Setup Adjustment Under Unknown Process Parameters and Fixed Adjustment Cost

Zilong Lian and Enrique Del Castillo *

Dept. of Industrial & Manufacturing Engineering
The Pennsylvania State University
University Park, PA 16802

Abstract

Consider a machine that can start production off-target where the initial offset is unknown and unobservable. The goal is to determine the optimal series of machine adjustments that minimize the expected value of the sum of quadratic off-target costs and fixed adjustment costs. Apart of the unknown initial offset, the process is supposed to be in a state of statistical control, so the process model is applicable to discrete-part production processes. The process variance is also assumed unknown. We show, using a dynamic programming formulation based on the Bayesian estimation of all unknown process parameters, how the optimal process adjustment policy is of a deadband form where the width of the deadband is time-varying and U-shaped. Computational results and implementation details are presented. The simpler case of a known process variance is also solved using a dynamic programming approach. It is shown that the solution to this case is a good approximation to the first case, when the variance is actually unknown. The unknown process variance solution, however is the most robust with respect to variation in the process parameters.

Key words: Statistical Process Control, Deadband Adjustment, Feedback Control

1 Introduction

Adjusting a discrete-part manufacturing process is frequently necessary when the setup of the machine is improperly done. An incorrect setup can result in an offset in the quality characteristic (usually, some dimension) of the parts

* Corresponding author. e-mail: exd13@psu.edu

produced in the batch of product made subsequently to the setup. The offset cannot be observed directly, and sequential adjustments must take place after fabrication and measurement of a series of parts.

In the mechanical engineering literature, this adjustment problem has received attention under the name "process positioning" (e.g., see Bjorke, 1989). In the statistical quality control literature, this problem, called the setup adjustment problem, was originally studied by F.E. Grubbs (1954, in a paper reprinted in 1983). Grubbs considers only off-target quadratic costs and studies two cases depending on whether the offset is an unknown constant or a random variable with known distribution. In the first case, he shows the optimal adjustment policy to have the form $U_i = Y_i/i$, where U_i is the adjustment made after part i is processed and Y_i is the observed deviation from target of the quality characteristic. The sequence of adjustments then follow a harmonic rule, so Trietsch (1998) refers to this adjustment policy as the "harmonic rule". A review and unification of these and other rules for setup adjustment have been given recently by Del Castillo et al. (2003a, 2003b). These authors show how the previous setup adjustment rules are all cases of Linear Quadratic Gaussian control.

In the literature on setup adjustment, process parameters are all assumed known. These includes the offset moments (if the setup is assumed random) and the part-to-part variance. Furthermore, only off-target costs have been considered before, but no reference to the important case of fixed adjustment costs has been made. In this paper, a solution to the setup adjustment problem is given for the case all process parameters are unknown and there are fixed adjustment and quadratic off-target costs.

Process control problems under fixed adjustment costs have been studied by Box and Jenkins (1963), Crowder (1992), and Box and Luceño (1997), among others. Solution to these problems is of the form of a *deadband*, where the process is adjusted only if it is predicted to be far enough from target, with the deadband width denoting the action limits that depend on the costs involved. These authors consider a process that would drift off-target if let uncontrolled. They all consider known process parameters. In contrast, the present paper considers a simpler stochastic process, adequate to model discrete-part processes, but assumes no a priori knowledge of the process parameters. The justification for adjustments is not a drifting process, but rather an incorrect setup operation that results in an offset. Such production processes are called *setup dominant* by Gryna (2001). Some examples include cutting processes, drilling, and presswork.

The paper is organized as follows. We next present the process assumptions and the bayesian approach utilized for parameter inferences. Section 3 then shows the dynamic programming (DP) formulation for the case all parameters

need to be estimated. Section 4 gives details about the computer implementation of the DP solution. Then, in Section 5, the sensitivity of the optimal adjustment solution to changes in the true process parameters and on the prior estimates utilized is studied. The simpler case when the part to part variance is known is solved and analyzed in Section 6. This is a case that results in a simpler DP solution due to the smaller state-space. Finally, Section 7 compares the two cases, for known and unknown part-to-part variance.

2 Process Model, Cost Criterion and Parameter Estimates

Suppose N parts are processed sequentially in time and suppose the machine setup results in an unknown initial offset θ_0, which, if unadjusted, will affect all N parts. Let us denote by Y_i the deviation from target of the quality characteristic of interest for the i^{th} part. We assume a controllable factor u_i exists to adjust the process in a linear manner, and denote by $U_i = u_{i+1} - u_i$ the adjustment made before processing part $i+1$. The adjustment U_i is then computed based on observations obtained up to and including part i. Then, the equations describing the observed quality characteristic can be expressed as follows:

$$Y_i = \theta_i + \upsilon_i \tag{1}$$
$$\theta_i = \theta_{i-1} + U_{i-1},\ i = 1, 2, ..., N \tag{2}$$

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

$$L = E\left\{ \sum_{i=1}^{N} (Y_i^2 + c\delta(U_{i-1})) \right\}, \tag{3}$$

where $\delta(x) = 1$ if $x \neq 0$ and $\delta(x) = 0$ otherwise. Following Crowder (1991), the cost c can be understood as the ratio of the adjustment cost to the per unit quadratic off-target cost. The objective is to find the optimal adjustments $\{U_i\}_{i=1}^N$ based on the observed deviations from target in order to minimize the cost criterion in (3).

We use a Bayesian model to make inferences about the two unknown parameters, θ_0 and σ_υ, and be able to adjust the process based on these inferences. For this purpose, a useful model, discussed by Gelman et al. (2003), is the two-parameter normal conjugate model with prior distributions given by:

$$\theta_0|\sigma_\upsilon \sim N(\mu_0, \sigma_\upsilon^2/\kappa_0) \tag{4}$$
$$\sigma_\upsilon^2 \sim Inv-\chi^2(\nu_0, \sigma_0^2), \tag{5}$$

where the second distribution is a scaled inverse chi-square distribution with degrees of freedom v_0 and scale parameter σ_0. The joint model is frequently denoted $(\theta_0, \sigma_\upsilon^2) \sim N-Inv-\chi^2(\mu_0, \sigma_0^2/\kappa_0; \nu_0, \sigma_0^2)$. Note that the parameters θ_0 and σ_υ^2 are assumed dependent in their joint prior distribution such that a large value of σ_υ^2 induces a high-variance in the prior distribution of the initial mean θ_0. As discussed by Gelman et al. (2003), this is useful since prior beliefs about θ_0 can be calibrated by the scale of measurement of Y so that they are equivalent to κ_0 prior measurements on this scale. This facilitates the assessment and incorporation of prior information, if any. Alternatively, these conjugate priors can also be set to be noninformative or "objective".

The Bayesian model on the unknown parameters allows to make predictions on the quality characteristic Y, which is directly observable in contrast to the unknowns θ_0 and σ_υ. Our interest is to make inferences in future values of Y at time i, for which we need to obtain an expression for the posterior predictive density of Y. For this purpose, we use the conjugate prior distributions in (4-5) and equations (1-2) to get the posterior densities:

$$Y_i|\theta_i, \sigma_\upsilon^2 \sim N(\theta_i, \sigma_\upsilon^2) \tag{6}$$

or alternatively

$$Y_{i+1}|\theta_i, \sigma_\upsilon^2 \sim N(\theta_i + U_i, \sigma_\upsilon^2), \tag{7}$$

where

$$(\theta_i, \sigma_\upsilon^2)|Y^{(i)}, U^{(i)} \sim N-Inv-\chi^2(\mu_i, \sigma_i^2/\kappa_i; \nu_i, \sigma_i^2). \tag{8}$$

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

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

$$\int_0^\infty \int_{-\infty}^\infty f(Y_{i+1}|\ \theta_i, \sigma_\upsilon^2) f(\theta_i, \sigma_\upsilon^2 \mid Y^{(i)}, U^{(i)})\ d\theta_i\ d\sigma_\upsilon^2.$$

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

$$f(Y_i|\ Y^{(i)}, U^{(i)}) = \Psi(\cdot|\ \mu_i, \kappa_i, \nu_i, \sigma_i^2) \tag{9}$$

where the mean of the distribution is adjusted by U_i. It is not difficult to show (see Appendix 1) that the posterior density Ψ is a Student t distribution with v_i degrees of freedom, mean equal to $\mu_i + U_i$, and scale parameter equal to $\sigma_i^2(k_i+1)/k_i$. This is a valuable result we use in the next section to determine an optimal setup adjustment policy.

A nice feature of the two-parameter normal conjugate model is that the four state variables $(\mu_i, \kappa_i, \nu_i, \sigma_i^2)$ are easily updated recursively. From expressions in Gelman et al. (2003) in conjunction with the posterior density (8) written at time $i-1$ (used as the prior density at time i), we obtain the recursive updating expressions:

$$\mu_i = \frac{\kappa_{i-1}}{\kappa_{i-1}+1}(\mu_{i-1} + U_{i-1}) + \frac{1}{\kappa_{i-1}+1} y_i \tag{10}$$

$$\kappa_i = \kappa_{i-1} + 1 \tag{11}$$

$$\nu_i = \nu_{i-1} + 1 \tag{12}$$

$$\nu_i \sigma_i^2 = \nu_{i-1}\sigma_{i-1}^2 + \frac{\kappa_{i-1}}{\kappa_{i-1}+1}(y_i - \mu_{i-1} - U_{i-1})^2, \tag{13}$$

Where y_i is the observed value of Y_i at the end of period i. In these equations it is assumed an adjustment of magnitude U_i modifies the updated mean after each observation y_i is obtained. The expressions will be used in Section 3 to build a Dynamic Programming formulation to solve the setup adjustment problem.

3 Dynamic programming solution to the setup adjustment problem

Suppose at stage i, i.e. after part i is observed and before part $i+1$ is processed, the posterior predictive density in (9) is obtained by updating four state variables κ_i, ν_i, μ_i and σ_i^2. Assuming κ_0 and ν_0 are fixed when the prior information is given, the κ_i's and ν_i's are all constants determined by the stage index i, that is

$$\kappa_i = \kappa_0 + i, \ \nu_i = \nu_0 + i, \ \text{for } \forall\, i. \tag{14}$$

Therefore, only two state variables, μ_i and σ_i^2, are needed to describe the predictive density.

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

$$\Psi(x|\mu,\kappa,\nu,\sigma^2) = \Psi(-x|\mu,\kappa,\nu,\sigma^2)$$
$$\Psi(x_1|\mu,\kappa,\nu,\sigma^2) > \Psi(x_2|\mu,\kappa,\nu,\sigma^2), \text{ if } |x_1-\mu| < |x_2-\mu|. \tag{15}$$

In addition, its expectation and variance are given by

$$E\{X\} = \mu$$
$$Var\{X\} = \sigma^2 \frac{\nu}{\nu-2}(1+\frac{1}{\kappa}) \doteq V(\kappa,\nu,\sigma^2), \text{ for } \nu > 2. \tag{16}$$

Since the state variables κ and ν are uniquely defined by stage index i, we introduce the simpler notation

$$V(\kappa_i,\nu_i,\sigma^2) = V_i(\sigma^2)$$
$$\text{and } \Psi(\cdot|\mu,\kappa_i,\nu_i,\sigma^2) = \Psi_i(\cdot|\mu,\sigma^2). \tag{17}$$

Define $R_i(\mu_i,\sigma_i^2)$ to be the minimum cost from parts $(i+1)$ to N given the current density $\Psi_i(\cdot|\mu_i,\sigma_i^2)$. Then $R_0(\mu_0,\sigma_0^2)$ is the expected cost of the optimal solution we seek to the problem.

At the boundary, i.e., at stage $N-1$, we have that $Y_N|y_{N-1},u_{N-1} \sim \Psi_{N-1}(\cdot|\mu_{N-1}+U_{N-1},\sigma_{N-1}^2)$, so

$$\begin{aligned} R_{N-1}(\mu_{N-1},\sigma_{N-1}^2) &= \min_{U_{N-1}} E\{Y_N^2 + c\delta(U_{N-1})\} \\ &= \min_{U_{N-1}} \{V(\kappa_{N-1},\nu_{N-1},\sigma_{N-1}^2) + (\mu_{N-1}+U_{N-1})^2 + c\delta(U_{N-1})\} \\ &= V_{N-1}(\sigma_{N-1}^2) + \min\{\mu_{N-1}^2, c\}. \end{aligned} \tag{18}$$

The optimal adjustment which minimizes this last expression is clearly

$$U_{N-1} = \begin{cases} -\mu_{N-1} & \text{if } |\mu_{N-1}| > c^{1/2} \\ 0 & \text{if } |\mu_{N-1}| \le c^{1/2} \end{cases} \tag{19}$$

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

$$R_i(\mu_i,\sigma_i^2) = \min_{U_i}\{V_i(\sigma_i^2) + (\mu_i+U_i)^2 + c\delta(U_i) + E\{R_{i+1}(\mu_{i+1},\sigma_{i+1}^2)|U_i\}\}, \tag{20}$$

where

$$\begin{aligned} E\{R_{i+1}(\mu_{i+1},\sigma_{i+1}^2)|U_i\} &= \int R_{i+1}(\mu_{i+1},\sigma_{i+1}^2) f(Y_{i+1}|\, Y^{(i)},U^{(i)}) dY_{i+1} \\ &= \int R_{i+1}(\mu_{i+1},\sigma_{i+1}^2)\Psi_i(\mu_i+U_i,\sigma_i^2) dY_{i+1}, \end{aligned} \tag{21}$$

and from (10) and (13),

$$\mu_{i+1} = \frac{\kappa_i(\mu_i + U_i) + Y_{i+1}}{\kappa_{i+1}} \tag{22}$$

$$\sigma^2_{i+1} = \frac{\nu_i\sigma_i^2 + \frac{\kappa_i}{\kappa_{i+1}}(Y_{i+1} - \mu_i - U_i)^2}{\nu_{i+1}}. \tag{23}$$

The optimal adjustment rule or controller U_i^* depends on μ_i and σ_i^2, and is the optimal solution to (20). For each possible U_i, the expected cost can be calculated numerically using backwards induction and Monte Carlo integration. U_i^* is chosen to be the U_i yielding minimum cost, and this is a numerically intensive computation. The computational complexity can be reduced by narrowing the candidates $U_i(\mu_i, \sigma_i^2)$ we evaluate.

3.1 Deadband structure of the optimal adjustment policy

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

$$R_i(\mu_i, \sigma_i^2) = \min_{U_i}\{c\delta(U_i) + R_i'(\mu_i + U_i, \sigma_i^2)\} \tag{24}$$

If we can find $\mu_i = \mu^*_{i,\sigma_i}$ that minimizes $R_i'(\mu_i, \sigma_i^2)$ for given σ_i, (24) can be rewritten as:

$$R_i(\mu_i, \sigma_i^2) = \min\{R_i'(\mu_i, \sigma_i^2), c + R_i'(\mu^*_{i,\sigma_i}, \sigma_i^2)\}. \tag{25}$$

The optimal controller is thus

$$U_i^*(\mu_i, \sigma_i^2) = \begin{cases} \mu^*_{i,\sigma_i} - \mu_i & \text{if } R_i'(\mu_i, \sigma_i^2) - R_i'(\mu^*_{i,\sigma_i}, \sigma_i^2) > c \\ 0 & \text{o.w.} \end{cases} \tag{26}$$

Thus the decision is made based on a tradeoff between the adjustment cost and the cost savings gained by adjusting at the current stage. In such manner, we can find the optimal adjustment for each state (μ_i, σ_i) by numerically calculation.

However, the optimal controller (26) is not in a "deadband" form, which would be easier to implement. To get a deadband form for the optimal controller, consider two states at stage i, $(\mu_i = x_1, \sigma_i^2)$ and $(\mu_i = x_2, \sigma_i^2)$, where the posterior densities are the same except for their means. It is easy to show that $R_i'(x_1, \sigma_i^2) = R_i'(x_2, \sigma_i^2)$ when $|x_1| = |x_2|$ due to the symmetries in the process

and in the cost structure. Suppose $|x_1| > |x_2|$, i.e., the first state represents a process with a larger expected offset while both processes have the same offset variance. The expected offset of the first process will either remain or be brought down at some cost, so it has a higher expected cost than the process at the second state. Hence $R_i'(x_1, \sigma_i^2) > R_i'(x_2, \sigma_i^2)$, i.e., $R_i'(\mu_i, \sigma_i^2)$ is symmetric around $\mu_i = 0$ and increasing with $|\mu_i|$. Following (25) and (26) we have

$$R_i(\mu_i, \sigma_i^2) = \min\{R_i'(\mu_i, \sigma_i^2), c + R_i'(0, \sigma_i^2)\} \tag{27}$$

and the optimal controller is

$$U_i^*(\mu_i, \sigma_i^2) = \begin{cases} -\mu_i & \text{if } |\mu_i| > \alpha_{i,\sigma_i} \\ 0 & \text{if } |\mu_i| \leq \alpha_{i,\sigma_i} \end{cases} \tag{28}$$

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

The adjustment rule is therefore of the *deadband* type (Box and Jenkins, 1963, Box and Luceño, 1997, Crowder, 1992). This means that only for process states with a mean far enough from target an adjustment is justified. The action limits defining the deadband are clearly a function of the adjustment cost c.

According to the foregoing discussion, only two choices are compared for U_i^* in calculating each $R_i(\mu_i, \sigma_i^2)$. During the backwards calculation, α_{i,σ_i} can be obtained by finding the minimum $|\mu_i|$ with $|U_i^*(\mu_i, \sigma_i^2)| > 0$. Once the limits α_{i,σ_i} are obtained, the control policy in (28) can be applied to adjust the initial offset for the process described by (1)- (2).

4 Computer implementation of the solution

The complete procedure to adjust process (1-2) with the the optimal controller (28) can be divided into two steps: *i)* generating the control table with the adjustment limits, a computation that can be done off-line, that is, prior to start producing parts, and *ii)* on-line adjustment of the process. We now explain each of these two steps in detail. An R program that implements this procedure (called chart.R) can be downloaded from http://www.ie.psu.edu/researchlabs/EngineeringStatistics/software.htm.

4.1 Generating the control table

The state variables (μ_i, σ_i) belong to an unbounded continuous space $R \oplus R^+$ where $\oplus$ denotes Cartesian set product. Such state space need to be mapped to

a bounded discrete space D with finite elements so that for each $(\mu_i, \sigma_i) \in D$, $R_i(\mu_i, \sigma_i)$ can be approximated numerically by backwards induction. A typical form of D will be $D = D_\mu \oplus D_\sigma$, where $D_\mu = \{\mu | \mu = i \times d_\mu, i = -n_1, -n_1 + 1, ...n_1 - 1, n_1\}$ and $D_\sigma = \{\sigma | \sigma = i \times d_\sigma, i = 0, 1, ..., n_2\}$. d_μ and d_σ are positive numbers and n_1, n_2 are positive integers. A state (μ, σ) is mapped to $(\mu', \sigma') \in D$ such that μ' is the closest element to μ in D_μ and σ' is the closest element to σ in D_σ. Increments d_μ and d_σ are set to be small enough for accuracy. Upper bounds $n_1 \times d_\mu$ and $n_2 \times d_\sigma$ are set to be large enough in order to include most of the possible states that can occur.

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

- Step 1. Specify the value of c, the relative adjustment cost. Set the prior state variables, κ_0, ν_0, μ_0 and σ_0, according to the prior information of the process. A proper selection of the prior distribution is discussed in the next section.

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

- Step 3. Calculate $R_i(\mu_i, \sigma_i)$ for each $(\mu_i, \sigma_i) \in D$ backwards for each stage i, by computing (20) using Monte Carlo integration.

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

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

As shown in Table 1, the limits α_{i,σ_i} are generally increasing with σ_i, reflecting the fact that a higher σ_i value provides less precision in the process mean or offset θ_i. This can be better seen from the unconditional variance of the process

Table 1
Control table showing the adjustment limits α_{i,σ_i} calculated for the example.

Part i	σ_i										
	0	1	2	3	4	5	6	7	8	9	10
1	1.0	1.9	2.3	2.6	2.7	2.8	2.9	2.9	2.9	3.0	3.0
2	1.0	1.5	2.0	2.3	2.5	2.6	2.7	2.7	2.8	2.9	2.9
3	1.1	1.3	1.8	2.0	2.2	2.4	2.5	2.6	2.7	2.7	2.8
4	1.2	1.3	1.6	1.9	2.1	2.2	2.3	2.4	2.5	2.6	2.7
5	1.3	1.3	1.5	1.7	1.9	2.1	2.2	2.3	2.4	2.5	2.5
6	1.4	1.4	1.5	1.6	1.8	2.0	2.1	2.2	2.3	2.4	2.4
7	1.5	1.6	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.3
8	1.8	1.8	1.8	1.8	1.9	1.9	2.0	2.0	2.1	2.2	2.2
9	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.3	2.3
10	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0

mean, $Var(\theta_i)$, which equals to:

$$Var(\theta_i) = E[Var(\theta_i|\sigma_v)] + Var(E[\theta_i|\sigma_v]) = E[\sigma_v^2/\kappa_i] + Var(\mu_i) = \frac{\sigma_i^2}{\kappa_i}\frac{\nu_i}{\nu_i - 2}. \tag{29}$$

The "U" shape of the deadband implies that the optimal controller is more likely to suggest adjusting when the precision is higher. This fact is also reflected by another tendency observed from the table, that for a fixed σ_i, α_{i,σ_i} is "U" shaped when seen as a function of the part number i. This implies the adjustment limits are "U" shaped. This is explained as follows (see also Figure 1). The more observations, the higher the precision of the predictive density of the process mean. Since κ_i increases linearly as a function of i, the variance $Var(\theta_i)$ is decreasing with i. However, the role of the precision becomes unimportant when the process approaches its end, when high precision is already obtained by the large number of observations and it can not be significantly increased further by simply increasing the number of observations. At that point, all α_{i,σ_i} limits are close to each other for a given stage i regardless of σ_i. This is so because as κ_i gets larger it reduces the differences in precision obtained from the different σ_i's. Near the end, if σ_i is fixed, α_{i,σ_i} is increasing with i because the future benefit of an adjusted process is not justified, given there are few parts to produce until the end of the batch of N parts. A similar "funneling out" of the adjustment limits due to an end of horizon effect was reported by Crowder (1992) for a different process model with known parameters.

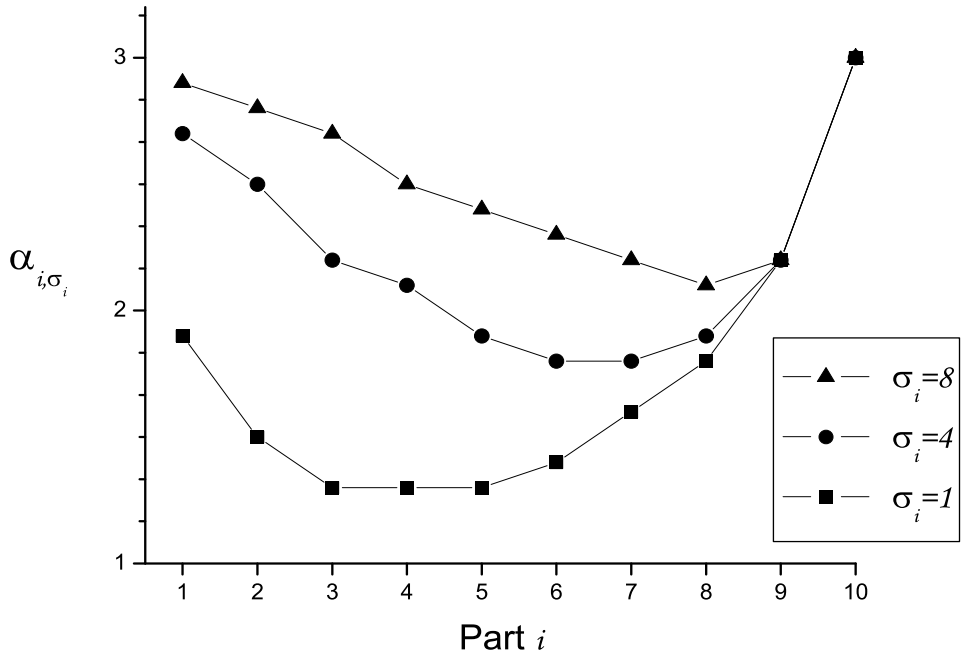

Fig. 1. Adjustment limits α_{i,σ_i} for all parts i and $\sigma_i = 1, 4, 8$ in the numerical example.

4.2 Adjusting a Process with the Control Table

Once the control table is created, the control policy described in (28) can be applied to adjust the initial offset of a process. At each stage i of the process, $i = 1, 2, ..., N-1$, the following procedure is repeated.

- Step 1. Update the state variables (μ_i, σ_i) as in (10-13) given the new observation y_i.
- Step 2. Map σ_i to $\sigma_i' \in D_\sigma$, i.e., find the σ_i' on the control table closest to σ_i.
- Step 3. If $|\mu_i| > \alpha_{i,\sigma_i'}$, adjust the process by $U_i = -\mu_i$; otherwise make no adjustment $(U_i = 0)$.

Example (cont.) An example is shown to illustrate the procedure to adjust a process with a generated table. The process is simulated such that it has 10 parts and the initial offset is $\theta_0 = 6$. Assume the adjustment cost is $c = 9$ and the prior state variables are $\mu_0 = 0$, $\sigma = 10$, $\kappa_0 = 1$, and $\nu_0 = 2.01$. The same discrete state space D is chosen as in section 4.1, so table 1 is used to adjust

the process at each stage.
At stage 0, i.e. at the beginning of the process, the prior state variable $\mu_0 = 0$ implies that no adjustment is necessary. Therefore, $U_0 = 0$.
At stage 1, the first part is made and observed, and suppose we get $Y_1 = 4.28$. The new current state variables are:

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

The closest element to σ_1 in D_σ is $\sigma_1' = 8$. Referring to table 1, $\alpha_{1,8} = 2.8$ is larger than μ_1. Therefore, adjusting is not justified, so $U_1 = 0$.
At Stage 2, suppose we observe $Y_2 = 6.70$. The updated state variables are:

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

In this case, $\sigma_2' = 7$ in D_σ is the closest element to σ_2, and we find $\alpha_{2,7} = 2.6$ in the control table. Since $\mu_2 > \alpha_{2,7}$, we should adjust the process by $U_2 = -3.66$. Continuing in this form, at each stage from 3 to 9, we update the state variables and decide whether to adjust or not according to the control table. An illustration after ten hypothetical observation is shown in Figure 2. The figure shows all observations and how the true means θ_i change due to the adjustments.

As we can see in the figure, for this specific sample process, only two adjustments are made at times 2 and 5. Figure 3 shows an "adjustment plot" which indicates when the process mean μ_i exceeds the adjustment limits α_{i,σ_i}. As it can be seen, the adjustments efficiently drive the true mean closer to target.

5 Sensitivity Analysis and Extremal Situations

Thus far, only one example has shown how the adjustment rule works. In this section, more cases are investigated and for each case more replications are simulated to evaluate the average performance of the optimal policy. In each

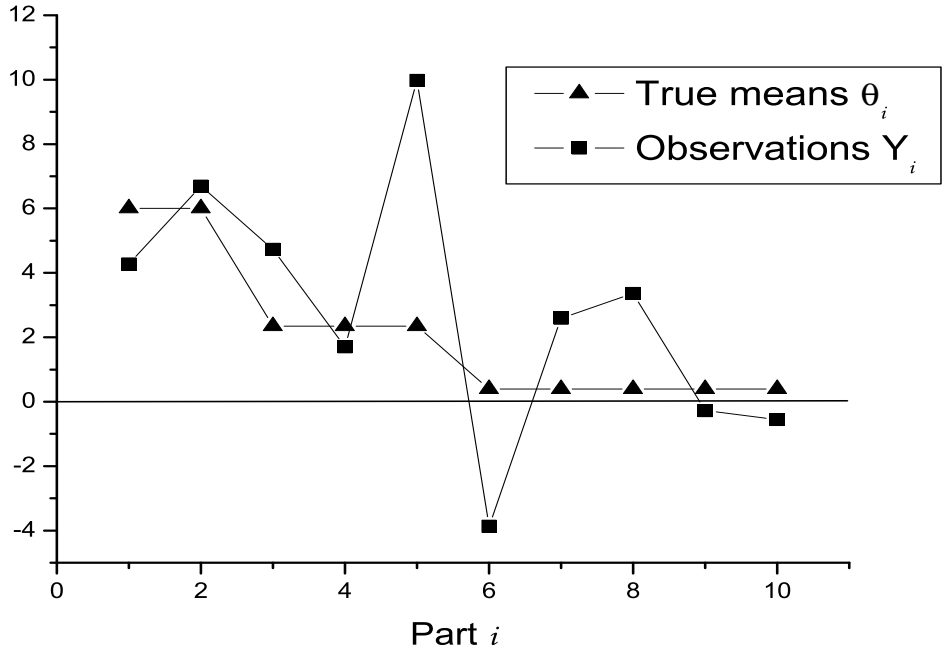

Fig. 2. Plot of the true means and observations

of the next subsections the sensitivity of the optimal policy with respect to different process and cost parameters is investigated.

5.1 System Parameters: θ_0,σ_ν

To show the impact of the proposed adjustment rule on a real processes, the control table for the example in Section 4.1 is applied to different types of processes with 10 parts and fixed adjustment cost $c = 9$, assuming the prior state variables are the same for all cases, $\mu_0 = 0$, $\sigma = 10$, $\kappa_0 = 1$, $\nu_0 = 2.01$. For each type of process, 1000 replications are made to estimate the expected total cost. The results are summarized in Table 2, where θ_0 is the true initial offset and σ_v is the standard deviation of the noise. $\bar{L}_A$ is the average loss when the adjustment rule was applied and $\bar{L}_N$ is the average loss without any adjustment. $\Delta = 1 - \bar{L}_A/\bar{L}_N$ is the saving rate made by the adjustment rule. The 99% confidence intervals of Δ are included in the table.

The behavior of the adjustment rule can also be seen in Table 3, where $\overline{|\theta_i|}$ is the average level of true offset for each part and P_{Ai} is the proportion of times an adjustment is made right before part i. P_{Ai} is obtained from the total number of adjustments made right before part i divided by the total number

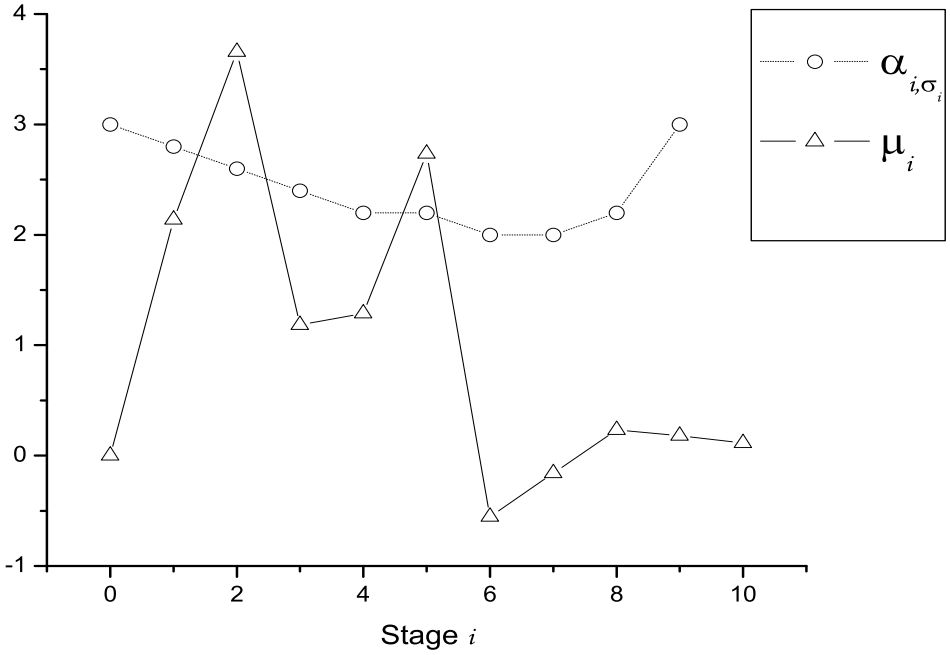

Fig. 3. Plot of μ_i and corresponding α_{i,σ_i}

Table 2
Savings made by the adjustment

Case	θ_0	σ_v	$\bar{L}_A$	$\bar{L}_N$	Δ
1	0	2	40.2	39.4	-1.1%±0.8%
2	2	2	75.3	81.1	3.2%±1.6%
3	4	2	99.2	201.3	49.2%±0.9%
4	0	3	100.1	89.7	-9.3%±2.0%
5	3	3	141.3	183.0	18.6%±1.9%
6	6	3	193.8	448.4	55.7%±0.8%

of replications.

As one can see in Tables 2 and 3, the proposed adjustment rule reduces the costs significantly as θ_0/σ_v goes up. Limited number of adjustments are made as early as the offsets are estimated with high confidence, and the initial offsets are brought down. The worst performance occurs when the initial offset θ_0 is actually 0 (cases 1 and 4). In such cases, any adjustment is redundant and will increase both the quadratic losses and the adjustment costs. However, if the rule is applied in these cases, very few adjustments are made and the cost increase is small. Therefore, the optimal adjustment rule is robust

Table 3
Summary of the behavior of the adjustment rule

Case		Part									
		1	2	3	4	5	6	7	8	9	10
1	$\overline{\lvert\theta_i\rvert}$	0	0.016	0.018	0.023	0.025	0.032	0.034	0.034	0.034	0.034
	P_{Ai}	0	0.005	0.001	0.002	0.005	0.005	0.003	0.000	0.000	0.000
2	$\overline{\lvert\theta_i\rvert}$	2	1.966	1.890	1.765	1.630	1.503	1.326	1.234	1.224	1.224
	P_{Ai}	0	0.043	0.073	0.093	0.086	0.075	0.099	0.049	0.006	0.000
3	$\overline{\lvert\theta_i\rvert}$	4	3.312	2.274	1.526	1.253	1.104	1.048	1.026	1.026	1.026
	P_{Ai}	0	0.208	0.329	0.259	0.104	0.066	0.027	0.015	0.000	0.000
4	$\overline{\lvert\theta_i\rvert}$	0	0.188	0.315	0.324	0.372	0.364	0.362	0.350	0.341	0.341
	P_{Ai}	0	0.054	0.047	0.031	0.046	0.027	0.034	0.027	0.006	0.000
5	$\overline{\lvert\theta_i\rvert}$	3	2.586	2.089	1.668	1.353	1.161	1.043	0.943	0.926	0.926
	P_{Ai}	0	0.188	0.196	0.164	0.133	0.091	0.062	0.056	0.013	0.000
6	$\overline{\lvert\theta_i\rvert}$	6	3.722	2.614	2.175	1.976	1.814	1.647	1.487	1.450	1.450
	P_{Ai}	0	0.565	0.310	0.149	0.099	0.080	0.090	0.080	0.018	0.000

with respect to different process parameters without requiring the parameter estimates before processing.

5.2 Changes in initial prior state variables: μ_0,σ_0,κ_0,ν_0

According to (5), ν_0 and σ_0^2 determine the precision of the initial prior density of σ_v^2. From (4) and (29), κ_0, ν_0 and σ_0 together determine the precision of the initial prior density of θ_0. Low precisions (i.e., large variances) should be utilized when no prior information about the process parameters (θ_0, σ_v^2) is available. In such case, σ_0 should be set to a reasonably large number and κ_0 and ν_0 should be as small as possible. κ_0 can be a small positive number like 0.01, and ν_0 may be a number slightly larger than 2, because $\nu_0 > 2$ is necessary for calculating the optimal expected loss $R_0(\mu_0, \sigma_0)$, where the variance of Y_1 described in equation (16) is needed. These choices for a noninformative prior are close to Jeffrey's rule (see, e.g., Gelman et al., 2003)

The parameter κ_0 determines (more than σ_0^2 and ν_0) the precision of the initial prior density for θ_0. A larger κ_0 indicates more confidence that the initial offset is close to μ_0. The parameter κ_0 can also be considered a tuning variable used to trade-off the performance when θ_0/σ_ν is small and when θ_0/σ_ν is large.

Example (cont.) We illustrate the effect of changing κ_0 in the example of Section 4 by setting it equal to 0.01 (as opposed to $\kappa = 1$). The other parameters were not changed: $\mu_0 = 0, \sigma_v = 10$, and $\nu_0 = 2.01$. The new performance statistics are shown in Table 4 for the same six cases as before.

Table 4
Savings made by the adjustment when $\kappa_0 = 0.01$ and $\mu_0 = 0$

Case	θ_0	σ_v	$\bar{L}_A$	$\bar{L}_N$	Δ
1	0	2	53.0	40.3	-25.8%±4.7%
2	2	2	77.1	79.2	-4.3%±3.3%
3	4	2	84.4	198.4	55.7%±1.5%
4	0	3	123.6	90.1	-35.0%±3.8%
5	3	3	146.9	180.9	13.2%±2.8%
6	6	3	172.6	446.0	59.9%±1.4%

Comparing Table 4 with Table 2, although the smaller value of κ_0 slightly improves the performance for processes with high θ_0/σ_0, the loss for low θ_0/σ_0 cases is dramatically increased. The reason lies in the forward updating of μ_i in equation (13). A smaller κ_0 gives more weight to new observations. When σ_ν is large, the fluctuation of the first several observations can drive μ_i high enough to justify adjustments no matter how small or non-existent the true offset is. A slightly larger value of κ_0 (around 1.0) makes the adjustment rule more robust.

5.3 Fixed Adjustment Cost: c

When c is very large, the limits α will also be large and hence no adjustment will be made. When $c \to 0$, the α's will converge to 0. In that case an adjustment will be made at every stage. It is easy to show that, when $c = 0$ and $\kappa_0 \to 0$, the proposed adjustment rule works exactly as Grubbs' Harmonic Rule (Grubbs, 1983).

6 A simpler case: σ_v^2 known.

Now we consider a simpler problem, in which the system equations and cost function are the same as in (1) - (3), but the system parameter σ_v is assumed to be known. A different Bayesian model is employed to estimate the unknown parameter θ_0, based on which we can adjust the process.

A conjugate prior distribution is given by

$$\theta_0 \sim N(\mu_0, \tau_0^2) \tag{30}$$

$$Y_1|\theta_0 \sim N(\theta_0 + U_0, \sigma_v^2). \tag{31}$$

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

$$\theta_i|Y^{(i)}, U^{(i)} \sim N(\mu_i, \tau_i^2), \tag{32}$$

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

$$Y_{i+1}|\theta_i \sim N(\theta_i + U_i, \sigma_v^2). \tag{33}$$

The posterior predictive density of Y_i is

$$Y_{i+1}|Y^{(i)}, U^{(i)} \sim N(\mu_i + U_i, \sigma_v^2 + \tau_i^2) \tag{34}$$

This density is characterized by two state variables, μ_i and τ_i^2. These two variables can be easily updated recursively. Using the prior density in stage $i-1$ as the prior, we obtain the recursive updating equations:

$$\frac{1}{\tau_i^2} = \frac{1}{\tau_{i-1}^2} + \frac{1}{\sigma_v^2} \tag{35}$$

$$\mu_i = \frac{\frac{\mu_{i-1}+U_{i-1}}{\tau_{i-1}^2} + \frac{y_i}{\sigma_v^2}}{\frac{1}{\tau_{i-1}^2} + \frac{1}{\sigma_v^2}} = \frac{\mu_{i-1} + U_{i-1} + \frac{\tau_{i-1}^2}{\sigma_v^2} y_i}{1 + \frac{\tau_{i-1}^2}{\sigma_v^2}} \tag{36}$$

6.1 Dynamic Programming

Once the prior state variable τ_0 is specified, all other state variables τ_i are constants determined by the stage index i. Therefore, the predictive density at stage i is described by only one state variable, μ_i. The predictive density $Y_{i+1}|Y^{(i)}, U^{(i)}$ can be denoted as $\phi_i(Y_i - \mu_i)$, a normal distributed density with mean μ_i and variance $\sigma_v^2 + \tau_i^2$.

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

$$
\begin{aligned}
R_{N-1}(\mu_{N-1}) &= \min_{U_{N-1}} E\{Y_N^2 + c\delta(U_{N-1})\} \\
&= \min_{U_{N-1}} \{\sigma_v^2 + \tau_{N-1}^2 + (\mu_{N-1} + U_{N-1})^2 + c\delta(U_{N-1})\} \\
&= \sigma_v^2 + \tau_{N-1}^2 + \min\{\mu_{N-1}^2, c\}. \qquad (37)
\end{aligned}
$$

The optimal adjustment which minimizes this last expression is clearly

$$
U_{N-1} = \begin{cases} -\mu_{N-1} & \text{if } |\mu_{N-1}| > c^{1/2} \\ 0 & \text{if } |\mu_{N-1}| \leq c^{1/2} \end{cases} \qquad (38)
$$

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

$$
R_i(\mu_i) = \min_{U_i} \{\sigma_v^2 + \tau_i^2 + (\mu_i + U_i)^2 + c\delta(U_i) + E\{R_{i+1}(\mu_{i+1})|U_i\}\}, \qquad (39)
$$

where

$$
E\{R_{i+1}(\mu_{i+1})|U_i\} = \int R_{i+1}(\mu_{i+1})\phi_i(Y_{i+1} - (\mu_i + U_i))dY_{i+1}, \qquad (40)
$$

and

$$
\mu_{i+1} = \frac{\mu_i + U_i + \frac{\tau_i^2}{\sigma_v^2} Y_{i+1}}{1 + \frac{\tau_i^2}{\sigma_v^2}} \qquad (41)
$$

Now we introduce a theorem which is proved in Appendix 2.

Theorem 1. For the problem described by equation (39), the following three statements are true:
i)$R_i(\mu_i)$ is a nonnegative function which is symmetric about $\mu_i = 0$ and nondecreasing in $|\mu|$.
ii) $E\{R_{i+1}(\mu_{i+1})|U_i\}$ is a function of $\mu_i + U_i$; it is symmetric about $\mu_i + U_i = 0$ and nondecreasing in $|\mu_i + U_i|$
iii) If the optimal controller U^* is not zero, $U^* = -\mu_i$.

According to the third statement in **Theorem 1**, (39) can be transformed to

$$
R_i(\mu_i) = \sigma_v^2 + \tau_i^2 + \min\{\mu_i^2 + E\{R_{i+1}(\mu_{i+1})|U_i = 0\}, c + E\{R_{i+1}(\mu_{i+1})|U_i = -\mu_i\}\} \qquad (42)
$$

The corresponding optimal controller is

$$
U_i^*(\mu_i) = \begin{cases} -\mu_i & \text{if } \mu_i^2 + E\{R_{i+1}(\mu_{i+1})|U_i = 0\} > c + E\{R_{i+1}(\mu_{i+1})|U_i = -\mu_i\} \\ 0 & \text{o.w.} \end{cases} \qquad (43)
$$

We notice that, in (43), $\mu_i^2 + E\{R_{i+1}(\mu_{i+1})|U_i = 0\}$ is an increasing symmetric continuous function of $|\mu_i|$, whose minimum value is at $E\{R_{i+1}(\mu_{i+1})|U_i = 0\}$ when $\mu_i = 0$ and a maximum value of infinite. The right hand side of the inequality is a constant for all μ_i, which is between the minimum and maximum values of the left hand side. So there exist a nonnegative critical number α_i such that both sides are equal when $|\mu_i| = \alpha_i$. Thus the optimal controller can be rewritten as

$$U_i^*(\mu_i) = \begin{cases} -\mu_i & \text{if } |\mu_i| > \alpha_i \\ 0 & |\mu_i| \leq \alpha_i \end{cases} \tag{44}$$

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

6.2 Computer Implementation

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

Example 2. Suppose $N = 10$ parts are to be processed and the process parameter $\sigma_\upsilon = 1$ is known. The adjustment cost is $c = 9$. The initial prior parameter is set at $\tau_0 = 1$. The discrete state variable space D is chosen to be $\{\mu_i|\mu_i = 0.1k\}$, where k is an integer such that $|a| \leq 50$. The control limits α_i are shown in Table 5. The usage of the control limits table is very similar

Table 5
Control table showing the adjustment limits α_i calculated for the example.

Stage i	0	1	2	3	4	5	6	7	8	9
α_i	2.7	2.3	2.1	1.9	1.7	1.6	1.7	1.8	2.2	3

to the procedure described in section 4.2. At each stage i, $i = 0, ..., N-1$, we compare the updated state variable μ_i with the control limit α_i. If $|\mu_i| > \alpha_i$, we make an adjustment $U_i = -\mu_i$. Otherwise, no adjustment is needed at this stage.

An R computer program that implements the variance known case (chart_n.R)

is also available from download from http://www.ie.psu.edu/researchlabs/ EngineeringStatistics/software.htm.

7 Comparison of known vs. unknown σ_υ^2 approaches

We have presented two methods for the adjustment problem described in (1) - (3): **1.** The method derived in sections 2 to 4, in which the system parameter σ_υ is unknown; **2.** The method described in section 6, where the system parameter σ_υ is assumed to be known. A comparison between each of these two methods and Grubbs' Harmonic Rule is conducted next.

Suppose a process is going to be adjusted when $N = 10$, and no prior information about the system parameters θ_0 and σ_υ is given. For method 2, a guess of σ_υ is required, and suppose this guess is $\sigma_\upsilon = 5$. The prior parameters are set to be $\mu_0 = 0$ and $\tau_0 = 5$. For method 1, the prior parameters are set as following: $\mu_0 = 0$, $\sigma_0 = 5$, $\kappa_0 = 1$, and $\nu_0 = 2.01$. Notice that from equations (10) and (36) if $\kappa_0 = \frac{\tau_0^2}{\sigma_\upsilon^2}$, the recursive updating of μ_i is essentially the same for methods 1 and 2. The only difference between the methods will be the estimate of true parameter σ_υ^* and the confidence of μ_i as a estimator of θ_i. (σ_υ^* here denotes the true parameter to differ from the parameter σ_υ used in method 2.) Several cases were investigated to compare method 1, method 2 and Grubbs' rule. The settings of the prior parameters were kept the same for all cases. The results are summarized in Table 6. Each saving rate in that table is the percentage savings induced by an adjustment rule, compared to the total loss incurred when there is no adjustment. The savings are average estimators based on 1000 replications. In each replication, the same random errors were used for method 1, method 2, Grubbs' rule, and no adjustment case. The 99% confidence intervals are also shown in the table.

Some general conclusions can be reached from Table 6:

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

Table 6
Average savings compared to total losses under no adjustment

c	$\frac{\theta_0}{\sigma_v^*}$	θ_0	σ_v^*	Savings		
				method 1	method 2	Grubbs
4		0	1	-0.2%±0.4%	0.0%±0.0%	-555%±24%
	0	0	5	-17.4%±1.6%	-16.8%±1.6%	-35.0%±1.7%
		0	7	-16.8%±1.5%	-16.7%±1.4%	-24.5%±1.4%
		1	1	-0.5%±1.0%	-0.6%±0.6%	-226.7%±9.9%
	1	5	5	29.3%±1.7%	29.3%±1.7%	23.4%±1.9%
		7	7	30.9%±1.8%	31.0%±1.7%	28.1%±1.8%
		2	1	38.8%±1.2%	32.0%±1.4%	-25.5%±3.0%
	2	10	5	62.5%±0.8%	62.4%±0.8%	60.9%±0.9%
		14	7	63.2%±0.8%	63.2%±0.8%	62.6%±0.8%
		3	1	59.5%±0.5%	58.6%±0.5%	34.1%±1.0%
	3	15	5	73.3%±0.4%	73.3%±0.4%	72.9%±0.4%
		21	7	73.7%±0.4%	73.8%±0.4%	73.6%±0.4%
9		0	1	0.0%±0.0%	0.0%±0.0%	-1149%±49%
	0	0	5	-19.6%±1.9%	-19.0%±2.0%	-61.2%±2.9%
		0	7	-19.2%±1.6%	-18.7%±1.6%	-37.8%±1.7%
		1	1	-0.4%±0.3%	-0.02%±0.3%	-532%±28%
	1	5	5	27.5%±1.9%	27.3%±1.8%	11.4%±2.2%
		7	7	31.5%±1.8%	31.3%±1.8%	24.1%±2.0%
		2	1	29.9%±1.4%	6.1%±1.0%	-147%±5.5%
	2	10	5	61.0%±0.8%	60.7%±0.8%	56.6%±1.0%
		14	7	62.3%±0.8%	62.2%±0.8%	60.3%±0.9%
		3	1	54.1%±0.5%	45.2%±0.8%	-19.7%±1.9%
	3	15	5	72.4%±0.4%	72.5%±0.4%	71.3%±0.4%
		21	7	72.9%±0.4%	73.0%±0.4%	72.4%±0.4%
16		0	1	0.0%±0.0%	0.0%±0.0%	-2065%±102%
	0	0	5	-19.0%±2.1%	-17.9%±2.1%	-100.9%±4.5%
		0	7	-19.6%±1.9%	-18.8%±1.8%	-57.5%±2.6%
		1	1	0.0%±0.0%	0.0%±0.1%	-931%±38%
	1	5	5	25.3%±1.8%	25.0%±1.9%	-5.4%±2.8%
		7	7	25.6%±1.8%	25.6%±1.8%	10.8%±2.3%
		2	1	1.6%±0.7%	0.0%±0.3%	-307.3%±9.5%
	2	10	5	58.6%±0.8%	58.5%±0.8%	50.3%±1.0%
		14	7	60.9%±0.8%	61.0%±0.8%	57.0%±0.9%
		3	1	35.3%±1.1%	24.9%±1.2%	-101%±3.5%
	3	15	5	71.2%±0.4%	71.3%±0.4%	68.4%±0.5%
		21	7	72.5%±0.4%	72.6%±0.4%	71.4%±0.5%

2 is negligible. We can conclude that method 1 is the most robust of the three methods considered.

- Grubbs' harmonic rule has a performance always dominated by the other two more complex policies. This is because it was not designed considering adjustment costs, thus it suggests adjustments for every part i. However,

given its notorious simplicity it should be the preferred method when the adjustment cost is very low or for very noisy processes.

Acknowledgement. This research was supported by NSF grant DMI 0200056.

References

Bjorke, O. (1989). *Computer-Aided Tolerancing*, 2nd. ed. New York: ASME Press.

Box, G.E.P. and Jenkins, G. M., (1963). "Further Contributions to Adaptive Quality Control Simultaneous Estimations of Dynamics Nonzero Costs," *Bulletin of the International Statistical Institute*, 34, 943-974

Box, G.E.P. and Luceño, A., (1997). *Statistical Control by Monitoring and Feedback Adjustment*. John Wiley & Sons, New York, NY.

Del Castillo, E., Pan, R., and Colosimo, B.M., (2003a). "A Unifying View of Some Process Adjustment Method", *Journal of Quality Technology*, 35, 286-293

Del Castillo, E., Pan, R., and Colosimo, B.M., (2003b). "Small Sample Performance of Some Statistical Adjustment Methods", *Communications in Statistics, Simulation and Computation*, 32(3), pp. 923-942.

Crowder, S. V. ,(1992). "An SPC Model for Short Production Runs: Minimizing Expected Cost", *Technometrics*, 34, 64-73

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B., (2003). *Bayesian Data Analysis*, 2nd ed. Chapman & Hall/CRC.

Grubbs, F.E., (1954)."An Optimum Procedure for Setting Machines or Adjusting Processes", *Industrial Quality Control*, July, reprinted in *Journal of Quality Technology* (1983), 15, 4, 186-189

Gyrna, F.M., (2001). *Quality Planning & Analysis*, 4th ed. New York: McGraw-Hill.

Trietsch, D., (1998). "The Harmonic Rule for Process Setup Adjustment with Quadratic Loss", *Journal of Quality Technology*, 30, 1, 75-84

Appendix 1. Derivation of the predictive density

We have that

$$f(Y_{i+1}|Y^{(i)},U^{(i)}) = \int_0^\infty \int_{-\infty}^\infty f(Y_{i+1}|\ \theta_i,\sigma_v^2) f(\theta_i,\sigma_v^2 \mid Y^{(i)},U^{(i)})\ d\theta_i\ d\sigma_v^2.$$

where the two densities in the integral are as in (6) and (8), respectively. Therefore,

$$f(Y_{i+1}|Y^{(i)},U^{(i)}) \propto \int_0^\infty \int_{-\infty}^\infty (\sigma_v^2)^{-\nu_i/2-2} \exp\left\{\frac{-1}{2\sigma_v^2}\left(k_i(\mu_i-\theta_i)^2 + (Y_{i+1}-\theta_i)^2 + \nu_i\sigma_i^2\right)\right\} d\theta_i\ d\sigma_v^2.$$

Integrating first with respect to θ_i we get

$$\begin{aligned} f(Y_{i+1}|Y^{(i)},U^{(i)}) &\propto \int_0^\infty \sigma_v^{-\nu_i/2-3/2} \exp\left\{-\frac{1}{2}\frac{\frac{\kappa_i(Y_{i+1}-\mu_i)^2}{\kappa_i+1} + \nu_i\sigma_i^2}{\sigma_v^2}\right\} d\sigma_v^2 \\ &\propto \left[\frac{\kappa_i(Y_{i+1}-\mu_i)^2}{\kappa_i+1} + \nu_i\sigma_i^2\right]^{-\left(\frac{\nu_i+1}{2}\right)} \\ &\propto \left[\frac{\kappa_i}{\kappa_i+1}\left(\frac{Y_{i+1}-\mu_i}{\sigma_i}\right)^2 \frac{1}{\nu_i} + 1\right]^{-\left(\frac{\nu_i+1}{2}\right)} \end{aligned}$$

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

Appendix 2. Proof of Theorem 1.

Lemma 1 Let Z be a random variable with probability density $\Psi(z)$, which is symmetric about $z=0$ and nonincreasing in $|z|$. Let $H(z)$ be a nonnegative function that is symmetric about $z=0$ and nondecreasing in $|z|$. Then, the function

$$G(\mu) = E\{H(\mu+Z)\} \tag{45}$$

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

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

$$
\begin{aligned}
G(-\mu) &= \int_{-\infty}^{\infty} H(-\mu+z)\Psi(z)dz \\
&= \int_{\infty}^{-\infty} H(-\mu-z)\Psi(-z)d(-z) \\
&= \int_{-\infty}^{\infty} H(\mu+z)\Psi(z)dz \\
&= G(\mu). \qquad (46)
\end{aligned}
$$

Now we are going to prove that $G(\mu)$ is nondecreasing in $|\mu|$.
Since $H(z)$ is symmetric and nondecreasing, there exits a nonnegative value β_r,

$$\beta_r = \inf z | H(z) > r, z \geq 0 \qquad (47)$$

such that

$$P\{H(z) > r\} = P\{|z| > \beta_r\} \qquad (48)$$

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

$$G(\mu) = E\{H(\mu+z)\} = \int_0^{\infty} P\{H(\mu+Z) > r\}dr. \qquad (49)$$

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

$$G(\mu_2) - G(\mu_1) = \int_0^{\infty} \{P\{H(\mu_2+Z) > r\} - P\{H(\mu_1+Z) > r\}\}dr \qquad (50)$$

Consider,

$$
\begin{aligned}
& P\{H(\mu_2+Z) > r\} - P\{H(\mu_1+Z) > r\} \qquad (51) \\
&= P\{|\mu_2+Z| > \beta_r\} - P\{|\mu_1+Z| > \beta_r\} \\
&= P\{\mu_2+Z < -\beta_r\} + P\{\mu_2+Z > \beta_r\} - P\{\mu_1+Z < \beta_r\} - P\{\mu_1+Z > \beta_r\} \\
&= P\{Z < -\beta_r-\mu_1-\delta\} + P\{Z > \beta_r-\mu_1-\delta\} - P\{Z < -\beta_r-\mu_1\} - P\{Z > \beta_r-\mu_1\} \\
&= P\{\beta_r-\mu_1-\delta < Z < \beta_r-\mu_1\} - P\{-\beta_r-\mu_1-\delta < Z < -\beta_r-\mu_1\} \\
&= \int_{\beta_1-\mu_1-\delta}^{\beta_r-\mu_1} \Psi(z)dz - \int_{-\beta_1-\mu_1-\delta}^{-\beta_r-\mu_1} \Psi(z)dz \\
&= \int_0^{\delta} \{\Psi(\beta_r-\mu_1-\delta+s) - \Psi(-\beta_r-\mu_1-\delta+s)\}ds. \qquad (52)
\end{aligned}
$$

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

$$|-\beta_r-\mu_1-\delta+s| = \beta_r+\mu_1+\delta-s \geq |\beta_r-\mu_1-\delta+s| \qquad (53)$$

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

$$\Psi(\beta_r-\mu_1-\delta+s) - \Psi(-\beta_r-\mu_1-\delta+s) \geq 0. \qquad (54)$$

Hence, (52)≥ 0, and therefore (50)≥ 0. So $G(\mu)$ in nondecreasing in $|\mu|$. Q.E.D.

Theorem 1. For the problem described by equation (39), the following three statements are true:
i)$R_i(\mu_i)$ is a nonnegative function which is symmetric about $\mu_i = 0$ and nondecreasing in $|\mu|$.
ii) $E\{R_{i+1}(\mu_{i+1})|U_i\}$ is a function of $\mu_i + U_i$; it is symmetric about $\mu_i + U_i = 0$ and nondecreasing in $|\mu_i + U_i|$
iii) If the optimal controller U^* is not zero, $U^* = -\mu_i$.
Proof. Obviously, statements *i)*, *ii)*, and *iii)* are true for stage $N-1$, where R_N is defined to be 0.
By induction, assume *i)*, *ii)*, and *iii)* hold true for stage $i+1$. At stage i, first consider

$$\begin{aligned} E\{R_{i+1}(\mu_{i+1})|U_i\} &= \int R_{i+1}\left(\frac{\mu_i + U_i + \frac{\tau_i^2}{\sigma_v^2}Y_{i+1}}{1+\frac{\tau_i^2}{\sigma_v^2}}\phi_i(Y_{i+1} - (\mu_i + U_i))\right) dY_{i+1} \\ &= \int R_{i+1}\left(\mu_i + U_i + \frac{1}{\sigma_v^2/\tau_i^2 + 1}x\right)\phi_i(x)dx \\ &= \int R_{i+1}(\mu_i + U_i + z)\phi_i'(z)dz, \end{aligned} \tag{55}$$

where,

$$x = Y_{i+1} - (\mu_i + U_i), \tag{56}$$

$$z = \frac{1}{\sigma_v^2/\tau_i^2 + 1}x, \tag{57}$$

and $\phi_i'(z)$ is the density function of a normal distribution with mean 0 and variance $\frac{\sigma_v^2+\tau_i^2}{(\sigma_v^2/\tau_i^2+1)^2}$., which is symmetric about $z = 0$ and nonincreasing in $|z|$. Lemma 1 implies that (55) is symmetric about $\mu_i + U_i = 0$ and nondecreasing in $|\mu_i + U_i|$. Therefore, *ii)* is proved.

Consider

$$R_i(\mu_i) = \min_{U_i}\{\sigma_v^2 + \tau_i^2 + (\mu_i + U_i)^2 + c\delta(U_i) + E\{R_{i+1}(\mu_{i+1})|U_i\}. \tag{58}$$

If the optimal control $U_i^* \neq 0$, obviously (58) is optimized when $U_i^* = -\mu_i$. Therefor, *iii)* is proved.

Finally, note that (58) can be rewritten as

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

(59)

Consider the terms in (59). μ_i^2 and $E\{R_{i+1}(\mu_{i+1})|U_i = 0\}$ are both functions that are symmetric about $\mu_i = 0$, and nondecreasing in $|\mu_i|$. All the other terms are constants for $\forall\mu_i$. So $R_i(\mu_i)$ is indeed a function that is symmetric about $\mu_i = 0$ and nondecreasing in $|\mu_i|$, and *i)* is proved to be true. Q.E.D.