model this closed-loop system and to seek any opportunity to optimize it.

Consider an open-loop process, where the relation between the process input x_t (controllable factor) and the process output y_t (process measure) can be characterized by a rational transfer function as

$$(1 - a_1\mathcal{B} - a_2\mathcal{B}^2 ... - a_r\mathcal{B}^r)y_t = (b_0 + b_1\mathcal{B} + ... + b_s\mathcal{B}^s)x_{t-b} \tag{2.8}$$

where $\mathcal{B}$ is a backshift operator and it is defined by $\mathcal{B}y_t = y_{t-1}$, and this transfer function is called of (r, s, b) order. Let $A_r(\mathcal{B}) = (1 - a_1\mathcal{B} - a_2\mathcal{B}^2 ... - a_r\mathcal{B}^r)$ and $B_s(\mathcal{B}) = (b_0 + b_1\mathcal{B} + ... + b_s\mathcal{B}^s)$, then the transfer function can be written as

$$y_t = \frac{A_r(\mathcal{B})}{B_s(\mathcal{B})}\mathcal{B}^b x_t. \tag{2.9}$$

However, it is a deterministic relation between y_t and x_t. By adding a process noise term, which is modelled by an ARIMA (AutoRegressive Integrated Moving Average) process, the stochastic open-loop process becomes

$$y_t = \frac{A_r(\mathcal{B})}{B_s(\mathcal{B})}\mathcal{B}^b x_t + \frac{C_q(\mathcal{B})}{D(\mathcal{B})}\varepsilon_t \tag{2.10}$$

where $\{\varepsilon_t\}$ is an i.i.d. noise sequence, $C_q(\mathcal{B}) = 1 - c_1\mathcal{B} - ... - c_q\mathcal{B}^q$ and $D(\mathcal{B}) = (1 - d_1\mathcal{B} - ... - d_p\mathcal{B}^p)(1 - \mathcal{B})^d$; therefore, the noise term is referred as an ARIMA(p,q,d) process and the whole open-loop process model is known as Box-Jenkins model with TF(r,s,b) plus ARIMA(p,q,d) (Box *et. al.* [13]).

However, many manufacturing processes are operated under the action of some feedback controller. One common type of the controller is the proportional-integral (PI) controller, which takes the form of

$$x_t = k_P y_t + k_I \sum_{j=1}^{t} y_j \tag{2.11}$$

where k_P and k_I are the weights for proportional portion and integral portion of the feedback control respectively. The aforementioned EWMA controller is a special case of PI controller, where only the integral control is in use, i.e., $k_P = 0$; therefore, it is also called the integral (I) controller. Frequently, these controller is activated on the process in a precautious manner, i.e., they are used to stabilize the process but are not optimal in improving the process performance. The conventional process identification methods used in open-loop processes (i.e., when the controllable factors are independent of the

quality characteristic measurements) makes use of the cross-correlation function between process input and output data (Box *et al.* [13]). For a closed-loop process, however, it is well-known that control actions and the estimated process disturbance are correlated, thus the conventional identification method becomes infeasible. Box and MacGregor [16, 17] proposed a method based on adding a "dither" signal to break the collinearity between the process input and output.

Identifying a closed-loop process by analyzing the autocorrelation of the process output was recently discussed by Del Castillo [30, 32]. He provided a family of disturbance models that consists of several types of models especially useful for discrete part manufacturing. He also showed that identifying a process transfer function in a closed-loop process is the same as identifying an ARMA model of the process output deviations, if the disturbance is a member of the assumed family. Chapter 7 will present a methodology of identifying closed-loop processes and fine-tuning the controller that extends and generalizes his results.

Chapter 3

An Unifying View of Some Process Adjustments Methods and Their Small-Sample Performance

In this chapter, some process adjustment rules for the machine setup error problem, which was previously introduced in Section 2.2.1, are investigated from a Bayesian point of view. The Bayesian formulation unifies some well-known process adjustment schemes, including Grubbs's harmonic and extended rules [44], adjustment methods based on stochastic approximation and recursive least squares, and a recent method on adaptive EWMA controllers due to Guo, Chen and Chen [45]. Small-sample performance comparisons between these methods and discrete integral feedback controllers (EWMA controllers) are provided. In case the offset is an unknown constant, performance measures for a single realization of the adjusted process are considered. If the offset is instead a random variable, performance over the ensemble of all possible realizations is studied.

3.1   A Kalman Filter Model for Process Adjustment

This section provides a Bayesian formulation to the machine setup adjustment problem based on a Kalman Filter estimator. As stated in Grubbs [44], the setup adjustment problem can be investigated in two cases – one is a single process where the setup error is an unknown constant and the other one is a group of similar processes (or process ensemble) where the setup error is better to be modelled by a probabilistic distribution. Consider the more general case when the setup error d is a random variable

with known mean $\bar{d}$ and known variance P_0. Then, for the first manufactured part the mean deviation from target of the quality characteristic is assumed equal to:

$$\mu_1 = d + x_0 \tag{3.1}$$

where x_0 is the initial setting of the machine setpoint (the controllable factor).

The first observed deviation from target will be given by

$$y_1 = \mu_1 + \varepsilon_1 = d + x_0 + \varepsilon_1 \tag{3.2}$$

where the ε_t's are assumed identically and independently distributed as a normal with mean 0 and variance σ_ε^2 and models both the part-to-part variability and the measurement error.

It will be assumed that adjustments affect the mean of the process. This is in sharp contrast with some EPC methods where the adjustment is supposed to modify the observed deviation, simply compensating for some inherent deviation that cannot be corrected (Del Castillo [29]). Thus, in this case, the first set point adjustment $\nabla x_1 = x_1 - x_0$ will result in a new process mean of

$$\mu_2 = \mu_1 + \nabla x_1$$

and the second deviation from target will be

$$y_2 = \mu_2 + \varepsilon_2.$$

Continuing in this form, the general expressions for the mean of the process and the deviations from target are, respectively,

$$\mu_t = \mu_{t-1} + \nabla x_{t-1} \tag{3.3}$$

and

$$y_t = d + x_{t-1} + \varepsilon_t \tag{3.4}$$

or, equivalently:

$$y_t = \mu_t + \varepsilon_t. \tag{3.5}$$

Suppose the objective is to find the process adjustments $\nabla x_1, \nabla x_2, ..., \nabla x_{n-1}$ that minimize

$$E\left[\sum_{t=1}^{n} \mu_t^2\right]. \tag{3.6}$$

In other words, it is assumed that there are quadratic off-target costs but no cost is incurred when performing the adjustments. Optimization of this type of criterion is based on the *separation principle* (Åström [94], Lewis [61]). For the setup adjustment problem under consideration, this principle indicates that the optimal solution can be found by solving separately the problem of estimating the μ_t's (process means) from the problem of finding the best adjustments $\{\nabla x_t\}$. If the optimal adjustment equation that is obtained through this separation is identical to what would have been obtained if the process were deterministic, the controller is said to be a certainty equivalence controller. This essentially means that the parameter estimates are used in the control equation as if they were the true parameters. In our case, if the means μ_t were known it is evident that

the best adjustment would be simply to set $\nabla x_t = -\mu_t$ which from equations (3.3-3.5) yields a minimum variance process which is on-target on average. In what follows, it is shown that Grubbs' extended rule is a simple case of a certainty equivalence controller in which we adjust by $\nabla x_t = -\widehat{\mu}_t$ where the mean μ_t is estimated separately using a Kalman Filter.

It has been shown that certainty equivalence holds for processes that obey linear difference equations (such as Equation (3.3)) when the criterion to optimize is a quadratic function of the state variable and the control factor (μ_t and ∇x_t in our case). This is the celebrated Linear Quadratic (LQ) control problem. The Appendix shows the solution to a more general LQ control problem applied to setup adjustment problems. We now show how Grubbs' extended rule is the optimal solution for criterion (3.6). Recently, Trietsch [100] pointed out the optimality of Grubbs' extended rule. We emphasize that this optimality property only holds if the first two moments of the setup error distribution are known.

The estimation problem is solved in a Bayesian framework using a simple Kalman Filter (Meinhold and Singpurwalla [72]). Given the model (3.3-3.5), define

$$y^t | \mu_t \sim N(\mu_t, \sigma_\varepsilon^2)$$

and define the prior distribution of μ_t:

$$\mu_t | y^{t-1} \sim N(\widehat{\mu_{t-1}} + \nabla x_{t-1}, P_{t-1})$$

where $y^t = \{y_1, y_2, ..., y_t\}$ are all available data at time t. The first mean has a prior distribution $\mu_1 \sim N(\overline{d} + x_0, P_0)$ where $\overline{d} = E[d]$, $\mathrm{Var}[d] = P_0$ and x_0 is the initial setpoint of the machine. Note that if $\overline{d}$ is known, then we should set $x_0 = -\overline{d}$ and get $\mu_1 \sim N(0, P_0)$. If suppose d is an unknown constant, then P_0 can be interpreted as our confidence (*a prior* variance) on the initial estimate of the setup error, $\hat{d}_0$, from Bayesian point of view.

The posterior mean of μ_t is $\widehat{\mu}_t = E[\mu_t | y^t]$, and the posterior variance of μ_t is defined as

$$P_t = \mathrm{Var}(\mu_t | y^t).$$

Given this setup, we have that

$$\mu_t | y^t \sim N(\widehat{\mu}_t, P_t)$$

with

$$\widehat{\mu}_t = \widehat{\mu_{t-1}} + \nabla x_{t-1} + \frac{P_{t-1}}{P_{t-1} + \sigma_\varepsilon^2}[y_t - (\widehat{\mu_{t-1}} + \nabla x_{t-1})]$$

and

$$P_t = \frac{P_{t-1}\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + P_{t-1}}$$

which is a recurrence equation easily solved by iteration yielding

$$P_t = \frac{\sigma_\varepsilon^2 P_0}{\sigma_\varepsilon^2 + tP_0}$$

where P_0 is known. The Kalman Filter estimate of the process mean given the data is

$$E[\mu_t|y^t] = \widehat{\mu_t} = \widehat{\mu_{t-1}} + \nabla x_{t-1} + K_t(y_t - (\widehat{\mu_{t-1}} + \nabla x_{t-1})) \tag{3.7}$$

where the quantities

$$K_t = \frac{P_{t-1}}{P_{t-1} + \sigma_\varepsilon^2} = \frac{1}{t + \frac{\sigma_\varepsilon^2}{P_0}} \tag{3.8}$$

are the "Kalman weights". Note that expression (3.7) is independent of the particular choice of the ∇x_t's (a consequence of the separation principle). Under the stated assumptions of normality, $\widehat{\mu_t}$ is the minimum mean square error (MMSE) estimator of μ_t. As shown by Duncan and Horn [35], if the normality assumptions are relaxed, $\widehat{\mu_t}$ is the MMSE *linear* estimator, i.e., among all estimators that are linear combinations of the observations it has smallest MSE, but there may be better nonlinear estimators.

To minimize $E[\sum_{t=1}^n \mu_t^2]$ one can argue as follows. Conditioning on all available data at time $t-1$ we have that

$$
\begin{aligned}
E[\mu_t^2|y^{t-1}] &= \mathrm{Var}(\mu_t|y^{t-1}) + [E(\mu_t|y^{t-1})]^2 \\
&= \frac{\sigma_\varepsilon^2 P_0}{\sigma_\varepsilon^2 + (t-1)P_0} + (\widehat{\mu_{t-1}} + \nabla x_{t-1})^2
\end{aligned}
$$

which is minimized by taking $\nabla x_{t-1} = -\widehat{\mu_{t-1}}$. From our earlier discussion, this is a certainty equivalence controller. Applying this adjustment rule at every point in time also minimizes the sum of the squared deviations (3.6). This can be formally proved by

showing how this problem is a particular instance of a LQ problem and utilizing the LQ solution (see Appendix A).

Substituting the control rule into the process mean estimate, we get

$$\widehat{\mu_t} = \widehat{\mu_{t-1}} + K_t y_t \qquad (3.9)$$

and the adjustment rule is

$$x_t = x_{t-1} - K_t y_t. \qquad (3.10)$$

We conclude that Grubbs' extended rule minimizes the expected sum of squared deviations provided the setup error mean and variance are known. If the errors are all normally distributed, Grubbs extended rule is the optimal solution for criterion (3.6). If the errors are not normal, Grubbs extended rule is the best *linear* control law that minimizes (3.6) (Åström [94], Lewis [61]). These additional facts can also be proved using LQ or LQG (linear control gaussian) theory.

The Kalman Filter equation (3.8) together with equation (3.10) provide four particular cases of interest:

1. **Grubbs harmonic rule and Robbins and Monro stochastic approximation.-**

Suppose one realization of the process is investigated. In this case, the process setup offset d is an unknown constant. Since it is lack of any *a priori* information on d before the process starts, it is reasonable to let $P_0 \to \infty$. Then, Grubbs'

harmonic rule is obtained, because under these conditions the Kalman weights

$$K_t = \frac{1}{t}.$$

(3.11)

The mean estimates become

$$\hat{\mu}_t = \frac{1}{t} y_t$$

and

$$\nabla x_t = -\frac{1}{t} y_t$$

or

$$x_t = x_{t-1} - \frac{1}{t} y_t$$

(3.12)

which is exactly Grubbs' harmonic rule.

Grubbs obtained (3.12) by solving a constrained optimization problem (i.e., min $Var(y_{t+1})$, subject to $E[y_{t+1}] = 0$) under the assumption the setup error d was an unknown constant (a machine offset).

Interestingly, equation (3.12) allows to see that Grubbs' harmonic rule is a special case of Robbins and Monro's [79] stochastic approximation algorithm for the sequential estimation of the offset d. Robbins and Monro's procedure was obtained by investigating the conditions under which the sequence K_t makes $\lim_{t\to\infty} E[d - \hat{d}_t]^2 = 0$ (i.e., the mean square convergence of the estimates $\hat{d}_t$ to d). Later, Blum [11] established the stronger result of convergence with probability one for the Robbins-Monro scheme. Since then, a wealth of asymptotic results have been

reported with respect to Robbins and Monro's procedure and its modifications. The setup adjustment application requires a study of the finite sample properties of these methods, which will be presented in the next section.

2. **Grubbs extended rule.-**

An extended adjustment rule was also proposed by Grubbs. In this case, the setup offset d is modelled as a random variable with the mean of value 0 and known variance σ_d^2. According to Grubbs, the "variance of the setup" (σ_d^2) is due to changes in the machine over many "occasions", where an occasion may be understood as one setup operation. Not surprisingly, Grubbs' extended rule is exactly same as what was derived from Kalman Filter technique, i.e.,

$$K_t = \frac{1}{t + \sigma_\varepsilon^2 / \sigma_d^2}. \tag{3.13}$$

Although there exists many optimal properties of this adjustment rule, in order to apply this rule, the second moments of both process error and setup offset must be known, which is quite unrealistic in practice. In the next section, it is shown that when these second moments are unknown the extended rule does *not* perform better than the harmonic rule in general.

3. **Recursive least squares.-**

If $\sigma_\varepsilon^2 = 1$ is fixed in (3.8), the so-called recursive least squares (RLS, Young [119], Ljung and Söderström [62]) estimate of d is obtained:

$$\hat{d}_t = \hat{d}_{t-1} + \frac{1}{1/P_0 + t} y_t.$$

It is of interest to study the performance of such an adjustment scheme given that recursive least square estimates are reputed to converge faster than those obtained using stochastic approximation methods in adaptive control applications.

4. **Unreliable measurements.-**

If $\sigma_\varepsilon^2 \to \infty$, i.e., if the measurements are completely unreliable, this implies that $K_t \to 0$ and $\hat{d}_t = \hat{d}_{t-1} = \cdots = \hat{d}_0$, with $\hat{d}_0$ being the prior estimate of the offset. In such case, it is optimal to let $x_t = -\hat{d}_0$ for all periods, implying no adjustments are made. Evidently, if we had $\hat{d}_0 = d$ then setting $x_t = -\hat{d}_0$ would be optimal. Unfortunately, if the prior estimate $\hat{d}_0$ is far from d, and σ_ε^2 tends to infinity, the prior estimate cannot be improved based on new measurements. If σ_ε^2 is large but finite and $\hat{d}_0$ far from d, the adjustment method will eventually bring the process to target, but the convergence will be very slow.

Although it is not a special case of (3.8), it is also of interest to consider the case when $K_t = \lambda$, in which case a discrete integral controller (Box and Luceño, [18]), also called an EWMA controller (Sachs *et. al.* [86], Del Castillo and Hurwitz [33]) is obtained:

$$x_t = x_{t-1} - \lambda y_t,$$

where the integration constant is the EWMA weight λ. This controller has the main advantage of compensating against sudden shifts that can occur at any point in time besides of the initial offset d. This means that the controller remains "alert" to compensate for shifts or other disturbances. A disadvantage is that it is not clear what value of λ to use. Because of this, some attempts have been made at developing adaptive techniques that modify λ as the control session evolves (see, for example, Patel and Jenkins [78]). In particular, Guo, Chen and Chen [45], apparently unaware of the paper by Grubbs, proposed to apply a "time varying" EWMA controller such that it minimizes the mean square deviation of the quality characteristic after a sudden shift occurs. Not surprisingly, the optimal weights obey, once again, Grubbs' harmonic rule:

$$\lambda_t^* = \frac{1}{t - \tau + 1}$$

where τ is the point in time when a shift in the mean occurs. Guo *et. al.* proposed to determine the change point τ using an SPC control chart. To achieve the permanent protection and robustness against a variety of disturbances that a discrete integral controller provides, together with the advantage of using temporarily larger weights after a shift occurs, Guo and co-workers actually suggested to use

$$\lambda_t = \max(\lambda_0, \lambda_t^*)$$

where λ_0 is typically a small value, say 0.1. The large and small values of λ correspond, respectively, to the "rapid control mode" and "gradual control mode" which have been applied on semiconductor manufacturing processes (Moyne *et. al.* [76]).

3.2 Small-sample Performance of Process Adjustment Rules

3.2.1 Performance indices for small samples

The performance indices that will be used in the remainder of this chapter are presented in this section for two cases, which were also considered by Grubbs.

Consider first the case where the setup error d is an unknown constant or "offset". For this case, the performance index considered is the scaled Average Integrated Square Deviation (AISD) incurred over m time instants or parts. This is defined for integer $m > 0$ as:

$$\text{AISD}(m) = \frac{1}{m\sigma_\varepsilon^2} \sum_{t=1}^{m} E[y_t^2] = \frac{1}{m\sigma_\varepsilon^2} \sum_{t=1}^{m} \left(Var(y_t) + E[y_t]^2 \right). \qquad (3.14)$$

The AISD is a common performance index in the control engineering literature. Since y_t models deviations from target, the AISD index is like an average "variance plus squared bias" calculation, and is a surrogate of a quadratic off-target "quality loss" function. We avoid dependency on σ_ε^2 by dividing by this quantity.

Consider now the case where the setup error d is a random variable. The performance measure to be used when d is random is once again the AISD but we need to

account for the additional variability in the setup error, so we define

$$\text{AISD}_d(m) = \frac{1}{m\sigma_\varepsilon^2} E_d \left[\sum_{t=1}^m E[y_t^2] \right] = \frac{1}{m\sigma_\varepsilon^2} \int_{-\infty}^\infty \sum_{t=1}^m E[y_t^2] f_d(x) dx, \qquad (3.15)$$

where the outer expectation is taken over the distribution of d. The case when d is normal with known mean and known variance was discussed by Trietsch [100]. Under such conditions, Grubbs' extended rule is optimal for the AISD_d criterion.

3.2.2 Performance for an unknown constant setup offset

Suppose d is an unknown constant, but unaware of this fact a user applies Grubbs' extended rule (i.e., the Kalman Filter adjustment scheme given by (3.10)) to the process. By Bayesian interpretation, P_0 is the user's confidence of the initial estimate of process offset d. Then, this rule applied to the process $y_t = d + x_{t-1} + \varepsilon_t$ results in

$$\frac{E[y_t]}{\sigma_\varepsilon} = \frac{A}{B_1(t-1)+1} \qquad (3.16)$$

and

$$\frac{Var(y_t)}{\sigma_\varepsilon^2} = 1 + \frac{t-1}{(1/B_1 + t - 1)^2} \qquad (3.17)$$

where $A = (d - \hat{d}_0)/\sigma_\varepsilon$ measures how far off the initial estimate of the offset was. The quantity $B_1 = P_0/\sigma_\varepsilon^2$ is a measure of the "confidence" on the initial offset estimate.

Taking these formulae (3.16) and (3.17) into the AISD equation, after some algebra, equation (3.14) can be written as:

$$\text{AISD}(m) = C_1 A^2 + C_2, \tag{3.18}$$

and C_1 and C_2 are defined as:

$$C_1 = \frac{\Psi'(1/B_1) - \Psi'(m + 1/B_1)}{B_1^2 m}$$

and

$$C_2 = \frac{B_1[\Psi(m + \frac{1}{B_1}) - \Psi(\frac{1}{B_1})] + \Psi'(m + \frac{1}{B_1}) - \Psi'(\frac{1}{B_1})}{B_1 m} + 1$$

where $\Psi(x) = d \ln \Gamma(x)/dx$ (Psi or digamma function), and $\Psi'(x) = d\Psi(x)/dx$ (trigamma function).

Taking limit as $B_1 \to \infty$, we get a simple formula for the AISD given by Grubbs harmonic rule:

$$\text{AISD}_G(m) = 1 + \frac{\Psi(m) + \gamma + A^2}{m} \equiv A^2/m + C_3 \tag{3.19}$$

where $C_3 = 1 + (\Psi(m) + \gamma)/m$ and $\gamma \approx 0.5772156$ is Euler's constant.

These analytical results can be easier to use if a software that computes the polygamma function is available (e.g., Mathematica or Maple).

For a discrete integral controller (or EWMA controller), it can be shown that

$$\frac{E[y_t]}{\sigma_\varepsilon} = (1 - \lambda)^{t-1} A \tag{3.20}$$

and

$$\frac{Var(y_t)}{\sigma_\varepsilon^2} = \frac{2 - \lambda(1 - \lambda)^{2(t-1)}}{2 - \lambda}. \tag{3.21}$$

From them, the closed-form expression of AISD(m) can be computed as:

$$\text{AISD}_{\text{EWMA}}(m) = \frac{2}{2 - \lambda} + \left(\frac{1 - (1 - \lambda)^{2m}}{m(2 - \lambda)}\right)\left(\frac{A^2}{\lambda} - \frac{1}{2 - \lambda}\right). \tag{3.22}$$

AISD expressions allow to study the trade-offs between the sum of the variances and the sum of squared expected deviations (squared bias). For the Kalman Filter scheme, as $B_1 = P_0/\sigma_\varepsilon^2 \to 0$, implying increasingly higher confidence in the *a priori* offset estimate, then $m^{-1}\sum_{t=1}^m Var(y_t)/\sigma_\varepsilon^2 \to 1$ (i.e., we get lower variance), but $m^{-1}\sum_{t=1}^m E[y_t]^2/\sigma_\varepsilon^2 \to A^2$ (i.e., we get larger bias). Similarly, for the EWMA controller, as $\lambda \to 0$, implying less weight given to the last observation, we have that $m^{-1}\sum_{t=1}^m Var(y_t)/\sigma_\varepsilon^2 \to 1$ (lower variance), but $m^{-1}\sum_{t=1}^m E[y_t]^2/\sigma_\varepsilon^2 \to A^2$ (larger bias).

The performance of the following adjustment rules has been evaluated based on the AISD criterion:

1. Grubbs harmonic rule, where $\hat{d}_t = \hat{d}_{t-1} + y_t/t$;

2. Kalman Filter rule 1 (assumes σ_ε^2 is known), where $\hat{d}_t = \hat{d}_{t-1} + \frac{y_t}{\sigma_\varepsilon^2/P_0 + t}$. This is equivalent to Grubbs' extended rule;

3. Kalman filter rule 2 which is same as above but σ_ε^2 is estimated on-line from $z_t = y_t - x_{t-1}$ using only the data available at time t;

4. Discrete integral controller (EWMA controller), where $\hat{d}_t = \hat{d}_{t-1} + \lambda y_t$.

	B_1 small	B_1 large
$|A|$ small	good choice (case 1)	bad choice (case 2)
$|A|$ large	bad choice (case 3)	good choice (case 4)

Table 3.1.   Scenarios of interested adjusting schemes, $A = (d - \hat{d}_0)/\sigma_\varepsilon$, $B_1 = P_0/\sigma_\varepsilon^2$.

There are two parameters that can be modified in Grubbs' adjustment rules: $\hat{d}_0$ and P_0. The effect of these parameters can be studied from looking at the effect of changes in A and B_1, as previously defined. Therefore, the four scenarios presented in Table 3.1 were investigated.

In the table, if the initial prior variance P_0 is large relative to σ_ε^2 (i.e., if B_1 is large), the weights K_t will be close to $1/t$ (Grubbs' harmonic rule), i.e., the initial estimate $\hat{d}_0$ will be discounted faster. This turns out to be a good decision if the initial offset estimate is far from d, where the distance between d and $\hat{d}_0$ is measured relative to σ_ε. A similar good decision is when P_0 is low and $\hat{d}_0$ is a good estimate of the offset (B_1 small, A small). In such case, $K_t < 1/t$, so there will be a slower discounting of the initial estimate $\hat{d}_0$. Cases (2) and (3) on the table indicate bad decisions, when the value of P_0 does not reflect how good the initial offset estimate really is. Since in the absence of historical information it is difficult to know *a priori* the value of d, it is of practical interest to study the four cases on the Table.

Table 3.2 contrasts the AISD performance of Grubbs' harmonic rule, the discrete integral controller (EWMA controller) and the Kalman Filter adjusting scheme (σ_ε^2 known). The table shows the values of AISD(m) for $m = 5, 10$ and 20. As can be

$m = 5$				
B_1 :	$|A| = 0$	$|A| = 1$	$|A| = 2$	$|A| = 3$
1/90	**1.00023**	1.95791	4.83092	9.61929
0.5	1.09344	1.48656	2.66589	4.63144
1	1.16394	**1.45667**	2.33483	3.79844
2	1.24138	1.47815	**2.18847**	3.37233
90	1.41043	1.61046	2.21056	**3.21074**
Grubbs	1.41667	1.61667	2.21667	3.21667
I controller ($\lambda = 0.1$)	1.01655	1.70215	3.75895	7.18696
I controller ($\lambda = 0.2$)	1.05601	1.55191	3.03962	5.51914
I controller ($\lambda = 0.3$)	1.10922	1.49030	2.63354	4.53894
$m = 10$				
B_1 :	$|A| = 0$	$|A| = 1$	$|A| = 2$	$|A| = 3$
1/90	**1.00049**	1.91004	4.63869	9.18644
0.5	1.09038	1.31359	1.98323	3.09930
1	1.13792	**1.29290**	1.75783	2.53271
2	1.18491	1.30578	**1.66840**	2.27276
90	1.27952	1.37954	1.67959	**2.17969**
Grubbs	1.28290	1.38290	1.68290	2.18290
I controller ($\lambda = 0.1$)	1.02830	1.49063	2.87761	5.18925
I controller ($\lambda = 0.2$)	1.08060	1.35518	2.17890	3.55178
I controller ($\lambda = 0.3$)	1.14190	1.33782	1.92558	2.90518
$m = 20$				
B_1 :	$|A| = 0$	$|A| = 1$	$|A| = 2$	$|A| = 3$
1/90	**1.00090**	1.82739	4.30685	8.43929
0.5	1.07243	1.19211	1.55117	2.14960
1	1.10008	**1.17989**	1.41931	1.81835
2	1.12585	1.18691	**1.37009**	1.67539
90	1.17564	1.22565	1.37568	**1.62573**
Grubbs	1.17739	1.22739	1.37739	1.62739
I controller ($\lambda = 0.1$)	1.03899	1.29825	2.07606	3.37240
I controller ($\lambda = 0.2$)	1.09568	1.23455	1.65116	2.34552
I controller ($\lambda = 0.3$)	1.15917	1.25721	1.55133	2.04152

Table 3.2. Kalman Filter adjusting scheme (σ_ε^2 known), Grubbs' harmonic rule and Integral controller AISD performance. $A = (d - \hat{d}_0)/\sigma_\varepsilon$, $B_1 = P_0/\sigma_\varepsilon^2$. Bold numbers are minimums by column.

seen from the Table, the "gap" between the column minimum and the AISD provided by Grubbs rule shrinks as the offset d gets much larger than σ_ε (i.e., as $|A|$ increases). This gap, however, is quite moderate except in the unrealistic case where one is very confident ($B_1 = P_0/\sigma_\varepsilon^2$ small) of our *a priori* offset estimate and the *a priori* offset estimate turns out to be quite accurate (i.e., $A = 0$). This is an unrealistic case because it implies we practically know the value of the offset d.

If $A = 0$, it can be seen from (3.16) and (3.17) that the AISD indices equal the average scaled variance since the deviations from target will always equal zero on average. If $d = \hat{d}_0 = 0$, the AISD quantifies the average *inflation in variance* we will observe for adjusting a process when there is no need to do so. Note that for $A = 0$ (no offset), one can get an inflation in variance equal to zero if $B_1 = 0$ in the Kalman Filter scheme or if $\lambda = 0$ in the integral control scheme. This inflation in variance has been studied, for discrete integral controllers, by Box and Luceño [18] and Del Castillo [30], although these authors looked at asymptotic variances, and not at small-sample variances as we do here.

Perhaps it should be pointed out that if one were extremely confident on the estimate of the offset of the machine ($B_1 \to 0$), simply setting $x_t = -\hat{d}_0$ for $t = 0, 1, \dots$. will result in an on-target process assuming we indeed have $\hat{d}_0 = d$. Thus, for most practical cases where a sequential adjustment rule is needed, the Kalman Filter rule (and Grubbs' extended rule) does not perform significantly better than Grubbs' harmonic scheme in the case of a constant unknown setup error.

Intuitively, if the variance σ_ε^2 is unknown the performance of the Kalman Filter scheme can only worsen. This was confirmed by estimating AISD using simulation. Thus

Grubbs harmonic rule is also superior, in the single realization case, to the Kalman Filter scheme with variance unknown.

Turning to the discrete integral controller, it can be seen that it also provides a very competitive scheme compared to the Kalman Filter scheme. The parameter λ has the effect of bringing the process back to target more rapidly the larger λ is. The trade-off is that there is an increase, for small A, of the AISD index as λ is increased. That is, the inflation in variance due to adjusting an on-target process increases as λ increases. From the Table, it appears the value $\lambda = 0.2$ provides a relatively good trade-off between fast return to target and inflation of variance if the process is really on-target (no offset).

3.2.3 Performance when the setup offset is a random variable

Suppose now that the offset d is a random variable such that $d \sim (\bar{d}_0, \sigma_d^2)$. Note that no assumption on the distribution of d is made. We wish to evaluate the performance of the different adjustment methods by averaging over the possible realizations of the random offset d. As mentioned earlier, if the mean and variance of d are known, then the Kalman Filter scheme, and hence, Grubbs extended rule are optimal for a quadratic loss function such as our AISD_d criterion. This was the case discussed by Grubbs [44] and Trietsch [100]. In this section we consider the more general case when the mean and variance of d are both unknown.

When d is random, we need to use *a prior* estimate $\hat{d}_0$ with associated variance P_0 (confidence on the *a prior* estimate $\hat{d}_0$) to start the Kalman Filter scheme (3.10). The situation is depicted in Figure 3.1. Using (3.15) as our performance index, the expression

Fig. 3.1. Starting the Kalman filter scheme, $\bar{d}_0$ and/or σ_d^2 unknown

for $\text{AISD}_d(m)$ is obtained from its definition as follows

$$
\begin{aligned}
\text{AISD}_d(m)_{KF} &= \int_{-\infty}^{\infty} \text{AISD}_{KF}(m) f_d(x) dx \\
&= C_1 \int_{-\infty}^{\infty} \frac{(d - \hat{d}_0)^2}{\sigma_\varepsilon^2} f_d(x) + C_2 \int_{-\infty}^{\infty} f_d(x) dx \\
&= \frac{C_1}{\sigma_\varepsilon^2} \int_{-\infty}^{\infty} (d^2 - 2d\hat{d}_0 + \hat{d}_0^2) f_d(x) dx + C_2
\end{aligned}
$$

Since $\int_{-\infty}^{\infty} d^2 f_d(x) dx = \sigma_d^2 - \bar{d}_0^2$ and $\int_{-\infty}^{\infty} d\, f_d(x) dx = \bar{d}_0$, then

$$
\text{AISD}_d(m)_{KF} = \frac{C_1}{\sigma_\varepsilon^2}(\sigma_d^2 + (\bar{d}_0 - \hat{d}_0)^2) + C_2 \tag{3.23}
$$

$$
= C_1(B_2 + A_2^2) + C_2, \tag{3.24}
$$

where $A_2 = (\bar{d}_0 - \hat{d}_0)/\sigma_\varepsilon$ is a measure of the average error incurred by the offset estimate, $B_2 = \sigma_d^2/\sigma_\varepsilon^2$ is a measure of the variability of the setup. The $\mathrm{AISD}_d(m)$ formula for Grubbs' harmonic rule is obtained in a similar way,

$$\mathrm{AISD}_d(m)_G = \frac{B_2 + A_2^2}{m} + C_3 \tag{3.25}$$

where C_1, C_2, and C_3 are functions of B_1 and m as shown in the last section. Recall that B_1 is a measure of confidence in $\hat{d}_0$, therefore, since σ_d^2 is not known, we have that in general $B_1 \neq B_2$.

For the discrete integral (or EWMA) controller, the corresponding closed-form expression for $\mathrm{AISD}_d(m)_{\mathrm{EWMA}}$ is

$$\mathrm{AISD}_d(m)_{\mathrm{EWMA}} = \frac{2}{2-\lambda} + \frac{1-(1-\lambda)^{2m}}{(2-\lambda)m}\left[\frac{B_2 + A_2^2}{\lambda} - \frac{1}{2-\lambda}\right]. \tag{3.26}$$

The AISD_d performance of the Kalman Filter approach, Grubbs harmonic rule, and that of an integral controller are evaluated using equations (3.23) (3.25) and (3.26). Figure 3.2 shows cases when the Kalman filter approach is better than Grubbs' harmonic rule for different values of B_1, B_2, A_2, and m. The shaded regions correspond to cases where $\mathrm{AISD}_d(m)_{KF} < \mathrm{AISD}_d(m)_G$. As it can be seen, for large average offsets (A_2 large) and/or large setup noise (B_2 large), Grubbs harmonic rule is better. Here "large" and "small" are terms relative to the process variance σ_ε^2. The advantage of the harmonic rule over the Kalman Filter scheme decreases with increasing value of $B_1 = P_0/\sigma_\varepsilon^2$. Note that under the assumptions in Case 1 above (when $\bar{d}_0$ and σ_ε^2 are known), i.e., when we

Fig. 3.2. Kalman Filter and Grubbs rule performance, random setup error. Shaded regions indicate cases for which $\text{AISD}_d(m)_{KF} < \text{AISD}_d(m)_G$. $B_1 = P_0/\sigma_\varepsilon^2, B_2 = \sigma_d^2/\sigma_\varepsilon^2, A_2 = |d_0 - \hat{d}_0|/\sigma_\varepsilon$

Fig. 3.3. Kalman Filter and discrete integral (EWMA) controller (with $\lambda = 0.2$) performance, random setup error. Shaded regions indicate cases for which $\mathrm{AISD}_d(m)_{KF} < \mathrm{AISD}_d(m)\mathrm{EWMA}$. $B_1 = P_0/\sigma_\varepsilon^2$, $B_2 = \sigma_d^2/\sigma_\varepsilon^2$, $A_2 = |d_0 - \hat{d}_0|/\sigma_\varepsilon$

Fig. 3.4. Kalman Filter and discrete integral (EWMA) controller (with $\lambda = 0.1$) performance, random setup error. Shaded regions indicate cases for which $\mathrm{AISD}_d(m)_{KF} < \mathrm{AISD}_d(m)_{\mathrm{EWMA}}$. $B_1 = P_0/\sigma_\varepsilon^2, B_2 = \sigma_d^2/\sigma_\varepsilon^2, A_2 = |d_0 - \hat{d}_0|/\sigma_\varepsilon$

Fig. 3.5. Grubbs harmonic rule and discrete integral (EWMA) controller performance, random setup error. Shaded regions indicate cases for which $\text{AISD}_d(m)_G < \text{AISD}_d(m)_{\text{EWMA}}$. $B_2 = \sigma_d^2/\sigma_\varepsilon^2, A_2 = |d_0 - \hat{d}_0|/\sigma_\varepsilon$

have that $B_1 = B_2$ and $A_2 = 0$, the Kalman Filter method always dominates Grubbs rule. This agrees with our earlier comment which indicated that the Kalman Filter scheme (and Grubbs extended rule) is optimal for the AISD_d criterion if the parameters are known.

Figures 3.3 and 3.4 compare the AISD_d performance of the Kalman Filter approach with that of an Integral controller with $\lambda = 0.2$ and $\lambda = 0.1$, respectively. As it can be seen, the Kalman Filter scheme is to be preferred in more cases as the number of observations m increases. The integral controller should be preferred when the average offset is large (A_2 large) and/or the setup is very variable (large B_2). This is even more true as the confidence in the initial offset mean decreases (i.e., the larger B_1). Observe how for cases where the average offset is very small the integral controller also dominates the Kalman Filter approach.

Finally, Figure 3.5 shows the AISD_d comparisons between Grubbs' harmonic rule and an integral controller. The integral controller outperforms the harmonic rule for cases near the origin, when A_2 is small (small average error in offset estimate) and B_2 is small (low setup variance). As the sample size increases, Grubbs' harmonic rule dominates the integral controller scheme.

3.3   Summary

In this chapter, an unifying point of view of some process adjustment procedures for setting up a machine was presented based on a Kalman Filter approach. A connection between Grubbs harmonic rule and stochastic approximation was made. The small sample properties of Grubbs adjustment schemes and that of an integral controller were

analyzed for the cases when a setup error is systematic (non-random) and when it is a random variable with unknown mean and variance. The performance metric used was a quadratic off-target process cost.

If the setup error is an unknown constant it was shown that for most practical cases when sequential adjustments are necessary, Grubbs' harmonic rule represents a better strategy than the Kalman Filter scheme. The even simpler integral or EWMA controller with weight $\lambda = 0.2$ provides a competitive alternative to the harmonic rule for cases when the offset is small (in the order of less than one standard deviation of the process). If the setup error is instead a random variable, an integral controller performs better than the Kalman Filter scheme when the setup noise is relatively high and the offset is very large on on average. When the offset is large and/or the setup noise is large, Grubbs harmonic rule outperforms the Kalman Filter scheme.

The analytic formulae presented in this chapter allow to obtain similar results for other values of the process and controller parameters without recourse to simulation. Further recommendations about when to use each method in the random setup error case can be reached by looking at Figures 3.2-3.5.

Chapter 4

Integration of Sequential Process Adjustments and SPC Control Charts

One essential task of Quality Engineers is to maintain a stable manufacturing process in the sense that the mean value of the quality characteristic of the process is kept right on its desired target level. However, random shocks on the initially on-target process may shift the process mean to an off-target value. Traditional Statistical Process Control (SPC) employs different types of control charts to detect such shifts in mean since the time of the shift is not predictable, but SPC techniques do not explicitly provide a process adjustment method. Process adjustment is usually regarded as a function pertaining to Engineering Process Control (EPC), an area that traditionally has belonged to process engineers rather than to quality engineers. The lack of adjustments existing in the SPC applications can cause large process off-target quality costs – a problem of particular concern in a short-run manufacturing process.

This chapter focuses on integrating different control charts with the sequential process adjustment method developed in Chapter 3 for monitoring and controlling a manufacturing process which experiences infrequent shifts in the process mean. Corresponding to the conditions prevalent in a short-run manufacturing environment, small-sample properties of this SPC/EPC integration scheme will be investigated. The performance of the integration scheme will be evaluated on both i.i.d. and autocorrelated processes.

4.1   Process Model

We assume a univariate process that consists of a measurable quality characteristic y and a single controllable factor x. The process mean is defined as the expectation of y and it is initially on its target level, but random shocks could shift it off-target. We also assume that the inherent process random errors are a sequence of identical independent distributed (i.i.d.) random numbers or a sequence of mildly autocorrelated random numbers.

If the process is an i.i.d. process, the process model is given by the following difference equation:

$$y_t = x_{t-1} + \mu_t + \varepsilon_t \qquad (4.1)$$

where μ_t is the process mean at sample or part t and $\{\varepsilon_t\}$ are a sequence of i.i.d. random errors, $\varepsilon_t \sim (0, \sigma_\varepsilon^2)$. In the simulation presented later, normality of the errors is assumed. Without loss of generality, the target of y_t is assumed to be zero, or y_t can be understood as a deviation from target. The process starts from the in-control state which is assumed to be such that the mean of the process equals the target, i.e., $\mu_1 = 0$ and

$$\mu_t = \mu_{t-1} + \delta(t), \quad \text{for } t = 2, 3, ... \qquad (4.2)$$

with

$$\delta(t) = \begin{cases} 0 & \text{if } t < t_0, \\ \delta \sim N(\mu_s, \sigma_s^2) & \text{if } t \geq t_0, \quad \text{where } t_0 = \text{shift time.} \end{cases}$$

Fig. 4.1. Step-type disturbance on the process mean

If the process is an autocorrelated process, an ARMA model for the process is used instead. However, the autocorrelation considered in this chapter is stationary and mild; otherwise, an automatic control scheme is recommended to be put on the process to overcome the autocorrelation. The stationary ARMA model of process observations is

$$A(\mathcal{B})(y_t - \mu_t) = B(\mathcal{B})\varepsilon_t, \tag{4.3}$$

where $A(\mathcal{B})$ and $B(\mathcal{B})$ are polynomials in the backshift operator $\mathcal{B}$ and μ is the process mean, which is again assumed to be 0 when the process is in control. So, at the in-control state, the process is modelled as

$$A(\mathcal{B})y_t = B(\mathcal{B})\varepsilon_t.$$

After a random shock, μ is shifted to an unknown level, say, s, then the process model changes to:

$$A(\mathcal{B})(y_t - s) = B(\mathcal{B})\varepsilon_t.$$

For example, for an ARMA(1,1) process, before the mean shift the process is modelled by

$$y_t = \phi y_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1},$$

and after the mean shift, the process changes to

$$y_t = (1 - \phi)s + \phi y_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}$$

and

$$y_{t_0} = s + \phi y_{t_0-1} + \varepsilon_{t_0} - \theta\varepsilon_{t_0-1},$$

where t_0 is the time of the mean shift. We still assume that the adjustment after the mean shift has been detected is additive to the whole process, that is, by applying adjustments, the whole process model is like

$$A(\mathcal{B})(y_t - s) = x_t + B(\mathcal{B})\varepsilon_t. \tag{4.4}$$

In this simple model one can see that the effect of the random shift in process mean can be eliminated by varying the controllable factor x after the shift is detected. In recent work (see, e.g, Chen and Elsayed [22], Crowder and Eshleman [27], Yashchin [117]), there is considerable emphasis on estimating a time-varying process mean instead

of adjusting for such variability. Because the true process mean is not observable directly, adjustments based only on one estimate are almost always biased. In this chapter, the sequential adjustment procedure presented in Chapter 3 will be integrated with several commonly used control charts for controlling a process with random mean shifts. In order to simplify the setup of the control chart, the process variance is assumed known in advance. It will be shown that this strategy – integrating control charts and sequential adjustments – is good at monitoring and adjusting a process under infrequent random shocks and it also simplifies the requirement for process mean estimation.

4.2   Control Charts and Adjustment Methods

As reviewed in Chapter 2, the commonly used control chart methods include Shewhart charts, EWMA charts and CUSUM charts. In the following sections, we will concentrate on comparing performance of integrating Shewhart charts and CUSUM charts with different adjustment methods. The adjustment methods that will be used include the single adjustment method based on the shift estimate from the control chart, sequential adjustment method and discrete integral (EWMA) control.

A control chart can be used not only to detect the time of the mean shift, but also to estimate the magnitude of the shift. Taguchi [97] advocated adjusting an opposite amount of observation whenever an observation on a Shewhart chart is out of control limits. This is criticized by Wiklund [111], since a single process observation could be severely biased from the true process mean. Wiklund proposed to use a maximum like- lihood estimate calculated from a conditional probability density function. His method and other methods are compared in Table 2.2. In addition, the following equation is

used for the CUSUM estimate of the mean:

$$\hat{\mu} = \begin{cases} K + \frac{c_t^+}{N^+} & \text{if } c_t^+ > H \\ -K - \frac{c_t^-}{N^-} & \text{if } c_t^- > H \end{cases} \tag{4.5}$$

where N^+ and N^- are the number of periods in which a run of non zero values of c^+

or c^- (as defined in (2.2)) were observed (Montgomery [73]). Shift detection and shift

size estimation are valuable for process adjustment purposes. If the shift size is precisely

known, it is obvious that by letting $x_{t+1} = -\mu_t$ the process will be reset back to its

target in view of equation (4.1). Nevertheless, due to the process disturbances $\{\varepsilon_t\}$, the

process mean is not directly observable.

One alternative to adjust this process was given in Chapter 3, namely, Grubbs'

harmonic and extended rules. Suppose the setup error can be adjusted directly, then the

solution which minimizes the variance of the next process observation y_{t+1} is obtained

by applying the adjustment rule as $x_t = x_{t-1} - a_t y_t$ with $a_t = 1/t$, which constitutes a

harmonic sequence. Clearly, this solution can be easily applied on a general manufactur-

ing process, where its process mean might be shifted to an unknown level at a random

time.

An alternative process adjustment method is to apply consistent process control

actions on the process regardless whether the process mean is off-target or not. One of

the most commonly used controllers is the discrete integral (EWMA) controller, which

adjust the process setting according to this equation: $x_t = x_{t-1} - \lambda y_t$, where λ is a

constant and $0 < \lambda < 1$. This automatic control scheme has been intensively discussed

in recent process control literature, especially for the semiconductor industry (see, for example, Sachs *et. al.* [86]). In the next section, its performance on adjusting for process mean shifts will be compared with other SPC/EPC integration schemes.

4.3 Integration of Control Charts and Sequential Adjustments for IID Processes

The proposed integrated process monitoring and adjustment scheme consists of three steps: monitor the process using a control chart, estimate the shift size when a shift in the process mean is detected, and finally apply the sequential adjustment procedure to bring the process mean back to target. The performance of the integrated scheme depends on the sensitivity of the control chart to detect shifts in the process mean, on the accuracy of the initial estimate of shift size and on the number of sequential adjustment that are made. To compare the performance of various combinations of control charts and adjustment methods, we first simulate an i.i.d. manufacturing process (4.1) for a total of 50 observations, and monitor and adjust it using one of the six methods listed on Table 4.1. The performance of each method is evaluated by the index AISD as defined in Chapter 3.

In the simulation, a shift in process mean occurs after the fifth observations and adjustments are conducted immediately after the shift is detected. The mean value of 10,000 simulation results are illustrated in Figure 4.2. The y axis in the figure represents the percentage improvement in the AISD of using some adjustment method compared to the AISD without adjustment, i.e., $\frac{\text{AISD}_{no\ adjust} - \text{AISD}_{method\ i}}{\text{AISD}_{no\ adjust}} \times 100$, so this is a "larger the better" value. This value is plotted with respect to the actual shift size which

Method	Shift detection	Shift size estimation	Adjustment
1	Shewhart chart for individuals (3σ limits)	Last observation (Taguchi's method)	one adjustment after an out-of-control alarm
2	Shewhart chart for individuals (3σ limits)	Maximum Likelihood Estimate (Wiklund's method)	one adjustment according to the MLE value
3	CUSUM chart for individuals (k=0.5 h=5)	CUSUM estimate (equation (4.5))	one adjustment according to the CUSUM estimate
4	Shewhart chart for individuals (3σ)	last observation (Taguchi's method)	5 sequential adjustments following with $a_t = 1/(t - t')$
5	Shewhart chart for individuals (3σ)	MLE (Wiklund's method)	5 sequential adjustments following with $a_t = 1/(t - t')$
6	CUSUM chart for individuals (k=0.5 h=5)	CUSUM estimate (equation (4.5))	5 sequential adjustments following with $a_t = 1/(t - t')$

Table 4.1. Six methods of integrating control charts and sequential adjustments. t' is the time of detecting a mean shift.

was varied from 0 to 4σ. Here the shift sizes are constant (i.e., $\sigma_s = 0$). One can see that the sequential adjustment methods (4 to 6) are superior to the one-step adjustment methods (1 to 3) for almost all shift sizes. More specifically, using a CUSUM chart and sequential adjustments (Method 6) has significant advantage over other methods when the shift size is small or moderate, and using a Shewhart chart and sequential adjustments (Method 4) is better for large shifts. Moreover, one-step adjustment methods, especially the Taguchi's method, may dramatically deteriorate a process when the shift size is small. No method can improve the AISD when the shift size is very small, but comparatively Method 6 is still better than others.

To study a general shifting process, the mean shift in the following simulation is changed to a stochastic process in which shifts occur randomly in time according to a geometric distribution. Specifically, the occurrence of a shift at each run is a Bernoulli trial with probability $p = 0.05$ and the shift size is normally distributed as $s \sim N(\mu_s, 1)$.

Fig. 4.2.   Performance of six integrated methods of control charts and adjustments (the process mean was shifted after the 5th observation)

Besides the previous six methods, an integral control scheme (i.e., an EWMA controller) was studied for comparison purposes. The convergence of EWMA schemes with a small control parameter for adjusting a step type disturbance has been shown by Sachs *et al.* [86]. The control parameter λ of the EWMA controller was set at 0.2. There is no process monitoring needed for the integral control scheme because the controller is always in action. The simulations were repeated 10,000 times.

For this general shift model, it can be seen from Figure 4.3 that sequential adjustment methods still out-perform any one-step adjustment method. Evidently, the EWMA controller performs better than any other sequential method when the shift size mean is small, which explains the popularity of EWMA controllers. However, one main advantage of the proposed SPC/EPC integrated methods is that they detect process changes using common SPC charts whereas the EWMA controller alone does not have this SPC function, in other words, there is no possibility for process improvement through correction of assignable causes if only an EWMA controller is utilized. Process improvement through human intervention is facilitated by having a monitoring (SPC) mechanism that triggers the adjustment procedure and keeps a time-based record of alarms useful for process diagnostics.

The step-type random shift process considered in this chapter is similar to Barnard's model (Barnard, [8]). When the shifts occur more and more frequently, the process approaches an IMA (Integrated Moving Average) disturbance. As it is well-known, the EWMA controller is the minimum variance controller for a responsive process with IMA disturbance (Box and Luceño, [18]). This explains why the EWMA controller works better when shifts occur more frequently (larger p) than less frequently (smaller p).

Fig. 4.3. Performance of EPC and SPC integration for a more general shift model (the shift occurs with probability $p = 0.05$ at each observation)

However, since this scheme conducts adjustments at every process run, it cannot be suitable for a process where adjustment costs are considerable. The economic consideration of adjustments will be discussed in Section 4.3.2.

Another drawback of the EWMA controller is that one has to decide what value of the control parameter λ to use. It is recommended that this parameter should be small in order to maintain the stability of the process, but small parameter values may not be optimal from an AISD point of view, especially when the mean shift size is large. Moreover, the high performance of the EWMA scheme comes from the frequent random shifts modeled in the previous simulation study (an average of 2.5 shifts per 50 runs). If the chance of shifts decreases, the inflation of variance which is caused by adjusting an on-target process will deteriorate the effectiveness of this scheme. The small-sample properties of the variance provided by EWMA and harmonic adjusting rules have been discussed in Chapter 3.

In Figure 4.4, the probability of random shifts p was decreased to 0.01 and the same simulation as in Figure 4.3 was conducted. Under these conditions, the EWMA method cannot compete well with the sequential adjustment methods combined with CUSUM or Shewhart chart monitoring. More simulation results for different probabilities of shifts p are listed in Table 4.2. It is found that the EWMA adjustment method is better for small shifts and Method 4 is better for large shifts when p is large; as p gets smaller ($p < 0.02$), i.e., the process is subject to infrequent random shocks, Method 6 gets harder to beat. Therefore, the proposed SPC/EPC integrated methods work better when p is small. This is relevant in certain types of manufacturing where process upsets occur very rarely, e.g., microelectronic and semiconductor industries.

Fig. 4.4. Performance of EPC and SPC integration for the general shift model, less frequent shifts (p=0.01).

% improvement on AISD		Mean of shift size				
		0	1σ	2σ	3σ	4σ
p=0.05	Method 4	11.20	36.05	62.27	**74.93**	**81.04**
		(0.30)	(0.38)	(0.36)	(0.31)	(0.29)
	Method 6	18.89	41.50	64.07	73.90	78.73
		(0.28)	(0.35)	(0.33)	(0.30)	(0.29)
	EWMA controller	**24.91**	43.11	60.47	67.76	70.71
	($\lambda = 0.1$)	(0.27)	(0.32)	(0.30)	(0.28)	(0.28)
	EWMA controller	24.51	**45.32**	65.26	73.31	76.68
	($\lambda = 0.2$)	(0.30)	(0.36)	(0.33)	(0.31)	(0.30)
	EWMA controller	21.16	44.02	**65.59**	74.21	78.38
	($\lambda = 0.3$)	(0.33)	(0.39)	(0.36)	(0.33)	(0.32)
p=0.035	Method 4	6.65	24.31	47.76	**62.41**	**68.85**
		(0.26)	(0.37)	(0.40)	(0.39)	(0.38)
	Method 6	13.80	30.35	50.56	61.91	66.68
		(0.25)	(0.33)	(0.36)	(0.36)	(0.36)
	EWMA controller	**18.31**	32.18	48.68	56.01	59.58
	($\lambda = 0.1$)	(0.25)	(0.32)	(0.34)	(0.34)	(0.34)
	EWMA controller	16.82	**32.81**	51.21	61.09	64.35
	($\lambda = 0.2$)	(0.29)	(0.36)	(0.39)	(0.38)	(0.39)
	EWMA controller	13.13	30.33	**51.76**	60.82	65.48
	($\lambda = 0.3$)	(0.32)	(0.40)	(0.41)	(0.41)	(0.42)
p=0.02	Method 4	1.48	11.85	28.86	41.60	**48.34**
		(0.24)	(0.32)	(0.39)	(0.43)	(0.45)
	Method 6	8.07	17.52	**32.53**	**41.94**	47.20
		(0.21)	(0.29)	(0.36)	(0.39)	(0.41)
	EWMA controller	**10.37**	**18.86**	30.68	38.05	41.06
	($\lambda = 0.1$)	(0.22)	(0.29)	(0.35)	(0.38)	(0.39)
	EWMA controller	7.35	17.09	31.40	39.49	43.57
	($\lambda = 0.2$)	(0.26)	(0.33)	(0.40)	(0.43)	(0.45)
	EWMA controller	2.16	13.03	28.90	38.19	42.28
	($\lambda = 0.3$)	(0.28)	(0.37)	(0.44)	(0.47)	(0.49)
p=0.005	Method 4	-3.36	-1.02	3.64	9.02	12.57
		(0.18)	(0.21)	(0.28)	(0.34)	(0.37)
	Method 6	**1.32**	**3.60**	**7.88**	**11.77**	**14.37**
		(0.12)	(0.16)	(0.23)	(0.28)	(0.32)
	EWMA controller	-0.36	1.55	5.53	7.72	9.95
	($\lambda = 0.1$)	(0.13)	(0.17)	(0.24)	(0.27)	(0.30)
	EWMA controller	-5.55	-2.91	1.42	4.89	7.19
	($\lambda = 0.2$)	(0.16)	(0.21)	(0.27)	(0.32)	(0.35)
	EWMA controller	-11.25	-8.47	-2.94	0.42	3.15
	($\lambda = 0.3$)	(0.18)	(0.23)	(0.31)	(0.36)	(0.39)

Table 4.2.   Performance of SPC/EPC integrated adjustment schemes and EWMA scheme when varying the probability of a shift. The numbers are the mean values and standard errors (in parenthesis) of the percentage improvement on AISD (compared to the process without adjustment) computed from 10,000 simulations. Bold numbers are largest improvement for each p and mean shift size combination.

4.3.1   An improved integrated SPC/EPC method

The performance of the different SPC/EPC integration methods studied herein depends on 1) their ability to detect a shift and 2) their ability to estimate the process mean. When a process is in the in-control state, an out-of-control alarm signaled by the control chart is called a false alarm. Adjustments triggered by false alarms will inflate the variance of the in-control process, although the inflation will decrease when sequential adjustments are used. On the other hand, if the control chart cannot signal an alarm quickly after a real shift has occurred, it will also impede a quick recovery through adjustment.

Since the detection properties of a CUSUM chart can be tuned by modifying its design parameters h and k, it is of interest to study Method 6 with different CUSUM chart parameters. In Figure 4.5, several different values of h were tried while fixing k at 0.5 to make the chart sensitive to small shifts. It was found that when h is small, the process will suffer from a large number of false alarms generated by the control chart; when h is large, the improvement in AISD will be limited for large shift sizes due to the lack of sensitivity that the CUSUM chart has to large shifts. A CUSUM chart with $h = 5$ seems to be the best choice since it gives fewer false alarms for a normal process and has comparatively short ARLs for large shift sizes.

In order to improve further the performance of Method 6 for large frequent shifts, we propose a hybrid monitoring scheme combined with a sequential adjustment scheme. A combined CUSUM-Shewhart chart is used, where the parameters on the CUSUM are $k = 0.5$ and $h = 5$ and the control limits on the Shewhart chart are set at $\pm 3.5\sigma$.

Fig. 4.5. Performance of Method 6 with different parameters in the CUSUM chart (the process mean was shifted after the 5th observation)

ARL	Shift size				
	0	1σ	2σ	3σ	4σ
CUSUM (h=3)	59	6.36	2.56	1.59	1.15
CUSUM (h=4)	169	8.34	3.22	1.98	1.44
CUSUM (h=5)	469	10.34	3.89	2.39	1.72
CUSUM-Shewhart	391	10.20	3.77	2.10	1.34
Shewhart (3σ)	370	43.96	6.30	2.00	1.19

Table 4.3. ARLs of CUSUM and CUSUM-Shewhart charts. The value $k = 0.5$ was used in the CUSUM charts and $k = 0.5$, $h = 5$ and $c = 3.5$(Shewhart control limit) were used in the CUSUM-Shewhart chart. The ARLs of CUSUM charts are approximated using equations given by Siegmund [88] and the ARLs of CUSUM-Shewhart chart are from Montgomery [73].

Fig. 4.6. Performance of a hybrid monitoring and adjusting method (the process mean was shifted after the 5th observation)

Whenever the combined chart signals an alarm, the initial estimate of the shift size will be given by the CUSUM estimate if it is smaller than 1.5σ; otherwise, it will be the negative value of y_t (Taguchi's method, Taguchi [97]). The average run lengths of this combined monitoring approach are contrasted with those of a CUSUM chart in Table 4.3. Comparing this new method to Methods 4 and 6 (see Figure 4.6), one can see that the new method makes a considerable improvement on the large shift size while sacrificing a little for small shift sizes. This trade-off cannot be avoided due to the nature of this hybrid monitoring method.

We finally point out in this section that a method for sequentially adjusting the parameter of an EWMA controller was recently proposed by Guo, Chen and Chen [45]. They use two EWMA control charts for detecting moderate (2σ) and large (3σ) shifts. After detection, a harmonic adjustment sequence is triggered when either chart signals an alarm. In Figure 4.7, the two-EWMA method with the suggested chart parameters by Guo *et. al.* is compared with Method 4, Method 6 and with the hybrid monitoring method proposed before by using the general shift model with the shift probability p equals 0.05. Clearly, the two-EWMA method performs worse than other methods, especially on large shift sizes. This can be explained by the insensitivity of EWMA chart on estimating a general shift size (see Table 2.2).

4.3.2 Cost justification

In quality control, the cost of adjusting a process usually can not be ignored, because the adjustment is a set of decisions and actions such as stopping the process, investigating the causes of out-of-control and resetting the process. In the simulations

Fig. 4.7. Comparing the two-EWMA method with other SPC/EPC integrated schemes. $\lambda_1 = 0.6$, $L_1 = 3.285$ and $\lambda_2 = 0.33$, $L_2 = 3.25$ were used for the two EWMA charts. Shifts occur with $p = 0.05$.

shown in the previous section, the number of sequential adjustments was arbitrarily selected as five. The economic consideration of the number of adjustments will be discussed in this section.

When the harmonic adjustment rule is applied on a shifted process with shift size s, $(x_t + s)$ is asymptotically normally distributed [84], that is

$$(x_t + \mu_s) \to^D N(0, \sigma^2/t).$$

Therefore, x_t will be likely in the interval of $\pm 3\sigma/\sqrt{t}$ around $-s$ and the effect of an adjustment will decrease rapidly when the number of adjustments grows. Figure 4.8 presents the results of the simulation studies where different numbers of adjustments are applied. The process was assumed to have a mean shift after the 5th run and the simulation was repeated 1,000 times. It was found that after four or five adjustments the AISDs of the process can not be further improved significantly. In Figure 4.9, a 3-D plot of the AISD improvement function of Method 6 (integrated CUSUM chart and sequential adjustments) is shown as a function of the mean shift size (μ_s) and the number of adjustments. The AISD improvement function is very flat on the adjustment number axis, as opposed to the AISD as a function of the shift size. Therefore, it is not worth to do many adjustments.

The optimal number of adjustments can be obtained if the cost elements of the off-target process and the cost of adjustments are known. By marginal analysis, adjustments should be conducted as long as the adjustment cost is lower than the savings obtained from decreasing the AISD by adjusting the process one more time. Suppose a process is

Fig. 4.8. Improvements on AISDs when the adjustment number increases (a general shift model with p=0.05, $\mu_s = 2$ and $\sigma_s = 1$ was used.)

Fig. 4.9. A 3-D view of the AISD improvements from Method 6 when both shift size and adjustment number change ($\sigma_s = 1$)

M/Ω	1	2	5	10
n	6	4	2	1

Table 4.4. Optimal number of adjustments

going to be run for N observations or parts, and it will be adjusted sequentially for the first n parts. Then the AISD is defined as:

$$AISD(n, N) = \frac{n \ AISD(n) + (N - n)(Var(y_{n+1}) + E[y_{n+1}]^2)}{N} \qquad (4.6)$$

So the adjustment is only profitable when

$$\Omega N\{AISD(n, N) - AISD(n + 1, N)\} > M, \qquad (4.7)$$

where Ω is the unit off-target quality cost and M is the adjustment cost. By using equations (3.18), (3.20), (3.21) and (4.6), we get

$$n < \frac{\sqrt{(M + \Omega\sigma^2)^2 + 4(N - 1)M\Omega\sigma^2} - (M + \Omega\sigma^2)}{2M} \qquad (4.8)$$

For example, with $N = 50$ and $\sigma = 1$, the optimal number of adjustments computed by equation (4.8) is given in Table 4.4.

4.4 Integrated SPC/EPC Methods for Autocorrelated Processes

When process observations are autocorrelated, it is well-known that traditional control charts that are designed for detecting changes in an i.i.d. process would have substantially deteriorated shifts detection properties. The main problem associated with autocorrelated data is that the autocorrelation may induce a large amount of false alarms, which, in turn, reduce the usability of a control chart for detecting true shifts in the process mean.

Two categories of SPC methods for autocorrelated data were reviewed in Chapter 2. In the first approach, filtering out the autocorrelation of the process (fitting an ARMA model to the data and inverting it), process mean shifts are also transformed into some complicated transient patterns which are hard to detect even when the process residuals are an i.i.d. sequence. In the second approach, changing the control chart limits allows to maintain the normal ARL property of a control chart and it is easy to apply since the ARMA model of the autocorrelated process is not necessary. Therefore, in this section, we will recalculate the control chart limits for AR(1) and MA(1) processes. Specifically, the method proposed by Yashchin [118] for CUSUM charts will be used and this modified chart will be applied on simulated mildly autocorrelated processes. Similar to the previous section, the shifts in the process mean occur at random times. Whenever this shift is detected by the control chart, five sequential adjustments are applied on the process.

Yashchin's design of CUSUM charts consists of two steps. First, use an i.i.d. sequence $\{y_t^*\}$ to substitute the original autocorrelated sequence $\{y_t\}$, where $\{y_t^*\}$ must

satisfy the following two requirements: i) its "runaway range" r (the run-length before

its CUSUM statistic exceeds control limits) should be approximately equal to that of the

original sequence; and ii) $\sum_{t=1}^{r} y_t^*$ and $\sum_{t=1}^{r} y_t$ should have the same first and second

moments. Then, a suitable CUSUM chart for this i.i.d. sequence can be designed.

For an ARMA(1,1) process, like $y_t = \phi y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}$, the mean of the i.i.d

sequence $\{y_t^*\}$ should be equal to the mean of $\{y_t\}$, which is μ, and the variance of $\{y_t^*\}$

can be computed from

$$\sigma_r^2 = \frac{1}{r} Var\left(\sum_{t=1}^{r} y_t\right) = \sigma_y^2\left[1 + \frac{2\phi d}{1-\phi} + \frac{2\phi d(1-\phi^r)}{r(1-\phi^2)}\right], \tag{4.9}$$

where σ_y^2 is the marginal variance of $\{y_t\}$, and

$$\sigma_y^2 = \frac{\sigma_\varepsilon^2(1+\theta^2 - 2\phi\theta)}{1-\phi^2} \quad \text{and} \quad d = \frac{(1-\phi\theta)(\phi-\theta)}{\phi(1+\theta^2 - 2\phi\theta)}.$$

The "runaway range", r, can then be obtained by considering a Brownian Motion process

with drift μ and standard deviation per unit time σ_r. The value r is the expected time

to absorption at the control limit h of this Brownian Motion process starting from 0 at

time 0, given it does not return to 0, i.e.,

$$r = \frac{\sigma_r^2}{\mu^2}\left(\frac{\mu h/\sigma_r^2}{th(\mu h/\sigma_r^2)} - 1\right) \tag{4.10}$$

where $th(.)$ is the hyperbolic tangent function and h is the control limit on a one-side

CUSUM chart. Using equations (4.9) and (4.10), we can calculate the variance of the

Coefficients of autocorrelation	$\phi=0.2$	$\phi = 0.15$	$\phi = 0.1$	$\phi = 0.05$
Design parameter for CUSUM charts	$h = 6.3$	$h = 5.8$	$h = 5.3$	$h = 4.9$
Shift size	ARL (SRL)	ARL (SRL)	ARL (SRL)	ARL (SRL)
0	299 (17.2)	304 (17.2)	299 (16.8)	304 (17.1)
0.5	42.2 (5.84)	40.6 (5.64)	37.7 (5.51)	35 (5.34)
1	13.1 (2.72)	12 (2.58)	11.1 (2.46)	10.4 (2.43)
1.5	7.31 (1.75)	6.71 (1.69)	6.17 (1.59)	5.7 (1.53)
2	4.95 (1.31)	4.65 (1.25)	4.27 (1.21)	3.93 (1.14)
2.5	3.83 (1.06)	3.57 (1.02)	3.28 (0.98)	3.05 (0.95)
3	3.15 (0.91)	2.95 (0.88)	2.72 (0.84)	2.53 (0.82)

Table 4.5. Run-length distributions of CUSUM charts with modified control limits for AR(1) processes, $y_t = \phi y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim N(0, 1)$

i.i.d. sequence according to various selections of h; thus, the ARLs of this sequence for the CUSUM chart with control limit h can be obtained.

In this research, Yashchin's method is used to find an initial design value of the CUSUM chart's limit, then the chart's ARL property is verified by simulation study and the chart's limit may be further modified until the simulation result is desired. Tables 4.5 and 4.6 list the ARLs and SRLs (standard deviations of run lengths) of several modified CUSUM charts for AR(1) and MA(1) processes by the simulation study (Here, another parameter of the CUSUM chart, k, is 0.5. Recall a usual CUSUM for an i.i.d. process, $y_t \sim^{iid} (0, 1)$, has $k = 0.5$ and $h = 5$). The in-control ARLs are set to be roughly 300.

The modified CUSUM charts are applied on an autocorrelated process. Whenever an "out-of-control" signal is triggered, this activates five sequential adjustments for removing the process mean shift. As in the previous section, fifty process observations were simulated. At each observation there is a small possibility (from 0.5% to 5%) such

Coefficients of autocorrelation	$\theta=-0.2$	$\theta =- 0.15$	$\theta =- 0.1$	$\theta = -0.05$
Design parameter for CUSUM charts	$h = 6.0$	$h = 5.6$	$h = 5.3$	$h = 4.9$
Shift size	ARL (SRL)	ARL (SRL)	ARL (SRL)	ARL (SRL)
0	299 (17.2)	304 (17.2)	299 (16.8)	304 (17.1)
0.5	42.2 (5.84)	40.6 (5.64)	37.7 (5.51)	35 (5.34)
1	13.1 (2.72)	12 (2.58)	11.1 (2.46)	10.4 (2.43)
1.5	7.31 (1.75)	6.71 (1.69)	6.17 (1.59)	5.7 (1.53)
2	4.95 (1.31)	4.65 (1.25)	4.27 (1.21)	3.93 (1.14)
2.5	3.83 (1.06)	3.57 (1.02)	3.28 (0.98)	3.05 (0.95)
3	3.15 (0.91)	2.95 (0.88)	2.72 (0.84)	2.53 (0.82)

Table 4.6. Run-length distributions of CUSUM charts with modified control limits for MA(1) processes, $y_t = \varepsilon_t - \theta \varepsilon_{t-1}$, where $\varepsilon_t \sim N(0, 1)$

that the process mean could be shifted to a random number s, where $s \sim N(\mu_s, 1)$. The simulation was repeated 10,000 times. Five methods were tested: modified CUSUM chart plus single adjustment, modified CUSUM chart plus five sequential adjustments, and EWMA controllers with the control parameter λ ($\lambda = 0.1, 0.2, 0.3$). The percentage of improvements on the AISDs of the adjusted process compared to the unadjusted process are listed in Tables 4.7-4.10. More results are included in Appendix B.

From these tables, one can see that clearly a scheme which integrates a CUSUM charts with sequential adjustments outperforms those where only a single adjustment is made. However, EWMA controllers with properly selected λ values are very competitive especially when shift sizes are not large. This result is reasonable since the EWMA controller keeps constant adjustments on the process, which will not only adjust for possible process mean shifts but also compensate for the autocorrelations in the original

process observations. The advantage of integration schemes becomes visible only when the process autocorrelations are small and the chance of mean shifts is small.

4.5 Summary

In this chapter, several combinations of process monitoring and adjusting methods were studied for both i.i.d. processes and autocorrelated processes. It was found that sequential adjustments are superior to single adjustment strategies for almost all types of process shifts and magnitudes considered. A CUSUM chart used together with a simple sequential adjustment scheme was found to reduce the AISDs of a shifted process more than any other combined scheme when the shift size is not very large. It was further proposed that a hybrid CUSUM-Shewhart monitoring method, when coupled with a sequential adjustment scheme, has a more competitive performance on both small and large shift sizes.

Unlike some commonly used automatic process control methods, the integrated SPC/EPC schemes that we proposed do not require continuous adjustments on the process. Therefore, these methods are suitable for process control when the process is subject to infrequent random shocks. The number of adjustments can be justified by comparing the cost and the benefit of the adjustment. Since sequential adjustments are applied, the effect of the initial estimate of the process mean is not as critical as in the single adjustment method, so this method requires much less computation effort and is easy to be implemented on the manufacturing floor.

% improvement on AISD		Mean of shift size				
		0	1σ	2σ	3σ	4σ
p=0.05	CUSUM chart +	9.18	34.56	59.73	71.95	78.73
	Single adj.	(0.35)	(0.40)	(0.37)	(0.33)	(0.30)
	CUSUM chart +	17.33	40.31	62.98	73.87	**80.04**
	Sequential adj.	(0.29)	(0.36)	(0.34)	(0.31)	(0.28)
	EWMA controller	24.87	42.76	59.78	66.96	70.70
	($\lambda = 0.1$)	(0.26)	(0.31)	(0.30)	(0.28)	(0.26)
	EWMA controller	**26.18**	**46.04**	65.13	73.33	77.76
	($\lambda = 0.2$)	(0.29)	(0.34)	(0.32)	(0.30)	(0.28)
	EWMA controller	25.09	46.01	**66.17**	**74.94**	79.76
	($\lambda = 0.3$)	(0.30)	(0.36)	(0.34)	(0.32)	(0.29)
p=0.035	CUSUM chart +	4.96	22.80	45.47	58.63	65.86
	Single adj.	(0.32)	(0.39)	(0.40)	(0.39)	(0.39)
	CUSUM chart +	12.53	28.70	49.22	61.29	**67.67**
	Sequential adj.	(0.26)	(0.34)	(0.37)	(0.37)	(0.37)
	EWMA controller	19.13	31.97	47.62	55.70	59.60
	($\lambda = 0.1$)	(0.24)	(0.31)	(0.33)	(0.33)	(0.34)
	EWMA controller	**19.39**	**33.79**	51.67	61.01	65.60
	($\lambda = 0.2$)	(0.27)	(0.35)	(0.37)	(0.37)	(0.37)
	EWMA controller	17.66	32.92	**52.03**	**62.07**	67.05
	($\lambda = 0.3$)	(0.29)	(0.37)	(0.39)	(0.39)	(0.39)
p=0.02	CUSUM chart +	1.00	10.12	26.97	38.42	45.60
	Single adj.	(0.27)	(0.34)	(0.39)	(0.43)	(0.45)
	CUSUM chart +	7.13	16.04	31.26	**41.69**	**47.97**
	Sequential adj.	(0.21)	(0.29)	(0.36)	(0.40)	(0.43)
	EWMA controller	**11.64**	19.38	30.92	38.14	42.08
	($\lambda = 0.1$)	(0.22)	(0.28)	(0.34)	(0.37)	(0.38)
	EWMA controller	10.55	**19.48**	**32.80**	41.25	45.89
	($\lambda = 0.2$)	(0.25)	(0.31)	(0.38)	(0.41)	(0.43)
	EWMA controller	8.03	17.64	31.95	41.10	46.13
	($\lambda = 0.3$)	(0.27)	(0.34)	(0.40)	(0.44)	(0.46)
p=0.005	CUSUM chart +	-4.25	-2.01	3.49	7.57	10.13
	Single adj.	(0.21)	(0.24)	(0.29)	(0.33)	(0.37)
	CUSUM chart +	0.13	2.39	7.47	**11.02**	**13.28**
	Sequential adj.	(0.13)	(0.17)	(0.25)	(0.29)	(0.33)
	EWMA controller	**1.86**	**3.85**	**7.94**	10.22	11.56
	($\lambda = 0.1$)	(0.13)	(0.17)	(0.23)	(0.27)	(0.29)
	EWMA controller	-0.73	1.55	6.34	9.08	10.68
	($\lambda = 0.2$)	(0.15)	(0.20)	(0.27)	(0.31)	(0.33)
	EWMA controller	-4.09	-1.64	3.54	6.53	8.28
	($\lambda = 0.3$)	(0.17)	(0.21)	(0.29)	(0.33)	(0.36)

Table 4.7.   Performance of integrated SPC/EPC adjustment schemes and an EWMA adjustment scheme on an AR(1) process when varying the probability of a shift.

% improvement on AISD		Mean of shift size				
		0	1σ	2σ	3σ	4σ
p=0.05	CUSUM chart +	10.48	35.43	61.34	73.54	79.07
	Single adj.	(0.36)	(0.41)	(0.37)	(0.32)	(0.30)
	CUSUM chart +	18.77	41.40	64.51	**75.35**	**80.25**
	Sequential adj.	(0.30)	(0.36)	(0.34)	(0.30)	(0.29)
	EWMA controller	24.54	42.51	59.80	67.14	70.11
	($\lambda = 0.1$)	(0.26)	(0.32)	(0.30)	(0.28)	(0.27)
	EWMA controller	**24.70**	**44.85**	**64.69**	73.21	76.74
	($\lambda = 0.2$)	(0.30)	(0.36)	(0.33)	(0.31)	(0.30)
	EWMA controller	22.31	43.81	65.22	74.50	78.37
	($\lambda = 0.3$)	(0.32)	(0.38)	(0.35)	(0.32)	(0.32)
p=0.035	CUSUM chart +	5.74	24.26	47.34	59.71	66.02
	Single adj.	(0.33)	(0.39)	(0.40)	(0.40)	(0.39)
	CUSUM chart +	13.17	30.18	51.05	**62.15**	**67.75**
	Sequential adj.	(0.27)	(0.35)	(0.37)	(0.37)	(0.38)
	EWMA controller	**18.30**	31.94	47.79	55.17	58.51
	($\lambda = 0.1$)	(0.25)	(0.31)	(0.34)	(0.34)	(0.35)
	EWMA controller	17.23	**32.85**	**51.09**	59.88	63.89
	($\lambda = 0.2$)	(0.29)	(0.36)	(0.38)	(0.38)	(0.39)
	EWMA controller	14.05	30.89	50.64	60.29	64.75
	($\lambda = 0.3$)	(0.31)	(0.38)	(0.41)	(0.41)	(0.41)
p=0.02	CUSUM chart +	0.14	11.60	27.58	39.34	45.35
	Single adj.	(0.29)	(0.35)	(0.41)	(0.43)	(0.45)
	CUSUM chart +	6.91	17.51	**31.89**	**42.26**	**47.69**
	Sequential adj.	(0.23)	(0.31)	(0.37)	(0.41)	(0.43)
	EWMA controller	**10.27**	**19.08**	29.82	36.98	40.22
	($\lambda = 0.1$)	(0.22)	(0.29)	(0.35)	(0.37)	(0.39)
	EWMA controller	7.67	17.88	30.48	39.05	43.06
	($\lambda = 0.2$)	(0.25)	(0.33)	(0.40)	(0.43)	(0.44)
	EWMA controller	3.51	14.58	28.37	37.81	42.31
	($\lambda = 0.3$)	(0.28)	(0.36)	(0.43)	(0.46)	(0.48)
p=0.005	CUSUM chart +	-4.94	-1.96	2.90	7.34	10.01
	Single adj.	(0.23)	(0.25)	(0.30)	(0.34)	(0.38)
	CUSUM chart +	-0.13	**2.46**	**6.90**	**10.88**	**13.33**
	Sequential adj.	(0.15)	(0.19)	(0.25)	(0.30)	(0.34)
	EWMA controller	**0.28**	2.32	5.71	8.36	10.10
	($\lambda = 0.1$)	(0.14)	(0.18)	(0.23)	(0.27)	(0.30)
	EWMA controller	-4.13	-1.75	2.31	5.52	7.60
	($\lambda = 0.2$)	(0.16)	(0.21)	(0.27)	(0.32)	(0.35)
	EWMA controller	-9.44	-6.85	-2.36	1.22	3.50
	($\lambda = 0.3$)	(0.18)	(0.23)	(0.30)	(0.35)	(0.38)

Table 4.8. Performance of integrated SPC/EPC adjustment schemes and an EWMA adjustment scheme on an AR(1) process with $\phi = 0.05$ when varying the probability of a shift.

	% improvement on AISD	Mean of shift size				
		0	1σ	2σ	3σ	4σ
p=0.05	CUSUM chart + Single adj.	9.44 (0.35)	34.67 (0.40)	59.13 (0.37)	72.12 (0.33)	78.44 (0.30)
	CUSUM chart + Sequential adj.	17.74 (0.29)	40.46 (0.36)	62.34 (0.34)	74.15 (0.31)	**79.79** (0.28)
	EWMA controller ($\lambda = 0.1$)	24.68 (0.26)	42.50 (0.31)	58.74 (0.30)	67.03 (0.28)	70.37 (0.27)
	EWMA controller ($\lambda = 0.2$)	**25.74** (0.29)	**45.50** (0.35)	63.80 (0.33)	73.24 (0.30)	77.23 (0.29)
	EWMA controller ($\lambda = 0.3$)	24.41 (0.31)	45.22 (0.37)	**64.66** (0.35)	**74.76** (0.32)	79.12 (0.30)
p=0.035	CUSUM chart + Single adj.	5.16 (0.32)	22.58 (0.38)	45.93 (0.40)	58.74 (0.39)	64.74 (0.39)
	CUSUM chart + Sequential adj.	12.93 (0.26)	28.75 (0.34)	49.89 (0.37)	61.45 (0.37)	**66.58** (0.37)
	EWMA controller ($\lambda = 0.1$)	**18.92** (0.25)	31.42 (0.31)	47.68 (0.33)	55.62 (0.33)	58.28 (0.34)
	EWMA controller ($\lambda = 0.2$)	18.88 (0.28)	**33.04** (0.35)	51.54 (0.37)	60.79 (0.37)	63.95 (0.38)
	EWMA controller ($\lambda = 0.3$)	16.91 (0.30)	31.98 (0.37)	**51.74** (0.39)	**61.73** (0.39)	65.19 (0.40)
p=0.02	CUSUM chart + Single adj.	0.58 (0.28)	10.70 (0.34)	27.50 (0.40)	38.84 (0.42)	44.11 (0.45)
	CUSUM chart + Sequential adj.	7.17 (0.22)	16.19 (0.29)	31.82 (0.37)	**41.88** (0.40)	**46.74** (0.43)
	EWMA controller ($\lambda = 0.1$)	**11.44** (0.22)	**18.85** (0.28)	31.07 (0.34)	37.96 (0.37)	40.63 (0.38)
	EWMA controller ($\lambda = 0.2$)	10.11 (0.25)	18.55 (0.32)	**32.70** (0.38)	40.80 (0.41)	44.09 (0.43)
	EWMA controller ($\lambda = 0.3$)	7.38 (0.27)	16.41 (0.34)	31.62 (0.41)	40.44 (0.44)	44.08 (0.46)
p=0.005	CUSUM chart + Single adj.	-4.21 (0.22)	-2.22 (0.24)	2.84 (0.29)	6.97 (0.34)	10.43 (0.37)
	CUSUM chart + Sequential adj.	0.34 (0.14)	2.33 (0.17)	7.05 (0.24)	**10.73** (0.30)	**13.69** (0.33)
	EWMA controller ($\lambda = 0.1$)	**1.51** (0.14)	**3.32** (0.17)	**7.08** (0.23)	9.62 (0.27)	11.31 (0.29)
	EWMA controller ($\lambda = 0.2$)	-1.52 (0.16)	0.61 (0.19)	5.11 (0.26)	8.09 (0.31)	10.12 (0.34)
	EWMA controller ($\lambda = 0.3$)	-5.25 (0.18)	-2.94 (0.21)	2.00 (0.28)	5.24 (0.33)	7.46 (0.36)

Table 4.9. Performance of integrated SPC/EPC adjustment schemes and an EWMA adjustment scheme on a MA(1) process with $\theta = -0.2$ when varying the probability of a shift.

% improvement on AISD		Mean of shift size				
		0	1σ	2σ	3σ	4σ
p=0.05	CUSUM chart +	10.63	35.02	60.85	72.47	78.66
	Single adj.	(0.35)	(0.40)	(0.36)	(0.33)	(0.31)
	CUSUM chart +	18.96	40.78	63.95	**74.40**	**79.90**
	Sequential adj.	(0.30)	(0.36)	(0.34)	(0.31)	(0.29)
	EWMA controller	24.51	41.69	59.31	66.34	69.73
	($\lambda = 0.1$)	(0.26)	(0.32)	(0.30)	(0.29)	(0.28)
	EWMA controller	**24.66**	**44.03**	64.10	72.33	76.33
	($\lambda = 0.2$)	(0.30)	(0.35)	(0.34)	(0.31)	(0.30)
	EWMA controller	22.26	42.99	**64.58**	73.57	77.94
	($\lambda = 0.3$)	(0.32)	(0.38)	(0.36)	(0.33)	(0.32)
p=0.035	CUSUM chart +	5.45	24.11	47.45	60.14	66.61
	Single adj.	(0.33)	(0.39)	(0.40)	(0.39)	(0.39)
	CUSUM chart +	13.20	30.14	**51.18**	**62.56**	**68.41**
	Sequential adj.	(0.27)	(0.35)	(0.37)	(0.37)	(0.37)
	EWMA controller	**18.20**	31.79	47.65	55.54	59.00
	($\lambda = 0.1$)	(0.25)	(0.32)	(0.34)	(0.34)	(0.34)
	EWMA controller	17.11	**32.59**	50.97	60.22	64.49
	($\lambda = 0.2$)	(0.28)	(0.36)	(0.38)	(0.38)	(0.38)
	EWMA controller	13.91	30.55	50.52	60.63	65.40
	($\lambda = 0.3$)	(0.31)	(0.39)	(0.41)	(0.41)	(0.41)
p=0.02	CUSUM chart +	0.46	11.10	27.55	39.59	44.91
	Single adj.	(0.29)	(0.35)	(0.41)	(0.43)	(0.46)
	CUSUM chart +	7.05	16.86	**31.95**	**42.58**	**47.54**
	Sequential adj.	(0.23)	(0.31)	(0.38)	(0.41)	(0.43)
	EWMA controller	**10.42**	**18.59**	30.20	37.12	40.30
	($\lambda = 0.1$)	(0.22)	(0.29)	(0.35)	(0.37)	(0.39)
	EWMA controller	7.86	17.31	30.88	39.22	43.13
	($\lambda = 0.2$)	(0.25)	(0.33)	(0.40)	(0.43)	(0.44)
	EWMA controller	3.73	13.97	28.77	38.01	42.38
	($\lambda = 0.3$)	(0.28)	(0.36)	(0.43)	(0.46)	(0.48)
p=0.005	CUSUM chart +	-4.61	-1.40	3.11	7.57	9.97
	Single adj.	(0.22)	(0.25)	(0.30)	(0.35)	(0.38)
	CUSUM chart +	-0.28	**2.64**	**6.94**	**11.11**	**13.31**
	Sequential adj.	(0.15)	(0.20)	(0.26)	(0.31)	(0.34)
	EWMA controller	**-0.06**	2.63	5.85	8.77	9.97
	($\lambda = 0.1$)	(0.13)	(0.18)	(0.23)	(0.27)	(0.29)
	EWMA controller	-4.58	-1.37	2.47	6.03	7.52
	($\lambda = 0.2$)	(0.16)	(0.21)	(0.27)	(0.32)	(0.34)
	EWMA controller	-9.98	-6.43	-2.19	1.77	3.47
	($\lambda = 0.3$)	(0.17)	(0.23)	(0.30)	(0.35)	(0.38)

Table 4.10. Performance of integrated SPC/EPC adjustment schemes and an EWMA adjustment scheme on a MA(1) process with $\theta = -0.05$ when varying the probability of a shift.

In this study, it was also found that EWMA control schemes with proper control parameters can outperform other schemes for a process with frequent mean shifts, especially when the process is autocorrelated. But since this control scheme requires no process monitoring, it does not provides the opportunity for quality engineers to distinguish process mean shifts from process autocorrelations and to understand and correct the root cause of the shifts.

<div style="text-align:center">

Chapter 5

Sequential Adjustment Methods for Asymmetric Off-target Quality Cost Functions

</div>

In previous chapters, sequential process adjustment rules were evaluated by the criterion AISD, i.e., the accumulated off-target costs, assumed quadratic. Such criterion is similar to Taguchi's quadratic quality loss function (Taguchi [98]). This and the next chapters present some alternative process adjustment methods when other manufacturing cost functions are used directly as the performance criteria of the adjustment rules instead of the AISD.

In many industrial cases, some non-quadratic quality cost functions are prevalent. For example, 0-1 type quality costs are often used for modelling the quality of conforming and nonconforming products. In this chapter, a more general quality cost structure based on asymmetric off-target costs is considered. Under this cost structure, higher-than-target observations imply a different cost than lower-than-target observations. This is especially relevant when material is removing during a manufacturing process, such as when drilling a hole, where an undersized hole can be reworked but an oversized hole may result in scrapping the whole part.

5.1 Asymmetric Cost Models

In this section, the machine setup error problem is once again studied; however, the process off-target cost is assumed to be an asymmetric function around the target

value as opposed to the previously considered quadratic cost function where the costs of over target and under target are symmetric. Suppose the quality characteristic y of each machined part is measured with reference to a nominal value, which is assumed, without loss of generality, to be equal to 0. After an erroneous start-up, the process is assumed to be off-target by d units, but this value cannot be observed directly due to the inherent production variability and the error of measurement. After the quality characteristic is measured a control parameter x can be set, which is assumed to have an immediate effect on the process output. Therefore, the process is the same as Equation (2.5) and it is repeated here

$$y_t = d + x_{t-1} + \varepsilon_t \tag{5.1}$$

where $t = 1, ..., N$ denotes a discrete time index or part number, and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ are normally distributed i.i.d. sequence.

To evaluate the costs associated with the control procedure, two cost models often adopted in industrial practice will be considered. In the first case, costs are assumed to arise only when the part processed is non-conforming, i.e., when the quality characteristic is out of the specification limits. In particular, it will be assumed that nonconforming to the lower or the upper specification limit could lead to different costs. For example, consider the case of a quality characteristic related with a dimension obtained after a finishing operation. In such operation, the costs associated with oversized and undersized items, which are mainly determined by either scrapping or re-working, are almost always different.

Fig. 5.1.   The asymmetric constant cost function with different costs when the quality characteristic is below the LSL or above the USL.

Therefore, two constants, c_1^c and c_2^c, are used to represent the costs associated with the violation of the LSL and USL, respectively. The superscript c indicates the *constant* cost model, thus we have

$$C_n^c = \begin{cases} c_1^c & \text{if } y_t < LSL \\ 0 & \text{if } LSL \leq y_t \leq USL \\ c_2^c & \text{if } y_t > USL \end{cases} \qquad (5.2)$$

(see Figure 5.1).

Another asymmetric cost model of interest is based on a piecewise quadratic cost function. In this case, the cost function can be more properly considered as a penalty function, in which the loss is assumed to be proportional to the square of the distance

of the quality characteristic from its nominal value. The asymmetry in the cost function is modeled through two constants, c_1^q and c_2^q, where the superscript q indicates the *quadratic* cost model, given by

$$C_n^q = \begin{cases} c_1^q Y_n^2 & \text{if } y_t < 0 \\ c_2^q Y_n^2 & \text{if } y_t \geq 0 \ . \end{cases} \tag{5.3}$$

The value of the constants c_1^q and c_2^q can be computed with reference to the Specification Limits as suggested by Wu and Tang [115]. The distance between the nominal value and the LSL or USL is denoted by Δ, and the cost corresponding to a quality characteristic equal to LSL or USL is L_1 or L_2, respectively. The constants, c_1^q and c_2^q, are given by:

$$c_1^q = \frac{L_1}{\Delta^2} \quad \text{and} \quad c_2^q = \frac{L_2}{\Delta^2} \tag{5.4}$$

(see Figure 5.2). The correspondence between the coefficients adopted with the constant and the quadratic cost models can be found from (5.4) by letting $L_1 = c_1^c$ and $L_2 = c_2^c$. The traditional symmetric cost models are therefore special cases of the above models, i.e., $c_1^c = c_2^c$ and $c_1^q = c_1^q$.

Since most of the recently developed devices for on-line inspection and measurement can transmit the data acquired to the controller of the machine, the assumption of an automatic feedback procedure is realistic. In this scenario, the cost of the adjustments can be neglected and therefore has not been considered in the following analysis.

The asymmetry in the cost function implies two issues that have to be considered in designing the adjustment rule. The first is related to the long-term or steady-state

Fig. 5.2. The asymmetric quadratic cost function with different costs when the quality characteristic is below the LSL or above the USL.

target $T^{\bullet}$ that has to be entered on the machine at start-up, where the superscript $\bullet$ is replaced by either c or q to indicate either a constant or a quadratic cost function. The problem of determining this value, referred to in the literature as the *optimum target point*, has been addressed for asymmetric cost functions in manufacturing by Ladany [58] and Wu and Tang [115].

The second issue is related to the way in which, starting from an initial offset, the quality characteristic should converge to the target as determined by the adjustment procedure. Both of these issues are considered in the remainder of this chapter. In particular, the steady-state target $T^{\bullet}$ will be derived by minimizing the long term expected costs, and the adjustment rule will be determined by considering all the costs associated with the transient period, evaluating the Average Integrated Expected Cost

(AIEC) performance index:

$$AIEC^\bullet = \frac{1}{N} \sum_{n=1}^{N} E(C_n^\bullet) \qquad (5.5)$$

where $E(C_n^\bullet)$ indicates the expected value of the costs at the nth step of the adjustment procedure.

5.2 A Biased Feedback Adjustment Rule

Since a control variable is available for removing a possible start-up error of a process, it is necessary to design a feedback adjustment rule to manipulate this variable. A common feedback linear adjustment rule is one of the form:

$$x_t = x_{t-1} - K_t(y_t - T^\bullet) \ . \qquad (5.6)$$

That is, the adjustments $x_t - x_{t-1}$ are proportional to the latest measured deviation of the quality characteristic y_t from the steady-state target $T^\bullet$. In Chapter 3, it was shown that, depending on the selection of the sequence $\{K_n\}$, this sequential adjustment results in Grubbs' harmonic and extended rules [44], which in turn is a direct application of Robbins and Monro's stochastic approximation techniques [79], the EWMA or integral controller, and an approach based on Recursive Least Squares. The performance of all the rules mentioned above was studied with respect to symmetric cost functions only.

Since the asymmetry in the cost model induces different losses depending on the side from which the quality characteristic approaches the steady-state target, the

performance of the linear adjustment rule could be enhanced by introducing a bias term in (5.6). Anbar [5] proposed a biased stochastic approximation procedure, further studied by Krasulina [57], for the problem of one-sided convergence. In this model, a bias term b_t is introduced into the adjustment rule, i.e.,

$$x_t = x_{t-1} - K_n(y_t - T^\bullet + b_t) \ . \tag{5.7}$$

Using the law of the repeated logarithm, Anbar demonstrated the convergence of Y_n as $n \to \infty$ when b_n converges to zero in $n^{\frac{1}{2}}(log(log\ n))^{-\frac{1}{2}}$.

Equation (5.7) is the adjustment rule we will consider in what follows; however, the conditions of the process variables outlined in Anbar [5] do not give insight on the selection of the sequence $\{b_t\}$ with reference to a specific asymmetric cost function. The adjustment procedure proposed in this paper is instead oriented to derive a sequence of bias coefficients $\{b_t\}$ that minimize the costs incurred during the transient phase of convergence of the quality characteristic to its steady-state target. In order to preserve its easiness of use, the bias sequence $\{b_t\}$ should be able to be computed off-line even when the process measurements are not available. This condition assures the control rule to be applicable to any manufacturing process, independently from the time units characterizing its dynamics.

By recursively substituting (5.7) in (5.1), the general expression of the quality characteristic at the nth step of the procedure is given by:

$$y_t = \prod_{i=1}^{t-1}(1-K_i)d - \sum_{i=1}^{t-1}\left[K_i(\varepsilon_i + b_i)\prod_{j=i+1}^{t-1}(1-K_j)\right] + T^\bullet\sum_{i=1}^{t-1}\left[K_i\prod_{j=i+1}^{t-1}(1-K_j)\right] + \varepsilon_t$$

(5.8)

where:

$$\prod_{j=t}^{t-1}(1-K_j) = 1 \ .$$

Since process errors are normally distributed, the quality characteristic y_t at each step of the procedure is also normally distributed, i.e., $y_t \sim N(\mu_t, \sigma_t^2)$, with mean and variance equal to:

$$\mu_t = \prod_{i=1}^{t-1}(1-K_i)d - \sum_{i=1}^{t-1}\left[K_i b_i\prod_{j=i+1}^{t-1}(1-K_j)\right] + T^\bullet\sum_{i=1}^{t-1}\left[K_i\prod_{j=i+1}^{n-1}(1-K_j)\right]$$

(5.9)

$$\sigma_t^2 = \sigma_\varepsilon^2\left[1 + \sum_{i=1}^{t-1}K_i^2\prod_{j=i+1}^{t-1}(1-K_j)^2\right] \ .$$

(5.10)

As it can be observed, the sequence of bias terms $\{b_i\}$ affects only the mean value μ_t of the quality characteristic. Therefore, for a given selection of $\{K_i\}$, the bias terms $\{b_i\}$ can be determined by equating the right hand side of expression (5.9) to the *optimal mean* at the nth step $m_t^\bullet$, i.e., the nth component of the vector $\mathbf{m}^\bullet = \{m_n^\bullet, n = 1, ..., N\}$ that minimizes the $AIEC^\bullet$ given by (5.5). The computation of $\mathbf{m}^\bullet$ will be addressed in the next section.

Although the approaches in Anbar [5] and Krasulina [57] utilize the harmonic sequence for $\{K_t\}$, i.e., $K_t = 1, 1/2, 1/3, ...$, it is in principle possible to consider a

different sequence, while maintaining the form of the controller given by (5.7). For example, besides considering the harmonic sequence (Grubbs' approach), a constant sequence (the EWMA or integral control approach) can be considered instead.

In the case when $K_i = 1/i$, $i = 1, 2, ..., t - 1$ (a harmonic series), the value of the mean and the variance of the quality characteristic at each step are given by:

$$\mu_t = T^\bullet - \frac{1}{t-1} \sum_{i=1}^{t-1} b_i \tag{5.11}$$

$$\sigma_t^2 = \sigma_\varepsilon^2 \left(\frac{t}{t-1} \right) . \tag{5.12}$$

If K_i is instead set equal to a constant λ, as in the EWMA approach, the resulting mean and variance are:

$$\mu_t = T^\bullet + (1-\lambda)^{t-1}(d - T^\bullet) - \lambda(1-\lambda)^{t-1} \sum_{i=1}^{t-1} \frac{b_i}{(1-\lambda)^i} \tag{5.13}$$

$$\sigma_t^2 = \sigma_\varepsilon^2 \left[\frac{2 - \lambda(1-\lambda)^{2(t-1)}}{2-\lambda} \right] . \tag{5.14}$$

It is noticed that in Equation (5.11) the value of μ_t does not depend on the initial unknown offset d, thus an off-line computation of b_n is possible. For the biased EWMA approach, μ_t is a function of the unknown offset d, so the sequence of biased coefficients $\{b_t\}$ can not be computed off-line. Therefore, we will only consider the biased harmonic adjustment rule in what follows. From Equation (5.11), the general expression for b_t can be obtained by equating the mean of the response to the optimal mean at the tth and the $t + 1^{th}$ steps, i.e.,

$$-\frac{1}{t-1}\sum_{i=1}^{t-1}b_i + T^\bullet = m_t^\bullet$$

and

$$-\frac{1}{t}(\sum_{i=1}^{t-1}b_i + b_t) + T^\bullet = m_{t+1}^\bullet \ ,$$

from where the general expression for the bias term b_t is given by

$$b_t = t(T^\bullet - m_{t+1}^\bullet) - (t-1)(T^\bullet - m_t^\bullet) \ . \tag{5.15}$$

5.3   Formulae for the Optimal Target and the Sequence of Bias Term

To complete the adjustment rule, the optimal steady-state target $T^\bullet$ and the sequence of bias terms $\{b_t\}$ have to be specified. As previously mentioned, the first value represents the optimal mean $m_t^\bullet$ as $t \to \infty$, while the sequence of bias terms can be computed using (5.15), once the vector of optimal means $\mathbf{m}^\bullet$ is known. To compute this vector, the minimization problem that has to be solved can be stated as:

$$\min_{\boldsymbol{\mu}} \ AIEC^\bullet \tag{5.16}$$

where $\boldsymbol{\mu} = \{\mu_t, t = 1, ..., N\}$ is the vector composed by the means of the response at each step of the procedure, and $AIEC^\bullet$ is the performance index given by equation (5.5). As showed in Appendix C, when the linear control rule (5.7) is in use, the optimization in

(5.16) is equivalent to the following set of minimization problems:

$$\min_{\mu_t} \; E(C_t^{\bullet}) \; , \; t = 1, 2, ..., N \; . \tag{5.17}$$

Problems in (5.17) will be solved for the two types of cost functions studied. Consider first the constant asymmetric cost function. The expected cost at time t is given by:

$$
\begin{aligned}
E(C_t^c) \; &= \; c_1^c \int_{-\infty}^{LSL} f_N(y_t; \mu_t, \sigma_t^2) dy_t + c_2^c \int_{USL}^{\infty} f_N(y_t; \mu_t, \sigma_t^2) dy_t \\
&= \; c_1^c \Phi\left(\frac{LSL - \mu_t}{\sigma_t}\right) + c_2^c \left[1 - \Phi\left(\frac{USL - \mu_t}{\sigma_t}\right)\right] \tag{5.18}
\end{aligned}
$$

where $f_N(\cdot)$ is the normal density function and $\Phi(\cdot)$ is the standard normal distribution function. The minimum of this function with respect to μ_t can be derived by computing the first and second order derivatives of $E(C_t^c)$. As reported in Appendix D, the optimal mean m_t^c, obtained by equating the first derivative of $E(C_t^c)$ to zero, is given by

$$m_t^c = \frac{\sigma_t^2 \ln(c_1^c / c_2^c)}{(USL - LSL)} + \frac{1}{2}(USL + LSL). \tag{5.19}$$

As a special case, when the cost function is symmetric, i.e. $c_1^c = c_2^c$, the result obtained is $m_t^c = (USL + LSL)/2$, which is equal to 0 when USL and LSL are symmetric around the nominal value. Since the second derivative with respect to μ_t (see Appendix

D) is always greater than zero when the condition $LSL < \mu_t < USL$ is satisfied, the value of m_t^c obtained is the minimum for the expected cost $E(C_t^C)$.

The steady-state target T^C can be derived as a particular case of the general expression (5.19) by considering the limit, as $t \to \infty$, of σ_t^2 given by (5.12). Since this limit is equal to σ_ε^2, we get

$$T^C = \frac{\sigma_\varepsilon^2 \ln(c_1^c/c_2^c)}{(USL - LSL)} + \frac{1}{2}(USL + LSL) . \tag{5.20}$$

Substituting (5.20) and (5.19) into the expression of the bias term, given by (5.15), the values of the bias terms b_t for the asymmetric constant cost function can be directly computed. In this case, all b_t's except the first one equal to zero, i.e.,

$$b_t = \begin{cases} -\dfrac{\ln(c_1^c/c_2^c)\sigma_\varepsilon^2}{(USL-LSL)} & \text{if t=1} \\[2mm] 0 & \text{if t=2,...,N .} \end{cases} \tag{5.21}$$

Although the feedback adjustment procedure has a non-zero bias b_t only at the first step, b_1 affects the following adjustments through the x_{t-1} term in the expression of the controller (5.7).

Consider now the quadratic asymmetric cost function. The expected cost at the tth step of the procedure is given by

$$E(C_t^q) = c_1^q \int_{-\infty}^{0} y_t^2 f_N(t_t; \mu_t, \sigma_t^2) dy + c_2^q \int_{0}^{\infty} y_t^2 f_N(y_t; \mu_t, \sigma_t^2) dy_t.$$

By solving the two integrals (as reported in Appendix E), the following expression for the expected value of the cost is obtained:

$$E(C_t^q) = c_2^q(\mu_t^2 + \sigma_t^2) + (c_2^q - c_1^q)\left[\sigma_t\mu_t\phi\left(\frac{\mu_t}{\sigma_t}\right) - (\mu_t^2 + \sigma_t^2)\Phi\left(-\frac{\mu_t}{\sigma_t}\right)\right] . \qquad (5.22)$$

Computing the first derivative with respect to μ_t and equating it to zero, the optimal mean m_t^q is determined by the following equation:

$$2c_2^q m_t^q + 2(c_2^q - c_1^q)\left[\sigma_t\phi\left(\frac{m_t^q}{\sigma_t}\right) - m_t^q\Phi\left(-\frac{m_t^q}{\sigma_t}\right)\right] = 0 \qquad (5.23)$$

where $\phi(\cdot)$ is the standard normal density function and $\Phi(\cdot)$ is the standard normal distribution function. Although there is no closed form expression for m_t^q, it can be computed numerically off-line, since all the quantities in expression (5.23) do not depend on the actual observations of the quality characteristic. Similarly as the constant cost function, if the quadratic cost function is symmetric, i.e., $c_1^q = c_2^q$, the optimal mean m_t^q is zero for $t = 1, 2, ..., N$.

The second derivative of $E(C_t^q)$ with respect to μ_t is always positive (see Appendix E), so m_t^q given by equation (5.23) determines a minimum of the expected cost. Again, the steady-state target T^q can be computed as a special case by considering $lim_{t\to\infty}\sigma_t = \sigma_\varepsilon$, in equation (5.23), so T^q is the solution of

$$2c_2^q T^q + 2(c_2^q - c_1^q)\left[\sigma_\varepsilon\phi\left(\frac{T^q}{\sigma_\varepsilon}\right) - T^q\Phi\left(-\frac{T^q}{\sigma_\varepsilon}\right)\right] = 0 . \qquad (5.24)$$

Therefore, in the case of the quadratic cost model, the feedback adjustment rule can be obtained by evaluating numerically the optimal means m^q_t that satisfy equation (5.23) for $t = 1, 2, ..., N$, and the optimal target T^q can be obtained from equation (5.24). Substituting these values in equation (5.15), we obtain the sequence of bias coefficients $\{b_t\}$.

In summary, the biased linear adjustment procedure for constant and quadratic cost functions are as follows:

Solution to the Asymmetric Constant Cost Model

Given: c^c_1, c^c_2, USL, LSL, σ_ε, N.

1. Compute the steady-state target T^C using (5.20);

2. Compute the bias coefficient b_1 using (5.21);

3. Adjust the control variable on-line according to the following equation:

$$
x_t = \begin{cases} -[y_1 - T^C + b_1] & \text{if } t = 1 \\ x_{t-1} - \frac{1}{t}[y_t - T^C] & \text{if } t = 2, ..., N \ . \end{cases}
\tag{5.25}
$$

Solution to the Asymmetric Quadratic Cost Model

Given: c^q_1, c^q_2, σ_ε, N.

1. Compute the steady-state target T^q by solving numerically equation (5.24);

2. Find the sequence of bias terms $\{b_t\}$ for $t = 1, ..., N$:

   • Compute the optimal mean m^q_t by solving numerically equation (5.23) where $\sigma_t = \sigma_\varepsilon \sqrt{\frac{t}{t-1}}$;

- Substitute T^q_t and m^q_t into (5.15) to obtain b_t;

3. Adopt the biased linear adjustment rule for on-line process adjustment:

$$x_t = x_{t-1} - \frac{1}{t}(y_t - T^q + b_t) \ .$$

5.4 An Application to a Real Machining Process

In this section, the biased linear adjustment procedure for start-up errors will be applied to a real machining problem. The performance of the biased rule will be compared with that of Grubbs' rule and with the EWMA (integral) controller. The latter two procedures follow the adjustment rules of the form (5.6) where K_t is equal to $1/t$ for Grubbs' rule and equal to a constant λ for the EWMA controller.

A hole-finishing operation is performed on a pre-existing hole in a raw aluminum part made by pressure casting. The specification limits on the final hole diameter are at $57.000 \pm 0.030mm$. After the execution of the operation, the diameter of the hole (D) is measured in an automatic inspection station constituted by a probe that acquires the diameter while the workpiece rotates 360 degrees around the axis of the hole. The mean diameter is computed and recorded. Due to the materials machined and the tools used (polycrystalline inserts), the tool wear can be neglected and no trend is present in the data collected. We let the quality characteristic of this process be the difference between measurement D_t and the nominal value of the hole diameter, i.e., $y_t = D_t - 57000$, in microns. The standard deviation of the process σ_ε is estimated through $\sqrt{MSE}$ (the square root of Mean Square Error), which is obtained from an ANOVA analysis of

historical process data after start-ups and which is equal to 10 microns (thus the process capability ratio, PCR, is 1). From the ANOVA analysis, it is also found that after setup or maintenance operations the process mean often exhibits a shift or offset, which is, on average, in the order of $3\sigma_\varepsilon$. In this case, parts are produced in lots of size 15.

The costs related to non-conforming items are different depending on whether the diameter obtained is below the lower or above the upper specification limit. Indeed, when the hole diameter is less than the LSL, an additional machining operation can correct the defect by opportunely selecting the depth of cut. On the other hand, when the diameter obtained is greater than USL, the part has to be scrapped, since there is no possibility to recover the nonconforming workpiece. The cost of an undersized hole, c_1^c, is determined by considering the additional repairing operation while the cost of an oversized hole, c_2^c, is equal to the margin lost minus the value of the scrap. In this case, the asymmetric ratio r $(r = c_2^c/c_1^c)$ is 6.5. If a quadratic cost model is assumed, by adopting the relation outlined in expressions (5.4), the same ratio between c_2^q and c_1^q can be obtained.

As showed in Figure 5.3, the controllable variable x_t is the radial position of the tool. In fact, by opportunely selecting this variable, the depth of cut can be changed, thus modifying the dimension of the diameter obtained. Furthermore, the adoption of a parametric part program can in principle allow for an automatic adjustment procedure: once a diameter is measured, the value of the controllable variable can be determined and transmitted to the control unit of the machining center that will process the next part accordingly. In a real-life application of an adjustment procedure, the resolution of the machine in setting the tool position should be considered in order to derive the

Fig. 5.3. The hole finishing operation.

approximation of the adjustment size. In this case a precision in the order of microns determines that we round the adjustment to zero decimal places.

Assuming the asymmetric constant cost function model, the expected value of the cost reported in (5.18) can be rewritten as a function of r, thus a scaled form of the expected costs at each step of the adjustment procedure is obtained as:

$$\frac{E(C^c_t)}{c^c_1} = \Phi\left(\frac{LSL - \mu_t}{\sigma_t}\right) + r\left[1 - \Phi\left(\frac{USL - \mu_t}{\sigma_t}\right)\right] \qquad (5.26)$$

Therefore, the performance comparisons among the different control rules will be evaluated using as performance index the Scaled Average Integrated Expected Cost (SAIEC), defined as:

$$SAIEC^c = \frac{1}{N}\sum_{t=2}^{N} \frac{E(C^c_t)}{c^c_1} \qquad (5.27)$$

where the index in the summation starts from 2, since the quality characteristic of the first part machined does not depend on the adjustment procedure. To define the biased adjustment rule, the steady-state target T^C and the biased coefficients b_t need to be computed and rounded to the closest integer. Using equation (5.20), the steady-state target results $T^C = -3$ micron. Therefore, according to (5.21), the biased coefficients are given by:

$$b_t = \begin{cases} 3 & t = 1 \\ 0 & t = 2, ..., 15 \end{cases} \qquad (5.28)$$

Figure 5.4 reports the plots of the expected value of the quality characteristic obtained with both the biased and Grubbs' procedures. In particular, the piecewise behavior of the biased mean converging to the target value is due to the approximation (rounding) adopted to consider the precision of the machine in setting the tool position. In fact, changing the precision of the approximation to the second decimal place, the mean at each step of the biased procedure is represented by the dotted line in Figure 5.4. As it can be observed, the adoption of the biased procedure induces a convergence of the mean to the steady-state target value T^C from the side of lower nonconforming costs.

The savings in cost obtained by the biased rule are shown in Figure 5.5, where the percentage difference in $SAIEC^C$ determined by the biased and Grubbs' procedures is reported as a function of the items processed (computed from data in Table 5.1).

A further comparison between the biased and different EWMA control rules, characterized by values of the parameter λ ranging from 0.2 to 0.8 has been carried out. Since the performance of an EWMA controller depends on the initial offset d, a constant

Fig. 5.4. Trajectory of the optimal mean of the quality characteristic (m_t^c) using Grubbs' and the biased procedures (considering 0 and 2 decimal places) under the constant cost model ($r = 6.5$ and $N = 15$).

n	EWMA0.2	EWMA0.4	EWMA0.6	EWMA0.8	Grubbs	Biased
2	1.227	0.532	0.234	0.119	0.092	0.080
3	0.898	0.341	0.149	0.088	0.064	0.057
4	0.689	0.245	0.112	0.076	0.051	0.046
5	0.553	0.192	0.092	0.070	0.043	0.039
6	0.461	0.158	0.080	0.067	0.037	0.034
7	0.393	0.135	0.072	0.065	0.033	0.031
8	0.342	0.118	0.066	0.063	0.030	0.028
9	0.303	0.106	0.062	0.062	0.028	0.026
10	0.271	0.096	0.058	0.061	0.026	0.024
11	0.246	0.089	0.055	0.060	0.024	0.023
12	0.225	0.082	0.053	0.059	0.023	0.022
13	0.207	0.077	0.051	0.059	0.022	0.021
14	0.193	0.072	0.050	0.058	0.021	0.020
15	0.180	0.068	0.048	0.058	0.020	0.019

Table 5.1. The $SAIEC^C$ adopting different control rules ($r = 6.5$, $N = 15$ and $A = 3$).

Fig. 5.5. The percentage savings in $SAIEC^c$ ($\dfrac{SAIEC^c_G - SAIEC^c_B}{SAIEC^c_G} \times 100$) obtained by using the biased procedure compared to Grubbs' rule under the constant cost function model ($r = 6.5$ and $N = 15$).

$A = (d - T^{\bullet})/\sigma_{\varepsilon}$, i.e., the difference between d and the target value in standard deviation units is assumed equal to 3 according to the practical case we have discussed.

The Scaled Average Integrated Expected Costs $SAIEC^C$ obtained with the biased procedure and the EWMA controllers are reported in Table 5.1 and plotted in Figure 5.6. As it can be observed, the biased procedures has the smallest expected cost compared to all the EWMA controllers and the advantage reduces as λ increases, So a value $\lambda = 0.8$ was used in the next comparison. It should be pointed out that much smaller values of λ are recommended in the literature (Box and Luceño [15]), but for these values of λ the EWMA performs relatively worse.

The cost comparison between the EWMA controller with $\lambda = 0.8$ and the biased controller is given in Figure 5.7, where the percentage saving in $SAIEC^C$ induced by the biased procedure over the EWMA is plotted. It is interesting to find that the advantage induced by the biased procedure is even higher as the number of parts produced increases. The reason for this behavior lies on the long-term performance of the EWMA control rule. In fact, as t tends to infinity, the mean of the y_t regulated by the EWMA controller approaches zero, but the variance approaches to the value $2\sigma_{\varepsilon}^2/(2 - \lambda)$, which is greater than σ_{ε}^2. This inflation in variance has been discussed in Chapter 3.

For the quadratic cost function model, an analogous comparison was performed. In this case, the expected cost reported in equation (5.22) can be rewritten in scaled form by manipulating the expression as follows:

$$E(C_t^q) = c_1^q \sigma_t^2 \left\{ \frac{c_2^q}{c_1^q} \left(\frac{\mu_t^2}{\sigma_t^2} + 1 \right) + \left(\frac{c_2^q}{c_1^q} - 1 \right) \left[\frac{\mu_t}{\sigma_t} \phi \left(\frac{\mu_t}{\sigma_t} \right) - \left(\frac{\mu_t^2}{\sigma_t^2} + 1 \right) \Phi \left(-\frac{\mu_t}{\sigma_t} \right) \right] \right\} .$$

Fig. 5.6. Comparison of $SAIEC^{c}$'s determined by the EWMA controllers (with different values of λ) and the biased procedure under the constant cost function model ($r = 6.5$, $N = 15$ and $A = 3$).

Fig. 5.7. The percentage savings in $SAIEC^c$ $(\frac{SAIEC^c_{EWMA0.8} - SAIEC^c_B}{SAIEC^c_{EWMA0.8}} \times 100)$ obtained by using the biased procedure compared to the EWMA rule with $\lambda = 0.8$ under the constant cost function model ($r = 6.5$, $N = 15$ and $A = 3$).

Considering that the variance at each step of the adjustment procedure (5.12) is proportional to the variance of the error σ_ε^2, the expected cost at the tth step of the procedure is given by:

$$\frac{E(C_t^q)}{c_1^q} = s_t \sigma_\varepsilon^2 \left\{ r(\delta_t^2 + 1) + (r-1) \left[\delta_t \phi(\delta_t) - (\delta_t^2 + 1)\Phi(-\delta_t) \right] \right\} \ ,$$

where $s_t = \left[1 + \sum_{i=1}^{t-1} K_i^2 \prod_{j=i+1}^{t-1} (1 - K_j)^2 \right]$ represents the ratio between σ_t^2 and σ_ε^2 in equation (5.12), r denotes the ratio between c_2^q and c_1^q and δ_t the ratio between μ_t and σ_t.

As in the constant cost function case, the performance index considered is related to the Scaled Average Integrated Expected Cost defined as:

$$SAIEC^q = \frac{1}{N} \sum_{t=2}^{N} \frac{E(C_t^q)}{c_1^q} \tag{5.29}$$

Figure 5.8 reports the plot of the mean of the quality characteristic obtained with Grubbs' rule (in which the mean is constant and equal to the steady-state target value), and the biased rule (in which the mean is set to m_t^q and converges to the target value). Similarly as in the constant cost model case, the mean induced by the biased procedure is computed by considering the assumption on the control variable resolution (in Figure 5.8 the theoretical behavior of one-sided convergence of m_t^q is reported with a dotted line, which was obtained by rounding m_t^q to the second decimal place). The values of the biased coefficients b_t are also shown in Table 5.2. As it can be observed, when the precision of the machine is considered, the sequence $\{b_t\}$ adopted is basically the same as

t	b_t	b_t (2 decimal precision)
1	3	3.06
2	0	0.26
3	0	0.11
4	0	0.06
5	0	0.04
6	0	0.03
7	0	0.02
8	0	0.01
9	0	0.01
10	0	0.01
11	0	0.01
12	0	0.01
13	0	0.01
14	0	0.00
15	0	0.00

Table 5.2.   The bias coefficients b_n computed under the quadratic cost function rounding to the nearest integer or considering the second decimal place ($r = 6.5$ and $N = 15$).

obtained with the asymmetric constant cost model (5.28), but the computation of b_t in the constant cost model is much easier because of the closed form expressions. Therefore, this numerical result permits to outline an approximated way to compute the b_n that does not require the numerical solution of equation (5.23).

Data on the $SAIEC^q/\sigma_\varepsilon^2$ obtained with the Grubbs' rule, the biased rule and the EWMA controller are reported in Table 5.3. The percentage in savings from adopting the biased procedure instead of Grubbs' rule are reported in Figure 5.9. Figures 5.10 and 5.11 report respectively the $SAIEC^q/\sigma_\varepsilon^2$ obtained with the biased procedure and the EWMA controllers and the detail on the percentage savings obtained over the EWMA controller with $\lambda = 0.8$.

Fig. 5.8. Trajectory of the optimal mean of the quality characteristic (m_t^q) using Grubbs' and the biased procedures (considering 0 and 2 decimal places) under the quadratic cost model ($r = 6.5$ and $N = 15$).

t	EWMA0.2	EWMA0.4	EWMA0.6	EWMA0.8	Grubbs	Biased
2	25.461	14.947	8.674	5.625	4.818	4.539
3	20.721	11.046	6.545	4.851	4.158	3.970
4	17.300	8.761	5.560	4.544	3.798	3.657
5	14.829	7.452	5.028	4.391	3.566	3.451
6	13.035	6.579	4.685	4.298	3.402	3.305
7	11.635	5.974	4.457	4.237	3.280	3.195
8	10.505	5.543	4.294	4.193	3.184	3.109
9	9.614	5.219	4.172	4.160	3.107	3.039
10	8.889	4.958	4.077	4.135	3.043	2.982
11	8.287	4.749	4.001	4.114	2.990	2.933
12	7.794	4.579	3.939	4.098	2.944	2.892
13	7.370	4.436	3.887	4.084	2.904	2.856
14	7.011	4.316	3.843	4.072	2.869	2.824
15	6.696	4.213	3.805	4.062	2.838	2.796

Table 5.3. The $SAIEC^q/\sigma_\varepsilon^2$ adopting different control rules ($r = 6.5$, $N = 15$ and $A = 3$)

Fig. 5.9. The percentage savings in $SAIEC^q$ ($\dfrac{SAIEC^q_G - SAIEC^q_B}{SAIEC^q_G} \times 100$) obtained by using the biased procedure compared to Grubbs' rule under the quadratic cost function model ($r = 6.5$ and $N = 15$).

Fig. 5.10. Comparison of $SAIEC^q/\sigma_\varepsilon^2$'s determined by the EWMA controllers (with different values of λ) and the biased procedure under the quadratic cost function model ($r = 6.5$, $N = 15$ and $A = 3$).
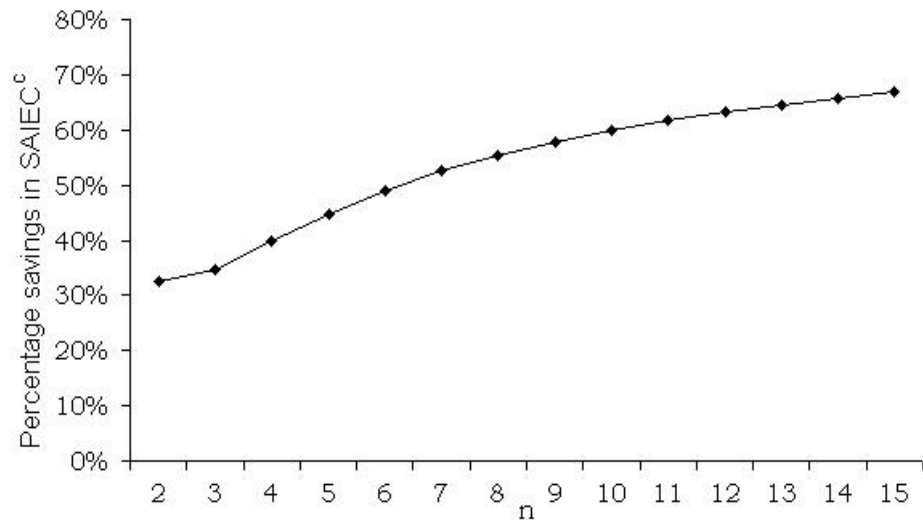
Fig. 5.11. The percentage savings in $SAIEC^q$ ($\dfrac{SAIEC^q_{EWMA0.8} - SAIEC^q_B}{SAIEC^q_{EWMA0.8}} \times 100$) obtained by using the biased procedure compared to the EWMA rule with $\lambda = 0.8$ under the quadratic cost function model ($r = 6.5$, $N = 15$ and $A = 3$).

The comparisons between the biased control rule, Grubbs' procedure and the EWMA indicate the same conclusions as for the constant asymmetric cost model, but quantitatively, the magnitude of the percentage advantage obtained with the biased rule is greater when adopting the constant cost model.

5.4.1   Sensitivity analysis

A numerical comparison of the performance obtained with the biased procedure, Grubbs' rule and the EWMA controllers was conducted to characterize situations in which the adoption of the feedback adjustment could be more profitable. The comparison has been carried out first for the biased procedure versus Grubbs' rule, since the performance in this case does not depend on the initial offset. The variables affecting the results in this case are the coefficient r, representing the asymmetry of the cost function, and N, the number of parts processed in each lot. The value of r was varied from 1 to 11 as in Ladany [58]. We point out that two real cases of asymmetric cost functions considered in Wu and Tang [115] and Moorhead and Wu [75] have r to be 4 and 6, respectively, and they are inside the range examined. The number of parts in the lot, N, was varied from 1 to 40.

Figures 5.12 and 5.13 present the savings in cost obtained with the biased procedure over Grubbs' rule for the constant and quadratic cost models, respectively. As it can be observed, the biased procedure has an advantage especially on the first parts produced (this suggests the adoption of the biased rule when parts are produced in small lots) and this advantage increases as the asymmetry in the function becomes more evident (i.e., as r increases).

Fig. 5.12. Sensitivity analysis: the percentage saving in $SAIEC^c$ ($\dfrac{SAIEC^c_G - SAIEC^c_B}{SAIEC^c_G} \times$

100) obtained by using the biased procedure compared to Grubbs' procedure under the constant cost function model, when the asymmetry ratio $r = c^c_2/c^c_1$ is varied.

Fig. 5.13. Sensitivity analysis: the percentage saving in $SAIEC^q$ ($\dfrac{SAIEC^q_G - SAIEC^q_B}{SAIEC^q_G} \times 100$) obtained by using the biased procedure compared to Grubbs' rule under the quadratic cost function model, when the asymmetry ratio $r = c^q_2 / c^q_1$ is varied.

Since the performance of the EWMA controllers depend on the initial offset d, standardized by the constant $A = (d - T^{\bullet})/\sigma_{\varepsilon}$, the comparison between the biased rule and the EWMA controller has been performed by considering A ranging from -4 to 4. Figures 5.14 and 5.15 report the difference in the Scaled Average Integrated Expected costs obtained with the EWMA and the biased controller, under the constant and the quadratic cost models, respectively. In particular, the difference is reported for the two extreme values of λ (i.e., $\lambda = 0.2$ and $\lambda = 0.8$) and the lot size (i.e., $N = 5$ and $N = 40$).

Depending on the initial offset, the advantage of using the biased procedure varies dramatically. Considering the case in which $\lambda = 0.2$, when A is greater than 1 the performance of the biased procedure dominates that of the EWMA controller, but the difference between the two procedures is almost negligible as A is close to zero. Furthermore, the advantage is asymmetric too. In particular, if A is positive, i.e. the offset d arises from the side in which non-conforming items are more expensive, the advantage of adopting the biased procedure is significantly greater, compared with the case in which the initial shift has the same magnitude but different sign. As the number of parts processed in the lot increases, the difference between the two procedures maintains the same behavior while reducing in magnitude (both approaches tend to reach their asymptotic performance, which are different only with respect to the variance σ_t^2 of the quality characteristic).

In the case when $\lambda = 0.8$ is used in the EWMA control rule, the advantage determined by the biased approach is reduced but is always greater than zero, regardless of the direction of the initial offset. Also in this case, the effect of A becomes even less significant when N increases.

Fig. 5.14. Sensitivity analysis: the difference in $SAIEC^c$ ($SAIEC^c_{EWMA} - SAIEC^c_B$) obtained under the constant cost model, when $r = c^c_2/c^c_1$, $A = (d - T^c)/\sigma_\varepsilon$ and N are varied. a. $\lambda = 0.2$ and $N = 5$; b. $\lambda = 0.8$ and $N = 5$; c. $\lambda = 0.2$ and $N = 40$; d. $\lambda = 0.8$ and $N = 40$.

Fig. 5.15. Sensitivity analysis: the difference in $SAIEC^q/\sigma_\varepsilon^2$ ($\dfrac{SAIEC^q_{EWMA} - SAIEC^q_B}{\sigma_\varepsilon^2}$) obtained under the quadratic cost model, when $r = c^q_2/c^q_1$, $A = (d - T^q)/\sigma_\varepsilon$ and N are varied. a. $\lambda = 0.2$ and $N = 5$; b. $\lambda = 0.8$ and $N = 5$; c. $\lambda = 0.2$ and $N = 40$; d. $\lambda = 0.8$ and $N = 40$.

5.5   Summary

Previous approaches to the setup adjustment problem considered only symmetric quadratic quality loss functions. This chapter presented a feedback adjustment rule that can be adopted when an asymmetric cost model can better represent the process quality losses entailed. Two asymmetric cost functions that are often encountered in manufacturing were considered. In the first case, the cost of a non-conforming item is assumed constant but changes depending whether the quality characteristic is below the lower or above the upper specification limit. In the second case, costs are supposed to be proportional to the square of the distance of the quality characteristic from the nominal value, but the proportional constant is allowed to change with the sign of this difference. Starting from the general form of a linear controller, a biased feedback adjustment rule was derived by minimizing the quality cost incurred during the transient phase in which the quality characteristic converges to its steady-state target. A numerical comparison of the cost incurred by the adjustment rule proposed and other rules discussed in the literature showed that the proposed procedure is effective, especially when the asymmetry in the cost function or the initial process offset are significant. Compared to Grubbs' harmonic rule, the proposed biased adjustment rule is recommended especially for manufacturing expensive parts which usually are produced in small lots.

Chapter 6

Process Adjustments with Significant Measurement and Adjustment Costs

As mentioned briefly in Chapter 4, the number of sequential adjustments should be determined based on the adjustment cost. In this chapter, the effects of the adjustment cost and measurement cost are systematically investigated. A new adjustment method, which apply sequential adjustments at nonconsecutive process runs, is proposed and scheduling methods to program the adjustments are discussed. A manufacturing cost function, which includes process off-target cost, measurement cost and adjustment cost, is used as the performance criterion for various adjustment rules.

6.1 Cost Model and Sample-Average Adjustment Procedure

The cost criterion used in previous discussions for the process adjustment of a machine setup error considers only an off-target quality cost, either with a symmetric quadratic function or other an asymmetric function. However, in quality control, measurement and adjustment costs usually cannot be ignored. In previous chapters, adjustments were suggested on every single run, but this is clearly not the most economic strategy when the measurement and adjustment costs are significant. In this chapter, we wish to minimize the total manufacturing cost which it is assumed consists of the following components:

- Expected off-target quality cost, C_q, which is the expectation of the sum of a quadratic function of y_t around its target, i.e., $C_q = \sum_{t=1}^{N} E[\Omega(y_t - T)^2]$. Here, a symmetric cost function is used and without loss of generality, the target, T, is assumed to be 0. The number, N, is the number of parts that need to be manufactured in the lot and Ω is the quadratic cost per unit.

- Adjustment cost,C_a, which is assumed to be fixed and independent of the magnitude of the adjustment, i.e., $C_a = M \times (\sum_{t=1}^{n} \delta(t))$, where $\delta(t)$ equals to 0 when no adjustment is scheduled and is 1 otherwise.

- Measurement cost, C_m, which is assumed to be proportional to the number of adjustments, i.e., $C_m = G \times m$, where m is the time of the last adjustment. Obviously, when the last adjustment has been executed and no more adjustments are needed till the end of production, measurements on the following runs are not necessary.

For the machine setup error problem, it is supposed that an error d can happen on the machine before the manufacturing process starts. Since there is a controllable factor x_t on the machine which can be adjusted to eliminate the effect of d, the essential question is how to estimate d based on the process observations $\{y_t\}$. Chapter 3 provides a Kalman Filter estimate of d_t, namely

$$\hat{d}_t = \hat{d}_{t-1} + \frac{1}{\sigma_\varepsilon^2/P_0 + t} y_t, \tag{6.1}$$

where $\hat{d}_0$ is an *a priori* estimate of d and P_0 is a measure of confidence on this initial estimator. The Kalman Filter estimator is the minimum mean square error linear estimator if the process variance is known (Duncan and Horn [35]), so the "controller" $x_t = -\hat{d}_t$ is optimal in that sense.

This adjustment procedure requires adjustments at every time period, an adjustment policy that may be undesirable if the costs of measurement and adjustment are significant. This implies that it is possible to design an adjustment schedule which skips some time periods between two successive adjustments and maintains the cost optimality of the whole procedure (Trietsch [100]). This idea has been illustrated in Chapter 1, that is, if there are no adjustments between time i and j, then $\bar{Y}_{ij}$, the average of $Y_{i+1}...Y_j$, is the unbiased MLE (maximum likelihood estimate) of $(d + X_i)$; therefore, it is intuitive to change the estimate of d to the following equation when the simple adjustment rule, $x_j = -\hat{d}_j$, is applied:

$$\hat{d}_j = \hat{d}_i + \frac{1}{\sigma_\varepsilon^2/P_0 + j} \sum_{t=i+1}^{j} y_t. \tag{6.2}$$

In the remaining of this section, we will show that the adjustment procedure based on equation (6.2) provides the same general expression for y_{j+1} as a function of the adjustments as given by the procedure based on equation (6.1). The performance of this adjustment rule will then be studied in subsequent sections.

Suppose y_i, the value of the quality characteristic at time i, is known. Then, from (6.1) we have that x_i, which is a function of y_i, is also known. We first use the procedure based on (6.1) to derive the function of y_{j+1}. From the process model and adjustment

function, we have

$$y_{i+1} = d + x_i + \varepsilon_{i+1},$$

$$x_{i+1} = x_i - K_{i+1} y_{i+1},$$

where $K_{i+1} = 1/(\sigma_\varepsilon^2/P_0 + i + 1)$. So,

$$y_{i+2} = d + x_{i+1} + \varepsilon_{i+2} = (1 - K_{i+1})(d + x_i) - K_{i+1}\varepsilon_{i+1} + \varepsilon_{i+2}.$$

By substituting $x_{i+2}, x_{i+3}, \ldots$ into $y_{i+3}, y_{i+4}, \ldots$, we find after some algebra that

$$y_{j+1} = (d + x_i) \prod_{l=i+1}^{j} (1 - K_l) - \sum_{l=i+1}^{j} K_l \varepsilon_l \prod_{r=l+1}^{j} (1 - K_r) + \varepsilon_{j+1}.$$

This can be simplified to:

$$y_{j+1} = \frac{i + \sigma_\varepsilon^2/P_0}{j + \sigma_\varepsilon^2/P_0}(d + x_i) - \frac{1}{j + \sigma_\varepsilon^2/P_0} \sum_{l=i+1}^{j} \varepsilon_l + \varepsilon_{j+1}. \tag{6.3}$$

Now consider using the estimation procedure based on (6.2) to derive a general expression for y_{j+1}. We have that

$$y_l = d + x_i + \varepsilon_l,$$

for $l = i+1, i+2, ..., j$. So, the sum of $y_{i+1}...y_j$ is

$$\sum_{l=i+1}^{j} y_l = (j-i)(d+x_i) + \sum_{l=i+1}^{j} \varepsilon_l.$$

Substituting this last result into (6.2) and taking the result back into $y_{j+1} = d + x_j + \varepsilon_{j+1} = d - \hat{d}_j + \varepsilon_{j+1}$, we have that

$$y_{j+1} = d + x_i - \frac{j-i}{j + \sigma_\varepsilon^2/P_0}(d + x_i + \frac{1}{j-i}\sum_{l=i+1}^{j} \varepsilon_l) + \varepsilon_{j+1}.$$

One can see that this simplifies and equals to (6.3).

Notice that equation (6.2) can be written as

$$\hat{d}_j = \hat{d}_i + \frac{j-i}{\sigma_\varepsilon^2/P_0 + j}\bar{y}_{ij}$$

where $\bar{y}_{ij}$ is the arithmetic average of $y_{i+1}...y_j$. Thus, we call this adjustment method "sample-average adjustment".

6.2   Algorithmic and Heuristic Methods for Optimal Adjustment Schedule

Since the sample-average adjustment procedure provides the opportunity of skipping adjustment actions between two arbitrary adjusting times, this procedure is especially useful when there are fixed adjustment costs independent of the magnitude of the adjustments. The goal is to find an optimal adjustment schedule that minimizes the total

Fig. 6.1. The graphical representation of an adjustment schedule

cost when N is given and all of the cost parameters are known. The optimal schedule can be obtained by using dynamic programming. The formulation is analogous to what in the inventory control literature is the well-known Wagner-Whitin (W-W) algorithm (Wagner and Whitin, [108]). In Figure 6.1, we represent the starting time of manufacturing each part as a node on a network used to determine the production timeline, so the optimal schedule is equivalent to a minimal cost path from node 1 to node n.

As before, the target of y_t is assumed to be zero, so $C_q = \Omega \sum_{t=1}^{N} (E^2[y_t] + Var(y_t))$. From (6.3), it is easy to derive that

$$\frac{E[y_t]}{\sigma_\varepsilon} = \frac{d - \hat{d}_0}{P_0/\sigma_\varepsilon^2(t-1) + 1}$$

and

$$\frac{Var(y_t)}{\sigma_\varepsilon^2} = 1 + \frac{t-1}{(\sigma_\varepsilon^2/P_0 + t - 1)^2}.$$

If $\hat{d}_0$ is an unbiased estimate of d (recall that d is an unknown constant), then $E[y_t] = 0$,

i.e., y_t is also unbiased. Define C_{ij} to be the cost from node $i + 1$ to node j. Then, we

have that

$$C_{ij} = M + \Omega(j - i)[1 + \frac{i}{(\sigma_\varepsilon^2/P_0 + i)^2}]\sigma_\varepsilon^2 + (j - i)G. \qquad (6.4)$$

The last item on the right hand side of the equation is dropped when $j = N$. The

W-W algorithm requires computation of the cost between pairs of nodes according to

the recursion:

$$C^{w-w}(j) = min\{C_{ij} + C^{w-w}(i), j = i + 1, \ i + 2, \ ..., \ N\} \quad for \ i = N - 1, N - 2, ..., 1,$$

and

$$C^{w-w}(N) = 0,$$

where $C^{w-w}(i)$ is the minimum cost from node i to j.

The computational effort of the W-W algorithm consists of at least $N(N + 1)/2$

calculations for the C_{ij}'s and $N(N - 1)/2$ comparisons. Therefore, it has a 2nd-order

polynomial complexity. Many heuristic methods have been proposed in the inventory

control literature to overcome some of the difficulties of the W-W method (Silver, Pyke

and Peterson, [90]). For example, a simpler method, based on the Silver-Meal (S-M)

heuristic which has been proved to have close to optimal performance in inventory control

applications (Silver and Meal, [89], Simpson, [91]), can be applied to this problem. This

method searches for the minimum unit cost, $C_{ij}/(j - i)$ by fixing period i and increasing

period j until a local minimum is obtained. Obviously, in equation (6.4), if i is fixed and

j increased, the unit cost will decrease consistently, that is, no adjustment is scheduled except for the first period. This result is rather uninteresting. So we work out the procedure in the backward direction: we fix j and decrease i to find the minimum unit cost. We call this searching method the *backward S-M method*. It has a 1st-order polynomial computational complexity.

To illustrate the backward S-M method, consider the case where $N = 20$, $G = 0$, $M = 0.5$ and $\sigma_\varepsilon = P_0 = 1$. Substituting these values into equation (6.4), setting $j = 20$ and decreasing i from 19 to 1, it is easy to find that $C_{ij}/(j - i)$ is minimized to 1.1319 when $i = 11$. Then setting $j = 11$ and decreasing i from 10 to 1, it is found that $C_{ij}/(j - i)$ is minimized to 1.2222 when $i = 5$. Finally, after letting $j = 5$ and recalculating $C_{ij}/(j - i)$ for $i < j$, the minimal value is found at $i = 1$. Thus, the adjustment steps given by this method are 1, 5 and 11.

Trietsch [101] proposed another approximating method for the optimal adjustment schedule, in which the time of adjustment, t, is treated as a continuous number and the results need to be rounded to the closest integers. To use this method, one needs to solve two complicated equations numerically when the procedure starts and also needs to solve another equation numerically at every subsequent iteration, so the computation effort greatly depends on the initial values selected for these equations. Furthermore, a improper selection of initial values used in this method can cause incorrect results.

6.3 Comparison of Numerical Results

Adjustment schedules with different total production runs and different measurement and adjustment costs were studied. Notice that the total cost is a linear combination of three components, so we can assume without loss of generality that the off-target quality cost parameter, Ω, is one unit. In this section, we assume that the first adjustment, if it is needed, is always applied at time 1, i.e., before the start of the manufacturing process, based on our previous knowledge of a possible setup error (d is usually assumed to be 0, so in this case no adjustment is needed before manufacturing starts). It is also assumed that the initial estimate of the mean of d, $\hat{d}_0$, is unbiased, so the first adjustment will bring the process to target on average. More importantly, the inherent process variance σ_ε^2 is assumed to be known in this section. Sections 4.1 and 4.2 show further analysis where some of these assumptions were removed.

In Tables 6.1-6.4, we contrast results from the three methods mentioned in the last section. The number of parts produced varies from 20 to 500, the costs of measurement and adjustment vary from 0 to 2 units and from 0.5 to 2 units, respectively. For those cases in the table, $\sigma_\varepsilon = P_0 = 1$. The W-W algorithm provides the optimal adjustment schedule for each case, but the computation effort of this algorithm is much greater than that of the other methods, especially when the total number of production runs, n, is large. By comparing the S-M and Trietsch's methods with the W-W method, we can make the following remarks.

Remark 1. The cost of measurement and adjustment will greatly affect the total number of adjustments. Generally, when these costs increase, the number of adjustments

N	G	M	method	cost	time(sec.)	adjustments
20	0	0.5	WW	22.92	~ 0	1-7
			SM	23.02	~ 0	1-5-11
			Trietsch's	23.36	0.21	1-3-6-11
		1	WW	23.92	~ 0	1-7
			SM	24.62	~ 0	1-3-9
			Trietsch's	25.36	0.22	1-3-6-11
		2	WW	25.75	~ 0	1
			SM	25.92	~ 0	1-7
			Trietsch's	27.78	0.14	1-2-7
	1	0.5	WW	24.25	0.02	1
			SM	24.25	~ 0	1
			Trietsch's	25.69	0.13	1-3
		1	WW	24.75	~ 0	1
			SM	24.75	~ 0	1
			Trietsch's	26.25	0.1	1-2
		2	WW	25.75	~ 0	1
			SM	25.75	~ 0	1
			Trietsch's	28.25	0.1	1-2
	2	0.5	WW	24.25	0.02	1
			SM	24.25	~ 0	1
			Trietsch's	26.25	0.11	1-2
		1	WW	24.75	~ 0	1
			SM	24.75	~ 0	1
			Trietsch's	27.25	0.1	1-2
		2	WW	25.75	~ 0	1
			SM	25.75	~ 0	1
			Trietsch's	29.25	0.11	1-2

Table 6.1. Comparison of costs, time and adjustment schedules of 3 schedule design methods, when the total number of parts produced is 20. The numbers in the columns G and M can be viewed as the ratios of the per unit measurement cost to the off-target quality cost (G/Ω), and of the unit adjustment cost to the off-target quality cost (M/Ω).

N	G	M	method	cost	time(sec.)	adjustments
50	0	0.5	WW	54.78	0.09	1-7-20
			SM	55.08	~ 0	1-4-8-16-28
			Trietsch's	55.46	0.29	1-3-7-14-22-33
		1	WW	56.28	0.09	1-7-20
			SM	56.86	~ 0	1-4-11-24
			Trietsch's	57.62	0.26	1-3-7-14-27
		2	WW	58.45	0.08	1-12
			SM	59.29	~ 0	1-7-19
			Trietsch's	61.15	0.18	1-2-7-19
	1	0.5	WW	61.11	0.08	1-4
			SM	61.11	~ 0	1-4
			Trietsch's	61.63	0.17	1-3-5
		1	WW	62.11	0.08	1-4
			SM	62.11	~ 0	1-4
			Trietsch's	62.25	0.13	1-5
		2	WW	63.25	0.08	1
			SM	64.11	~ 0	1-4
			Trietsch's	64.11	0.1	1-4
	2	0.5	WW	61.75	0.09	1
			SM	61.75	~ 0	1
			Trietsch's	64.11	0.13	1-4
		1	WW	62.25	0.08	1
			SM	62.25	~ 0	1
			Trietsch's	65.11	0.13	1-4
		2	WW	63.25	0.08	1
			SM	63.25	~ 0	1
			Trietsch's	66.31	0.1	1-3

Table 6.2.   Comparison of costs, time and adjustment schedules of 3 schedule design methods, when the total number of parts produced is 50.

N	G	M	method	cost	time(sec.)	adjustments
100	0	0.5	WW	106.22	0.32	1-7-18-43
			SM	106.52	~ 0	1-5-10-18-33-58
			Trietsch's	106.89	0.33	1-3-7-14-27-42-65
		1	WW	108.06	0.32	1-10-33
			SM	108.82	0.02	1-4-10-23-49
			Trietsch's	109.6	0.28	1-3-8-19-33-58
		2	WW	111.06	0.32	1-10-33
			SM	112.28	~ 0	1-5-15-40
			Trietsch's	114.12	0.23	1-2-6-16-40
	1	0.5	WW	117.67	0.32	1-7
			SM	117.92	~ 0	1-3-7
			Trietsch's	118.02	0.17	1-3-8
		1	WW	118.67	0.32	1-7
			SM	118.67	~ 0	1-7
			Trietsch's	119.67	0.13	1-2-8
		2	WW	120.67	0.32	1-7
			SM	120.67	~ 0	1-7
			Trietsch's	122.67	0.13	1-2-8
	2	0.5	WW	122.11	0.32	1-4
			SM	122.11	~ 0	1-4
			Trietsch's	123.15	0.11	1-2-6
		1	WW	123.11	0.32	1-4
			SM	123.11	~ 0	1-4
			Trietsch's	124.65	0.13	1-2-6
		2	WW	125.11	0.32	1-4
			SM	125.11	~ 0	1-4
			Trietsch's	127.65	0.13	1-2-6

Table 6.3.   Comparison of costs, time and adjustment schedules of 3 schedule design methods, when the total number of parts produced is 100.

N	G	M	method	cost	time(sec.)	adjustments
500	0	0.5	WW	509.65	7.93	1-7-18-43-99-223
			SM	510.12	0.08	1-5-10-18-33-57-99-170-229
			Trietsch's	510.18	0.4	1-3-7-15-31-54-95-166-288
		1	WW	512.37	7.92	1-8-25-70-188
			SM	513.2	0.07	1-5-13-29-60-123-249
			Trietsch's	514.03	0.36	1-3-8-19-44-99-170-292
		2	WW	516.62	7.93	1-11-42-147
			SM	518.8	0.06	1-4-13-34-84-206
			Trietsch's	520.64	0.29	1-2-7-22-63-126-251
	1	0.5	WW	544.16	7.92	1-7-19
			SM	544.29	0.04	1-5-10-19
			Trietsch's	544.39	0.21	1-3-8-20
		1	WW	545.66	7.99	1-7-19
			SM	546.37	0.04	1-3-8-19
			Trietsch's	546.55	0.16	1-2-7-20
		2	WW	548.28	7.94	1-18
			SM	548.66	0.03	1-7-18
			Trietsch's	549.35	0.14	1-3-19
	2	0.5	WW	558.91	7.93	1-6-13
			SM	559.21	0.04	1-3-7-13
			Trietsch's	559.26	0.18	1-3-7-14
		1	WW	560.3	7.93	1-13
			SM	560.41	0.04	1-5-13
			Trietsch's	560.49	0.15	1-5-14
		2	WW	562.3	7.92	1-13
			SM	563.49	0.03	1-4-13
			Trietsch's	563.8	0.13	1-3-14

Table 6.4.   Comparison of costs, time and adjustment schedules of 3 schedule design methods, when the total number of parts produced is 500.

decreases, and the measurement cost only affects the time of the last adjustment. This occurs because after the last adjustment no more measurements are needed, whereas before that time measurements should be conducted at every period (or part). Therefore, when the production runs are short and significant measurement or adjustment costs exist, it is optimal to adjust just once.

Remark 2. The distance between adjacent adjustments increases steadily in the sequence of adjustments. This can be explained as follows: when the process has been adjusted close but not exactly on target, we need stronger evidence from the process to demonstrate there is still an offset and to obtain an unbiased estimate of it. Such evidence is only obtained with longer runs of observations between adjustments.

Remark 3. Comparing the backward S-M and Trietsch's methods, the backward S-M method always give fewer or equal number of adjustments than the Trietsch's method does, and the schedule from S-M is usually closer to the one given by W-W. Thus the backward S-M method is a better heuristic in terms of minimizing total cost.

Remark 4. The S-M method has the advantage of reducing computing effort significantly comparing to the W-W algorithm and its computing time is consistently less than that of the Trietsch's method. In fact, this method can be easily implemented by using a handheld calculator or a spreadsheet software to support on-line process adjustment decisions.

6.3.1   The case when the process variance (σ_ε) is unknown

In the following discussion, we vary the true value of σ_ε to 0.8 and to 1.2, that is, the estimate $\hat{\sigma}_\varepsilon$, which was assumed to be 1 in the previous discussion and was used

in the three methods to obtain the optimal or near-optimal adjustment schedules is now an over- or under-estimate, respectively, of the true σ_ε. The minimum cost and the optimal adjustment schedule can be obtained by introducing the true value of σ_ε into the W-W algorithm. This was used as a benchmark to compare cost increments due to a poorly estimated σ_ε. The cost increments induced by over- or under-estimating σ_ε are investigated and presented in Figures 6.2 and 6.3.

Remark 5. When the production run, n, is large ($n = 100$, or 500), the cost increments by introducing over- or under-estimated σ_ε of using the three methods are very small and can be ignored.

Remark 6. The backward S-M method generally performs better than Trietsch's method, except in some particular cases (see, e.g., in Figure 6.2, when $n = 50$, $G = 2$, $M = 0.5$, 1 and 2) if σ_ε is over-estimated. The W-W method still performs well when $G = 0$ (no measurement cost). But when the measurement cost is considered, the W-W method is not better than the backward S-M method in general.

Remark 7. Under-estimating σ_ε ($\sigma_\varepsilon = 1.2$) will induce larger increases in cost than when over-estimating σ_ε for all rules. Normally when the true value of σ_ε is larger than expected, longer runs of parts between adjustments are needed since these provide more measurements for estimating the setup error magnitude; therefore, the total number of adjustments will decrease. Since Trietsch's method always provides more adjustments than the other two methods, it will incur in a higher cost. Conversely, when σ_ε is over-estimated, we find that Trietsch's method has an advantage in three cases: $N = 50$, $G = 2$, and $M = 0.5$, 1 and 2.

Fig. 6.2. Performance of adjustment scheduling methods for the case when σ_ε is over-estimated ($\sigma_\varepsilon = 0.8$, $\hat{\sigma}_\varepsilon = 1$). All of the cases presented in Tables 6.1-6.4 are investigated and compared.

Fig. 6.3. Performance of adjustment scheduling methods for the case when σ_ε is underestimated ($\sigma_\varepsilon = 1.2$, $\hat{\sigma}_\varepsilon = 1$) .

6.3.2 The case when a biased initial estimate of the offset is used

In the previous calculations, we assume the initial estimation, $\hat{d}_0$, equals to d; however, this assumption is hardly realistic since if d is exactly known, a one-step adjustment (or calibration) is enough for removing the offset. A biased initial estimate of d will lead to an increase in cost in the adjustment schedule calculated by each method. Similarly as in Section 4.1, the robustness of the three methods under such situation are compared and presented in Figures 6.4 and 6.5. We varied $|d - \hat{d}_0|$ to 1 and to 2.

Remark 8. Similarly as in remark 5, when the production run is large, the cost increments caused by the biased initial estimate of d are insignificant.

Remark 9. When n is small, Trietsch's method is generally the best one and the W-W method is worst. We observe that when $\hat{d}_0$ is close to d the backward S-M method can out-perform the Trietsch's method, e.g., see the cases when $|d - \hat{d}_0| = 1\sigma_\varepsilon$, $N = 50$ and $G = 0$. But when $\hat{d}_0$ is strongly biased for d or when the measurement cost is high, the S-M method may lead to a high cost increase, e.g., see the cases when $|\hat{d}_0 - d| = 2\sigma_\varepsilon$, $N = 50$ and $G = 2$.

Remark 10. When $|d - \hat{d}_0|$ is large, the cost increments are more severe and Trietsch's method is more robust compared to other methods since it adjusts more often than the other two methods.

It was found that when $|d - \hat{d}_0| \neq 0$, more adjustments are required by the optimal schedule, especially in the first few periods. Since Trietsch's method happens to provide more adjustments, it results in the lowest cost increments.

Fig. 6.4. Performance of adjustment scheduling methods for the case when $|d - \hat{d}_0| = 1\sigma_\varepsilon$.

Fig. 6.5.   Performance of adjustment scheduling methods for the case when $|d - \hat{d}_0| = 2\sigma_\varepsilon$.

Fig. 6.6. Comparison of performance of the modified S-M heuristic and other methods when $|d - \hat{d}_0| = 1\sigma_\varepsilon$ or $|d - \hat{d}_0| = 2\sigma_\varepsilon$ and $N = 20$ or $N = 50$.

To enhance the performance of the backward S-M method when $|d - \hat{d}_0| \neq 0$, we suggest to add one more adjustment at the beginning of the production run. The small-sample ($N = 20$ and $N = 50$) performance comparison of this modified S-M heuristic and other methods is presented in Figure 6.6 for some inaccurate initial estimate of the offset. The results show that the adjustment schedule given by this heuristic method has a similar cost as that given by the Trietsch's method. However, Trietsch's method seems to still have a slight cost advantage over the modified S-M approach but this is not enough, in our view, to justify its more complex computations over the modified S-M method. To illustrate the modified S-M heuristic, consider the case when $n = 50$, $G = 2$ and $M = 0.5$. The backward S-M method suggests a single adjustment at node 1. This schedule will incur in more than a 50% cost increase compared to the minimal cost if the initial estimate of the setup error is two units from the true value, i.e., if $|d - \hat{d}_0| = 2$. By inserting a second adjustment after the first one, i.e., by changing the adjustment steps to 1-2, the cost increment reduces to only 14%.

6.4 Summary

In this chapter, a sample-average adjustment procedure for adjusting the setup error was introduced when process measurement and process adjustment cost cannot be ignored. We compare the performance of three adjustment scheduling methods, which can achieve the optimal or close-to-optimal expected total manufacturing cost: the Wagner-Whitin method, a backward implementation of the Silver-Meal method and a method due to Trietsch [101]. It was found that when the production run is long (i.e., large lots of product), there is not significant difference between the performance of

the three methods. For a short-run manufacturing process, the proposed backward S-M method has the advantage of providing a close-to-optimal solution with small computational effort, even when the process variance estimate is biased. However, when there exists a significant bias on the initial estimate of the setup error and when the adjustment or the measurement costs are relatively high, the schedule provided by the backward S-M method may incur in a much higher cost increase than Trietsch's method. Finally, it is demonstrated that simply adding one more adjustment close to the beginning of the schedule enhances the robustness of the backward S-M method.

Chapter 7

Identification and Fine-Tuning of Closed-Loop Processes

In previous chapters, sequential process adjustment methods were discussed as an alternative to some process control schemes that continuously adjust the process, such as the discrete integral (or EWMA) controller. It was shown that if the autocorrelation in the original process observations is pronounced, or if the random shocks on the process mean are frequent, it is better to use a time series process model, such as an ARIMA (AutoRegressive Integrated Moving Average) model, and design a controller accordingly to minimize the process variance. Methods for designing such controllers for a process operating in open loop, that is, without the actions of any controller, are discussed extensively in the time series and control literature (see, e.g., Box, Jenkins and Reinsel [13]). However, in this chapter, a methodology for identifying a process already operating under the actions of a given controller (closed-loop operation) is proposed. Furthermore, it is suggested how to tune the controller based on the identified process to optimize the closed-loop performance of the process.

Conventional process identification techniques of an open-loop process use the cross-correlation function between historical values of the process input and of the process output. If the process is operated under the actions of a linear feedback controller, however, the cross-correlation function has no information on the process transfer function because of the linear dependency of the input on the output. In this chapter, several

circumstances where a closed-loop system can be identified by the autocorrelation function of the output are discussed. It is assumed that a proportional integral controller or an EWMA controller with known parameters is acting on the process while the process observations (output data) were collect. The disturbance is assumed to be a member of a simple yet useful family of stochastic models. It is shown that, with these general assumptions, it is possible to identify some dynamic process models commonly encountered in manufacturing. After process identification, tuning the controller parameters to optimize its performance becomes possible.

7.1 Process Model and Disturbance Dynamics

In this chapter, a more complicated process model will be used, where, unlike the simple adjustment models assumed in previous chapters, the process output (observations) is connected to the process input (controllable factor) through a non-trivial transfer function (ratio of two polynomial in the backshift operator), so the input can have delayed and dynamic transient effects on the output. Also, the process disturbance is assumed to be autocorrelated and it is modelled by an instance of an ARIMA model. Following the modelling approach due to Box *et al.* [13], it is assumed that an observed output deviation from target, y_t, consists of two components – a process "signal", S_t, and disturbance, N_t. That is, the signal generated by the underlying manufacturing mechanism can only be observed under the presence of a disturbance as follows:

$$y_t = S_t + N_t \qquad\qquad (7.1)$$

where S_t is written as a rational transfer function and the disturbance N_t is an ARIMA

process. The assumed process is then

$$y_t = \alpha + \frac{(\beta_0 + \beta_1 \mathcal{B} + \cdots + \beta_s \mathcal{B}^s)}{(1 - \phi_1 \mathcal{B} - \cdots - \phi_r \mathcal{B}^r)} \mathcal{B}^b x_t + N_t \tag{7.2}$$

and

$$N_t = \delta + N_{t-1} - \theta \varepsilon_{t-1} + \varepsilon_t, \quad |\theta| \leq 1 \tag{7.3}$$

where $\mathcal{B}$ is the backshift operator (defined as $\mathcal{B} e_t = e_{t-1}$), $\{\varepsilon_t\}$ are i.i.d. random numbers

and α is a constant (not necessarily zero) representing the expected deviation from target

when the input is set at a value of zero.

According to the Box-Jenkins taxonomy, the process is an (r,s,b) order transfer

function plus an IMA(1,1) with drift disturbance. In practice, r, s and b are rarely larger

than 2. According to Del Castillo [32], the disturbance model contains a useful family

of models. Depending on the different values δ and θ take, the disturbance is one of

the processes listed on Table 7.1. Note that we allow θ to be equal to one; then the

disturbance model is either a deterministic trend disturbance, which is useful to model

wearing-off of a tool in a variety of manufacturing processes, or a white noise process in

case $\delta = 0$.

δ	θ	Disturbance
0	0	Random Walk
$\neq 0$	0	Random Walk with Drift
0	$\neq 0$	IMA(1,1)
$\neq 0$	$\neq 0$	IMA(1,1) with Drift
$\neq 0$	1	Deterministic Trend
0	1	White Noise

Table 7.1.   Disturbance models described by Equation (7.3)

7.2   Process Identification

7.2.1   ARMA modelling of process observations from closed-loop data

Suppose that for a manufacturing process operating in closed-loop, no first principle knowledge of the process dynamic mechanism is available. Therefore, it is necessary to identify an empirical transfer function model that best describes the process behaviour. On the other hand, the feedback controller functioning on the input data is intently designed and installed by control engineers, so its adjustment scheme is assumed to be known. When the controller is not optimal in the sense of not minimizing the mean square error of the process output, the output deviations will exhibit certain autocorrelation patterns that are useful for process identification. In this section, we will derive the ARMA models that describe the output deviations for some processes that are commonly encountered in manufacturing and show how to use some advanced statistical techniques to identify them. Here, we assume that one of two types of controller - either an EWMA or a PI controller - is in use in the closed-loop while the data were collected.

The EWMA controller has attracted considerable attention in recent years, especially for the run-to-run control of batch productions in semiconductor manufacturing (see Ingolfsson and Sachs [51], Del Castillo and Hurwitz [33]). In such a closed-loop system, the effect of the process adjustments will be fully observed at the next output value, i.e., the process can be described by the following model:

$$y_t = \alpha + \beta x_{t-1} + N_t, \tag{7.4}$$

where α is the process offset, and β is the process gain. Note that in previous chapters, β was assumed to be 1. This is a particular case of model (7.2) with a (r,s,b)=(0,0,1) transfer function. For the EWMA controller, the process gain, β, is estimated by off-line experiments and is represented by b, and the initial estimate of α is a_0. The control scheme is as follows:

$$x_t = -\frac{a_t}{b}$$

and

$$a_t = \lambda(y_t - bx_{t-1}) + (1 - \lambda)a_{t-1} \qquad (0 \le \lambda \le 1)$$

where λ is a parameter that can be adjusted to achieve a desired behavior. The EWMA controller updates a_t in order to reduce the estimation error of the process offset, α. In fact, it is easy to show that the adjustment at each step is proportional to the present output deviation, that is,

$$\nabla x_t = -\frac{\nabla a_t}{b} = -\frac{\lambda(y_t - bx_{t-1}) + (1 - \lambda)a_{t-1} - a_{t-1}}{b} = -\frac{\lambda y_t}{b}. \tag{7.5}$$

One can compare this controller with a PI controller, a widely used industrial controller, which is a combination of two control schemes - proportional control and integral control:

$$x_t = k_P y_t + k_I \sum_{i=1}^{t} y_i \qquad (7.6)$$

where k_P and k_I are the proportional and integral control constants respectively. An equivalent form of equation (7.6) would make the input adjustment depend linearly on the last two output deviations:

$$\nabla x_t = c_1 y_t + c_2 y_{t-1} \qquad (7.7)$$

where $c_1 = k_P + k_I$ and $c_2 = -k_P$. As one can see from Equation (7.5), the EWMA controller is actually a special PI controller, i.e., a pure I controller with $c_1 = k_I = -\frac{\lambda}{b}$.

To derive the "closed-loop description" of the output deviations, we take first-order differences on the process and disturbance equations, then substitute the controller and disturbance functions into the process equation to obtain an ARMA model of the deviation. For instance, by taking first-order differences on process equation (7.4) and disturbance function (7.3), we have

$$\nabla y_t = \beta \nabla x_{t-1} + \nabla N_t, \qquad (7.8)$$

and

$$\nabla N_t = \delta + (1 - \theta \mathcal{B}) \varepsilon_t. \qquad (7.9)$$

Substituting Equations (7.9) and (7.5) into (7.8), we get

$$(1 - (1 - \lambda\xi)\mathcal{B})y_t = \delta + (1 - \theta\mathcal{B})\varepsilon_t \qquad (7.10)$$

where, $\xi = \frac{\beta}{b}$, is a measure of the bias in the gain estimate. Therefore, under the adjustment of an EWMA controller, the sequence of deviations from this closed-loop system is an ARMA(1,1) process with an asymptotic mean value of $\frac{\delta}{\lambda\xi}$. This result had been reported in Del Castillo [32].

Equation (7.4) describes a simple manufacturing process where the process output is fully determined by the most recent value of the controllable factor. In some more complicated processes, the delay between the input adjustment and output observation could be longer than one time period and also the effect of adjustments could extend to several subsequent time periods. Therefore, it is important to study all possible ARMA models that describe the output deviations for a class of transfer functions. Commonly found process transfer functions include one-time or two-time delay and first-order dynamic models. The ARMA models that describe the output deviations can be derived by using the same procedure as we did for the transfer function of order (0,0,1). For brevity, our results for EWMA and PI controllers are summarized on Tables 7.2 and 7.3 respectively. On these tables, a transfer function is given first, followed by the ARMA model of the deviations from target and by the asymptotic process mean. Note that the order of the ARMA models does not exceed two for an EWMA controller and three for a PI controller. This implies that to identify a closed-loop process under these

controllers, we should focus on searching a low order ARMA pattern from the process output data.

7.2.2   Stationarity of ARMA(2,q) process

Identifying a closed-loop process only based on the ARMA model of the output deviations could be ambiguous, because, as found in Table 7.2, there exist more than one transfer function corresponding to ARMA models of the same order, like in the case of an ARMA(2,1) or an ARMA(2,2). Carefully comparing the estimated values of parameters in these models may help to distinguish different processes if reliable parameter estimates are available. Since it is assumed that the closed-loop process has been stabilized by the adjustments of a suboptimal controller, the stationarity conditions of an ARMA(2,q) provide some additional constraints for the process parameters that are useful for identification purposes.

As it is well-known, a stationary ARMA(2,q) process $(1 - a_1 \mathcal{B} - a_2 \mathcal{B}^2) y_t = \Theta(\mathcal{B}) \varepsilon_t$ must satisfy the following conditions:

$$a_1 + a_2 < 1 \qquad a_2 - a_1 < 1 \qquad |a_2| < 1.$$

Applying these conditions to the two ARMA(2,2) processes listed on Table 7.2, for the transfer function of order (1,0,1), we get

$$0 < \lambda \xi < 2(1 + \phi) \qquad -1 < \phi < 1.$$

EWMA Controller		
Transfer function	r=0, s=0, b=1	$y_t = \alpha + \beta x_{t-1}$
Output deviation from target	ARMA(1,1)	$(1 - (1 - \lambda\xi)\mathcal{B})y_t = \delta + (1 - \theta\mathcal{B})\varepsilon_t$
Process mean	$\mu = \frac{\delta}{\lambda\xi}$	
Transfer function	r=1, s=0, b=1	$(1 - \phi\mathcal{B})y_t = \alpha + \beta x_{t-1}$
Output deviation	ARMA(2,2)	$(1 - (1 + \phi - \lambda\xi)\mathcal{B} - (-\phi)\mathcal{B}^2)y_t =$ $(1 - \phi)\delta + (1 - (\theta + \phi)\mathcal{B} - (-\theta\phi)\mathcal{B}^2)\varepsilon_t$
Process mean	$\mu = \frac{(1-\phi)\delta}{\lambda\xi}$	
Transfer function	r=0, s=1, b=1	$y_t = \alpha + \beta_1 x_{t-1} + \beta_2 x_{t-2}$
Output deviation	ARMA(2,1)	$(1 - (1 - \lambda\xi_1)\mathcal{B} - (-\lambda\xi_2)\mathcal{B}^2)y_t =$ $\delta + (1 - \theta\mathcal{B})\varepsilon_t$
Process mean	$\mu = \frac{\delta}{\lambda(\xi_1 + \xi_2)}$	
Transfer function	r=0, s=0, b=2	$y_t = \alpha + \beta x_{t-2}$
Output deviation	ARMA(2,1)	$(1 - \mathcal{B} - (-\lambda\xi)\mathcal{B}^2)y_t = \delta + (1 - \theta\mathcal{B})\varepsilon_t$
Process mean	$\mu = \frac{\delta}{\lambda\xi}$	
Transfer function	r=1, s=0, b=2	$(1 - \phi\mathcal{B})y_t = \alpha + \beta x_{t-2}$
Output deviation	ARMA(2,2)	$(1 - (1 + \phi)\mathcal{B} - (-\lambda\xi - \phi)\mathcal{B}^2)y_t =$ $(1 - \phi)\delta + (1 - (\theta + \phi)\mathcal{B} - (-\theta\phi)\mathcal{B}^2)\varepsilon_t$
Process mean	$\mu = \frac{(1-\phi)\delta}{\lambda\xi}$	

Table 7.2. ARMA models describing the deviations from target from different EWMA controlled processes. In all cases, the disturbance is $N_t = \delta + N_{t-1} - \theta\varepsilon_{t-1} + \varepsilon_t$.

PI Controller		
Transfer function	r=0, s=0, b=1	$y_t = \alpha + \beta x_{t-1}$
Output deviation from target	ARMA(2,1)	$(1 - (1 + c_1\beta)\mathcal{B} - c_2\beta\mathcal{B}^2)y_t = \delta + (1 - \theta\mathcal{B})\varepsilon_t$
Process mean	$\mu = \frac{\delta}{(c_1 + c_2)\beta}$	
Transfer function	r=1, s=0, b=1	$(1 - \phi\mathcal{B})y_t = \alpha + \beta x_{t-1}$
Output deviation	ARMA(2,2)	$(1 - (1 + \phi + c_1\beta)\mathcal{B} - (c_2\beta - \phi)\mathcal{B}^2)y_t =$ $(1 - \phi)\delta + (1 - (\theta + \phi)\mathcal{B} - (-\theta\phi)\mathcal{B}^2)\varepsilon_t$
Process mean	$\mu = \frac{(1-\phi)\delta}{(c_1 + c_2)\beta}$	
Transfer function	r=0, s=1, b=1	$y_t = \alpha + \beta_1 x_{t-1} + \beta_2 x_{t-2}$
Output deviation	ARMA(3,1)	$(1 - (1 + c_1\beta_1)\mathcal{B} - (c_2\beta_1 + c_1\beta_2)\mathcal{B}^2 - c_2\beta_2\mathcal{B}^3)e_t = \delta + (1 - \theta\mathcal{B})\varepsilon_t$
Process mean	$\mu = \frac{\delta}{(c_1 + c_2)(\beta_1 + \beta_2)}$	
Transfer function	r=0, s=0, b=2	$y_t = \alpha + \beta x_{t-2}$
Output deviation	ARMA(3,1)	$(1 - \mathcal{B} - c_1\beta\mathcal{B}^2 - c_2\beta\mathcal{B}^3)y_t = \delta + (1 - \theta\mathcal{B})\varepsilon_t$
Process mean	$\mu = \frac{\delta}{(c_1 + c_2)\beta}$	
Transfer function	r=1, s=0, b=2	$(1 - \phi\mathcal{B})y_t = \alpha + \beta x_{t-2}$
Output deviation	ARMA(3,2)	$(1 - (1 + \phi)\mathcal{B} - (c_1\beta - \phi)\mathcal{B}^2 - c_2\beta\mathcal{B}^3)y_t =$ $(1 - \phi)\delta + (1 - (\theta + \phi)\mathcal{B} - (-\theta\phi)\mathcal{B}^2)\varepsilon_t$
Process mean	$\mu = \frac{(1-\phi)\delta}{(c_1 + c_2)\beta}$	

Table 7.3. ARMA models describing the deviations from target of different PI controlled processes. In all cases, the disturbance model is $N_t = \delta + N_{t-1} - \theta\varepsilon_{t-1} + \varepsilon_t$

Fig. 7.1. Stability region of two ARMA(2,2) processes

For the transfer function of order (1,0,2), we have

$$0 < \lambda\xi \qquad 2(1+\phi) < \lambda\xi \qquad -1-\phi < \lambda\xi < 1-\phi.$$

The feasible regions of ϕ and $\lambda\xi$ for these two processes are shown in Figure 7.1. These two regions nicely separate from each other, which indicates that one would be able to select a correct transfer function after comparing the estimated process parameters with their feasible regions. For example, if ϕ and $\lambda\xi$ estimated from an ARMA(2,2) process are 0.5 and 1 in an EWMA controlled process, then the transfer function of order (1,0,1) is the only one that should be accepted. Similarly, the feasible regions for the two ARMA(2,1) models on Table 7.2 are drawn in Figure 7.2. Again, there is little difficulty in identifying these two processes from their parameter estimates, because one process has the coefficient of $\mathcal{B}$ equal to one unit and the other does not.

Fig. 7.2.   Stability region of two ARMA(2,1) processes

As shown in Table 7.3, a process controlled by a PI controller may have a closed-loop description equal to an ARMA process of order as high as 3. It is difficult to draw the feasible parameter regions for ARMA(3,q), since there are three parameters involved. But one can see that for the two ARMA(3,1) processes, one has the coefficient of $\mathcal{B}$ equal to one and the other does not. Therefore, distinguishing between these alternatives based on the parameter estimates is possible in principle.

In practice, when the manufacturing process is exposed to substantial random noise, precise estimation of parameters in ARMA models is unrealistic. In addition, an accurate form of the model is not guaranteed and there is the possibility of model bias. However, as long as the model is a reasonable approximation of the true process, closed-loop identification will provide useful information for tuning the controller and improving performance as desired.

7.2.3   Techniques for identification of ARMA processes

Identifying an ARMA model from the sequence of output deviations is an essential procedure for obtaining a good process model. However, it is frequently difficult to identify a mixed ARMA process. The traditional approach of model identification utilizes the autocorrelation and partial autocorrelation functions of the process data and compares them with some theoretical patterns. This approach is only effective when dealing with pure AR or MA processes; but it is very difficult to determine a mixed ARMA model this way. Since the late 1980's, more identification techniques have been developed to handle this problem (see, e.g., Gray, Kelley and McIntire [43], Hannan and Rissanen [46]). Most of them compare the process data with a series of tentative models and select the one that fits best. The extended sample autocorrelation function (ESACF) and smallest canonical correlation (SCAN) proposed by Tsay and Tiao [102, 103] are two identification techniques that have been implemented in some statistical analysis program, such as SAS. They will be used for closed-loop identification purposes.

In the ESACF method, data are filtered through an AR model, whose autocorrelation coefficients are determined by a candidate ARMA model. The residuals of this filter are called the extended samples. It has been shown that the autocorrelation of these extended samples follows an MA(q) model if the true model is an ARMA(p,q), and the ARMA candidate we entertain has a MA polynomial of order higher or equal to q. More specifically, Tsay and Tiao had shown how to arrange the different condidate information in a table. The ESACF table will exhibit a triangular pattern of zeros when the candidate ARMA model has higher order than the true ARMA model. The

SCAN method is based on some consistency properties of suitably normalized second-order sample moment equations and make use of the method of canonical correlation in standard multivariate analysis. The zeroes pattern on a SCAN table will be rectangular for any candidate model having higher order than the true model.

We use an example to illustrate the usage of ESACF and SCAN tables on identifying a mixed ARMA process. This example also appeared in Tsay and Tiao [102], but they only gave the ESACF table. The data are taken from the Series A in Box *et. al.* [13] and a fitted ARMA(1,1) model given by authors is as

$$(1 - 0.92\mathcal{B})y_t = 1.45 + (1 - 0.58\mathcal{B})\varepsilon_t$$

The ESACF and SCAN tables computed by SAS are as in Figure(7.3). Clearly, one can see the triangular zeros pattern and rectangular zeros pattern starting from AR(1) and MA(1) block in the ESACF and SCAN respectively. It is very unlikely that the order of an ARMA model from real manufacturing data is higher than 2 for EWMA control or 3 for PI control; therefore, one should search the zero pattern starting from AR(1) or AR(2) block. We suggest to use both methods because closed-loop data may not exhibit a nice clear pattern in one table.

7.3 Tuning the Controller

After the process transfer function has been identified and its parameters have been estimated, it is possible to tune the controller to a near-optimal control state. By tuning a closed-loop process, it is meant to adjust the parameters of the controller to

Squared Canonical Correlation Estimates

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	0.3263	0.2479	0.1654	0.1387	0.1183	0.1417
AR 1	0.0643	0.0012	0.0028	<.0001	0.0051	0.0002
AR 2	0.0061	0.0027	0.0021	0.0011	0.0017	0.0079
AR 3	0.0072	<.0001	0.0007	0.0005	0.0019	0.0021
AR 4	0.0049	0.0010	0.0014	0.0014	0.0039	0.0145
AR 5	0.0202	0.0009	0.0016	<.0001	0.0126	0.0001

Extended Sample Autocorrelation Function

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	0.5702	0.4951	0.3980	0.3557	0.3269	0.3498
AR 1	-0.3907	-0.0425	-0.0605	-0.0083	-0.0651	-0.0127
AR 2	-0.2859	-0.2699	-0.0449	0.0089	-0.0509	-0.0140
AR 3	-0.5030	-0.0106	0.0946	-0.0137	-0.0148	-0.0302
AR 4	-0.4785	-0.0176	0.0827	-0.0244	-0.0149	-0.0421
AR 5	-0.3878	-0.4101	-0.1651	0.0103	-0.1741	-0.0231

Fig. 7.3.   SCAN and ESACF tables of Series A

improve its performance. Usually, a controller has some parameters that can be reset by process engineers, such as λ in an EWMA controller or (c_1, c_2) in a PI controller. The performance of the controller can be evaluated by the Asymptotic Mean Square Deviation (AMSD) of the process outputs, which is similar to the quadratic quality cost of the process as used in previous chapters. However, here, the transient effect of process adjustment is ignored since the closed-loop process is identified at its stable status. The AMSD criterion is valid when there is little cost associated with changing the controllable factor; otherwise, the variance of the controllable factor needs to be taken into consideration.

In this section, the AMSD is used as the controller performance criterion. The variance of the controllable factor will be discussed in Section 7.5. It is well known that the AMSD of a process consists of two parts: the variance of output deviations and the square of the bias. The calculation formula for the variance of an ARMA(2,q) process is given in Appendix F. The goal of tuning a closed-loop system is to minimize the AMSD subject to the process stationarity condition, that is,

$$Min \quad AMSD(y_t) = Var(y_t) + Bias(y_t)^2$$

$$S.T. \quad stationarity\ conditions.$$

By consulting the variance formula in Appendix F, exact AMSD expressions of many closed-loop processes can be obtained. For instance, for a process with the transfer function of order (1,0,1) being operated under an EWMA controller, its output deviations follow an ARMA(2,2) process with asymptotic mean value of $(1-\phi)\delta/\lambda\xi$. Therefore, its

AMSD is

$$AMSD(y_t) = \frac{(1+\phi)(1-\phi)^2(1-\theta)^2 + 2\phi\lambda\xi(1-\theta)^2 + 2\theta\lambda\xi(1-\phi^2+\phi\lambda\xi)}{\lambda\xi(1-\phi)(2+2\phi-\lambda\xi)}\sigma_\varepsilon^2$$

$$+(\frac{\delta(1-\phi)}{\lambda\xi})^2. \tag{7.11}$$

If instead the process has a transfer function (1,0,2), the output deviations relate to another ARMA(2,2) model with different parameters and its AMSD is

$$AMSD(y_t) = \frac{(1+\phi)(1-\phi)^2(1-\theta)^2 + \lambda\xi(1+\theta^2)(1+\phi^2) - 2\theta\phi\lambda\xi(2\phi+\lambda\xi)}{\lambda\xi(1-\phi-\lambda\xi)(2+2\phi+\lambda\xi)}\sigma_\varepsilon^2$$

$$+(\frac{\delta(1-\phi)}{\lambda\xi})^2. \tag{7.12}$$

In a previous section, it was shown that in principle, process transfer functions can be identified correctly even when more than one closed-loop process follow an ARMA model of the same order. However, this strongly depends on the quality of our parameter estimates. Therefore, it is of interest to investigate the possibility of improving the controller's performance when the transfer function has been mis-identified. In the following example, we will show that a controller can be tuned to a near-optimal state as long as the estimated model is a reasonable approximation of the true one.

7.4 Example: A Simulated Controlled Process with Real Disturbance Data

Boyles [20] recently reported an uncontrolled process in which the fill weight deviation from target for a powdered food product was recorded with the controller turned off. He mentioned that the process was unstable and autocorrelated, because powder density was affected by several uncontrollable variables, such as batch-to-batch variations, and he also suggested that an integral-type controller should be used. In this section, those data of fill weight deviations are regarded as the disturbance, to which a non-optimal EWMA controller is applied. We then apply the closed-loop identification and tuning methodology described in previous sections and optimize the controller.

First, the data set reported by Boyles is identified as an IMA(1,2) process, which is $\nabla N_t = (1 - \underset{(0.07)}{0.61} \; \mathcal{B} - \underset{(0.07)}{0.26} \; \mathcal{B}^2)\varepsilon_t$, where the numbers in paranthesis below the coefficients of $\mathcal{B}$ and $\mathcal{B}^2$ are the corresponding standard errors. The white noise sequence, $\{\varepsilon_t\}$, has an estimated variance of 207.5. This model will be used as the true disturbance in a simulated manufacturing process. Suppose the true process is repeatedly adjusted by an EWMA controller with control parameter $\lambda = 0.4$. The adjusting action may be thought as twisting a valve that directly determines the powder volume per time unit, hence, the powder weight. Normally, the effect of this type of adjustment can be realized only partially during one time interval. This results in a first order process transfer function with a one time delay. So the transfer function is characterized by the equation, $(1 - \phi\mathcal{B})y_t = \beta x_{t-1}$. Here, let us assume that $\phi = 0.4$, $\beta = 1$ and $b = 0.8$ (b is

Fig. 7.4. A controlled closed-loop process

the off-line estimate of β), so $\xi = \frac{\beta}{b} = 1.25$. By adding the same disturbance sequence as that in the open-loop process, we reconstruct a controlled process of fill weight deviation data as shown in Figure 7.4. It is evident that the process has been stabilized with mean value around 0. The estimated process variance is 292.2. We now illustrate our closed-loop identification methodology assuming the true model description is unknown.

SCAN and ESACF methods are applied to the simulated process output data to identify an ARMA model from which we can identify the process. The SCAN and ESACF tables are shown in Figure 7.5. One can see from the SCAN table that the pattern of rectangular zeroes starts from the AR(2) row and MA(1) column. This means that any ARMA process with order higher or equal to an ARMA(2,1) could be a candidate. The ESACF table does not shows a clear triangular pattern at low AR or MA order. Based on the parsimony principle, it is reasonable to guess that the closed-loop description of the process is ARMA(2,1). By fitting an ARMA(2,1) model to the output deviations from

target, we have that the maximum likelihood estimators of the AR and MA parameters are $\hat{a}_1 = 0.688(0.179)$, $\hat{a}_2 = -0.338(0.131)$ and $\hat{b}_1 = 0.808(0.189)$, where the standard errors of these estimates appear in parenthesis. All of these estimates are significant by t-test. Therefore, the identified ARMA model is $(1 - 0.69\mathcal{B} + 0.34\mathcal{B}^2)y_t = (1 - 0.8\mathcal{B})\varepsilon_t$. After consulting the list of EWMA closed-loop descriptions in Table 7.2, we speculate that the process transfer function is either $y_t = \beta_1 x_{t-1} + \beta_2 x_{t-2}$ or $y_t = \alpha + \beta x_{t-2}$. However, since for the latter alternative a_1 must be equal to 1 and our estimate $\hat{a}_1$ indicates this is not true, we conclude the process transfer function is $y_t = \beta_1 x_{t-1} + \beta_2 x_{t-2}$. Note that this model is a reasonable approximation of the true model, since the complete true parametric model is $(1 - 0.9\mathcal{B} + 0.4\mathcal{B}^2)y_t = (1 - \mathcal{B} - 0.02\mathcal{B}^2 - 0.10\mathcal{B}^3)\varepsilon_t$ if the controller and disturbance functions are substituted into the assumed process transfer function. Of course, the real process is never known to process engineers.

From the ARMA(2,1) model parameter estimates, the parameters in the identified transfer function and disturbance models are estimated as $\hat{\xi}_1 = 0.775$ (because, $1 - \lambda\hat{\xi}_1 = 0.69$), $\hat{\xi}_2 = 0.85$ (because, $-\lambda\hat{\xi}_2 = -0.34$), $\hat{\theta} = 0.8$, and $\hat{\delta} = 0$. Substituting them into the process equation, we can optimize the $AMSD(y_t)$ subject to the process stationarity conditions, that is,

$$Min \ \ AMSD(y_t) = \frac{1.64(1 + 0.85\lambda) - 1.6(1 - 0.775\lambda)}{1.625\lambda(1 - 0.85\lambda)(2 - 0.075\lambda)}$$

$$S.T. \ \ 0 < \lambda < 1.$$

Solution to this problem yields an optimal solution of λ equal to 0.12. By resetting λ in the EWMA controller and running the process with the same 190 disturbance data in

Squared Canonical Correlation Estimates

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	0.0086	0.1180	0.0697	0.0295	0.0047	0.0352
AR 1	0.1253	0.1062	0.0003	0.0243	0.0375	0.0185
AR 2	0.0479	0.0120	0.0219	0.0114	0.0005	0.0003
AR 3	0.0922	0.0253	0.0229	0.0120	0.0004	0.0005
AR 4	0.0129	0.0005	0.0030	0.0092	0.0175	0.0159
AR 5	0.0042	0.0010	<.0001	0.0088	<.0001	0.0027

Extended Sample Autocorrelation Function

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	0.0927	-0.3405	-0.2616	-0.1685	0.0670	0.1828
AR 1	0.2432	-0.2825	-0.0286	-0.2209	0.0734	0.1583
AR 2	-0.4617	-0.2453	0.0534	-0.1709	-0.0307	0.0221
AR 3	-0.4885	-0.4594	0.3032	-0.1962	-0.1993	0.0043
AR 4	-0.3301	-0.0885	-0.2734	-0.1840	0.0284	-0.0316
AR 5	-0.4328	0.0781	-0.0227	-0.2435	0.0078	-0.0696

Fig. 7.5.   SCAN and ESACF tables of the simulated example

Boyles under the re-tuned EWMA controller, we find that the estimated variance of the process output is reduced to 231.4. Note that this value is very close to the minimum variance one can achieve for the output of this process, namely, $\hat{\sigma}_y^2 = 228.0$, but the minimum variance can only be obtained when the correct ARMA model is identified and estimated perfectly.

7.4.1   Including the cost of adjustments

Sometimes the cost of adjustments cannot be ignored, so the objective function of the optimization model should be changed to a combination of that balances output variability and adjustment effort as proposed by Box and Luceño [18]:

$$Min \quad J = \frac{AMSD(y_t)}{\sigma_\varepsilon^2} + \rho\frac{Var(\nabla x_t)}{\sigma_\varepsilon^2} \tag{7.13}$$

For this example, Table 7.4 lists the optimal λ, the associated cost functions, the AMSD, and the adjustment variance for different values of ρ (which is a quantity defined by the user). Suppose from the table the value $\lambda = 0.08$ is chosen. If the process is controlled by such EWMA controller and the same disturbance data as in Boyles [20], we find that $\widehat{AMSD}(y_t)/\sigma_\varepsilon^2 = 1.0001$, and $\widehat{Var}(\nabla x_t)/\sigma_\varepsilon^2 = 0.0100$, which closely agree with the table.

7.5   Summary

In this chapter, a method for identifying a process operating under the actions of a feedback controller was proposed. This method works for processes regulated with PI

ρ	λ	J	$\dfrac{AMSD}{\sigma^2_\varepsilon}$	$\dfrac{Var(\nabla x_t)}{\sigma^2_\varepsilon}$
0	0.12175	1.0122	1.0122	0.0234
0.1	0.11934	1.0145	1.0122	0.0225
0.5	0.11154	1.0229	1.0130	0.0198
1	0.10441	1.0321	1.0148	0.0173
2	0.09460	1.0477	1.0192	0.0143
5	0.07886	1.0831	1.0329	0.0100

Table 7.4.   Optimal solutions to problem (7.13) in the example

or EWMA controllers under the assumption that the disturbance is IMA(1,1) with drift. It was shown that when the disturbance function is one from the proposed disturbance family, it is possible to identify some dynamic process models commonly encountered in manufacturing. ARMA models of the output deviations from these processes are provided. After identification, the approach suggests to tune the controller to a near-optimal setting according to a well-known performance criterion, which is either the AMSD of the process output, or a weighted sum of AMSD of the output and variance of the adjustments.

Chapter 8

Research Contributions and Future Work

This chapter summarizes the research contributions of this thesis and comments on some possible extensions of the current research. The contributions of this thesis are summarized in Section 7.1. Section 7.2 describes the further research on using Markov Chain Monte Carlo (MCMC) methods for setup error adjustments.

8.1 Research Contributions

The contributions of this thesis are summarized as follows:

- The setup adjustment problem was discussed and a new unifying point of view for this problem was presented based on a Kalman Filter approach. The Bayesian interpretation of the sequential adjustment procedure was given both for a single realization of the process and for a process ensemble. A connection between this procedure with the Linear Quadratic Gaussian (LQG) controller, a well-known method in the control engineering literature, and the Stochastic Approximation method was made.

- Small-sample properties of various process adjustment rules, including Grubbs' harmonic rule, an adjustment based on a Kalman Filter approach and an discrete integral (or EWMA) controller were analyzed in response to the concern of the performance of these rules on a short-run manufacturing process. The performance

metric used was a quadratic off-target quality cost. It was shown that if the setup error is an unknown constant, Grubbs' harmonic rule represents a better strategy than the Kalman Filter scheme (equivalent to Grubbs' extended rule). The even simpler integral controller with weight $\lambda = 0.2$ provides a competitive alternative to the harmonic rule for cases when the process offset is small. If the setup error is instead a random variable with known first and second moments, then the Kalman Filter scheme is optimal in the sense of minimizing the total quadratic quality cost. But when the first and second moments of the offset are unknown (a situation common in practice), the harmonic rule and the integral controller can often outperform the Kalman Filter scheme when the setup noise is relatively high and/or the offset is large on average. Based on these results, a quality engineer can select select his/her own process adjustment strategy according to the case he/she is dealing with.

- Sequential adjustment strategies were further studied beyond the application to the process setup adjustment problem. For on-line quality control, guarding the process from a step-type shift in the process mean is an important task. Such a shift can occur at any point in time, not necessarily at startup. The process upset is detected with an SPC chart. It was shown that sequential adjustments are superior to using one single adjustment. When the shift in the process mean is frequent, it is better to install an automatic process controller, such as the integral controller, to adjust the process constantly without monitoring the process mean. On the other hand, when the shift is infrequent (which is a valid assumption in many modern

industries), the integrated SPC/EPC scheme (with several sequential adjustments) has a better performance. Furthermore, such integrated scheme provides the opportunity of recording the time and frequency of shifts, thus the root cause of the shift in the process mean can be analyzed later on. For a process with i.i.d. noise, a CUSUM chart used in conjunction with the sequential adjustment method is recommended, while it was also pointed out that a hybrid Shewhart-CUSUM chart with sequential adjustments has an improved performance on large shift sizes while slightly sacrificing performance on small shift sizes. For a process with autocorrelated noise, the control chart limits are modified in order to give a run-length performance comparable to that of the chart for an i.i.d. process, following recommendation by Yashchin [118]. The performance of modified CUSUM charts with sequential adjustments was also evaluated for a variety of autocorrelated processes and compared with the performance of integral controllers. It was found that this integrated scheme has similar performance as that for the i.i.d. process, but as the process autocorrelation increases, the integral controller with a proper parameter λ becomes to dominate other adjustment methods.

- A modified sequential adjustment strategy for the asymmetric off-target quality cost was proposed. Asymmetric quality cost functions are associated with many manufacturing processes, for example, a machining process where materials are removed from the bulk part. In this case, a value of the quality characteristic above its target has a different cost implication from it below the target. Therefore, it is desirable that, during the process adjustment period, the likelihood of the value

of the quality characteristic falling into the lower cost region will be higher than the likelihood of the opposite case. It was shown that by adding a bias term into the sequential adjustment scheme, it is possible to let the process mean converge to its target from the low cost side. The bias term was computed for the cases of a constant asymmetric cost function and a quadratic asymmetric cost function. The performance of the biased adjustment rule was compared with that of Grubbs' harmonic rule and of integral controllers, and significant cost savings were found for short-run processes.

- Another modified sequential adjustment strategy was proposed when process measurement and process adjustment costs are significant. The adjustment rule skips several process runs after one adjustment has been executed, while still maintaining the optimality of the adjustment rule based on the Kalman Filter approach. The scheduling method of determining the optimal time instants for adjusting for a setup offset was discussed. In particular, the performance of three scheduling methods – one based on the Wagner-Whitin algorithm, one based on the Silver-Meal heuristic, and the other one based on a method due to Trietsch [101] – was compared. A backward implementation of the Silver-Meal heuristic was recommended and a slight modification of this method was shown to be robust to with respect to an unknown process variance. This strategy can be easily applied for on-line process control.

- In the final part of this thesis, a closed-loop process scenario was presented. In such case, an automatic controller is put on the process for stabilizing the process

instead of optimizing the process. In order to improve the process performance, it is necessary to obtain the process transfer function and process noise models. However, conventional open-loop process identification methods cannot be used because of the dependency of the process input on the output. Under certain assumptions on the process controller and noise models, a catalog of possible ARMA models of the closed-loop processes was given. These provide a mapping to the open-loop transfer functions. It was found that the information obtained from fitting ARMA models to the output obtained during closed-loop operation can be used to fine-tune the process controller to improve the process even if the true transfer function cannot be exactly obtained. Therefore, this controller updating method is suitable for an established process where overhauling the whole process is impractical.

In summary, this thesis discussed several statistical adjustment strategies for a variety of process abnormalities and manufacturing cost considerations. The adjustment tools can empower a quality engineer to actively conduct continuous process quality improvements and they also broaden the manufacturing applications of statistical and probabilistic principles.

8.2 Future Work

As discussed in this thesis, in a modern manufacturing environment products are often manufactured in short-run processes, i.e., similar parts are processed in small

batches. In such case, two sources of variability – "within-batch" variability and "between-batches" variability – become relevant. The between-batch and with-batch variances are also important for a quality engineer to determine the process capability of the current manufacturing process and to conduct actions for further quality improvements. Also, frequent process setups due to short runs may lead to a higher possibility of setup errors on the machine. The setup error can be modelled as a random variable. If its first and second moments are known and the variance of process is known exactly, it was shown in Chapter 3 that Grubbs' extended rule is optimal when the only cost of interest is a quadratic off-target quality cost. However, it was also shown in Chapter 3 that when no information on the moments of the setup offset is available, the simpler Grubbs' harmonic rule or an EWMA controller may have a better performance than the extended rule for minimizing the total off-target quality cost of short-run processes.

Future work can be undertaken to develop a process adjustment approach over a set of batches, where no previous knowledge on the setup offset distribution and on the process variance is available. The resulting adjustment procedure proposed would be very useful for the quality control of an experimental manufacturing process or a newly installed process. In this section we sketch the main idea behind such an adjustment procedure based on a hierarchical Bayesian model and the use of Markov Chain Monte Carlo (MCMC) methods.

Let y_{ij} be the value of quality characteristic of the process observed for batch i, part j, where $i = 1, ..., I$ and $j = 1, ..., J$. This quality characteristic is measured with reference to a target value T which is assumed, without loss of generality, equal to zero. The initial (unknown) process offset of the ith batch is denoted by d_i . It is assumed that

this offset is a random occurrence from a normal distribution with mean μ and variance σ^2_d, i.e.,

$$d_i | \mu, \sigma^2_d \sim N(\mu, \sigma^2_d) \quad . \tag{8.1}$$

where both μ and σ^2_d are assumed unknown. Here, μ can be seen as a common offset which is due to some systematic setup error that is of interest to be removed. The process model is

$$y_{ij} = d_i + x_{i(j-1)} + \varepsilon_{ij} \quad , \tag{8.2}$$

where ε_{ij} is process noise with 0 mean and unknown variance σ^2_ε. Considering that at the time y_{ij} is observed, $x_{i(j-1)}$ is known, a different variable z_{ij} can be derived as:

$$z_{ij} = y_{ij} - x_{i(j-1)} = d_i + \varepsilon_{ij} \quad . \tag{8.3}$$

This allows to derive an analogy with a one-way random effects model, i.e.,

$$\begin{aligned} z_{ij} | d_i, \sigma^2_\varepsilon &\sim N(d_i, \sigma^2_\varepsilon) \quad , \\ d_i | \mu, \sigma^2_d &\sim N(\mu, \sigma^2_d) \quad . \end{aligned} \tag{8.4}$$

Finally, the adjustment can be written as

$$x_{ij} = -\widehat{d}_i | \mathbf{x^{ij}} \quad , \tag{8.5}$$

where $\mathbf{x^{ij}} = \{x_{11}, x_{12}, ..., x_{1J}, ..., x_{i1}, ..., x_{ij}\}$ represent all (transformed) data observed at the time the estimate of d_i is computed and $\widehat{d}_i | \mathbf{x^{ij}}$ represents the estimate of the ith

mean in the random effect model, given all available data. As clear from the last equation, selecting the adjustments at each step involves means in an unbalanced one-way random effect model.

From a Bayesian perspective, the one-way random effects model is a special case of a *hierarchical model*, used in describing multi-parameter problems in which parameters are related with some structure, depending on the specific problem addressed (Box and Tiao [19]). The choice of the number of levels or stages in the hierarchy is strictly related with the problem addressed, although most of the applications, as the one way random-effect model, require a three-level model (Carlin and Louis [21]).

In a three-stage hierarchical model, the first stage in the hierarchy represents the observed data, conditionally on a set of unknown parameters. The second stage of the model specifies the distributions of the parameters. A traditional Bayesian model can be considered as a hierarchical model consisting of these two stages, in which the second stage is constituted by priors on unknown parameters. However, when at least one of the parameters in the second stage has a probabilistic specification which depends on a further set of unknown parameters, called *hyperparameters*, a third stage in the hierarchy is required.

For the problem of adjustments in a batch-to-batch process with initial process setup offsets, the first stage of the hierarchy models the distribution of the observations conditionally on unknown parameters as given by the first equation in (8.4). The second stage in the hierarchy specifies the distribution of the current process mean d_i and the between-batch variance σ_ε^2. The parameter σ_ε^2 does not have any further hierarchical structure. Adopting conjugacy at each step of the hierarchical model (a common choice

Fig. 8.1.   The hierarchical model of data, parameters and hyperparameters

for the random effects model, Gelfand *et al.* [42]; Gelfand and Smith, [41]), the prior distribution for σ^2_ε is given by:

$$\sigma^2_\varepsilon | a_2, b_2 \sim IG(a_2, b_2) \ , \tag{8.6}$$

where IG represents an Inverse-Gamma distribution and a_2, b_2 are assumed known and are typically very small numbers to model "vague" prior information (Spiegelhalter *et al.* [92]). Another unknown parameter modelled at the second stage in the hierarchy is the initial off-set d_i, given by equation (8.1). Its distribution is given conditionally on other two random parameters μ and σ^2_d. These *hyperparameters* are modelled at the third stage in the hierarchy. Adopting once again conjugacy, priors on these hyperparameters

are given by:

$$\mu|\mu_0, \sigma^2_0 \quad \sim \quad N(\mu_0, \sigma^2_0) \quad , \tag{8.7}$$

$$\sigma^2_d|a_1, b_1 \quad \sim \quad IG(a_1, b_1) \quad , \tag{8.8}$$

where $\mu_0, \sigma^2_0, a_1, b_1$ are again assumed known, but "vague".

The traditional inference procedure on a random effects model is performed by the analysis of variance (ANOVA). However, ANOVA is not appropriate for the problem under study, where data are collected from a short-run process, since ANOVA requires considerable data. Furthermore, it is possible that the variance components estimated by ANOVA be negative. The Bayesian method is a natural alternative to this problem, since the Bayesian estimator is based on process data and prior distribution only. The Bayesian approach also permits to easily tackle the lack of normality and/or independence and possible heterogeneity of variances (Box and Tiao [19]).

In future research, a Bayesian approach could be developed for setting adjustments on the batch by batch production. In particular, Markov Chain Monte Carlo (MCMC) techniques, which perform Monte Carlo integration using Markov Chains, can be utilized. A Gibbs sampler can be applied on-line to predict the process setup offset of each batch and to suggest an adjustment accordingly prior to the first part of each batch. Although a Bayesian approach allows to overcome difficulties induced by sampling theory and to derive easy extensions to more complex problems, this approach requires intensive computational efforts in calculation of marginal posterior densities. Therefore, how

to simplify the computation and how to automate this approach for on-line process adjustment application are challenges worth of further study.

Appendix A

The Linear Quadratic Setup Adjustment Problem

In this appendix we give the solution to a general linear quadratic setup adjustment problem. It corresponds to the classic LQG control problem with incomplete state information, and it is derived, for example, in Åström [94]. Suppose there are p controllable factors (machine setpoints) and p outputs modelled as deviations from target. Assume the process is described by the linear equation

$$\boldsymbol{\mu}_{\mathbf{t}} = \boldsymbol{\mu}_{\mathbf{t-1}} + \nabla\mathbf{x}_{\mathbf{t-1}} + \mathbf{w}_{\mathbf{t-1}}, \quad \boldsymbol{\mu}_{\mathbf{1}} \sim \mathbf{N}(\overline{\mathbf{d}}, \boldsymbol{\Sigma}_{\mathbf{d}}), \quad \mathbf{w}_{\mathbf{t}} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{w}})$$

and

$$\mathbf{y}_{\mathbf{t}} = \boldsymbol{\mu}_{\mathbf{t}} + \boldsymbol{\varepsilon}_{\mathbf{t}}, \quad \boldsymbol{\varepsilon}_{\mathbf{t}} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon}).$$

The criterion to minimize is quadratic and equals:

$$E\left[\boldsymbol{\mu}_{\mathbf{n}}' \mathbf{Q}\boldsymbol{\mu}_{\mathbf{n}} + \sum_{\mathbf{t=1}}^{\mathbf{n-1}} (\boldsymbol{\mu}_{\mathbf{t}}' \mathbf{Q}\boldsymbol{\mu}_{\mathbf{t}} + \nabla\mathbf{x}_{\mathbf{t}}' \mathbf{R}\nabla\mathbf{x}_{\mathbf{t}})\right].$$

The optimal solution is as follows (Åström [94], Lewis [61]):

1. Let $S_n = Q$. Compute (off-line) for $t = n-1, n-2, ..., 1$ the controller gain matrix L_t as follows:

$$L_t = (S_{t+1} + R)^{-1} S_{t+1}$$

where

$$S_t = S_{t+1} - L_t'(S_{t+1} + R)L_t + Q.$$

2. For $t = 1, 2, ..., n$ compute (can be done off-line as well) the Kalman weight matrices

K_t as follows:

$$K_t = (P_{t-1} + \Sigma_w)(P_{t-1} + \Sigma_w + \Sigma_\varepsilon)^{-1}$$

where

$$P_t = [I - (P_{t-1} + \Sigma_w + \Sigma_\varepsilon)^{-1}](P_{t-1} + \Sigma_w)'$$

with $P_0 = \Sigma_d$.

3. Let $\widehat{\mu_1} = \overline{d} + x_0$. Compute (on-line), for $t = 1, 2, ..., n - 1$ the adjustments as

follows:

$$\nabla x_t = -L_t \widehat{\mu_t}$$

where

$$\widehat{\mu_t} = \widehat{\mu_{t-1}} + \nabla x_{t-1} + K_t(y_t - (\widehat{\mu_{t-1}} + \nabla x_{t-1})).$$

From the LQG model, Grubbs' extended rule is obtained by letting $Q = 1, R = 0, \Sigma_w = 0$.

Appendix B

Additional Tables on the Performance of Different Adjustment Schemes for Autocorrelation Processes

% improvement on AISD		Mean of shift size				
		0	1σ	2σ	3σ	4σ
p=0.05	CUSUM chart + Single adj.	9.83 (0.35)	35.18 (0.40)	60.29 (0.37)	72.76 (0.32)	78.24 (0.31)
	CUSUM chart + Sequential adj.	17.97 (0.29)	41.12 (0.36)	63.60 (0.33)	74.69 (0.30)	**79.54** (0.29)
	EWMA controller ($\lambda = 0.1$)	24.62 (0.26)	43.16 (0.31)	59.84 (0.30)	67.36 (0.28)	69.99 (0.27)
	EWMA controller ($\lambda = 0.2$)	**25.53** (0.29)	**46.13** (0.35)	64.98 (0.33)	73.65 (0.30)	76.85 (0.29)
	EWMA controller ($\lambda = 0.3$)	23.98 (0.31)	45.74 (0.37)	**65.83** (0.34)	**75.16** (0.32)	78.70 (0.31)
p=0.035	CUSUM chart + Single adj.	5.26 (0.32)	23.34 (0.39)	46.17 (0.40)	58.99 (0.39)	65.98 (0.39)
	CUSUM chart + Sequential adj.	12.94 (0.26)	29.42 (0.34)	49.85 (0.37)	**61.44** (0.37)	**67.71** (0.37)
	EWMA controller ($\lambda = 0.1$)	**18.69** (0.25)	32.16 (0.31)	47.58 (0.34)	55.35 (0.34)	59.17 (0.34)
	EWMA controller ($\lambda = 0.2$)	18.48 (0.28)	**33.73** (0.35)	51.27 (0.38)	60.37 (0.38)	64.95 (0.38)
	EWMA controller ($\lambda = 0.3$)	16.26 (0.30)	32.51 (0.37)	**51.30** (0.40)	61.17 (0.40)	66.20 (0.40)
p=0.02	CUSUM chart + Single adj.	0.63 (0.28)	10.46 (0.34)	26.59 (0.39)	38.98 (0.43)	44.23 (0.45)
	CUSUM chart + Sequential adj.	7.16 (0.22)	16.24 (0.29)	30.65 (0.37)	**42.12** (0.41)	**46.77** (0.43)
	EWMA controller ($\lambda = 0.1$)	**11.13** (0.22)	**18.84** (0.28)	29.81 (0.34)	38.06 (0.37)	40.45 (0.38)
	EWMA controller ($\lambda = 0.2$)	9.58 (0.25)	18.38 (0.32)	**31.17** (0.38)	40.86 (0.42)	43.80 (0.44)
	EWMA controller ($\lambda = 0.3$)	6.55 (0.27)	15.99 (0.34)	29.82 (0.41)	40.37 (0.45)	43.64 (0.47)
p=0.005	CUSUM chart + Single adj.	-4.27 (0.22)	-1.64 (0.24)	3.45 (0.29)	7.94 (0.34)	10.81 (0.37)
	CUSUM chart + Sequential adj.	0.16 (0.13)	2.59 (0.18)	**7.34** (0.25)	**11.26** (0.30)	**13.86** (0.33)
	EWMA controller ($\lambda = 0.1$)	**1.16** (0.13)	**3.41** (0.17)	7.11 (0.23)	9.81 (0.27)	11.21 (0.29)
	EWMA controller ($\lambda = 0.2$)	-2.11 (0.16)	0.48 (0.20)	4.83 (0.27)	8.10 (0.31)	9.82 (0.34)
	EWMA controller ($\lambda = 0.3$)	-6.16 (0.17)	-3.39 (0.22)	1.32 (0.29)	4.97 (0.34)	6.84 (0.37)

Table B.1. Performance of integrated SPC/EPC adjustment schemes and an EWMA adjustment scheme on an AR(1) process with $\phi = 0.15$ when varying the probability of a shift.

% improvement on AISD		Mean of shift size				
		0	1σ	2σ	3σ	4σ
p=0.05	CUSUM chart + Single adj.	10.38 (0.36)	35.49 (0.40)	60.14 (0.37)	72.66 (0.32)	78.99 (0.30)
	CUSUM chart + Sequential adj.	18.60 (0.30)	41.22 (0.36)	63.46 (0.34)	**74.64** (0.30)	**80.31** (0.29)
	EWMA controller ($\lambda = 0.1$)	24.81 (0.26)	42.62 (0.31)	59.27 (0.30)	66.83 (0.28)	70.40 (0.27)
	EWMA controller ($\lambda = 0.2$)	**25.36** (0.30)	**45.31** (0.35)	64.26 (0.33)	73.00 (0.31)	77.20 (0.29)
	EWMA controller ($\lambda = 0.3$)	23.40 (0.32)	44.62 (0.37)	**64.94** (0.35)	74.39 (0.32)	78.96 (0.31)
p=0.035	CUSUM chart + Single adj.	5.41 (0.32)	23.99 (0.39)	46.83 (0.40)	59.73 (0.39)	66.19 (0.39)
	CUSUM chart + Sequential adj.	12.87 (0.27)	29.83 (0.35)	50.74 (0.37)	**62.20** (0.37)	**68.01** (0.37)
	EWMA controller ($\lambda = 0.1$)	**18.24** (0.25)	**31.82** (0.32)	47.82 (0.34)	55.52 (0.34)	59.04 (0.34)
	EWMA controller ($\lambda = 0.2$)	17.54 (0.28)	32.97 (0.36)	**51.41** (0.38)	60.46 (0.38)	64.67 (0.38)
	EWMA controller ($\lambda = 0.3$)	14.77 (0.31)	31.31 (0.38)	51.25 (0.40)	61.12 (0.40)	65.74 (0.40)
p=0.02	CUSUM chart + Single adj.	0.90 (0.28)	11.45 (0.35)	28.46 (0.40)	38.92 (0.43)	45.26 (0.45)
	CUSUM chart + Sequential adj.	7.35 (0.23)	17.08 (0.30)	**32.51** (0.37)	**42.01** (0.41)	**47.59** (0.43)
	EWMA controller ($\lambda = 0.1$)	**11.14** (0.22)	**18.98** (0.29)	30.93 (0.35)	37.31 (0.37)	40.61 (0.39)
	EWMA controller ($\lambda = 0.2$)	9.08 (0.25)	18.09 (0.33)	32.04 (0.39)	39.73 (0.42)	43.75 (0.44)
	EWMA controller ($\lambda = 0.3$)	5.47 (0.28)	15.22 (0.35)	30.37 (0.42)	38.85 (0.45)	43.33 (0.47)
p=0.005	CUSUM chart + Single adj.	-4.04 (0.21)	-1.97 (0.25)	3.09 (0.30)	7.89 (0.35)	10.21 (0.37)
	CUSUM chart + Sequential adj.	0.08 (0.14)	2.39 (0.18)	**7.01** (0.25)	**11.37** (0.31)	**13.36** (0.34)
	EWMA controller ($\lambda = 0.1$)	**0.45** (0.13)	**2.65** (0.17)	6.42 (0.23)	9.35 (0.27)	10.38 (0.29)
	EWMA controller ($\lambda = 0.2$)	-3.49 (0.16)	-0.88 (0.20)	3.63 (0.27)	7.09 (0.32)	8.50 (0.34)
	EWMA controller ($\lambda = 0.3$)	-8.24 (0.17)	-5.38 (0.22)	-0.44 (0.29)	3.37 (0.35)	5.02 (0.37)

Table B.2. Performance of integrated SPC/EPC adjustment schemes and an EWMA adjustment scheme on an AR(1) process with $\phi = 0.1$ when varying the probability of a shift.

% improvement on AISD		Mean of shift size				
		0	1σ	2σ	3σ	4σ
p=0.05	CUSUM chart + Single adj.	9.81 (0.35)	35.07 (0.40)	60.58 (0.36)	71.83 (0.33)	78.45 (0.30)
	CUSUM chart + Sequential adj.	18.12 (0.29)	40.75 (0.35)	63.76 (0.33)	73.83 (0.31)	**79.71** (0.29)
	EWMA controller ($\lambda = 0.1$)	24.69 (0.26)	42.43 (0.31)	59.74 (0.30)	66.38 (0.28)	70.06 (0.27)
	EWMA controller ($\lambda = 0.2$)	**25.47** (0.29)	**45.22** (0.35)	64.84 (0.33)	72.54 (0.31)	76.79 (0.30)
	EWMA controller ($\lambda = 0.3$)	23.80 (0.31)	44.66 (0.37)	**65.66** (0.35)	**73.97** (0.32)	78.57 (0.31)
p=0.035	CUSUM chart + Single adj.	5.59 (0.32)	23.96 (0.39)	46.32 (0.40)	59.07 (0.39)	65.93 (0.39)
	CUSUM chart + Sequential adj.	13.09 (0.26)	30.04 (0.34)	50.02 (0.37)	**61.62** (0.37)	**67.73** (0.37)
	EWMA controller ($\lambda = 0.1$)	**18.52** (0.25)	32.35 (0.31)	47.34 (0.34)	55.50 (0.34)	59.17 (0.34)
	EWMA controller ($\lambda = 0.2$)	18.12 (0.28)	**33.80** (0.35)	**50.91** (0.38)	60.39 (0.38)	64.80 (0.38)
	EWMA controller ($\lambda = 0.3$)	15.73 (0.30)	32.48 (0.38)	50.84 (0.40)	61.11 (0.40)	65.95 (0.40)
p=0.02	CUSUM chart + Single adj.	0.63 (0.28)	10.77 (0.34)	27.96 (0.40)	38.54 (0.43)	45.71 (0.45)
	CUSUM chart + Sequential adj.	7.13 (0.22)	16.50 (0.30)	32.33 (0.37)	**41.86** (0.40)	**47.94** (0.43)
	EWMA controller ($\lambda = 0.1$)	**11.04** (0.22)	**18.83** (0.28)	31.10 (0.34)	37.46 (0.37)	41.40 (0.39)
	EWMA controller ($\lambda = 0.2$)	9.28 (0.25)	18.26 (0.32)	**32.52** (0.39)	40.03 (0.42)	44.64 (0.44)
	EWMA controller ($\lambda = 0.3$)	6.07 (0.27)	15.76 (0.35)	31.19 (0.41)	39.37 (0.45)	44.39 (0.47)
p=0.005	CUSUM chart + Single adj.	-4.00 (0.22)	-1.53 (0.24)	3.28 (0.29)	7.36 (0.34)	10.09 (0.37)
	CUSUM chart + Sequential adj.	0.28 (0.14)	2.57 (0.18)	**7.07** (0.24)	**10.98** (0.30)	**13.35** (0.33)
	EWMA controller ($\lambda = 0.1$)	**1.16** (0.14)	**3.06** (0.17)	6.47 (0.23)	9.35 (0.27)	10.78 (0.29)
	EWMA controller ($\lambda = 0.2$)	-2.28 (0.16)	-0.08 (0.20)	4.01 (0.26)	7.43 (0.31)	9.19 (0.34)
	EWMA controller ($\lambda = 0.3$)	-6.48 (0.18)	-4.13 (0.22)	0.37 (0.29)	4.10 (0.34)	6.08 (0.37)

Table B.3. Performance of integrated SPC/EPC adjustment schemes and an EWMA adjustment scheme on a MA(1) process with $\theta = -0.15$ when varying the probability of a shift.

% improvement on AISD		Mean of shift size				
		0	1σ	2σ	3σ	4σ
p=0.05	CUSUM chart +	9.50	35.37	60.59	72.70	78.31
	Single adj.	(0.35)	(0.40)	(0.36)	(0.32)	(0.31)
	CUSUM chart +	17.64	41.15	63.72	**74.53**	**79.56**
	Sequential adj.	(0.29)	(0.36)	(0.34)	(0.30)	(0.29)
	EWMA controller	23.89	42.47	59.52	66.68	69.73
	($\lambda = 0.1$)	(0.26)	(0.31)	(0.30)	(0.28)	(0.28)
	EWMA controller	**24.27**	**45.12**	64.43	72.79	76.36
	($\lambda = 0.2$)	(0.29)	(0.35)	(0.33)	(0.31)	(0.30)
	EWMA controller	22.17	44.37	**65.04**	74.13	78.03
	($\lambda = 0.3$)	(0.31)	(0.37)	(0.35)	(0.33)	(0.32)
p=0.035	CUSUM chart +	5.81	23.66	46.31	59.41	66.37
	Single adj.	(0.32)	(0.39)	(0.40)	(0.39)	(0.39)
	CUSUM chart +	13.66	29.74	50.19	**61.88**	**68.11**
	Sequential adj.	(0.27)	(0.35)	(0.37)	(0.37)	(0.37)
	EWMA controller	**18.91**	31.73	47.43	55.18	59.03
	($\lambda = 0.1$)	(0.25)	(0.31)	(0.34)	(0.34)	(0.34)
	EWMA controller	18.27	**32.79**	**50.85**	60.02	64.62
	($\lambda = 0.2$)	(0.28)	(0.35)	(0.38)	(0.38)	(0.38)
	EWMA controller	15.54	31.08	50.58	60.61	65.66
	($\lambda = 0.3$)	(0.31)	(0.38)	(0.40)	(0.41)	(0.40)
p=0.02	CUSUM chart +	0.73	11.58	28.45	39.46	45.74
	Single adj.	(0.28)	(0.34)	(0.40)	(0.43)	(0.45)
	CUSUM chart +	7.16	17.17	**32.67**	**42.48**	**48.11**
	Sequential adj.	(0.22)	(0.30)	(0.37)	(0.41)	(0.43)
	EWMA controller	**10.84**	**18.91**	31.22	37.66	41.03
	($\lambda = 0.1$)	(0.22)	(0.29)	(0.34)	(0.37)	(0.39)
	EWMA controller	8.66	17.97	32.34	40.09	44.11
	($\lambda = 0.2$)	(0.25)	(0.33)	(0.39)	(0.42)	(0.44)
	EWMA controller	4.99	15.04	30.67	39.23	43.65
	($\lambda = 0.3$)	(0.27)	(0.35)	(0.42)	(0.45)	(0.47)
p=0.005	CUSUM chart +	-4.19	-1.89	3.38	7.83	10.46
	Single adj.	(0.21)	(0.25)	(0.29)	(0.34)	(0.37)
	CUSUM chart +	0.06	**2.60**	**7.34**	**11.27**	**13.47**
	Sequential adj.	(0.14)	(0.18)	(0.25)	(0.30)	(0.34)
	EWMA controller	**0.40**	2.56	6.30	9.14	10.47
	($\lambda = 0.1$)	(0.13)	(0.17)	(0.23)	(0.27)	(0.29)
	EWMA controller	-3.57	-1.08	3.38	6.83	8.44
	($\lambda = 0.2$)	(0.15)	(0.20)	(0.27)	(0.31)	(0.34)
	EWMA controller	-8.37	-5.67	-0.76	3.07	4.85
	($\lambda = 0.3$)	(0.17)	(0.22)	(0.29)	(0.34)	(0.37)

Table B.4.   Performance of integrated SPC/EPC adjustment schemes and an EWMA adjustment scheme on a MA(1) process with $\theta = -0.1$ when varying the probability of a shift.

Appendix C

Equivalence Between Minimization of the $AIEC^{\bullet}$ criterion and Minimization of Each $E(C_t^{\bullet})$

Consider the minimization problem

$$Min_{\boldsymbol{\mu}} \; AIEC^{\bullet}$$

where $\boldsymbol{\mu} = \{\mu_t, \; t = 1, ..., N\}$ is the $N \times 1$ vector of the means of the quality characteristic and $AIEC^{\bullet}$ is given by (5.5). The assumption of a linear feedback adjustment rule induces an affine relation among the means of the response variable, which can be generally expressed as $\boldsymbol{\mu} = \mathbf{R}\boldsymbol{\mu} + \mathbf{s}$, where $\mathbf{R}$ is a $N \times N$ matrix and $\mathbf{s}$ is a $N \times 1$ vector. In particular the mean at the tth step μ_t can be written as:

$$\mu_t = \sum_{i=1}^{N} r_{ti} \mu_i + s_t \tag{C.1}$$

where r_{ti} is the entry in row t and column i of the matrix $\mathbf{R}$ and s_t is the tth component of $\mathbf{s}$. In particular, the mean at each step is a function only of the previous ones, therefore $r_{ti} = 0$ for $i \geq t$.

To minimize $AIEC^{\bullet}$ the first order condition consists in equating to zero all the components of the gradient vector, i.e.:

$$\nabla AIEC^{\bullet} = \mathbf{0} \; ,$$

where $\mathbf{0}$ is a $N \times 1$ vector of zeros. Considering the expression of $AIEC^{\bullet}$ given in Chapter 5, the ith component of the gradient can be rewritten as:

$$\frac{\partial AIEC^{\bullet}}{\partial \mu_i} = \frac{1}{N} \sum_{t=1}^{N} \frac{\partial E(C_t^{\bullet})}{\partial \mu_i} = \frac{1}{N} \sum_{t=1}^{N} \frac{\partial E(C_t^{\bullet})}{\partial \mu_t} \frac{\partial \mu_t}{\partial \mu_i} = \frac{1}{N} \sum_{t=1}^{N} \frac{\partial E(C_t^{\bullet})}{\partial \mu_t} r_{ti} .$$

Therefore, the first order condition is satisfied when

$$\frac{\partial E(C_t^{\bullet})}{\partial \mu_t} = 0 \ , \ t = 1, 2, ..., N \ . \tag{C.2}$$

The second order condition can be determined by considering two theorems, derived by extending to strictly convex functions results reported in (Bazaraa $et\ al.$ [10]) for convex functions.

Theorem 1

Let $f_1, f_2, ..., f_k : E_n \rightarrow E_1$ be strictly convex functions. Then, the function f defined as $f(\mathbf{x}) = \sum_{j=1}^{k} \alpha_j f_j(\mathbf{x})$, where $\alpha_j > 0$ for $j = 1, ..., k$ is strictly convex.

Theorem 2

Let $g : E_m \rightarrow E_1$ be a strictly convex function and let $\mathbf{h}: E_n \rightarrow E_m$ be an affine function of the form $\mathbf{h}(\mathbf{x}) = \mathbf{Ax} + \mathbf{b}$, where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b}$ is an $m \times 1$ vector. Then, the composite function $f : E_n \rightarrow E_1$, defined as $f(\mathbf{x}) = g[\mathbf{h}(\mathbf{x})]$, is strictly convex.

Consider that $AIEC^{\bullet}$ can be seen as a linear combination of $E(C_t^{\bullet})$ with weights $1/N$. Therefore the first theorem allows to assert that $AIEC^{\bullet}$ is strictly convex when each component $E(C_t^{\bullet})$ is also a strictly convex function of the vector of means μ. On the other hand, the expected cost at time n, $E(C_t^{\bullet})$, is a composite function, since it is

directly related only to one component of the vector, namely μ_t, which in turn depends on the whole vector μ through an affine relation (given by equation C.1). Therefore, considering the second theorem, the strictly convexity of $E(C_t^\bullet)$ as a function of the whole vector μ, is proved once it is showed that $E(C_t^\bullet)$ is a strictly convex function of the scalar μ_t. Merging the results from the first and second theorems, the second order condition can be stated as:

$$\frac{\partial^2 E(C_t^\bullet)}{\partial^2 \mu_t} > 0 \ . \tag{C.3}$$

This condition implies that $AIEC^\bullet$ is a strictly convex function, thus characterized by a unique and global minimum. Considering both the first and the second order conditions, given respectively by (C.2) and (C.3), the minimization of $AIEC^\bullet$ considered can be replaced by the following set of minimization problems:

$$\min_{\mu_t} E(C_t^\bullet) \ , \ t = 1, 2, ..., N \ .$$

Appendix D

Minimization of $E(C_t^c)$ for the Asymmetric Constant Cost Function

In the case of the asymmetric constant cost function

$$E(C_t^c) = c_1^c \Phi\left(\frac{LSL - \mu_t}{\sigma_t}\right) + c_2^c\left[1 - \Phi\left(\frac{USL - \mu_t}{\sigma_t}\right)\right] \ ,$$

taking the first derivative with respect to μ_t and equating it to zero, we get

$$\frac{\partial}{\partial \mu_t} E(C_t^c) = -\frac{c_1^c}{\sigma_t}\phi\left(\frac{LSL - \mu_t}{\sigma_t}\right) + \frac{c_2^c}{\sigma_t}\phi\left(\frac{USL - \mu_t}{\sigma_t}\right) = 0 \ .$$

Therefore, the condition for the optimal target at time t, m_t^c, is given by

$$\frac{\phi\left(\frac{USL - m_t^c}{\sigma_t}\right)}{\phi\left(\frac{LSL - m_t^c}{\sigma_t}\right)} = \frac{c_1^c}{c_2^c} \ . \tag{D.1}$$

Considering the analytical expression of the normal density $\phi(\cdot)$, equation (D.1) can be rewritten as:

$$\frac{\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{USL-m_t^c}{\sigma_t}\right)^2\right]}{\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{LSL-m_t^c}{\sigma_t}\right)^2\right]} = \exp\left\{-\frac{1}{2}\left[\left(\frac{USL-m_t^c}{\sigma_t}\right)^2 - \left(\frac{LSL-m_t^c}{\sigma_t}\right)^2\right]\right\} = \frac{c_1^c}{c_2^c}$$

(D.2)

By taking the logarithm on both sides of equation (D.2), the closed form expression of the optimal target m_t^c can be obtained as follows:

$$\frac{USL^2 + m_t^{c\,2} - 2\,USL\,m_t^c - LSL^2 - m_t^{c\,2} + 2\,LSL\,m_t^c}{\sigma_t^2} = -2\ln\left(\frac{c_1^c}{c_2^c}\right)$$

$$USL^2 - LSL^2 - 2(USL - LSL)m_t^c = -2\sigma_t^2\ln\left(\frac{c_1^c}{c_2^c}\right)$$

$$m_t^c = \frac{2\sigma_t^2\ln(\frac{c_1^c}{c_2^c}) + USL^2 - LSL^2}{2(USL - LSL)} = \frac{\sigma_t^2\ln(\frac{c_1^c}{c_2^c})}{(USL - LSL)} + \frac{1}{2}(USL + LSL)\,.$$

(D.3)

In order to evaluate if the optimal mean m_t^c obtained determines a minimum of the cost function, the second order derivative has to be considered:

$$
\begin{aligned}
\frac{\partial^2}{\partial \mu_t^2} E(C_t^c) &= \frac{\partial}{\partial \mu}\left[-\frac{c_1^c}{\sigma_t}\phi\left(\frac{LSL - \mu_t}{\sigma_t}\right) + \frac{c_2^c}{\sigma_t}\phi\left(\frac{USL - \mu_t}{\sigma_t}\right)\right] \\
&= \frac{c_1^c}{\sigma_t^3}(\mu_t - LSL)\phi\left(\frac{LSL - \mu_t}{\sigma_t}\right) + \frac{c_2^c}{\sigma_t^3}(USL - \mu_t)\phi\left(\frac{USL - \mu_t}{\sigma_t}\right)
\end{aligned}
\tag{D.4}
$$

As it can be observed, this is always greater than zero as long as the condition $LSL < \mu_t < USL$ is satisfied.

Appendix E

Minimization of $E(C_t^c)$ for the Asymmetric Quadratic Cost Function

Consider the expected cost at time t given by:

$$E(C_t^q) = c_1^q \int_{-\infty}^{0} y_t^2 f_N(y_t; \mu_t, \sigma_t^2) dy_t + c_2^q \int_{0}^{\infty} y_t^2 f_N(y_t; \mu_t, \sigma_t^2) dy_t \ . \qquad (E.1)$$

Since c_1^q and c_2^q are constants, the expression of the expected value of cost at time t is completely defined by solving the generic integral:

$$\int_{a}^{b} y^2 f_N(y; \mu, \sigma^2) dy = \frac{1}{\sqrt{2\pi}\sigma} \int_{a}^{b} y^2 \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] dy \ . \qquad (E.2)$$

Let $z = \frac{y-\mu}{\sigma}$, thus $y = \mu + \sigma z \rightarrow dy = \sigma dz$, $y = a \rightarrow z = \frac{a-\mu}{\sigma} = c$ and $y = b \rightarrow z = \frac{b-\mu}{\sigma} = d$. Hence, the integral in (E.2) can be rewritten as:

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{c}^{d} (\mu + \sigma z)^2 \exp\left(-\frac{z^2}{2}\right) \sigma dz = \qquad (E.3)$$

$$\frac{1}{\sqrt{2\pi}} \left\{ \mu^2 \int_{c}^{d} \exp\left(-\frac{z^2}{2}\right) dz + \sigma^2 \int_{c}^{d} z^2 \exp\left(-\frac{z^2}{2}\right) dz + 2\mu\sigma \int_{c}^{d} z \exp\left(-\frac{z^2}{2}\right) dz \right\} \ .$$

The first term on the right hand side of (E.3) can be simply calculated as:

$$\frac{\mu^2}{\sqrt{2\pi}} \int_c^d \exp\left(-\frac{z^2}{2}\right) dz = \mu^2 \left[\Phi(d) - \Phi(c)\right] \ ,$$

where $\Phi(\cdot)$ represents the cumulative standard normal distribution function. The second term can be evaluated integrating by parts as follows:

$$
\begin{aligned}
\frac{\sigma^2}{\sqrt{2\pi}} \int_c^d z^2 \exp\left(-\frac{z^2}{2}\right) dz &= -\frac{\sigma^2}{\sqrt{2\pi}} \int_c^d z(-z) \exp\left(-\frac{z^2}{2}\right) dz \\
&= -\frac{\sigma^2}{\sqrt{2\pi}} \int_c^d z \left[\frac{d}{dz} \exp\left(-\frac{z^2}{2}\right)\right] dz \\
&= -\frac{\sigma^2}{\sqrt{2\pi}} \left[z \exp\left(-\frac{z^2}{2}\right)\right]_c^d + \frac{\sigma^2}{\sqrt{2\pi}} \int_c^d \exp\left(-\frac{z^2}{2}\right) dz \\
&= -\frac{\sigma^2}{\sqrt{2\pi}} \left[d \exp\left(-\frac{d^2}{2}\right) - c \exp\left(-\frac{c^2}{2}\right)\right] + \sigma^2 \left[\Phi(d) - \Phi(c)\right] \ .
\end{aligned}
$$

Finally, the third term in (E.3) can be computed as follows:

$$
\begin{aligned}
\frac{2\mu\sigma}{\sqrt{2\pi}} \int_c^d z \exp\left(-\frac{z^2}{2}\right) dz &= -\frac{2\mu\sigma}{\sqrt{2\pi}} \int_c^d (-z) \exp\left(-\frac{z^2}{2}\right) dz \\
&= -\frac{2\mu\sigma}{\sqrt{2\pi}} \int_c^d \left[\frac{d}{dz} \exp\left(-\frac{z^2}{2}\right)\right] dz \\
&= -\frac{2\mu\sigma}{\sqrt{2\pi}} \left[\exp\left(-\frac{z^2}{2}\right)\right]_c^d \\
&= -\frac{2\mu\sigma}{\sqrt{2\pi}} \left[\exp\left(-\frac{d^2}{2}\right) - \exp\left(-\frac{c^2}{2}\right)\right] = -2\mu\sigma \left[\phi(d) - \phi(c)\right] \ .
\end{aligned}
$$

Therefore:

$$\int_a^b y^2 f_N(y;\mu,\sigma^2)dy \;=\; (\mu^2+\sigma^2)\left[\Phi(d)-\Phi(c)\right]+ \tag{E.4}$$

$$-\frac{\sigma^2}{\sqrt{2\pi}}\left[d\exp\left(-\frac{d^2}{2}\right)-c\exp\left(-\frac{c^2}{2}\right)\right]+-2\mu\sigma\left[\phi(d)-\phi(c)\right]\;,$$

where $\frac{a-\mu}{\sigma}=c$ and $\frac{b-\mu}{\sigma}=d$. With this result, the first integral in (E.1) can be computed by evaluating (E.4) when $c\to-\infty$ and $d=-\frac{\mu}{\sigma}$. We have that:

$$\int_{-\infty}^{0} y_t^2 f_N(y_t;\mu_n,\sigma_t^2)dy_t \;=\; (\mu_t^2+\sigma_t^2)\Phi\left(-\frac{\mu_t}{\sigma_t}\right)+$$

$$-\frac{\sigma_t^2}{\sqrt{2\pi}}\left[-\frac{\mu_t}{\sigma_t}\exp\left(-\frac{\mu_t^2}{2\sigma_t^2}\right)-\lim_{c\to-\infty}c\exp\left(-\frac{c^2}{2}\right)\right]-2\mu_t\sigma_t\phi\left(\frac{\mu_t}{\sigma_t}\right)$$

and, by using De L'Hospital's rule:

$$\lim_{c\to-\infty}c\exp\left(-\frac{c^2}{2}\right)=\lim_{c\to-\infty}\frac{c}{\exp\left(\frac{c^2}{2}\right)}=\lim_{c\to-\infty}\frac{1}{c\exp\left(\frac{c^2}{2}\right)}=0\;.$$

Hence,

$$\int_{-\infty}^{0} y_t^2 f_N(y_t;\mu_t,\sigma_t^2)dy_t=(\mu_t^2+\sigma_t^2)\Phi\left(-\frac{\mu_t}{\sigma_t}\right)-\sigma_t\mu_t\phi\left(\frac{\mu_t}{\sigma_t}\right)\;.$$

The second integral in (E.1) can be analogously computed, considering that in this case $c = -\frac{\mu}{\sigma}$ and $d \to \infty$. This is given by:

$$\int_0^\infty y_t^2 f_N(y_t; \mu_t, \sigma_t^2) dy_t = (\mu_t^2 + \sigma_t^2) \left[1 - \Phi\left(-\frac{\mu_t}{\sigma_t} \right) \right] + \sigma_t \mu_t \phi\left(\frac{\mu_t}{\sigma_t} \right) .$$

Therefore, the expected costs in equation (E.1) can be rewritten as:

$$E(C_t^q) = c_2^q(\mu_t^2 + \sigma_t^2) + (c_2^q - c_1^q) \left[\sigma_t \mu_t \phi\left(\frac{\mu_t}{\sigma_t} \right) - (\mu_t^2 + \sigma_t^2)\Phi\left(-\frac{\mu_t}{\sigma_t} \right) \right] . \tag{E.5}$$

In order to minimize $E(C_t^q)$, given by expression (E.5), the first and the second order derivatives with respect to μ_t are given by:

$$\begin{aligned}
\frac{\partial}{\partial \mu_t} E(C_t^q) &= 2c_2^q \mu_t + (c_2^q - c_1^q) \left[\sigma_t \phi\left(\frac{\mu_t}{\sigma_t} \right) + \sigma_t \mu_t \frac{\partial}{\partial \mu_t} \phi\left(\frac{\mu_t}{\sigma_t} \right) + \right. \tag{E.6} \\
&\quad \left. -2\mu_t \Phi\left(-\frac{\mu_t}{\sigma_t} \right) - (\mu_t^2 + \sigma_t^2) \frac{\partial}{\partial \mu_t} \Phi\left(-\frac{\mu_t}{\sigma_t} \right) \right] = 0 ,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2}{\partial \mu_t^2} E(C_t^q) &= 2c_2^q + (c_2^q - c_1^q) \left[2\sigma_t \frac{\partial}{\partial \mu_t} \phi\left(\frac{\mu_t}{\sigma_t} \right) + \sigma_t \mu_t \frac{\partial^2}{\partial \mu_t^2} \phi\left(\frac{\mu_t}{\sigma_t} \right) + \right. \tag{E.7} \\
&\quad \left. -2\Phi\left(-\frac{\mu_t}{\sigma_t} \right) - 4\mu_t \frac{\partial}{\partial \mu_t} \Phi\left(-\frac{\mu_t}{\sigma_t} \right) - (\mu_t^2 + \sigma_t^2) \frac{\partial^2}{\partial \mu_t^2} \Phi\left(-\frac{\mu_t}{\sigma_t} \right) \right] .
\end{aligned}$$

Equations (E.6) and (E.7) can be computed from the first and second order derivatives of $\phi\left(\frac{\mu_t}{\sigma_t}\right)$ and $\Phi\left(-\frac{\mu_t}{\sigma_t}\right)$. With respect to $\phi\left(\frac{\mu_t}{\sigma_t}\right)$, given by

$$\phi\left(\frac{\mu_t}{\sigma_t}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{\mu_t^2}{\sigma_t^2}\right),$$

these are:

$$\frac{\partial}{\partial\mu_t}\phi\left(\frac{\mu_t}{\sigma_t}\right) = \frac{1}{\sqrt{2\pi}}\left(-\frac{\mu_t}{\sigma_t^2}\right)\exp\left(-\frac{1}{2}\frac{\mu_t^2}{\sigma_t^2}\right) = -\frac{\mu_t}{\sigma_t^2}\phi\left(\frac{\mu_t}{\sigma_t}\right)$$

$$\frac{\partial^2}{\partial\mu_t^2}\phi\left(\frac{\mu_t}{\sigma_t}\right) = -\frac{1}{\sigma_t^2}\phi\left(\frac{\mu_t}{\sigma_t}\right) + \frac{\mu_t^2}{\sigma_t^4}\phi\left(\frac{\mu_t}{\sigma_t}\right).$$

While considering $\Phi(-\frac{\mu_t}{\sigma_t})$, given by:

$$\Phi(-\frac{\mu_t}{\sigma_t}) = \int_{-\infty}^{-\frac{\mu_t}{\sigma_t}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{\mu_t^2}{\sigma_t^2}\right),$$

the derivatives are obtained as follows:

$$\frac{\partial}{\partial\mu_t}\Phi\left(-\frac{\mu_t}{\sigma_t}\right) = -\frac{1}{\sigma_t}\phi\left(-\frac{\mu_t}{\sigma_t}\right) \quad \text{and} \quad \frac{\partial^2}{\partial\mu_t^2}\Phi\left(-\frac{\mu_t}{\sigma_t}\right) = \frac{\mu_t}{\sigma_t^3}\phi\left(-\frac{\mu_t}{\sigma_t}\right).$$

Therefore, the first derivative of the expected costs is:

$$
\begin{aligned}
\frac{\partial}{\partial \mu_t} E(C_t^q) &= 2c_2^q \mu_t + (c_2^q - c_1^q) \left[\sigma_t \phi \left(\frac{\mu_t}{\sigma_t} \right) - \frac{\mu_t^2}{\sigma_t} \phi \left(\frac{\mu_t}{\sigma_t} \right) + \right. \\
&\qquad \left. -2\mu_t \Phi \left(-\frac{\mu_t}{\sigma_t} \right) + \frac{\mu_t^2}{\sigma_t} \phi \left(\frac{\mu_t}{\sigma_t} \right) + \sigma_t \phi \left(\frac{\mu_t}{\sigma_t} \right) \right] \\
&= 2c_2^q \mu_t + 2(c_2^q - c_1^q) \left[\sigma_t \phi \left(\frac{\mu_t}{\sigma_t} \right) - \mu_t \Phi \left(-\frac{\mu_t}{\sigma_t} \right) \right] ,
\end{aligned}
\tag{E.8}
$$

while the second order optimality condition is given by:

$$
\begin{aligned}
\frac{\partial^2}{\partial \mu_t^2} E(C_t^q) &= 2c_2^q + (c_2^q - c_1^q) \left[-2\frac{\mu_t}{\sigma_t} \phi \left(\frac{\mu_t}{\sigma_t} \right) - \frac{\mu_t}{\sigma_t} \phi \left(\frac{\mu_t}{\sigma_t} \right) + \frac{\mu_t^3}{\sigma_t^3} \phi \left(\frac{\mu_t}{\sigma_t} \right) + \right. \\
&\qquad \left. -2\Phi \left(-\frac{\mu_t}{\sigma_t} \right) + \frac{4\mu_t}{\sigma_t} \phi \left(\frac{\mu_t}{\sigma_t} \right) - \frac{\mu_t^3}{\sigma_t^3} \phi \left(\frac{\mu_t}{\sigma_t} \right) - \frac{\mu_t}{\sigma_t} \phi \left(\frac{\mu_t}{\sigma_t} \right) \right] \\
&= 2c_2^q + (c_2^q - c_1^q) \left[-2\Phi \left(-\frac{\mu_t}{\sigma_t} \right) \right] \\
&= 2c_2^q \left[1 - \Phi \left(-\frac{\mu_t}{\sigma_t} \right) \right] + 2c_1^q \left[\Phi \left(-\frac{\mu_t}{\sigma_t} \right) \right] > 0 .
\end{aligned}
\tag{E.9}
$$

Appendix F

Variance of an ARMA(2,q) Process, where $q \leq 2$.

For an ARMA(2,1) process such as

$$(1 - a_1 \mathcal{B} - a_2 \mathcal{B}^2)z_t = (1 - b_1 \mathcal{B})\varepsilon_t, \tag{F.1}$$

its autocovariance is computed by multiplying both sides of the above equation by z_{t-k} and taking expectation:

$$(1 - a_1 \mathcal{B} - a_2 \mathcal{B}^2)\gamma_k = \gamma_{z\varepsilon}(k) - b_1 \gamma_{z\varepsilon}(k-1)$$

where γ_k is the autocorrelation coefficient of z, and $\gamma_{z\varepsilon}$ is the cross-correlation coefficient of z and ε.

When $k = 0$, we have that

$$\gamma_0 = a_1 \gamma_1 + a_2 \gamma_2 + \sigma_\varepsilon^2 - b_1 \gamma_{z\varepsilon}(-1).$$

When $k = 1$,

$$\gamma_1 = a_1 \gamma_0 + a_2 \gamma_1 - b_1 \sigma_\varepsilon^2$$

and when $k \geq 2$,

$$\gamma_k = a_1 \gamma_{k-1} + a_2 \gamma_{k-2}.$$

Also, multiplying both sides of equation (F.1) by ε_{t-1} and taking expectation, we get

$$\gamma_{z\varepsilon}(-1) = (a_1 - b_1)\sigma_\varepsilon^2.$$

Therefore, the variance of z_t is

$$\gamma_0 = \frac{(1-a_2)(1+b_1^2) - 2b_1 a_1}{(1+a_2)(1-a_1-a_2)(1+a_1-a_2)} \sigma_\varepsilon^2. \tag{F.2}$$

Following a similar derivation, the variance of an ARMA(2,2) process $(1 - a_1\mathcal{B} - a_2\mathcal{B}^2)z_t = (1 - b_1\mathcal{B} - b_1\mathcal{B}^2)\varepsilon_t$ can be shown to be equal to

$$\gamma_0 = \frac{(1-a_2)(1+b_1^2+b_2^2) - 2b_1(1-b_2)a_1 - 2b_2 a_2 - 2b_2(a_1^2 - a_2^2)}{(1+a_2)(1-a_1-a_2)(1+a_1-a_2)} \sigma_\varepsilon^2. \tag{F.3}$$

ARMA(2,1) model	$a_1\mathcal{B}$	$a_2\mathcal{B}^2$
$b=1,\ r=0,\ s=1$	$1 - \lambda\xi_1$	$-\lambda\xi_2$
$b=2,\ r=0,\ s=0$	1	$\lambda\xi$
ARMA(2,2) model	$a_1\mathcal{B}$	$a_2\mathcal{B}^2$
$b=1,\ r=1,\ s=0$	$1 + \phi - \lambda\xi$	$-\phi$
$b=2,\ r=1,\ s=0$	$1 + \phi$	$-\lambda\xi - \phi$

Table F.1.   Parameters of four ARMA models

As one can see from Table 7.2, each closed-loop process has a different expression for the coefficients of $\mathcal{B}$ and $\mathcal{B}^2$ in its ARMA model of the output deviations. The coefficients are listed in Table F.1.

References

[1] B.M. Adams and W.H. Woodall. An analysis of Taguchi's on-line process control procedure under a random-walk model. *Technometrics*, 31(4):401–413, 1989.

[2] B.M. Adams, W.H. Woodall, and C.R. Superville. Discussion. *Technometrics*, 36(1):19–22, 1994.

[3] L.C. Alwan. Effects of autocorrelation on control chart performance. *Communications in Statistics - Theory and Methodology*, 21(4):1025–1049, 1992.

[4] L.C. Alwan and H.V. Roberts. Time-series modelling for statistical process control. *Journal of Business and Economic Statistics*, 6(1):87–95, 1988.

[5] D. Anbar. A modified Robbins-Monro procedure approximating the zero of a regression function from below. *The Annals of Statistics*, 5(1):229–234, 1977.

[6] D.W. Apley and J. Shi. The GLRT for statistical process control of autocorrelated processes. *IIE Transactions*, 31(12):1123–1134, 1999.

[7] O. O. Atienza, L. C. Tang, and B. W. Ang. A SPC procedure for detecting level shifts of autocorrelated processes. *Journal of Quality Technology*, 30(4):340–351, 1998.

[8] G.A. Barnard. Control charts and stochastic processes. *J. Royal Stat. Soc., Series B*, 21(2):239–271, 1959.

[9] M. Basseville and I.V. Nikiforov. *Detection of Abrupt Changes: Theory and Application.* Prentice Hall, 1993.

[10] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty. *Nonlinear Programming: theory and algorithms - Second Edition.* John Wiley and Sons, 1993.

[11] J.R. Blum. Approximation methods which converge with probability one. *Annals of Mathematical Statistics*, 25(2):382–386, 1954.

[12] G.E.P. Box and G.M. Jenkins. Further contributions to adaptive quality control: Simultaneous estimation of dynamics non-zero costs. *Bulletin of the International Statistical Institute*, 34:943–974, 1963.

[13] G.E.P. Box, G.M. Jenkins, and G. Reinsel. *Time Series Analysis, Forecasting, and Control, 3rd edtion.* Englewood Cliffs: Prentice Hall, 1994.

[14] G.E.P. Box and T. Kramer. Statistical process monitoring and feedback adjustment - a discussion. *Technometrics*, 34(3):251–267, 1992.

[15] G.E.P. Box and A. Luceño. Discrete proportinal-integral control with constrained adjustment. *The Statistician*, 44(4):479–495, 1995.

[16] G.E.P. Box and J.F. MacGregor. The analysis of closed-loop dynamics-stochastic systems. *Technometrics*, 16(3):391–398, 1974.

[17] G.E.P. Box and J.F. MacGregor. Parameter estimation with closed-loop operating data. *Technometrics*, 18(4):371–384, 1976.

[18] G.E.P. Box and A. Luce no. *Statistical Control by Monitoring and Feedback Adjustment.* John Wiley & Sons, Inc., 1997.

[19] G.E.P. Box and G.D. Tiao. *Bayesian Inference in Statistical Analysis.* John Wiley and Sons, Inc., 1973.

[20] R.A. Boyles. Phase I analysis for autocorrelated processes. *Journal of Quality Technology*, 32(4):395–409, 2000.

[21] B.P. Carlin and T.A. Louis. *Bayes and Empirical Bayes Methods for Data Analysis.* Chapman and Hall, 1996.

[22] A. Chen and E.A. Elsayed. An alternative mean estimator for processes monitored by spc charts. *International Journal of Production Research*, 38(13):3093–3109, 2000.

[23] K.L. Chung. On a stochastic approximation method. *Ann. Math. Statist.*, 25(3):463–483, 1954.

[24] J.P. Comer Jr. Some stochastic approximation procedures for use in process control. *Annals of Mathmatical Statistics*, 35(3):1136–1146, 1964.

[25] S.V. Crowder. *Kalman Filtering and Statistical Process Control.* PhD thesis, Statistics, Iowa State University, 1986.

[26] S.V. Crowder. An SPC model for short production runs: Minimizing expected cost. *Technometrics*, 34(1):64–73, 1992.

[27] S.V. Crowder and L. Eshleman. An alternative mean estimator for processes monitored by SPC charts. *International Journal of Production Research*, 38(3):3093–3109, 2000.

[28] W. Davis. Using corrective action to make matters worse by not considering variation. *Quality Progress*, 33(10):56–61, 2000.

[29] E. Del Castillo. A note on two process adjustment models. *Quality and Reliability Engineering International*, 14(1):23–28, 1998.

[30] E. Del Castillo. Some properties of EWMA feedback quality adjustment schemes for drifting disturbances. *Journal of Quality Technology*, 33(2):153–166, 2001.

[31] E. Del Castillo. *Statistical Process Adjustment for Quality Control*. John Wiley & Sons, Inc., New York, 2002.

[32] E. Del Castillo. Closed-loop disturbance identification and controller tuning for discrete manufacturing processes. *Technometrics*, 2002 to appear.

[33] E. Del Castillo and A. Hurwitz. Run to run process control: a review and some extensions. *Journal of Quality Technology*, 29(2):184–196, 1997.

[34] W.E. Deming. *Out of the Crisis*. MIT, Center for Advanced Engineering Study, Cambridge, Mass., 1994.

[35] D.B. Duncan and S.D. Horn. Linear dynamic recursive estimation from the viewpoint of regression analysis. *Journal of the American Statistical Association*, 67(340):815–821, 1972.

[36] A. Dvoretzky. On stochastic approximation. *Proc. 3rd Berkeley Sym. on Math. Stat. and Prob., J. Neyman (ed.), Berkeley: University of California Press*, pages 39–55, 1956.

[37] J.R. English, S. Lee, T.W. Martin, and C. Tilmon. Detecting changes in autoregressive processes with X-bar and EWMA charts. *IIE Transactions*, 32(12):1103–1113, 2000.

[38] W.D. Ewan and K.W. Kemp. Sampling inspection of continuous processes with no autocorrelation between sucessive results. *Biometrika*, 47(3 and 4):363, 1960.

[39] V. Fabian. Stochastic approximation. *Optimzing Methods in Statistics*, 1971. J.S. Rustags, Academic Press, NY, 1971.

[40] E.W. Frees and D. Ruppert. Estimation following a sequential designed experiment. *Journal of American Statistical Association*, 85(412):1123–1129, 1990.

[41] A.E. Gelfand, S.E. Hills, A. Racine-Poon, and A.F. Smith. Illustration of bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, 85(412):972–985, 1990.

[42] A.E. Gelfand and A.F.M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.

[43] H.L. Gray, G.D. Kelley, and D.D. McIntire. A new approach to ARMA modeling. *Commu. Statist. - Simula. Computa.*, B7(1):1–77, 1978.

[44] F.E. Grubbs. An optimum procedure for setting machines or adjusting processes. *Industrial Quality Control, July*, 1954. reprinted in Journal of Quality Technology, 1983, 15, 4, pp. 186-189.

[45] R.S. Guo, A. Chen, and J.J. Chen. An enhanced EWMA controller for processes subject to random disturbances. *Run to run control in semiconductor manufacturing, Moyne, J., Del Castillo, E., and Hurwitz, A., CRC press*, 2000.

[46] E.J. Hannan and J. Rissanen. Recursive estimation of mixed autoregressive moving average order. *Biometrika*, 69(1):81–94, 1982.

[47] H.J. Harrington. *Business Process Improvement*. New York: McGraw-Hill, 1991.

[48] T.J. Harris. Optimal controllers for nonsymmetric and nonquadratic loss functions. *Technometrics*, 34(3):297–306, 1992.

[49] T.J. Harris and W.H. Ross. Statitical process control procedures for correlated observations. *The Canadian Journal of Chemical Engineering*, 69:48–57, 1991.

[50] S.J. Hu and C. Roan. Change patterns of time series-based control charts. *Journal of Quality Technology*, 28(3):302–312, 1996.

[51] A. Ingolfsson and E. Sachs. Stability and sensitivity of an EWMA controller. *Journal of Quality Technology*, 25(4):271–287, 1993.

[52] W. Jiang and K. L. Tsui. An economic model for integrated APC and SPC control charts. *IIE Transactions*, 32(6):505–513, 2000.

[53] R.E. Kalman. A new approach to linear filtering and prediction problems. *Transaction ASME Journal of Basic Engineering*, 82:35–45, 1960.

[54] W.D. Kelton, W.M. Hancock, and D.P. Bischak. Adjustment rules based on quality control charts. *Int. J. Prod. Res.*, 28(2):365–400, 1990.

[55] H. Kesten. Accelerated stochastic approximation. *Annals of Mathematical Statistics*, 29(1):41–59, 1958.

[56] T.T. Kramer. *Process Control From an Economic Point of View.* PhD thesis, Statistics, University of Wisconsin - Madison., 1989.

[57] T.P. Krasulina. On the probability of not exceeding a desired threshold by a stochastic approximation algorithm. *Automation and Remote Control*, 59(10):1424–1427, 1998.

[58] S.P. Ladany. Optimal set-up of a manufacturing process with unequal revenue from oversized and undersized items. *IEEE 1995 Engineering Management Conference*, pages 428–432, 1995.

[59] T.L. Lai and H. Robbins. Adaptive design in regression and control. *Proceedings of the National Academy of Sciences of the United States of America*, 75(2):586–587, 1978.

[60] T.L. Lai and H. Robbins. Adaptive design and stochastic approximation. *Annals of Statistics*, 7(6):1196–1221, 1979.

[61] F.L. Lewis. *Optimal Estimation with An Introduction to Stochastic Control Theory.* John Wiley and Sons, 1986.

[62] L. Ljung and T. Söderström. *Theory and Practice of Recursive Identification.* Cambridge, Mass., The MIT Press, 1987.

[63] G. Lorden. Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics*, 41(2):520–527, 1971.

[64] C. Lu and M.R. Reynolds Jr. EWMA control charts for monitoring the mean of autocorrelated processes. *Journal of Quality Technology*, 31(2):166–188, 1999.

[65] C. Lu and M.R. Reynolds Jr. Cusum charts for monitoring an autocorrelated processes. *Journal of Quality Technology*, 33(3):316–334, 2001.

[66] J.M. Lucas. Combined Shewhart-CUSUM quality control schemes. *Journal of Quality Technology*, 14(2):51–59, 1982.

[67] J.M. Lucas and M.S. Saccucci. Exponentially weighted moving average control schemes: Properties and enhancement. *Technometrics*, 32(1):1–12, 1990.

[68] A. Luceño. Performance of EWMA versus last observation for feedback control. *Commu. Statist. - Theory Meth.*, 22(1):241–255, 1993.

[69] J.F. MacGregor. On-line statistical process control. *Chemical Engineering Progress, Oct.*, pages 21–31, 1988.

[70] J.F. MacGregor. Discussion. *Journal of Quality Technology*, 23(3):198–199, 1991.

[71] S. Maghsoodloo and M.C. Li. Optimal asymmetric tolerance design. *IIE Transactions*, 32(12):1127–1137, 2000.

[72] R.J. Meinhold and N.D. Singpurwalla. Understanding the Kalman Filter. *The American Statistician*, 37(2):123–127, 1983.

[73] D.C. Montgomery. *Introduction to Statistical Quality Control, 3rd edition.* New York, Wiley, 1996.

[74] D.C. Montgomery and C.M. Mastrangelo. Some statistical process control methods for autocorrelated data. *Journal of Quality Technology*, 23(3):179–193, 1991.

[75] P.R. Moorhead and C.F.J. Wu. Cost-driven parameter design. *Technometrics*, 40(2):111–119, 1998.

[76] J. Moyne, E. del Castillo, and A. Hurwitz. *Run to Run Process Control in Semiconductor Manufacturing.* CRC Press, Boca Raton, FL, 2000.

[77] E.S. Page. Continuous inspection schemes. *Biometrika*, 41(1-2):100–115, 1954.

[78] N.S. Patel and S.T. Jenkins. Adaptive optimization of run-to-run controllers: The ewma example. *IEEE Transactions on Semiconductor Engineering*, 13(1):97–107, 2000.

[79] H. Robbins and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22(3):400–407, 1951.

[80] H. Robbins and D. Siegmund. A convergence theorem for non-negative almost positive supermartingales and some applications. *Optimizing Methods in Statistics, J.S. Rustagi*, pages 237–257, 1971. Academic Press, NY.

[81] N.H. Ruhhal, G.C. Runger, and M. Dumitrescu. Control charts and feedback adjustments for a jump disturbance model. *Journal of Quality Technology*, 32(4):379–394, 2000.

[82] D. Ruppert. A new dynamic stochastic approximation procedure. *The Annals of Statistics*, 7(6):1179–1195, 1979.

[83] D. Ruppert. Stochastic approcimation of an implicitly defined function. *The Annals of Statistics*, 9(3):555–566, 1981.

[84] D. Ruppert. Stochastic approximation. *Handbook of Sequential Analysis*, 1991. B. K. Ghosh and P. K. Sen, Marcel Dekker, NY, 1991.

[85] T.P. Ryan. Discussion. *Journal of Quality Technology*, 23(3):200–202, 1991.

[86] E. Sachs, A. Hu, and A. Ingolfsson. Run by run process control: Combining spc and feedback control. *IEEE Transactions on Semiconductor Manufacturing*, 8(1):26–43, 1995.

[87] W.A. Shewhart. *Economic Control of Quality of Manufactured Product*. Van Nostrand, New York, 1931.

[88] D. Siegmund. *Sequential Analysis: Tests and Confidence Intervals*. Springer-Verlag, New York, 1985.

[89] E.A. Silver and H.C. Meal. A heuristic for selecting lot size quantities for the case of a deterministic time-varying demand rate and discrete opportunities for replenishment. *Production and Inventory Management*, 10(4):64–74, 1973.

[90] E.A. Silver, D.F. Pyke, and R. Peterson. *Inventory Management and Production Planning and Scheduling, 3rd Edition.* John Wiley and Sons, Inc., 1998.

[91] N.C. Simpson. Questioning the relative virtues of dynamic lot sizing rules. *Computers and Operations Research*, 28(9):899–914, 2001.

[92] D.J. Spiegelhalter, A. Thomas, N.G. Best, and W.R. Gilks. *BUGS: Bayesian inference Using Gibbs Sampling, version 0.30.* Cambridge: Medical Research Council Biostatistics Unit, 1994.

[93] A.P. Stamboulis. First order autoregressive model applied to quality control. *New York University memorandum*, 1971.

[94] K.J. Åström. *Introduction to Stochastic Control Theory.* San Diega, CA: Academic Press, 1970.

[95] K.J. Åström and B. Wittenmark. *Computer Controlled Systems Theory and Design, 3rd edition.* Englewood Cliffs: Prentice Hall, 1997.

[96] P. Sullo and M. Vandeven. Optimal adjustment strategies for a process with run-to-run variation and 0-1 quality loss. *IIE Transactions*, 31(12):1135–1145, 1999.

[97] G. Taguchi. Quality engineering in Japan. *Commu. Statist. - Theory Meth.*, 14:2785–2801, 1985.

[98] G. Taguchi, E.A. Elsayed, and T. Hsiang. *Quality Engineering in Production System*. New York: McGraw-Hill, 1989.

[99] D.H. Timmer, J. Pignatiello Jr., and M. Longnecker. The development and evaluation of CUSUM-based control charts for an AR(1) process. *IIE Transactions*, 30(6):525–534, 1998.

[100] D. Trietsch. The harmonic rule for process setup adjustment with quadratic loss. *Journal of Quality Technology*, 30(1):75–84, 1998.

[101] D. Trietsch. Process setup adjustment with quadratic loss. *IIE Transactions*, 32(4):299–307, 2000.

[102] R.S. Tsay and G.C. Tiao. Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *Journal of the American Statistical Association*, 79(385):84–96, 1984.

[103] R.S. Tsay and G.C. Tiao. Use of canonical analysis in time series model identification. *Biometrika*, 72(2):299–315, 1985.

[104] W.T. Tucker, F.W. Faltin, and S.A. Vander-Wiel. Algorithmic statistical process control: An elaboration. *Technometrics*, 35(4):363–375, 1993.

[105] S.A. Vander-Wiel, W.T. Tucker, F.W. Faltin, and N. Doganaksoy. Algorithmic statistical process control: Concepts and an application. *Technometrics*, 34(3):286–297, 1992.

[106] A.V. Vasilopoulos and A.P. Stamboulis. Modification of control chart limits in the presence of data correlation. *Journal of Quality Technology*, 10(1):20–30, 1978.

[107] J.H. Venter. An extension of the Robbins-Monro procedure. *Annals of Mathematical Statistics*, 38(1):181–190, 1967.

[108] H. Wagner and T.M. Whitin. Dynamic version of the economic lot size model. *Managemnet Science*, 5(1):89–96, 1958.

[109] D.G. Wardell, H. Moskowitz, and R.D. Plante. Rejoinder. *Technometrics*, 36(1):23–27, 1994.

[110] D.G. Wardell, H. Moskowitz, and R.D. Plante. Run-length distributions of special-cause control charts for correlated processes. *Technometrics*, 36(1):3–17, 1994.

[111] S.J. Wiklund. Estimating the process mean when using control charts. *Economic Quality Control*, 7:105–120, 1992.

[112] S.J. Wiklund. Adjustment strategies when using Shewhart charts. *Economic Quality Control*, 8:3–21, 1993.

[113] W.H. Woodall and B.M. Adams. The statistical design of CUSUM charts. *Quality Engineering*, 5(4):559–570, 1993.

[114] C.M. Wright, D.E. Booth, and M.Y. Hu. Joint estimation: SPC method for short-run autocorrelated data. *Journal of Quality Technology*, 33(3):365–378, 2001.

[115] C.C. Wu and G.R. Tang. Tolerance design for products with asymmetric quality losses. *International Journal of Production Research*, 36(9):2592–2541, 1998.

[116] E. Yashchin. Analysis of CUSUM and other Markov-type control schemes by using empirical distributions. *Technometrics*, 34(1):54–63, 1992.

[117] E. Yashchin. Estimating the current mean of a process subject to abrupt changes. *Technometrics*, 37(3):311–323, 1992.

[118] E. Yashchin. Performance of CUSUM control schemes for serially correlated observations. *Technometrics*, 35(1):37–52, 1993.

[119] P. Young. *Recursive Estimation and Time-Series Analysis*. New York: Springer-Verlag, 1984.

[120] N.F. Zhang. A statistical control chart for stationary process data. *Technometrics*, 40(1):24–38, 1998.

Vita

Rong Pan, Ph.D.

Research Interests

- Applied Statistics and Quality Control, including: Time Series Forecasting and Control, Stochastic Processes, Bayesian Statistics, Statistical Quality Control and Design of Experiments.

Education

- Ph.D., Industrial and Manufacturing Engineering, Minor in Statistics, Penn State University, August 2002.

- M.S., Industrial Engineering, College of Engineering of Florida A&M and Florida State Universities, August 1999.

- B.E., Materials Science and Engineering, Shanghai Jiao Tong University, July 1995.

Research Experience

- 9/2000-5/2002: Research Assistant, Applied Statistics Laboratory, Department of Industrial and Manufacturing Engineering, Penn State University.

  Conducted research in *Time Series Control and Process Adjustment*, supported by NSF grant DMI 99-96031.

- 1/1999-5/2000: Research Assistant, Department of Industrial and Manufacturing Engineering, Penn State University.

  Quality consulting, supported by General Nutrition Products.

- 9/1997-12/1998: Research Assistant, Department of Industrial Engineering, College of Engineering, Florida A&M and Florida State Universities.

  Statistical analysis of fiber permeability for the resin transfer manufacturing (RTM) process of fiber-reinforced polymer composites, supported by Air Force fund.

Industrial Experience

- 6/2000-8/2000: CAD Engineer, REMCOM CO., Pennsylvania.

- 5/1999-8/1999: Quality Consultant, General Nutrition Products, South Carolina.

- 1/1996-5/1997: Quality Technician, General Electric Plastics Division, China.

- 9/1995-1/1996: Trainee, General Electric, Singapore.