

Center for Quality and Productivity Improvement
University of Wisconsin-Madison
610 Walnut Street
Madison, Wisconsin 53705

Report No. 43

**PROCESS CONTROL FROM AN
ECONOMIC POINT OF VIEW**

**Chapter 2:
Fixed Monitoring and Adjustment Costs**

Tim Kramer*

February, 1990

This research was sponsored by the National Science Foundation under Grant No. DDM-8808138

*Applied Statistician, Hewlett Packard, Avondale, Pennsylvania

This report is part of a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Statistics) at the University of Wisconsin-Madison (1989). Thesis advisor: George E.P. Box.

**CENTER FOR QUALITY AND PRODUCTIVITY IMPROVEMENT
UNIVERSITY OF WISCONSIN-MADISON**

Report No. 43

**PROCESS CONTROL FROM AN
ECONOMIC POINT OF VIEW**

**CHAPTER 2:
FIXED MONITORING AND ADJUSTMENT COSTS**

Tim Kramer

February 1990

PRACTICAL SIGNIFICANCE

In this chapter an economic balance is found between the cost of regulating a process about a target and the cost of deviating from target. Specifically, we discuss the problem of choosing a control scheme that minimizes the combined costs of monitoring, adjustment and being off-target. In addition to the various costs, the choice depends on the nature of the disturbance and the dynamic relationship between control actions and their effects.

Key Words: Off-Target Costs, Monitoring Cost, Adjustment Cost, Non-Stationary Disturbance

This research was sponsored by the National Science Foundation under Grant No. DDM-8808138.

This report is part of a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Statistics) at the University of Wisconsin-Madison (1989). Thesis advisor: George E.P. Box.

The Center for Quality and Productivity Improvement cares about your reactions to our reports. Please send comments (general or specific) to: Report Feedback, Center for Quality and Productivity Improvement, 610 Walnut Street, Madison, WI 53705. All replies will be forwarded to the authors.

PROCESS CONTROL FROM AN ECONOMIC POINT OF VIEW
CHAPTER 2: FIXED MONITORING AND ADJUSTMENT COSTS

TABLE OF CONTENTS

| | | |
|------|---|-------------|
| 2.1 | Introduction | 26 |
| 2.2 | The Machine Tool Control Problem..... | 29 |
| 2.3 | Approximations to the Standardized ARL and MSD..... | 36 |
| 2.4 | The Solution of the Machine Tool Control Problem..... | 45 |
| 2.5 | Estimating the Disturbance Parameters..... | 59 |
| 2.6 | Comparison with Other Control Charts..... | 65 |
| 2.7 | Control When There is a Monitoring Cost | 77 |
| 2.8 | Fixed Monitoring and Adjustment Costs | 86 |
| 2.9 | Aids for Implementing an MCC Scheme..... | 92 |
| 2.10 | Taguchi's Solution to the Machine Tool Adjustment Problem | 110 |
| 2.11 | Summary..... | 123 |
| | BIBLIOGRAPHY..... | 124a |

CHAPTER 2: FIXED MONITORING AND ADJUSTMENT COSTS

2.1 Introduction

The goal in this chapter is to find an economic balance between the cost of regulating a process about a target and the cost of deviating from target. Specifically, we discuss the problem of choosing a control scheme that minimizes the combined costs of monitoring, adjustment and being off-target. In addition to the various costs, the choice depends on the nature of the disturbance and the dynamic relationship between control actions and their effects.

Adjustments with no dynamics and no delay

We restrict consideration in this chapter to adjustments which have an appreciable cost and whose effect is fully realized within one time unit. Examples of such changes include a die change and the replacement of a filter membrane, each characterized by having an appreciable cost associated with the change and, once made, the effects of these changes were fully and immediately realized. Adjustments, if any, by necessity are made after the data are collected and so the class of difference equations relating input and output are

$$x_{t+1} = c_0 + gw_t \quad (2.1)$$

Note that $t+1$ is the first opportunity to observe the effect of a change made at time t and so equation (2.1) is a special case of equation (1.10) with the delay parameter, b , set to one.

Fixed costs for adjusting and monitoring

The cost of adjustment is assumed to be a constant, C_a , which represents all costs associated with making an adjustment including, for example, labor, lost productivity, and new or refurbished parts. Similarly, the cost to monitor the process is assumed to be a constant, C_m . This cost would include the data collector's time and the use of equipment and materials to obtain the data. There are many examples of fixed monitoring costs ranging from the negligible costs of having on-line detectors obtain size and weight data, the marginal costs of having a quality inspector obtain tissue softness or particle size data, to the substantial costs of having dimensions verified on a coordinate measurement machine or chemical contents determined by a procedure which uses expensive chemicals.

The disturbance

The disturbance is assumed to follow

$$z_{t+1} - z_t = a_{t+1} - \theta_1 a_t \quad 0 \leq \theta_1 \leq 1 \quad (2.2)$$

where $\{a_t\}$ is a white noise sequence of "random shocks", that is, a sequence normally and independently distributed about a zero mean with variance σ_1^2 .

Depending on the value of θ_1 it was shown in the last chapter that this model could represent a drifting nonstationary disturbance as well as uncorrelated noise about a fixed mean representing a process in a state of control. The subscript 1 on both the smoothing constant, θ_1 , and the variance of the random shocks, σ_1^2 , is needed in what follows as we refer to the disturbance as observed at the base interval (unit interval) and the monitoring interval. The base interval parameters always have the 1 subscript

and the monitoring interval parameters have an m subscript.

Off-target cost

We assume that the cost (or loss) from being off-target is a quadratic function of the deviation $y_t - T$ from target. Thus

$$\text{Off-target cost} = k_T(y - T)^2 = \frac{C_T(y - T)^2}{\gamma_1^2 \sigma_1^2} \quad (2.3)$$

where $\gamma_1 = 1 - \theta_1$ and σ_1^2 is the variance of the random shocks underlying the disturbance model. The "standardized form" involving C_T , as is shown later, is a natural parameter for determining the action criteria and monitoring interval. The value for C_T is readily determined once the disturbance parameters γ_1 and σ_1 and the off-target cost for a single deviation $y - T$ are known. Such a value may be obtained by considering, for example, the average loss that would result from supplying a component which was unacceptable to the customer. Thus, a deviation $y_u - T = d_u$ which produces a cost c_u would yield

$$C_T = \frac{c_u \gamma_1^2 \sigma_1^2}{d_u^2} \quad (2.4)$$

Preview of chapter results

We consider first the *machine tool control problem* in which the monitoring cost are negligible and the monitoring interval is regarded as fixed but the adjustment cost are substantial. We next consider the situation where the adjustment cost is negligible but the monitoring cost is substantial and the optimal monitoring interval is to be determined. Finally, we consider the situation where both the monitoring and

adjustment costs are substantial. The results derived are compared to those of Taguchi.

2.2 The Machine Tool Control Problem

In this section we consider the problem of designing a control scheme when the system has no dynamics but the cost of adjustment is important. Though Box and Jenkins [9] solved this problem more than 25 years ago, it is presented here for several reasons. First, their formulation and analysis form the foundation for the extension considered later in this chapter in which a monitoring cost is included. Second, some controversy between the SPC and APC practitioners can be resolved by comparing the results of this analysis to the use of a Shewhart control chart. Box and Jenkins mentioned the similarity between the minimal cost procedure and the Shewhart control chart but it is more fully explored here. Third, new charts and figures that may simplify (and thereby expand) the use of this procedure have been developed and included.

The machine tool control problem occurs in the mass production of components, such as piston rods, where some quality characteristics, such as rod length, is measured periodically. There is a loss associated with deviating from target which is assumed to be proportional to the square of the deviation, as in equation (2.3). The disturbance is assumed to follow the nonstationary process defined by equation (2.2). The disturbance can be counteracted only by making adjustments, $-w_i$, to a manipulated input variable (position of a cutting bar). The accumulation of all the adjustments up to time t has brought this input variable to a level

$$-W_t = -\sum w_{t-j} \quad (2.5)$$

An adjustment, $-w_t$, possibly after considerable modification in passage through the process, emerges as an incremental change $-x_t$ in the output (change in cutting length). The accumulated effect up to time t of these induced output changes is $-X_t = -\sum x_{t-j}$. The induced change, X_t , will be called the adjustable *set point* at time t . The extent to which the set point X_t fails to compensate for the accumulated disturbance z_t is evidenced by e_t , the observed deviation of the output from its target value. Thus

$$e_t = z_t - X_t \quad (2.6)$$

or in differenced form,

$$e_{t+1} = e_t + a_{t+1} - \theta_1 a_t - x_{t+1} \quad (2.7)$$

Suppose $\hat{z}_{t+1} = \hat{z}_t(1)$ is the predicted disturbance one step ahead. Then if it costs nothing to make an adjustment we would set $X_{t+1} = \hat{z}_{t+1}$ at each stage. But we must pay $\$C_a$ to make an adjustment and so the predicted deviation, $\hat{z}_{t+1} - X_{t+1}$, at time t must be sufficiently large to warrant paying C_a dollars. Because of the adjustment cost the set point is usually kept at a constant level for considerable periods (i.e., many of the "adjustments," $x_j, x_{j+1}, x_{j+2}, \dots$ are zero and the corresponding levels of the set point $X_j, X_{j+1}, X_{j+2}, \dots$ are identical).

The problem is, knowing k_T, C_a, θ_1 and σ_1 , to choose an optimal policy that tells us (a) *when* to change, and (b) by *how much* to change so that the overall cost in running the control scheme is minimized.

Choice of optimal policy

Box and Jenkins used the framework of dynamic programming to look at this sequential decision making problem. First consider the situation where the control procedure must terminate after one further item and one wants to know whether to make an adjustment. The expected cost viewed from time t for making this additional item is

$$\begin{aligned} k_T E(z_{t+1} - X_t)^2 &= k_T (\hat{z}_{t+1} - X_t)^2 + k_T \sigma_1^2 && \text{if no change is made} \\ k_T E(z_{t+1} - X_{t+1})^2 + C_a &= k_T \sigma_1^2 + C_a && \text{if a change is made} \end{aligned} \quad (2.8)$$

The second equality holds when X_{t+1} is set equal to \hat{z}_{t+1} which minimizes that expression. Hence, the rule which minimizes the expected one step loss is to *adjust if*

$$|\hat{z}_{t+1} - X_t| \geq \left[\frac{C_a}{k_T} \right]^{1/2} = \lambda_1 \quad (2.9)$$

and continue at the current level if

$$|\hat{z}_{t+1} - X_t| \leq \lambda_1 \quad (2.10)$$

In the above equations, λ_1 is the action limit to employ when the process continues for just one more item. Note that the action criteria expressed in equations (2.9) and (2.10) depends only on the absolute value of the predicted deviation from target, $|\hat{z}_{t+1} - X_{t+1}|$, and that the appropriate control action, when one is justified, is to adjust the set point so that this predicted deviation is zero. We now show that the optimal N step strategy is to adjust when the absolute value of the predicted deviation from target exceeds some limit and, when justified, the set point X_{t+1} is adjusted so that the predicted deviation is zero, $\hat{z}_{t+1} - X_{t+1} = 0$

Let $L_N(\hat{z}_{t+1}, X_t)$ be the minimal expected loss if the procedure terminates after N further items and suppose that it has already been established that $L_{N-1}(\hat{z}_{t+1}, X_t)$ is a function of $|\hat{z}_{t+1} - X_{t+1}|$. Then the expected N -step loss if no change is made at time t , $L_N^{\text{no}}(\hat{z}_{t+1}, X_t)$, is the expected loss one step ahead plus the expected minimal loss if we started at $z_{t+1} - X_t$ and there were $N - 1$ further items. This is

$$L_N^{\text{no}}(\hat{z}_{t+1}, X_t) + \int_{-\infty}^{\infty} L_{N-1}(\hat{z}_{t+2}, X_t) \cdot \text{Prob}(\hat{z}_{t+2} | \hat{z}_{t+1}) d\hat{z}_{t+2} \quad (2.11)$$

Since the disturbance follows equation (2.2)

$$z_{t+2} = z_{t+1} + a_{t+2} - \theta_1 a_{t+1} \quad (2.12)$$

the predicted values satisfy

$$\hat{z}_{t+2} = \hat{z}_{t+1} + \gamma_1 a_{t+1} \quad (2.13)$$

where a_j is the random shock at time j and $\gamma_1 = 1 - \theta_1$. The random shocks are assumed to be independent normal random variables with mean 0 and variance σ_1^2 .

Using the above with the fact that

$$L_N^{\text{no}}(\hat{z}_{t+1}, X_t) = k_T(\hat{z}_{t+1} - X_t)^2 + k_T \sigma_1^2 \quad (2.14)$$

one obtains

$$L_N^{\text{no}}(\hat{z}_{t+1}, X_t) = k_T(\hat{z}_{t+1} - X_t)^2 + k_T \sigma_1^2 + \int_{-\infty}^{\infty} L_{N-1}(\hat{z}_{t+1} + \gamma_1 \sigma_1 u, X_t) p(u) du \quad (2.15)$$

where $p(u)$ is the unit normal distribution. From the original supposition one gets

$$L_{N-1}(\hat{z}_{t+1} + \gamma_1 \sigma_1 u, X_t) = f_{N-1}(|\hat{z}_{t+1} + \gamma_1 \sigma_1 u - X_t|) \quad (2.16)$$

which, upon integration according to equation (2.15), must be some function of

$|\hat{z}_{t+1} - X_t|$. Hence, $L_N^{\text{no}}(\hat{z}_{t+1}, X_t)$ is some function of $|\hat{z}_{t+1} - X_t|$.

Now consider the loss if an adjustment is made at time t . The loss, $L_N^{\text{adj}}(\hat{z}_{t+1}, X_t)$, corresponding to adjustment is

$$L_N^{\text{adj}}(\hat{z}_{t+1}, X_t) = k_T \sigma_1^2 + C_a + \int_{-\infty}^{\infty} L_{N-1}(\hat{z}_{t+1} + \gamma_1 \sigma_1 u, X_{t+1}) p(u) du \quad (2.17)$$

The integral is minimized by setting X_{t+1} equal to \hat{z}_{t+1} which renders $L_N^{\text{adj}}(\hat{z}_{t+1}, X_t)$ independent of either \hat{z}_{t+1} or X_t . Hence, the optimal rule is to adjust when $L_N^{\text{no}}(\hat{z}_{t+1}) \geq L_N^{\text{adj}}(\hat{z}_{t+1})$ which can be expressed in the form $|\hat{z}_{t+1} - X_t| \geq \lambda_N$. Box and Jenkins showed that the sequence $\{\lambda_j\}$ approaches a limiting value λ corresponding to the practical case where the time of operation of the control procedure is effectively infinite. Crowder [17] showed that these limiting values are reached quickly—three decimal place convergence is usually reached with N as small as five.

Finding the best value of λ

The preceding analysis shows that the minimal cost strategy is to adjust when the predicted deviation exceeds some *action limit* λ that depends on the disturbance parameters and the costs of adjustment and being off-target. We now proceed to find an explicit relationship for the action limit as a function of these parameters.

Suppose that an adjustment was made at time t . Then the deviation $z_{t+n} - X_{t+1}$ if no change has occurred by time $t+n$ is

$$e_{t+n} = z_{t+n} - X_{t+1} = a_{t+n} + \gamma_1(a_{t+n-1} + \dots + a_{t+1}) \quad (2.18)$$

where $a_{t+j} = z_{t+j} - \hat{z}_{t+j}$. If an adjustment is made at time $t+n$, then

$$|\hat{z}_{t+n+1} - X_{t+1}| = |\gamma_1(a_{t+n} + a_{t+n-1} + \dots + a_{t+1})| \geq \lambda \quad (2.19)$$

Let the term *run* denote the number of periods between adjustments. For fixed λ ,

equation (2.19) has a run length distribution and an average run length (ARL). The average loss per run as a result of being off-target is

$$k_T E(S_n) = k_T E \left[\sum_{j=1}^n (z_{t+j} - X_{t-1})^2 \right] \quad (2.20)$$

where the expectation is over the distribution of z and the run lengths and is subject to the boundary condition (2.19). The overall expected loss per run is $C_a + k_T E(S_n)$ and on a per observation basis this is

$$L = \frac{C_a}{\text{ARL}} + \frac{k_T E(S_n)}{\text{ARL}} \quad (2.21)$$

Expressions for the ARL and $E(S_n)$ as functions of λ are needed to find the best value of λ . The action criteria of equation (2.19) can be rewritten relating the random shocks $\{a_j\}$ to unit normal deviates $\{u_j\}$ by factoring out the standard deviation.

Doing this, one obtains that adjustments should be made when

$$| u_{t+n} + u_{t+n-1} + \cdots + u_{t+1} | \geq \lambda / \gamma_1 \sigma_1 \quad (2.22)$$

The action criteria (adjustment rule) is now expressed above in a form that is

equivalent to passage through a barrier for a random walk. Hence, the ARL between adjustments is equivalent to the first passage time for a standardized random walk, for which a numerical approximation is given in the next section. Using the notation $h(\Lambda)$ to represent the first passage time of the standardized random walk through a barrier Λ , one sees that for the disturbance model with parameters $\gamma_1 = 1 - \theta_1$ and σ_1 , the ARL for a limit λ is equivalent to $h(\lambda / \gamma_1 \sigma_1)$.

Finally, the *mean squared deviation* (MSD) for a run is again related to the random walk. In terms of the previous notation,

$$\text{MSD} = \frac{E(S_n)}{\text{ARL}} \quad (2.23)$$

Equation (2.20) may be rewritten in terms of the unit normal deviates and when combined with equation (2.23) gives

$$\text{MSD} = \sigma_1^2 + \gamma_1^2 \sigma_1^2 \frac{E\left(\sum_{j=1}^{n-1} U_j^2\right)}{\text{ARL}} \quad (2.24)$$

where

$$U_{j-1} = u_{t+j-1} + \dots + u_{t+1} \quad (2.25)$$

The action criteria is now seen to be equivalent to changing when $U_{j-1} \geq \lambda/\gamma_1 \sigma_1$. Let

$g(\Lambda)$ represent $E\left(\sum_{j=1}^{n-1} U_j^2\right)/\text{ARL}$. It is seen that $g(\Lambda)$ is related to the MSD of the

standardized random walk that is renewed (set to zero) after the process passes through a barrier at $\pm\Lambda$:

$$\text{MSD} = g(\Lambda) + 1 \quad (2.26)$$

As with the ARL function, $h(\Lambda)$, a numerical approximation to $g(\Lambda)$ is given in the next section.

Using the substitutions for the standardized ARL and MSD one obtains the following expression for the loss due to process regulation:

$$L = \frac{C_a}{h(\lambda/\gamma_1 \sigma_1)} + k_T \sigma_1^2 + k_T \gamma_1^2 \sigma_1^2 g(\lambda/\gamma_1 \sigma_1) \quad (2.27)$$

Letting $\Lambda = \lambda/\gamma_1 \sigma_1$ and $C_T = k_T \gamma_1^2 \sigma_1^2$, gives

$$L = k_T \sigma_1^2 + \frac{C_a}{h(\Lambda)} + C_T g(\Lambda) \quad (2.28)$$

An expression for the best choice of λ is obtained by setting the partial derivative of

the loss with respect to Λ to zero, giving

$$\frac{g'(\Lambda)h(\Lambda)^2}{h'(\Lambda)} = \frac{C_a}{C_T} \quad (2.29)$$

The right side of equation (2.27) is the ratio of the adjustment cost to an off-target cost (which is shown later to be the rate at which the off-target costs accrue when no adjustments are made). This ratio is called the *relative adjustment cost* and be denoted by R_a . Using the substitution $f(\Lambda) = g'(\Lambda)h(\Lambda)^2/h'(\Lambda)$, one gets

$$f(\Lambda) = R_a \quad (2.30)$$

or

$$\Lambda = f^{-1}(R_a) \quad (2.31)$$

and

$$\lambda = f^{-1}(R_a)\gamma_1\sigma_1 \quad (2.32)$$

The best choice of λ is some function of the relative adjustment cost where this function is related to both the standardized ARL and MSD functions. Numerical approximations are given in the next section and these are used to obtain estimates of the best value of λ in §2.4.

2.3 Approximations to the Standardized ARL and MSD

In the last section we showed that the minimal cost control scheme required that adjustments to the process be made only when the predicted deviation from target exceeded an action limit that depended on the adjustment and off-target costs as well as the disturbance parameters. The overall loss can be expressed as a function of the ARL and MSD when equations (2.23) and (2.25) are combined.

$$L = \frac{C_a}{ARL} + k_T \cdot MSD \quad (2.34)$$

We also showed that both the ARL and MSD for a control scheme with action limits at $\pm \lambda$ and disturbance parameters (θ_1, σ_1) can be related to the ARL and MSD of a standardized random walk crossing a boundary. The following two equations show this relationship:

$$\text{Control scheme ARL} = h(\lambda/\gamma_1 \sigma_1) \quad (2.35)$$

where $\gamma_1 = 1 - \theta_1$ and $h(\Lambda)$ is the average run length of the standardized random walk crossing a boundary, Λ ; and

$$\text{Control scheme MSD} = \sigma_1^2 + \gamma_1^2 \sigma_1^2 g(\lambda/\gamma_1 \sigma_1) \quad (2.36)$$

where $1 + g(\Lambda)$ is the MSD of the standardized random walk. Finally, we showed that the action limit which leads to a minimal cost control scheme is given by

$\lambda = f^{-1}(R_a) \gamma_1 \sigma_1$ where f is defined as

$$f(\Lambda) = \frac{g'(\Lambda)h(\Lambda)^2}{h'(\Lambda)} \quad (2.37)$$

What is needed now are equations for $g(\Lambda)$ and $h(\Lambda)$ so that the action limit λ may be determined for a specified relative adjustment cost, R_a .

Box and Jenkins provided the following approximations for the functions $h(\Lambda)$ and $g(\Lambda)$:

$$h(\Lambda) = \Lambda^2 + 1.25331\Lambda + 0.66667 \quad (2.38)$$

and

$$g(\Lambda) = \frac{\Lambda^4 + 2.50663\Lambda^3 + \Lambda^2 - 0.4}{6\Lambda^2 + 7.51988\Lambda + 4} \quad (2.39)$$

These approximations improve as Λ gets larger and are still within 5% of simulated

values with Λ as small as one. The inaccuracies for small Λ have little meaning in the realm of the machine tool control problem as large relative adjustment costs lead to large action limits. However, this chapter considers an array of problems including those in which the relative adjustment cost is small, leading to narrow action limits and, correspondingly, frequent adjustments. Hence, approximations for $g(\Lambda)$ and $h(\Lambda)$ are needed that are good whether Λ is small or large.

The equations that follow were developed to provide a smooth approximation to simulated values for the standardized ARL and MSD. The control procedure (adjust when the predicted deviation exceeds the action limit) was simulated for a range of Λ using the IMSL routine GGNML for generating the normal random shocks. For each value of Λ the control scheme was applied for a minimum of 30,000 periods and a maximum of 800,000 periods. The number of periods was increased as Λ got larger, reflecting the larger ARLs, so as to generate a minimum of 1000 separate runs. This 1000 run minimum was not achieved for Λ greater than 30 as the number of periods (and computer run time) became excessive.

The equation for $h(\Lambda)$ was chosen to be a monotonically increasing function of Λ with $h(0) = 1$ and $h(\Lambda) \approx \Lambda^2$ for large Λ , matching theoretical values for both large Λ and $\Lambda = 0$. A simple quadratic function of Λ was initially chosen. The ratio of the simulated ARLs to the quadratic approximation was plotted versus Λ and a second function was chosen to fit these ratios. The chosen approximation to the standardized ARLs is the product of these two functions:

$$h(\Lambda) = (1 + 1.1\Lambda + \Lambda^2) \cdot (1 - 0.115\exp(-9.2(\Lambda^{0.3} - 0.88)^2)) \quad (2.40)$$

The equation for $g(\Lambda)$ was chosen in a similar manner. The function was chosen to match the theoretical values at zero, $g(0) = 0$ and for large Λ , $g(\Lambda) \approx \Lambda^2/6$. As with the approximation to the ARLs, a quadratic function provided an initial approximation, and a second function was obtained to compensate for the ratio between the simulated MSDs and the quadratic approximation. The following function provides an acceptable approximation to the simulated MSDs:

$$g(\Lambda) = \frac{(1 + 0.06\Lambda^2)}{1 - 0.647 \cdot \Phi(1.35(\ln(\Lambda) - 0.67))} - 1 \quad (2.41)$$

where $\Phi(\cdot)$ is the normal cdf. One may notice that this function approaches $.16997\Lambda^2$ as Λ gets large as opposed to the desired value of $.16666\Lambda^2$. This discrepancy is immaterial for the range of Λ considered in the problems described in this chapter.

A listing of the simulated and estimated (approximate) values of MSD and ARL for a range of Λ is given in table 2.1. Figures 2.1 and 2.2 provide plots of the simulated and estimated values versus Λ of the ARL and MSD, respectively. These plots show how good equations (2.40) and (2.41) are for estimating the ARL and MSD, generally within 1% of a smoothed value of the simulations.

One of the benefits of these approximating equations is that numeric solutions to equations of the form $f(\Lambda) = R_a$ are readily obtained using standard root finder computer routines, equations (2.40) and (2.41), and their derivatives. This specific example where Λ is determined for a range of R_a is discussed more fully in the next section.

| Simulated and Estimated Values of MSD and ARL for a Range of Action Limits | | | | | |
|---|----------------|---------------|---------------|---------------|---------------|
| Action Limit | Items (1000's) | Simulated MSD | Estimated MSD | Simulated ARL | Estimated ARL |
| 0.05 | 30 | 0.996 | 1 | 1.04 | 1.04 |
| 0.1 | 30 | 0.997 | 1.001 | 1.09 | 1.09 |
| 0.15 | 30 | 0.996 | 1.002 | 1.13 | 1.13 |
| 0.2 | 30 | 0.991 | 1.003 | 1.19 | 1.18 |
| 0.25 | 30 | 1.014 | 1.006 | 1.24 | 1.24 |
| 0.3 | 30 | 1.019 | 1.009 | 1.31 | 1.3 |
| 0.35 | 30 | 1 | 1.014 | 1.38 | 1.37 |
| 0.4 | 30 | 1.015 | 1.02 | 1.45 | 1.44 |
| 0.45 | 30 | 1.03 | 1.028 | 1.51 | 1.52 |
| 0.5 | 30 | 1.044 | 1.037 | 1.6 | 1.6 |
| 0.55 | 30 | 1.049 | 1.048 | 1.68 | 1.69 |
| 0.6 | 30 | 1.055 | 1.06 | 1.79 | 1.79 |
| 0.65 | 30 | 1.062 | 1.073 | 1.89 | 1.89 |
| 0.7 | 30 | 1.084 | 1.088 | 2.02 | 2 |
| 0.75 | 30 | 1.104 | 1.104 | 2.1 | 2.12 |
| 0.8 | 30 | 1.109 | 1.121 | 2.22 | 2.24 |
| 0.85 | 30 | 1.13 | 1.14 | 2.35 | 2.37 |
| 0.9 | 30 | 1.14 | 1.159 | 2.51 | 2.5 |
| 0.95 | 30 | 1.156 | 1.18 | 2.65 | 2.64 |
| 1 | 30 | 1.191 | 1.202 | 2.8 | 2.79 |
| 1.1 | 50 | 1.241 | 1.25 | 3.08 | 3.1 |
| 1.2 | 50 | 1.294 | 1.301 | 3.42 | 3.44 |
| 1.3 | 50 | 1.35 | 1.357 | 3.76 | 3.79 |
| 1.4 | 50 | 1.4 | 1.417 | 4.18 | 4.18 |
| 1.5 | 50 | 1.47 | 1.48 | 4.59 | 4.58 |
| 1.6 | 50 | 1.548 | 1.548 | 4.98 | 5.01 |
| 1.7 | 50 | 1.618 | 1.619 | 5.47 | 5.46 |
| 1.8 | 50 | 1.69 | 1.694 | 5.96 | 5.93 |
| 1.9 | 50 | 1.772 | 1.773 | 6.46 | 6.42 |
| 2 | 50 | 1.881 | 1.855 | 6.82 | 6.93 |
| 2.1 | 50 | 1.907 | 1.941 | 7.64 | 7.47 |
| 2.2 | 50 | 2.036 | 2.031 | 7.94 | 8.02 |
| 2.3 | 50 | 2.124 | 2.124 | 8.68 | 8.59 |
| 2.4 | 50 | 2.232 | 2.221 | 9.15 | 9.19 |
| 2.5 | 50 | 2.348 | 2.322 | 9.81 | 9.8 |
| 2.6 | 50 | 2.427 | 2.426 | 10.4 | 10.43 |
| 2.7 | 50 | 2.565 | 2.534 | 10.97 | 11.09 |
| 2.8 | 50 | 2.654 | 2.645 | 11.57 | 11.76 |
| 2.9 | 50 | 2.746 | 2.76 | 12.59 | 12.45 |
| 3 | 50 | 2.909 | 2.878 | 12.87 | 13.16 |
| 3.1 | 50 | 2.983 | 3 | 13.92 | 13.89 |
| 3.2 | 50 | 3.112 | 3.125 | 14.55 | 14.64 |
| 3.3 | 50 | 3.228 | 3.254 | 15.34 | 15.41 |
| 3.4 | 50 | 3.417 | 3.387 | 15.81 | 16.2 |
| 3.5 | 50 | 3.497 | 3.522 | 17.14 | 17.01 |
| 3.6 | 50 | 3.638 | 3.662 | 17.88 | 17.83 |
| 3.7 | 50 | 3.812 | 3.804 | 18.29 | 18.68 |
| 3.8 | 50 | 3.981 | 3.95 | 18.95 | 19.55 |
| 3.9 | 50 | 4.067 | 4.1 | 20.41 | 20.43 |

Table 2.1

| Simulated and Estimated Values of MSD and ARL for a Range of Action Limits | | | | | |
|---|----------------|---------------|---------------|---------------|---------------|
| Action Limit | Items (1000's) | Simulated MSD | Estimated MSD | Simulated ARL | Estimated ARL |
| 4 | 50 | 4.272 | 4.253 | 21.21 | 21.34 |
| 4.1 | 50 | 4.402 | 4.409 | 21.9 | 22.27 |
| 4.2 | 50 | 4.606 | 4.568 | 23.25 | 23.21 |
| 4.3 | 50 | 4.757 | 4.731 | 23.92 | 24.17 |
| 4.4 | 50 | 4.889 | 4.897 | 24.55 | 25.16 |
| 4.5 | 50 | 5.032 | 5.067 | 26.29 | 26.16 |
| 4.6 | 50 | 5.177 | 5.24 | 27.11 | 27.19 |
| 4.7 | 50 | 5.307 | 5.416 | 27.72 | 28.23 |
| 4.8 | 50 | 5.528 | 5.595 | 29 | 29.29 |
| 4.9 | 50 | 5.748 | 5.778 | 29.59 | 30.37 |
| 5 | 50 | 5.983 | 5.963 | 31.17 | 31.48 |
| 5.25 | 27.6 | 6.282 | 6.442 | 33.99 | 34.32 |
| 5.5 | 30.3 | 6.828 | 6.941 | 37.44 | 37.29 |
| 5.75 | 33.1 | 7.586 | 7.46 | 40.67 | 40.38 |
| 6 | 36 | 8.101 | 7.999 | 43.22 | 43.59 |
| 6.25 | 39.1 | 8.415 | 8.558 | 47.58 | 46.93 |
| 6.5 | 42.3 | 9.151 | 9.137 | 51.84 | 50.39 |
| 6.75 | 45.6 | 9.848 | 9.736 | 52.61 | 53.98 |
| 7 | 49 | 10.403 | 10.355 | 58.47 | 57.7 |
| 7.25 | 52.6 | 11.058 | 10.994 | 59.87 | 61.54 |
| 7.5 | 56.3 | 11.762 | 11.652 | 64.29 | 65.5 |
| 7.75 | 60.1 | 11.931 | 12.331 | 69.76 | 69.59 |
| 8 | 64 | 13.023 | 13.03 | 73.23 | 73.8 |
| 8.25 | 68.1 | 13.439 | 13.748 | 79.24 | 78.14 |
| 8.5 | 72.3 | 14.467 | 14.487 | 76.62 | 82.6 |
| 8.75 | 76.6 | 15.499 | 15.246 | 86.81 | 87.19 |
| 9 | 81 | 15.974 | 16.024 | 87.85 | 91.9 |
| 9.25 | 85.6 | 16.443 | 16.824 | 96.25 | 96.74 |
| 9.5 | 90.3 | 17.46 | 17.643 | 103.14 | 101.7 |
| 9.75 | 95.1 | 18.363 | 18.482 | 104.24 | 106.79 |
| 10 | 100 | 19.423 | 19.342 | 111.48 | 112 |
| 10.25 | 105.1 | 20.119 | 20.222 | 118.05 | 117.34 |
| 10.5 | 110.3 | 21.316 | 21.123 | 120.36 | 122.8 |
| 10.75 | 115.6 | 21.676 | 22.044 | 132.83 | 128.39 |
| 11 | 121 | 23.497 | 22.985 | 128.86 | 134.1 |
| 11.25 | 126.6 | 23.614 | 23.947 | 143.82 | 139.94 |
| 11.5 | 132.3 | 25.546 | 24.929 | 141.75 | 145.9 |
| 11.75 | 138.1 | 25.863 | 25.932 | 150.56 | 151.99 |
| 12 | 144 | 27.176 | 26.956 | 154.67 | 158.2 |
| 12.25 | 150.1 | 28.056 | 28 | 169.75 | 164.54 |
| 12.5 | 156.3 | 28.695 | 29.065 | 174.97 | 171 |
| 12.75 | 162.6 | 30.128 | 30.15 | 178.44 | 177.59 |
| 13 | 169 | 31.215 | 31.257 | 179.41 | 184.3 |
| 13.25 | 175.6 | 32.484 | 32.384 | 194.85 | 191.14 |
| 13.5 | 182.3 | 33.369 | 33.531 | 204.78 | 198.1 |
| 13.75 | 189.1 | 35.262 | 34.7 | 202.64 | 205.19 |
| 14 | 196 | 35.18 | 35.889 | 212.58 | 212.4 |
| 14.25 | 203.1 | 37.291 | 37.099 | 223.15 | 219.74 |
| 14.5 | 210.3 | 38.493 | 38.33 | 224.87 | 227.2 |
| 14.75 | 217.6 | 39.249 | 39.582 | 233.94 | 234.79 |

Table 2.1

| Simulated and Estimated Values of MSD and ARL for a Range of Action Limits | | | | | |
|---|----------------|---------------|---------------|---------------|---------------|
| Action Limit | Items (1000's) | Simulated MSD | Estimated MSD | Simulated ARL | Estimated ARL |
| 15 | 225 | 41.486 | 40.854 | 236.1 | 242.5 |
| 15.25 | 232.6 | 43.654 | 42.148 | 250.07 | 250.34 |
| 15.5 | 240.3 | 44.36 | 43.462 | 254.23 | 258.3 |
| 15.75 | 248.1 | 45.374 | 44.798 | 265.02 | 266.39 |
| 16 | 256 | 46.098 | 46.154 | 270.61 | 274.6 |
| 16.25 | 264.1 | 49.159 | 47.531 | 273.36 | 282.94 |
| 16.5 | 272.3 | 48.897 | 48.929 | 288.1 | 291.4 |
| 16.75 | 280.6 | 50.222 | 50.349 | 304.63 | 299.99 |
| 17 | 289 | 52.056 | 51.789 | 300.73 | 308.7 |
| 17.25 | 297.6 | 52.969 | 53.25 | 314.55 | 317.54 |
| 17.5 | 306.3 | 55.98 | 54.732 | 316.05 | 326.5 |
| 17.75 | 315.1 | 55.703 | 56.235 | 354 | 335.59 |
| 18 | 324 | 57.514 | 57.76 | 337.85 | 344.8 |
| 18.25 | 333.1 | 61.461 | 59.305 | 344.07 | 354.14 |
| 18.5 | 342.3 | 62.1 | 60.871 | 364.87 | 363.6 |
| 18.75 | 351.6 | 63.592 | 62.459 | 374.8 | 373.19 |
| 19 | 361 | 64.69 | 64.067 | 375.26 | 382.9 |
| 19.25 | 370.6 | 68.367 | 65.697 | 380.06 | 392.74 |
| 19.5 | 380.3 | 65.981 | 67.347 | 420.17 | 402.7 |
| 19.75 | 390.1 | 68.59 | 69.019 | 417.18 | 412.79 |
| 20 | 400 | 69.714 | 70.712 | 437.64 | 423 |
| 21 | 800 | 79.253 | 77.694 | 454.29 | 465.1 |
| 22 | 800 | 86.356 | 85.015 | 500.63 | 509.2 |
| 23 | 800 | 93.24 | 92.674 | 578.45 | 555.3 |
| 24 | 800 | 100.457 | 100.671 | 618.72 | 603.4 |
| 25 | 800 | 109.523 | 109.007 | 673.4 | 653.5 |
| 26 | 800 | 119.08 | 117.682 | 680.85 | 705.6 |
| 27 | 800 | 127.505 | 126.697 | 776.7 | 759.7 |
| 28 | 800 | 136.59 | 136.05 | 833.33 | 815.8 |
| 29 | 800 | 146.89 | 145.743 | 855.62 | 873.9 |
| 30 | 800 | 160.4 | 155.775 | 927 | 934 |
| 31 | 800 | 168.468 | 166.147 | 991.33 | 996.1 |
| 32 | 800 | 176.724 | 176.858 | 1026.96 | 1060.2 |
| 33 | 800 | 189.131 | 187.909 | 1152.74 | 1126.3 |
| 34 | 800 | 199.485 | 199.299 | 1210.29 | 1194.4 |
| 35 | 800 | 208.949 | 211.029 | 1186.94 | 1264.5 |
| 36 | 800 | 225.035 | 223.099 | 1232.67 | 1336.6 |
| 37 | 800 | 241.841 | 235.509 | 1374.57 | 1410.7 |
| 38 | 800 | 258.947 | 248.258 | 1451.91 | 1486.8 |
| 39 | 800 | 254.916 | 261.347 | 1489.76 | 1564.9 |
| 40 | 800 | 273.566 | 274.776 | 1670.15 | 1645 |
| 41 | 800 | 275.671 | 288.545 | 1673.64 | 1727.1 |
| 42 | 800 | 297.201 | 302.653 | 1909.31 | 1811.2 |
| 43 | 800 | 310.63 | 317.102 | 1970.44 | 1897.3 |
| 44 | 800 | 341.768 | 331.89 | 1960.78 | 1985.4 |
| 45 | 800 | 348.118 | 347.018 | 1990.05 | 2075.5 |
| 46 | 800 | 365.568 | 362.486 | 2222.22 | 2167.6 |
| 47 | 800 | 380.132 | 378.294 | 2150.54 | 2261.7 |
| 48 | 800 | 389.189 | 394.442 | 2197.8 | 2357.8 |
| 49 | 800 | 396.647 | 410.93 | 2439.02 | 2455.9 |
| 50 | 800 | 411.437 | 427.757 | 2523.66 | 2556 |

Table 2.1

Simulated and Estimated ARLs vs. Action Limits

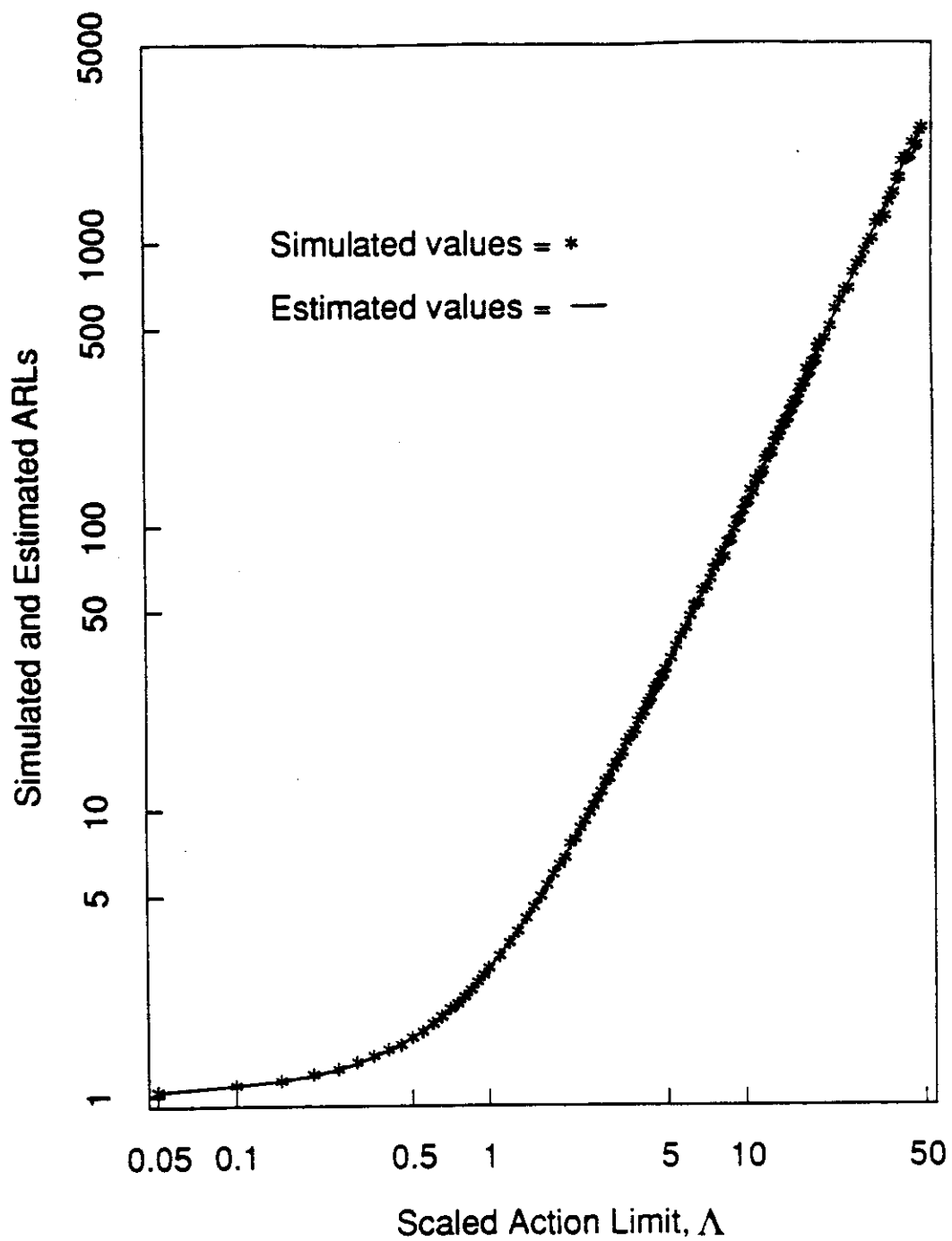


Figure 2.1

Simulated and Estimated MSDs vs. Action Limits

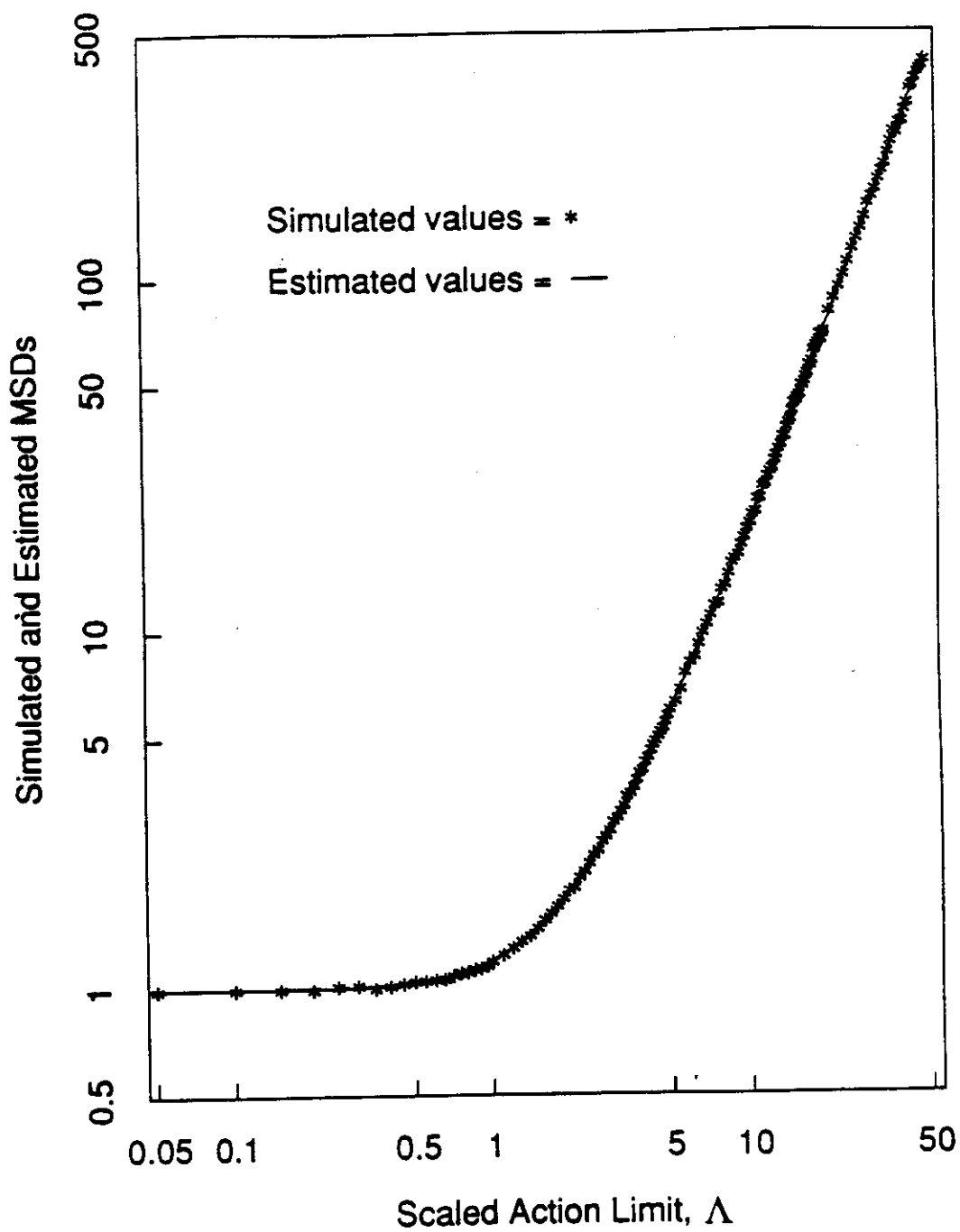


Figure 2.2

2.4 The Solution to the Machine Tool Control Problem

The previous sections showed the theoretical relationship between the relative adjustment cost, R_a , and the action limit, λ , as well as equations which provide an approximation to this theoretical relationship. Computer routines were written in the programming language C to solve numerically for Λ as a function of R_a , where $\Lambda = \lambda/\gamma_1\sigma_1$. In a manner similar to that of Shewhart control charts where the limits are determined as a multiple of σ , namely 3σ , the action limits in the machine tool control problem are calculated as multiples of $\gamma_1\sigma_1$, specifically $\Lambda\gamma_1\sigma_1$. But, unlike the Shewhart control chart in which the multiple is a constant, 3, independent of the particular application, the multiple in the machine tool control problem, Λ , depends critically on the application through the relative adjustment cost.

Table 2.2 lists values of Λ for a range of R_a . These values were obtained as output of the C computer routines and are consistent with the values listed by Box and Jenkins and also those by Crowder [17]. Crowder considered the machine tool control problem from the viewpoint of a Kalman filtering and obtained recursive equations for the minimal costs strategy with an increasing horizon. These values were compared to his when the horizon was effectively infinite and showed excellent agreement.

A simple, approximate relation between Λ and R_a is given by

$$\Lambda = (6R_a)^{1/4} - 0.593 \quad (2.42)$$

which is a slightly modified version of the equation given by Box and Jenkins [9].

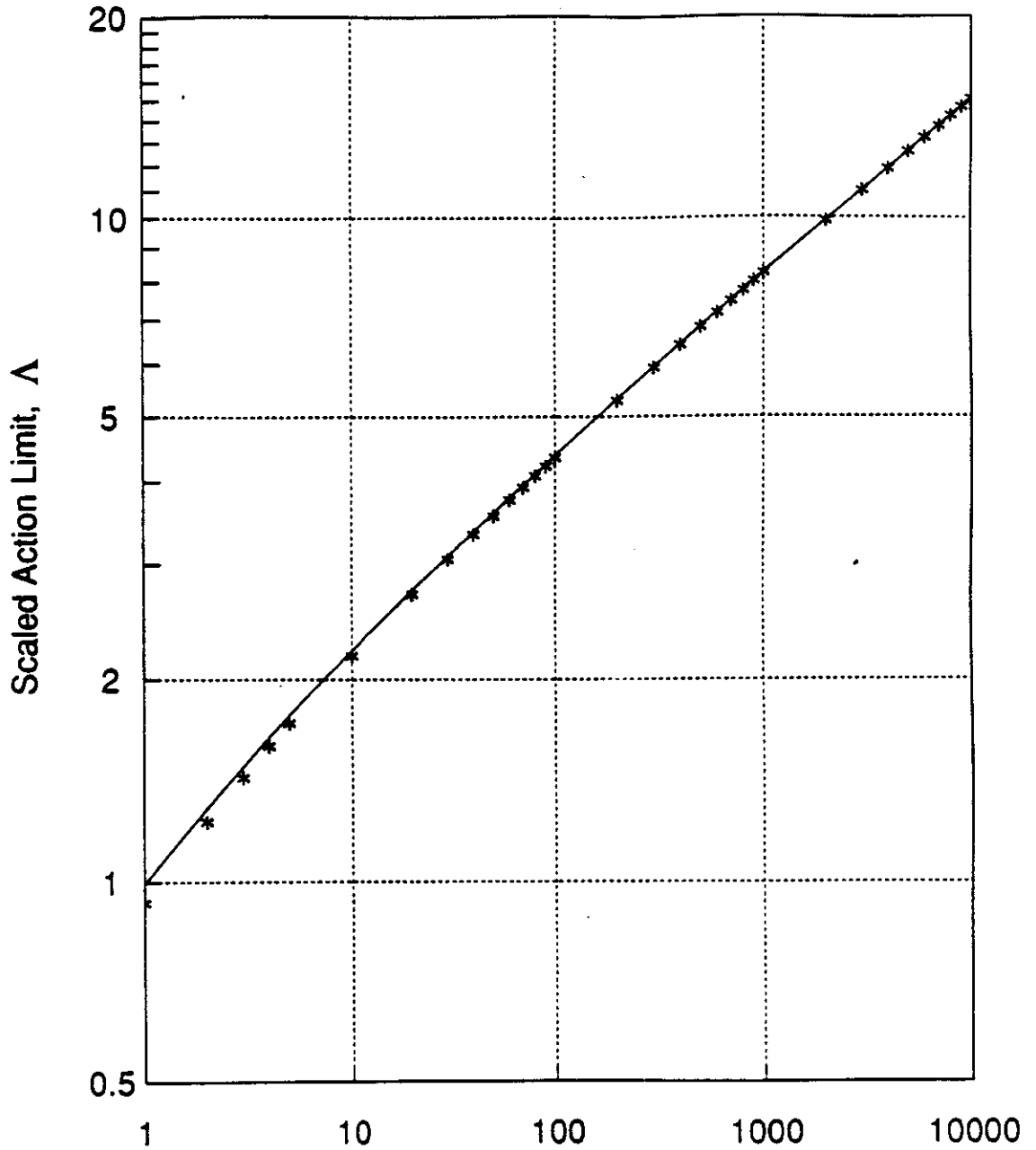
One might have guessed the order of the relationship between Λ and R_a by noticing, for large Λ , that $g(\Lambda) \approx \Lambda^2/6$, $h(\Lambda) \approx \Lambda^2$, and $f(\Lambda) \approx \Lambda^4/6$. Hence $\Lambda^4/6 \approx R_a$ and

Table 2.2: Scaled Action Limits for Relative Adjustment Costs For the Machine Tool Control Problem

| R_a | Λ | R_a | Λ |
|-------|-----------|-------|-----------|
| 1 | 0.93 | 600 | 7.15 |
| 2 | 1.23 | 700 | 7.45 |
| 3 | 1.43 | 800 | 7.73 |
| 4 | 1.59 | 900 | 7.98 |
| 5 | 1.72 | 1000 | 8.21 |
| 10 | 2.17 | 2000 | 9.87 |
| 20 | 2.70 | 3000 | 10.99 |
| 30 | 3.05 | 4000 | 11.85 |
| 40 | 3.33 | 5000 | 12.57 |
| 50 | 3.55 | 6000 | 13.18 |
| 60 | 3.75 | 7000 | 13.72 |
| 70 | 3.92 | 8000 | 14.21 |
| 80 | 4.08 | 9000 | 14.65 |
| 90 | 4.22 | 10000 | 15.06 |
| 100 | 4.34 | | |
| 200 | 5.28 | | |
| 300 | 5.91 | | |
| 400 | 6.40 | | |
| 500 | 6.80 | | |

$\Lambda \approx (6R_a)^{1/4}$. Figure 2.3 shows a plot of the tabled values of Λ , termed the *scaled action limit*, against the relative adjustment cost, R_a . The smooth line is obtained from equation (2.42) and shows excellent agreement with the tabled values of Λ for R_a greater than 10. For smaller values of R_a equation (2.42) overestimates slightly the scaled action limit, being biased high by 0.04 when $R_a = 1$. One potential problem with equation (2.42) is that *negative* values of Λ are obtained for very small values of R_a (less than 0.02). This problem of negative limits is avoided when the original

Action Limits for Machine Tool Control Problem



$$R_a = C_a / C_T = C_a / k_T \gamma_1^2 \sigma_1^2$$

Figure 2.3

Action Limits for Machine Tool Control Problem

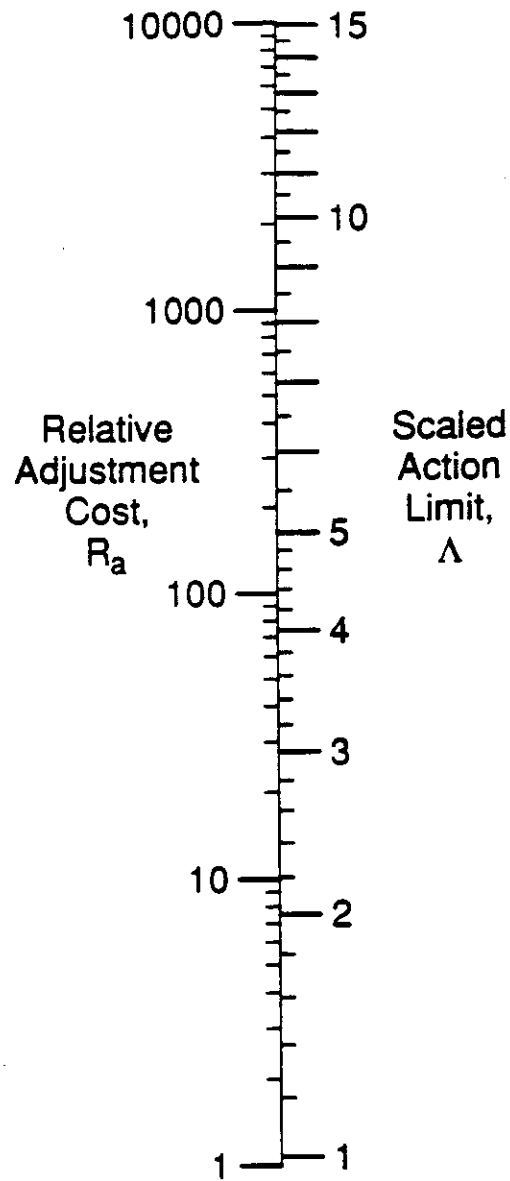


Figure 2.4

An Outline For Implementing the Machine Tool Control Scheme

1. Get Disturbance Parameters and Costs

- a) Smoothing constant, θ_1
- b) Variance of random shocks, σ_1^2
- c) $\gamma_1 = 1 - \theta_1$
- d) Adjustment cost, C_a
- e) Off-target proportionality constant, k_T . k_T may be estimated by taking the ratio of the average cost of unacceptable product, c_u , to the square of the corresponding deviation from target, d_u , $k_T = c_u / d_u^2$

2. Calculate Intermediate Constants

- a) Off-target cost constant, $C_T = k_T \gamma_1^2 \sigma_1^2$
- b) Relative adjustment cost, $R_a = C_a / C_T$

3. Get action limits

- a) Calculate scaled action limits as $\Lambda \approx (6R_a)^{1/4} - 0.593$. Λ can also be estimated from table 2.2, figure 2.3, or figure 2.4
- b) Calculate actual action limit as $\lambda = \Lambda \gamma_1 \sigma_1$

4. Adjust When Predicted Deviation Exceeds Action Limit

- a) Let e_t be the observed deviation at time t , $e_t = y_t - T$. Predicted deviations may be calculated by using the recursive equation, $\hat{e}_{t+1} = \gamma_1 e_t + \theta_1 \hat{e}_t$, where \hat{e}_{t+1} is the predicted deviation one step ahead at time t . If $|\hat{e}_{t+1}| \geq \lambda$ adjust the set point to bring predicted deviation to 0. Reset \hat{e}_{t+1} to 0 in recursion equation.

problem is rephrased such that the adjustment costs are negligible.

Figure 2.4 provides one additional way to obtain Λ as a function of R_d . This figure shows, perhaps better than the previous figure or table, the (in)sensitivity of Λ to changes in R_d . For example, if the true value of R_d is 1000 but its estimate is 20% low or high, the error is only about 5% in the estimate of Λ (7.73 or 8.62 compared to 8.21).

An outline for implementing the machine tool control scheme

The preceding theory and numerical analysis allows one to obtain action limits which minimize the control scheme costs subject to the assumptions of the nature of the disturbance and costs. Figure 2.5 provides an outline of the steps involved in implementing such a control scheme. This outline has four major elements: (1) disturbance parameters and costs are estimated, (2) intermediate constants are calculated, (3) the action limits are determined, and (4) adjustments are made when predicted deviations exceed the action limits. Two aspects of the outline have yet to be fully described, namely, the estimation of the disturbance parameters and the estimation of the predicted deviations. Three methods of estimating the disturbance parameters are described in the next section whereas the estimation of the predicted deviations is described below.

Obtaining the predicted deviations

The prediction made at time t of the one step ahead deviation, $\hat{e}_{t+1} = \hat{e}_t(1)$, is the minimum mean square error estimate of how much the process will deviate from the

desired target, $e_{t+1} = y_{t+1} - T$, at time $t+1$. When this predicted deviation exceeds the action limit, λ , an adjustment is made to the set point, X_t , so that the process is brought back to target, namely $X_{t+1} = X_t + \hat{e}_{t+1}$. After making this change to the set point the best estimate of the predicted deviation is zero. All this action hinges on being able to predict future deviations.

The observed deviations are the sum of two components, the disturbance that affects the process, z_t , and the adjustments to the set point to keep the process near the desired target, X_t . This relation was given in equation (2.6) as $e_t = z_t - X_t$ and in the differenced form, equation (2.7), as $e_{t+1} = e_t + a_{t+1} - \theta_1 a_t - x_{t+1}$. The symbol x_{t+1} represents the *change* in the set point made in the interval t to $t+1$. Using this last equation with the fact that the best estimate of a_{t+1} at time t is zero, one gets

$$\hat{e}_{t+1} = e_t - \theta_1 a_t - x_{t+1} \quad (2.43)$$

This expression depends on the unobservable random shock, a_t , which must be eliminated to obtain a useful prediction equation. One can use the relation

$$\begin{aligned} a_t &= z_t - \hat{z}_t = (z_t - X_t) - (\hat{z}_t - X_t) \\ &= e_t - \hat{e}_t \end{aligned} \quad (2.44)$$

to get the useful prediction equation

$$\hat{e}_{t+1} = \gamma_1 e_t + \theta_1 \hat{e}_t - x_{t+1} \quad (2.45)$$

This equation allows predicted deviations to be determined as a function of the observed deviation, the previous prediction, and any change made to the process to bring it back to target. Because of the equation's simple dependence on one past prediction and the current deviation, it can be readily programmed into inexpensive

calculators. Process operators can be taught to plot these predicted deviations on a chart with limits at $\pm \lambda$ and react when the limits are exceeded, just as they do now with Shewhart control charts.

An example

Suppose that the assumptions of the machine tool control problem are reasonably satisfied in an operation where a metal bar is cut to size. Specifically, assume that the monitoring cost (cost of measuring bar length) is negligible but adjustments are relatively expensive, costing \$100 every time the cutting tools are replaced. An adjustment, replacement and alignment of cutting tools, is considered as having no dynamics and no delay—there is no delay between the decision to replace the cutting tools and the start of the change. The specification for the bar size is $42.80 \pm 0.05\text{cm}$. The process normally cuts one bar every 15 seconds. Bars that are out of specification are either reworked or recycled and historical records show that the average loss for a reworked or recycled bar is \$1. Bar lengths tend to be correlated because of machine vibrations, cutting tool wear, inaccuracies in the bar positioning, and the like. One bar was measured every 5 minutes from production during the evening shift each day for a week and, using the methods of the next section, the disturbance parameters of this process were identified as $\theta_1 = 0.6$ and $\sigma_1 = 0.01$. With these data one is able to determine the minimal cost control scheme.

The outline of figure 2.5 will be used to determine the appropriate control scheme. The smoothing constant is given above as $\theta_1 = 0.6$ and the variance of the random shocks as $\sigma_1^2 = 0.0001$. $\gamma_1 = 1 - \theta_1 = 0.4$. The adjustment cost is \$100. The

off-target proportionality constant can be calculated as follows. The disturbance parameters were determined on the basis of a five minute interval. Twenty bars would be normally produced during this time and are assumed to have essentially identical lengths. Therefore, if a measured bar is at the specification limit, one assumes that the 20 bars produced during the last interval all are at or near the specification limit and need to be reworked or recycled, generating a loss of \$20. Hence, the off-target proportionality constant is $k_T = c_u/d_u^2 = 20/(.05)^2 = 8000$ and the off-target cost constant is $C_T = k_T \gamma_1^2 \sigma_1^2 = (8000)(.4)^2(.01)^2 = .128$. The relative adjustment cost is $R_a = C_a/C_T = 100/.128 = 781.25$. Equation (2.42) gives the scaled action limit as $\Lambda = (6R_a)^{1/4} - 0.593 = 7.68$ and the actual action limit as $\lambda = \Lambda \gamma_1 \sigma_1 = 0.03$. Therefore, the cutting tools should be replaced when the predicted deviation from target exceeds 0.03 cm. Equation (2.45) can be used to calculate predicted deviations as $\hat{e}_{t+1} = 0.4e_t + 0.6\hat{e}_t - x_{t+1}$. The initial value for the predicted deviation has little effect and is usually set to zero, $\hat{e}_0 = 0$.

One may wonder what properties this control scheme would have. For instance, how often would the cutting tools need to be replaced? How do the control costs split between those associated with adjustment and those with being off-target? How much are these costs per piece of production? How often are defects made? These questions are all answered below as well as simulating a potential realization of the control scheme.

Equation (2.34) separated the costs of the control scheme into the portion due to adjustment, C_a/ARL and the portion due to being off-target, $k_T \cdot MSD$. The ARL is

equivalent to $h(\lambda/\gamma_1\sigma_1) = h(\Lambda)$ and, since $\Lambda = 7.68$, $h(7.68) = 68.4$ (interpolate in table 2.1 or use equation (2.40)). This ARL is the average number of five minute intervals between adjustments so that the adjustment cost *per piece* is

$$\begin{aligned}
 &(\$100/\text{adjustment})(1 \text{ adjustment}/68.4 \text{ intervals})(1 \text{ interval}/20 \text{ pieces}) \\
 &= \$ 0.073 / \text{piece}
 \end{aligned}$$

The average of 68.4 intervals between adjustments is 342 minutes or just short of six hours.

The cost of being off-target depends on the MSD which is given by equation (2.36) as $\sigma_1^2 + \gamma_1^2\sigma_1^2g(\Lambda)$. Equation (2.41) or interpolation in table 2.1 ($g(\Lambda) = \text{MSD} - 1$ for the standardized random walk) for $\Lambda = 7.68$ gives $\hat{g}(7.68) = 10.14$ and $\text{MSD} = 0.000262$. The average loss *per interval* due to deviating from target is $k_T \cdot \text{MSD} = \$2.098$ and *per piece*, \$0.105.

The question of how often defective bars are produced is not so easily answered. The theory is developed with the philosophy that any deviation from target has some associated loss, and this loss increases as the deviation gets larger. There are no mythical limits for which bars just within the limits are “perfect” and those just outside the limits are “junk”. Rather, the quadratic loss associated with deviations provides a smooth transition between good, fair and poor quality, and motivates keeping processes regulated about a target. Having said all this, one may obtain a crude approximation to the percent defective by assuming that the mean of bars produced within an interval is uniformly distributed between $\pm \lambda$ and that the standard

deviation about this mean is σ_1 . For the problem considered here, this becomes

$$\text{Percent Defective} = 100 \left[1 - \int_{-0.03}^{+0.03} \frac{1}{0.06} \int_{-0.05}^{+0.05} \frac{1}{0.01\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{0.01}\right)^2} dx d\mu \right]$$

$$= 0.3\%$$

This estimate is likely too high as it gives too much probability to the mean being near the adjustment limits. A comparison can be made with simulating the control scheme using the above disturbance parameters and limits. This was done for 20,000 intervals and produced 16 defectives, for a simulated percentage defective of 0.08.

Figures 2.6 through 2.8 show the results of applying the control scheme for 200 intervals on a simulated data set. The observations (deviations from target bar length) are plotted as points in figure 2.6 and the predicted deviations are connected with line segments. When the predicted deviation exceeds ± 0.03 cm, an adjustment is made to bring the process back to target. Four adjustments were necessary in the simulated example, at intervals 81, 102, 138 and 188. There are very few observations near ± 0.05 cm as would be expected by the above estimates of percentage defective.

Figure 2.7 shows the set point position that would be necessary to achieve the deviations shown in figure 2.6. For most intervals, no adjustment is justified so the set point often remains at a constant level. Figure 2.8 shows the underlying disturbance. The observed process would have been equivalent to this disturbance had no adjustments been made. One can see that some adjustment is justified during the first 200 intervals as only defectives (deviation from target greater than 0.05 cm) are

Observations and Predictions with Adjustments

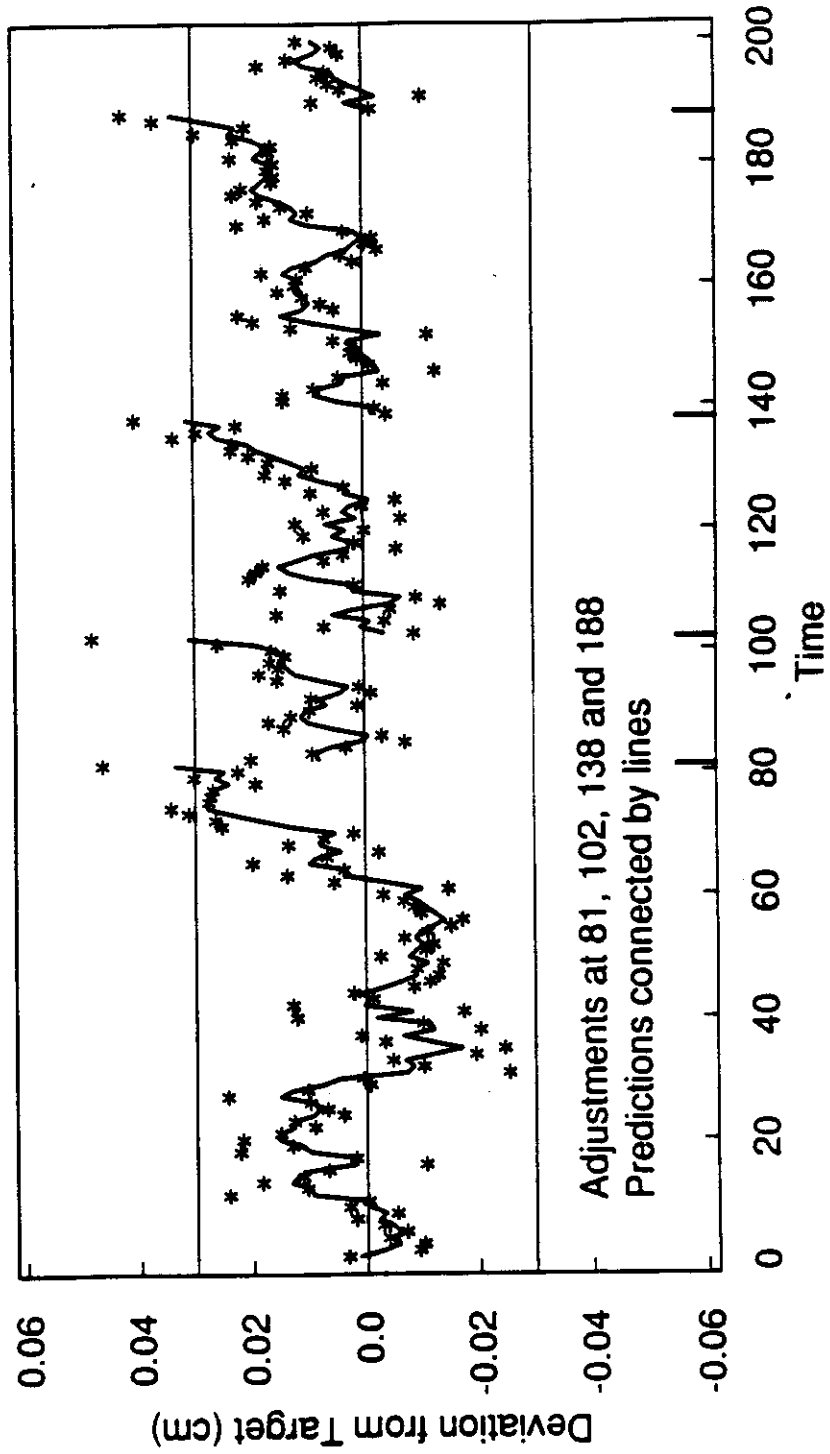


Figure 2.6

Set Point Position after Adjustments

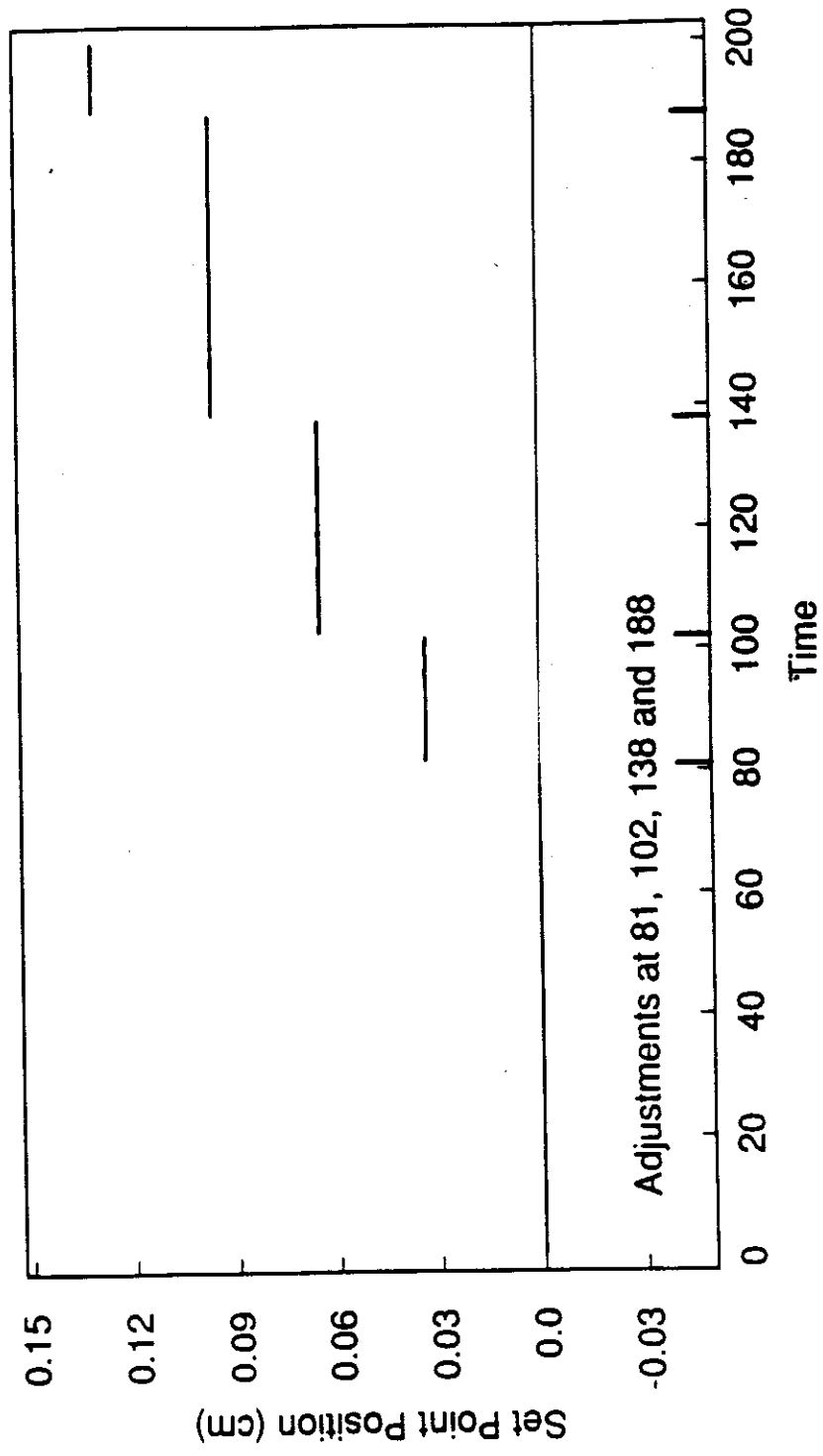


Figure 2.7

Simulated Disturbance

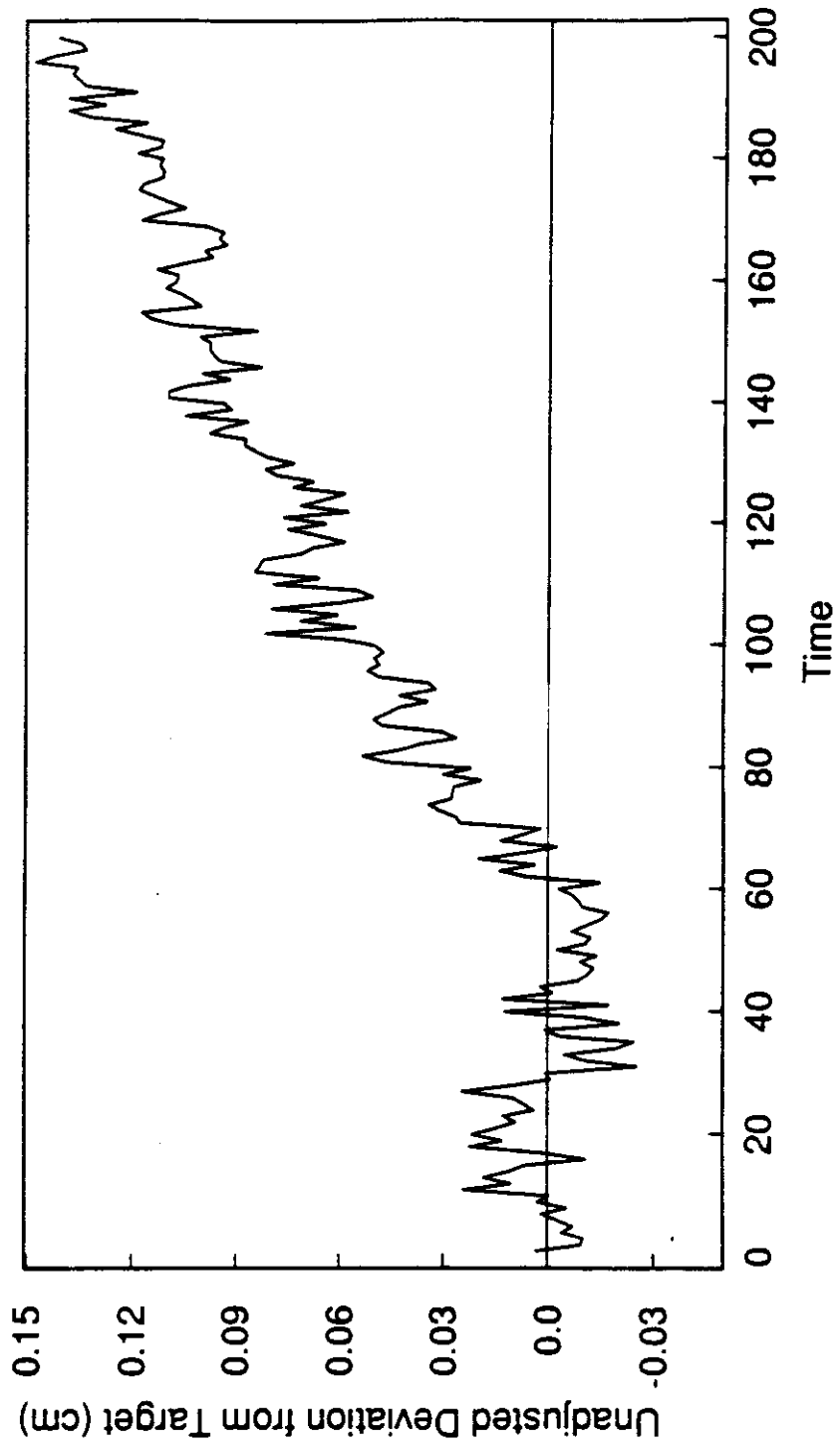


Figure 2.8

produced during the last 100 intervals.

2.5 Estimating the Disturbance Parameters

The disturbance model of equation (2.2) has several properties which are critically related to the choice of control scheme. One property is that, except when $\theta_1 = 1$ and the disturbance model is the stationary model where the errors are IID about a fixed mean, the process is nonstationary, i.e., the mean is time-dependent. A second property, related to the first, is that the variance is also time-dependent. These two properties may at first glance present an imposing barrier to maintaining a process about a target. However, the disturbance model can be separated into two components, the predictable part which depends on the smoothing constant, θ_1 , and the unpredictable part that we have called the random shock which has a mean of zero and variance, σ_1^2 . This separation into predictable and unpredictable aspects is the key to estimating the disturbance parameters while the predictability is essential to process regulation.

The separation of the disturbance model into predictable and unpredictable components is seen by expressing equation (1.4) in the following form:

$$z_{t+1} = \bar{z}_t + a_{t+1} \quad (2.46)$$

where

$$\bar{z}_t = \begin{cases} \mu & \text{when } \theta_1 = 1 \\ \gamma_1 z_t + \gamma_1 \theta_1 z_{t-1} + \gamma_1 \theta_1^2 z_{t-2} + \dots & \text{when } 0 < \theta_1 < 1 \\ z_t & \text{when } \theta_1 = 0 \end{cases} \quad (2.47)$$

The predictable part is captured in the term \bar{z}_t and the unpredictable part in a_{t+1} . This

formulation can now be used to obtain estimates of θ_1 and σ_1^2 once a sequence of the disturbance values has been obtained.

Suppose that data on deviations from target have been collected at regular intervals for the quality characteristic of interest. Estimates of the disturbance \bar{z}_t are needed and must be *reconstructed* from these data. This is accomplished by adding the value of the set point to each of the observed deviations as $z_t = e_t + X_t$. If no adjustments had been made during the data collection period reconstruction is unnecessary as constant level changes in z_t are immaterial. If data were collected on several separate occasions, one may estimate the disturbance parameters for each individual set of data and then pool these estimates to obtain some aggregate estimates, or one may piece the separate series together into one long series after allowing for possible level changes between the end of one series and the beginning of the next. (If $z_{1,n}$ is the end of the first series and $z_{2,1}$ is the beginning of the next, a constant may be added so that $z_{1,n} = z_{2,1}$. Alternatively, a constant may be added such that $z_{1,n} = z_{2,1}$ and then drop $z_{2,1}$ from the series.)

Once a sequence of the disturbance values has been obtained, one may compute $\text{Var}(z_{t+1} - \bar{z}_t)$ for a range of values of θ_1 . The value of θ_1 that minimizes this variance is an estimate of the true smoothing constant, and the variance of the differences, the residual variance, is an estimate of σ_1^2 . This method of estimating θ_1 and σ_1^2 was used for the disturbance data of figure 2.8. To get the procedure started, \bar{z}_1 was set equal to z_1 . A plot of the residual variance versus θ_1 is given in figure 2.9 with the minimum of $\hat{\sigma}_1^2 = 9.306e-5$ (where e-5 means the number is multiplied by

10^{-5}) found at $\hat{\theta}_1 = 0.525$. These data were simulated with $\theta_1 = 0.60$ and $\sigma_1^2 = 10e^{-5}$. The flatness of the residual variance curve in figure 2.9 indicates that θ_1 is not well determined. In fact, the residual variance is just 0.7% larger than the minimum when $\theta_1 = 0.6$. This uncertainty in the estimate of θ_1 has little effect on the action limits (λ is roughly proportional to $\gamma_1^{1/2}$) and only a moderate effect on the predicted deviations, $\hat{e}_{t+1} = \hat{e}_t + \gamma_1(e_t - \hat{e}_t)$.

An alternative way to get estimates of θ_1 and σ_1^2 is to use a standard time series package and fit a first-order integrated moving average model, ARIMA(0,1,1). Most time series packages make it easy to get the parameter estimates as well as check model adequacy. When the disturbance data of figure 2.8 were analyzed with the ARIMA command in MINITAB, the following estimates were obtained: $\hat{\theta}_1 = 0.529$ and $\hat{\sigma}_1^2 = 9.26e^{-5}$, supporting the results of the previous analysis.

A third method of getting disturbance parameter estimates involves plotting the variance of lagged differences, $\text{Var}(z_t - z_{t-j})$, versus the amount of differencing. This plot, developed by Jowett [24,25] to identify time series, is called a *variogram*. One of the benefits of this method is that the lagged differences for nonstationary series have finite variances. Another benefit is that it is a graphical procedure which displays patterns specific to classes of time series. For the disturbance model of equation (2.2), one gets

$$\text{Var}(z_t - z_{t-j}) = \sigma_1^2(2 - 2\gamma_1 + j\gamma_1^2), \quad \text{for } j = 1, 2, 3, \dots \quad (2.48)$$

Hence, the variogram for this model has a slope of $\gamma_1^2 \sigma_1^2$ and an intercept of $2\theta_1 \sigma_1^2$. If one lets r be the ratio of the slope to the intercept in the variogram, one can get the

Residual Variance for a Range of θ_1
Using Disturbance Data of Figure 2.8

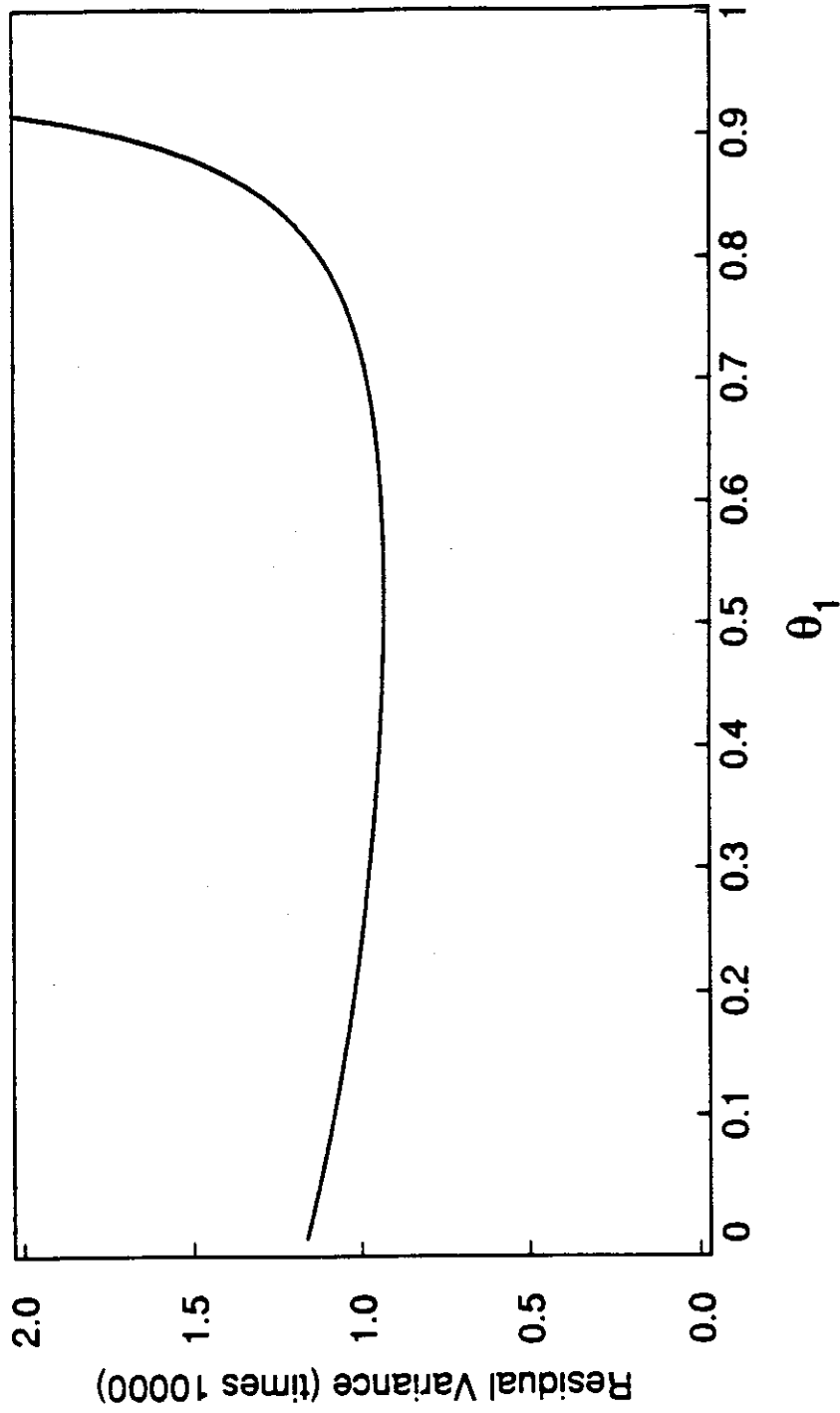


Figure 2.9

disturbance parameter estimates as

$$\begin{aligned}
 r &= \frac{\text{slope}}{\text{intercept}} \\
 \hat{\gamma}_1 &= -r + \sqrt{r^2 + 2r} \\
 \hat{\theta}_1 &= 1 - \hat{\gamma}_1 \\
 \hat{\sigma}_1^2 &= \frac{\text{intercept}}{2\hat{\theta}_1}
 \end{aligned}
 \tag{2.49}$$

These estimates may not be as good as those obtained by either of the first two methods, but they may be easier to get. This variogram also has an intuitive connection with the machine tool control scheme as the off-target cost constant, C_a , is equivalent to the proportionality constant, k_T , times $\gamma_1^2 \sigma_1^2$. This last term, $\gamma_1^2 \sigma_1^2$, is the rate at which the variance of lagged differences increases. Hence, C_a may be interpreted as the rate at which the off-target cost increases when no adjustments are made, or an *incremental non-adjustment cost*. The action limits then may be expressed as a function of the adjustment cost to the non-adjustment cost.

Figure 2.10 presents the variogram for the disturbance data of figure 2.8 through lag 20. The least squares estimates of the variogram's slope and intercept are $1.26e-5$ and $1.18e-4$, respectively. This gives a slope to intercept ratio of $r = 0.107$, and $\hat{\gamma}_1 = 0.368$, $\hat{\theta}_1 = 0.632$, $\hat{\sigma}_1^2 = 9.34e-5$. These values are consistent with those obtained by the other methods.

This section briefly described three different methods for estimating the disturbance parameters. These estimates are essential for constructing the action chart through the determination of action limits, and for utilizing the action chart through the determination of the predicted deviations. Moderate inaccuracies in the

Variogram for Disturbance Data of Figure 2.8

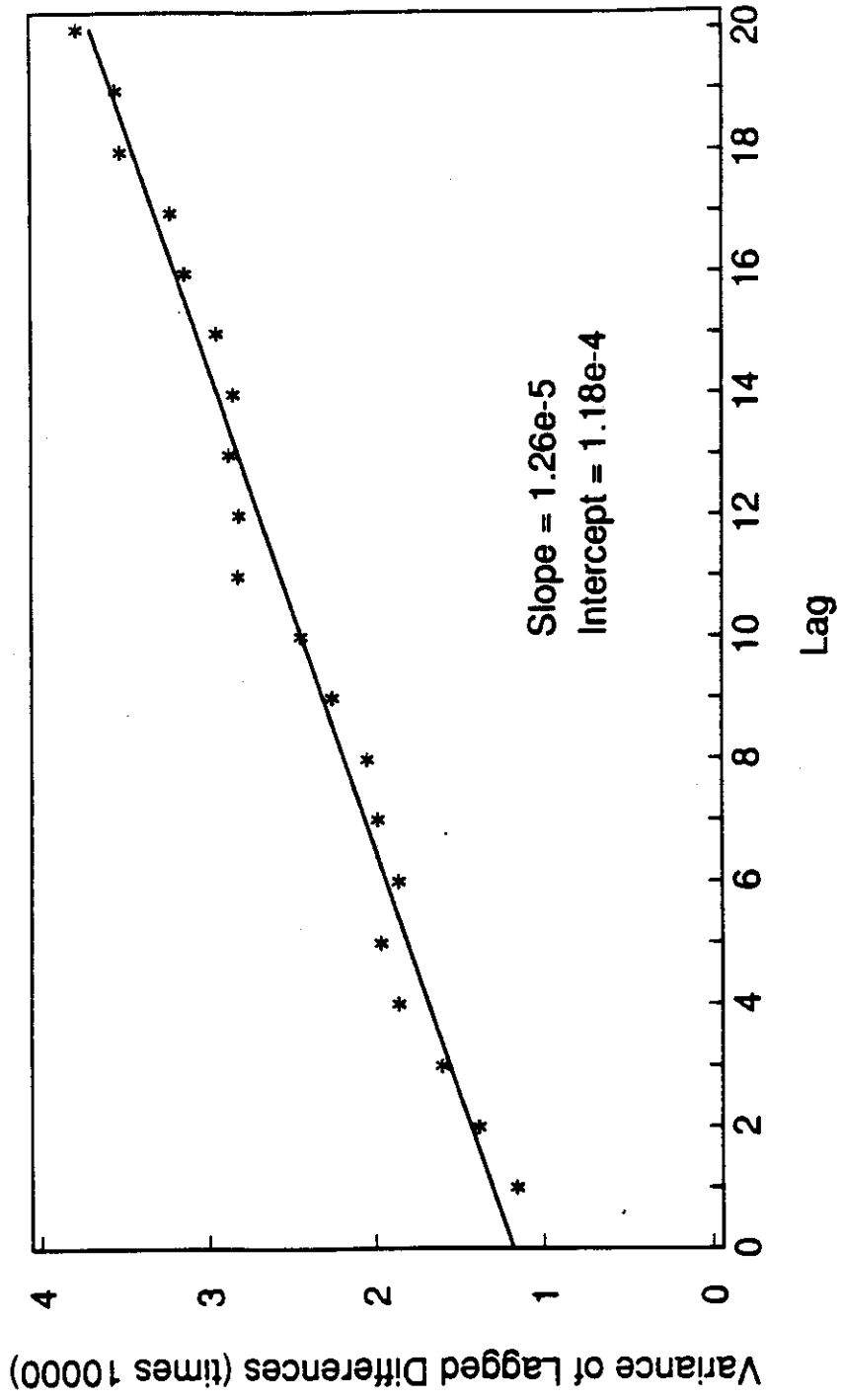


Figure 2.10

disturbance parameters should only slightly increase the cost of the control scheme relative to the minimal cost scheme as the actual action limits, $\pm\lambda$, are roughly proportional to $(\gamma_1 \sigma_1)^{1/2}$. More importantly, the formal analysis of the machine tool control costs should lead to a scheme which is right in nature even though there may be minor errors in the actual details of implementation. In addition, one may infer that, because the control scheme is insensitive to small discrepancies in θ_1 and σ_1^2 , the disturbance parameters need not be exactly constant to apply the control scheme. Hence, the control scheme should still work well in the case where the disturbance parameters are dynamic.

2.6 Comparison with Other Control Charts

The machine tool adjustment (MTA) chart that has just been described was developed to solve a specific problem, namely, when to adjust a process subject to a disturbance so as to minimize the total cost of regulation. This section discusses how this chart compares to three other control charts, the Shewhart, the exponentially weighted moving average, and the cusum chart.

Comparison with the Shewhart control chart

The Shewhart control chart, likely the most common control chart used in industry, was developed for reasons other than process regulation yet has many aspects in common with the MTA chart developed by Box and Jenkins for that purpose in the machine tool control scenario. In fact, the two charts are identical except for two aspects: (1) the Shewhart chart action criteria depends on the observed

deviation from target and not the predicted deviation, and (2) the Shewhart limits are typically set at $\pm 3\sigma$ whereas the process regulation chart's limits are set at $\pm \Lambda \gamma_1 \sigma_1$ where Λ depends on costs and disturbance parameters. Although the intentions of the two charts will always be different, one may want to know when they "appear" to be identical, i.e., when the action criteria are equivalent.

In the fullest sense, the charts can only be equivalent if the predicted deviation, \hat{e}_{t+1} , is equal to the observed deviation, e_t . From equation (2.47) one finds that this is true only when $\theta_1 = 0$, indicating that the disturbance follows a random walk. If a Shewhart chart were applied to a process in which the disturbance followed a random walk, one would quickly see patterns indicative of assignable causes (either the 3σ limits would be quickly exceeded or there would be long runs above or below the mean). Typically, the process operators, after repeatedly investigating the process to find the non-existent assignable causes, would become frustrated and remove the chart. Their reaction is justified. If the disturbance truly follows a random walk one should use a chart for process regulation, possibly in addition to a chart for detecting assignable causes. The chart for assignable causes could be a suitably modified Shewhart chart to account for the wandering character of the process level and thereby reduce the probability of searching for non-existent assignable causes.

The question of when the action limits for the Shewhart chart and the MTA chart are equal is still of interest. At a recent visit with the Statistical Consulting group at 3M, I was asked whether it was "acceptable to bring the control limits in" for a process where there were clear patterns in the deviations from target and there was an

adjustable set point. The implication behind the question was that more uniform product would be produced if one regulated the process before 3σ deviations were observed. I responded by saying that the action limits are a function of the costs, but from the given description, it indeed sounded reasonable to bring the control limits in.

One of the problems in comparing limits between the two charts is that the Shewhart chart depends on the standard deviation of rational subgroup data whereas the process regulation chart depends on the standard deviation of the random shocks. One needs an expression that relates the two when the disturbance model follows equation (2.2). This relation depends on the way in which the rational subgroup is taken which, in many cases, is left largely to the operator. Usually a rational subgroup is comprised of a small sample of items (between 2 and 6 items) taken from the process at about the same time while the process is deemed to be operating in a "typical" manner. For the purposes of obtaining explicit equations, assume that a rational subgroup is composed of a sequential set of observed deviations from the process when no adjustments were made. For example, a rational subgroup of size five may be composed of the deviations $e_{12}, e_{13}, \dots, e_{16}$ and another of deviations $e_{53}, e_{54}, \dots, e_{57}$. An estimate of σ^2 is obtained by calculating the within subgroup variance, S_n^2 :

$$S_n^2 = \frac{\sum_{i=1}^n (e_i - \bar{e})^2}{n-1}, \quad \text{where} \quad \bar{e} = \frac{\sum_{i=1}^n e_i}{n} \quad (2.50)$$

One is able to show, after much careful algebra, that the expected value of this sample variance is

$$E(S_n^2) = \sigma_1^2(1 - \gamma_1 + \gamma_1^2(n + 1)/6) \quad (2.51)$$

Figure 2.11 plots the ratio, $E(S_n^2)/\sigma_1^2$, for subgroup sizes between 2 and 10 and γ_1 between 0 and 1. For subgroups of five items or smaller, the expected subgroup variance is less than or equal to the variance of the random shocks for all values of γ_1 . When γ_1 is near zero, the ratio is approximately $\sigma_1^2(1 - \gamma_1) = \theta_1 \sigma_1^2$. Using the relation that is plotted in figure 2.11 and expressed in equation (2.51), one can find the relative adjustment cost, R_a , that justifies 3σ limits when the rational subgroups are taken as described above and the disturbance model of equation (2.2) is followed.

The relative adjustment cost that justifies $k\sigma$ limits, where $\sigma^2 = E(S_n^2)$, is

$$R_a = \frac{1}{6} \left[0.593 + \frac{k}{\gamma_1} \left[1 - \gamma_1 + \gamma_1^2 \frac{(n + 1)}{6} \right]^{1/2} \right]^4 \quad (2.52)$$

This equation depends on k , the multiple of σ , the sample size of the rational subgroups, n , and the disturbance parameter, γ_1 . Implicitly, the relative adjustment constant also depends on the disturbance parameters, γ_1 and σ_1 , but only to the extent that a ratio of adjustment to non-adjustment costs is determined. Table 2.3 lists the relative adjustment costs that justify 3σ limits for rational subgroups of size two through six. This table shows that the relative adjustment cost justifying 3σ limits increases as θ_1 gets larger. This behavior reflects the fact that, as θ_1 increases, the action limits get smaller and smaller for fixed Λ (and, correspondingly, fixed R_a). To get the equivalent of 3σ limits, both Λ and R_a must be increased as θ_1 is increased. The example given in §2.4 had $\theta_1 = 0.6$ and $R_a = 781.25$ and, for these values, table 2.3 shows that the action limits are wider than 3σ limits for rational subgroups of from two to six items. In fact, had σ been determined from a rational subgroup of five

Ratio of Rational Subgroup Variance to Random Shock Variance for Disturbance Model of Equation (2.2)

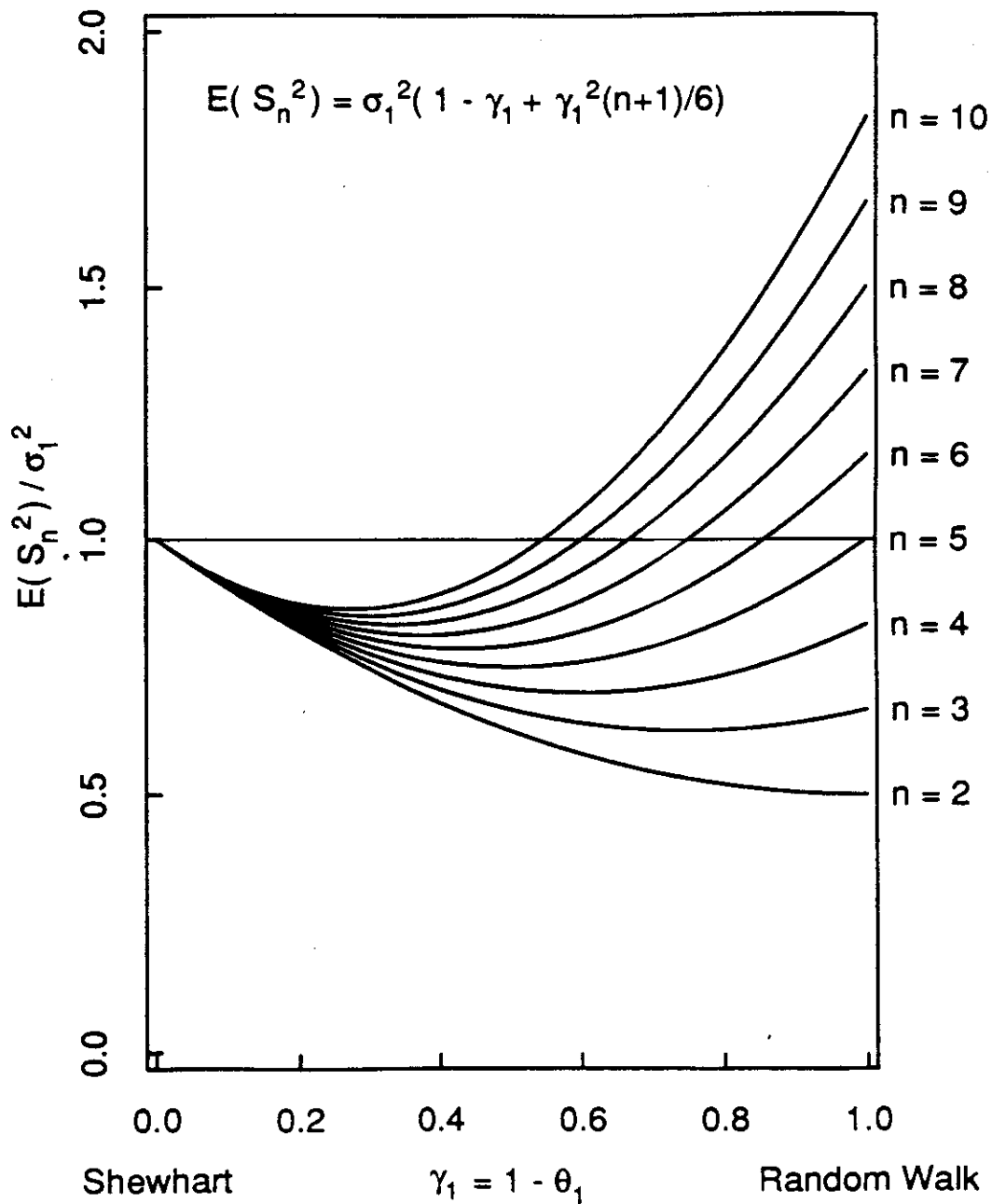


Figure 2.11

items the action limits of that example would correspond to 3.44σ . That example was contrived to generate $3\sigma_1$ limits and not the traditional Shewhart 3σ limits.

Table 2.3 can also be used to obtain a defensible answer to the question posed by the people at 3M, namely, when can the limits be brought in closer than 3σ ? Without formally analyzing cost data associated with adjustments and being off-target one can guess that many processes with an adjustable set point will have a relative adjustment cost that is quite a bit less than 781, the value for the example in §2.4. Suppose that one believes the relative adjustment cost is less than 100. Then table 2.3 indicates that the limits should be tighter than 3σ for all values of θ_1 greater than 0.5. My experience shows that this range of θ_1 includes a large portion of industrial processes.

Table 2.3: R_a That Justifies 3σ Limits

| θ_1 | γ_1 | Size of Rational Subgroup | | | | |
|------------|------------|---------------------------|----------|----------|----------|----------|
| | | 2 | 3 | 4 | 5 | 6 |
| 0.0 | 1.0 | 9 | 14 | 21 | 28 | 36 |
| 0.1 | 0.9 | 13 | 19 | 26 | 34 | 43 |
| 0.2 | 0.8 | 20 | 27 | 35 | 44 | 54 |
| 0.3 | 0.7 | 33 | 42 | 52 | 63 | 74 |
| 0.4 | 0.6 | 63 | 74 | 87 | 100 | 115 |
| 0.5 | 0.5 | 135 | 152 | 169 | 187 | 206 |
| 0.6 | 0.4 | 352 | 377 | 404 | 431 | 459 |
| 0.7 | 0.3 | 1207 | 1253 | 1299 | 1347 | 1395 |
| 0.8 | 0.2 | 6731 | 6836 | 6942 | 7049 | 7157 |
| 0.9 | 0.1 | 120048 | 120482 | 120916 | 121351 | 121787 |
| 1.0 | 0.0 | ∞ | ∞ | ∞ | ∞ | ∞ |

Comparison with the EWMA chart

The EWMA chart is a chart in which the plotted values are exponentially weighted averages of past data and action is taken when limit lines are exceeded. The limit lines are usually determined as a multiple of the standard deviation of the weighted average of IID normal variables. The motivation for using the EWMA chart is that it detects small changes (less than one σ) in a process mean more quickly than a Shewhart chart. Although it was originally developed to be a chart to detect assignable causes, it is especially sensitive to a very specific assignable cause, drifting mean. The MTA chart also uses an exponentially weighted average of past data to determine when the process mean has drifted (too far). The differences between the MTA and EWMA chart are (1) a slight difference in the motivation for each chart, (2) the way in which the exponential weight (smoothing constant) is chosen, and (3) the manner in which the limits are determined.

For the MTA chart, the limits depend on both the costs and the disturbance parameters while the smoothing constant depends only on (is equivalent to) the disturbance parameter, θ_1 . For the EWMA chart the weights and limits are chosen to give an acceptable ARL profile (a set of average run lengths associated with shifts in mean level). Hence, the EWMA chart *implicitly* associates costs with deviations from target. For example, suppose one has to choose between two EWMA charts: chart 1 has ARL=50 when the mean has shifted 0.5σ and ARL=3 when the mean has shifted 3σ ; chart 2 has ARL=200 when the mean has shifted 0.5σ and ARL=7 when the mean has shifted 3σ . Choosing chart 2 implies that one can *afford* some shifting in the

mean whereas choosing chart 1 implies that it is important to detect quickly even small changes in the mean. The relation between variation in the process level and costs is just made more explicit with the MTA chart. In practice, though, one may treat the MTA chart and the EWMA chart as being identical.

Comparison with cusum charts

The cusum chart has been used frequently in industry recently, in part, because it detects small changes in the process level more quickly than the Shewhart chart. Roberts [38] showed that a cusum's performance, as measured by the ARL for detecting shifts in mean, is very similar to that of EWMA charts. Lucas [29] and Lucas and Crosier [30] have studied the effect on the ARL of modifications to the cusum including the use of a headstart value and the tandem use of a cusum and a Shewhart chart.

There are two common, theoretically equivalent, variants of the cusum in use today. The first, the V-mask, plots the cumulative sum (cusum) of deviations from target, $C_n = \sum_{i=1}^n e_i$, on a chart. If the deviations vary about a constant level, the plot of the cusum has a slope that is equivalent to the constant level. A V-mask, actually a figure the shape of the "greater than" symbol, >, is used to detect when the slope is significantly different from 0. The V-mask is placed some distance ahead of the most recent sum and if any past sum falls outside the rays of the V, action is called for. The ARL profile depends on the distance used for the placement of the V-mask and the angle between the rays.

The second method of applying a cusum involves keeping track of two cumulative sums, one for positive deviations and the other for negative deviations.

The equations for the sums are

$$\begin{aligned} C_{H,i} &= \text{MAX}(0, C_{H,i-1} + e_i - \Delta) \\ C_{L,i} &= \text{MAX}(0, C_{L,i-1} - e_i - \Delta) \end{aligned} \quad (2.53)$$

When the process deviations are consistently positive, $C_{H,i}$ gets larger and $C_{L,i}$ goes to zero. The opposite happens when consistently negative deviations occur. Action is required if either sum exceeds an action limit, L . The action limit, L , and the tolerance, Δ , together determine the ARL profile. This second formulation of the cusum is easier to program than the first and therefore simplifies computer evaluation of the ARL.

The tolerance, Δ , tends to force both the high sum and low sum toward zero when the process level varies near zero. When the process level is larger than the tolerance, say μ_0 , then a plot of the high cusum has a constant slope up with slope $\mu_0 - \Delta$. This property allows one to identify constant level changes in the process and to estimate when the process changed to this new level (when the cusum started being non-zero).

As mentioned above, Roberts showed that the ARL profile for a cusum chart was very similar to that of an appropriately chosen EWMA chart. An alternative way to show the similarity between the two charts is to consider the *effective weight* given to past deviations, i.e., the weight given to individual deviations when a decision is made. For the EWMA chart, these weights fall off exponentially: $1 - r$, $(1 - r)r$, $(1 - r)r^2$, and so on, for the smoothing constant, r . For the cusum chart, one may

think that all deviations get equal weight as one is calculating an equally weighted sum of all past deviations. However, in the second formulation, this sum can go to zero and therefore ignore all prior data. In the first formulation, only the deviations back to the first one outside one of the rays are important. To evaluate the effective weight given to individual deviations, one may simulate a set of data and apply a cusum scheme to it. Suppose for a simulated set of data, the high cusum is first to exceed the action limit, at period 35. Suppose also that the high cusum was last zero at period 18. Then, deviations for periods one to 18 all have zero weight (since they contribute no information to the action criterion), and deviations from periods 19 to 35 all have equal weight, namely $1/17$ (since they are all equally important). Hence, for this one run, the effective weight is $1/17$ for the 17 most recent periods and 0 for all others. A different set of data would likely generate different weights. However, if this procedure was repeated for many runs and an average taken of the weights associated with the lagged periods, one could get an effective weight for a given lag associated with a specific cusum chart.

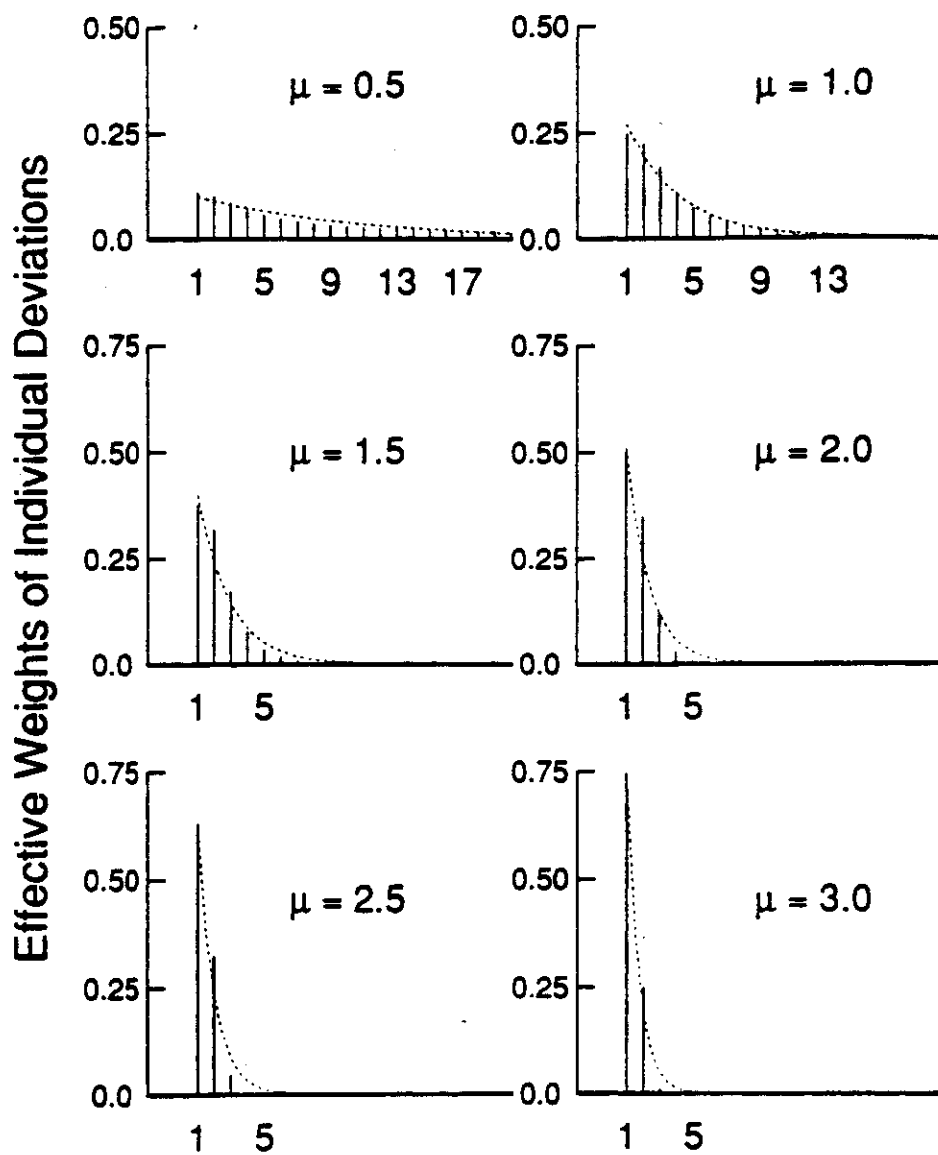
The above procedure for determining effective weights was performed for a cusum chart with a tolerance, $\Delta = 0.5$, and an action limit of 5.0. Data were generated as IID normal variables with mean, μ , and variance, 1. The effective weights were determined by calculating the average of the weights for 1000 runs. Six levels of the mean were chosen: 0.5, 1.0, 1.5, 2.0, 2.5, 3.0. The cusums were given an initial value of 2.5, corresponding to half of the action limit. The purpose of this headstart is to reduce the time to detection when the mean is clearly not zero without unduly

penalizing the ARL when the mean is zero (cf. Lucas and Crosier, [30]). The results of these simulations are plotted in figure 2.12. Superimposed on the plots of effective cusum weights are dashed curves corresponding to appropriately chosen exponential weights. The smoothing constant for each curve was chosen to give a “best” fit with the cusum weights where best means minimum sum of squared differences between the weights for the first 20 lags. Figure 2.12 shows that, dependent on the process level, this cusum chart has a pattern of weights similar to those of an EWMA chart. Though not proven here, one might presume that other cusum charts could also be paired with EWMA charts by matching effective weights.

Why the similarities?

Although the MTA chart has some similarities with the Shewhart chart, one is especially struck by the MTA chart’s similarities with the EWMA and cusum chart. The similarities between these three charts likely have some common, underlying cause or causes. I believe the charts are similar because they share two basic assumptions. First, these charts were all designed to detect shifts from target in the process level and, though the assumptions about the process and costs differ, all these charts penalize deviations from target with smaller penalties associated with small deviations and larger penalties with large deviations. Secondly, they share the assumption that information about the current process level is predominantly contained in recent observations. Taken together, these two aspects of penalizing deviations and discounting past information, leads to something like an exponentially weighted moving average with some detection limit. Rather than be surprised by the

Effective Weights of Individual Deviations
 For a Cusum Chart of $N(\mu, 1.0)$ Observations
 Limit = 5.0, Tolerance = 0.5, Headstart = 2.5



Horizontal Axis: Lag of Deviation

Figure 2.12

similarities between the charts, one should be surprised by the dominant character that these two assumptions have on the nature of the control scheme. The author was.

2.7 Control When There is a Monitoring Cost

The preceding analysis began with the assumption that the cost to adjust the process was substantial, but the cost to monitor the process, i.e., the cost associated with gathering data about the process, was negligible. In this section the assumptions are reversed. We now assume there is a substantial, fixed cost for monitoring the process, C_m , and a negligible cost for making adjustments. The disturbance is assumed to follow equation (2.2) and adjustments, when they are made, are made without dynamics or delay. The goal, as before, is to minimize the total cost of process regulation.

The effect of sampling the disturbance

When a monitoring cost is included in the process regulation framework the question of *how often to monitor* is important as well as the previous two questions of *when to adjust* and *how much to adjust*. The *monitoring interval*, m , is defined as the multiple of the initial short *base interval* at which the process is experimentally monitored. As mentioned in chapter one, only fixed monitoring intervals are considered here.

When the disturbance model of equation (2.2) is monitored (or sampled) at some multiple m of the base interval, then (as shown by Box and Jenkins [12]) the monitored disturbance, $n_t = z_{tm}$, again follows the disturbance model of equation (2.2) but with different disturbance parameters:

$$n_t = n_{t-1} + u_t - \theta_m u_{t-1} \quad (2.54)$$

where $\{u_t\}$ is a sequence of independent random shocks with variance σ_m^2 . One can relate the monitored and base disturbance parameters to obtain

$$\theta_m \sigma_m^2 = \theta_1 \sigma_1^2 \quad (2.55)$$

and

$$\frac{m\gamma_1^2}{\theta_1} = \frac{\gamma_m^2}{\theta_m} \quad (2.56)$$

where $\gamma_m = 1 - \theta_m$. These equations are used in what follows to relate regulation costs obtained for a monitoring interval to those of the base interval. Three additional equations, which are derived from the above two, are used frequently enough to justify listing them separately. The first of these derived equations

$$\gamma_m^2 \sigma_m^2 = m\gamma_1^2 \sigma_1^2 \quad (2.57)$$

may be thought of in terms of the variogram introduced in §2.5. It was shown there that the plot of $\text{Var}(z_t - z_{t-j})$ had slope $\gamma_1^2 \sigma_1^2$. Equation (2.57) can be thought of as relating the slope of this variance function for m lags to that for a unit lag. The second of these derived equations

$$\theta_m = \frac{2\theta_1 + m\gamma_1^2 - (m^2\gamma_1^4 + 4m\theta_1\gamma_1^2)^{1/2}}{2\theta_1}, \quad \text{when } 0 < \theta_1 \leq 1 \quad (2.58)$$

$$\theta_m = 0, \quad \text{when } \theta_1 = 0$$

allows one to directly obtain values for θ_m given the monitoring interval and the base parameter, θ_1 . The final equation evaluates the expected value of the cross-product between random shocks for the base and for the monitoring interval :

$$E(a_{1m}u_t) = \sigma_1^2 \quad (2.59)$$

One property that can be seen from the above equations is that the smoothing constant, θ_m , becomes smaller as the monitoring interval increases, asymptotically approaching zero. Hence, any disturbance model that satisfies equation (2.2) when monitored at long enough intervals appears to be a random walk. Figure 2.13 is a nomogram which shows how quickly the random walk is approached. Starting at θ_1 in figure 2.13, by moving to the right a distance corresponding to the monitoring interval on the m scale, we obtain the value θ_m .

The process regulation costs for a fixed monitoring interval

The preceding equations were presented so that the expected process regulation costs for a fixed monitoring interval may be determined. Once determined, the monitoring interval that minimizes the expected cost may be found.

Begin by supposing that a process is monitored every m periods and adjustments, if any, are made at that time. The expected loss for a monitoring interval is then

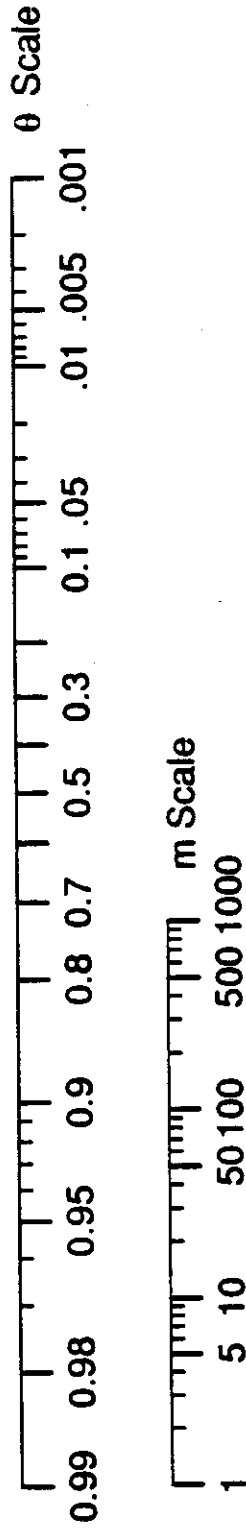
$$L = C_m + k_T \sum_{i=1}^m E(e_{tm+i}^2) \quad (2.60)$$

where the fixed monitoring cost, C_m , is accrued once per interval, there is no adjustment cost, and the sum gives the total off-target cost for the interval. The deviations from target are defined as $e_{tm+i} = z_{tm+i} - X_{tm+i}$. But since the adjustments are made only at the monitoring times, $X_{tm+1} = X_{tm+2} = \dots = X_{tm+m}$, and letting \hat{n}_{t+1} represent the predicted disturbance one monitoring interval ahead, one gets

$$E(e_{tm+i}^2) = E((z_{tm+i} - \hat{n}_{t+1})^2) + (\hat{n}_{t+1} - X_{tm+m})^2 \quad (2.61)$$

The last term is the squared difference between the predicted disturbance and the set

Determination of θ_m given θ_1 and m



Start on top scale at θ_1

Move to right distance corresponding to m

Value at this point is θ_m

For example, if $\theta_1 = 0.9$, θ_{100} is approximately 0.36

Figure 2.13

point which may be set to zero. (Hence, the optimal adjustment strategy is to adjust the set point every monitoring period so that $X_{tm+1} = \hat{n}_{t+1}$. In addition, one can show algebraically that

$$z_{tm+i} - \hat{n}_{t+1} = a_{tm+i} + \sum_{j=1}^{i-1} \gamma_1 a_{tm+j} - \theta_1 a_t + \theta_m u_t \quad (2.62)$$

Since the expectation on the right hand side is zero, \hat{n}_{t+1} is an unbiased estimate of z_{tm+i} . Therefore, the first term in equation (2.61) measures the variability of the disturbance about its predicted value. This variability may be evaluated, giving

$$\begin{aligned} E((z_{tm+i} - \hat{n}_{t+1})^2) &= \sigma_1^2 + (i-1)\gamma_1^2 \sigma_1^2 + \theta_1^2 \sigma_1^2 + \theta_m^2 \sigma_m^2 - 2\theta_1 \theta_m \sigma_1^2 \\ &= \sigma_1^2 \left[1 + (i-1)\gamma_1^2 + \theta_1(\theta_1 - \theta_m) \right] \end{aligned} \quad (2.63)$$

From these expressions we get the expected process regulation cost for one monitoring interval m as

$$L_m = C_m + mk_T \sigma_1^2 \left[1 + \frac{(m-1)}{2} \gamma_1^2 + \theta_1(\theta_1 - \theta_m) \right] \quad (2.64)$$

whence the regulation cost per base interval is

$$L_1 = \frac{C_m}{m} + k_T \sigma_1^2 \left[1 + \frac{(m-1)}{2} \gamma_1^2 + \theta_1(\theta_1 - \theta_m) \right] \quad (2.65)$$

Finding the minimal cost monitoring interval

Equation (2.65) expresses the process regulation costs as a function of the base interval disturbance parameters and the monitoring interval m . The minimal cost monitoring interval may be obtained by setting the derivative with respect to m to zero, giving

$$0 = -\frac{C_m}{m^2} + k_T \sigma_1^2 \left[\frac{\gamma_1^2}{2} - \theta_1 \frac{\partial \theta_m}{\partial m} \right] \quad (2.66)$$

where

$$\frac{\partial \theta_m}{\partial m} = \frac{\gamma_1^2}{2\theta_1} - \frac{m\gamma_1^4 + 2\theta_1\gamma_1^2}{2\theta_1(m^2\gamma_1^4 + 4m\theta_1\gamma_1^2)^{1/2}} \quad (2.67)$$

Attempts to express m in equation (2.66) as an explicit function of the disturbance parameters were fruitless. However, one can proceed by using the simpler formulation

$$m^2 \left[\frac{m\gamma_1^2 + 2\theta_1}{(m^2\gamma_1^4 + 4m\theta_1\gamma_1^2)^{1/2}} \right] = \frac{2C_m}{k_T\gamma_1^2\sigma_1^2} = 2R_m \quad (2.68)$$

where $R_m = C_m/k_T\gamma_1^2\sigma_1^2$ is the relative monitoring cost, i.e., the ratio of the fixed monitoring cost to the rate at which the off-target costs increase by not adjusting.

The monitoring interval which generates the minimal process regulation costs may now be obtained numerically as a function of θ_1 and R_m . Figure 2.14 shows optimal monitoring intervals as functions of R_m and θ_1 . The contours show, as expected, that the monitoring interval gets longer as R_m increases. They also show, possibly contrary to one's intuition, that the monitoring interval gets shorter as the disturbance takes on a more stable character (θ_1 gets larger). The reason that one would have to monitor a relatively stable process *more frequently* than a relatively unstable one for fixed R_m is connected with the definition of R_m . If R_m is to remain constant as θ_1 increases, then either the monitoring cost must decrease (making it cheaper to monitor) or the off-target constant, k_T , must increase (making it more expensive to be off-target) or both. Both of these changes justify the movement to a

Contours of Monitoring Interval
When the Monitoring Cost is Fixed
For Values of R_m and θ_1

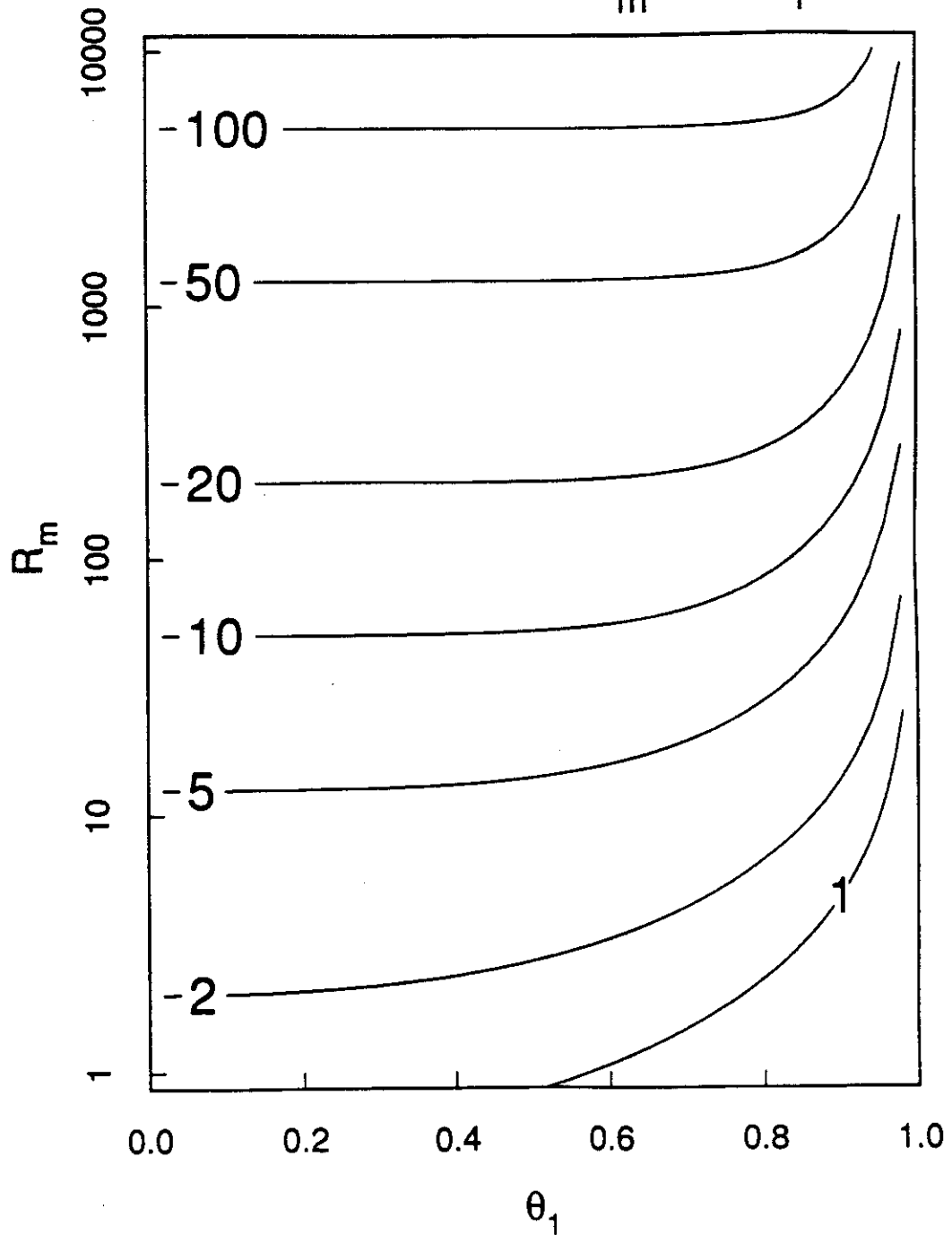


Figure 2.14

shorter, more frequent monitoring interval.

Another striking pattern in figure 2.14 is that the contours are essentially independent of the smoothing constant, θ_1 , when it is less than 0.5 or so. The reason for this behavior is seen by looking at the fraction on the left side of equation (2.68). This fraction approaches one as $m\gamma_1^2$ gets large. When this fraction is nearly one, equation (2.68) can be simplified to give

$$m = (2R_m)^{1/2} \quad (2.69)$$

which correspond to the plotted contours.

Kartha's treatment of this problem

Kartha [26] considered the previous process regulation problem but approached it in a slightly different manner. Kartha approximated the loss of equation (2.60) by treating the off-target costs as being approximately equal to $k_T(\sigma_1^2 + \sigma_m^2)/2$. This approximation is equivalent to saying that the MSD for an interval is nearly equal to the average of the variation just after adjustment, σ_1^2 , and the variation at the end of the monitoring interval, σ_m^2 . The expected loss relative to the base interval is then

$$L_K = \frac{C_m}{m} + \frac{k_T(\sigma_1^2 + \sigma_m^2)}{2} \quad (2.70)$$

Kartha set the partial derivative with respect to m to zero and, after much manipulation, was able to express m as the solution to a cubic equation. Kartha erroneously concluded that a specific root to the cubic equation was the minimal cost monitoring interval but, except for this error, Kartha's results are similar to those obtained here.

Summary of the fixed monitoring cost results

The analysis in this section showed that the minimal cost regulation scheme when there is a fixed monitoring cost and no adjustment cost is obtained by using figure 2.14. Alternatively, one may obtain a numerical root to

$$m^5 + 4\beta_1 m^4 + 4\beta_1^2 m^3 - 4R_m^2 m - 16\beta_1 R_m^2 = 0 \quad (2.71)$$

which is a reformulation of equation (2.68) with the substitution $\beta_1 = \theta_1 / \gamma_1^2$. If R_m is large or θ_1 is small, equation (2.69) provides a reasonable estimate of the minimal cost monitoring interval.

We have employed an exact expression for the loss due to being off-target whereas Kartha employed an approximation. There were several reasons for preferring the exact expression: (1) an exact expression is easily obtained, (2) it wasn't known whether Kartha's approximation was good enough, and (3) this expression is critical for the results throughout this chapter and the next.

In the two situations considered so far, in which there was either an adjustment cost or a monitoring cost, the optimal strategy depended on the ratio of this cost to the incremental cost of not adjusting the process. For the case of determining action limits, the limits depend on the fourth root of the relative adjustment cost. For the case of determining the monitoring interval, the interval depends on the square root of the relative monitoring cost. When both costs are included in one scenario, as they are in the next section, the order of these dependencies remains the same.

2.8 Fixed Monitoring and Adjustment Costs

In the previous sections we have considered process regulation when there was either a monitoring cost or an adjustment cost. Now we consider the problem of finding a control scheme that minimizes the process regulation costs when there are both costs of monitoring and of adjustment. The minimal cost control scheme will determine (1) the *monitoring interval* at which the data are collected, (2) the *action criteria* which defines when adjustments to the set point are required, and (3) the adjustment rule which specifies *how much change* to make to the set point.

Process regulation costs for a fixed monitoring interval

The minimal cost regulation scheme is determined by finding the cost associated with a fixed monitoring interval, and then finding the monitoring interval that minimizes these costs. Begin by supposing that the disturbance corresponding to the base interval, z_t , follows equation (2.2). Also suppose that the process is monitored at a multiple m of this base interval. Then, as described in the previous section, the disturbance corresponding to the monitoring interval, $n_t = z_{tm}$, follows

$$n_t = n_{t-1} + u_t - \theta_m u_{t-1} \quad (2.54)$$

where $\{u_t\}$ is a sequence of independent random shocks with variance σ_m^2 . If one is interested only in the process performance at the monitoring periods, the theory developed for the machine tool control problem applies and the process regulation cost per monitoring interval is

$$L_m = C_m + \frac{C_a}{h(\Lambda)} + k_T m \sigma_m^2 + k_T m \gamma_m^2 \sigma_m^2 g(\Lambda) \quad (2.72)$$

This equation corresponds to equation (2.27) with the following three changes: (1) the monitoring interval disturbance parameters are substituted for the base interval disturbance parameters, (2) the monitoring cost, C_m , is included, and (3) the off-target proportionality constant, k_T , has been multiplied by m because the deviation at the end of a monitoring interval now represents the deviations throughout the m periods within the interval. The first term is the cost of monitoring the process, the second is the cost of making adjustments, and the last two together are the cost of being off-target. The *per base interval* process regulation cost is therefore

$$L_1 = \frac{C_m}{m} + \frac{C_a}{mh(\Lambda)} + k_T \sigma_m^2 \left[1 + \gamma_m^2 g(\Lambda) \right] \quad (2.73)$$

Equations (2.72) and (2.73) evaluate the off-target costs of the deviations at the monitoring periods, e_{im} , without directly accounting for the deviations between monitoring periods. That is, the deviations $e_{im+1}, e_{im+2}, \dots, e_{im+m-1}$ are not explicitly included in the above costs. Rather, the equations above approximate all deviations in a monitoring interval with the deviation at the end of the interval, e_{im+m} . This approximation leads to overstating the true off-target cost. The necessary correction is derived below by evaluating the expected difference between the within interval deviations and the end interval deviation.

Correcting for the deviations between monitorings

The MSD for a monitoring interval of m periods is

$$\begin{aligned}
\frac{1}{m} \sum_{i=1}^m E(e_{tm+i}^2) &= E(e_{tm+m}^2) \\
&+ \frac{1}{m} \sum_{i=1}^m E\left[(e_{tm+m} - e_{tm+i})^2\right] \\
&- \frac{2}{m} \sum_{i=1}^m E\left[(e_{tm+m} - e_{tm+i})(e_{tm+m})\right]
\end{aligned} \tag{2.74}$$

The first line on the right of equation (2.74) corresponds to a deviation that is explicitly accounted for in the two previous equations, i.e., a deviation that occurs at the end of a monitoring interval. The second and third lines on the right correspond to the average correction needed *per base interval* to get the true MSD. The value of the set point remains constant within a monitoring interval so that one can relate the difference between deviations to random shocks:

$$\begin{aligned}
e_{tm+m} - e_{tm+i} &= z_{tm+m} - z_{tm+i} \\
&= a_{tm+m} + \gamma_1 \sum_{j=i+1}^{m-1} a_{tm+j} - \theta_1 a_{tm+i}
\end{aligned} \tag{2.75}$$

Hence, one can obtain an expression for the second line of equation (2.74):

$$\frac{1}{m} \sum_{i=1}^m E\left[(e_{tm+m} - e_{tm+i})^2\right] = \frac{(m-1)\sigma_1^2}{m} \left[\theta_1^2 + 1 + \frac{(m-2)\gamma_1^2}{2} \right] \tag{2.76}$$

Using the relation

$$e_{tm+m} = a_{tm+m} + \gamma_1 \sum_{j=1}^{m-1} a_{tm+m-j} + z_{tm} - \theta_1 a_{tm} - X_{tm} \tag{2.77}$$

one can show that the third line of equation (2.74) is

$$\begin{aligned}
-\frac{2}{m} \sum_{i=1}^m E\left[(e_{tm+m} - e_{tm+i})(e_{tm+m})\right] &= \\
&\frac{-2\sigma_1^2(m-1)}{m} \left[1 - \theta_1 \gamma_1 + \frac{(m-2)\gamma_1^2}{2} \right]
\end{aligned} \tag{2.78}$$

These equations can be combined to relate the MSD for an interval with the squared

deviation at the end of the monitoring interval:

$$\frac{1}{m} \sum_{i=1}^m E(e_{im+i}^2) = E(e_{im+m}^2) - \frac{(m-1)\gamma_1^2 \sigma_1^2}{2} \quad (2.79)$$

Hence, the amount of the overestimate is $(m-1)\gamma_1^2 \sigma_1^2 / 2$.

Note that the expected bias caused by considering only the deviations at the end of the monitoring intervals does not depend on e_{im} , X_{im} , or z_{im} . If there had been dependence then the nature of the optimal control strategy would have had to be derived for the situation of fixed adjustment and monitoring costs. However, since the bias is independent of the deviations, set point, or disturbance, the theory underlying the machine tool control problem applies, i.e., the optimal strategy is to adjust when the predicted deviation exceeds some action limits and, when justified, the adjustment to the set point should be such that it negates the predicted deviation from target. Expressions for the action limit and monitoring interval still need to be obtained but, in practical terms, the proper control scheme is obtained by using the *machine tool adjustment* (MTA) chart.

Determining the monitoring interval and action limits

The corrected expression for the process regulation costs can now be used to find the control parameters that give minimum costs. The minimum is identified in the usual way—the partial derivatives with respect to the variable parameters, m and Λ , is set to zero and the resulting equations solved simultaneously.

The corrected regulation cost *per base interval* is

$$L_1 = \frac{C_m}{m} + \frac{C_a}{mh(\Lambda)} + k_T \sigma_m^2 \left[1 + \gamma_m^2 g(\Lambda) \right] - \frac{k_T(m-1)\gamma_1^2 \sigma_1^2}{2} \quad (2.80)$$

This loss may be scaled by $C_T = k_T \gamma_1^2 \sigma_1^2$ and simplified to give

$$\frac{R_m}{m} + \frac{R_a}{mh(\Lambda)} + mg(\Lambda) + \beta_1 + \frac{(m^2 + 4m\beta_1)^{1/2}}{2} + \frac{1}{2} \quad (2.81)$$

where $\beta_1 = \theta_1 / \gamma_1^2$. Equation (2.81) shows that the optimal monitoring interval and action limits depends on R_m , the relative monitoring cost, R_a , the relative adjustment cost, and β_1 (or θ_1 or γ_1), a function of the smoothing constant. The exact dependencies are now obtained. The partial derivative with respect to Λ when set to zero gives

$$\frac{\partial L}{\partial \Lambda} = mg'(\Lambda) - \frac{R_a}{m} \frac{h'(\Lambda)}{h(\Lambda)^2} = 0 \quad (2.82)$$

which leads to

$$\frac{R_a}{m^2} = \frac{g'(\Lambda)h(\Lambda)^2}{h'(\Lambda)} = f(\Lambda) \quad (2.83)$$

Compare the above expression to equation (2.30), $f(\Lambda) = R_a$, obtained for the machine tool control problem in §2.2. These expressions relate the *scaled action limit*, Λ , to the relative adjustment cost. In §2.4, it was shown that $f(\Lambda) = \Lambda^4/6$, and for the machine tool control problem, $\Lambda = (6R_a)^{1/4}$ and $\lambda = (6R_a)^{1/4} \gamma_1 \sigma_1$. Here, one gets $\Lambda = (6R_a/m^2)^{1/4}$ and $\lambda = (6R_a/m^2)^{1/4} \gamma_m \sigma_m$. This last action limit, λ , may be simplified using equation (2.57), $\gamma_m^2 \sigma_m^2 = m \gamma_1^2 \sigma_1^2$, to give $\lambda = (6R_a)^{1/4} \gamma_1 \sigma_1$. These expressions show that the action limit, λ , is only marginally dependent on the relative monitoring cost and the monitoring interval.

The partial derivative with respect to m when set to zero gives

$$\frac{\partial L}{\partial m} = -\frac{R_m}{m^2} - \frac{R_a}{h(\Lambda)m^2} + g(\Lambda) + \frac{m + 2\beta_1}{2(m^2 + 4m\beta_1)^{1/2}} = 0 \quad (2.84)$$

Equation (2.83) may be used to eliminate m in the above expression, giving

$$0 = -\frac{R_m}{R_a}f(\Lambda) - \frac{f(\Lambda)}{h(\Lambda)} + g(\Lambda) + \frac{1 + 2\frac{\beta_1}{R_a^{1/2}}f(\Lambda)}{2\left[1 + 4\frac{\beta_1}{R_a^{1/2}}f(\Lambda)\right]^{1/2}} \quad (2.85)$$

Note that this equation can be expressed in the form

$$F\left(\frac{R_m}{R_a}, \frac{\beta_1}{R_a^{1/2}}, \Lambda\right) = 0 \quad (2.86)$$

where F is a function obtained by manipulating equation (2.85). The important aspect of equation (2.86) is that Λ depends on two parameters, R_m/R_a and $\beta_1/R_a^{1/2}$, that are functions of the original three parameters, R_m , R_a , and β_1 . This dependency of Λ on two parameters allows contours of Λ to be drawn for a grid of values of R_m/R_a and $\beta_1/R_a^{1/2}$. Such a contour plot is given in the next section along with a description of its use. One final observation about the determination of Λ —the ratio R_m/R_a is equivalent to C_m/C_a —so that one of the parameters is the ratio of the monitoring cost to the adjustment cost. This ratio will be shown to be the dominant parameter in the determination of Λ for many situations.

Equation (2.83) expressed the monitoring interval m as an implicit function of Λ and R_a . An explicit equation for m is

$$m = \left[\frac{R_a}{f(\Lambda)} \right]^{1/2} \quad (2.87)$$

An approximate equation for $f(\Lambda)$, obtained from equation (2.42), is

$$f(\Lambda) = \frac{(\Lambda + 0.593)^4}{6} \quad (2.88)$$

which leads to

$$m = \frac{2.449R_a^{1/2}}{(\Lambda + 0.593)^2} \quad (2.89)$$

Summary

This section has shown the nature of the minimal cost regulation scheme and has developed equations for determining the monitoring interval and action limits. Details and aids for implementing appropriate charts when there are monitoring and adjustment costs are given in the next section. An example is presented there including an investigation of the sensitivity of the action limits and monitoring interval to uncertainty in the costs and disturbance parameters.

2.9 Aids for Implementing an MCC Scheme

The last section derived the equations necessary to determine the action limit and monitoring interval for the minimal cost control scheme. This section presents graphical aids for implementing a minimal cost control (MCC) scheme and an example which illustrates their use. The section concludes with some simple approximations to the optimal scheme when $\beta_1 / R_a^{1/2}$ is negligibly small.

An example

The following example is hypothetical but serves to illustrate the issues surrounding the implementation of an MCC scheme. We suppose that at one stage in a

manufacturing operation involves welding two metal plates together along a common edge, forming a V-shaped trough. The usual production rate is one welded piece per minute. The strength of the joint is a critical characteristic and historical records have shown that the joint strength is directly related to weld depth. Weld depth is obtained by cutting a cross section from the trough and measuring the depth under a microscope. The specification for weld depth is 0.140 ± 0.010 mm (140 ± 10 μ m). Welded pieces that are outside of the specification must be scrapped at a cost of \$4 each. The total cost of getting a weld depth measurement is estimated as \$20, made up of \$4 for the piece that is cut, \$8 for the labor, and \$8 for the use of the cutting machine and microscope. The weld depth is adjusted by stopping the machine, replacing and aligning the tip of the welding rod. The total estimated cost for an adjustment to the welding machine is \$60. This total cost is made up of \$24 for downtime, \$6 for labor, and \$30 for the new rod tip. One is interested in knowing the characteristics of the machine tool adjustment scheme for this example.

An outline for implementing the machine tool adjustment scheme

Figure 2.15 is an outline for implementing the machine tool adjustment scheme. This outline separates the implementation into four basic steps, namely: (1) get disturbance parameters and costs, (2) calculate intermediate constants, (3) determine action limit and monitoring interval, and (4) adjust when predicted deviation exceeds this action limit. This outline will be used to guide the analysis of this example.

The first step in implementing the machine tool adjustment scheme is to get the disturbance parameters and costs. Although the manager for the welding operation

An Outline For Implementing the Minimum Cost Control Scheme When There Are Monitoring and Adjustment Costs

1. Get Disturbance Parameters and Costs
 - a) Smoothing constant, θ_1
 - b) Variance of random shocks, σ_1^2
 - c) $\gamma_1 = 1 - \theta_1$
 - d) Monitoring cost, C_m
 - e) Adjustment cost, C_a
 - f) Off-target proportionality constant, k_T . k_T may be estimated by taking the ratio of the average cost of unacceptable product, c_u , to the square of the corresponding deviation from target, d_u , $k_T = c_u / d_u^2$

2. Calculate Intermediate Constants
 - a) Off-target cost constant, $C_T = k_T \gamma_1^2 \sigma_1^2$
 - b) Relative adjustment cost, $R_a = C_a / C_T$
 - c) Function of smoothing constant, $\beta_1 = \theta_1 / \gamma_1^2$
 - d) First parameter for determining Λ , the ratio of monitoring to adjustment cost, C_m / C_a
 - e) Second parameter for determining Λ , $\beta_1 / R_a^{1/2}$

Figure 2.15

3. Determine action limits and monitoring interval

a) Calculate scaled action limit, Λ , by referring to figures 2.16, 2.17 or 2.18. One may also interpolate in table 2.5 or solve for Λ numerically using equation (2.85).

b) Determine the monitoring interval as

$$m = \left[\frac{R_a}{f(\Lambda)} \right]^{1/2} = \frac{2.449R_a^{1/2}}{(\Lambda + 0.593)^2} \quad (2.89)$$

c) Determine θ_m using figure 2.13 or equation (2.58).

d) Calculate actual action limit as $\lambda = \Lambda \gamma_m \sigma_m = \Lambda m^{1/2} \gamma_1 \sigma_1$.

4. Adjust When Predicted Deviation Exceeds Action Limit

a) Let e_{tm} be the observed deviation at time tm . Predicted deviations may be calculated by using the recursive equation, $\hat{e}_{tm+m} = \gamma_m e_{tm} + \theta_m \hat{e}_{tm}$, where \hat{e}_{tm+m} is the predicted deviation m periods ahead at time tm . If $|\hat{e}_{tm+m}| \geq \lambda$ adjust the set point to bring predicted deviation to 0. Reset \hat{e}_{tm+m} to 0 in recursion equation.

Figure 2.15, continued

Some Useful Equations for Implementing an MTA Chart

$$h(\Lambda) = \frac{(1 + 1.1\Lambda + \Lambda^2)}{(1 - 0.115\exp(-9.2(\Lambda^{0.3} - 0.88)^2))} \quad (2.40)$$

$$g(\Lambda) = \frac{(1 + 0.06\Lambda^2)}{1 - 0.647 \cdot \Phi(1.35(\ln(\Lambda) - 0.67))} - 1 \quad (2.41)$$

where $\Phi(\cdot)$ is the normal cdf.

$$\theta_m = \frac{2\theta_1 + m\gamma_1^2 - (m^2\gamma_1^4 + 4m\theta_1\gamma_1^2)^{1/2}}{2\theta_1}, \quad \text{when } 0 < \theta_1 \leq 1 \quad (2.58)$$

$$\theta_m = 0, \quad \text{when } \theta_1 = 0$$

$$0 = -\frac{R_m}{R_a}f(\Lambda) - \frac{f(\Lambda)}{h(\Lambda)} + g(\Lambda) + \frac{1 + 2\frac{\beta_1}{R_a^{1/2}}f(\Lambda)}{2\left[1 + 4\frac{\beta_1}{R_a^{1/2}}f(\Lambda)\right]^{1/2}} \quad (2.85)$$

$$f(\Lambda) = g'(\Lambda)h(\Lambda)^2/h'(\Lambda) = \frac{(\Lambda + 0.593)^4}{6} \quad (2.88)$$

Figure 2.15, continued

believes that monitoring once per hour ought to be sufficient, to be on the safe side, has one weld joint analyzed every 15 minutes for a week. The methods of §2.5 are used to analyze these weld depth data and estimates of the disturbance parameters are obtained, $\hat{\theta}_1 = 0.7$ and $\hat{\sigma}_1 = 3 \mu\text{m}$. Although these values are estimates of the true disturbance parameters, they are treated as the true values. When $\theta_1 = 0.7$, $\gamma_1 = 0.3$. The monitoring cost, C_m , is given as \$20 and the adjustment cost, C_a , as \$60. The proportionality constant, k_T , is not given directly but may be determined from the above data. Fifteen minutes of production represents fifteen troughs at \$4 each, or \$60. These troughs are defective when the weld depth deviates from target by $10\mu\text{m}$ or more, leading to an estimate of k_T as $k_T = 60/10^2 = 0.6$ dollars per squared μm . This completes the first step in the outline in which the disturbance parameters and costs are obtained; the values are listed in table 2.4.

| Disturbance Parameter and Costs for the Example | |
|--|---------------------------------|
| Parameter | Value |
| Smoothing constant, θ_1 | 0.7 |
| Variance of random shocks, σ_1^2 | $9\mu\text{m}^2$ |
| $\gamma_1 = 1 - \theta_1$ | 0.3 |
| Monitoring cost, C_m | \$20 |
| Adjustment cost, C_a | \$60 |
| Off-target proportionality constant, k_T . | 0.6 dollars per μm^2 |
| Off-target cost constant, $C_T = k_T \gamma_1^2 \sigma_1^2$ | \$0.486 |
| Relative adjustment cost, $R_a = C_a / C_T$ | 123.5 |
| Function of smoothing constant, $\beta_1 = \theta_1 / \gamma_1^2$ | 7.778 |
| Ratio of monitoring to adjustment cost, C_m / C_a | 0.333 |
| Second parameter for determining Λ , $\beta_1 / R_a^{1/2}$ | 0.700 |

Table 2.4

The second step in the outline is to calculate the intermediate constants, the first of which is the off-target constant, $C_T = k_T \gamma_1^2 \sigma_1^2 = 0.486$ dollars. The second constant is the relative adjustment cost, $R_a = C_a / C_T = 123.5$. The third constant is β_1 , a function of the smoothing constant, $\beta_1 = \theta_1 / \gamma_1^2 = 7.778$. The ratio of the monitoring to adjustment cost is $C_m / C_a = 0.333$. Finally, the second parameter for calculating Λ is obtained as $\beta_1 / R_a^{1/2} = 0.700$. This completes the second step and these values are also given in table 2.4.

The third step in the outline is to determine the action limit and monitoring interval. The scaled action limit is obtained by referring to figures 2.16, 2.17, 2.18 or table 2.5. When $\beta_1 / R_a^{1/2} = 0.700$ and $C_m / C_a = 0.333$, figure 2.16 gives Λ as approximately 1. Table 2.5 gives $\log_{10}(\Lambda)$ as a function of $\log_{10}(C_m / C_a)$ and $\log_{10}(\beta_1 / R_a^{1/2})$. The logarithms of the parameters and scaled action limit are listed. This listing format is chosen so that a wide range of costs could be represented in a single table and, for such a wide range of costs, the contours of the logarithm of the scaled action limits lie nearly on a planar surface. This formulation enables linear interpolation to provide good estimates of the scaled action limit. Linear interpolation in table 2.5 with $\log_{10}(C_m / C_a) = -0.4776$ and $\log_{10}(\beta_1 / R_a^{1/2}) = -0.155$ gives $\log_{10}(\Lambda) = 0.0172$ or $\Lambda = 1.04$. Estimates of Λ from the figures or table 2.5 are usually acceptable approximations. However, the most accurate estimate of Λ is found by evaluating equation (2.85) giving $\Lambda = 1.021$. Note that all the alternative ways of estimating Λ give values of one or so. This value of Λ can now be used to determine the monitoring interval. When the approximation to $f(\Lambda)$ given in

Λ as a Function of
 $\beta_1 / R_a^{0.5}$ and C_m / C_a

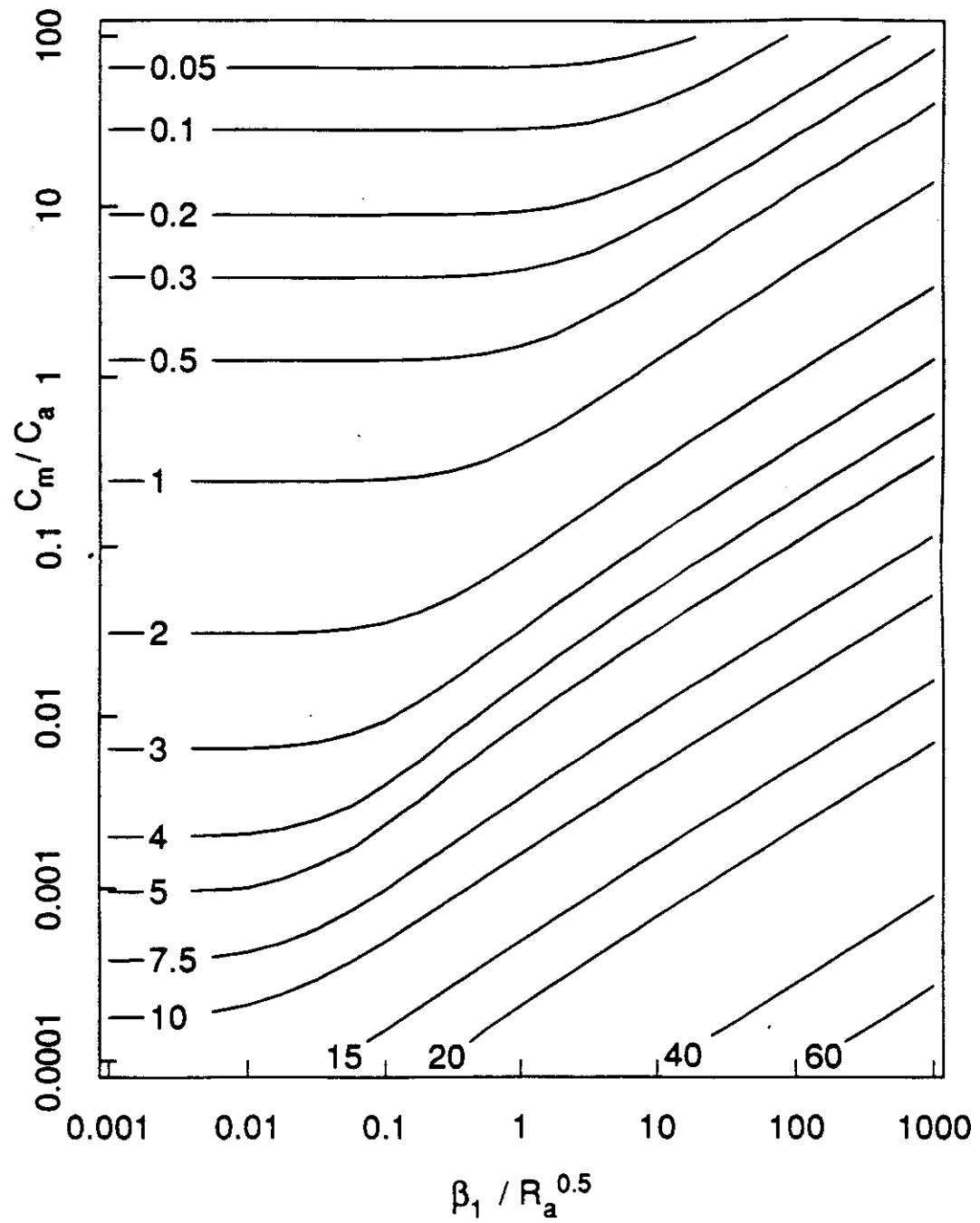


Figure 2.16

Contours of Λ versus C_m / C_a
for Values of $\beta_1 / R_a^{0.5}$

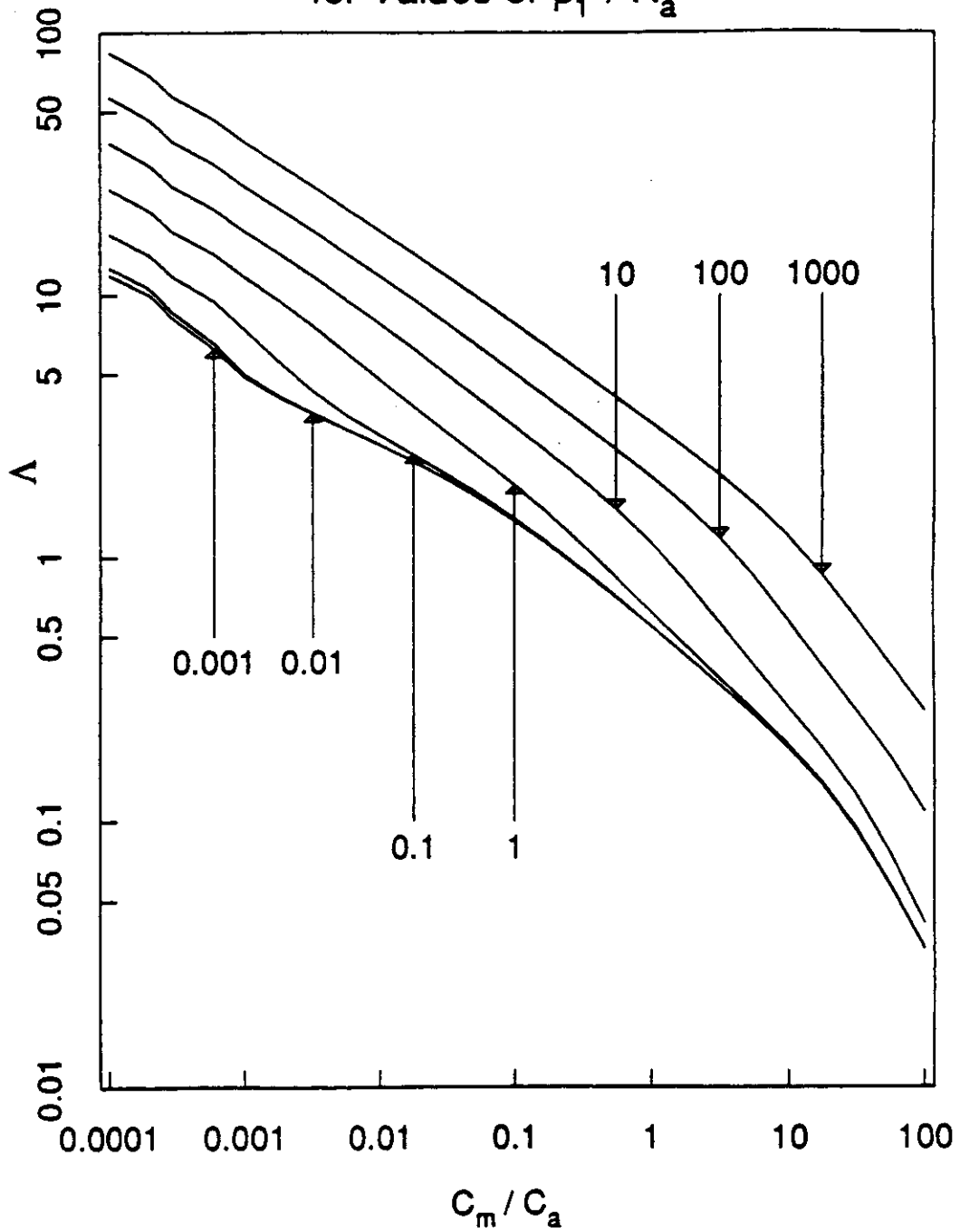


Figure 2.17

Contours of Λ versus $\beta_1 / R_a^{0.5}$
 for Values of C_m / C_a

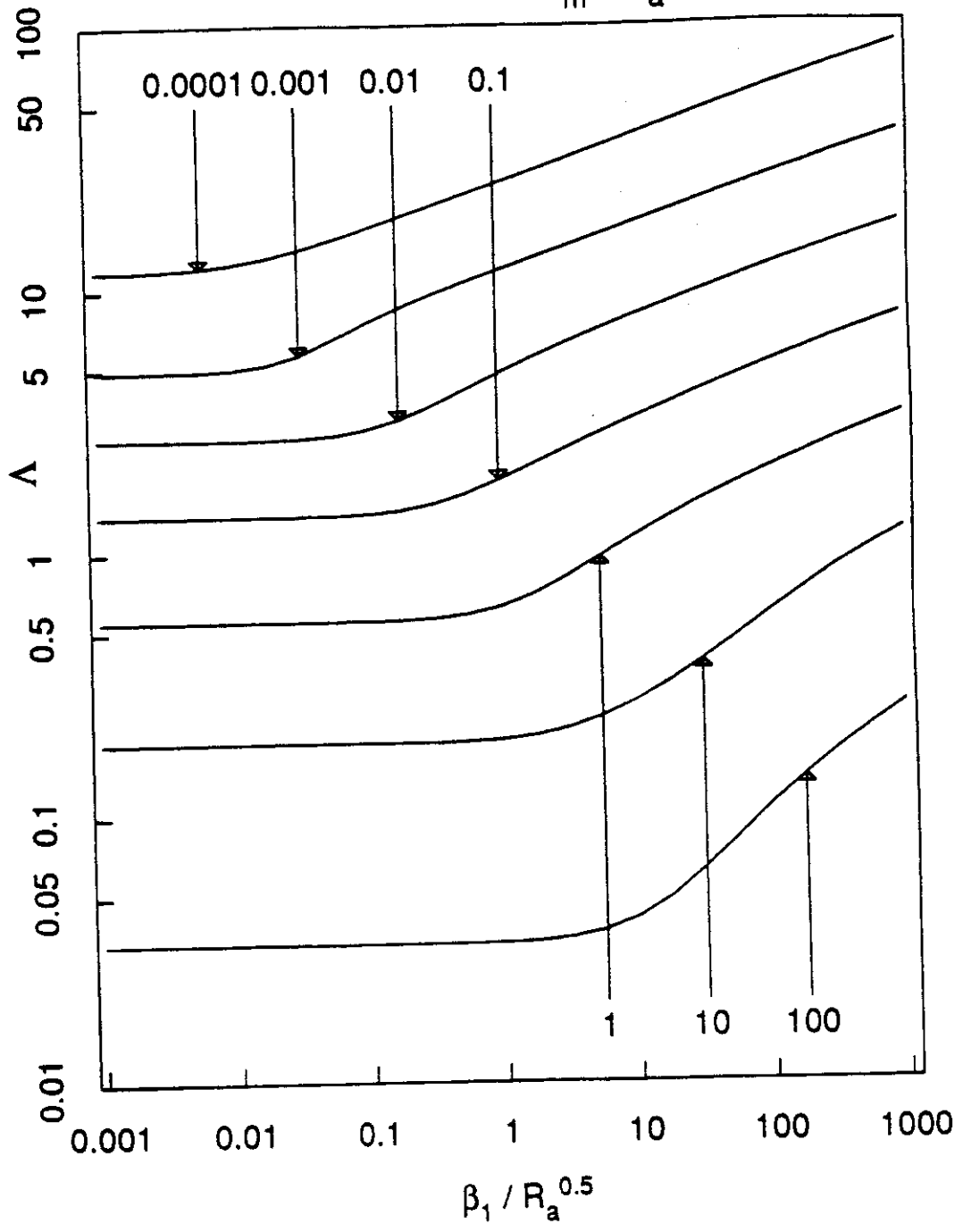


Figure 2.18

$\log_{10}(\Lambda)$ as a function of $\log_{10}(C_m/C_a)$ and $\log_{10}(\beta_1/R_a^{1/2})$

| $\log_{10}(C_m/C_a)$ | $\log_{10}(\beta_1/R_a^{1/2})$ | | | | | | |
|----------------------|--------------------------------|--------|--------|--------|--------|--------|--------|
| | -3.0 | -2.5 | -2.0 | -1.5 | -1.0 | -0.5 | 0.0 |
| -4.0 | 1.074 | 1.080 | 1.102 | 1.155 | 1.231 | 1.316 | 1.404 |
| -3.5 | 0.911 | 0.915 | 0.934 | 0.990 | 1.071 | 1.156 | 1.241 |
| -3.0 | 0.690 | 0.691 | 0.701 | 0.754 | 0.872 | 0.979 | 1.072 |
| -2.5 | 0.550 | 0.550 | 0.553 | 0.568 | 0.637 | 0.769 | 0.885 |
| -2.0 | 0.430 | 0.430 | 0.431 | 0.437 | 0.468 | 0.557 | 0.680 |
| -1.5 | 0.298 | 0.298 | 0.298 | 0.300 | 0.316 | 0.374 | 0.479 |
| -1.0 | 0.140 | 0.140 | 0.140 | 0.141 | 0.149 | 0.186 | 0.278 |
| -0.5 | -0.047 | -0.047 | -0.047 | -0.047 | -0.043 | -0.023 | 0.050 |
| 0.0 | -0.261 | -0.261 | -0.261 | -0.260 | -0.259 | -0.250 | -0.206 |
| 0.5 | -0.486 | -0.486 | -0.486 | -0.486 | -0.486 | -0.482 | -0.461 |
| 1.0 | -0.725 | -0.725 | -0.725 | -0.725 | -0.725 | -0.724 | -0.713 |
| 1.5 | -1.037 | -1.037 | -1.037 | -1.037 | -1.037 | -1.036 | -1.030 |
| 2.0 | -1.481 | -1.481 | -1.481 | -1.481 | -1.481 | -1.480 | -1.477 |

| $\log_{10}(C_m/C_a)$ | $\log_{10}(\beta_1/R_a^{1/2})$ | | | | | | |
|----------------------|--------------------------------|--------|--------|--------|--------|--------|--------|
| | 0.0 | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 |
| -4.0 | 1.404 | 1.492 | 1.580 | 1.667 | 1.753 | 1.838 | 1.922 |
| -3.5 | 1.241 | 1.327 | 1.414 | 1.500 | 1.585 | 1.670 | 1.755 |
| -3.0 | 1.072 | 1.159 | 1.246 | 1.331 | 1.417 | 1.502 | 1.587 |
| -2.5 | 0.885 | 0.983 | 1.074 | 1.161 | 1.247 | 1.333 | 1.418 |
| -2.0 | 0.680 | 0.793 | 0.893 | 0.985 | 1.074 | 1.162 | 1.248 |
| -1.5 | 0.479 | 0.594 | 0.701 | 0.801 | 0.895 | 0.986 | 1.075 |
| -1.0 | 0.278 | 0.393 | 0.504 | 0.609 | 0.708 | 0.804 | 0.896 |
| -0.5 | 0.050 | 0.174 | 0.298 | 0.410 | 0.514 | 0.613 | 0.710 |
| 0.0 | -0.206 | -0.093 | 0.057 | 0.192 | 0.309 | 0.416 | 0.517 |
| 0.5 | -0.461 | -0.390 | -0.249 | -0.079 | 0.072 | 0.200 | 0.313 |
| 1.0 | -0.713 | -0.671 | -0.570 | -0.417 | -0.239 | -0.068 | 0.078 |
| 1.5 | -1.030 | -0.998 | -0.904 | -0.753 | -0.586 | -0.412 | -0.231 |
| 2.0 | -1.477 | -1.459 | -1.384 | -1.205 | -0.969 | -0.764 | -0.585 |

Table 2.5

equation (2.88) is used, one gets $m = 10.70$. A better estimate of m is obtained by fully evaluating the functions $h(\Lambda)$, $g'(\Lambda)$, and $h'(\Lambda)$ and using these to evaluate $f(\Lambda)$. Doing this gives $f(\Lambda) = 1.232$ and $m = 10.01$. Hence the monitoring interval should be ten times the base interval, or every $2\frac{1}{2}$ hours (150 minutes). The smoothing constant corresponding to this monitoring interval, θ_{10} , is evaluated using equation (2.58) as 0.34. Finally, the actual action limit, λ , is evaluated as $2.9\mu\text{m}$.

The monitoring interval and action limit may now be used to construct the minimal cost process regulation scheme. The process is monitored every $2\frac{1}{2}$ hours and the predicted deviation from target is plotted on a chart with parallel action limit lines at $\pm 2.9\mu\text{m}$. The process is adjusted when the predicted deviation from target exceeds these limits. The predicted deviation is equal to 0.66 times the current deviation plus 0.34 times the previous predicted deviation. Figure 2.19 shows a potential realization of this scheme. The plotted points are the deviations from the target of $140\mu\text{m}$ for troughs that could have been measured every 15 minutes. The weld depth for every tenth one of these points is actually measured and these points have been circled. The predicted values are calculated using the weld depths of the circled points and are denoted by the connected short horizontal lines. When these horizontal lines are beyond the limit lines at $\pm 2.9\mu\text{m}$, an adjustment is made bringing the predicted weld depth back to target.

Table 2.1 may be used to estimate how often the welding rod tips will have to be replaced. When $\Lambda = 1$, the ARL is estimated as 2.79. This means that on average an adjustment (rod tip replacement) will be required every 2.79 monitorings. With a

Hypothetical Data for Example in Section 2.9

Circled Points are Monitored Weld Depths

Connected Bars are Predicted Deviations

Adjustments Made at Periods 60, 80, 140, 170 and 190

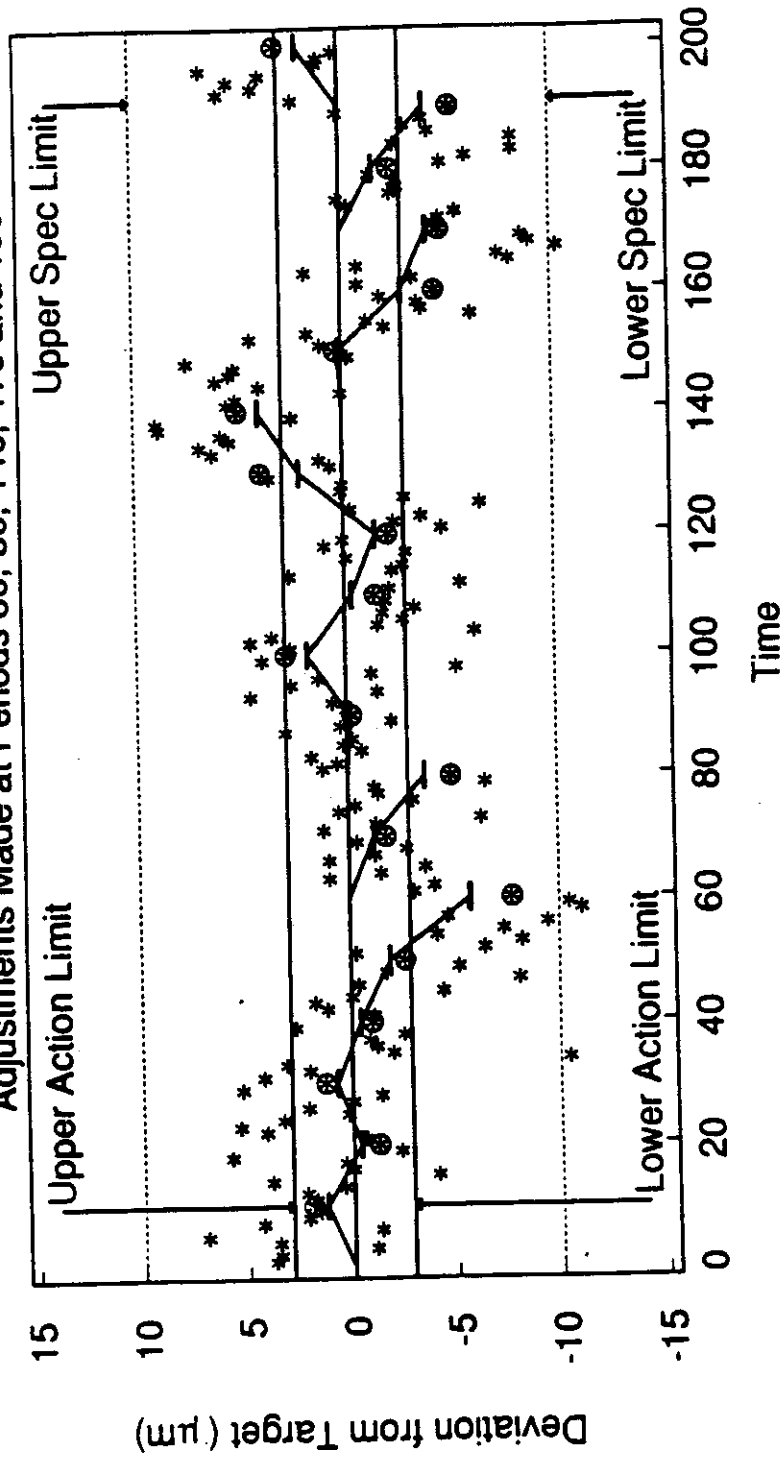


Figure 2.19

monitoring interval of 2½ hours, this leads to the expectation that a rod tip replacement will be required about every 7 hours or so. For the simulated data of figure 2.19, 5 adjustments were required during the 200 fifteen minute periods, or one adjustment every 10 hours. Hence, the simulated data showed slightly fewer adjustments than expected.

Figure 2.19 also shows that four troughs, though not monitored, would have had weld depths below the lower specification limit of 130µm. Simulation (50,000 periods) shows that this control scheme leads to an expected percent defective rate of 1.5%. This defective rate may be unacceptably high even though it was justified by the input costs and disturbance parameters. If a regulation scheme with such a high defective rate was found in practice, one might question whether the costs were adequately modeled as a lower defective rate is obtained with a lower monitoring cost, a lower adjustment cost, or a higher off-target cost. In this particular example, the off-target costs were approximated by including only the cost of scrapping the troughs. However, screening of troughs is destructive and cannot be used to prevent out-of-spec product from reaching downstream operations. Hence, one might approximate the off-target costs by estimating the total cost of downstream defect costs.

Sensitivity to assumptions

One may be interested in how changes in the monitoring interval and action limits effect the expected cost of process regulation. Equation (2.80) allows the expected per interval cost to be determined for combinations of m and $\lambda = \Lambda\gamma_m^2\sigma_m^2$:

$$L = \frac{C_m}{m} + \frac{C_a}{mh(\Lambda)} + k_T \sigma_m^2 \left[1 + \gamma_m^2 g(\Lambda) \right] - \frac{k_T(m-1)\gamma_1^2 \sigma_1^2}{2} \quad (2.80)$$

Figure 2.20 presents cost contours for combinations of λ and m in the area of the optimum, $\lambda = 2.90$ and $m = 10$. This plot was created with $C_m = 20$, $C_a = 60$, and $k_T = 0.6$. One sees that the expected process regulation cost per 15 minute period is slightly less than \$14.10, or approximately \$0.94 per trough. This cost can be broken down into the per period monitoring cost (\$2.00), adjustment cost (\$2.11) and off-target cost (\$9.97). The contours, being tall, slight angled ovals, show that the action limit is not critical in this example and nearly independent of the chosen monitoring interval. The contours also show that longer monitoring intervals can be chosen with only a slight increase in the per period regulation cost. A monitoring interval of 20 periods (300 minutes) with an action limit of 2.9 is less than 8% more expensive than the least cost combination. A contour plot similar to this one could be created for other problems to evaluate the sensitivity of the regulation cost to the optimum combination of monitoring interval and action limit.

Another sensitivity issue for the process regulation scheme is how critical is proper specification of the disturbance parameters and costs. One method of evaluating this sensitivity is to compare two expected costs, the first associated with a specific combination of monitoring interval and action limit (chosen for a particular combination of costs and disturbance parameters), and the other associated with a different combination of costs and disturbance parameters. The minimal cost scheme for the example has $m = 10$ and $\lambda = 2.9$ when $R_a = 123.5$, $R_m = 41.15$, and $\theta_1 = 0.7$. Figure 2.21 shows contours of the percent increase in expected regulation costs for

Regulation Costs in Dollars Per Period
 for Section 2.9 Example
 As a Function of Monitoring Interval and Action Limit

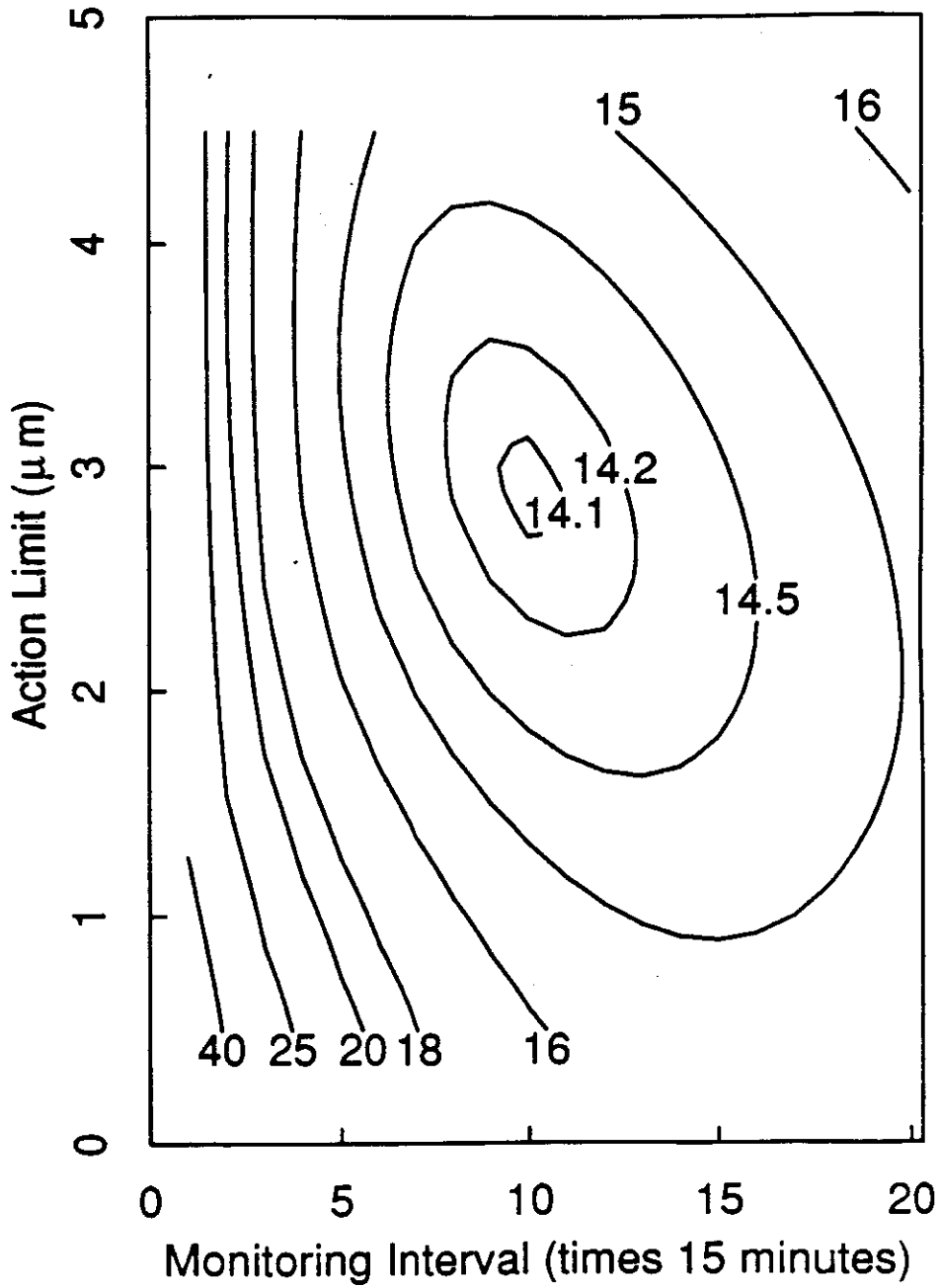


Figure 2.20

using a control scheme with $m = 10$ and $\lambda = 2.9$ instead of the minimal cost scheme. These contours show that mild misspecification of the relative adjustment and monitoring costs or the smoothing constant is not critical to the resulting cost of the control scheme in this example. These contours were created with relative adjustment and monitoring costs varying between four times smaller and four times larger than the "identified" costs and for θ_1 equal to 0.55, 0.70, and 0.85.

Simplification when $\beta_1 / R_a^{1/2}$ is negligible

Three parameters are generally needed to determine the minimal cost monitoring interval and action limit. However, figure 2.16 show that when the ratio $\beta_1 / R_a^{1/2}$ is small, it has only a marginal effect on the regulation scheme. For small values of this ratio, Λ is determined exclusively by $C_m / C_a = R_m / R_a$. The monitoring interval, being dependent on R_a and Λ , is then determined by R_a and R_m . A simplification of the determination of the monitoring interval and action limit is therefore possible when $\beta_1 / R_a^{1/2}$ is negligible. This ratio is small whenever θ_1 is small or R_a is large or both, representing a process that is closer to a random walk than a Shewhart process or an adjustment cost that is large in relation to the rate at which off-target cost increases.

The determination of the scaled action limit when $\beta_1 / R_a^{1/2}$ is negligible is found by manipulating equation (2.86), giving

$$\frac{R_m}{R_a} = \frac{-1}{g(\Lambda)} + \frac{g(\Lambda)}{f(\Lambda)} + \frac{1}{2f(\Lambda)} \quad (2.90)$$

This is readily solved numerically to give Λ as a function of R_m / R_a . Figure 2.22 shows the correspondence between the ratio of monitoring to adjustment cost and the

Percent Increase in Regulation Costs
 when $m = 10$ and $\lambda = 2.9$
 Compared to Minimal Cost Scheme

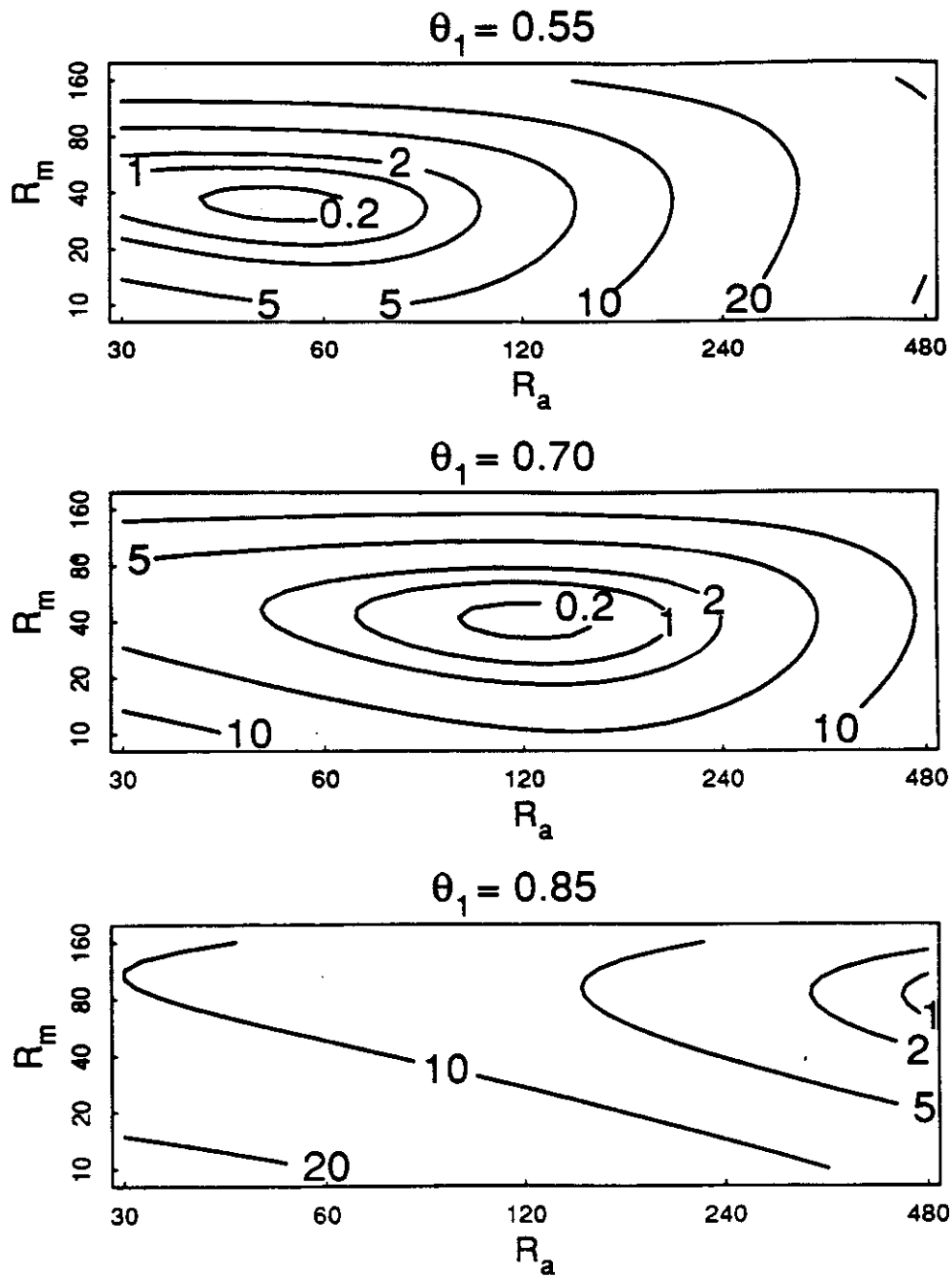


Figure 2.21

scaled action limit, Λ . As this ratio gets smaller, indicating that data collection is increasingly cheaper than adjusting, the scaled action limit gets bigger, making adjustments less frequent.

Figure 2.23 plots contours of the monitoring interval for a range of R_a and R_m . These contours are, for the most part, vertical lines that are only slightly tipped to the left, implying that R_m determines the monitoring interval to a large extent. The computer routine which determined the optimum had numerical difficulties whenever the ratio R_m/R_a was less than 0.0005 or greater than 100. Hence the contours in the upper left and lower right corner are only approximate.

Taken together, figures 2.22 and 2.23 show that the scaled action limit is determined by the ratio R_m/R_a and the monitoring interval largely determined by R_m . For many applications, the relative adjustment cost is large so that these figures allow for a quick determination of the monitoring interval and scaled action limit. When the ratio $\beta_1/R_a^{1/2}$ is not negligible, the outline of figure 2.15 may be used to determine the regulation scheme.

2.10 Taguchi's Solution to the Machine Tool Adjustment Problem

The previous sections have considered the machine tool adjustment problem when there are fixed costs of monitoring and adjustment and the disturbance follows equation (2.2). Taguchi [40,41,42] also considered this problem but from a somewhat different approach than that described in previous sections or the approach by Box and Jenkins [9]. This section describes Taguchi's analysis of the machine tool adjustment problem and compares his results with those we have obtained. In particular, we show

Scaled Action Limit, Λ
 When $\beta_1 / R_a^{1/2}$ is Negligible

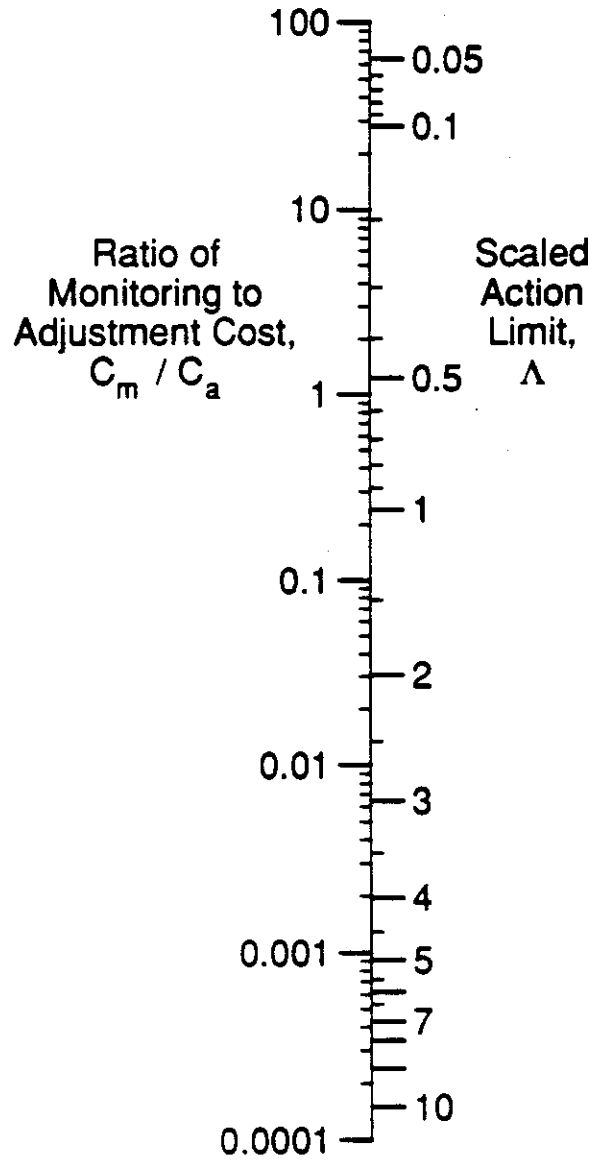


Figure 2.22

Monitoring Interval
When $\beta_1 / R_a^{1/2}$ Is Negligible

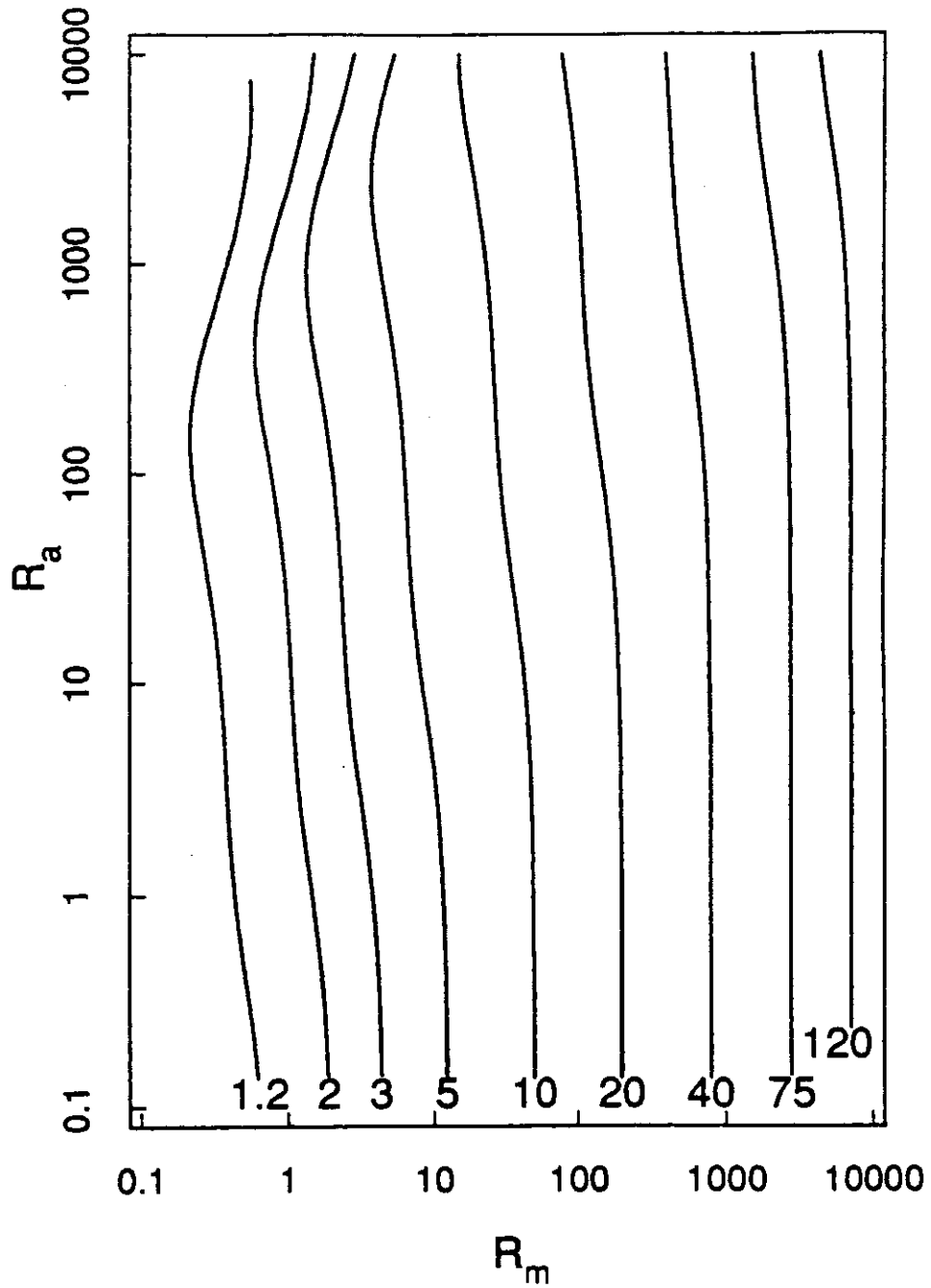


Figure 2.23

that the process regulation costs for many of his schemes are much more expensive than they need to be.

Taguchi's assumptions

Taguchi assumes that there are fixed costs associated with monitoring, C_m , and adjustment, C_a , and an off-target cost that is proportional to the mean squared deviation, with proportionality constant k_T . He also assumes that the disturbance follows a special case of equation (2.2), specifically, that of a random walk in which $\theta_1 = 0$. Hence, the disturbance is modeled by

$$z_{t+1} - z_t = a_t \tag{2.91}$$

where $\{a_t\}$ is a white noise sequence of random shocks with zero mean and variance σ_1^2 . The process is monitored at a multiple, m , of the base interval and adjustments are made when the *observed deviation* exceeds action limits at $\pm\lambda$. The monitoring interval and scaled action limit, $\Lambda = \lambda/\sigma_1$ are chosen to minimize the total process regulation costs:

$$L = \frac{C_a}{ARL} + \frac{C_m}{m} + k_T \cdot MSD \tag{2.92}$$

Taguchi also allows that a delay between monitoring the process and making an adjustment is possible, but that extension is not be considered here. These assumptions then differ from those considered earlier in three important ways, (1) the disturbance is assumed to follow a random walk, (2) the adjustments are made when the *observed deviation* exceeds an action limit, λ , and (3) the action limit is determined as $\lambda = \Lambda\sigma_1$ as opposed to $\lambda = \Lambda\gamma_1\sigma_1$. If the disturbance truly followed a

random walk, Taguchi's assumptions and those we have made would coincide and the *predicted* deviation would equal the *observed* deviation.

Taguchi's approximation to the ARL and MSD

The estimate of the average run length between adjustments is obtained by noting that the random walk is the discrete analog of Brownian motion. The ARL for a Brownian motion process with variance parameter σ_1^2 first passing through a barrier at $\pm\Lambda\sigma_1$ is

$$\text{ARL} = \Lambda^2 \quad (2.93)$$

Taguchi uses the ARL of a Brownian motion process to approximate the ARL of a random walk disturbance. However, this approximation *underestimates* the true ARL, as a Brownian motion process when observed at discrete intervals may pass a barrier at an unobserved time. When the process is next observed, it may or may not be within the action limits. In either case, the observed run length will be greater than the first passage time.

Taguchi employs a crude approximation for the mean squared deviation (MSD), obtained as follows. He treats the deviations from target at monitoring times as if they are uniformly distributed between the lower and upper action limits, $-\Lambda\sigma_1$ and $\Lambda\sigma_1$, respectively. This leads to an estimate of the MSD as $\Lambda^2\sigma_1^2/3$. For deviations between monitorings, Taguchi relates these deviations to the deviation at the beginning of the monitoring interval. Suppose e_{lm} is the deviation at the beginning of a monitoring interval. Then the deviations in this interval may be related using equation (2.91), giving

$$\begin{aligned}
e_{tm+1} &= e_{tm} + a_{tm+1} \\
e_{tm+2} &= e_{tm} + a_{tm+1} + a_{tm+2} \\
&\vdots \\
&\vdots \\
e_{tm+m} &= e_{tm} + a_{tm+1} + \dots + a_{tm+m}
\end{aligned}
\tag{2.94}$$

The expected squared deviation for this monitoring interval is $E(e_{tm}^2) + \sigma_1^2(m+1)/2$.

Thus, on Taguchi's assumption of a uniform distribution, the MSD is

$$\text{MSD} = \frac{\Lambda^2 \sigma_1^2}{3} + (m+1)\sigma_1^2
\tag{2.95}$$

Obtaining the monitoring interval and action limit

Using Taguchi's approximations, the total process regulation cost are estimated as

$$L = \frac{C_m}{m} + \frac{C_a}{\Lambda^2} + k_T \left[\frac{\Lambda^2 \sigma_1^2}{3} + \frac{(m+1)\sigma_1^2}{2} \right]
\tag{2.96}$$

The optimal action limit and monitoring interval are obtained by setting the partial derivatives of equation (2.96) to zero:

$$m = \left[\frac{2C_m}{k_T \sigma_1^2} \right]^{1/2}
\tag{2.97}$$

$$\Lambda = \left[\frac{3C_a}{k_T \sigma_1^2} \right]^{1/4}
\tag{2.98}$$

These expressions have some interesting characteristics. First note that if one approximates the mean squared deviation solely by uniformly distributed deviations, then equation (2.96) leads to an optimal strategy of never monitoring ($m = \infty$) and never adjusting ($\Lambda = \infty$). The implication of these uniformly distributed deviations is

that process regulation cannot reduce the off-target costs and so monitoring and adjustment costs can never be justified. Hence, Taguchi's "correction" for the deviations between monitoring periods, $(m + 1)\sigma_1^2/2$, whether accurate or not, is critical in his determination of a control scheme with a finite monitoring interval.

Also, note that equations (2.97) and (2.98) can be rewritten in terms of the previously defined relative adjustment and monitoring costs:

$$m = (2R_m \gamma_1^2)^{1/2} \quad (2.99)$$

$$\Lambda = (3R_a \gamma_1^2)^{1/4} \quad (2.100)$$

When the disturbance is a random walk, $\gamma_1 = 1$, these equations are solely dependent on R_a and R_m . When γ_1 is near one, the equation for the monitoring interval matches equation (2.69), $m \approx (2R_m)^{1/2}$, which was developed for the situation of a monitoring cost alone. Similarly, equation (2.100) has the same form as the equation developed for an adjustment cost only, equation (2.42), $\Lambda = (6R_a)^{1/4} - 0.593$. These expressions are roughly equivalent when $R_a = 32$ giving a scaled action limit of $\Lambda \approx 3.13$.

Therefore, Taguchi's approach leads to control schemes similar to those obtained if (a) the adjustment and monitoring costs could be considered separately, (b) θ_1 could always be assumed to be near zero, and (c) the scaled action limit, Λ , could be assumed close to 3. However, one cannot always treat the costs of adjustment and monitoring separately, θ_1 is not always near zero, and Λ is not always near 3. We show below that the use of these sweeping assumptions can lead to control schemes with process regulation costs much greater than the minimum cost control scheme.

A comparison between Taguchi's scheme and the minimal cost scheme

When comparing the minimal cost scheme and Taguchi's scheme, one must be aware that the two schemes work somewhat differently when θ_1 is not zero. Taguchi's scheme adjusts on observed deviations instead of predicted deviations and it is unclear how large an adjustment is required with Taguchi's theory. However, approximate comparisons are possible by treating Taguchi's action limit as if it applied to predicted deviations instead of observed deviations, and assuming that adjustments are made to bring the predicted deviation back to zero. This method of evaluating Taguchi's cost leads to underestimating the adjustment costs and overestimating the off-target costs. The discrepancy disappears as θ_1 approaches zero.

Table 2.6, comparing Taguchi's control scheme to the minimal cost control scheme, was created for a range of R_a , R_m , and θ_1 . The monitoring interval and action limit for Taguchi's scheme and the minimal cost scheme are listed as well as the percent increase in expected process regulation costs when using Taguchi's scheme as opposed to the minimal cost scheme. The minimal cost scheme was generally determined by using equations (2.85) and (2.89). When the minimal cost scheme called for a monitoring interval shorter than the base interval, m was set to one and the optimal action limit was chosen using the fixed adjustment cost only equation, (2.32). When the action limit was very small (less than $0.02\gamma_1\sigma_1$), it was set to zero and the monitoring interval determined by using the fixed monitoring cost only equation, (2.68).

**Comparison of Control Schemes:
Taguchi's scheme versus minimal cost scheme**

| $\theta_1 = 0.9$ | | | | | | |
|------------------|-------|---------|--------------------|--------------|--------------------|---------------|
| R_a | R_m | Taguchi | | Minimal Cost | | % Increase |
| | | m | λ/σ_1 | m | λ/σ_1 | In Reg. Costs |
| 1 | 1 | 1.00 | 0.416 | 1.00 | 0.097 | 2.9 |
| 1 | 10 | 1.00 | 0.416 | 1.69 | 0.072 | 4.0 |
| 1 | 100 | 1.41 | 0.416 | 7.52 | 0.028 | 35.3 |
| 1 | 1000 | 4.47 | 0.416 | 32.63 | 0.000 | 90.1 |
| 1 | 10000 | 14.14 | 0.416 | 127.32 | 0.000 | 185.3 |
| 10 | 1 | 1.00 | 0.740 | 1.00 | 0.219 | 8.0 |
| 10 | 10 | 1.00 | 0.740 | 1.71 | 0.198 | 8.7 |
| 10 | 100 | 1.41 | 0.740 | 7.78 | 0.103 | 39.7 |
| 10 | 1000 | 4.47 | 0.740 | 32.81 | 0.034 | 94.2 |
| 10 | 10000 | 14.14 | 0.740 | 127.32 | 0.000 | 188.0 |
| 100 | 1 | 1.00 | 1.316 | 1.00 | 0.436 | 21.7 |
| 100 | 10 | 1.00 | 1.316 | 1.74 | 0.412 | 21.6 |
| 100 | 100 | 1.41 | 1.316 | 8.20 | 0.317 | 50.2 |
| 100 | 1000 | 4.47 | 1.316 | 34.07 | 0.153 | 104.2 |
| 100 | 10000 | 14.14 | 1.316 | 128.04 | 0.046 | 195.6 |
| 1000 | 1 | 1.00 | 2.340 | 1.00 | 0.821 | 56.1 |
| 1000 | 10 | 1.00 | 2.340 | 1.73 | 0.800 | 54.0 |
| 1000 | 100 | 1.41 | 2.340 | 8.43 | 0.701 | 75.8 |
| 1000 | 1000 | 4.47 | 2.340 | 38.15 | 0.483 | 123.7 |
| 1000 | 10000 | 14.14 | 2.340 | 132.74 | 0.242 | 213.3 |
| 10000 | 1 | 1.00 | 4.162 | 1.00 | 1.506 | 121.2 |
| 10000 | 10 | 1.00 | 4.162 | 1.67 | 1.495 | 117.0 |
| 10000 | 100 | 1.41 | 4.162 | 8.72 | 1.379 | 126.8 |
| 10000 | 1000 | 4.47 | 4.162 | 40.46 | 1.177 | 158.2 |
| 10000 | 10000 | 14.14 | 4.162 | 153.69 | 0.760 | 237.8 |

Table 2.6

**Comparison of Control Schemes:
Taguchi's scheme versus minimal cost scheme**

| $\theta_1 = 0.45$ | | | | | | |
|-------------------|-------|---------|--------------------|--------------|--------------------|---------------|
| R_a | R_m | Taguchi | | Minimal Cost | | % Increase |
| | | m | λ/σ_1 | m | λ/σ_1 | In Reg. Costs |
| 1 | 1 | 1.00 | 0.976 | 1.42 | 0.437 | 7.7 |
| 1 | 10 | 2.46 | 0.976 | 4.45 | 0.229 | 16.2 |
| 1 | 100 | 7.78 | 0.976 | 14.09 | 0.069 | 17.4 |
| 1 | 1000 | 24.60 | 0.976 | 44.68 | 0.000 | 17.8 |
| 1 | 10000 | 77.78 | 0.976 | 141.41 | 0.000 | 18.1 |
| 10 | 1 | 1.00 | 1.736 | 1.55 | 1.114 | 9.3 |
| 10 | 10 | 2.46 | 1.736 | 5.16 | 0.718 | 21.3 |
| 10 | 100 | 7.78 | 1.736 | 14.59 | 0.398 | 24.0 |
| 10 | 1000 | 24.60 | 1.736 | 44.90 | 0.122 | 20.5 |
| 10 | 10000 | 77.78 | 1.736 | 141.41 | 0.000 | 18.8 |
| 100 | 1 | 1.00 | 3.086 | 1.76 | 2.263 | 8.8 |
| 100 | 10 | 2.46 | 3.086 | 5.85 | 1.909 | 19.0 |
| 100 | 100 | 7.78 | 3.086 | 16.94 | 1.251 | 29.7 |
| 100 | 1000 | 24.60 | 3.086 | 46.44 | 0.706 | 27.9 |
| 100 | 10000 | 77.78 | 3.086 | 141.99 | 0.217 | 21.6 |
| 1000 | 1 | 1.00 | 5.489 | 1.68 | 4.408 | 6.0 |
| 1000 | 10 | 2.46 | 5.489 | 6.71 | 3.951 | 13.0 |
| 1000 | 100 | 7.78 | 5.489 | 19.34 | 3.360 | 23.6 |
| 1000 | 1000 | 24.60 | 5.489 | 53.88 | 2.217 | 33.6 |
| 1000 | 10000 | 77.78 | 5.489 | 146.91 | 1.255 | 29.4 |
| 10000 | 1 | 1.00 | 9.760 | 1.32 | 8.269 | 4.7 |
| 10000 | 10 | 2.46 | 9.760 | 7.38 | 7.675 | 7.1 |
| 10000 | 100 | 7.78 | 9.760 | 22.14 | 6.996 | 14.6 |
| 10000 | 1000 | 24.60 | 9.760 | 61.57 | 5.964 | 25.3 |
| 10000 | 10000 | 77.78 | 9.760 | 170.53 | 3.941 | 34.9 |

Table 2.6

**Comparison of Control Schemes:
Taguchi's scheme versus minimal cost scheme**

| $\theta_1 = 0.0$ | | | | | | |
|------------------|-------|---------------------------------------|--------|--|--------|-----------------------------|
| R_d | R_m | Taguchi $m \quad \lambda/\sigma_1$ | | Minimal Cost $m \quad \lambda/\sigma_1$ | | % Increase In Reg. Costs |
| 1 | 1 | 1.41 | 1.316 | 1.71 | 0.717 | 5.6 |
| 1 | 10 | 4.47 | 1.316 | 4.65 | 0.406 | 4.2 |
| 1 | 100 | 14.14 | 1.316 | 14.22 | 0.124 | 1.2 |
| 1 | 1000 | 44.72 | 1.316 | 44.72 | 0.000 | 0.3 |
| 1 | 10000 | 141.42 | 1.316 | 141.42 | 0.000 | 0.1 |
| 10 | 1 | 1.41 | 2.340 | 1.95 | 1.928 | 2.6 |
| 10 | 10 | 4.47 | 2.340 | 5.39 | 1.274 | 6.5 |
| 10 | 100 | 14.14 | 2.340 | 14.70 | 0.721 | 4.5 |
| 10 | 1000 | 44.72 | 2.340 | 44.97 | 0.221 | 1.3 |
| 10 | 10000 | 141.42 | 2.340 | 141.42 | 0.000 | 0.3 |
| 100 | 1 | 1.41 | 4.162 | 2.23 | 4.019 | 1.2 |
| 100 | 10 | 4.47 | 4.162 | 6.16 | 3.428 | 2.8 |
| 100 | 100 | 14.14 | 4.162 | 17.05 | 2.266 | 6.9 |
| 100 | 1000 | 44.72 | 4.162 | 46.48 | 1.283 | 4.6 |
| 100 | 10000 | 141.42 | 4.162 | 142.20 | 0.394 | 1.3 |
| 1000 | 1 | 1.41 | 7.401 | 2.54 | 7.802 | 1.6 |
| 1000 | 10 | 4.47 | 7.401 | 7.05 | 7.147 | 1.2 |
| 1000 | 100 | 14.14 | 7.401 | 19.49 | 6.097 | 2.9 |
| 1000 | 1000 | 44.72 | 7.401 | 53.93 | 4.030 | 7.0 |
| 1000 | 10000 | 141.42 | 7.401 | 146.98 | 2.281 | 4.6 |
| 10000 | 1 | 1.41 | 13.161 | 1.60 | 14.965 | 3.1 |
| 10000 | 10 | 4.47 | 13.161 | 8.02 | 13.874 | 1.6 |
| 10000 | 100 | 14.14 | 13.161 | 22.30 | 12.710 | 1.2 |
| 10000 | 1000 | 44.72 | 13.161 | 61.63 | 10.842 | 2.9 |
| 10000 | 10000 | 141.42 | 13.161 | 170.53 | 7.166 | 7.0 |

Table 2.6

Table 2.6 shows that Taguchi's control scheme does quite well in comparison with the minimal cost scheme when $\theta_1 = 0.0$. For this situation, Taguchi's scheme is generally less than 10% more expensive than the minimal cost scheme with the biggest increase arising when R_a and R_m are of the same magnitude. However, when $\theta_1 = 0.45$, the increases in regulation costs are on the order of 10 to 30 percent. Again, the largest increases are typically observed when R_a and R_m are of similar size. When $\theta_1 = 0.9$, Taguchi's scheme does especially poor with increases in regulation costs of 100 percent or more being common for large R_a or large R_m . Hence, Taguchi's scheme does reasonably well if the disturbance follows a random walk but does poorly when θ_1 is near one.

We should remember that Taguchi's assumption of a universal random walk is in many cases an unreal one. Specifically, it implies that the forecast of where the process will be in the next interval depends *only* on where it is in the present interval. Our own experience is that many processes met in practice have θ greater than 0.5 and sometimes as high as 0.9. This seems reasonable in view of the fact that θ is the exponential smoothing constant which determines how rapidly past data is being discounted. When $\theta = 0.9$, 21 previous deviations have at least 10% of the weight of the most recent deviation in the forecast. When $\theta = 0.7$, 6 previous deviations contribute similarly to the forecast, and when $\theta = 0.5$, 3 previous deviations have a similar contribution. Only in the limit does the forecast depend solely on the most recent deviation, as Taguchi assumes.

Taguchi uses equation (2.96) as both an estimate of the regulation costs and as a function which justifies the determination of the monitoring interval and action limit. In practice, though, Taguchi recommends that the monitoring interval and action limit be derived by modifying some existing limits. Specifically, let λ_1 be the initial action limit and u_1 be the average run length of the observed deviation exceeding the action limit. Taguchi estimates the *mean square drift per unit product*, $\hat{\sigma}_1^2$, by relating the action limit and average run length.

$$\hat{\sigma}_1^2 = \frac{\lambda_1^2}{u_1} \quad (2.101)$$

Then a new monitoring interval and action limit is obtained by substituting $\hat{\sigma}_1^2$ for σ_1^2 :

$$m = \left[\frac{2C_m}{k_T \hat{\sigma}_1^2} \right]^{1/2} \quad (2.102)$$

$$\lambda = \left[\frac{3C_a}{k_T \hat{\sigma}_1^2} \right]^{1/4} \hat{\sigma}_1 \quad (2.103)$$

Using this method to obtain the monitoring interval and action limit may be generally superior to using equations (2.99) and (2.100) as λ_1^2/u_1 is, to a crude approximation, an estimate of $\gamma_1^2 \sigma_1^2$. This leads to $m \approx (2R_m)^{1/2}$ and $\lambda \approx (3R_a)^{1/4} \gamma_1 \sigma_1$, which are similar to the equations for calculating the minimal cost control scheme. One benefit of the above strategy is that nothing needs to be known about the underlying disturbance model except the run length obtained for a chosen set of action limits. However, one drawback is that the strategy of using an action limit and its corresponding run length to generate a new action limit can lead to a divergent sequence of action limits. The potential for a divergent sequence of limits was shown

independently by Kramer [27] and Adams and Woodall [2].

Adams and Woodall analyzed the accuracy of Taguchi's approximations and the practical effect of using these approximations to determine a monitoring interval and action limit for random walk processes. They showed that, for random walks, the approximations can be grossly inaccurate whenever the monitoring interval is large and the action limit is small, a situation which results from a large monitoring cost and a small adjustment cost. This scenario is common in the process industries where adjustments may involve turning a control knob but monitoring may require an expensive analysis. Woodall and Adams did not extend their results to the general disturbance model given by equation (2.2) but table 2.6 shows that Taguchi's approximations are worse for the general disturbance model than the special case of a random walk.

In summary, Taguchi's approach is simple but relatively inefficient and costly to implement relative to the minimal cost control scheme. His concept of modifying existing action limits to determine a lower cost process regulation scheme has benefits but, in its present form, can lead to divergent limits. Further, the implicit restriction to a random walk disturbance has severe shortcomings in practice.

2.11 Summary

This chapter considered process regulation from the viewpoint of minimizing the total costs associated with being off-target, making adjustments and monitoring a process. The purpose of regulation is to negate the effects of a disturbance, flexibly modeled here by an equation that can represent a random walk and a stationary,

Shewhart process. Adjustments in this chapter were assumed to have no associated dynamics or delay. Minimal cost control schemes were developed for schemes with: significant adjustment cost but negligible monitoring cost, significant monitoring cost but negligible adjustment cost, and, finally, for significant monitoring and significant adjustment costs. It was shown that when there is a significant adjustment cost, adjustments should be made when predicted deviations from target exceed an action limit. The magnitude of the adjustment is such that it negates the predicted deviation. When there is a monitoring cost an optimal monitoring interval was determined as a multiple of a base interval at which values of θ_1 and σ_1 had been obtained. The determination of a monitoring interval and action limit depend on three parameters, θ_1 , a disturbance parameter, R_a , a relative adjustment cost that measures the cost of adjustment to non-adjustment, and R_m , a relative monitoring cost that measures the cost of monitoring to non-adjustment.

Graphical aids and outlines were provided to assist in the implementation of a process control scheme.

Bibliography

1. Abraham, B. and G. E. P. Box. 1979. Sampling interval and feedback control. *Technometrics*, 21 (no. 1): 1-8.
2. Adams, B. M. and W. H. Woodall. An analysis of Taguchi's on-line process control procedure under a random walk model. *Technometrics*, to appear.
3. Alwan, L. and H. V. Roberts. 1986. Time series modeling for statistical process control. *Proceedings of the Bus. and Econ. Stat. Section, Amer. Stat. Assoc.*, 315-320.
4. Åström, K. J and B. Wittenmark. 1984. *Computer Controlled Systems: Theory and Design*. Prentice-Hall, New Jersey.
5. Åström, K. J. 1970. *Introduction to Stochastic Control*. Mathematics in Science and Engineering Series, 70. Academic Press.
6. Barnard, G. A. 1959. Control charts and stochastic processes. *J. Royal Stat. Soc., Series B*, 21 (no. 2): 239-271.
7. Box, G. E. P. and N. R. Draper. 1987. *Empirical Model-Building and Response Surfaces*, Wiley.
8. Box, G. E. P. 1980. Sampling and Bayes' inference in scientific modelling and robustness. *J. Royal Stat. Soc., Series A*, 143 (part 4): 383-430.
9. Box, G. E. P. and G. M. Jenkins. 1963. Further contributions to adaptive optimization and control: simultaneous estimation of dynamics: non-zero costs. *Bulletin of the International Statistical Institute*, 34th Session, Ottawa.
10. Box, G. E. P. and G. M. Jenkins. 1965. Mathematical models for adaptive control and optimisation. *A. I. Chem. E. - I. Chem. E. Symposium Series*, 4: 61-68.
11. Box, G. E. P. and G. M. Jenkins. 1968. Some recent advances in forecasting and control. *J. Royal Stat. Soc., Series C (Applied Statistics)*, 17 (no. 2): 91-109.
12. Box, G. E. P. and G. M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.
13. Box, G. E. P., G. M. Jenkins and J. F. MacGregor. 1974. Some recent advances in forecasting and control, part II. *J. Royal Stat. Soc., Series C (Applied Statistics)*, 23 (no. 2): 158-179.
14. Box, G. E. P. and J. G. Ramírez. 1989. Personal communication.

15. Clarke, D. W. and P. J. Gawthorpe. 1975. Self-tuning controller. *Proceedings of the Institution of Electrical Engineers*, 122 (no. 9): 929-934.
16. Clarke, D. W. and R. Hastings-James. 1971. Design of digital controllers for randomly disturbed systems. *Proceedings of the Institution of Electrical Engineers* 118, (no. 10): 1503-6.
17. Crowder, S. V. 1986. *Kalman Filtering and Statistical Process Control*. Ph.D. diss., Iowa State University.
18. Crowder, S. V. 1987. A simple method for studying run-length distributions of exponentially weighted moving average charts. *Technometrics*, 29 (no.4): 401-408.
19. Crowder, S. V. 1987. Average run lengths of exponentially weighted moving average charts. *Journal of Quality Technology*, 19 (no. 3): 161-164.
20. Deming, W. E. 1986. *Out of the Crisis*, MIT, Center for Advanced Engineering Study, Cambridge, Mass.
21. Ewan, W. D. and K. W. Kemp. 1960. Sampling inspection of continuous processes with no autocorrelation between successive results. *Biometrika*, 47 (3 and 4): 363.
22. Hunter, J. S. 1986. The exponentially weighted moving average. *Journal of Quality Technology*, 18 (no. 4): 203-210.
23. Imai, M. 1986. *Kaizen: The Key to Japan's Competitive Success*. Random House.
24. Jowett, G. H. 1955. The comparison of means of sets of observations from sections of independent stochastic series. *J. Royal Stat. Soc., Series B*, 17 (no. 2): 208-227.
25. Jowett, G. H. 1957. Statistical analysis using local properties of smoothly heteromorphic stochastic series. *Biometrika*, 44: 454-463
26. Kartha, P. 1973. Some Statistical and Economic Problems in Quality Control. Ph.D. diss., University of Wisconsin, Madison.
27. Kramer, T. 1986. An extension of the Box-Jenkins approach to process regulation. Speech given at the 3rd Quality and Productivity Research Conference in Oakland, Michigan.
28. Lorenzen, T. J. and L. C. Vance. 1986. The economic design of control chart: a unified approach. *Technometrics*, 28 (no. 1): 3-10.
29. Lucas, J. M. 1976. The design and use of V-mask control schemes. *Journal of Quality Technology*, 8 (no. 1): 1-12.

30. Lucas, J. M. and R. B. Crosier. 1982. Fast initial response for cusum quality-control schemes: give your cusum a head start. *Technometrics*, 24 (no. 3): 199-206.
31. MacGregor, J. F. 1972. Topics in the Control of Linear Processes Subject to Stochastic Disturbances. Ph.D. diss., University of Wisconsin, Madison.
32. MacGregor, J. F. 1976. Optimal choice of the sampling interval for discrete process control. *Technometrics*, 18 (no. 2): 151-160.
33. MacGregor, J. F. 1987. Interfaces between process control and online statistical process control. *Computing and Systems Technology Division Communications*, 10 (no. 2): 9-20.
34. MacGregor, J. F., T. J. Harris and J. D. Wright. 1984. Duality between control of processes subject to randomly occurring deterministic disturbances and ARIMA stochastic disturbances. *Technometrics*, 26 (no. 2): 389-397.
35. MacGregor, J. F. and P. W. Tidwell. 1977. Correspondence: Discrete stochastic control with input constraints. *Proceedings of the Institution of Electrical Engineers*, 124 (no. 8): 732-734.
36. Mayr, O. 1970. *The Origins of Feedback Control*. M.I.T. Press, Cambridge, Massachusetts.
37. Roberts, S. W. 1959. Control chart test based on geometric moving averages. *Technometrics*, 1 (no. 3): 239-50.
38. Roberts, S. W. 1966. A comparison of some control chart procedures. *Technometrics*, 8 (no. 3): 411-430.
39. Shewhart, W. A. 1931. *Economic Control of Quality of Manufactured Product*. D. Van Nostrand Company, Inc.
40. Taguchi, G. 1986. *Introduction to Quality Engineering: Designing Quality Into Products and Processes*. Asian Productivity Organization.
41. Taguchi, G. 1984. Specification value and quality control, part 6. *The International QC Forum*.
42. Taguchi, G. 1984. Specification value and quality control, part 7. *The International QC Forum*.