Run lengths of control charts for correlated output of feedback processes

de Fokkert, J.I.; Di Bucchianico, A.; Wijnen, J.T.M.; Praagman, J.

Published: 01/01/1999

Document Version
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the author’s version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 14. Dec. 2018
Run Lengths of Control Charts for Correlated Output of Feedback Processes


Eindhoven, June 1999
The Netherlands
Run Lengths of Control Charts for Correlated Output of Feedback Processes

J.I. van Zante - de Fokkert *  A. Di Bucchianico†  J.Th.M. Wijnen‡  J. Praagman§

Abstract
We study the influence of a shift in the mean level of the disturbance in one of the feedback models introduced in Box and Kramer (1992). We show that MMSE feedback control cannot completely remove the effect of this shift. Therefore we investigate the effect of Shewhart and CUSUM control charts for the correlated output of general linear feedback mechanisms. We present a practical way to compute with arbitrary precision the average run length and the standard deviation of the run length of a Shewhart control chart on residuals. For the corresponding CUSUM chart, we derive integral equations.

Keywords  SPC, APC, Shewhart control chart, CUSUM control chart

AMS Subject Classification  62N10

*Centre for Quantitative Methods, P.O. Box 414, 5600 AK Eindhoven, The Netherlands, email: vanZante@cqm.nl
†Corresponding author. EURANDOM and Eindhoven University of Technology, Department of Mathematics, both P.O. Box 513, 5600 MB Eindhoven, The Netherlands, email: sandro@win.tue.nl
‡Eindhoven University of Technology, Department of Mathematics, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, email: jacttmw@win.tue.nl
§Centre for Quantitative Methods, P.O. Box 414, 5600 AK Eindhoven, The Netherlands, email: Praagman@cqm.nl
1 Introduction

Statistical Process Control (SPC) is one of the methods used to monitor the variation of product quality and to reduce it if possible. Control charts, like Shewhart, CUSUM, and EWMA charts, form a considerable part of the techniques associated with SPC. A control chart represents the course of a quality characteristic versus time. By observing the quality characteristic in this way, it can be judged whether the production process is in statistical control or not. This judgement is based on control limits, indicated on the chart. A process is said to be in control if it is only influenced by common causes, i.e., causes that are responsible for variation that is inherent to the process. In fact, by using a model of the process's inherent variation, SPC aims at a timely detection of variation in the process that goes beyond this common cause behaviour. Such deviations are interpreted as signals for special cause variability and should be followed by identification and removal of the particular special cause. SPC monitors the process and tries to minimize the consequences of special causes.

On the other hand, Automatic Process Control (APC) (also called Engineering Process Control) can be used in situations where production processes are subject to disturbances that cannot be removed (e.g., a trend caused by wear-out). APC aims at maintaining the product quality as close as possible to a target value, by carrying out control actions that compensate for the predicted influence of the disturbance. This control action can be derived from a transfer model that describes the relation between the output variable(s) and the input variable(s). In this case the automatic control mechanism is part of the normal process behaviour.

An excellent overview of the different aims, models and backgrounds of SPC and APC can be found in Göb (1988). Although SPC and APC have different aims (SPC is monitoring the process while APC is adjusting it), there are certain circumstances under which feedback control corresponds to an EWMA control chart (cf. Box and Kramer (1992), MacGregor (1988), and MacGregor and Harris (1990)). For many years SPC and APC have been used separately of each other. The idea of implementing SPC and APC simultaneously in a single production process has been put forward from time to time (see e.g., Barnard (1963)), but it is safe to say that this idea did not catch until the papers (Hahn (1989) and MacGregor and Harris (1990)) appeared. The ideas of these papers have been taken up by several people, resulting in papers like Box and Kramer (1992), Vander Wiel et al. (1992), Tucker et al. (1993), Montgomery et al. (1994), Faltin et al. (1997) and a monograph: Box and Lukeño (1997). Several issues must be taken care of when combining SPC and APC. First of all one has to carefully choose a model for the disturbance. Some authors argue that a standard model with independent identically distributed observations may not be appropriate. Zhang (1998) advocated the use of weakly stationary processes, while Box and Kramer (1992) and Vander Wiel (1996) propose to use integrated moving average models. Usually single persistent shifts of the mean are investigated, an exception being Göb et al. (1998) where also a sequence of shifts at random times is included in the model. Another issue is what to chart and how to chart (i.e., the choice of control chart). Since feedback control results in correlated data, standard control charts are not appropriate. A general technique for monitoring a process with correlated data is to use a control chart on residuals of a fitted time series (see e.g., Alwan and Roberts (1988), Harris and Ross (1991), Montgomery and Mastrangelo (1991), Longnecker and Ryan (1992), Yashchin (1993), Wardell et al. (1994) and Faltin et al. (1997)). The use of cuscore charts in this context is advocated by Box and Lukeño (1997) and Shao (1998). An alternative is to use charts based on run sums as proposed in Willemain and
Yet another possibility is to chart the raw correlated data and modify the control limits, see e.g., Vasilopoulos and Stamboulis (1978) and Schmid (1995). Finally, the choice of controller is important. In the statistical literature, one usually considers MMSE feedback control or PI controllers (see e.g., Box and Luceño (1997) and Tsung et al. (1998)). However, we should bear in mind that more sophisticated controllers have been developed during the last decades by the control community.

In this article we consider a slight extension of one of the feedback models discussed in Box and Kramer (1992). In particular, we study Minimum Mean Squared Error (MMSE) feedback control. We present an industrial example where this model is applied in Section 2. Section 2 also contains our model assumptions. We apply a shift in the mean level of the disturbance and explicitly calculate its influence on both the output measurements and the control actions in Section 3. It turns out that MMSE feedback control cannot completely remove the effect of this shift, which results in a nonzero expected deviation of the output from its target value. This emphasizes the importance of applying SPC to a process that is automatically controlled. This also puts doubt on the emphasis in the statistical literature on MMSE feedback control. For a delay period equal to one (this is the number of periods that goes by before a control action influences the output), we show in Section 4 how to compute with arbitrary precision the average run length (ARL) and standard deviation of the run length (SRL) of a Shewhart chart on the residuals of an ARMA model under a shift of the mean of the disturbance. This extends the results of Longnecker and Ryan (1992) and Wardell et al. (1994). We also derive integral equations for the ARL of a corresponding CUSUM chart. A small simulation study compares the performance of Shewhart charts with CUSUM charts. Finally, three appendices contain derivations of our formulas.

This paper is partly based on Van Zante (1993). In particular, Sections 2 and 3 are direct extensions of Van Zante (1993).

### 2 Feedback Control

In this section we discuss feedback control. We start with an example that one of us encountered during work for a major producer of optical discs. One of the process steps is to put a thin layer of tellurium on a glass plate. This is accomplished by placing batches of 22 plates into a 'sputter room', in which 'target material' (tellurium) is placed. The room is filled with Argon, which is ionised by an electrical charge. The ions possess sufficient energy to extract atoms from the target. Due to inhomogeneity of the target material, there are fluctuations in the number of atoms that precipitate on the glass plates.

Let $Y_t$ be the thickness of the tellurium layer if the target material were homogeneous. However, due to the inhomogeneity of the target material, the actual thickness $U_t$ of the tellurium layer equals $Y_t + Z_t$, where $Z_t$ is a disturbance. The thickness of the tellurium layer is crucial for the quality of the optical disc. In order to control this quality, the thickness of the tellurium layer of one plate is used to adjust the power $X_t$ of the electrical charge. Practical experience shows that the disturbance $Z_t$ can be modelled by an AR(1) process. Thus the following equations form an adequate model for the sputter process:
\[ Y_t = \mu + g X_{t-b} \]  
\[ U_t = \mu + g X_{t-b} + Z_t \]  
\[ Z_t = \varphi Z_{t-1} + \alpha_t, \]  

where \( \mu \) and \( g \) are constants, \( b \) is the delay period and \( \alpha_t \sim \text{NID}(0, \sigma^2_\alpha) \). A schematic view of the sputter process is given in Figure 1.

The control action \( X_t \) is carried out by an MMSE controller. The following simulation (see Figure 2) indicates that this controller is not able to cope with a small shift in the mean of disturbance \( Z_t \). We took the following (typical) values for the model parameters: \( b = 1, X_0 = 1.82, Z_0 = 1.5, \varphi = -0.25, g = 16.6, \mu = 30.2, \) and \( \sigma_\alpha = 0.17 \) (and hence, \( \sigma_Z = \sigma_\alpha / \sqrt{1 - \varphi^2} = 0.1756 \)). We imposed a mean shift of size 0.1756 in \( Z_t \) after 150 periods. Note that the vertical axis of Figure 2 does not start at zero.

![Figure 1: Schematic view of the sputter process](image)

![Figure 2: Simulation of sputter process](image)

We now proceed with a general analysis of the model given by (1) and (2). Note that if \( b = 1 \), then model (1) is model (14) of Box and Kramer (1992). This model is also studied in...
a series of papers by Vander Wiel and collaborators, starting with Vander Wiel et al. (1992). We may and will assume without loss of generality, that $\mu$ equals 0. Instead of using the AR(1) process $Z_t = \varphi Z_{t-1} + a_t$ of the example, we assume that $Z_t$ can be represented by an ARMA($p,q$)-model (Box and Jenkins (1976)):

$$\Phi(B) Z_t = \Theta(B) a_t,$$

where $a_t \sim \text{NID}(0, \sigma^2)$. Here

$$\Phi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$$

and

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

are polynomials in $B$, the backward shift operator ($B^k a_t := a_{t-k}$). We moreover assume that the roots of both $\Phi$ and $\Theta$ lie outside the unit circle. Hence, there exist power series $\Gamma = \Theta/\Phi$ and $\Xi = \Phi/\Theta$ such that

$$Z_t = \Gamma(B) a_t = \left(1 - \sum_{i=1}^{\infty} \gamma_i B^i\right) a_t,$$

and

$$a_t = \Xi(B) Z_t = \left(1 - \sum_{i=1}^{\infty} \xi_i B^i\right) Z_t.$$

The control action $X_t$ aims to keep the expected output deviation close to zero. As controller we take a linear causal filter on the error terms $a_t$, i.e. we set the control action at time $t$ equal to:

$$X_t = \sum_{j=0}^{\infty} c_j a_{t-j} = \left(\sum_{j=0}^{\infty} c_j B^j\right) a_t.\ (6)$$

Note that because of the linear relations between $a_t$, $U_t$, and $Z_t$, a linear filter on any of these is also a linear on the others. Combining (6) with (2) and (4), we obtain that for $b > 1$, the output $U_t$ equals

$$U_t = \left[1 - \sum_{j=1}^{b-1} \gamma_j B^j + \sum_{j=b}^{\infty} (g c_{j-b} - \gamma_j) B^j\right] a_t.\ (7)$$

while for $b = 1$, we have

$$U_t = a_t + \sum_{j=1}^{\infty} (g c_{j-1} - \gamma_j) a_{t-j}.\ (8)$$

Since the MMSE $b$-steps ahead linear forecast $\hat{Z}_{t+b}(t)$ at time $t$ is given by (see Box and Jenkins (1976))

$$\hat{Z}_{t+b}(t) = -\sum_{j=0}^{\infty} \gamma_{b+j} a_{t-j},\ (9)$$
it follows that MMSE-controllers are included in this setup by choosing \( c_j = \gamma_{b+j}/g \) in (6). Thus MMSE controllers are special in the sense that in this case control and monitoring are largely related matters, because they are based on (forecasts of) disturbances. For MMSE controllers, (7) reduces for \( b > 1 \) to

\[
U_t = a_t - \sum_{j=1}^{b-1} \gamma_j a_{t-j},
\]

which is the \( b \)-steps ahead forecast error of the disturbance \( Z_t \). Furthermore, if MMSE control is applied with \( b = 1 \), then \( U_t = a_t \), which implies that the output is uncorrelated.

Because the disturbance \( Z_t \) (and hence \( a_t \)) cannot be measured directly, we rewrite (6) in terms of current and previous output deviations \( U_t \), by substituting (7) into (6):

\[
X_t = \frac{\sum_{j=0}^{b-1} c_j B^j}{1 - \sum_{j=1}^{b} \gamma_j B^j + \sum_{j=b}^{\infty} (g C_j - \gamma_j) B^j} U_t,
\]

while for \( b = 1 \) we have

\[
X_t = \frac{\sum_{j=0}^{\infty} c_j B^j}{1 + \sum_{j=1}^{\infty} (g C_{j-1} - \gamma_j) B^j} U_t.
\]

Since the \( U_t \) are correlated for \( b > 1 \) and for \( b = 1 \) if the controller is not an MMSE controller (cf. (7)), we cannot apply standard control charts to \( U_t \). However, it follows from (2) and (9) that \( e_{U_t}(1) \), the one-step ahead forecast error of \( U_t \) based on minimum MMSE forecasts, satisfy

\[
e_{U_t}(1) = Z_t - \tilde{Z}_t(t-1) = a_t,
\]

where \( \tilde{Z}_t(t-1) \) is the forecast of \( Z_t \) at time \( t-1 \). Hence, the forecasts \( e_{U_t}(1) \) are uncorrelated.

In the following section we use these formulas to compute the influence of a shift in the mean of the disturbance.

### 3 The influence of a shift in the disturbance

In this section we study the effect of a change in the disturbance \( Z_t \) on the output \( U_t \) and the control action \( X_t \), in order to show the importance of monitoring an automatically controlled process. The simplest change is a persistent shift of the mean. When a shift occurs in the mean level of the disturbance, then feedback control reacts to the increased (or decreased) output deviation. Because of the compensating influence of the control action on the produced output, the shift in the disturbance will not become visible as one single shift in the mean level of the output deviation. Formulas (14) and (16) show the influence of the shift on
the produced output $U_t$ and on the control action $X_t$, respectively. The special case of ARMA(1,1)-disturbances $Z_t$ and MMSE controllers $X_t$ can be found in Vander Wiel et al. (1992). We assume in the sequel that all parameters in (2) and (3) are known and do not change in time. We will show that even in this ideal situation, serious problems occur.

We assume that there is no shift of the mean level of the disturbance until time $c$, i.e., $Z_t = \delta_t + \Gamma(B) a_t$, with $\delta_t = 0$ for $t < c$ and $\delta_t = \delta$ for $t \geq c$. Note that the $a_t$ are still assumed to be NID(0, $\sigma^2_a$). Hence, the theoretical controller (6) and the practical controller (11) are not equivalent when $\delta_t \neq 0$.

In order to compute the influence of a shift, we first derive the influence of a general change $Z_t = \delta_t + \Gamma(B) a_t$ with no restrictions on $\delta_t$. We use $U_{t-b} = B^b U_t$ and substitute the above formulas and (11) into (2). This yields for $b > 1$:

$$U_t = \left(1 - \sum_{j=1}^{b-1} \gamma_j B^j + \sum_{j=b}^{\infty} (g c_{j-b} - \gamma_j) B^j\right) (\Xi(B) \delta_t + a_t),$$

(14)

while for $b = 1$ we have

$$U_t = \left(1 + \sum_{j=1}^{\infty} (g c_{j-1} - \gamma_j) B^j\right) (\Xi(B) \delta_t + a_t).$$

(15)

Similarly, by substituting (14) into (11) or (15) into (12), we find the effect of the shift on the control action $X_t$ for $b > 1$:

$$X_t = \left(\sum_{j=0}^{\infty} c_j B^j\right) (\Xi(B) \delta_t + a_t).$$

(16)

Note that (16) reduces to (6) if $\delta_t \equiv 0$.

The following special cases of (14) are of interest. If $\delta_t \equiv 0$ (no shift), then (14) reduces to (7). If $X_t$ is an MMSE controller, then (14) reduces to

$$U_t = \left(1 - \sum_{j=1}^{b-1} \gamma_j B^j\right) (\Xi(B) \delta_t + a_t)$$

(17)

and (15) reduces to

$$U_t = \Xi(B) \delta_t + a_t.$$  

(18)

We now return to our goal of computing a persistent change of size $\delta$ starting at time $c$, i.e. $\delta_t = \delta I[t \geq c]$, where

$$I[t \geq c] = \begin{cases} 
1 & \text{if } t \geq c \\
0 & \text{otherwise}
\end{cases}.$$

Then (18) becomes:

$$U_t = \Xi(B) \delta I[t \geq c] + a_t,$$

(19)

where $B^k I[t \geq c] = I[t-k \geq c]$. This has the following important consequence. Although we applied a shift to the mean of $Z_t$, the feedback controller causes the output to be the same as
the output of an uncontrolled ARMA process with a shift in the mean of the driving noise \( \alpha_t \). In other words, the output \( U_t \) is the same as the output of an uncontrolled ARMA-process with a shift of the mean of residuals \( \alpha_t \).

In order to show the vulnerability of MMSE feedback control, let us first consider the special case of white noise disturbance \( (\Xi(B) = 1) \). In this case there is no control \((X_t = 0)\), and thus there is no compensation for the shift \((U_t = \delta I[t > c] + \alpha_t)\). If \( Z_t = \Gamma(B) \alpha_t \), then it follows from (17) that the change in the output deviation due to the shift in the disturbance equals:

\[
\Xi(B) \left( 1 - \sum_{j=1}^{b-1} \gamma_j B^j \right) \delta I[t \geq c] = \left( 1 + \Xi(B) \sum_{j=0}^{\infty} \gamma_{b+j} B^{b+j} \right) \delta I[t \geq c].
\]

The right-hand side shows that the mean level of the output equals \( \delta \) during the first \( b \) periods after \( c \), whereas the left-hand side shows that the mean level of the output converges to

\[
(1 - \xi_1 - \xi_2 - \cdots) (1 - \gamma_1 - \cdots - \gamma_{b-1}) \delta.
\]

Since \( \Xi(B) \) has only roots outside the unit circle, this asymptotic level is unequal to zero unless \( \gamma_1 + \cdots + \gamma_{b-1} = 1 \). Hence, MMSE feedback control is not able to remove the influence of the shift completely, which results in an increased output deviation of the target value. This is another proof of the vulnerability of MMSE controllers, which is well-known to control engineers but seems to be not generally known to statisticians. Note however, that if the disturbance would follow an ARIMA process \((\Xi\) has a root at 1), then MMSE feedback control is able of removing the disturbance. It is useful to note that the performance of the controller depends on the type of disturbance.

The following section discusses the application of control charts for the system described in (1)-(3).

### 4 Control charts for \( U_t \)

In this section we discuss control charts for the system (1)-(3) with MMSE control. Note that monitoring \( U_t, X_t, \) and \( Z_t \) have different aims. Apart from the gain constant \( g \), the portion of the disturbance compensated for by the MMSE controller is moved from \( U_t \) to \( X_t \). Thus, monitoring \( U_t \) is looking for evidence of unsuccessful compensation, while monitoring \( X_t \) is looking for evidence of successful compensation. However, in practice, one cannot always measure the control action \( X_t \). Ideally, one would like to monitor the disturbance \( Z_t \), since this picks up the shift, whether successfully compensated for or not. The disturbance \( Z_t \) can only be retrieved by using (2). Therefore, we restrict ourselves to control charts for \( U_t \).

Depending on the size of the delay period, two different situations for \( U_t \) exist. It follows from (7) that for MMSE controllers, the output measurements are uncorrelated if and only if \( b = 1 \). Therefore, we have to distinguish between the cases \( b = 1 \) and \( b > 1 \).

#### 4.1 Control charts for \( U_t \) if \( b = 1 \)

To apply SPC to the system described by (1)-(3) with \( b = 1 \) and MMSE feedback control, Box and Kramer (1992) suggest Shewhart control charts for the observed output \( U_t \), the disturbance \( Z_t \), and the control action \( X_t \).

MacGregor and Harris (1990) advise to plot the
difference $X_t - X_{t-1}$ or the disturbance $Z_t$. Vander Wiel et al. (1992) applied a CUSUM control chart to the output $U_t$. In the following two subsections we study a Shewhart and a CUSUM control chart for the output $U_t$, if $b = 1$. Both control charts have proven to be useful in traditional SPC (uncorrelated data). Shewhart control charts are easy to implement and perform well in detecting large shifts, while CUSUM control charts outperform Shewhart control charts in detecting small shifts. We investigate the performance of these charts for the detection of a one standard deviation shift in the mean of the disturbance process. Note that if $b = 1$, then it follows from (8) that $U_t = a_t$. Hence, control charts for the output $U_t$ coincide with control charts for residuals of ARMA($p,q$)-processes. These charts have been studied by several authors in order to monitor correlated data (Alwan and Roberts (1988), Harris and Ross (1991), Montgomery and Mastrangelo (1991), Longnecker and Ryan (1992), Wardell et al. (1994), Runger et al. (1995), and Faltin et al. (1997)). Note that although we applied a shift of the mean of $Z_t$, the feedback controller causes the output to be the same as the output of an uncontrolled ARMA process with a shift in the mean of the driving noise $a_t$. We will use the average run length (ARL) and the standard deviation of the run length (SRL) to judge the performance of a control chart. The ARL$_{\text{out}}$, which is the average number of observations (periods) between the shift and its detection, has to be as small as possible. On the other hand, the ARL$_{\text{in}}$, which is the average number of periods between two false alarms, has to be reasonably large. In order to get an idea of the spread of the run length distribution, we also study its standard deviation. E.g., a 3-$\sigma$ Shewhart control chart for uncorrelated observations from a normal distribution, has an ARL$_{\text{in}}$ of 370 and an SRL$_{\text{in}}$ of 370. In case of a one sigma shift in the mean level, the ARL$_{\text{out}}$ equals 43.9 and the SRL$_{\text{out}}$ equals 43.4. Subsections 4.1.1 and 4.1.2 discuss the Shewhart and the CUSUM control charts for the output $U_t$, respectively.

### 4.1.1 A Shewhart Control Chart for $U_t$ if $b = 1$

In this subsection we study the performance of Shewhart control charts. The central line of this chart represents the centre of the distribution of the quality characteristic that we want to monitor. The control limits are usually placed three sigma above and below the central line. Thus for MMSE control with $b = 1$, a Shewhart control chart for $U_t = a_t$ has the following form:

Upper control limit : $EU_t + 3 \sqrt{\text{Var}U_t} = 3\sigma_a$

Central line : $EU_t = 0$

Lower control limit : $EU_t - 3 \sqrt{\text{Var}U_t} = -3\sigma_a$.

Although we only use these limits in the sequel, it is easy to adapt our formulas to other limits.

Alwan and Roberts (1988) suggested the idea of monitoring residuals of correlated data instead of the original data. Longnecker and Ryan (1992) gave a formula for the ARL$_{\text{out}}$ of a Shewhart control chart for the residuals of an AR($p$)-process. For the same case, Wardell et al. (1994) gave formulas for both the ARL$_{\text{out}}$ and the SRL$_{\text{out}}$. Moreover, an exact formula for the corresponding probability generating function is contained in formula (D.4) of Wardell et al. (1994), by taking $g(x) = x^z$. For the general ARMA($p,q$)-case, Wardell et al. (1994) only gave approximate results for the ARL$_{\text{out}}$ and SRL$_{\text{out}}$ without indication of the accuracy.
while Longnecker and Ryan (1992) gave upper and lower bounds for the ARL\textsubscript{\text{out}} with a non-computable estimate of the error. We improve on these papers by giving a closed formula for the probability generating function of out-of-control run length, from which we derive computable exact upper and lower bounds for ARL and SRL that can be made arbitrarily close. The approximations of Wardell et al. (1994) turn out to coincide in special cases with our lower or upper bounds (depending on the signs of the coefficients of $\Xi$).

As mentioned before, $U_t = a_t$ (see (10)) and the ARL\textsubscript{\text{lin}} and SRL\textsubscript{\text{lin}} are both approximately equal to 370. By (19), a shift of size $\delta$ in the mean level of $Z_t$, will not result in just one change in the mean level of the output $U_t$, but in a sequence of level changes. Let $F$ be the cumulative distribution function of the standard normal distribution and let and $\Gamma$ and $\Xi$ be as in (4) and (5). Assume that a shift of size $\delta$ occurs in the disturbance at time $c$. Define $p_{k+1}$ ($k \geq 0$) to be the probability that the output $U_t$ measured at the end of period $c + k$, falls outside the control limits. Since $p_{k+1} = P(|U_{c+k}| > 3\sigma_n)$, it follows from (19) that (cf. (8) of Wardell et al. (1994)):

$$p_{k+1} = 1 - F \left( 3 - (1 - \xi_1 - \cdots - \xi_k) \frac{\delta}{\sigma_n} \right) + F \left( -3 - (1 - \xi_1 - \cdots - \xi_k) \frac{\delta}{\sigma_n} \right).$$

(20)

Closed formulas for the $\xi_k$ can be obtained easily (by hand or using a computer algebra system like Mathematica) using partial fractions and power series expansions, the approach taken by Wardell et al. (1994) being unnecessarily complicated. Using Lemma A1 of Appendix A, $P(\text{RL}_{\text{out}} \geq 1) = 1$, and

$$P(\text{RL}_{\text{out}} \geq j) = \prod_{i=1}^{j-1} (1 - p_i)$$

for $j = 2, \ldots$, we obtain:

$$P_{\text{RL}_{\text{out}}} (z) = z + \frac{z-1}{z} \sum_{j=2}^{\infty} \prod_{i=1}^{j-1} (1 - p_i) z^j,$$

(21)

where $P_{\text{RL}_{\text{out}}} (z)$ is the probability generating function of the out-of-control run length. Part b) of Lemma A1 of Appendix A yields that

$$\text{ARL}_{\text{out}} = 1 + \sum_{j=1}^{\infty} \prod_{i=1}^{j} (1 - p_i)$$

(22)

and

$$\text{VRL}_{\text{out}} = 1 + \sum_{j=1}^{\infty} (2j + 1) \prod_{i=1}^{j} (1 - p_i) - (\text{ARL}_{\text{out}})^2.$$  

(23)

Elementary calculus yields that $F(3 - x) - F(-3 - x) \leq 2F(3) - 1$ for all $x$, hence $1 - p_k \leq 2F(3) - 1 < 1$ for all $k$. It thus follows from the Ratio Test that the right-hand sides of (22) and (23) are convergent, i.e. both the ARL\textsubscript{\text{out}} and the SRL\textsubscript{\text{out}} are finite.

The above formulas are exact, but in general not useful in this form, since there is no directly computable formula for $p_k$. These formulas are useful, however, for the important
special case of AR(p)-processes. For these processes, $\Xi(B)$ is a polynomial in $B$ of degree $p$, and thus $p_k = p_{p+1}$ for $k \geq p + 1$. Hence, in this case (22) and (23) reduce to

$$\text{ARL}_{\text{out}}(\text{AR}(p)) = 1 + \sum_{j=1}^{p} \prod_{i=1}^{j} (1 - p_i) + \frac{\prod_{i=1}^{p+1} (1 - p_i)}{p_{p+1}},$$

and

$$\text{VRL}_{\text{out}}(\text{AR}(p)) = 1 + \sum_{j=1}^{p} (2j + 1) \prod_{i=1}^{j} (1 - p_i) + \left( \frac{2 + (2p + 1)p_{p+1}}{p_{p+1}^2} \right) \prod_{i=1}^{p+1} (1 - p_i) - (\text{ARL}_{\text{out}})^2,$$

which were obtained earlier by Longnecker and Ryan (1992) (only ARL) and Wardell et al. (1994) (see Lemma B3 in Appendix B for another derivation). Figures 3 and 4 gives contour plots of out-of-control ARL’s and SRL’s for AR(2) processes with $\delta = \sigma_z$ and the standard 3σ control limits. Contour plots of out-of-control ARL’s for other control limits can be found in Longnecker and Ryan (1992).

![Figure 3: ARL out of AR(2) with MMSE feedback and one sigma shift ($\delta = \sigma_z$)](image)

For general ARMA($p,q$)-processes, Wardell et al. (1994) resorted to approximations without investigating accuracy. Their approximations are based on setting $p_t$ constant from a certain cutoff value (cf. the AR(p) case). Longnecker and Ryan (1992) give upper and lower bounds for the ARL out for general ARMA($p,q$)-processes, expressed in terms of the difference between the probabilities $p_k + 1$ from (20) and their limit. We improve on the bounds given by Longnecker and Ryan (1992) by using certain monotonicity properties of the probabilities $p_k$, thus avoiding the numerical problems of bounding the difference between the probabilities $p_k + 1$ and their limit. Note that both our results and those from Longnecker and Ryan are hard bounds, not approximations. Moreover, we extend their results by giving bounds for the SRL out. Our bounds are based on first computing $p_1, \ldots, p_n$, and then replacing the remaining $p_k$’s by upper and lower bounds. Fix $n$ and let $p_{\text{low}}$ and $p_{\text{up}}$ be such that

$$p_{\text{low}} \leq p_k \leq p_{\text{up}}.$$
for $k > n$. In this way we obtain the following hard bounds for $\text{ARL}_{\text{out}}$ and $\text{VRL}_{\text{out}}$:

$$1 + \sum_{j=1}^{n-1} \prod_{i=1}^{j} (1 - p_i) + \frac{\prod_{i=1}^{n} (1 - p_i)}{\text{UB}_{\text{up}}} \leq \text{ARL}_{\text{out}} \leq 1 + \sum_{j=1}^{n-1} \prod_{i=1}^{j} (1 - p_i) + \frac{\prod_{i=1}^{n} (1 - p_i)}{\text{UB}_{\text{low}}}$$

(24)

and

$$\text{LB}(\text{VRL}_{\text{out}}) \leq \text{VRL}_{\text{out}} \leq \text{UB}(\text{VRL}_{\text{out}})$$

(25)

where

$$\text{LB}(\text{VRL}_{\text{out}}) = 1 + \sum_{j=1}^{n-1} (2j + 1) \prod_{i=1}^{j} (1 - p_i) + \left(2 + \frac{(2n - 1)p_{\text{up}}}{p_{\text{up}}^2}\right) \prod_{i=1}^{n} (1 - p_i) - (\text{UB}(\text{ARL}_{\text{out}}))^2,$$

and

$$\text{UB}(\text{VRL}_{\text{out}}) = 1 + \sum_{j=1}^{n-1} (2j + 1) \prod_{i=1}^{j} (1 - p_i) + \left(2 + \frac{(2n - 1)p_{\text{low}}}{p_{\text{low}}^2}\right) \prod_{i=1}^{n} (1 - p_i) - (\text{LB}(\text{ARL}_{\text{out}}))^2,$$

with $\text{LB}(\text{ARL}_{\text{out}})$ and $\text{UB}(\text{ARL}_{\text{out}})$ the lower and upper bound of $\text{ARL}_{\text{out}}$ as given in (24)).

We refer to appendix B for proofs of these bounds.

At first sight, these formulas seem hard to use since they require hard bounds on the probabilities $p_k$ of (20) for large $k$. However, we will give two examples that illustrate how to find such bounds. We refer to Chapter 3 of Box and Jenkins (1976) for the conditions that the parameters must satisfy for the disturbance to be causal and invertible.

**MA(1)-model** An MA(1)-model can be described by:

$$Z_t = (1 - \theta B) a_t; \quad \sigma_z^2 = (1 + \theta^2) \sigma_a^2.$$
Thus $\xi_k = -\theta^k$. Since $\theta < 1$, we have
\[
\frac{1 - |\theta|^{k+1}}{1 - \theta} \leq 1 + \theta + \cdots + \theta^k \leq \frac{1 + |\theta|^{k+1}}{1 - \theta}.
\]
Note that these bounds for $1 + \theta + \cdots + \theta^k$ are positive and monotone in $k$. Since $F$ is increasing, it is thus possible to write down explicit bounds for $p_k$ for $k > n$.

**ARMA(1,1)-model** An ARMA(1,1)-model can be described by:
\[
(1 - \varphi B) Z_t = (1 - \theta B) a_t ; \quad \sigma_z^2 = \frac{1 + \theta^2 - 2\varphi\theta}{1 - \varphi^2} a_a^2.
\]
We proceed in a similar way as for the MA(1)-model. Here $\xi_k = (\varphi - \theta) \theta^{k-1}$, which yields the bounds
\[
\frac{1 - \varphi - |\varphi - \theta| |\theta|^k}{1 - \theta} \leq 1 - (\varphi - \theta) \left(1 + \theta + \cdots + \theta^{k-1}\right) \leq \frac{1 - \varphi + |\varphi - \theta| |\theta|^k}{1 - \theta}.
\]
Note that the lower bound is positive for large $k$. Again, the bounds are monotone in $k$ and we can write down explicit bounds for $p_k$ for $k > n$. Obviously, these bounds coincide with the bounds for the MA(1)-model if $\varphi = 0$.

Our method of obtaining bounds for the ARL and SRL of a Shewhart chart can in principle also be applied to other models than ARMA models. The key property that make our method work, is that one can compute monotone bounds for the probabilities $p_k$ as in the above examples.

We checked the values of the ARL's and SRL's for various ARMA(1,1) parameters in Table 3 of Wardell et al. (1994). Although their table is based on approximations, most of the values do not differ too much from our exact results (where we increased $n$ until our lower and upper bounds coincided; often $n = 30$ sufficed). For values of $\varphi$ and $\theta$ close to 1 and/or -1, $n$ had to be increased to values in the range of 100-150.

In the next subsection we study CUSUM control charts. For uncorrelated data, CUSUM control charts detect small shifts faster than Shewhart control charts (cf. Subsection 4.1.3).

### 4.1.2 A CUSUM Control Chart for $U_t$ if $b = 1$

In this subsection we study the performance of CUSUM control charts. The tabular form of the CUSUM control chart uses $S_H(t)$ and $S_L(t)$, the upper one sided CUSUM and the lower one sided CUSUM for period $t$, respectively. The CUSUM $S_H(t)$ and $S_L(t)$ for a CUSUM control chart for $U_t$ ($b = 1$) are given by (see. e.g. Van Dobben de Bruin (1962) or Hawkins and Olwell (1988)):
\[
S_H(t) = \max(0, U_t - K + S_H(t-1)) \quad S_L(t) = \max(0, -U_t - K + S_L(t-1)) \quad S_H(0) = S_L(0) = 0,
\]
where $K$ is the reference value. Lucas (1982) proposed $K = \delta/2$ for uncorrelated data. We follow the recommendation of Runger et al. (1995) and choose $K$ equal to half the final change in the output level. It follows from (19) that:

$$K = \frac{\delta}{2} \left(1 - \sum_{i=1}^{\infty} \xi_i\right).$$  \hspace{1cm} (26)

If $S_H(t) > H$ or $S_L(t) > H$, we conclude that the mean level of $U_t$ has experienced a positive shift or a negative shift, respectively. One should choose the decision value $H$ in such a way that both the ARL_{out} and the ARL_{in} have acceptable values. Page (1954) derived integral equations for the ARL_{in} and the ARL_{out} of a one-sided CUSUM control chart for uncorrelated data (see also Van Dobben de Bruin (1962)). We extend Page’s results by giving formulas for the ARL_{out} of a one-sided CUSUM control chart for residuals of ARMA(p, q)-processes. Appendix C shows these formulas. We assume that $S_L(c) = S_H(c) = 0$. Lucas and Crosier (1982) gave the ARL_{in} and the ARL_{out} of the two-sided, symmetrical CUSUM control chart under this assumption (see Appendix C). Note that for an in-control situation $U_i \overset{d}{=} U_0$ for $i > 0$. Hence, ARL_{H,i}(i, s) = ARL_{H,0}(0, s) for $i \geq 0$, where ARL_{H,i}(i, s) and ARL_{L,i}(i, s), $i = 0, 1, \ldots$, are the average number of periods that the upper and lower one-sided CUSUM takes to give a false alarm, given $S_H(i-1) = s$ and $S_L(i-1) = s$, respectively. Contrary to the formula for the ARL_{in} of Page (1954) which consists of a single integral equation, our formula for the ARL_{out} consists of an infinite system of integral equations for general ARMA(p, q)-processes. For AR(p)-processes, however, the number of integral equations is finite and equals $p + 1$ (Lemma C4). In this case, we solve the integral equations for ARL_{H,\delta}(c + p, s) and ARL_{L,\delta}(c + p, s) numerically by using Gaussian quadrature with 24 Gaussian points (cf. Vance (1986)). This solution was used to successively numerically solve the integral equations for ARL_{H,\delta}(c + p - 1, s), ARL_{L,\delta}(c + p - 1, s), \ldots, ARL_{H,\delta}(c + 1, s) and ARL_{L,\delta}(c + 1, s), which finally gives ARL_{H,\delta}(c, 0) and ARL_{L,\delta}(c, 0). Note that the Markov approach of Brook and Evans (1972) as extended by Runger et al. (1995) to the AR(p)-case, essentially consists in replacing the integrands in Lemma C3 by piecewise constant functions. If we assume that $U_{c+i} \overset{d}{=} U_{c+n}$ for $i \geq n$ for general ARMA(p, q)-processes, then we can use the above approach also, to find an approximation for the ARL_{out} of a one-sided CUSUM control chart. A choice for $n$ may be a value of $n$ such that the corresponding bounds for Shewhart charts in Subsection 4.1.1 are close. Note, however, that solving the system of integral equations for large values of $n$ may lead to numerical instability.

### 4.1.3 Simulations

We now investigate the behaviour of (two-sided) CUSUM charts through a small simulation study. We use the choice of $K$ as put forward in (26). For all CUSUM charts, we choose $H$ such that ARL_{in} equals approximately 370. Since run length distributions of control charts are usually skewed, we look at both the ARL and SRL to judge the performance of a control chart. Extensive computations for AR(1) processes can be found in Runger et al. (1995). We see that the choice of $H$ changes considerably for different choices of the ARMA(1,1) parameters $\varphi$ and $\theta$.

For uncorrelated data, CUSUM charts perform better than Shewhart charts for small (1 standard deviation or less) shifts of the mean. The simulation presented in table 2 shows that this situation also holds for correlated data.
### Table 1: CUSUM chart for ARMA(1,1) processes

<table>
<thead>
<tr>
<th>$\varphi$</th>
<th>$\theta$</th>
<th>$\delta$</th>
<th>$K$</th>
<th>$H$</th>
<th>ARL$_{in}$</th>
<th>SRL$_{in}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.25</td>
<td>1</td>
<td>0.50</td>
<td>4.81</td>
<td>376</td>
<td>365</td>
</tr>
<tr>
<td>0.25</td>
<td>0.25</td>
<td>2</td>
<td>1.00</td>
<td>2.48</td>
<td>356</td>
<td>360</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.25</td>
<td>1</td>
<td>0.83</td>
<td>3.05</td>
<td>383</td>
<td>377</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.25</td>
<td>2</td>
<td>1.67</td>
<td>1.40</td>
<td>396</td>
<td>414</td>
</tr>
<tr>
<td>0.75</td>
<td>0.25</td>
<td>1</td>
<td>0.17</td>
<td>9.90</td>
<td>346</td>
<td>330</td>
</tr>
<tr>
<td>0.75</td>
<td>0.25</td>
<td>2</td>
<td>0.33</td>
<td>6.40</td>
<td>328</td>
<td>325</td>
</tr>
<tr>
<td>0.25</td>
<td>0.75</td>
<td>1</td>
<td>1.50</td>
<td>1.62</td>
<td>384</td>
<td>363</td>
</tr>
<tr>
<td>0.25</td>
<td>0.75</td>
<td>2</td>
<td>3.00</td>
<td>0.01</td>
<td>359</td>
<td>365</td>
</tr>
<tr>
<td>-0.75</td>
<td>0.25</td>
<td>1</td>
<td>1.17</td>
<td>2.15</td>
<td>373</td>
<td>277</td>
</tr>
<tr>
<td>-0.75</td>
<td>0.25</td>
<td>2</td>
<td>2.33</td>
<td>0.67</td>
<td>351</td>
<td>354</td>
</tr>
<tr>
<td>0.25</td>
<td>-0.75</td>
<td>1</td>
<td>0.21</td>
<td>9.10</td>
<td>398</td>
<td>376</td>
</tr>
<tr>
<td>0.25</td>
<td>-0.75</td>
<td>2</td>
<td>0.43</td>
<td>5.40</td>
<td>381</td>
<td>387</td>
</tr>
</tbody>
</table>

### Table 2: Comparison of Shewhart and CUSUM charts for ARMA(1,1) processes

<table>
<thead>
<tr>
<th>$\varphi$</th>
<th>$\theta$</th>
<th>$\delta$</th>
<th>CUSUM</th>
<th>Shewhart</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$K$</th>
<th>$H$</th>
<th>ARL$_{out}$</th>
<th>SRL$_{out}$</th>
<th>ARL$_{out}$</th>
<th>SRL$_{out}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.25</td>
<td>1</td>
<td>0.50</td>
<td>4.81</td>
<td>376</td>
</tr>
<tr>
<td>0.25</td>
<td>0.25</td>
<td>2</td>
<td>1.00</td>
<td>2.48</td>
<td>356</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.25</td>
<td>1</td>
<td>0.83</td>
<td>3.05</td>
<td>383</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.25</td>
<td>2</td>
<td>1.67</td>
<td>1.40</td>
<td>396</td>
</tr>
<tr>
<td>0.75</td>
<td>0.25</td>
<td>1</td>
<td>0.17</td>
<td>9.90</td>
<td>346</td>
</tr>
<tr>
<td>0.75</td>
<td>0.25</td>
<td>2</td>
<td>0.33</td>
<td>6.40</td>
<td>328</td>
</tr>
<tr>
<td>0.25</td>
<td>0.75</td>
<td>1</td>
<td>1.50</td>
<td>1.62</td>
<td>384</td>
</tr>
<tr>
<td>0.25</td>
<td>0.75</td>
<td>2</td>
<td>3.00</td>
<td>0.01</td>
<td>359</td>
</tr>
<tr>
<td>-0.75</td>
<td>0.25</td>
<td>1</td>
<td>1.17</td>
<td>2.15</td>
<td>373</td>
</tr>
<tr>
<td>-0.75</td>
<td>0.25</td>
<td>2</td>
<td>2.33</td>
<td>0.67</td>
<td>351</td>
</tr>
<tr>
<td>0.25</td>
<td>-0.75</td>
<td>1</td>
<td>0.21</td>
<td>9.10</td>
<td>398</td>
</tr>
<tr>
<td>0.25</td>
<td>-0.75</td>
<td>2</td>
<td>0.43</td>
<td>5.40</td>
<td>381</td>
</tr>
</tbody>
</table>

### 4.2 Control charts for $U_t$ ($b > 1$)

If $b > 1$, then the output $U_t$ is correlated. This correlation offers us the opportunity to predict future observations and to chart the uncorrelated one-step ahead forecast errors of $U_t$.

We know from (13) that the one-step ahead forecast error of $U_t$ based on minimum MMSE forecasts are uncorrelated.

We assume that a shift of size $\delta$ occurs in the mean level of the disturbance at time $c$, i.e., $Z_t = \delta I[t \geq c] + \Gamma(B) a_t$. Substitution of this formula and (3) into (13) yields the effect of the shift on $e_{U_t}(1)$:

$$e_{U_t}(1) = \Xi(B) \delta I[t \geq c] + a_t.$$  

Note that the above formula is identical to (19). This implies that the results of Subsection 4.1 are also valid for Shewhart and CUSUM control charts on the one-step ahead forecast error of $U_t$ in case $b > 1$. 

14
5 Conclusions

We have shown that MMSE feedback control is not capable of completely removing the effect of a relatively shift in the mean of the disturbance. In our example in Section 2, a Shewhart control chart would have detected the shift after 27 periods (see Figure 5) when we set up the standard $3\sigma$ control limits using the true value of $\sigma_a$ (the ARL and SRL are 24.2, 23.2 respectively). A two-sided CUSUM control chart with $K = \delta (1 - \varphi)/2 = 0.1098$ (cf. our discussion of (26)) and $H = 0.67$ has an ARL$_{in}$ of approximately 370. This CUSUM chart would have detected the shift in our simulated data after only 5 periods (see Figure 6), when using the true value of $\sigma_z$ (the ARL$_{out}$ and SRL$_{out}$ are 8.3 and 3.7, respectively). Of course, in practice the parameters need to be estimated, which affects these run lengths.

![Figure 5: Shewhart chart for simulation data](image)

We have also shown how to compute with arbitrary precision the ARL and SRL of Shewhart charts for residuals of ARMA processes. In principle, our technique may also be applied

![Figure 6: CUSUM chart for simulation data](image)
to Shewhart charts for residuals of other processes. For corresponding CUSUM charts, we derived integral equations.

Appendices

A Generating functions

In this appendix we present a general lemma, which is useful for calculating the probability generating function of the run length (RL) of Shewhart control charts for residuals of correlated data. Since RL is expressed most conveniently in terms of $P(\text{RL} \geq j)$, we use the following modified form of the probability generating function.

**Lemma A.1** Let $X$ be a discrete random variable taking values on $1, 2, \ldots$ and define

$$\bar{P}(z) := \sum_{j=1}^{\infty} P(X \geq j) z^j.$$  

Then we have:

a) $P(z) = 1 + \frac{z-1}{z} \bar{P}(z)$ and $\bar{P}(z) = z \frac{P(z) - 1}{z - 1}$, where $P(z)$ is the probability generating function of $X$.

b) $EX_{(n)} = P^{(n)}(1) = -\sum_{k=0}^{n-1} \frac{n!}{k!} (-1)^{n-k} \bar{P}^{(k)}(1)$, where $EX_{(n)} = EX(X-1)(X-2) \cdots (X-n+1)$ and $P^{(n)}$ denotes the $n^{th}$ derivative of $P$.

In particular, $EX = P'(1) = \bar{P}'(1)$ and $\text{Var}(X) = 2\bar{P}'(1) - \bar{P}(1) - (\bar{P}(1))^2$.

Proof.

a) We have

$$\bar{P}(z) = \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} P(X = k) z^j = \sum_{k=1}^{\infty} P(X = k) z^k,$$

rewriting the above formula yields a).
b) The first equality is a basic property of probability generating functions. The second equality follows by applying Leibniz’s formula for derivatives of products:

\[(f \cdot g)^{(n)} = \sum_{k=0}^{n} \binom{n}{k} f^{(k)} g^{(n-k)}\]

to a) and evaluating at \(z = 1\).

**B Shewhart charts**

In this appendix we derive upper and lower bounds for the ARL_out and the VRL_out of Shewhart control charts for residuals of correlated data studied in Section 4.1.1, to which we refer for unexplained notation. These bounds are based on first computing \(P_1, \ldots, P_n\), and then replacing the remaining \(p_k\)'s by upper and lower estimates. Fix \(n\) and let \(p_{\text{low}}\) and \(p_{\text{up}}\) be such that \(p_{\text{low}} \leq p_k \leq p_{\text{up}}\) for \(k > n\).

**Lemma B.1**

\[
1 + \sum_{j=1}^{n-1} \prod_{i=1}^{j} (1 - p_i) + \frac{\prod_{i=1}^{n} (1 - p_i)}{p_{\text{up}}} \leq \text{ARL}_{\text{out}} \leq 1 + \sum_{j=1}^{n} \prod_{i=1}^{j} (1 - p_i) + \frac{\prod_{i=1}^{n} (1 - p_i)}{p_{\text{low}}}.
\]

**Proof.**

Using \(p_{\text{low}} \leq p_k\) and (22), we obtain the following upper bound for \(\text{ARL}_{\text{out}}\):

\[
\text{ARL}_{\text{out}} = 1 + \sum_{j=1}^{n} \prod_{i=1}^{j} (1 - p_i) + \prod_{i=1}^{n} (1 - p_i) + \sum_{j=n+1}^{\infty} \prod_{i=1}^{j} (1 - p_i) \leq 1 + \sum_{j=1}^{n} \prod_{i=1}^{j} (1 - p_i) + \prod_{i=1}^{n} (1 - p_i) \sum_{j=n+1}^{\infty} (1 - p_{\text{low}})j
\]

\[
= 1 + \sum_{j=1}^{n} \prod_{i=1}^{j} (1 - p_i) + \prod_{i=1}^{n} (1 - p_i) \frac{1 - p_{\text{low}}}{p_{\text{low}}}
\]

The lower bound for \(\text{ARL}_{\text{out}}\) can be found in a similar way.

**Remarks**

a) Note that we may always obtain a lower bound, by choosing the trivial upper estimate \(p_{\text{up}} = 1\).
b) The following argument shows that the upper and lower bounds converge to the true ARL_{out} if \( n \) tends to \( \infty \). Since \( \lim_{i \to \infty} p_i > 0 \), it follows that the product \( \prod_{i=1}^{n} (1 - p_i) \) converges to 0 as \( n \) tends to \( \infty \). Thus the upper bound is bounded from above. Moreover, the difference between the upper and lower bound converges to 0 as \( n \) tends to \( \infty \). Putting everything together, we may conclude that both bounds converge to the true ARL_{out}.

Lemma B.2 \( LB(VRL_{out}) \leq VRL_{out} \leq UB(VRL_{out}) \), where

\[
LB(VRL_{out}) = 1 + \sum_{j=1}^{n-1} (2j+1) \prod_{i=1}^{j} (1 - p_i) + \left( \frac{2 + (2n-1)p_{up}}{p_{up}^2} \right) \prod_{i=1}^{n} (1 - p_i) - (UB(ARL_{out}))^2,
\]

\[
UB(VRL_{out}) = 1 + \sum_{j=1}^{n-1} (2j+1) \prod_{i=1}^{j} (1 - p_i) + \left( \frac{2 + (2n-1)p_{low}}{p_{low}^2} \right) \prod_{i=1}^{n} (1 - p_i) - (LB(ARL_{out}))^2,
\]

and \( LB(ARL_{out}) \) and \( UB(ARL_{out}) \) the lower and upper bound of the ARL_{out}, respectively (as given in lemma B.1).

Proof.
We only show how to obtain an upper bound. A lower bound can be found in a similar way. Using \( p_{low} \leq p_k \) for \( k > n \), it follows that:

\[
\sum_{j=1}^{\infty} j \prod_{i=1}^{j} (1 - p_i) = \sum_{j=1}^{n-1} j \prod_{i=1}^{j} (1 - p_i) + \sum_{j=n+1}^{\infty} j \prod_{i=1}^{j} (1 - p_i) + n \prod_{i=1}^{n} (1 - p_i)
\]

\[
\leq \sum_{j=1}^{n-1} j \prod_{i=1}^{j} (1 - p_i) + \left( \sum_{j=n+1}^{\infty} j \prod_{i=1}^{j} (1 - p_{low}) + n \right) \prod_{i=1}^{n} (1 - p_i)
\]

\[
= \sum_{j=1}^{n-1} j \prod_{i=1}^{j} (1 - p_i) + \left( \sum_{j=1}^{\infty} j (1 - p_{low})^{j-1} + n \sum_{j=1}^{\infty} (1 - p_{low})^{j} + n \right) \prod_{i=1}^{n} (1 - p_i)
\]

Substituting this result into (23) and using Lemma B.1, we obtain UB(VRL_{out}) after some simplification. \( \square \)
Remark In a similar way as for the bounds of the ARL_{out}, we may show that the bounds of the variance converge to the true variance.

**Lemma B.3** For an AR(p)-process, we have:

$$ARL_{out}(AR(p)) = 1 + \sum_{j=1}^{p} \prod_{i=1}^{j} (1 - p_i) + \prod_{i=1}^{p+1} \frac{(1 - p_i)}{p_{p+1}}$$

and

$$VRL_{out}(AR(p)) = 1 + \sum_{j=1}^{p} (2j + 1) \prod_{i=1}^{j} (1 - p_i) + \left(\frac{2 + (2p + 1)p_{p+1}}{p_{p+1}^2}\right) \prod_{i=1}^{p+1} (1 - p_i) - (ARL_{out})^2.$$

**Proof.** Recall that for AR(p)-processes $p_k = p_{p+1}$ for $k > p$. Now apply Lemmas B.1 and B.2 with $n = p$, $p_{low} = p_{up} = p_{p+1}$, and simplify the formula for the VRL. \hfill $\square$

## C CUSUM charts

In this appendix we derive formulas for the ARL of CUSUM control charts for residuals of ARMA(p, q)-processes. This extends the results of Runger et al. (1995) who calculated the ARL for the AR(p)-case using a Markov chain approach. Our approach uses integral equations in the spirit of Page (1954). The Markov chain approach used in Runger et al. (1995) is a special choice of discretization of these integral equations. With the results of Yashchin (1985) it is in principle possible to extend our approach to calculate the SRL as well. We refrained from deriving formulas for the SRL, since these formulas are very complicated and hence, possibly numerically unstable.

We assume that a shift of size $\delta$ occurs at time $c$. Define $P_{H,\delta}(c + i, s, n), i = 0, 1, \ldots$, to be the probability that it takes $n$ periods after time $c + i$ before the shift is detected, given $S_H(c + i) = s$.

For an MMSE controller with $b = 1$ it follows from (19) that

$$U_c \sim N(\delta, \sigma^2) \quad \text{and} \quad U_{c+i} \sim N((1 - \xi_1 - \cdots - \xi_{c+i})\delta, \sigma^2) \quad \text{for} \quad i \geq 1.$$ 

We denote the normal density and cumulative distribution function of $U_{c+i}$ by $f_{U_{c+i}}$ and $F_{U_{c+i}}$, respectively.

**Lemma C.1** For $i = 0, 1, \ldots$, we have:

$$P_{H,\delta}(c + i, s, 1) = 1 - F_{U_{c+i+1}}(H + K - s)$$

and

$$P_{H,\delta}(c + i, s, n) = P_{H,\delta}(c + i + 1, 0, n - 1)F_{U_{c+i+1}}(K - s) + \int_0^H P_{H,\delta}(c + i + 1, y, n - 1)f_{U_{c+i+1}}(y + K - s)dy.$$
Proof.

For $n = 1$ we have

\[ P_{H,\delta}(c+i,s,1) = \text{Prob}(S_H(c+i) \geq H \mid S_H(c+i) = s) = \text{Prob}(U_{c+i+1} \geq H + K - s) = 1 - F_{U_{c+i+1}}(H + K - s). \]

For $n > 1$, we condition on $S_H(c+i+1)$. We have to distinguish between $S_H(c+i+1) = 0$ and $0 < S_H(c+i+1) < H$. This yields:

\[
P_{H,\delta}(c+i,s,n) = P_{H,\delta}(c+i+1,0,n-1)F_{U_{c+i+1}}(K-s) + \int_0^H P_{H,\delta}(c+i+1,y,n-1)f_{U_{c+i+1}}(y+K-s)\,dy,
\]

as required. \hfill \Box

In order to obtain moments of the run length, we now derive integral equations for the following generating function:

\[
M_{H,\delta}(c+i,s,x) := \sum_{n=1}^{\infty} P_{H,\delta}(c+i,s,n)e^{nx}, i = 0,1,\ldots
\]

The $n^{th}$ moment of the run length given $S_H(c+i) = s$, equals the $n^{th}$ derivative of the moment generating function $M_{H,\delta}(c+i,s,x)$ with respect to $x$ evaluated at $x = 0$. Note that knowledge of the first two moments yields the variance.

**Lemma C.2** For $i = 0,1,\ldots$, we have:

\[
e^{-x}M_{H,\delta}(c+i,s,x) = 1 - F_{U_{c+i+1}}(H + K - s) + M_{H,\delta}(c+i+1,0,x)F_{U_{c+i+1}}(K-s) + \int_0^H M_{H,\delta}(c+i+1,y,x)f_{U_{c+i+1}}(y+K-s)\,dy.
\]

Proof. Substitute lemma C.1 into the definition of $M_{H,\delta}(c+i,s,x)$.

**Lemma C.3** For $i = 0,1,\ldots$, we have:

\[
ARL_{H,\delta}(c+i,s) = 1 + ARL_{H,\delta}(c+i+1,0)F_{U_{c+i+1}}(K-s) + \int_0^H ARL_{H,\delta}(c+i+1,y)f_{U_{c+i+1}}(y+K-s)\,dy
\]

and

\[
Mom_{2,H,\delta}(c+i,s) = -1 + 2ARL_{H,\delta}(c+i,s) + Mom_{2,H,\delta}(c+i+1,0)F_{U_{c+i+1}}(K-s) + \int_0^H Mom_{2,H,\delta}(c+i+1,y)f_{U_{c+i+1}}(y+K-s)\,dy.
\]
Proof. Differentiate the results of lemma C.2 with respect to $x$ and set $x = 0$.

**Lemma C.4** For an AR($p$)-process and an MMSE controller, we have:

$$A_{RLH, \delta}(c+i, s) = 1 + A_{RLH, \delta}(c+i+1, 0)F_{U_{c+i+1}}(K-s) + \int_0^H A_{RLH, \delta}(c+i+1, y)f_{U_{c+i+1}}(y+K-s) dy$$

with $U_{c+i} \sim N((1 - \varphi_1 - \cdots - \varphi_i)\delta, \sigma_\delta^2)$ for $i = 0, 1, \ldots, p-1$, and

$$A_{RLH, \delta}(c+i, s) = 1 + A_{RLH, \delta}(c+p, 0)F_{U_{c+p}}(K-s) + \int_0^H A_{RLH, \delta}(c+p, y)f_{U_{c+p}}(y+K-s) dy$$

with $U_{c+p} \sim N((1 - \varphi_1 - \cdots - \varphi_p)\delta, \sigma_\delta^2)$ for $i \geq p$.

Proof. Recall that for ARL($p$)-processes $\xi_i = \varphi_i$ for $i = 1, \cdots, p$, $\xi_i = 0$ for $i > p$ and $U_{c+i} \overset{d}{=} U_{c+p}$ for $i \geq p$. Hence, $A_{RLH, \delta}(c+i, s) = A_{RLH, \delta}(c+p, s)$ for $i \geq p$. Substituting this result into lemma C.3 we obtain our result.

References


