Research Report
Statistical Research Unit
Department of Economics
University of Gothenburg
Sweden

Minimax Optimality of CUSUM for an Autoregressive Model

Knoth, S. and Frisén, M.

Minimax Optimality of CUSUM for an Autoregressive Model

Sven Knoth[*]

*Institute of Mathematics and Statistics, Department of
Economics and Social Sciences, Helmut Schmidt
University Hamburg, PO Box 700822, 22008 Hamburg,
Germany*

Marianne Frisén[†]

*Statistical Research Unit, Department of Economics,
University of Gothenburg, PO Box 640, SE 405 30
Göteborg, Sweden*

Different change point models for $AR(1)$ processes are reviewed. For some models, the change is in the distribution conditional on earlier observations. For others the change is in the unconditional distribution. Some models include an observation before the first possible change time — others not. Earlier and new CUSUM type methods are given and minimax optimality is examined. For the conditional model with an observation before the possible change there are sharp results of optimality in the literature. The unconditional model with possible change at (or before) the first observation is of interest for applications. We examined this case and derived new variants of four earlier suggestions. By numerical methods and Monte Carlo simulations it was demonstrated that the new variants dominate the original ones. However, none of the methods is uniformly minimax optimal.

Keywords and Phrases: Autoregressive; Change point; Monitoring; Online detection.

[a]Sven.Knoth@hsu-hh.de
[b]Marianne.Frisen@statistics.gu.se

1 Introduction

The aim of sequential surveillance is the timely detection of important changes in the process that generates the data.

Surveys and bibliographies on statistical surveillance are given for example by Lai (1995), who gives a full treatment of the field and concentrates on the minimax properties of stopping rules, by Woodall and Montgomery (1999) and Ryan (2000), who concentrate on control charts, and by Frisén (2009) and the following discussion who consider optimality for different applications.

2 Specifications and notations

The variable under surveillance depends on the application. We denote the process by $X = \{X(t) : t = 0, 1, 2, \ldots\}$, where $X(t)$ is the observation made at time t, which is here discrete.

The purpose of the monitoring is to detect a possible change. The time for the change is denoted by τ. We consider only one change. Before the change, the distribution belongs to the family f^0, and after the time τ, the distribution belongs to the family f^1. The change point τ can be regarded either as a random variable or as a deterministic but unknown value, depending on what is most relevant for the application. Here, we use the latter approach.

At each decision time s, we want to discriminate between the two events, $C(s)$ and $D(s)$. For most applications, these can be further specified as $C(s) = \{\tau \leq s\}$ (at or before the decision time, there has been a change) and $D(s) = \{\tau > s\}$ (at the decision time, no change has occurred yet), respectively. The (possibly random) process that determines the state of the system is denoted by $\mu(t)$. This is here a parameter in the distribution. We consider a step change, where a parameter changes from one constant level, say $\mu(t) = \mu^0$, to another constant level, $\mu(t) = \mu^1$. Then, we have $\mu(t) = \mu^0$ for $t = 0, \ldots, \tau - 1$ and $\mu(t) = \mu^1$ for $t \geq \tau$.

We use the observations $X_s = \{X(t) : t \leq s\}$ to form an alarm criterion which, when fulfilled, is an indication that the process is in state $C(s)$, and an alarm is triggered. We use an alarm statistic, $H(X_s)$, and a control limit, $L(s)$, where the time of an alarm, t_A, is

$$t_A = \min \left\{ s : H(X_s) > L(s) \right\}.$$

Here, only the properties up to the first alarm will be considered.

3 Models

Both models where the change from one distribution to another is in the conditional distribution (on the earlier distribution) and in the unconditional distribution are of interest. Since an $AR(1)$ process depends on an earlier observation, the case of no observation before the possible change time will be treated separately. In the sequel it is assumed that the correlation coefficient, α, and the variance of the observations are known.

3.1 U: Unconditional Change Point Models

Here we have one (unconditional) distribution for X before the change and another after. The distribution $f^0\big(X(t)\big)$ is true for $t < \tau$ and another $f^1\big(X(t)\big)$ for $t \geq \tau$. This framework was already deployed in Goldsmith and Whitfield (1961) and Bagshaw and Johnson (1975). Essentially, the shift at the change point τ is added to the unconditional mean of the data.

3.1.1 UA: At Least One Observation Before the Possible Change

The available observations are $X(0), X(1), \ldots$ The values of τ which we consider are $1, 2, \ldots$ We will study the model UA:

$$X(t) - \mu(t) = \alpha\big(X(t-1) - \mu(t-1)\big) + \varepsilon(t),$$

where the independent $\varepsilon(t) \sim \mathcal{N}(0, \sigma^2)$ and there is a step change in the unconditional mean $\mu(t)$,

$$E\big(X(t)\big) = \mu(t) = \begin{cases} 0 & , \ t < \tau \\ \delta & , \ t \geq \tau \end{cases}.$$

Thus, the distribution for $X(0)$ has mean zero. The (unconditional) variance for all observations is

$$Var\big(X(t)\big) = \gamma_0 = \sigma^2/(1 - \alpha^2).$$

Since we have one observation for $t = 0$, before the possible change, and thus $E\big(X(0)\big) = 0$, we have

$$X(1) - \mu(1) = \alpha X(0) + \varepsilon(1).$$

Observe that though this is a change between the unconditional distributions, the conditional distribution has two change points (τ and $\tau + 1$) and the conditional

problem does not fit in the usual surveillance theory for one change point. The conditional mean is

$$
E\big(X(t)\big|X(t-1)\big) = \begin{cases} \alpha X(t-1) & , t < \tau \\ \alpha X(t-1) + \delta & , t = \tau \\ \alpha X(t-1) + (1-\alpha)\delta & , t > \tau \end{cases}.
$$

Yashchin (1993) mentioned this specific pattern for $t = \tau$ at the end of his Section 2. He considered this contribution as a head start for the CUSUM sequence right at the change point. See also Harris and Ross (1991) and Atienza et al. (1998) for some more references in SPC literature on the use of the unconditional model. Another way to express this UA model (see, e. g., Schmid, 1997) is

$$
X(t) = \alpha X(t-1) + \varepsilon(t)\,,
$$

$$
Y(t) = \begin{cases} X(t) & , t < \tau \\ X(t) + \delta & , t \geq \tau \end{cases}.
$$

For $Y(t) = X(t) + d(t)$ we get

$$
\begin{aligned} Y(t) &= \alpha X(t-1) + \varepsilon(t) + d(t) = \alpha\big(Y(t-1) - d(t-1)\big) + \varepsilon(t) + d(t) \\ &= \alpha Y(t-1) + d(t) - \alpha d(t-1) + \varepsilon(t) \end{aligned}
$$

with

$$
d(t) = \begin{cases} 0 & , t < \tau \\ \delta & , t \geq \tau \end{cases}.
$$

Note that case UA is the framework deployed quite recently in Han and Tsung (2009b).

3.1.2   UB: No Observation Before the Possible Change

This is the most frequently discussed model in the literature and it is relevant for many applications.

Here we need a special specification of the distribution of the first observation. It is common, as in, e. g., Knoth and Schmid (2004), to rely on the marginal distribution and assume

$$
X(1) \sim \mathcal{N}\big(\mu(1), \gamma_0\big)\,.
$$

In Section 5 we present several CUSUM type methods for this model.

3.2   C: Conditional Change Point Models

Here the conditional distribution $f^0\big(X(t)\big|X(t-1)\big)$ is true for $t < \tau$ and another $f^1\big(X(t)\big|X(t-1)\big)$ for $t \geq \tau$.

3.2.1 CA: At Least One Observation Before the Possible Change

By deploying $d(t)$ directly we get immediately to the AR(1) model with a potential change in the conditional distribution. Thus, by writing

$$X(t) - d(t) = \alpha\big(X(t-1) - d(t)\big) + \varepsilon(t)$$

we conclude that

$$X(t) = \alpha X(t-1) + (1-\alpha)d(t) + \varepsilon(t)\,,$$

and finally for the conditional expected value

$$E\big(X(t)\big|X(t-1)\big) = \begin{cases} \alpha X(t-1) & ,\ t < \tau \\ \alpha X(t-1) + (1-\alpha)\delta & ,\ t \geq \tau \end{cases}.$$

Note that $E\big(X(t)\big) = \alpha E\big(X(t-1)\big) + (1-\alpha)d(t)$ so that $E\big(X(t)\big) = \alpha E\big(X(t-1)\big)$ for $t < \tau$. This is stationary only for a zero mean process. Additionally, $E\big(X(\tau)\big) = \delta$ and $E\big(X(\tau+i)\big) = (1-\alpha^{i+1}) \times \delta$ for $i = 0, 1, \ldots$ Hence, the series $\{X(t)\}_{t=\tau}^{\infty}$ is not stationary anymore. This demonstrates the great difference between the conditional and unconditional models. They will be appropriate for different applications.

3.2.2 CB: No Observation Before the Possible Change

Like for UB we would need some special definition for $X(1)$. We have not seen any work on this case.

4 Minimax optimality

A minimax solution, with respect to τ, avoids the requirements of information about the distribution of τ. Lorden (1971) and later Moustakides (1986) treat the "worst possible case", by using not only the least favorable value of change time, τ, but also the least favorable outcome of $X_{\tau-1}$ before the change occurs. Their minimax criterion is the minimum of:

$$\mathcal{W} = \sup_{\tau \geq 1} \operatorname{ess\,sup} E_\tau\big\{[t_A - \tau + 1]^+ | X_{\tau-1}\big\}\,,$$

where $E_\tau()$ denotes the expectation for the change point position τ. The minimization is for a fixed value of $\mathrm{ARL}^0 = E_\infty(t_A)$, where $\tau = \infty$ denotes the in control case where no change ever appears. Formulas for easier calculation of $\mathcal{W}$ are described in Appendix B. Numerical methods for approximate calculation are described in Appendix C.

5 The CUSUM method

5.1 CUSUM for general models

The CUSUM method was first suggested by Page (1954) and is reviewed for example in the book by Hawkins and Olwell (1998). The alarm condition of the method, for detection of a change at time τ from distribution f^0 to distribution f^1, can be expressed by the partial likelihood ratios as (this is simply the transfer of Page (1954) or Siegmund (1985) to dependent observations – see Subsection A.1 for some details regarding Page's definitions)

$$ t_A = \min \left\{ s : \max_{1 \leq t \leq s} L(s,t) > G \right\}, \tag{1} $$

where G is a constant and

$$ L(s,t) = \frac{f^1(X_s | \tau = t)}{f^0(X_s | \tau = t)} . $$

The method is sometimes called *the* likelihood ratio method, but this combination of likelihood ratios should not be confused with the full likelihood ratio method, which is a weighted average of the partial likelihood ratios. For the optimality properties of methods constructed by different combinations of the partial likelihood ratios see Frisén (2003).

For *iid* data the CUSUM method satisfies the minimax criterion of optimality described in Section 4. Other positive properties of the method have been confirmed for example by Srivastava and Wu (1993). It was demonstrated by Frisén and Wessman (1999) that the CUSUM method works almost as well with respect to the expected delay as the full likelihood ratio method and Shiryaev-Roberts methods which are constructed to meet the expected delay criterion.

Lai (1995) and Lai and Shan (1999) point out that the general structure and the good minimax properties of generalizations of the CUSUM method make this technique suitable for complicated problems.

5.2 CUSUM for autocorrelated Data

In many applications, there are time dependencies. For examples in finance see, e. g., Okhrin and Schmid (2007a) and Okhrin and Schmid (2007b). For an example of an environmetric application see Pettersson (1998). The theory of surveillance of dependent data is not simple. The most common approach to surveillance of models with dependencies is to adjust the alarm limit of an ordinary method, so that the false alarm risk is controlled, see, e. g., Schmid and Schöne (1997). Another approach is to monitor the process of residuals as in, for example, Kramer

and Schmid (1997). It is natural to modify the CUSUM method by use of the conditional likelihood (conditioned on earlier observations). Several suggestions about how to handle the first statistic and how to get less computational burden are given. For discussions about different variants of CUSUM for dependent data see, e.g., Nikiforov (1979), Basseville and Nikiforov (1993), Yashchin (1993), Schmid (1997), and Lai (1998).

For the model CA we have

$$
L(s,t) = \frac{f^0\big(X(0)\big) \times \prod_{i=1}^{t-1} f^0\big(X(i)\big|X(i-1)\big) \times \prod_{i=t}^{s} f^1\big(X(i)\big|X(i-1)\big)}{f^0\big(X(0)\big) \times \prod_{i=1}^{s} f^0\big(X(i)\big|X(i-1)\big)}
$$

$$
= \exp\left\{ \frac{1}{2\sigma^2} \sum_{i=t}^{s} -\big(X(i) - \alpha X(i-1) - (1-\alpha)\delta\big)^2 \right. \tag{2}
$$

$$
\left. + \big(X(i) - \alpha X(i-1)\big)^2 \right\}
$$

with $t = 1, 2, \ldots, s$. For the model CB we have the same expression for $t > 1$ but

$$
L(s,1) = \frac{f^1\big(X(1)\big)}{f^0\big(X(1)\big)} = \exp\left\{ \frac{1}{2\gamma_0} \left[-\big(X(1) - \delta\big)^2 + \big(X(1)\big)^2 \right] \right\} \tag{3}
$$

For the model UA, we have a third distribution $f^{0,1}$ besides f^0 and f^1.

$$
L(s,t) = \frac{f^{0,1}\big(X(t)\big|X(t-1)\big) \times \prod_{i=t+1}^{s} f^1\big(X(i)\big|X(i-1)\big)}{\prod_{i=t}^{s} f^0\big(X(i)\big|X(i-1)\big)}
$$

$$
= \exp\left\{ \frac{1}{2\sigma^2} \left[-\big(X(t) - \alpha X(t-1) - \delta\big)^2 + \big(X(t) - \alpha X(t-1)\big)^2 \right.\right.
$$

$$
\tag{4}
$$

$$
\left.\left. + \sum_{i=t+1}^{s} -\big(X(i) - \alpha X(i-1) - (1-\alpha)\delta\big)^2 + \big(X(i) - \alpha X(i-1)\big)^2 \right] \right\}
$$

For the model UB we have the same expression as for UA for $t > 1$ but $L(s,1)$ is the same as for CB. This statistic combined with the CUSUM rule (1) is in the sequel named method **M1**.

For all the models the CUSUM statistic is based on

$$\max_{t=1,2,\ldots,s} L(s,t)\,.$$

For application and numerical evaluations, other equivalent expressions are more convenient. We define

$$H(s) = \max_{t=1,2,\ldots,s} \frac{\sigma^2 \log L(s,t)}{\delta} \tag{5}$$

as the CUSUM statistic. The same operation applied to G leads to the alarm limit $h = \sigma^2 \log G/\delta$. The iteration rules of $H(s)$ for the different CUSUM methods are derived in Appendix A.

5.3 Alternative CUSUM-like Methods for Model UB

Several alternatives to method **M1** are suggested here and in the literature. The statistics of **M1** described in Section 5.2 by (3) and (4) can be expressed as

$$
\begin{aligned}
H(1) &= (1-\alpha^2)\big(X(1)-k\big)\,, \\
H(s) &= \max\big\{H(s-1) + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big), \tilde{\varepsilon}(s) - k\big\}
\end{aligned}
$$

where $\tilde{\varepsilon}(s) = X(s) - \alpha X(s-1)$ and $k = \delta/2$. As is seen in Appendix A, this leads to a sequence with reflecting barrier

$$
H^*(s) = \max\{H(s), z_r\} \quad \text{with} \quad z_r = \begin{cases} -\alpha k & ,\ \alpha > 0 \\ \alpha h - \alpha(1-\alpha)k & ,\ \alpha < 0 \end{cases}
$$

and

$$
H^*(s) = \max\big\{H^*(s-1) + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big), \tilde{\varepsilon}(s) - k, z_r\big\}\,.
$$

The alarming behavior is, for $h > 0$, equivalent to that of **M1**. Reflection at 0, the usual CUSUM lower barrier and deployed in Knoth and Schmid (2004), leads to

$$
\begin{aligned}
H^0(1) &= \max\{H(1), 0\}\,, \\
H^0(s) &= \max\big\{H^0(s-1) + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big), \tilde{\varepsilon}(s) - k, 0\big\}\,.
\end{aligned}
$$

$H^0(s)$ does not exhibit the same alarming behavior as $H(s)$ and $H^*(s)$. Note that in Schmid (1997) the iteration rule for $H(s)$ was proved, while in Knoth and Schmid (2004) it was wrongly assumed (without proving it) that $H^0(s) = \max\{H(s), 0\}$. The method based on the statistic H^0 in combination with the CUSUM alarm rule is here named **M4**.

The next two competing CUSUM methods are quite similar. If we would consider the UA instead of the UB framework, then they would coincide. Essentially, they differ in the treatment of the first observation. They stem from concepts such as introducing CUSUM as sequence of repeated SPRTs as in Basseville and Niki-forov (1993) or deploying the standard CUSUM to the (empirically recaptured) residuals, $\tilde{\varepsilon}(s)$, like in Runger et al. (1995). To get the method based on repeated SPRTs, in the sequel called **M2**, apply the CUSUM method derived for CB – based on (2) and (3) – to UB. This leads to the same sequence for $s > 1$ as for the repeated SPRTs and the same choice for $s = 1$ as for method **M4**.

$$O(1) = H^0(1) = \max\{(1 - \alpha^2)(X(1) - k), 0\},$$
$$O(s) = \max\{O(s-1) + (1-\alpha)(\tilde{\varepsilon}(s) - (1-\alpha)k), 0\}.$$

Eventually, the residual CUSUM derived as in Knoth and Schmid (2004) — see also Harris and Ross (1991), Runger et al. (1995), or Wieringa (1999) for earlier works on residual CUSUM schemes — provides method **M3**. Here we have

$$R(1) = \max\left\{(1-\alpha^2)\left(\sqrt{\frac{1+\alpha}{1-\alpha}}X(1) - \frac{1-\alpha}{1+\alpha}k\right), 0\right\}$$

and

$$R(s) = \max\{R(s-1) + (1-\alpha)(\tilde{\varepsilon}(s) - (1-\alpha)k), 0\}.$$

The basic idea of the residual schemes is to apply a standard method to an *iid* stream of new variables that are normalized in the in-control.

For all methods described above, their worst-case performance differs between $\tau = 1$ and $\tau > 1$. We will examine modifications of their first CUSUM statistic which makes the worst-case performance equal between $\tau = 1$ and $\tau > 1$. The basic idea is to mimic at $\tau = 1$ the worst-case for $\tau > 1$. The latter is based on the minimal value for the last CUSUM value prior the change point τ. For **M1** it is z_r, while for the other methods it is 0. Fortunately, the value $H^*(\tau - 1) = z_r$ yields the following nice property:

$$H^*(\tau) = \max\{z_r + (1-\alpha)(\tilde{\varepsilon}(\tau) - (1-\alpha)k), \tilde{\varepsilon}(\tau) - k, z_r\} = \max\{\tilde{\varepsilon}(\tau) - k, z_r\}.$$

See (11) and (12) in the Appendix and recall the specific structure of z_r for $\alpha > 0$ and $\alpha < 0$. Thus, it is sufficient to equip $H^*(1)$ for $\tau = 1$ with a similar distribution as the one of $\tilde{\varepsilon}(\tau) - k$ for $\tau > 1$. Consequently, replacing $H(1) = \max\{(1-\alpha^2)(X(1) - k), z_r\}$ by

$$\max\left\{\sqrt{1-\alpha^2}\left(X(1) - \left[2 - 1/\sqrt{1-\alpha^2}\right]k\right), z_r\right\}$$

9

does the trick. Recall also the singular nature of the residual $\tilde{\varepsilon}(\tau)$: $E_1\big(X(1)\big) = E_\tau\big(\tilde{\varepsilon}(\tau)\big) = \delta$ while $E_\tau\big(\tilde{\varepsilon}(s)\big) = (1-\alpha)\delta$ for $s > \tau$, and $\delta/2 = k$. Denote the new method with $\mathbf{M1}^e$.

For $\mathbf{M2}$ and $\mathbf{M3}$ it is simpler. Now, the new first value of the CUSUM sequence should behave like $(1-\alpha)\big(\tilde{\varepsilon}(\tau) - (1-\alpha)k\big)$. Hence, replace $O(1)$ and $R(1)$ with

$$\max\left\{(1-\alpha)\sqrt{1-\alpha^2}\Big(X(1) - \big[2 - (1+\alpha)/\sqrt{1-\alpha^2}\big]k\Big), 0\right\}.$$

Originally, $\mathbf{M2}$ and $\mathbf{M3}$ differed only for the first CUSUM statistic. Thus, the above modification leads only to one new CUSUM method. Denote it with $\mathbf{M2}^e$. The most subtle case of modification is for method $\mathbf{M4}$. Here, we have to distinguish between $\alpha > 0$ and $\alpha < 0$. For positive α, the contribution of $(1-\alpha)\big(\tilde{\varepsilon}(\tau) - (1-\alpha)\big)$ dominates $\tilde{\varepsilon}(\tau) - k$ for $\tilde{\varepsilon}(\tau) < (2-\alpha)k$. Furthermore, for $\tilde{\varepsilon}(\tau) < (1-\alpha)k$ the reflecting barrier is the maximum. On the contrary, for negative α it becomes dominant for $(2-\alpha)k < \tilde{\varepsilon}(\tau)$. Eventually, $H^0(1)$ is replaced by

$$\max\Big\{\ \sqrt{1-\alpha^2}\Big(X(1) - \big[2 - 1/\sqrt{1-\alpha^2}\big]k\Big),$$
$$(1-\alpha)\sqrt{1-\alpha^2}\Big(X(1) - \big[2 - (1+\alpha)/\sqrt{1-\alpha^2}\big]k\Big), z_r\Big\}.$$

This scheme, the last one on our list, is denoted with $\mathbf{M4}^e$.

Table 1: CUSUM type schemes for UB

code	description	Reference (example)
$\mathbf{M1}$	CUSUM defined by likelihood ratio	Schmid (1997)
$\mathbf{M2}$	Repeated SPRTs	Basseville and Nikiforov (1993)
$\mathbf{M3}$	Residual CUSUM	Runger et al. (1995)
$\mathbf{M4}$	modified $\mathbf{M1}$ — reflection at 0	Knoth and Schmid (2004)
$\mathbf{M}*^e$	modified first statistic	This section

6 Minimax properties of the CUSUM method

Lai (1998) gave results under certain restrictions for asymptotic minimax of the CUSUM method when the average run length tends to infinity. He exemplified with a Markov chain. This gives proofs for the asymptotic minimax optimality of the model CA since the restrictions are fulfilled. Han and Tsung (2009a) proved

the minimax optimality under certain other restrictions but without the require-
ment of the average run length tending to infinity. They exemplified with an AR
model with a bounded state space and found that the requirements are satisfied.
Theoretical results for the unconditional models are hard to achieve since we can-
not rely on the extensive theory of surveillance for one change in distribution as
explained in Section 3.1. However, case UB is of great interest for application.
Thus we report results by numerical approximations and Monte Carlo simulations.

6.1 Effect of the ^{e}-Adjustment

The ^{e}-versions described in Section 5.3 are uniformly better than the original ones
as seen in Figure 1. This can heuristically be explained. We change the first
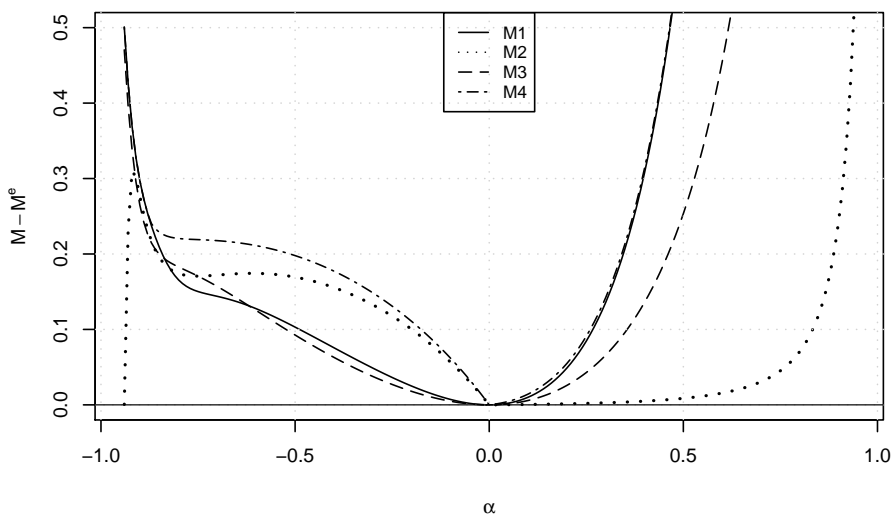

Figure 1: Difference in $\mathcal{W}$ between the ^{e}-versions and the original versions.

CUSUM statistics to ensure $\mathcal{W}_1 = \mathcal{W}_{>1}$ for the considered shift. This leads to
only a tiny change of h (see Figure 2), since this is dominated by what happens
for later times. Note that **M2** is the only scheme, where $\mathcal{W}_1$ is smaller than $\mathcal{W}_{>1}$.
The change is in the direction which makes the worst of $\mathcal{W}_1$ and $\mathcal{W}_{>1}$ better. Thus
$\mathcal{W}$ gets better.
Since our results are numerical, we support our claim by demonstrating that the
differences are statistically significant at a Monte Carlo study. Thus the difference
is not due to random errors induced by the Monte-Carlo study. Whether the dif-
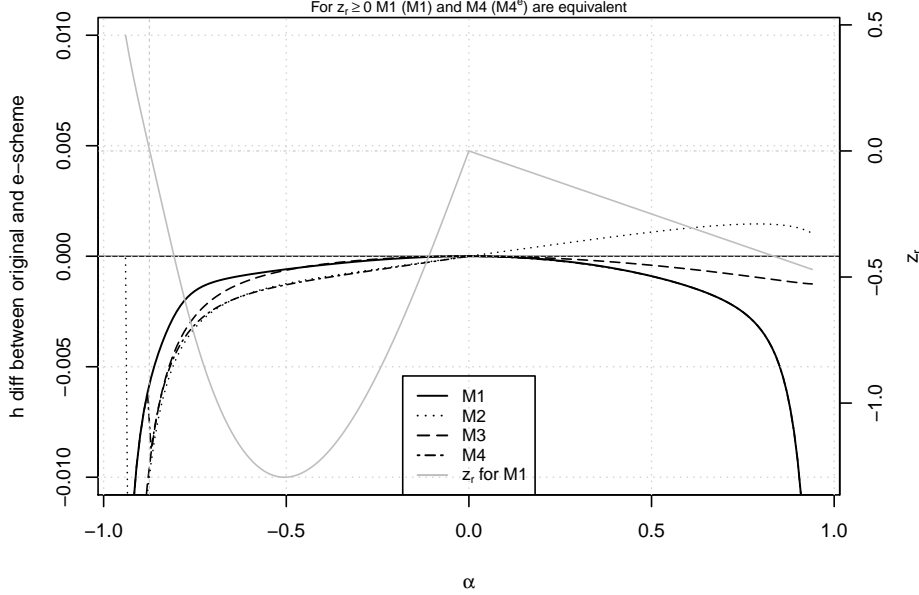ference is of significant importance for a specific application is another story. This

11

Figure 2: Changes in the alarm threshold from **M*** to **M***^{e}.

depends heavily on the value of α.

We describe the technique in detail for one case. To demonstrate that **M1** is worse than **M1**^{e} for some value of α (in this case $\alpha = -0.65$), we chose alarm limits numerically such that the ARL^0 is larger for **M1**^{e} than for **M1**. We check by simulations that this is so with a p-value less than 1%. The related results, based on 10^9 replicates, are $\widehat{\text{ARL}}^0_{\textbf{M1}} = 499.9837$ with corresponding standard error 0.015747 for $h_1 = 4.397069$ and $\widehat{\text{ARL}}^0_{\textbf{M1}^e} = 500.6255$ with standard error 0.015793 for $h_{1e} = 4.4$ for **M1** and **M1**^{e}, respectively. Thus, the ARL$^0_{\textbf{M1}^e}$ is significantly larger, on the 1 % level, than ARL$^0_{\textbf{M1}}$. Next, we test on the 1% level that the value of $\mathcal{W}$ is larger for **M1** than for **M1**^{e} for the selected value of α. Here we got, for 10^{11} replicates, $\widehat{\mathcal{W}}_{\textbf{M1}} = 3.367583$ with standard error 0.000004 and $\widehat{\mathcal{W}}_{\textbf{M1}^e} = 3.231286$ with standard error 0.000004 for **M1** and **M1**^{e}, respectively. The p-value for a test of equal (or better) value of $\mathcal{W}$ for **M1** (for the same or less ARL^0) is thus at most $1 - 0.99^2 = 0.02$. For the considered example, the p-value would be much smaller.

6.2 Comparison Between the Different ^{e}-Methods

In order to give evidence that none of the methods **M1**^{e}, **M2**^{e}=**M3**^{e} or **M4**^{e} is optimal with respect to $\mathcal{W}$ uniformly over α, we chose interesting values of α based on the numerical results in Figure 3. To show that **M1**^{e} is not optimal we

demonstrated that $\mathcal{W}$ is significantly worse than $\mathbf{M2}^e$=$\mathbf{M3}^e$ for $\alpha = 0.65$ in spite of a significantly less ARL^0. We got the result $\mathcal{W} = 27.639$ for $\mathbf{M2}^e$=$\mathbf{M3}^e$ with slightly increased h (ARL^0=500.40), and $\mathcal{W} = 27.686$ for $\mathbf{M1}^e$ (ARL^0=500.00), all with a standard error smaller than 0.0001. To show that $\mathbf{M2}^e$=$\mathbf{M3}^e$ is not optimal we demonstrated that $\mathcal{W}$ is significantly worse than $\mathbf{M1}^e$ for $\alpha = -0.65$. The result was $\mathcal{W} = 3.2570$ for $\mathbf{M2}^e$=$\mathbf{M3}^e$ (ARL^0=500.00) and $\mathcal{W} = 3.2313$ for $\mathbf{M1}^e$ (ARL^0=500.84), all with a standard error smaller than 0.00001. Eventually, to indicate that $\mathbf{M4}^e$ is not optimal we demonstrated that $\mathcal{W}$ is significantly worse than $\mathbf{M1}^e$ for $\alpha = -0.65$ ($\mathcal{W} = 3.2404$ for $\mathbf{M4}^e$). Thus none of the ^{e}-methods is uniformly optimal with respect to $\mathcal{W}$. See Figure 3.
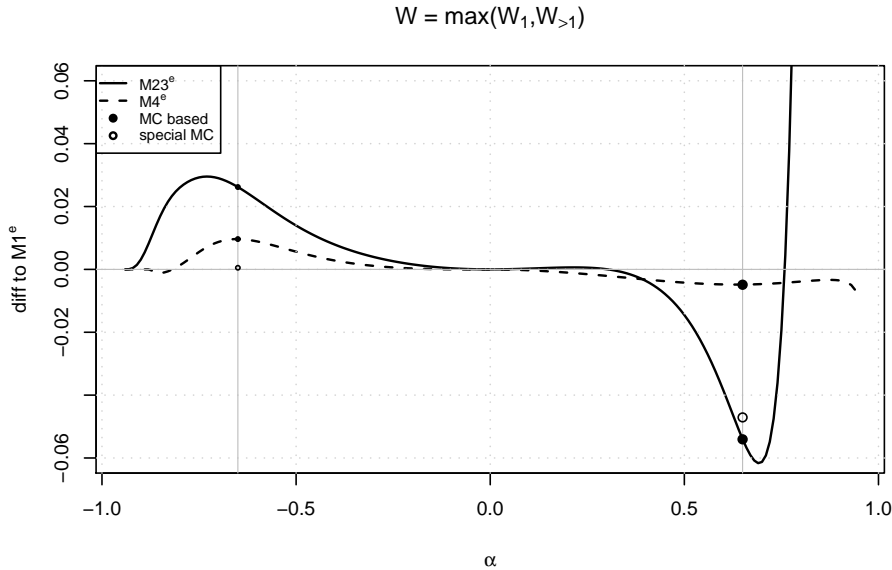

Figure 3: Differences to the values of $\mathcal{W}$ of $\mathbf{M1}^e$ for the methods $\mathbf{M2}^e$=$\mathbf{M3}^e$ and $\mathbf{M4}^e$, respectively. The lines are numerical results. The circles are results by Monte-Carlo simulations. The diameters of the circles are roughly equal to 6 times the standard error. The solid circles are for the schemes with calibrated ARL^0. The empty circles are for the special schemes with higher values of ARL^0 for the method with the least value of $\mathcal{W}$, to allow the empirical proof of lack of uniform optimality, given in the text.

7 Discussion

An advantage of the conditional C-models is that there is a change between two distributions while there is a change between three ones in the unconditional mod-

13

els. An advantage with the A-models where an observation before the possible change is available is that no special test is necessary for the first decision but all are motivated by the same conditional likelihood ratio. In an application we might know that the system is in control at $t = 0$ and we look for possible future changes. For this CA model, sharp optimality results are available.

However, this CA model is not suitable for all applications. Model UB is commonly used in practice – see, for example, Runger et al. (1995), Schmid (1997), Wieringa (1999), Sparks (2000). In an application, it might be possible that the change has happened before the monitoring is started.

A drawback as concerns possibilities of analytical uniform results for model UB is that the first statistic gives little information for large α, and the combination with an aim to detect also $\tau = 1$ makes the problem unbalanced with a great emphasis on $\tau = 1$. One determines the alarm limit for ARL^0 where the behavior for large values of s dominates but the worst case is often for $\tau = 1$, for which small values of s dominate.

We used statistical inference based on a simulation study to prove our main results. In Section 6.1 we demonstrated that the four earlier methods are dominated by the ^{e}-versions and in Section 6.2 we demonstrated that none of the ^{e}-methods is uniformly best. Thus the conclusion is that none of the seven methods examined is uniformly (with respect to α) minimax optimal but that improvement over commonly used methods is possible.

Appendix A: Iteration rules

For sequential methods, it is convenient with iteration rules, which allow updates of the statistic obtained at previous time points. For all four models, convenient formulas for calculation can be found. We will first look at the conditional models CA and CB, where results are available in the literature but express the iteration rules in a way that makes comparison with the unconditional ones easy. Then, we will derive new results for the unconditional models UA and UB.

A.1   Iteration Rules for the Conditional Models

For CA, the normalized logarithm of the likelihood ratio $\sigma^2 \log L(s,t)/\delta$ in $H(s)$ is equivalent to

$$(1 - \alpha) \sum_{i=t}^{s} \big(X(i) - \alpha X(i-1) - (1-\alpha)\delta/2\big).$$

Remember that $\tilde{\varepsilon}(s) = X(s) - \alpha X(s-1)$ are the empirical residuals and $k = \delta/2$. Now we have the iteration rule

$$
\begin{aligned}
H(s) &= \max_{t=1,2,\dots,s} (1-\alpha) \sum_{i=t}^{s} \big(\tilde{\varepsilon}(i) - (1-\alpha)k\big) \\
&= \max\left\{ \max_{t=1,2,\dots,s-1} (1-\alpha) \sum_{i=t}^{s-1} \big(\tilde{\varepsilon}(i) - (1-\alpha)k\big), 0 \right\} \\
&\qquad\qquad\qquad\qquad\qquad + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big) \\
&= \max\big\{ H(s-1), 0 \big\} + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big) .
\end{aligned}
\tag{6}
$$

Note that this iteration rule is slightly different to a common one which would be here

$$
H(s)' = \max\big\{ H(s-1)' + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big), 0 \big\} .
\tag{7}
$$

The latter agrees with *Rule 2*, $S_n' = \max(S_{n-1}' + x_n, 0)$ by Page (1954), and can be seen as a sequence of sequential probability ratio test statistics. The main advantage of (7) is that $H(s)'$ is bounded from below. This helps at setting up numerical algorithms for calculating performance measures. However, it is recursion (6) which is the CUSUM method (1) and which was considered in Moustakides (1986) and, quite recently, in Han and Tsung (2009a) at their proofs for optimality. It corresponds to Page's *Rule 1*: "Take action if $S_n - \max_{0 \le i < n} S_i \ge h$". Replacing $0 \le i < n$ by $0 \le i \le n$ would give again (7). Because of $H(s)' = \max\{H(s), 0\}$ for all s, the replacement changes the alarming behavior only for $h < 0$. This would, however, imply a very small $E_\infty(t_A)$ as pointed out by Moustakides (1986). Thus, the difference between the two methods is seldom important for applications.

In case of CB, only the first value of the CUSUM sequence, $H(1)$, has to be changed. From the definition of $L(s,1)$ in Section 5.2 it follows

$$
H(1) = (1 - \alpha^2)\big(X(1) - k\big) .
\tag{8}
$$

Note that with this start-up $H(1)$ depends only on the first observation, that is, there is no potential head start parameter involved.

In summary for CA and CB, the resulting CUSUM sequence is simply the usual CUSUM applied to the (empirical) residuals with an adjusted reference value. A further modified $H(1)$ for CB leads directly to the so-called residual CUSUM scheme as, e. g., in Knoth and Schmid (2004).

A.2 Iteration Rules for the Unconditional Models

It is more difficult for the cases UA and UB. In the sequel, it is shown that for models UA and UB the barrier 0 has to be replaced by a different one in order to

maintain equivalent alarming behavior as for CA/CB and the classical *iid* CUSUM schemes. At first sight, a similar recursion rule with reflection at zero should be possible (as claimed in Knoth and Schmid, 2004). A more detailed analysis of the CUSUM for UA or UB provides different reflection borders. For UA the CUSUM statistic becomes

$$
\begin{aligned}
H(s) &= \max_{t=1,2,\ldots,s} \left\{ \tilde{\varepsilon}(t) - k + (1-\alpha) \sum_{i=t+1}^{s} \big(\tilde{\varepsilon}(i) - (1-\alpha)k \big) \right\} \\
&= \max \left\{ \max_{t=1,2,\ldots,s-1} \left\{ \tilde{\varepsilon}(t) - k + (1-\alpha) \sum_{i=t+1}^{s-1} \big(\tilde{\varepsilon}(i) - (1-\alpha)k \big) \right\} + \right. \\
&\qquad\qquad \left. (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k \big), \tilde{\varepsilon}(s) - k \right\} \\
&= \max \left\{ H(s-1) + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k \big), \tilde{\varepsilon}(s) - k \right\}. \qquad (9)
\end{aligned}
$$

For UB, the only change from UA is for $s = 1$. There are many suggestions in the literature as was seen in Section 5.3.

A.2.1 Iteration Rules with Reflecting Barriers

Now it looks like the sequence $H(s)$ has no lower reflection barrier for UA. Essentially, this is caused by the contribution of time s that is different for $t = s$ than for $t < s$. However, for a slightly changed method, a recursion with reflection barrier can be established. For reasonable alarm thresholds, $h \geq 0$, the new sequence yields the same alarm behavior as the CUSUM method (9). By introducing

$$
z_r = \begin{cases} -\alpha k & , \ \alpha > 0 \\ \alpha h - \alpha(1-\alpha)k & , \ \alpha < 0 \end{cases}
$$

we define the statistic
$$
H(s)^* := \max\{H(s), z_r\}.
$$

The barrier z_r is smaller than zero for $\alpha > 0$, or $\alpha < 0$ and $h > (1-\alpha)k$. For $z_r > 0$, the methods **M1** and **M4** (also **M1**^{e} and **M4**^{e}) are equivalent. For the sequence $H(s)^*$, the following iteration rule is valid:

$$
H(s)^* = \max \left\{ H(s-1)^* + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k \big), \tilde{\varepsilon}(s) - k, z_r \right\} \qquad (10)
$$

Proof of the iteration rule for H^*:
Consider first case $\alpha > 0$. Deploying $\alpha > 0$, and the definitions of $H()$ and $H()^*$

for s and $s-1$ yields:

$$
\begin{aligned}
H(s)^* &= \max\big\{H(s), -\alpha k\big\}\\
&= \max\big\{H(s-1) + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big), \tilde{\varepsilon}(s) - k, -\alpha k\big\}\\
&= \max\big\{H(s-1) + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big),\\
&\qquad\qquad\qquad\qquad - \alpha k + \tilde{\varepsilon}(s) - (1-\alpha)k, -\alpha k\big\}\\
&= \max\big\{H(s-1) + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big),\\
&\qquad - \alpha k + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big), -\alpha k + \tilde{\varepsilon}(s) - (1-\alpha)k, -\alpha k\big\}\\
&= \max\big\{H(s-1)^* + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big), \tilde{\varepsilon}(s) - k, -\alpha k\big\} \quad (11)
\end{aligned}
$$

For case $\alpha < 0$ the validity of the recursion is proved differently. Now, a simple property of the sequence $H(s)$ is utilized. For any s along the unstopped series, $\tilde{\varepsilon}(s) - k$ is not larger than h. Any s with $\tilde{\varepsilon}(s) - k > h$ would flag an alarm on both $H(s)$ and $H(s)^*$ (if not already stopped earlier). From $\tilde{\varepsilon}(s) - k \leq h$ the following inequalities can be derived:

$$
\begin{aligned}
&\alpha h - \alpha\big(\tilde{\varepsilon}(s) - k\big) \leq 0\\
\rightarrow\quad &\alpha h + (1-\alpha)\big(\tilde{\varepsilon}(s) - k\big) \leq \tilde{\varepsilon}(s) - k\\
\rightarrow\quad &\alpha h - \alpha(1-\alpha)k + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big) \leq \tilde{\varepsilon}(s) - k
\end{aligned}
$$

Based on the last inequality above and on the definitions of $H()$ and $H()^*$ for s and $s-1$ we have:

$$
\begin{aligned}
H(s)^* &= \max\big\{H(s), \alpha h - \alpha(1-\alpha)k\big\}\\
&= \max\big\{H(s-1) + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big),\\
&\qquad\qquad\qquad\qquad \tilde{\varepsilon}(s) - k, \alpha h - \alpha(1-\alpha)k\big\}\\
&= \max\Big\{H(s-1) + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big),\\
&\quad \alpha h - \alpha(1-\alpha)k + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big), \tilde{\varepsilon}(s) - k, \alpha h - \alpha(1-\alpha)k\Big\}\\
&= \max\big\{H(s-1)^* + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big),\\
&\qquad\qquad\qquad\qquad \tilde{\varepsilon}(s) - k, \alpha h - \alpha(1-\alpha)k\big\} \qquad (12)
\end{aligned}
$$

This completes the proof.

Appendix B: Integral equations and other formulas for calculation of performance measures

This section provides methods to calculate measures of expected delay numerically. Note that they are determined under the change point model UB. It is

easily adopted for UA. The results under CA and CB are obtained similarly to the *iid* case. Under UB (also under CB), the schemes initialize with $H(1) = (1 - \alpha^2)(X(1) - k)$ or with the maximum of this value and some reflection barrier, or with the special start-up of the residual CUSUM method. Throughout this section we use the term $H(s)$ not as a well-defined likelihood expression as before, but as a dummy for the alarm statistic for each method. To calculate $\mathcal{W}$, we have to distinguish between different change point positions. Essentially, the cases $\tau = 1$ and $\tau > 1$ are different. The interesting quantities are

$$
\begin{aligned}
\mathcal{W}_1 &= \mathcal{L}_1 = E_1(t_A)\,, \\
\mathcal{L}_{>1}(z) &= E_\tau\big(t_A - \tau + 1 \mid H(\tau - 1) = z, t_A \geq \tau\big) \quad \text{for } \tau > 1, z \in [z_r, h]\,, \\
\mathcal{W}_{>1} &= \mathcal{L}_{>1}(z_r)\,,
\end{aligned}
$$

where $\mathcal{W}_1$ and $\mathcal{W}_{>1}$ denote $\mathcal{W}$ restricted to change point positions $\tau = 1$ and $\tau > 1$, respectively. $\mathcal{L}_{>1}(z)$ resembles the ARL function depending on $H(\tau - 1) = z$, the last value of the CUSUM sequence before the change. The total $\mathcal{W}$ is simply the max of $\mathcal{W}_1$ and $\mathcal{W}_{>1}$.

The distribution of $\tilde{\varepsilon}_\tau$ differs from the one of $\tilde{\varepsilon}_t$ for $t > \tau$ (and, of course, for $t < \tau$). The following function allows to deal with this situation. Let

$$
A(z) := E_\tau\big(t_A - t + 1 \mid H(t - 1) = z, t_A \geq t\big)\,, \ t > \tau\,, \ z \in [z_r, h]\,.
$$

For $t > \tau$ (cf Page (1954))

$$
A(z) = 1 + F(z \to z_r)\, A(z_r) + \int_{z_r}^{h} A(\tilde{z}) f(z \to \tilde{z})\, d\tilde{z}\,.
$$

Thereby, $F(z \to z_r)$ and $f(z \to \tilde{z})$ denotes the transition cdf and pdf, respectively, of the related CUSUM sequences. For the CA/CB schemes ($z_r = 0$) it could be solved as usually. For the UA/UB schemes a new approach is needed as demonstrated below. Denote $\Phi()$ and $\phi()$ the cdf and pdf of a standard normal distribution, respectively. Then given $A(z)$, the above ARLs could be calculated for the *B schemes (remember (8))

$$
\mathcal{L}_1 = 1 + F_1(z_r)A(z_r) + \int_{z_r}^{h} f_1(x)A(x)\, dx\,,
$$

$$
F_1(z_r) = \Phi\left(\frac{z_r}{1 - \alpha^2} + k - \delta \right)\,,
$$

$$
f_1(x) = \phi\left(\frac{x}{1 - \alpha^2} + k - \delta \right) / (1 - \alpha^2)\,.
$$

and for all schemes (for *A schemes $\mathcal{L}_1 = \mathcal{L}_{>1}(z_r)$)

$$\mathcal{L}_{>1}(z) = 1 + F_{>1}(z \to z_r)A(z_r) + \int_{z_r}^{h} f_{>1}(z \to x)A(x)\,dx$$

with the related functions $F_{>1}()$ and $f_{>1}()$ for the considered CUSUM sequences. For UA and UB in the most interesting case $z = z_r$ ($\mathcal{W}_{>1} = \mathcal{L}_{>1}(z_r)$), the two functions are

$$F_{>1}(z_r \to z_r) = \Phi\left(\frac{z_r + k - \delta}{\sqrt{1 - \alpha^2}}\right),$$

$$f_{>1}(z_r \to x) = \phi\left(\frac{x + k - \delta}{\sqrt{1 - \alpha^2}}\right) / \sqrt{1 - \alpha^2}.$$

The corresponding functions for CA and CB ($z_r = 0$) are

$$F_{>1}(0 \to 0) = \Phi\left(\frac{(1 - \alpha)k - \delta}{\sqrt{1 - \alpha^2}}\right),$$

$$f_{>1}(0 \to x) = \phi\left(\frac{x/(1 - \alpha) + (1 - \alpha)k - \delta}{\sqrt{1 - \alpha^2}}\right) / \left(\sqrt{1 - \alpha^2}(1 - \alpha)\right).$$

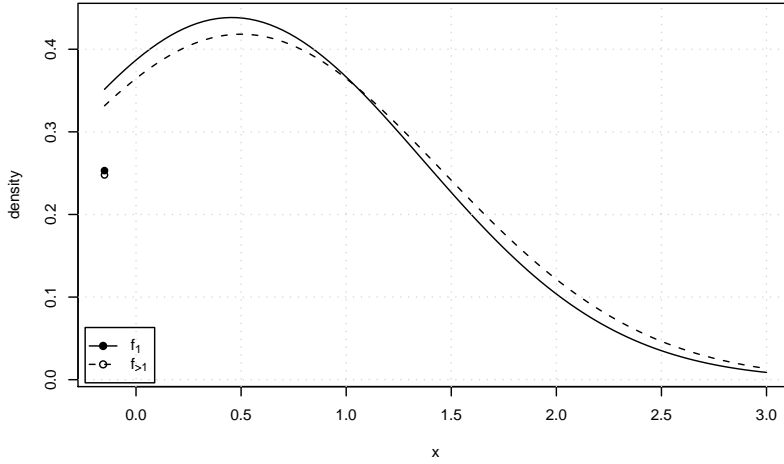Figure 4 illustrates the two different distributions used for the UB scheme.

Figure 4: Probability and densities $F_1(z_r)$, $f_1(x)$ and $F_{>1}(z_r \to z_r)$, $f_{>1}(z_r \to x)$ for UB with $k = 0.5, h = 3, \alpha = 0.3$

To identify integral equations for the function $A(z)$ in case of UA and UB we have to take into account the three-fold structure of the iteration rule. Essentially,

depending on z the component $\tilde{\varepsilon}(s) - k$ is smaller or larger than the competing one $z + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big)$. Eventually, for $\alpha > 0$ the equations for $z \in [z_r, h]$ look like

$\alpha h - \alpha(1-\alpha)k < z < h$:

$$A(z) = 1 + \Phi\big((z_r - z)/(1-\alpha) + (1-\alpha)k\big)A(z_r)$$
$$+ \int_{z_r}^{h} A(\tilde{z})\frac{\phi\big((\tilde{z} - z)/(1-\alpha) + (1-\alpha)k\big)}{1-\alpha}\,d\tilde{z}\,,$$

$z_r < z < \alpha h - \alpha(1-\alpha)k$:

$$A(z) = 1 + \Phi\big((z_r - z)/(1-\alpha) + (1-\alpha)k\big)A(z_r)$$
$$+ \int_{z_r}^{z/\alpha+(1-\alpha)k} A(\tilde{z})\frac{\phi\big((\tilde{z} - z)/(1-\alpha) + (1-\alpha)k\big)}{1-\alpha}\,d\tilde{z}$$
$$+ \int_{z/\alpha+(1-\alpha)k}^{h} A(\tilde{z})\phi\big(\tilde{z} + k\big)\,dx\,,$$

$z = z_r$:

$$A(z) = 1 + \Phi\big((1-\alpha)k\big)A(z_r) + \int_{z_r}^{h} A(\tilde{z})\phi\big(\tilde{z} + k\big)\,d\tilde{z}\,.$$

The decomposition is driven by the competition between the components $z + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big)$ and $\tilde{\varepsilon}(s) - k$ in (9) or (10). Let $\tilde{z}$ (as in the integrals above) be the new CUSUM value. Then it would be generated by the first component if

$$\tilde{z} = z + (1-\alpha)\big(\tilde{\varepsilon}(s) - (1-\alpha)k\big) \geq \tilde{\varepsilon}(s) - k$$
$$\Leftrightarrow \tilde{z} \geq \frac{\tilde{z} - z}{1-\alpha} + (1-\alpha)k - k$$
$$\Leftrightarrow z \geq \alpha\tilde{z} - \alpha(1-\alpha)k\,.$$

For instance, if $z \geq \alpha h - \alpha(1-\alpha)k$, then the condition above is fulfilled by all $\tilde{z}$ in $[z_r, h]$.

Similar ideas lead to the scheme for $\alpha < 0$. Here, the integral equations are:

$\alpha z_r - \alpha(1-\alpha)k < z < h$:

$$A(z) = 1 + \Phi\big((z_r - z)/(1-\alpha) + (1-\alpha)k\big)A(z_r)$$
$$+ \int_{z_r}^{h} A(x)\frac{\phi\big((x - z)/(1-\alpha) + (1-\alpha)k\big)}{1-\alpha}\,dx\,,$$

$z_r < z < \alpha z_r - \alpha(1-\alpha)k$:

$$A(z) = 1 + \Phi\big(z_r + k\big)A(z_r)$$
$$+ \int_{z_r}^{z/\alpha + (1-\alpha)k} A(x)\frac{\phi\big((x-z)/(1-\alpha) + (1-\alpha)k\big)}{1-\alpha}\, dx$$
$$+ \int_{z/\alpha + (1-\alpha)k}^{h} A(x)\phi\big(x+k\big)\, dx\,,$$

$z = z_r$:

$$A(z) = 1 + \Phi\big(z_r + k\big)A(z_r) + \int_{z_r}^{h} A(x)\phi\big(x+k\big)\, dx\,.$$

These integral equations will be solved numerically. The details are collected in the next section.

Appendix C: Numerical methods

The specific nature of the integral equations in the previous section – the issue with the integral border $z/\alpha + (1-\alpha)k$ and the associated decomposition in two integrals impedes the deployment of standard methods such as the Nyström method that is frequently used for ARL calculation – calls for methods like collocation. It was successfully used in Knoth (2006) for the ARL calculation of CUSUM schemes on the sample variance S^2. To sketch the ideas we start with the case $\alpha > 0$. Recall that the reflecting barrier is $z_r = -\alpha k$ for scheme $H^*()$ (and 0 for $H^0()$). Denote $B = \alpha h - \alpha(1-\alpha)k$. Consider the following set of nodes: $h > y_1 > y_2 > \ldots > y_b \geq B > y_{b+1} > y_N > y_{N+1} = z_r$. Thus, we have $N+1$ nodes total with b nodes larger than B. For collocation, the true solution of the integral equation is approximated by linear combination of certain base functions. Here, we utilize Chebyshov polynomials $T_j()$. The true function $A(z)$ is approximated by $\sum_{j=1}^{N} c_j T_j(z)$. The unit Chebyshov polynomials $\cos\big(j\,\arccos(z)\big)$ are rescaled from $[-1, 1]$ to $[z_r, h]$, that is,

$$T_j(z) = \cos\left[j\,\arccos\left(\frac{2z - z_r - h}{h - z_r}\right)\right]. \tag{13}$$

The nodes are derived from the extrema of the polynomial $T_N(z)$:

$$y_i = z_r + \frac{h - z_r}{2}\left[1 + \cos\left(\frac{(N-1-i)\,\pi}{N-1}\right)\right]\,, \quad i = 1, 2, \ldots, N\,.$$

Note that simpler rules to get both the nodes and the polynomials are possible, but the chosen ones ensure numerical stability, e. g., for increasing N.

To calculate the weights c_j we evaluate the following linear equations.

$i = 1, 2, \ldots, b :$

$$\sum_{j=1}^{N} c_j T_j(y_i) = 1 + \Phi\big((z_r - y_i)/(1 - \alpha) + (1 - \alpha)k\big) A(z_r)$$

$$+ \sum_{j=1}^{N} c_j \int_{z_r}^{h} T_j(x) \frac{\phi\big((x - y_i)/(1 - \alpha) + (1 - \alpha)k\big)}{1 - \alpha} \, dx \, ,$$

$i = b + 1, b + 2, \ldots, N :$

$$\sum_{j=1}^{N} c_j T_j(y_i) = 1 + \Phi\big((z_r - y_i)/(1 - \alpha) + (1 - \alpha)k\big) A(z_r)$$

$$+ \sum_{j=1}^{N} c_j \int_{z_r}^{y_i/\alpha + (1-\alpha)k} T_j(x) \frac{\phi\big((x - y_i)/(1 - \alpha) + (1 - \alpha)k\big)}{1 - \alpha} \, dx$$

$$+ \sum_{j=1}^{N} c_j \int_{y_i/\alpha + (1-\alpha)k}^{h} T_j(x) \phi\big(x + k\big) \, dx \, ,$$

$i = N + 1 \leftrightarrow z = z_r$:

$$\mathcal{L}(z_r) = 1 + \Phi\big((1 - \alpha)k\big) A(z_r) + \sum_{j=1}^{N} c_j \int_{z_r}^{h} T_j(x) \phi\big(x + k\big) \, dx \, .$$

The integrals are evaluated numerically (Gauss quadrature). It turns out that piecewise collocation, for $z \in [z_r, B]$ and $[B, h]$, provides much better performance. See Figure 5 for some illustration of the accuracy performance.

Acknowledgments

References

ATIENZA, O., TANG, L., AND ANG, B. (1998), A SPC procedure for detecting level shifts of autocorrelated processes, *Journal of Quality Technology* **30**, 340–351.
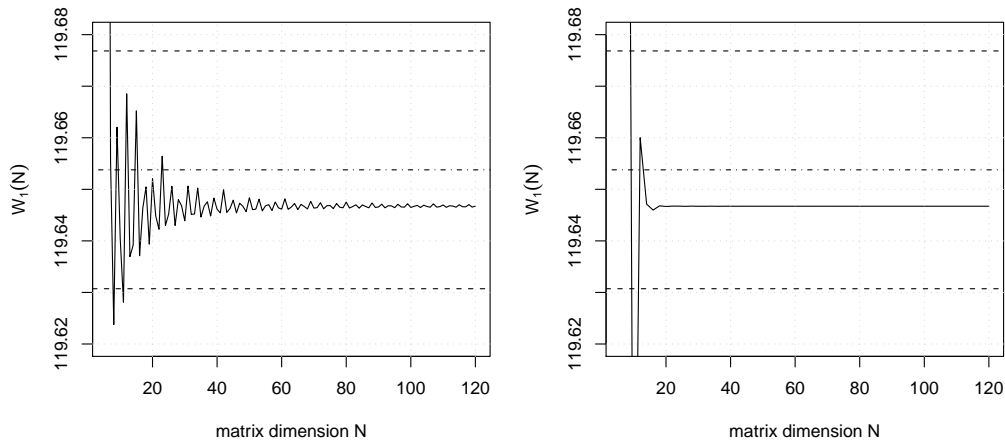
Figure 5: Approximation $W_1(N)$ of $\mathcal{W}_1$ vs. node number N for method **M1** with $\alpha = 0.3, k = 0.5, h = 3$. The left figure illustrates the initial collocation approach whereas the right figure does the same for the piecewise defined collocation polynomials. Results of a Monte-Carlo study are added as dashed lines (roughly the 95% confidence interval for $\mathcal{W}_1$ based on normal approximation).

BAGSHAW, M. AND JOHNSON, R. A. (1975), The effect of serial correlation on the performance of cusum tests II, *Technometrics* **17**, 73–80.

BASSEVILLE, M. AND NIKIFOROV, I. V. (1993), *Detection of Abrupt Changes: Theory and Application*, Prentice-Hall.

FRISÉN, M. (2003), Statistical Surveillance. Optimality and Methods, *International Statistical Review* **71**, 403–434.

FRISÉN, M. (2009), Optimal sequential surveillance for finance, public health, and other areas. *Sequential Analysis* **28**, 310–337.

FRISÉN, M. AND WESSMAN, P. (1999), Evaluations of likelihood ratio methods for surveillance. Differences and robustness, *Commun. Stat. Simula. Comput.* **28**, 597–622.

GOLDSMITH, P. L. AND WHITFIELD, H. (1961), Average run lengths in cumulative chart quality control schemes, *Technometrics* **3**, 11–20.

HAN, D. AND TSUNG, F. (2009), The optimal stopping time for detecting changes in discrete time markov processes, *Sequential Analysis* **28**, 115–135.

23

HAN, D. AND TSUNG, F. (2009b), Run length properties of the CUSUM and EWMA for a stationary linear process, *Statistica Sinica* **19**, 473–490.

HARRIS, T. J. AND ROSS, W. H. (1991), Statistical process control procedures for correlated observations. *Can. J. Chem. Eng.* **69**, 48–57.

HAWKINS, D. M. AND OLWELL, D. H. (1998), *Cumulative Sum Charts and Charting for Quality Improvement*, Springer.

KNOTH, S. (2006), Computation of the ARL for CUSUM-S^2 schemes, *Comput. Stat. Data Anal.* **51**, 499–512.

KNOTH, S. AND SCHMID, W. (2004), Control charts for time series: A review, In Lenz, H.-J. and Wilrich, P.-T., editors, *Frontiers of Statistical Quality Control*, volume 7, 210–236. Physica Verlag, Heidelberg, Germany.

KRAMER, H. G. AND SCHMID, W. (1997), EWMA charts for multivariate time series, *Sequential Analysis* **16**, 131–154.

LAI, T. L. (1995), Sequential changepoint detection in quality control and dynamical systems, *J. R. Stat. Soc., Ser. B* **57**, 613–658.

LAI, T. L. (1998), Information bounds and quick detection of parameter changes in stochastic systems, *IEEE Transactions on Information Theory* **44**, 2917–2929.

LAI, T. L. AND SHAN, J. Z. (1999), Efficient recursive algorithms for detection of abrupt changes in signals and control systems, *IEEE Transactions on Automatic Control* **44**, 952–966.

LORDEN, G. (1971), Procedures for reacting to a change in distribution, *Ann. Math. Stat.* **42**, 1897–1908.

MOUSTAKIDES, G. V. (1986), Optimal stopping times for detecting changes in distributions, *Ann. Stat.* **14**, 1379–1387

NIKIFOROV, I. V. (1979), Cumulative sums for detection of changes in random process characteristics, *Autom. Remote Control* **40**, 192–200.

OKHRIN, Y. AND SCHMID, W. (2007), *Surveillance of Univariate and Multivariate Linear Time Series*, 115–152, Wiley.

OKHRIN, Y. AND SCHMID, W. (2007), *Surveillance of Univariate and Multivariate Nonlinear Time Series*, 153–177, Wiley.

PAGE, E. S. (1954), Continuous inspection schemes, *Biometrika* **41**, 100–115.

PETTERSSON, M. (1998), Monitoring a freshwater fish population: statistical surveillance of biodiversity, *Environmetrics* **9**, 139–150.

RUNGER, G. C., WILLEMAIN, T. R., AND PRABHU, S. S. (1995), Average run lengths for CUSUM control charts applied to residuals, *Commun. Stat., Theory Methods* **24**, 273–282.

RYAN, T. P. (2000), *Statistical Methods for Quality Improvement*, John Wiley & Sons, 2nd edition.

SCHMID, W. (1997), CUSUM control schemes for Gaussian processes, *Statistical Papers* **38**, 191–217.

SCHMID, W. AND SCHÖNE, A. (1997), Some properties of the EWMA control chart in the presence of autocorrelation, *Ann. Stat.* **25**, 1277–1283.

SIEGMUND, D. (1985), *Sequential Analysis: Tests and Confidence Intervals*, Springer-Verlag.

SPARKS, R. S. (2000), CUSUM charts for AR1 data: are they worth the effort? *Austral. J. Statist.* **42**, 25–42.

SRIVASTAVA, M. S. AND WU, Y. (1993), Comparison of EWMA, CUSUM and Shiryayev-Roberts procedures for detecting a shift in the mean, *Ann. Stat.* **21**, 645–670.

WIERINGA, J. E. (1999), *Statistical Process Control for Serially Correlated Data*, PhD thesis, Rijksuniversiteit Groningen.

WOODALL, W. H. AND MONTGOMERY, D. C. (1999), Research issues and ideas in statistical process control, *Journal of Quality Technology* **31**, 376–386.

YASHCHIN, E. (1993), Performance of Cusum control schemes for serially correlated observations, *Technometrics* **35**, 37–52.

Research Report

2008:2	Jonsson, R.	When does Heckman's two-step procedure for censored data work and when does it not?
2008:3	Andersson, E.	Hotelling´s T2 Method in Multivariate On-Line Surveillance. On the Delay of an Alarm.
2008:4	Schiöler, L. & Frisén, M.	On statistical surveillance of the performance of fund managers.
2008:5	Schiöler, L.	Explorative analysis of spatial patterns of influenza incidences in Sweden 1999 — 2008.
2008:6	Schiöler, L.	Aspects of Surveillance of Outbreaks.
2008:7	Andersson, E & Frisén, M.	Statistiska varningssystem för hälsorisker
2009:1	Frisén, M., Andersson, E. & Schiöler, L.	Evaluation of Multivariate Surveillance
2009:2	Frisén, M., Andersson, E. & Schiöler, L.	Sufficient Reduction in Multivariate Surveillance
2010:1	Schiöler, L	Modelling the spatial patterns of influenza incidence in Sweden
2010:2	Schiöler, L. & Frisén, M.	Multivariate outbreak detection
2010:3	Jonsson, R.	Relative Efficiency of a Quantile Method for Estimating Parameters in Censored Two-Parameter Weibull Distributions
2010:4	Jonsson, R.	A CUSUM procedure for detection of outbreaks in Poisson distributed medical health events
2011:1	Jonsson, R.	Simple conservative confidence intervals for comparing matched proportions
2011:2	Frisén, M	On multivariate control charts
2011:3	Frisén, M	Methods and evaluations for surveillance in industry, business, finance, and public health