A Bivariate Dead band Process Adjustment Policy

Enrique Del Castillo
Department of Industrial and Manufacturing Engineering,
The Pennsylvania State University,
University Park, PA 16802, USA

and

Rainer Göb
Institut für Angewandte Mathematik und Statistik, Universität Würzburg,
Sanderring 2, D-97070 Würzburg, Germany

Abstract

A bivariate extension to Box and Jenkins (1963) “machine tool” problem is presented in this paper. The model balances the fixed cost of making an adjustment, which is independent of the magnitude of the adjustment, with the cost of running the process off-target, which is assumed quadratic. It is assumed that two controllable factors are available to compensate for the deviations from target of two quality characteristics or responses. Analytical formulae are presented for the computation of the loss function that combines off-target and adjustment costs per time unit. This includes expressions for the average adjustment interval and the scaled mean square deviations from target. The optimization of the model and the practical use of the resulting “dead band” adjustment strategy is illustrated with an application to semiconductor manufacturing.

Keywords: Time Series Control, Feedback Adjustment, Fixed Adjustment Cost, Bivariate IMA model.

1 Introduction.

In a landmark paper, Box and Jenkins (1963) contrasted adjustment policies for a “chemical” process with those of a “machine tool” process. The latter kind of process usually involves a large adjustment cost that is independent of the magnitude of the adjustments, in contrast to the former where off-target costs typically dominate. Assuming quadratic off-target costs, Box and Jenkins showed that the sum of off-target and fixed adjustment costs per time unit is minimized by a policy that has the form of what we will refer to in this paper as a *dead band adjustment policy*. In the univariate version of this type of policy, the process is not adjusted as long as the one step ahead minimum mean square error (MMSE) forecast, if no adjustment is made, falls inside two "control lines" placed symmetrically around the process target. The policy resembles a Shewhart chart applied to the forecasts, but the width of the control (or adjustment) limits is based on balancing the costs of running the process off-target and adjusting the process. In the present paper, we extend the Box-Jenkins univariate dead band model to the case there are two quality characteristics of interest, possibly cross-correlated, and there are two controllable factors available to adjusting the process.

Properties of univariate dead band adjustment policies have been studied by other authors more recently. Crowder (1992) solves Box and Jenkins's univariate machine tool problem using dynamic programming techniques when there is a finite number of periods in the planning horizon for the process. This is in contrast to Box and Jenkins (1963), who use a renewal reward process to minimize the long-run average cost per time unit. Crowder shows how the optimal dead band control limits funnel out as the end of the production run approaches, when a renewal of the process is assumed to occur. For the initial periods, the dead band limit width approaches the long-run solution obtained by Box-Jenkins as the planning horizon increases. Box and Kramer (1992) added a sampling cost component to the univariate machine tool model and discuss dead band policies when samples are not taken at every period. Jensen and Varderman (1993) studied the finite-horizon model in Crowder (1992) but considered the possibility that adjustment errors occur when setting the controllable factor. They show that even when there is no fixed adjustment cost (only quadratic off-target costs are present), a dead band-like policy is the optimal policy in the presence of adjustment errors. Srivastava and Wu (1991) consider the machine tool problem under the presence of inspection costs which were not included in the original model by Box and Jenkins.

Multivariate extensions of the univariate dead band control models are evidently of practical interest, given that most real-life process have multiple responses to control and multiple controllable factors. The present paper is a first attempt in a particular case which is relatively tractable yet considerably useful in practice, when only two responses are influenced by two controllable factors. The paper is organized in several sections. Sections 2 and 3 present the assumptions behind the process and the assumptions behind the control policy (or "controller"). Sections 4 and 5 discuss the loss function to be minimized and the cost assumptions involved. Section 6 gives the form of the optimal (dead band) bivariate policy. The optimal solution depends on knowing the second and fourth moments of a standardized bivariate time series, and these are derived in sections 7 and 8. With the moment formulae derived, an approximation to the loss function is given in section 9, and the accuracy of the approximation is studied in section 10. The numerical minimization of the loss function is addressed in section 11. This section contains a realistic scenario taken from the manufacturing of semiconductors where two responses are typically of interest.

2 The Process Model.

By extension to the univariate Box-Jenkins machine tool model, pairs of disturbances $\boldsymbol{z}_t = (z_{t,1}, z_{t,2})'$ are assumed to follow a bivariate IMA(1,1) process, i.e.,

$$\boldsymbol{z}_t - \boldsymbol{z}_{t-1} \quad = \quad \boldsymbol{\alpha}_t - \boldsymbol{\Theta}\boldsymbol{\alpha}_{t-1}. \tag{1}$$

Here, $\boldsymbol{\Theta} = (\theta_{ij})_{1 \le i,j \le 2}$ is a known 2×2 matrix, and the pairs $\boldsymbol{\alpha}_t = (\alpha_{t,1}, \alpha_{t,2})'$ constitute a bivariate cross correlated Gaussian white noise, i.e., normally distributed pairs with stationary variance-covariance matrix

$$\mathbf{C}_\alpha \quad = \quad \begin{pmatrix} \sigma_{1,\alpha}^2 & \kappa_\alpha \\ \kappa_\alpha & \sigma_{2,\alpha}^2 \end{pmatrix}$$

for each time t and without serial correlation, i.e., $\mathrm{Cov}[\alpha_{t,l}, \alpha_{s,m}] = 0$ for $s \neq t,\, l, m \in \{1, 2\}$. The components of the variance-covariance matrix are assumed to be known. For purposes of the control policy described in Section 3, below, it is necessary to forecast the disturbances. The minimum mean square error (MMSE) one step ahead forecast $\hat{\boldsymbol{z}}_{t+1}$ computed at time t for the disturbance vector $\boldsymbol{z}_{t+1}$ follows the EWMA recursion

$$\hat{\boldsymbol{z}}_{t+1} \quad = \quad \mathbf{L}\boldsymbol{z}_{t,l} \,+\, (\mathbf{I} - \mathbf{L})\hat{\boldsymbol{z}}_t \tag{2}$$

where $\mathbf{L} = \mathbf{I} - \boldsymbol{\Theta}$. The vector of the one step ahead forecast errors for time $t+1$ is just the white noise vector at time $t+1$, i.e.

$$\boldsymbol{z}_{t+1} - \hat{\boldsymbol{z}}_{t+1} \quad = \quad \boldsymbol{\alpha}_{t+1}. \tag{3}$$

3 The Control Model.

It is assumed that the process can be controlled via two *control factors* $X_{s+1,1}$, $X_{s+1,2}$ which are set at the *adjustment time* (*intervention time*) s, i.e., there is a delay of one time unit until the adjustment takes effect. $\boldsymbol{X} = (X_{s+1,1}, X_{s+1,2})'$ is the *control vector*. The control variables are supposed to compensate for the disturbances $z_{t,1}$, $z_{t,2}$ acting on the two process components at times $t = s+1, s+2, \ldots.$ The vector of deviations from target under the effect of the control variables is $\boldsymbol{d}_t = \boldsymbol{z}_t - \boldsymbol{X}_{s+1}$, if no adjustments are made at times $s+1, \ldots, t$. This model implies that a unit change on each control factor $X_{s+1,l}$ causes a unit change in the response, the deviations from target $d_{t,l}$. In other words, the "gain" matrix $\mathbf{G}$ in

$\boldsymbol{d}_t = \boldsymbol{z}_t - \mathbf{G}\boldsymbol{X}_{\mathbf{s+1}}$ equals the identity. There is no loss of generality with this, since if $\mathbf{G}$ is not the identity we simply use $\boldsymbol{X}_i = \mathbf{G}\boldsymbol{X}_i^{(0)}$ in what follows, where $\boldsymbol{X}_i^{(0)}$ is the vector of original control factors.

An *intervention* into the process (or an adjustment) at time s amounts to adjusting both control variables to values $X_{s+1,1}$, $X_{s+1,2}$. Interventions provoke costs due to factors like labour, material, process downtime and loss of production volume. Let $C > 0$ be the cost of an intervention. This is a fixed adjustment cost regardless of the magnitude of the adjustment made. To reduce intervention costs it is reasonable not to intervene permanently but only at selected intervention times s.

On the other hand, omitted adjustment leads to an increasing impact of the disturbances, and consequently to increasing deviations from target and increasing off target costs. In many cases the off target cost can be measured as a linear function of the square deviation from target. We assume costs $a_l > 0$ per unit of the square deviation from target in the lth production component, i.e., at times $t = s+1, s+2, ...$ the off target cost from component l is $a_l(z_{t,l} - X_{s+1,l})^2$.

In view of the off target cost it is reasonable to use the predicted amount of deviation from target as an intervention criterion. Let the last adjustment be made at time s with a resulting adjustment vector $\boldsymbol{X}_{s+1}$ at time $s+1$. At times $s+k$, $k = 1, 2, ...$ no intervention occurs as long as the vector of predicted deviations from target

$$\widehat{\boldsymbol{d}}_{s+k+1} \quad = \quad \hat{\boldsymbol{z}}_{s+k+1} - \boldsymbol{X}_{s+1}, \tag{4}$$

is inside a noncritical region $D \subset \mathbb{R}^2$ of the plane, which we will refer to as a *dead area*. At the first time $s+n$ with

$$\hat{\boldsymbol{d}}_{s+n+1} \notin D \quad \text{or} \quad n \geq n_0 \tag{5}$$

an alarm is given, and the control variables are adjusted so as to compensate the predicted disturbance at time $s+n+1$, i.e., $\boldsymbol{X}_{s+n+1} = \hat{\boldsymbol{z}}_{s+n+1}$. The upper limit n_0 for the length of periods without adjustment is prescribed for technical or security reasons, or it is a trivial upper limit, e.g., the lifetime of machinery or production equipment. In any case n_0 is a large upper limit, and alarms will generally result from the first condition in formula (5). The random time of the next intervention after the last recorded intervention time s is

$$N \quad = \quad N_D \quad = \quad \min\Big\{ \min\{n \in \mathbb{N} \,|\, \hat{\boldsymbol{d}}_{s+n+1} \notin D\Big\},\ n_0\Big\}. \tag{6}$$

Over the period $s+1, s+2, ..., s+N$ the control vectors remain constant at $\boldsymbol{X}_{s+1}$ and the vectors of the deviations from target are

$$\boldsymbol{z}_{s+1} - \boldsymbol{X}_{s+1},\ \boldsymbol{z}_{s+2} - \boldsymbol{X}_{s+1},\ \ldots,\ \boldsymbol{z}_{s+N} - \boldsymbol{X}_{s+1}. \tag{7}$$

The dead area $D \subset \mathbb{R}^2$ in the bivariate case corresponds to the univariate dead interval considered by Box and Jenkins (1963). Shifted along the time axis the dead interval induces a *dead band*. The appropriate shape of the dead area D in the bivariate case will be discussed in Section 5, below.

4 The Loss Function.

A good control policy has to establish a balance between the adjustment cost and the off target cost. Rare alarms reduce adjustment costs, but increase off target costs, and vice versa. From an economic point of view, the best policy is the one which minimizes the overall loss per time unit resulting from adjustments and from being off target. Under the assumptions of Section 3, a specific control policy is determined by the dead area $D \subset \mathbb{R}^2$. Hence we have to evaluate the loss incurred from running a process under the policy described in Section 3 as a function $L(D)$ of the dead area $D \subset \mathbb{R}^2$.

Consider a process run starting at time 0, controlled according to the policy described in Section 3. Adjustments are made at the end of periods 1,2,3,... of random length $N_1, N_2, N_3, ...$ at times $S_1 = N_1$, $S_2 = N_1 + N_2$, $S_3 = N_1 + N_2 + N_3$ and so on. For each time unit $S_k + 1, ..., S_k + N_{k+1}$ in a period between two successive adjustments at times S_k and $S_{k+1} = S_k + N_{k+1}$ the off target cost is evaluated by the quadratic cost function $a_l(z_{S_k+i,l} - X_{S_k+1,l})^2$. Hence the overall loss per time unit in the kth period is

$$V_k \quad = \quad \sum_{l=1}^{2} a_l \sum_{i=1}^{N_{k+1}} (z_{S_k+i,l} - X_{S_k+1,l})^2 \; + \; C. \tag{8}$$

For time t, let $K(t)$ be the number of periods elapsed until time t. Then the loss per time unit until time t is $V(t) = \frac{1}{t}\sum_{k=1}^{K(t)} V_k$. The process is assumed to run over a long time. Hence it is reasonable to evaluate the expected overall loss per time unit by the limit $\lim_{t\to\infty} E[V(t)]$. To calculate the latter quantity we observe that the pairs (N_k, V_k), $k = 1, 2, ...$ are serially independent and identically distributed, i.e., they constitute a *renewal reward process*, see the proof in appendix A. Hence an application of the *renewal reward theorem*, see *Ross* (1970), provides the limit $\lim_{t\to\infty} E[V(t)] = \frac{E[V]}{E[N]}$. The expected hitting time $E[N]$ is what Box and Luceño (1997) call the *average adjustment interval*, or AAI. Calculating $E[V]$ from equation (8) we obtain the following loss function

$$\begin{aligned} L(D) \quad &= \quad \frac{E[V]}{E[N]} \quad = \quad \frac{1}{E[N]} \sum_{l=1}^{2} a_l \sum_{i=1}^{N} E\Big[(z_{i,l} - X_{1,l})^2\Big] \; + \; \frac{C}{E[N]} \quad = \\ & \frac{1}{E[N]} \sum_{l=1}^{2} a_l \sum_{i=1}^{N} E\Big[(\hat{z}_{i,l} - X_{1,l})^2\Big] \; + \; \frac{C}{E[N]} \; + \; a_1\sigma_{1,\alpha}^2 \; + \; a_2\sigma_{2,\alpha}^2 \end{aligned} \tag{9}$$

as a function of the dead area $D \subset \mathbb{R}^2$. For the sake of convenience, in formula (9) and in subsequent calculations we use the first period starting at time 1 after adjustment at time

$s = 0$ to express the expectation $E[V]$. L depends on D through the time $N = N_D$ between successive interventions, where N_D is defined by formula (6).

5 The Standardized Loss Function.

Using the loss function (9), we might define the optimum control policy, i.e., the optimum dead area $D = D^\star$, as the one which minimizes $L(D)$ over $D \subset \mathbb{R}^2$. However, determining an optimal solution without restrictions on admissible shapes of the dead area $D \subset \mathbb{R}^2$ will be cumbersome. By considering an appropriately standardized version of the loss function $L(D)$ we get a more definite idea about reasonable shapes of D. This will lead to a concise restriction on D which is appropriate for determining specific optimum control policies.

The random variables $\hat{z}_{i,l} - X_{1,l}$, $l = 1, 2$, which are necessary for calculating the loss function $L(D)$ in formula (9), are the components of the vectors $\hat{\boldsymbol{z}}_i - \boldsymbol{X}_1$. From formulae (2), (3) and from $\boldsymbol{X}_1 = \hat{\boldsymbol{z}}_1$ we obtain

$$\hat{\boldsymbol{z}}_i - \boldsymbol{X}_1 \quad = \quad \sum_{j=1}^{i-1} \mathbf{L}\boldsymbol{\alpha}_{\mathbf{j}}. \tag{10}$$

The components of the random vectors $\boldsymbol{\beta}_j = \mathbf{L}\boldsymbol{\alpha}_j = (\beta_{j,1}, \beta_{j,2})'$ have the variance-covariance matrix

$$\begin{pmatrix} \sigma^2_{1,\beta} & \kappa_\beta \\ \kappa_\beta & \sigma^2_{2,\beta} \end{pmatrix} \quad = \quad \mathbf{C}_\beta \quad = \quad \mathbf{L}\mathbf{C}_\alpha\mathbf{L}' \tag{11}$$

and are serially uncorrelated. With respect to their cross-covariance $\mathbf{C}_\beta$, two cases have to be distinguished.

First, consider the case $\det \mathbf{C}_\beta = 0$. Then, a linear relation holds between $\beta_{j,1}$ and $\beta_{j,2}$ with probability 1, i.e., there exist reals c_1, c_2, c_3 such that $\mathrm{P}(c_1\beta_{j,1} + c_2\beta_{j,2} = c_3) = 1$, see *Schmetterer* (1974). In this case, we are dealing essentially with a single univariate problem which can be solved with the results of Box and Jenkins (1963).

In the sequel we assume $\det \mathbf{C}_\beta \neq 0$. Then each random vector $\boldsymbol{\beta}_j = \mathbf{L}\boldsymbol{\alpha}_j = (\beta_{j,1}, \beta_{j,2})'$ has a bivariate normal distribution with variance-covariance matrix $\mathbf{C}_\beta$, see *Schmetterer* (1974). The vectors $\boldsymbol{u}_j = (u_{j,1}, u_{j,2})'$ with $u_{j,l} = \frac{1}{\sigma_{l,\beta}}\beta_{j,l}$ constitute a bivariate cross correlated (but serially uncorrelated) Gaussian unit white noise, i.e., normally distributed pairs with stationary variance-covariance matrix

$$\mathbf{C}_u \quad = \quad \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

for each time t, where $\rho = \rho_\beta = \frac{\kappa_\beta}{\sigma_{1,\beta}\sigma_{2,\beta}}$, and without serial correlation, i.e., $\mathrm{Cov}[u_{t,l}, u_{s,m}] = 0$ for $s \neq t$, $l, m \in \{1, 2\}$. Note that assuming $\det \mathbf{C}_\beta \neq 0$ is equivalent to assuming $|\rho| < 1$.

Letting $\boldsymbol{U}_i = \boldsymbol{u}_1 + ... + \boldsymbol{u}_i$ we obtain from formula (10)

$$\hat{z}_{i,l} - X_{1,l} \quad = \quad \sigma_{l,\beta} U_{i-1,l} \qquad \text{for } \; i = 1, 2, ..., \;\; l = 1, 2. \tag{12}$$

Hence we can express the loss function $L(D)$ in the form

$$L(D) \quad = \quad a_1 \sigma_{1,\alpha}^2 \; + \; a_2 \sigma_{2,\alpha}^2 \; + \; a_1 \sigma_{1,\beta}^2 G_1(D) \; + \; a_2 \sigma_{2,\beta}^2 G_2(D) \; + \; \frac{C}{E[N]}, \tag{13}$$

where

$$G_l(D) \quad = \quad \frac{E\Big[\sum_{j=1}^{N_D} U_{j-1,l}^2\Big]}{E[N_D]} \qquad \text{for } \; l = 1, 2 \tag{14}$$

will be referred to as the *scaled mean square deviation* (or MSD). The predicted deviations from target $\hat{d}_{k+1,l} = \hat{z}_{k+1,l} - X_{1,l}$ (see equation 4), can be expressed as $\hat{d}_{k+1,l} = \sigma_{l,\beta} U_{k,l}$. Hence, equivalently to in (5), the time N of the next intervention is given by

$$N \quad = \quad N_{D'} \quad = \quad \min \Big\{ \min\{ n \in \mathbb{N} \,|\, \boldsymbol{U}_n \notin D' \Big\}, \; n_0 \Big\} \tag{15}$$

where

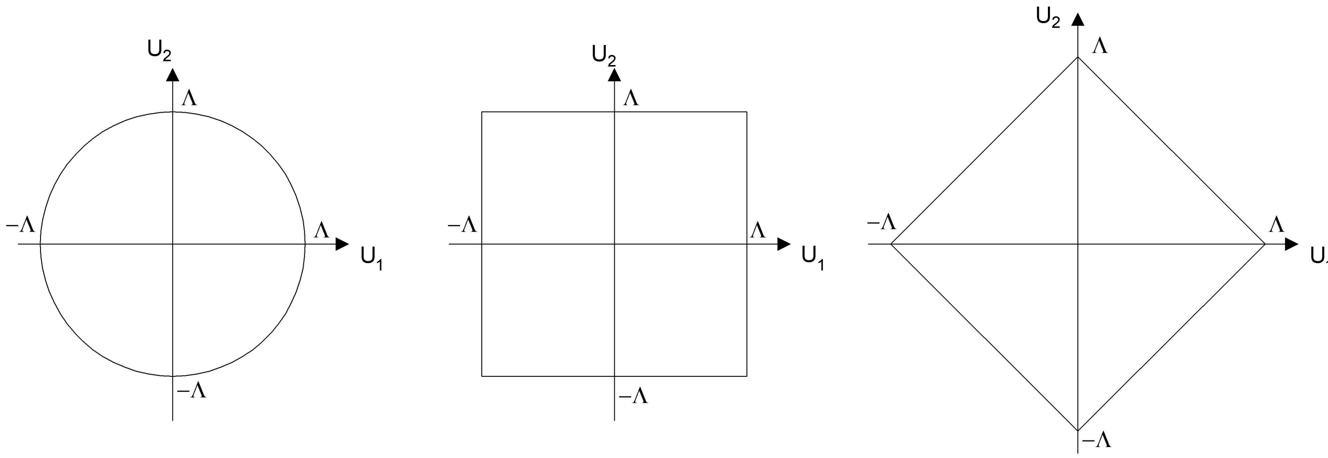
$$D' \quad = \quad \left\{ \left(\frac{x_1}{\sigma_{1,\beta}}, \frac{x_2}{\sigma_{2,\beta}} \right) \; | \; (x_1, x_2) \in D \right\}. \tag{16}$$

Now we are able to impose reasonable restrictions on the shape of the dead area D. Recall that an alarm signal $\boldsymbol{U}_n \notin D'$ entails adjustments in *both* compensating variables. Accordingly, neither of the two components should have a more prominent inclination to provoke an alarm. Hence, since the bivariate distribution of the vectors $\boldsymbol{U}_k = (U_{k,1}, U_{k,2})'$ is symmetric, symmetricity should also hold for the dead area D':

(DA) $D' \subset \mathbb{R}^2$ should be invariant under permutations $(x_1, x_2) \mapsto (x_2, x_1)$ of the coordinates.

To make the resulting control scheme practical for implementation in an industrial setting, a further reasonable requirement is that D' should be a convex area of a simple geometric nature on the plane. Three simple approaches to select D' are shown in figure 1: a circle, a square, and a rotated square. Each of these areas conforms to the above requirements, and each is a reasonable adaptation of the univariate dead interval considered by Box and Jenkins (1963) to the bivariate case. From an economic point of view the best choice among these three approaches is the one which guarantees a maximum dead area, i.e., a minimum of interventions, at a prescribed level $c = L(D) = L(D')$ of the loss function. We conjecture that in this sense the optimum shape is a circle. However, determining the values $L(D')$ for circles D' will be difficult from a mathematical point of view. To provide a practical solution for application in an industrial setting, we use a square shaped dead area D'. In the following Section 6 we shall see that a rotated square D'_Λ as on the right-hand side of Figure 1 is most convenient for calculations.

Figure 1: Some possible forms of “dead areas” on the plane.



6 The Standardized Dead Area and the Optimum Control Policy.

As the dead area with respect to the standardized predicted deviations from target $(U_{k,1}, U_{k,2})$, $k = 1, 2, ...$ we consider the interior

$$D'_\Lambda \quad = \quad \Big\{(U_1, U_2) \mid -\Lambda < U_1 + U_2 < \Lambda,\ -\Lambda < U_1 - U_2 < \Lambda\Big\} \tag{17}$$

of a rotated square with vertices $(0, \Lambda)$, $(\Lambda, 0)$, $(0, -\Lambda)$, $(-\Lambda, 0)$ as illustrated by the right-hand side of Figure 1. Hence the formula (15) for the time $N = N_\Lambda = N_{D'_\Lambda}$ of the next intervention amounts to

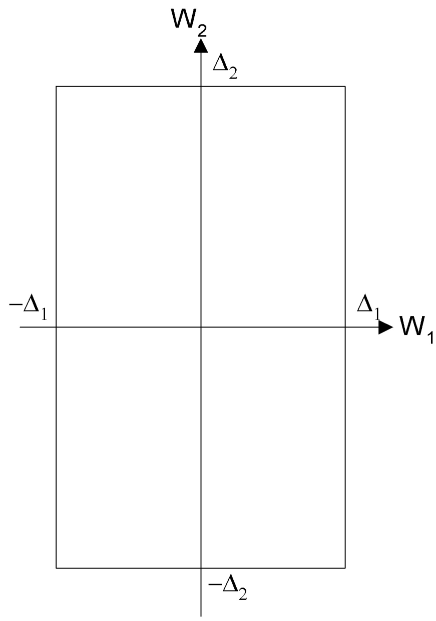
$$\begin{aligned} N \quad &= \quad N_\Lambda \quad = \\ &\min\Big\{ \min\{n \in \mathbb{N} \,|\, |W_{n,1}| \geq \frac{\Lambda}{\sqrt{2(1+\rho)}} \quad \text{or} \quad |W_{n,2}| \geq \frac{\Lambda}{\sqrt{2(1-\rho)}}\Big\},\ n_0 \Big\} \end{aligned} \tag{18}$$

where

$$W_{k,1} \quad = \quad \frac{U_{k,1} + U_{k,2}}{\sqrt{2(1+\rho)}} \quad = \quad w_{1,1} + ... + w_{k,1}, \quad w_{i,1} \quad = \quad \frac{u_{i,1} + u_{i,2}}{\sqrt{2(1+\rho)}}, \tag{19}$$

$$W_{k,2} \quad = \quad \frac{U_{k,1} - U_{k,2}}{\sqrt{2(1-\rho)}} \quad = \quad w_{1,2} + ... + w_{k,w}, \quad w_{i,2} \quad = \quad \frac{u_{i,1} - u_{i,2}}{\sqrt{2(1-\rho)}}, \tag{20}$$

Figure 2: Dead rectangle with respect to $(W_{k,1}, W_{k,2})$ on the plane.



It is easy to verify that the pairs $(w_{t,1}, w_{t,2})$ constitute a bivariate uncorrelated Gaussian unit white noise, i.e., normally distributed pairs with stationary variance-covariance matrix

$$\mathbf{C}_w = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

for each time t and without serial correlation, i.e., $\mathrm{Cov}[w_{t,l}, w_{s,l}] = 0$ for $s \neq t$, $l = 1, 2$, $\mathrm{Cov}[w_{t,1}, w_{s,2}] = 0$ for $s \neq t$. By formula (18), the time $N = N_\Lambda$ of the next intervention is expressed as the first exit time of the cross-independent bivariate random walk $(W_{k,1}, W_{k,2})$ from the open rectangle $(-\Delta_1; \Delta_1) \times (-\Delta_2; \Delta_2)$, where $\Delta_1 = \frac{\Lambda}{\sqrt{2(1+\rho)}}$, $\Delta_2 = \frac{\Lambda}{\sqrt{2(1-\rho)}}$, as illustrated by figure 2.

Hence we have two equivalent descriptions of the dead area:

- The dead area with respect to the standardized observations $(U_{k,1}, U_{k,2})$, $k = 1, 2, ...$, is the interior D'_Λ of a rotated rectangle as defined by formula (17).
- The dead area with respect to the uncorrelated pairs $(W_{k,1}, W_{k,2})$, $k = 1, 2, ...$, defined by formulae (19) and (20) is the interior

$$D''_\Lambda = \left\{ (y_1, y_2) \mid |y_1| < \frac{\Lambda}{\sqrt{2(1+\rho)}},\ |y_2| < \frac{\Lambda}{\sqrt{2(1-\rho)}} \right\} \tag{21}$$

of a rectangle parallel to the axes.

The loss function can be expressed as a function

$$L(\Lambda) \;=\; L(D'_\Lambda) \;=\; L(D''_\Lambda)$$

of the parameter Λ ranging over $(0;+\infty)$. Hence the optimum control policy can be defined by a value $\Lambda^\star$ which minimizes $L(\Lambda)$ for all $\Lambda > 0$. In analogy to the univariate case, the dead areas D'_Λ (rotated square centered in the origin) and D''_Λ (rectangle parallel to the axes, centered in the origin), when shifted along the time axis induce *dead bars.* From a practical point of view, displaying $U_{t,1}+U_{t,2}$ and $U_{t,1}-U_{t,2}$ on "adjustment" charts with limits at $\pm\Lambda$ is probably preferred, as we illustrate in Section 11. We first consider the moments needed to compute the standardized loss function.

7 Relations among Moments of $U_{N,l}$ and $W_{N,m}$.

To evaluate the loss function $L(D'_\Lambda) = L(\Lambda)$ we need the moments $E[U^2_{N,l}]$, $E[U^4_{N,l}]$ of the standardized accumulated deviations from target, which we use for this purpose in Section 9. By the choice of the dead area D_Λ, $U_{N,1}$ and $U_{N,2}$ have the same distribution. Hence

$$E[U^q_{N,1}] \;=\; E[U^q_{N,2}] \quad \text{for } \; q \in \mathbb{N}_0. \tag{22}$$

From the symmetry of the underlying bivariate normal distribution and from the symmetry of the dead area it is clear that

$$E[U^q_{N,1}] \;=\; 0 \;=\; E[U^q_{N,2}] \quad \text{for } \; q = 1,3,5,.... \tag{23}$$

Because of the correlation among the variables $U_{k,1}$, $U_{k,2}$, direct calculation of the moments $E[U^q_{N,l}]$, $q = 2,4,...$, is rather involved. It is more convenient to calculate the moments $E[W^q_{N,m}]$, see Section 8, and then to derive $E[U^2_{N,l}]$ and $E[U^4_{N,l}]$. For this purpose, we establish relations among the moments of $U_{N,l}$ and the moments of $W_{N,m}$. From formulae (19) and (20) we obtain

$$E[U^2_{N,l}] \;=\; \frac{1}{2}\Big\{(1+\rho)E[W^2_{N,1}] \;+\; (1-\rho)E[W^2_{N,2}]\Big\}, \tag{24}$$

$$E[U^4_{N,l}] \;+\; 3E[U^2_{N,1}U^2_{N,2}] \;=\; (1+\rho)^2E[W^4_{N,1}] \;+\; (1-\rho)^2E[W^4_{N,2}], \tag{25}$$

$$E[U^4_{N,l}] \;-\; E[U^2_{N,1}U^2_{N,2}] \;=\; 2(1-\rho^2)E[W^2_{N,1}W^2_{N,2}]. \tag{26}$$

Combining equations (25) and (26) we obtain

$$E[U^4_{N,l}] \;=\; \frac{1}{4}\Big\{(1+\rho)^2E[W^4_{N,1}] \;+\; 6(1-\rho^2)E[W^2_{N,1}W^2_{N,2}] \;+\; (1-\rho)^2E[W^4_{N,2}]\Big\}. \tag{27}$$

8 Moments of $W_{N,m}$.

Because of the independence of the variables $W_{k,1}$, $W_{k,2}$ we can adapt a method used by Box and Jenkins (1963) for the univariate case to determine an approximation of the moments $E[W_{N,m}^q]$ and of $E[W_{N,1}^p W_{N,2}^r]$. In this derivation, we ignore the upper limit n_0 for the length of periods without adjustment. See the explanation on n_0 in Section 3.

From the symmetry of the dead area and of the underlying normal distribution it is clear that

$$E[W_{N,1}^q] \;=\; 0 \;=\; E[W_{N,2}^q] \quad \text{for } \; q = 1, 3, 5, ... \tag{28}$$

As in Section 6 we use the abbreviating notation $\Delta_1 = \frac{\Lambda}{\sqrt{2(1+\rho)}}$, $\Delta_2 = \frac{\Lambda}{\sqrt{2(1-\rho)}}$. For $k \in \mathbb{N}$, let $h_k\colon \mathbb{R} \to \mathbb{R}$ be the joint density of $W_{1,m}, ..., W_{k,m}$, and let $g_{k,m}\colon \mathbb{R} \to \mathbb{R}$ be defined by

$$g_{k,m}(y) \;=\; \int\limits_{|y_1|<\Delta_m,...,|y_{k-1}|<\Delta_m} h_k(y_1, ..., y_{k-1}, y)\, \mathrm{d}y_1 ... \mathrm{d}y_{k-1}. \tag{29}$$

Since $W_{k,m} = w_{k,m} + W_{k-1,m}$ with $w_{k,m}$ distributed according to $N(0,1)$, the functions $g_{k,m}$ follow the recursion

$$g_{k,m}(x_m) \;=\; \int_{|y_m|<\Delta_m} g_{k-1,1}(y_m)\varphi(x_m - y_m)\, \mathrm{d}y_m, \tag{30}$$

where $\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right)$ is the density function of the normal distribution $N(0,1)$. $g_{k,m}$ is a density of $W_{k,m}$ in the event $\{|W_{1,m}| < \Delta_m, ..., |W_{k-1,m}| < \Delta_m\}$, i.e.,

$$\mathrm{P}(W_{k,m} \in B,\ |W_{1,m}| < \Delta_m, ..., |W_{k-1,m}| < \Delta_m) \;=\; \int_B g_{k,m}(x_m)\, \mathrm{d}x_m \tag{31}$$

for Borel sets B. From formula (31) it follows that

$$\frac{g_{n-1,1} \cdot g_{n-1,2} \cdot \mathbb{1}_{(-\Delta_1;\Delta_1)\times(-\Delta_2;\Delta_2)}}{\mathrm{P}(N \geq n)}$$

is a joint conditional density of $W_{n-1,1}$ and $W_{n-1,2}$ under the condition $N \geq n$, where $\mathbb{1}_B$ is the indicator function of a set B. We follow the intuitively reasonable approach used by Box and Jenkins (1963) for the univariate case: We approximate the joint conditional distribution of $W_{n-1,1}$ and $W_{n-1,2}$ under the condition $N \geq n$ by a bivariate uniform distribution over the dead rectangle $(-\Delta_1; \Delta_1) \times (-\Delta_2; \Delta_2)$, i.e., we assume

$$\frac{g_{n-1,1} \cdot g_{n-1,2} \cdot \mathbb{1}_{(-\Delta_1;\Delta_1)\times(-\Delta_2;\Delta_2)}}{\mathrm{P}(N \geq n)} \;\approx\; \frac{1}{4\,\Delta_1\,\Delta_2}. \tag{32}$$

The accuracy of this approximation is studied in Section 10.

Under this approximation we obtain for $p, r \in \mathbb{N}_0$

$$
\begin{aligned}
& E[W_{N,1}^p W_{N,2}^r | N = n]\, \mathrm{P}(N = n) \quad =_{(18),\,(31)} \\
& \int\limits_{\substack{\{|x_1|\geq\Delta_1\}\cup\\ \{|x_2|\geq\Delta_2\}}} x_1^p x_2^r g_{n,1}(x_1) g_{n,2}(x_2)\, \mathrm{d}x_1 \mathrm{d}x_2 \quad =_{(30)} \\
& \int\limits_{\substack{\{|x_1|\geq\Delta_1\}\cup\\ \{|x_2|\geq\Delta_2\}}} x_1^p x_2^r \int\limits_{|y_1|<\Delta_1} \int\limits_{|y_2|<\Delta_2} g_{n-1,1}(y_1) g_{n-1,2}(y_2) \varphi(x_1 - y_1) \varphi(x_2 - y_2)\, \mathrm{d}y_2 \mathrm{d}y_1\, \mathrm{d}x_1 \mathrm{d}x_2 \quad \approx_{(32)} \\
& \frac{\mathrm{P}(N \geq n)}{4\ \Delta_1\ \Delta_2} \Bigg\{ \int\limits_{|x_1|\geq\Delta_1} x_1^p \int_{-\Delta_1 - x_1}^{\Delta_1 - x_1} \varphi(y_1) \mathrm{d}y_1\, \mathrm{d}x_1 \ \cdot \int\limits_{x_2\in\mathbb{R}} x_2^r \int_{-\Delta_2 - x_2}^{\Delta_2 - x_2} \varphi(y_2) \mathrm{d}y_2\, \mathrm{d}x_2 \quad + \\
& \int\limits_{|x_1|<\Delta_1} x_1^p \int_{-\Delta_1 - x_1}^{\Delta_1 - x_1} \varphi(y_1) \mathrm{d}y_1\, \mathrm{d}x_1 \ \cdot \int\limits_{|x_2|\geq\Delta_2} x_2^r \int_{-\Delta_2 - x_2}^{\Delta_2 - x_2} \varphi(y_2) \mathrm{d}y_2\, \mathrm{d}x_2 \Bigg\} \quad = \\
& \frac{\mathrm{P}(N \geq n)}{4\ \Delta_1\ \Delta_2} \Big\{ I(\Delta_1, p) \Big[I(\Delta_2, r) + J(\Delta_2, r) \Big] \ + \ J(\Delta_1, p) I(\Delta_2, r) \Big\},
\end{aligned}
$$

where for $\Delta \geq 0$, $q \in \mathbb{N}_0$

$$
I(\Delta, q) \;=\; \int_{|x|\geq\Delta} x^q \Big[\phi(x + \Delta) - \phi(x - \Delta) \Big]\, \mathrm{d}x, \tag{33}
$$

$$
J(\Delta, q) \;=\; \int_{|x|<\Delta} x^q \Big[\phi(x + \Delta) - \phi(x - \Delta) \Big]\, \mathrm{d}x. \tag{34}
$$

The integrals $I(\Delta, q)$ and $J(\Delta, q)$ are evaluated by proposition B.1 in the appendix B. In particular,

$$
I(\Delta, 0) \;=\; 2\Big\{ \varphi(0) - \varphi(2\Delta) + 2\Delta\Big[1 - \phi(2\Delta)\Big] \Big\}, \quad J(\Delta, 0) \;=\; 2\Delta - I(\Delta, 0). \tag{35}
$$

Letting $p = 0 = r$ into the above derivation we obtain from (35)

$$
\mathrm{P}(N = n) \quad \approx \quad \frac{\mathrm{P}(N \geq n)}{4\ \Delta_1\ \Delta_2} \Big\{ 2\Delta_2 I(\Delta_1, 0) \ + \ \Big[2\Delta_1 - I(\Delta_1, 0) \Big] I(\Delta_2, 0) \Big\}.
$$

Finally for $p, r \in \mathbb{N}_0$

$$E[W^p_{N,1} W^r_{N,2}] \quad = \quad \sum_{n=1}^{\infty} E[W^p_{N,1} W^r_{N,2} | N = n] \ \mathrm{P}(N = n) \quad \approx$$
$$\frac{I(\Delta_1, p)\Big[I(\Delta_2, r) + J(\Delta_2, r)\Big] \ + \ J(\Delta_1, p) I(\Delta_2, r)}{2\Delta_2 I(\Delta_1, 0) \ + \ \Big[2\Delta_1 - I(\Delta_1, 0)\Big] I(\Delta_2, 0)} \ . \tag{36}$$

In particular for $q \in \mathbb{N}_0$

$$E[W^q_{N,1}] \quad \approx \quad \frac{I(\Delta_1, q) 2\Delta_2 \ + \ J(\Delta_1, q) I(\Delta_2, 0)}{2\Delta_2 I(\Delta_1, 0) \ + \ \Big[2\Delta_1 - I(\Delta_1, 0)\Big] I(\Delta_2, 0)} \ . \tag{37}$$

9 An Approximation of the Loss Function $L(\Lambda)$.

The standardized accumulated deviations from target $\sum_{j=1}^{N_D} U^2_{j-1,1}$ and $\sum_{j=1}^{N_D} U^2_{j-1,2}$ have the same distribution. Hence from formula (14) $G_1(D'_\Lambda) = G_2(D'_\Lambda)$. Consider the martingales $(R_{n,l})_{n \in \mathbb{N}}$, $(Y_{n,l})_{n \in \mathbb{N}}$, $(Z_{n,l})_{n \in \mathbb{N}}$ defined by formulae (52) and (53) in Appendix C. Obviously, N_D is a stopping time for these martingales, uniformly bounded by $N_D \le n_0$. Hence the optional stopping theorem (see Rogers and Williams (1994) provides

$$E[N] \ = \ E[U^2_{N,l}], \qquad E\Big[\sum_{j=1}^{N_D} U^2_{j-1,l}\Big] \ = \ \frac{E[U^4_{N,l}]}{6} - \frac{E[N]}{2}. \tag{38}$$

We point out that these expressions are exact and not approximations, as suggested in Box and Jenkins (1963).

Inserting into formula (14) we obtain the scaled MSD:

$$G_1(D'_\Lambda) \quad = \quad G_2(D'_\Lambda) \quad = \quad \frac{E[U^4_{N,l}]}{6E[U^2_{N,l}]} - \frac{1}{2}. \tag{39}$$

Hence, from formula (13) we have that

$$L(\Lambda) \quad = \quad L(D'_\Lambda) \quad =$$
$$a_1 \sigma^2_{1,\alpha} \ + \ a_2 \sigma^2_{2,\alpha} \ + \ (a_1 \sigma^2_{1,\beta} + a_2 \sigma^2_{2,\beta}) \left(\frac{E[U^4_{N,l}]}{6E[U^2_{N,l}]} - \frac{1}{2} \right) \ + \ \frac{C}{E[U^2_{N,l}]} \ . \tag{40}$$

To obtain an approximation of the loss function $L(\Lambda) = L(D'_\Lambda)$, we insert into formula (40) the approximations for the moments $E[U^2_{N,l}]$ and $E[U^4_{N,l}]$ determined from formulae (24), (27), (36), (37).

10 Accuracy of the approximations

The expressions for the moments (24) and (27) are based on the approximation (32). Similarly as what Box and Jenkins (1963) reported for the univariate case, the assumption of a uniform distribution for the standardized bivariate process $\boldsymbol{W}_{n-1}$ before the process falls out of the dead area was found to be inaccurate, particularly for large values of Λ. The geometrical reason for this problem is that, for large Λ, the points $(W_{n-1,1}, W_{n-1,2})$ will gather closer to the boundaries of the dead area than to the center of the region. Therefore, a correction regression equation was developed empirically by computing, through simulation, the "real" moments $E_r[U_{N,l}^2]$ and $E_r[U_{N,l}^4]$ and computing the differences $D_2 = E[U_{N,l}^2] - E_r[U_{N,l}^2]$ and $D_4 = E[U_{N,l}^4] - E_r[U_{N,l}^4]$. Here, $E_r[U_{N,l}^2]$ and $E_r[U_{N,l}^4]$ were estimated by simulating 50,000 renewals for $|\rho| \in \{0.1, 0.25, 0.5, 0.75, 0.85, 0.95\}$ and $\Lambda \in \{1, 2, ..., 15\}$. The moments $E[U_{N,l}^2]$ and $E[U_{N,l}^4]$ were computed as in (24) and (27). Note from (18-20) that the moments are invariant with respect to the sign of the cross-correlation coefficient ρ.

The following correction model was fitted to the errors in the second order moment data:

$$D_2^{0.337} = 0.385 + 0.133\Lambda - 0.840|\rho| - 0.00172\Lambda^2 + 0.90\rho^2 + 0.0375\Lambda|\rho| \tag{41}$$

that is, a quadratic polynomial model in Λ and ρ was fitted after a Box-Cox power transformation was applied to the data (hence the exponent in the left hand side). This model was fitted for $\Lambda > 2$ since for small values of Λ the analytic formula provides a good approximation to the real moment. Fortunately, model (41) provides an excellent fit, with $R^2 = 0.997$ and the p-values associated with the tests for the significance of each regressor equal to zero up to three decimal places in all cases.

For the errors in the fourth order moments, the corresponding fitted model was:

$$D_4^{0.224} = 0.515 + 0.386\Lambda - 1.25|\rho| + 0.00118\Lambda^2 + 1.44\rho^2 + 0.0768\Lambda|\rho| \tag{42}$$

where similarly as before, a full quadratic polynomial in Λ and ρ was fitted after a Box-Cox transformation was applied to the errors. Values $\Lambda \leq 2$ were excluded from the regression, similarly as before. The fit again is excellent, giving $R^2 = 0.999$ and all p-values of the individual tests of significance for each model parameter smaller or equal to 0.001.

In order to minimize the standardized cost function, the correction formulae $(41 - 42)$ were used for $\Lambda \geq 2$. For $\Lambda < 2$, no correction was used and the analytical formulae (24), (27) were directly utilized instead.

11 Minimization of the standardized loss function

From (40), it is evident that the optimal solution Λ^* depends on the relative cost parameter

$$C' = \frac{C}{a_1\sigma_{1,\beta}^2 + a_2\sigma_{2,\beta}^2}.$$

The only other parameter that the optimal solution depends on is the value of $|\rho|$, the cross-correlation of the bivariate series $\boldsymbol{\beta}_j = \mathbf{L}\boldsymbol{\alpha}_\mathbf{j}$. To find Λ^*, the cost function

$$L'(\Lambda) = \frac{L(\Lambda)}{a_1\sigma^2_{1,\beta} + a_2\sigma^2_{2,\beta}} - \frac{a_1\sigma^2_{1,\alpha} + a_2\sigma^2_{2,\alpha}}{a_1\sigma^2_{1,\beta} + a_2\sigma^2_{2,\beta}} = \frac{E[U^4_{N,l}]}{6E[U^2_{N,l}]} - \frac{1}{2} + \frac{C'}{E[U^2_{N,l}]}$$

was minimized using Matlab's `fminbnd` function, which minimizes a non-linear function subject to bounds (bounds of 0.1 and 20 were used in all cases in the table below). For $\Lambda > 2$, the two moments were corrected using (41-42). The solutions reported in this section were confirmed to provide the unique minimizer of the function within the interval (the Matlab code used in this section is available from the first author upon request).

Table 1 shows the optimal solution Λ^*, the corresponding value of the loss function $L'(\Lambda^*)$, the scaled MSD value ($G_l(D'_{\Lambda^*})$), and the Average Adjustment Interval (AAI= $E[N]$) for a variety of values of $|\rho|$ and C'. From it, a potential user can select a solution by finding acceptable MSD and AAI values, without having to define an explicit cost C'. In general terms, the optimal limit Λ^* increases with increasing relative fixed adjustment cost (C') and with increasing correlation ($|\rho|$). The cost function was observed to be fairly flat around the minimum point, so small departures of Λ from the optimum value Λ^* will not be of practical importance.

Relation with Box and Jenkins' univariate optimal solution

Clearly, our formulation reduces to solving two separate univariate problems using Box and Jenkins (1963) formulation, one for each response and each controllable factor l, when $\kappa_\alpha = 0$ (which implies $\rho = 0$) and both $\mathbf{\Theta}$ and $\mathbf{G}$ are diagonal matrices. In such case the two responses are said to be *decoupled*. To see further relations between the bivariate and the univariate models, we could try to solve a single univariate problem with the procedure in this paper. Suppose we want to solve for the best Box-Jenkins (1963) univariate dead band rule when the white noise is σ^2_α, the IMA(1,1) parameter is θ, the off target cost is a and the adjustment cost C. Then we would set in our code $a_1 = a_2 = a$ and $\sigma^2_{1,\beta} = \sigma^2_{2,\beta} = (1-\theta)\sigma^2_\alpha$, apart from setting $\rho = 0$. The solution thus obtained from minimizing $L'(\Lambda)$ will be related to the optimal solution found by Box and Jenkins, Λ^{BJ}, by the relation $\Lambda^{BJ} = \Lambda^*/\sqrt{2}$. The reason of this is the rotated nature of our dead area (Figure 1): Λ^{BJ} is the width of the square but we are solving for Λ^*, half the length of the diagonal. We now illustrate the bivariate procedure with a practical example.

Table 1. Some optimal solutions.					
C'	$\|\rho\|$	Λ^*	$L'(\Lambda^*)$	Scaled MSD	AAI
1	0.0	1.01	0.63	0.10	1.91
1	0.3	1.02	0.63	0.11	1.91
1	0.6	1.21	0.63	0.16	2.11
1	0.9	1.39	0.54	0.22	2.69
4	0.0	2.85	1.66	0.69	4.10
4	0.3	2.73	1.69	0.71	4.06
4	0.6	2.73	1.65	0.68	4.12
4	0.9	2.99	1.44	0.56	4.57
7	0.0	3.40	2.30	0.99	5.35
7	0.3	3.29	2.33	1.02	5.33
7	0.6	3.33	2.28	1.00	5.47
7	0.9	3.61	2.00	0.84	6.01
10	0.0	3.79	2.81	1.24	6.35
10	0.3	3.69	2.85	1.27	6.35
10	0.6	3.74	2.78	1.24	6.51
10	0.9	4.04	2.46	1.06	7.14
20	0.0	4.66	4.12	1.87	8.88
20	0.3	4.58	4.15	1.92	8.93
20	0.6	4.64	4.05	1.87	9.17
20	0.9	5.00	3.63	1.63	10.02
50	0.0	6.06	6.76	3.16	13.91
50	0.3	6.00	6.76	3.21	14.06
50	0.6	6.11	6.59	3.14	14.47
50	0.9	6.56	5.95	2.78	15.75
80	0.0	6.91	8.66	4.10	17.54
80	0.3	6.86	8.65	4.14	17.76
80	0.6	7.01	8.42	4.05	18.30
80	0.9	7.52	7.64	3.62	19.90
100	0.0	7.35	9.74	4.63	19.58
100	0.3	7.31	9.71	4.67	19.85
100	0.6	7.47	9.45	4.57	20.47
100	0.9	8.02	8.59	4.09	22.24
400	0.0	10.71	19.98	9.73	39.04
400	0.3	10.74	19.79	9.72	39.74
400	0.6	11.04	19.21	9.48	41.13
400	0.9	11.89	17.56	8.62	44.75
700	0.0	12.44	26.59	13.06	51.71
700	0.3	12.52	26.28	13.00	52.70
700	0.6	12.89	25.47	12.66	54.64
700	0.9	13.90	23.31	11.56	59.57
1000	0.0	13.67	31.87	15.72	61.91
1000	0.3	13.79	31.46	15.62	63.15
1000	0.6	14.22	30.47	15.21	65.55
1000	0.9	15.36	27.89	13.92	71.58

Example.- As a practical application of the adjustment method developed and the optimal solutions obtained, consider a chemical mechanical planarization (CMP) process which is of critical importance in the manufacture of semiconductors. This is a polishing process in which there are typically two responses of interest (see, e.g., Moyne et al, 2000): the removal rate of silicon oxide (hereafter, $z_{t,1}$) which we suppose here to have a target equal to 2700, and the non-uniformity of the wafer (hereafter, $z_{t,2}$) with target equal to 500. Two controllable factors, down force ($X_{t,1}^{(0)}$) and table speed ($X_{t,2}^{(0)}$) can be adjusted to provide better control to target. The factors are in coded units. Here, the time index t denotes the wafer number, assuming a single-wafer CMP machine is in use. To illustrate the methodology, we simulate this process from a somewhat modified model obtained from real experiments as reported in Del Castillo and Yeh (1998). Simulating the behavior of the process will allow us to see what would have occurred in the absence of any adjustments.

The model that is simulated for this illustration has a gain matrix equal to

$$\mathbf{G} = \begin{pmatrix} 547.6 & 616.3 \\ -62.3 & -128.6 \end{pmatrix}$$

and an IMA parameter matrix equal to

$$\mathbf{\Theta} = \begin{pmatrix} 0.4 & 0.1 \\ 0.3 & 0.5 \end{pmatrix}$$

The covariance matrix of the bivariate normal white noise sequence is

$$\mathbf{C}_\alpha = \begin{pmatrix} 3600 & -1500 \\ -1500 & 900 \end{pmatrix}$$

In practice, estimates of the previous parameters could be obtained using multivariate Time Series techniques (Reinsel, 1994).

From the aforementioned data, we have that

$$\mathbf{C}_\beta = (\mathbf{I} - \mathbf{\Theta})\mathbf{C}_\alpha(\mathbf{I} - \mathbf{\Theta})' = \begin{pmatrix} 1501 & -1268 \\ -1268 & 1399 \end{pmatrix}$$

thus $\rho = \rho_\beta = -0.8750$. Let us assume it costs $a_1 = 0.1$ dollars to have a removal rate that deviates one unit (Amstrongs per time unit, in this case) from the desired target of 2700 during the processing of one wafer. Similarly, assume it costs $a_2 = 0.1$ dollars to have a wafer with a non-uniformity that deviates one unit (Amstrongs, in this case) from its desired target of 500. Assume the cost of making an adjustment in the "recipe" used in processing each wafer equals $C = 1000$ dollars, and includes the cost of re-starting the machine (sometimes test wafers are introduced after adjustments), machine downtime, and operator time. With the given cost structure and process information, we have that $C' = C/(a_1\sigma_{1,\beta}^2 + a_2\sigma_{2,\beta}^2) = 3.448$. Minimizing $L'(\Lambda)$ with respect to Λ we obtain the optimal limit $\Lambda^* = 2.78$ with loss $L'(2.78) = 1.3451$

and AAI= $E[N] = 4.1589$ (wafers between adjustments), or approximately 24 adjustment will be made on average every 100 wafers are produced.

The resulting process adjustment procedure is as follows. A vector EWMA with parameter matrix $\mathbf{L} = \mathbf{I} - \mathbf{\Theta}$ provides one step ahead forecasts $\widehat{z}_{k+1,1}, \widehat{z}_{k+1,2}$ based on the measurements of the two responses. At each time instant k, the standardized bivariate series $\boldsymbol{U}_k$ is computed as

$$\boldsymbol{U}_k = \begin{pmatrix} U_{k,1} \\ U_{k,2} \end{pmatrix} = \begin{pmatrix} \frac{\widehat{z}_{k+1,1} - X_{s,1}}{\sigma_{1,\beta}} \\ \frac{\widehat{z}_{k+1,2} - X_{s,2}}{\sigma_{2,\beta}} \end{pmatrix}$$

where $s < k$ is the last period an adjustment was made and where we use $\boldsymbol{X}_k = \mathbf{G}\boldsymbol{X}_k^{(0)}$ with $\boldsymbol{X}_k^{(0)}$ a vector containing the down force and table speed controllable factors as elements. Whenever $|U_{k,1} + U_{k,2}| > 2.78 = \Lambda^*$ or $|U_{k,1} - U_{k,2}| > 2.78$, the controllable factors are changed such that $\boldsymbol{X}_{k+1} = \widehat{\boldsymbol{z}}_{k+1}$, or, in terms of the original controllable factors, the new settings are $\boldsymbol{X}_{k+1}^{(0)} = \mathbf{G}^{-1}\widehat{\boldsymbol{z}}_{k+1}$.

Figure 3 shows the uncontrolled and controlled processes. Figure 4 shows the standardized quantities $U_{k,1}+U_{k,2}$ and $U_{k,1}-U_{k,2}$ on a "adjustment chart" with limits at $\pm\Lambda^* = \pm 2.78$. Finally, Figure 5 shows the corresponding values of the controllable factors. Horizontal segments imply no adjustments are made during such periods. In the particular simulation depicted, 30 adjustments were made. As it can be seen for the simulated data shown, the down force is reduced throughout the control session while the table speed was increased during the last few runs.

References.

Åström, K.J., 1970. *Introduction to Stochastic Control Theory.* Academic Press, San Diego, CA.

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

Box G.E.P., G.M. Jenkins, and Reinsel, G. (1994). *Time Series Analysis, Forecasting, and Control* 3rd. ed., Englewood Cliffs: Prentice Hall (1994).

Box, G.E.P., and Kramer, T., 1992. "Statistical Process Monitoring and Feedback Adjustment– a Discussion", *Technometrics*, 34, 3, pp. 251-267.

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

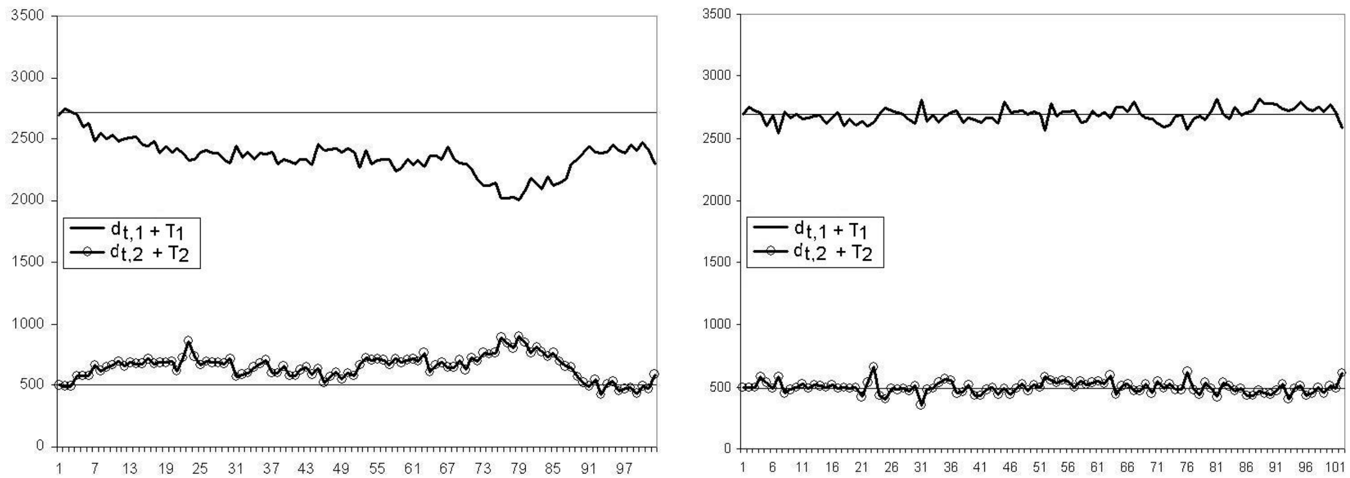


Figure 3: Realizations of the uncontrolled (left) vs. controlled (right) quality characteristics $(d_{t,l} + T_l, l = 1, 2)$ for the semiconductor example. Targets T_l equal 2700 and 500 units for $l = 1, 2$, respectively.

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

Del Castillo, E., and Yeh, J.Y., 1998. "An Adaptive Run-to-Run Optimizing Controller for Linear and Nonlinear Semiconductor Processes" *IEEE Transactions on Semiconductor Manufacturing*, 11, 2, pp. 285-295.

Jensen, K.L., and Vardeman, S.B., (1993). "Optimal Adjustment in the Presence of Deterministic Process Drift and Random Adjustment Error". *Technometrics*, 35, pp. 376-389.

Moyne, J., Del Castillo, E., and Hurwitz, A., eds. (2000). *Run to run control in semiconductor manufacturing*, CRC press, Boca Raton, FL.

Rogers, L.C.G., and Williams, D. (1994) *Diffusions, Markov Processes, and Martingales.* 2nd Edition. John Wiley & Sons, Chichester, New York.

Ross, S.M., 1970. *Applied Probability Models with Optimization Applications.* Holden-Day, San Francisco, Cambridge, London, Amsterdam.

Schmetterer, L., 1974. *Introduction to Mathematical Statistics.* Springer-Verlag, Berlin.

Srivastava, M. S., and Wu, Y. (1991). "A Second Order Approximation to Taguchi's On-

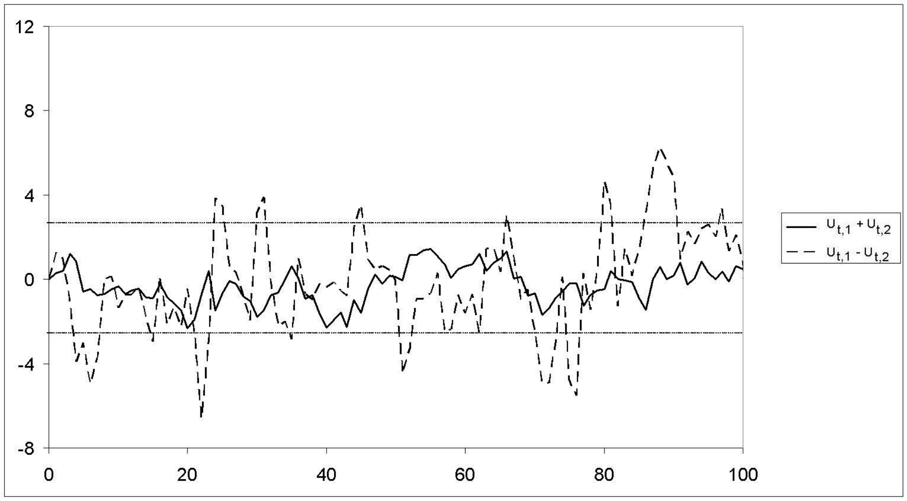


Figure 4: Standardized series $U_{t,1} + U_{t,2}$ and $U_{t,1} - U_{t,2}$ used to determine the time of the adjustments in the example. Adjustment limit $\Lambda^* = 2.78$.

line Control Procedure". *Communications in Statistics – Theory and Methods*, 20, 7, pp. 2149-2168.

A The Renewal Reward Process Property.

Consider an adjustment at time s. From the EWMA recursion (2) for the one step ahead predictors and from the adjustment formula $X_{s+1,l} = \hat{z}_{s+1,l}$ we obtain by induction for $k = 1, 2, ..$

$$\widehat{\boldsymbol{d}}_{s+k} \quad =_{(4)} \quad \hat{\boldsymbol{z}}_{s+k} - \hat{\boldsymbol{z}}_{s+1} \quad = \quad \mathbf{L} \sum_{\mathbf{j=1}}^{\mathbf{k-1}} \boldsymbol{\alpha}_{\mathbf{s+j}}. \tag{43}$$

From equation (43) and from the alarm rule (5) it is obvious that the lengths $N_1, N_2, N_3, ...$ of periods between adjustments are independent and identically distributed. From equation (1) we obtain for $i = 1, 2, ...$

$$z_{s+i,l} \quad = \quad z_{s+1,l} \; + \; \sum_{m=1}^{i-1} (\alpha_{s+m+1,l} - \theta_l \alpha_{s+m,l})$$

and hence

$$z_{s+i,l} - \hat{z}_{s+1,l} \quad =_{(3)} \quad \alpha_{s+1,l} \; + \; \sum_{m=1}^{i-1} (\alpha_{s+m+1,l} - \theta_l \alpha_{s+m,l}). \tag{44}$$

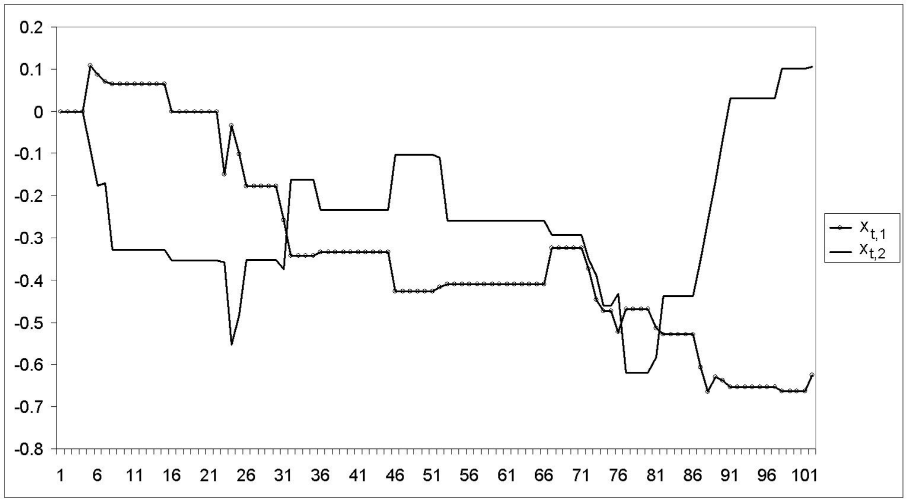


Figure 5: Levels for the controllable factors $X_{t,1}$ and $X_{t,2}$ for the example. Thirty adjustments were made.

From the assumptions on the white noise variables $\alpha_{t,l}$ and from equation (44) it follows that for $s_1 < s_2$, $1 \leq i \leq s_2 - s_1$, $j \geq 1$, the differences $z_{s_1+i,l} - \hat{z}_{s+1,l}$ and $z_{s_2+j,l} - \hat{z}_{s_2+1,l}$ are independent and normally distributed. Taking into account that the adjustment formula is $X_{s+1,l} = \hat{z}_{s+1,l}$, we can demonstrate that the vectors $(z_{S_k+1,l} - X_{S_k+1,l}, ..., z_{S_k+N_{k+1},l} - X_{S_k+1,l})$ of the deviations from target, indexed in $k \in \mathbb{N}$, are independent. Hence by definition (8), the overall losses V_k, $k = 1, 2, ...$, in the periods between interventions are independent. Since the lengths $N_1, N_2, N_3, ...$ of periods between adjustments are identically distributed, the losses V_k, $k = 1, 2, ...$, are also identically distributed.

Hence the pairs (N_k, V_k), $k = 1, 2, ...$ are serially independent and identically distributed, i.e., they constitute a *renewal reward process*, see Ross (1970).

B Integrals of the Normal Distribution Function.

The incomplete gamma integral is defined by

$$\Gamma(y, a) \;=\; \int_a^{+\infty} u^{y-1} \exp(-u)\, \mathrm{d}u \quad \text{for} \;\; a \in [0; +\infty). \tag{45}$$

For $a = 0$ we obtain the customary gamma function $\Gamma(y) = \Gamma(y, 0)$. Integrals of the normal probability density function $\varphi(y) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-y^2}{2}\right)$ can be expressed by means of the incomplete gamma integral:

$$\int_b^{+\infty} x^j \varphi(x)\,\mathrm{d}x \quad = \quad \frac{2^{\frac{j}{2}-1}}{\sqrt{\pi}}\,\Gamma\left(\frac{j+1}{2}, \frac{b^2}{2}\right) \qquad \text{for } \ b \in [0;+\infty),\ j \in \mathbb{N}_0. \tag{46}$$

To prove formula (46), substitute $u = \frac{x^2}{2}$. Integration by parts provides the formula

$$\int_A^{+\infty} y^j \Big[1-\phi(y)\Big]\,\mathrm{d}y \quad = \quad \frac{-A^{j+1}}{j+1}\Big[1-\phi(A)\Big] \ + \ \frac{1}{j+1}\int_A^{+\infty} y^{j+1}\varphi(y)\,\mathrm{d}y \tag{47}$$

for $j \in \mathbb{N}_0$, $A \geq 0$, and in particular

$$\int_0^{+\infty} y^j \Big[1-\phi(y)\Big]\,\mathrm{d}y \quad = \quad \frac{1}{j+1}\int_0^{+\infty} y^{j+1}\varphi(y)\,\mathrm{d}y \qquad \text{for } \ j \in \mathbb{N}_0. \tag{48}$$

From formulae (46), (47), (48) we obtain formulae for integrals of the normal distribution function $\phi(x) = \int_{-\infty}^{x} \varphi(y)\,\mathrm{d}y$.

B.1 Proposition. For $\Delta \geq 0$, $q \in \mathbb{N}_0$, let

$$I_{1,1}(\Delta,q) \ = \ \int_\Delta^{+\infty} x^q\Big[1-\phi(x-\Delta)\Big]\,\mathrm{d}x, \quad I_{1,2}(\Delta,q) \ = \ \int_\Delta^{+\infty} x^q\Big[1-\phi(x+\Delta)\Big]\,\mathrm{d}x, \tag{49}$$

$$I_{2,1}(\Delta,q) \ = \ \int_{-\infty}^{-\Delta} x^q \phi(x+\Delta)\,\mathrm{d}x, \quad I_{2,2}(\Delta,q) \ = \ \int_{-\infty}^{-\Delta} x^q \phi(x-\Delta)\,\mathrm{d}x, \tag{50}$$

$$J_0(\Delta,q) \ = \ \int_0^{2\Delta} (x-\Delta)^q \phi(x)\,\mathrm{d}x, \tag{51}$$

and let $I(\Delta,q)$, $J(\Delta,q)$ be defined by formulae (33), (34). Then we have:

(a) $\displaystyle (-1)^q I_{2,1}(\Delta,q) \ = \ I_{1,1}(\Delta,q) \ = \ \sum_{j=0}^{q}\binom{q}{j}\frac{\Delta^{q-j}}{j+1}\,\Gamma\left(\frac{j}{2}+1\right)\,\frac{2^{\frac{j-1}{2}}}{\sqrt{\pi}}.$

(b) $(-1)^q I_{2,2}(\Delta,q) \ = \ I_{1,2}(\Delta,q) \ =$

$$\sum_{j=0}^{q}\binom{q}{j}\frac{(-\Delta)^{q-j}}{j+1}\left\{-(2\Delta)^{j+1}\Big[1-\phi(2\Delta)\Big] \ + \ \Gamma\left(\frac{j}{2}+1, 2\Delta^2\right)\frac{2^{\frac{j-1}{2}}}{\sqrt{\pi}}\right\}.$$

(c) $J_0(\Delta,q) \ =$

$$\frac{\Delta^{q+1}}{q+1}\left(\phi(2\Delta)+\frac{(-1)^q}{2}\right)$$

$$- \ \frac{1}{q+1}\sum_{j=0}^{q+1}\binom{q+1}{j}(-\Delta)^{q+1-j}\frac{2^{\frac{j}{2}-1}}{\sqrt{\pi}}\ \left[\Gamma\left(\frac{j+1}{2}\right)-\Gamma\left(\frac{j+1}{2}, 2\Delta^2\right)\right].$$

(d) $I(\Delta,q) \;=\; \Big[1+(-1)^q\Big]\Big[I_{1,1}(\Delta,q) \,-\, I_{1,2}(\Delta,q)\Big]$, where $I(\Delta,q)$ is defined by formula (33).

(e) $J(\Delta,q) \;=\; \Big[1+(-1)^q\Big]\left[J_0(\Delta,q) \,-\, \dfrac{\Delta^{q+1}}{q+1}\right]$, where $J(\Delta,q)$ is defined by formula (34).

Proof of assertion (a) of proposition B.1. Substituting $y=-x$ we obtain

$$(-1)^q \int_{-\infty}^{-\Delta} x^q \phi(x+\Delta)\,\mathrm{d}x \quad=\quad -\int_{-\infty}^{-\Delta} (-1)(-x)^q \Big[1-\phi(-x-\Delta)\Big]\,\mathrm{d}x \quad=\quad I_{1,1}(\Delta,q).$$

Substituting $z=y-\Delta$ we obtain

$$I_{1,1}(\Delta,q) \quad=\quad \int_0^{+\infty} (z+\Delta)^q \Big[1-\phi(z)\Big]\,\mathrm{d}z \quad =_{(48)}$$
$$\sum_{j=0}^{q} \binom{q}{j} \frac{\Delta^{q-j}}{j+1} \int_0^{+\infty} y^{j+1}\varphi(y)\,\mathrm{d}y \quad =_{(46)} \sum_{j=0}^{q} \binom{q}{j} \frac{\Delta^{q-j}}{j+1}\, \Gamma\left(\frac{j}{2}+1\right) \frac{2^{\frac{j-1}{2}}}{\sqrt{\pi}}.$$

Assertion (b) of proposition B.1 is proved analogously: the first identity is obtained by substituting $y=-x$; the second identity is obtained by substituting $z=y+\Delta$, and then using formulae (46) and (47).

Proof of assertion (c) of proposition B.1. Integration by parts provides

$$J_0(\Delta,q) \quad=\quad \left[\frac{(x-\Delta)^{q+1}}{q+1}\phi(x)\right]\Big|_0^{2\Delta} \;-\; \frac{1}{q+1}\int_0^{2\Delta} (x-\Delta)^{q+1}\varphi(x)\,\mathrm{d}x \quad=$$
$$\frac{\Delta^{q+1}}{q+1}\left(\phi(2\Delta)+\frac{(-1)^q}{2}\right) \;-\; \frac{1}{q+1}\sum_{j=0}^{q+1}\binom{q+1}{j}(-\Delta)^{q+1-j}\int_0^{2\Delta} x^j\varphi(x)\,\mathrm{d}x.$$

By formulae (45) and (46) we obtain the assertion on $J_0(\Delta,q)$.

Proof of assertion (d) of proposition B.1. The symmetry relation $\phi(-y)=1-\phi(y)$ for the distribution function of the standard normal distribution provides

$$I(\Delta,q) \quad =_{(33),\,(49),\,(50)} \quad I_{1,1}(\Delta,q)-I_{1,2}(\Delta,q)+I_{2,1}(\Delta,q)-I_{2,2}(\Delta,q) \quad =_{(a),(b)}$$
$$\Big[1+(-1)^q\Big]\Big[I_{1,1}(\Delta,q) \,-\, I_{1,2}(\Delta,q)\Big].$$

Proof of assertion (e) of proposition B.1. The substitution $u=-z$ provides

$$\int_{-2\Delta}^{0} (z+\Delta)^q\phi(z)\,\mathrm{d}z \quad=\quad (-1)^q\int_0^{2\Delta}(u-\Delta)^q\Big[1-\phi(u)\Big]\,\mathrm{d}u \quad=$$

$$(-1)^q \left\{ \frac{\Delta^{q+1}}{q+1} \Big(1 - (-1)^{q+1} \Big) \; - \; J_0(\Delta, q) \right\}.$$

Using this result and substituting $y = x + \Delta$ and, respectively, $z = x - \Delta$ in the definition of $J(\Delta, q)$ in formula (51), we obtain

$$\begin{aligned} J(\Delta, q) \quad &= \quad \int_0^{2\Delta} (y - \Delta)^q \phi(y) \, \mathrm{d}y \; - \; \int_{-2\Delta}^{0} (z + \Delta)^q \phi(z) \, \mathrm{d}z \quad = \\ & J_0(\Delta, q) \; - \; (-1)^q \left\{ \frac{\Delta^{q+1}}{q+1} \Big(1 - (-1)^{q+1} \Big) \; - \; J_0(\Delta, q) \right\} \quad = \\ & \Big(1 + (-1)^q \Big) \left\{ J_0(\Delta, q) \; - \; \frac{\Delta^{q+1}}{q+1} \right\}. \end{aligned}$$

•

C Three Martingales.

Let the family $\Big((u_{n,1}, u_{n,2}) \Big)_{\mathbb{N}}$ of variables introduced in Section 5 be adapted to its natural filtration $(\mathcal{A}_n)_{\mathbb{N}}$, i.e., let $(\mathcal{A}_n)_{\mathbb{N}}$ be the sequence of smallest σ-algebras with $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \ldots$ where $(u_{n,1}, u_{n,2})$ is Borel-measurable with respect to $\mathcal{A}_n$ for $n \in \mathbb{N}$. Then $U_{n,l} = u_{1,l} + \ldots + u_{n,l}$ is Borel-measurable with respect to $\mathcal{A}_n$ for $n \in \mathbb{N}$, $l = 1, 2$. The martingale property with respect to the filtration $(\mathcal{A}_n)_{\mathbb{N}}$ is determined by the conditional expectations $E[\,\cdot\,|\mathcal{A}_n] = E[\,\cdot\,|u_{n,1}, u_{n,2}, \ldots, u_{1,1}, u_{1,2}]$. In this sense, the following sequences $(R_{n,l})_{n\in\mathbb{N}}$, $(Y_{n,l})_{n\in\mathbb{N}}$, $(Z_{n,l})_{n\in\mathbb{N}}$ with

$$R_{n,l} \;=\; U_{n,l}^2 - n, \qquad Y_{n,l} \;=\; \sum_{k=1}^{n-1} U_{k,l}^2 - n U_{n,l}^2 + \frac{n(n+1)}{2}, \tag{52}$$

$$Z_{n,l} \;=\; \frac{1}{6} U_{n,l}^4 - n U_{n,l}^2 + \frac{n^2}{2}, \tag{53}$$

are martingales. For the proof, we observe that

$$E[u_{k+1,l}^q U_{k,l}^r | \mathcal{A}_k] \quad = \quad U_{k,l}^r E[u_{k+1,l}^q] \quad = \quad U_{k,l}^r \cdot \begin{cases} 0, & \text{if } q \text{ is odd,} \\ \frac{q!}{\left(\frac{q}{2}\right)! 2^{q/2}}, & \text{if } q \text{ is even,} \end{cases} \tag{54}$$

since $U_{k,l}^r$ is measurable with respect to $\mathcal{A}_k$ and $u_{k+1,l}^q$ is independent of $\mathcal{A}_k$. Hence

$$E[R_{n+1,l}|\mathcal{A}_n] \quad = \quad E[R_{n,l} + 2u_{n+1,l}U_{n,l} + u_{n+1,l}^2 - 1 | \mathcal{A}_k] \quad =_{(54)} \quad R_{n,l},$$

$$E[Y_{n+1,l}|\mathcal{A}_n] \quad = \quad E[Y_{n,l} - (n+1)u_{n+1,l}^2 - 2u_{n+1,l}U_{n,l} + n + 1|\mathcal{A}_k] \quad =_{(54)} \quad Y_{n,l},$$

$$\begin{aligned} Z_{n+1,l} \quad &= \\ & Z_{n,l} + \frac{2}{3}u_{n+1,l}U_{n,l}^3 + u_{n+1,l}^2 U_{n,l}^2 \frac{2}{3}u_{n+1,l}^3 U_{n,l} + \\ & \frac{1}{6}u_{n+1,l}^3 - U_{n,l}^2 - 2(n+1)u_{n+1,l}U_{n,l} - (n+1)u_{n+1,l}^2 + n + \frac{1}{2} \end{aligned} \tag{55}$$

and hence by (54) $E[Z_{n+1,l}|\mathcal{A}_n] = Z_{n,l}$.