# Quality surveillance with EWMA control charts based on exact control limits

**Manuel Cabral Morais, Yarema Okhrin, Wolfgang Schmid**

**Angaben zur Veröffentlichung / Publication details:**

Morais, Manuel Cabral, Yarema Okhrin, and Wolfgang Schmid. 2015. "Quality surveillance with EWMA control charts based on exact control limits." *Statistical Papers* 56 (3): 863–85. https://doi.org/10.1007/s00362-014-0612-8.



# Quality surveillance with EWMA control charts based on exact control limits

**Manuel Cabral Morais · Yarema Okhrin · Wolfgang Schmid**

**Abstract** The control charts are main tools of statistical surveillance of quality in production processes. Exponentially weighted moving average charts that make use of exact control limits are discussed in detail in this paper. We start by assessing the impact of the smoothing constant $\lambda$ not only in the in-control average run length (ARL) of upper one-sided EWMA charts with exact control limits, but also in the range of the exact control limits of such charts with a common in-control ARL value (i.e. matched in-control). Based on the analytical results and on an extensive simulation study we conclude that the out-of-control ARL of matched in-control upper one-sided EWMA charts with exact control limits increases with $\lambda$. This in turn suggests the use of $\lambda$ values as close to the zero as possible and motivates what we called the (upper one-sided) limit chart. Its performance is extensively studied with regard to the ARL. Finally, we investigate the impact of $\lambda$ on the ARL of EWMA charts with asymptotic control limits; the (maximum) conditional average delay is also addressed as an additional performance measure.

**Keywords** Statistical process control · Run length · Stochastic ordering

M. C. Morais
Department of Mathematics & Center for Mathematics and its Applications (CEMAT), Instituto Superior Técnico, Lisboa, Portugal
e-mail: maj@math.ist.utl.pt

Y. Okhrin (✉)
Faculty of Business and Economics, University of Augsburg, Universitaetsstr. 16, 86159 Augsburg, Germany
e-mail: yarema.okhrin@wiwi.uni-augsburg.de

W. Schmid
Department of Statistics, European University Viadrina, 15207 Frankfurt (Oder), Germany
e-mail: schmid@euv-frankfurt-o.de

## 1 Background

Production is typically monitored using data collected on a regular basis: the observed values of a control statistic are sequentially plotted together with appropriate control limits in what is grandly termed as quality control chart. This graphical device is used to track process performance over time and identify assignable causes that may affect the quality of the output. Moreover, although proposed by Shewhart to his superiors at Bell Laboratories in a historic memorandum of May 16, 1924 (Juran 1997 and http://www.asq.org/about-asq/who-we-are/bio_shewhart.html), control charts are still among the most important and widely used devices in statistics (Stoumbos et al. 2000).

The simplicity and the one-size-fits-all character of the control charts pioneered by Shewhart (1931) are responsible for their immense popularity among practitioners. However, the fact that the Shewhart charts only use the information about the process given by the last observed value of the control statistic, and completely ignore any past information, is responsible for a serious limitation: Shewhart charts are not effective in the detection of assignable causes that lead to small and moderate shifts in the parameter being monitored.

The exponentially weighted moving average (EWMA) chart, introduced by Roberts (1959), incorporates all the information in the sequence of observed values of the control statistic and prove to be more effective than Shewhart chart for detecting small and moderate shifts, namely, in the process mean (Lucas and Saccucci 1990) or in the process variance (Crowder and Hamilton 1992) of independent output. EWMA has been also used to monitor the process mean and the process mean vector of dependent output with considerable advantage, as reported by several authors like Schmid (1997) and Kramer and Schmid (1997).

Moreover, apart from being a process monitoring device, EWMA can be viewed as a method for establishing real-time dynamical control of industrial processes (Hunter 1986), namely as the best linear forecast for a first order integrated moving average process (IMA(1)) with drifts, as promoted by Baxley (1990).

The EWMA chart for monitoring the process mean $\mu$ of independent output is often based on the statistic

$$Z_t = \begin{cases} z_0, & t = 0 \\ (1-\lambda)Z_{t-1} + \lambda X_t, & t = 1, 2, \ldots, \end{cases} \tag{1}$$

where: the initial value $z_0$ is frequently taken to be the target process mean $\mu_0$; $X_t$ is an estimator of $\mu$, usually the sample mean at time $t$; and $\lambda \in (0, 1]$ is a smoothing constant that corresponds to the weight given to the most recent sample. Having in mind that $Z_t$ can be equivalently written as the following moving average

$$Z_t = \lambda \sum_{i=0}^{t-1} (1-\lambda)^i X_{t-i} + (1-\lambda)^t z_0, \quad t = 1, 2, \ldots, \tag{2}$$

whose weights fall off geometrically, we immediately conclude that a $\lambda$ close to one leads to a short memory EWMA chart—in fact $\lambda = 1$ leads to nothing but a She-

whart chart—, whereas values of $\lambda$ close to zero lead to EWMA charts that give little importance to the most recent observations.

Clearly, the value of $\lambda$ and the control limits of the EWMA chart have a strong impact in its performance, as we can see in the next section, and should be carefully chosen by the user to give this chart desirable properties for both in-control and out-of-control situations.

## 2 EWMA charts with exact control limits

In what follows $\{Y_t\}$ denotes an i.i.d. target process normally distributed with mean $\mu_0$ and variance $\sigma^2$ which is related to an observed process $\{X_t\}$ as follows:

$$X_t = \begin{cases} Y_t & \text{for } t = \ldots, -1, 0 \\ Y_t + a\,\sigma & \text{for } t = 1, 2, \ldots, \end{cases} \tag{3}$$

where $-\infty < a < \infty$. If $a \neq 0$ then a sustained shift in the process mean $\mu$ has occurred at time $t = 1$. Needless to say, $\{X_t\}$ is said to be in-control if $a = 0$, and to be out-of-control otherwise.

The expectation (when $Z_0 = \mu_0$) and variance of this normally distributed control statistic with respect to (3) are equal to

$$E_a(Z_t) = E_0(Z_t) + a\,\sigma\left[1 - (1-\lambda)^t\right] = \mu_0 + a\,\sigma\left[1 - (1-\lambda)^t\right] \tag{4}$$

$$\mathrm{Var}_a(Z_t) = \mathrm{Var}_0(Z_t) = \frac{\lambda}{2-\lambda}\,\sigma^2\left[1 - (1-\lambda)^{2t}\right]. \tag{5}$$

The indexes “textita” and “0” mean, throughout the remainder of this paper, that the quantity (an expectation, a variance, a covariance, a probability, etc.) is calculated with respect to (3) and to the in-control situation, respectively.

Upward shifts (i.e. $a > 0$) can be detected by upper one-sided EWMA charts that give a signal at the sampling period $t$, suggesting that an increase in the process mean has occurred, if

$$Z_t > E_0(Z_t) + c\sqrt{\mathrm{Var}_0(Z_t)}\,, \tag{6}$$

for some fixed constant critical value $c$ that defines the range of these exact control limits. This signal is a valid one, in case the process is out-of-control, and it is called a false alarm, otherwise.

### 2.1 Assessing the performance of upper one-sided EWMA charts with exact control limits

To assess the performance of a control chart we have to consider the number of observations taken until a signal is triggered by this chart. This random variable is usually called the run length (RL) of the chart and it is defined as

$$N_a = \inf\left\{\, t \in I\!N : Z_t > E_0(Z_t) + c\sqrt{\mathrm{Var}_0(Z_t)} \,\right\} \tag{7}$$

for the upper one-sided EWMA charts with exact control limits. The properties of this performance measure not only depend on the magnitude of the shift $a$, but also on the values of $\lambda$ and $c$, therefore it is going to be often represented by $N_a(\lambda, c)$.

It should be also added that the asymptotic variance of $Z_t$, $\lim_{t\to+\infty} \mathrm{Var}_0(Z_t) = \frac{\lambda}{2-\lambda}\,\sigma^2$, is frequently preferred by practitioners to $\mathrm{Var}_0(Z_t)$, since this leads to constant control limits and therefore less computational effort in the implementation of the EWMA chart. Note, however, that the adoption of the asymptotic variance should be avoided when the value of $\lambda$ is small, since in this case $\mathrm{Var}_0(Z_t)$ slowly converges to its asymptotic value making the EWMA scheme substantially less sensitive to early process shifts (Steiner 1999).

Thus, we strongly recommend the use of exact control limits, also called time-varying control limits by Steiner (1999).

Now, before we go any further into the assessment of the performance of the upper one-sided EWMA chart with exact control limits, it is essential to add the joint distribution of $\mathbf{Z} = [Z_i]_{i=1,\ldots,k}$, a multivariate normal distribution with mean vector and covariance matrix given by

$$\begin{aligned}
\boldsymbol{\mu}_a &= \left[\mu_0 + a\,\sigma\,\left[1-(1-\lambda)^i\right]\right]_{i=1,\ldots,k} \qquad (8)\\
\Sigma_a &= \Sigma_0\\
&= \left[\mathrm{Cov}_0(Z_i, Z_j)\right]_{i,j=1,\ldots,t}\\
&= \left[\frac{\lambda}{2-\lambda}\,\sigma^2\,(1-\lambda)^{|j-i|}\left[1-(1-\lambda)^{2\,\min\{i,j\}}\right]\right]_{i,j=1,\ldots,k}, \qquad (9)
\end{aligned}$$

and distribution function represented by $F_{\mathcal{N}_k(\boldsymbol{\mu}_a,\Sigma_0)}(\mathbf{z})$.

In fact, the survival function of the first passage time $N_a(\lambda, c)$ is equal to

$$\begin{aligned}
P[N_a(\lambda,c) > k] &= P_a\left[Z_i \le E_0(Z_i) + c\,\sqrt{\mathrm{Var}_0(Z_i)},\quad i = 1,\ldots,k\right]\\
&= P_a\left[\frac{Z_i - E_a(Z_i)}{\sqrt{\mathrm{Var}_0(Z_i)}} \le c - a\,\sigma\,\frac{1-(1-\lambda)^i}{\sqrt{\mathrm{Var}_0(Z_i)}},\quad i = 1,\ldots,k\right]\\
&= F_{\mathcal{N}_k(\mathbf{0}_k,\mathbf{C}_k(\lambda))}\left(c - a\,\sigma\,\frac{1-(1-\lambda)^i}{\sqrt{\mathrm{Var}_0(Z_i)}},\quad i = 1,\ldots,k\right), \qquad (10)
\end{aligned}$$

for $k = 1, 2, \ldots$, where $\mathbf{0}_k$ is a column vector with $k$ zeroes and $\mathbf{C}_k(\lambda)$ is a correlation matrix with entries defined by

$$\mathrm{Corr}_0(Z_i, Z_j) = (1-\lambda)^{j-i}\sqrt{\frac{1-(1-\lambda)^{2i}}{1-(1-\lambda)^{2j}}},\quad 1 \le i \le j \le k. \qquad (11)$$

The average run length (ARL) is then given by

$$E[N_a(\lambda, c)] = 1 + \sum_{k=1}^{+\infty} P[N_a(\lambda, c) > k]. \qquad (12)$$

We rely on Monte-Carlo simulation methods to approximate the ARL of this upper one-sided EWMA chart with exact limits. This allows us to analyse all the charts in a methodologically consistent way.

Before we proceed we need to introduce two preparatory definitions, to make a comment and recall Theorem 5.1.7 by Tong (1990, p. 103) because they play a major role in the presentation of the main results that will soon follow.

The run length $RL$ is said to be stochastically smaller than the run length $RL'$ in the usual sense ($RL \leq_{st} RL'$) if and only if $P(RL > x) \leq P(RL' > x)$, for $-\infty < x < \infty$ (see Shaked and Shanthikumar 1994, p. 4)). Needless to say that $RL \leq_{st} RL'$ implies $E(RL) \leq E(RL')$.

Let $RL_\theta$ be a run length whose distribution depends on the parameter $\theta \in \Theta$. Then $RL_\theta$ stochastically increases with $\theta \in \Theta$ in the usual sense (for short, $RL_\theta \uparrow_{st}$ with $\theta$) if and only if $RL_\theta \leq_{st} RL_{\theta'}$, for $\theta \leq \theta'$ $(\theta, \theta' \in \Theta)$.

If $\theta$ represents the magnitude of a sustained shift in a specific parameter, then a stochastically decreasing $RL_\theta$ with regard to $\theta$ is what we hope to be dealing with since it means a stochastically increasing detection speed as the shift becomes more severe.

Finally, Theorem 5.1.7 by Tong (1990, p. 103) can also be stated as follows: let $\mathbf{Z} = [Z_i]_{i=1,\ldots,k} \sim \mathcal{N}_k(\boldsymbol{\mu}, \Sigma = [\sigma_{ij}]_{i,j=1,\ldots,k})$ and $\tilde{\mathbf{Z}} = [\tilde{Z}_i]_{i=1,\ldots,k} \sim \mathcal{N}_k(\boldsymbol{\mu}, \Gamma = [\gamma_{ij}]_{i,j=1,\ldots,k})$ be such that $\sigma_{ii} = \gamma_{ii}, i = 1, \ldots, k$ (i.e. $Z_i$ and $\tilde{Z}_i$ have the same marginal normal distribution for $i = 1, \ldots, k$). If $\sigma_{ij} \geq \gamma_{ij}$ for all $i \neq j$, then

$$P(Z_1 \leq c_1, \ldots, Z_k \leq c_k) \geq P(\tilde{Z}_1 \leq c_1, \ldots, \tilde{Z}_k \leq c_k), \tag{13}$$

for all vectors $[c_i]_{i=1,\ldots,k}$.

## 2.2 Monotonicity results

First note that the survival function of $N_a(\lambda, c)$ can be rewritten as

$$P[N_a(\lambda, c) > k] = F_{\mathcal{N}_k(\mathbf{0}_k, \mathbf{C}_k(\lambda))}\left(c - a\sqrt{\frac{2-\lambda}{\lambda} \times \frac{1-(1-\lambda)^i}{1+(1-\lambda)^i}}, \quad i = 1, \ldots, k\right), \tag{14}$$

for $k = 1, 2, \ldots$. Thus, we can immediately conclude that, for any fixed critical value $c$, any fixed smoothing constant $\lambda \in (0, 1]$ and any $k = 1, 2, \ldots$, $P[N_a(\lambda, c) > k]$ is a decreasing function of $a$, i.e.

$$N_a(\lambda, c) \downarrow_{st} \text{ with } a. \tag{15}$$

As a consequence, (15) means that the upper one-sided EWMA chart with exact control limits stochastically increases its detection speed as the upward shift becomes more severe. Not surprisingly, the detection speed stochastically decreases with the absolute value of $a$ when $a < 0$—after all, we are dealing with an upper one-sided chart which should never be used to detect downward shifts.

Another obvious result: for any fixed $a$, any fixed smoothing constant $\lambda \in (0, 1]$ and any $k = 1, 2, \ldots$, $P[N_a(\lambda, c) > k]$ is an increasing function of $c$, i.e.

$$N_a(\lambda, c) \uparrow_{st} \text{ with } c. \tag{16}$$

This result should be read as follows: if we increase the critical value $c$ we are bound to deal with a stochastically less sensitive upper one-sided EWMA chart with exact control limits.

Results concerning the stochastically monotone behaviour of $N_a(\lambda, c)$ in terms of $\lambda$ are difficult to achieve for two reasons, as we shall see later on: we are dealing with a multivariate normal distribution whose covariance matrix depends on $\lambda$; and $\sqrt{\frac{2-\lambda}{\lambda} \times \frac{1-(1-\lambda)^i}{1+(1-\lambda)^i}}$ is a decreasing function of $\lambda$.

However, in the absence of assignable causes ($a = 0$), we can state an important result concerning the in-control RL of all upper one-sided EWMA control charts with exact control limits and sharing the same critical value $c$; let us designate this family of control charts by $\mathcal{F}_{uEWMAe}(c)$.

**Theorem 1** *Assume that the random variables $\{Y_t\}$ are independent and identically distributed to $\mathcal{N}(\mu_0, \sigma^2)$. Then $P[N_0(\lambda, c) > k]$ is a decreasing function in $\lambda \in (0, 1]$, for any fixed $k = 1, 2, \ldots$ and any fixed $c > 0$, that is,*

$$N_0(\lambda, c) \downarrow_{st} \textit{ with } \lambda. \tag{17}$$

We can immediately interpret (17) as follows: increasing the weight given to the most recent observation yields within the family $\mathcal{F}_{uEWMAe}(c)$ leads to a chart with a stochastically smaller number of observations taken until a false alarm.

*Remark 2*

1. The in-control RL of the upper one-sided Shewhart control chart, $N_0(\lambda = 1, c)$, is the smallest in-control RL within $\mathcal{F}_{uEWMAe}(c)$, stochastically speaking.
2. The monotonicity result (17) implies that for any fixed critical value $c$ the in-control ARL of any control chart in the family $\mathcal{F}_{uEWMAe}(c)$ is a decreasing function in $\lambda \in (0, 1]$, as suggested but not proven by Frisén and Sonesson (2006). Therefore the Shewhart control chart has the smallest in-control ARL value within the family $\mathcal{F}_{uEWMAe}(c)$.
3. (17) is still valid for $c \leq 0$.

*Proof* (Theorem 1)—From (11) we get

$$\text{Corr}_0(Z_i, Z_j) = \sqrt{\frac{(1-\lambda)^{-2i} - 1}{(1-\lambda)^{-2j} - 1}}, \; 1 \leq i \leq j \leq k, \tag{18}$$

which is a decreasing function in $\lambda \in (0, 1]$.

Now, let $\mathbf{c}_k$ be a column vector with all its $k$ values equal to the critical value $c$. Then, from Theorem 5.1.7 by Tong (1990, p. 103), we can conclude that, for any fixed $k = 1, 2, \ldots$ and critical value $c$, the survival function $P[N_0(\lambda, c) > k] = F_{\mathcal{N}_k(\mathbf{0}_k, \mathbf{C}_k(\lambda))}(\mathbf{c}_k)$ decreases with $\lambda$, i.e. $N_0(\lambda, c) \downarrow_{st}$ with $\lambda$. □

The critical value $c$ in Theorem 1 is fixed. However, in order to compare control charts the critical value $c$ is usually chosen so that the in-control ARL is equal to a pre-specified value $\xi > 1$. This means that $c$ is the solution of the equation

$$E[N_0(\lambda, c)] = \xi. \tag{19}$$

The critical value $c$ is therefore a function of $\lambda$ and $\xi$. To make this dependence and the one of $N_a$ on this critical value even more obvious, we refer from now on to $c$ as $c(\lambda, \xi)$, and $N_a(\lambda, c)$ as $N_a(\lambda, c(\lambda, \xi))$. Moreover, capitalizing on the fact that

- $E[N_0(\lambda, c(\lambda, \xi))]$ is a continuous function in $c$, for fixed $\lambda$, because $\{Z_t\}$ is a collection of continuous random variables,
- $\lim_{c\to-\infty} E[N_0(\lambda, c(\lambda, \xi))] = 1$,
- $\lim_{c\to\infty} E[N_0(\lambda, c(\lambda, \xi))] = \infty$,

and on the mean-value theorem, we conclude that there is always a solution for Equation (19) and it is unique since $E[N_0(\lambda, c(\lambda, \xi))]$ is strictly increasing in $c$ (this monotone behaviour follows from (16) and the fact that we are dealing with a continuous random variable $Z_t$).

Since we rely on simulation to evaluate the ARL of the upper one-sided EWMA charts with exact control limits and on a numerical search for $c(\lambda, \xi)$, it is essential to characterize the critical value namely as a function of $\lambda$.

**Theorem 3** *Assume that the random variables $\{Y_t\}$ are independent and identically distributed to $\mathcal{N}(\mu_0, \sigma^2)$ and consider a fixed $\xi > 1$. Then the critical value $c(\lambda, \xi)$ is a continuous and increasing function in $\lambda \in (0, 1]$.*

Theorem 3 can be interpreted as follows: increasing the weight on the latest observation requires a larger critical value to attain the same in-control ARL.

*Remark 4* Note that we get $E[N_0(\lambda = 1, c)] = \frac{1}{\Phi(-c)}$ and $c(\lambda = 1, \xi) = \Phi^{-1}(1 - \xi^{-1})$, for the upper one-sided Shewart control chart.

It is interesting to note that $c(\lambda, \xi)$ may be also negative when $\lambda \in (0, 1)$, even for large values of $\xi$. Recall that

$$E[N_0(\lambda, c = 0)] = 1 + \sum_{k=1}^{\infty} P_0(Z_i \leq \mu_0,\ i = 1, \ldots, k) = \xi' > 1, \tag{20}$$

for $\lambda \in (0, 1)$, therefore we are drawn to the conclusion that we can always achieve a smaller in-control ARL value $\xi \in (1, \xi')$ by considering a smaller negative critical value, i.e. $c(\lambda, \xi)$ may take negative values.

*Proof* (Theorem 3)—$c(\lambda, \xi)$ is continuous in $\lambda \in (0, 1]$ because $E[N_0(\lambda, c(\lambda, \xi))]$ is also continuous.

Let $N_0(\lambda_1, c(\lambda_1, \xi))$ and $N_0(\lambda_2, c(\lambda_2, \xi))$ be the in-control RL of two upper one-sided EWMA with exact control limits and smoothing constants $\lambda_1$ and $\lambda_2$—$0 < \lambda_1 \leq \lambda_2 \leq 1$—, and such that they are matched in-control

$$E[N_0(\lambda_1, c(\lambda_1, \xi))] = E[N_0(\lambda_2, c(\lambda_2, \xi))] = \xi. \tag{21}$$

Now, recall the fact that the in-control ARL decreases with $\lambda \in (0, 1]$, for any fixed critical value, say $c(\lambda_1, \xi)$ (see second remark to Theorem 1). As a consequence

$$\xi = E[N_0(\lambda_1, c(\lambda_1, \xi))] \geq E[N_0(\lambda_2, c(\lambda_1, \xi)]. \tag{22}$$

Please note that, according to result (16), the only way to stochastically increase the in-control RL (and therefore the in-control ARL) of the upper one-sided EWMA chart—with smoothing parameter $\lambda_2$ and exact control limits dependent on $c(\lambda_1, \xi)$—is to increase the critical value to $c(\lambda_2, \xi)$, thus

$$c(\lambda_1, \xi) \leq c(\lambda_2, \xi), \tag{23}$$

i.e. $c(\lambda, \xi)$ is increasing in $\lambda$, for any fixed in-control ARL, $\xi > 1$. □

### 2.3 Illustrating some of the monotonicity results

FBy relying on the increasing behaviour of the in-control ARL in terms of the critical value $c(\lambda, \xi)$ (a consequence of result (16)), we are able to use the false position (or *regula falsi*) method (Conte and de Boor 1980, pp. 76–77) to find the root $c(\lambda, \xi)$, as shown in the next example.

We set $\xi = 500$ samples, simulated the in-control process from a normal distribution with zero mean and unit variance, considered the number of replications equal to $10^7$, and got the critical values $c(\lambda, \xi = 500)$, for different values of the smoothing parameter $\lambda = 0.0001, 0.0003, 0.0005, 0.0007, 0.001, 0.003, 0.005, 0.007, 0.01, 0.03, 0.05, 0.07, 0.1, 0.2, 0.3, 0.4, 0.5, 1$, and also the out-control (zero-state) ARL values for all these values of $\lambda$ and several magnitudes of the shift $a = 0.25, 0.5, 1.0, 1.5, 2.0, 3.0, 4.0$.

Please note that in any case the convergence criteria for the critical values is a relative error between the simulated in-control (zero-state) ARL and $\xi = 500$ smaller than 0.2 %. Moreover, any simulated RL is truncated if it is larger than 50,000 samples except for $\lambda = 0$ and $\lambda = 0.0001$. In these two cases 500000 samples were considered to avoid biased estimates of ARL. The reader should be aware that: the truncation of the simulated RL is absolutely crucial; the estimate of the ARL, based on *unlimited* simulated RL, and its standard error increase with the number of replications essentially because of the occurrence of some rare exceptional long runs. Suffice to say that a rigorous discussion of this phenomenon goes beyond the scope of this paper.

Also worthy of note is the fact that the ARL values of the upper one-sided Shewhart chart ($\lambda = 1$) were not obtained by simulation since there is a closed expression for them: $E[N_a(1, c(1, \xi))] = \frac{1}{1-\Phi(c(1,\xi)-a)}$, where $c(1, \xi) = \Phi^{-1}(1 - 1/\xi)$.

The results in Table 1 not only illustrate

- the decreasing behaviour of the out-of-control ARL when we fix both $\lambda$ and the critical value (a consequence of result (15)) and
- the increasing behaviour of the critical value $c(\lambda, \xi)$ in terms of $\lambda$ (see Theorem 3 and Fig. 2),

**Table 1** Critical values, in-control and out-of-control ARLs for upper one-sided EWMA charts with exact control limits and different smoothing parameters $\lambda$

| $\lambda$ | $c(\lambda, 500)$ | $a$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 0.25 | 0.5 | 1.0 | 1.5 | 2.0 | 3.0 | 4.0 |
| 0.0001 | 0.341738 | 500.986 | 5.395 | 2.611 | 1.455 | 1.158 | 1.053 | 1.004 | 1.000 |
| 0.0003 | 0.541806 | 499.034 | 7.316 | 3.223 | 1.629 | 1.227 | 1.081 | 1.007 | 1.000 |
| 0.0005 | 0.653826 | 499.790 | 8.665 | 3.641 | 1.746 | 1.274 | 1.102 | 1.010 | 1.000 |
| 0.0007 | 0.735931 | 499.955 | 9.785 | 3.991 | 1.843 | 1.314 | 1.119 | 1.012 | 1.001 |
| 0.001 | 0.829928 | 499.612 | 11.231 | 4.429 | 1.965 | 1.365 | 1.141 | 1.015 | 1.001 |
| 0.003 | 1.176433 | 500.046 | 18.079 | 6.498 | 2.531 | 1.605 | 1.254 | 1.035 | 1.002 |
| 0.005 | 1.368107 | 500.097 | 22.838 | 7.953 | 2.930 | 1.777 | 1.339 | 1.053 | 1.004 |
| 0.007 | 1.503701 | 499.703 | 26.523 | 9.124 | 3.253 | 1.920 | 1.412 | 1.070 | 1.006 |
| 0.01 | 1.654164 | 500.518 | 30.815 | 10.547 | 3.651 | 2.097 | 1.503 | 1.093 | 1.010 |
| 0.03 | 2.119538 | 499.941 | 45.136 | 15.552 | 5.131 | 2.771 | 1.867 | 1.205 | 1.028 |
| 0.05 | 2.311206 | 499.768 | 52.360 | 17.818 | 5.829 | 3.102 | 2.053 | 1.271 | 1.045 |
| 0.07 | 2.432397 | 499.922 | 58.745 | 19.531 | 6.298 | 3.325 | 2.180 | 1.319 | 1.069 |
| 0.1 | 2.543225 | 499.745 | 66.944 | 21.635 | 6.761 | 3.540 | 2.304 | 1.367 | 1.073 |
| 0.2 | 2.716605 | 500.155 | 90.816 | 28.487 | 7.838 | 3.937 | 2.519 | 1.453 | 1.101 |
| 0.3 | 2.789789 | 500.079 | 111.585 | 36.168 | 8.998 | 4.245 | 2.647 | 1.494 | 1.115 |
| 0.4 | 2.828317 | 499.949 | 130.705 | 44.673 | 10.503 | 4.606 | 2.768 | 1.522 | 1.123 |
| 0.5 | 2.850393 | 499.985 | 148.764 | 54.017 | 12.450 | 5.091 | 2.920 | 1.546 | 1.129 |
| 1 | 2.878162 | 500.000 | 232.971 | 114.948 | 33.135 | 11.894 | 5.265 | 1.823 | 1.151 |

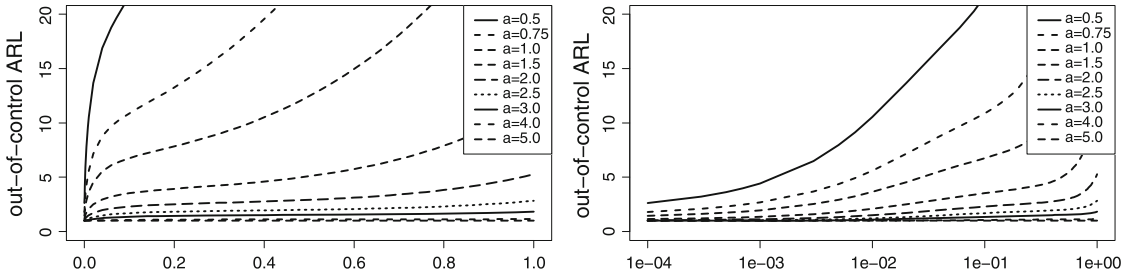


**Fig. 1** Out-of-control ARL of matched in-control upper one-sided EWMA charts with exact control limits as a function of $\lambda$ (on the *left*) and of $\log(\lambda)$ (on the *right*)

but also suggest that

- the out-of-control ARL values of matched in-control upper one-sided EWMA charts with exact limits increase with $\lambda$, for all the values we considered for the magnitude of the shift.

The results in Table 1 and Fig. 1 (but also in Figure 1 of Frisén and Sonesson 2006) support the conclusion that $\lambda$ should approach to zero to minimize the out-of-control ARL (for a fixed in-control ARL), regardless of the magnitude of the shift.

In Table 2 we can find additional Monte-Carlo simulation results. They refer to the in-control and the out-of-control median RL of the upper one-sided EWMA chart

**Table 2** Critical values and out-of-control median RLs for upper one-sided EWMA charts with exact control limits and different smoothing parameters $\lambda$

| $\lambda$ | $c(\lambda, 500)$ | $a$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 0.25 | 0.5 | 1.0 | 1.5 | 2.0 | 3.0 | 4.0 |
| 0.0001 | 0.341738 | 3 | 2 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0.0003 | 0.541806 | 4 | 2 | 2 | 1 | 1 | 1 | 1 | 1 |
| 0.0005 | 0.653826 | 6 | 3 | 2 | 1 | 1 | 1 | 1 | 1 |
| 0.0007 | 0.735931 | 8 | 3 | 2 | 1 | 1 | 1 | 1 | 1 |
| 0.001 | 0.829928 | 11 | 4 | 2 | 1 | 1 | 1 | 1 | 1 |
| 0.003 | 1.176433 | 57 | 7 | 4 | 2 | 1 | 1 | 1 | 1 |
| 0.005 | 1.368107 | 137 | 11 | 5 | 2 | 1 | 1 | 1 | 1 |
| 0.007 | 1.503701 | 205 | 14 | 6 | 2 | 2 | 1 | 1 | 1 |
| 0.01 | 1.654164 | 263 | 18 | 7 | 3 | 2 | 1 | 1 | 1 |
| 0.03 | 2.119538 | 332 | 33 | 12 | 4 | 2 | 2 | 1 | 1 |
| 0.05 | 2.311206 | 336 | 39 | 14 | 5 | 3 | 2 | 1 | 1 |
| 0.07 | 2.432397 | 344 | 43 | 16 | 5 | 3 | 2 | 1 | 1 |
| 0.1 | 2.543225 | 345 | 48 | 17 | 6 | 3 | 2 | 1 | 1 |
| 0.2 | 2.716605 | 346 | 64 | 21 | 7 | 3 | 2 | 1 | 1 |
| 0.3 | 2.789789 | 347 | 78 | 26 | 7 | 4 | 2 | 1 | 1 |
| 0.4 | 2.828317 | 347 | 91 | 32 | 8 | 4 | 2 | 1 | 1 |
| 0.5 | 2.850393 | 347 | 104 | 38 | 9 | 4 | 3 | 1 | 1 |
| 1 | 2.878162 | 347 | 161 | 80 | 23 | 8 | 4 | 1 | 1 |

with exact control limits and lead to a similar conclusion: the closer $\lambda$ is to zero, the smaller is the median RL of this chart.

The increasing behaviour of the out-of-control ARL (and median RL) in terms of $\lambda$ is the main motivation for the derivation of what we shall call the limit chart in the next section.

## 3 The limit chart

So far we studied the distribution of $N_a(\lambda, c)$ as a function of $\lambda$ assuming that this smoothing parameter belongs to $(0, 1]$, which is reasonable since $N_a(\lambda, c)$ and $c(\lambda, \xi)$ are only defined for $\lambda \in (0, 1]$.

In this section we extend some of the results of Morais et al. (2012) and discuss the limiting distribution of $N_a(\lambda, c)$ as $\lambda$ tends to zero in order to derive a control chart—the upper one-sided limit chart—whose RL has such a limiting distribution.

**Proposition 5** *Assume that the random variables $\{Y_t\}$ are independent and identically distributed to $\mathcal{N}(\mu_0, \sigma^2)$. Then, for any fixed $k = 1, 2, \ldots$ and any fixed critical value $c$,*

$$\lim_{\lambda \to 0+} P[N_a(\lambda, c) > k] = F_{\mathcal{N}_k(\mathbf{0}_k, \mathbf{C}_k(0))}\left(c - a\sqrt{i},\ i = 1, \ldots, k\right), \qquad (24)$$

*where*

$$\mathbf{C}_k(0) = \lim_{\lambda \to 0+} \mathbf{C}_k(\lambda) = \lim_{\lambda \to 0+} [\mathrm{Corr}_0(Z_i, Z_j)]_{i,j=1,\ldots,k} = \left[ \sqrt{\frac{\min\{i, j\}}{\max\{i, j\}}} \right]_{i,j=1,\ldots,k}. \tag{25}$$

Note that result (24) holds for any value of $a$, therefore for the in-control and the out-of-control states as described by (3).

*Proof* (Proposition 5)—Capitalizing on the L'Hôpital rule and on results (18) and (14) we successively get

$$\lim_{\lambda \to 0+} \mathrm{Corr}_0(Z_i, Z_j) = \lim_{\lambda \to 0+} \sqrt{\frac{(1-\lambda)^{-2i} - 1}{(1-\lambda)^{-2j} - 1}} = \sqrt{\frac{i}{j}}, \; 1 \le i \le j \le k, \tag{26}$$

that is, $\mathbf{C}_k(0) = \left[ \sqrt{\frac{\min\{i,j\}}{\max\{i,j\}}} \right]_{i,j=1,\ldots,k}$, and

$$\begin{aligned} &\lim_{\lambda \to 0+} P[N_a(\lambda, c) > k] \\ &\quad = \lim_{\lambda \to 0+} F_{\mathcal{N}_k(\mathbf{0}_k, \mathbf{C}_k(\lambda))} \left( c - a \sqrt{\frac{2-\lambda}{\lambda} \times \frac{1-(1-\lambda)^i}{1+(1-\lambda)^i}}, \; i = 1, \ldots, k \right) \\ &\quad = F_{\mathcal{N}_k(\mathbf{0}_k, \mathbf{C}_k(0))} \left( c - a\sqrt{i}, \; i = 1, \ldots, k \right), \; k = 1, 2, \ldots. \end{aligned} \tag{27}$$

□

(24) clearly suggests the use of a linear combination of $X_i$, $i = 1, \ldots, t$, as the control statistic at time $t$ of the upper one-sided limit chart, as stated in the next proposition.

**Proposition 6** *The upper one-sided limit chart makes use of the overall mean at sample time t,* $\bar{X}_t = \frac{1}{t} \sum_{i=1}^{t} X_i$*, and its RL,*

$$N_a(0, c) = \inf \left\{ t \in \mathbb{N} : \bar{X}_t > E_0(\bar{X}_t) + c\sqrt{Var_0(\bar{X}_t)} = \mu_0 + c\sigma/\sqrt{t} \right\}, \tag{28}$$

*has the survival function defined by* $\lim_{\lambda \to 0+} P[N_a(\lambda, c) > k]$ *in (24).*

*Remark 7*

1. $\bar{X}_t$ happens to be the best linear unbiased overall estimator of $\mu$.
2. The RL of the upper one-sided limit chart can be rewritten as $\inf\{t \in \mathbb{N} : \sum_{i=1}^{t} (X_i - \mu_0) > c\sigma\sqrt{t}\}$, which can be thought as the RL of an upper one-sided repeated significance test (please refer to Siegmund (1977) or Morais et al. 2012 for more details).

3. If we use the control statistic $\frac{1}{\sqrt{t}}\sum_{i=1}^{t}(X_i - \mu_0)$ instead of $\bar{X}_t$, we are now dealing with a chart similar to the CUSUM procedure from the structural change (econometrics) literature.
4. Moreover, if we note that $\bar{X}_t = (1-\frac{1}{t})\bar{X}_{t-1}+\frac{1}{t}X_t$, the limit chart can be interpreted as an upper one-sided EWMA chart with time-varying smoothing parameters $\lambda_t = \frac{1}{t}$ and exact control limits $\mu_0 + c\sigma/\sqrt{t}$.
5. It should be emphasized that the upper one-sided limit chart is not a moving average chart as defined by (Montgomery 2009, p. 419).

*Proof* (Proposition 6)—The result can be immediately proved. Since $E_a(\bar{X}_i) = \mu_0 + a\sigma$, $\mathrm{Var}_0(\bar{X}_i) = \frac{\sigma^2}{i}$ and

$$\mathrm{Corr}_0(\bar{X}_i, \bar{X}_j) = \frac{\frac{1}{i}\frac{1}{j}\sum_{l=1}^{\min\{i,j\}}\mathrm{Cov}_0(X_l, X_l)}{\sqrt{\frac{\sigma^2}{i}\frac{\sigma^2}{j}}} = \sqrt{\frac{\min\{i, j\}}{\max\{i, j\}}}, \tag{29}$$

for $i, j = 1, \ldots, k$, we consequently get $\left[\mathrm{Corr}_0(\bar{X}_i, \bar{X}_j)\right]_{i,j=1,\ldots,k} = \lim_{\lambda\to 0+}\mathbf{C}_k(\lambda) = \mathbf{C}_k(0)$ and

$$\begin{aligned} P[N_a(0, c) > k] &= P_a\left[\bar{X}_i \le E_0(\bar{X}_i) + c\sqrt{\mathrm{Var}_0(\bar{X}_i)},\ i = 1, \ldots, k\right] \\ &= P_a\left[\frac{\bar{X}_i - E_a(\bar{X}_i)}{\sqrt{\mathrm{Var}_0(\bar{X}_i)}} \le c - a\sqrt{i},\ i = 1, \ldots, k\right] \\ &= F_{\mathcal{N}_k(\mathbf{0}_k, \mathbf{C}_k(0))}\left(c - a\sqrt{i},\ i = 1, \ldots, k\right) \\ &= \lim_{\lambda\to 0+} P[N_a(\lambda, c) > k], \end{aligned} \tag{30}$$

for $k = 1, 2, \ldots$.

Assuming that (3) is valid, we can interpret the chart described in Proposition 6 as the limit of the upper one-sided EWMA chart with exact control limits when $\lambda$ tends to zero, thus, concluding the proof. □

### 3.1 Illustrating the performance of the limit chart

It is time to assess the performance of the (upper one-sided) limit chart and confront it with other matched in-control charts ($\xi = 500$). For that matter, let us consider the following charts and parameters:

A. upper one-sided limit chart—$c(0, 500) = -0.084701$;
B. upper one-sided EWMA chart with exact control limits—$\lambda = 0.1$ and $c(0.1, 500) = 2.543225$;
C. upper one-sided EWMA chart with asymptotic control limits—$\lambda = 0.1$ and $c = 2.532760$; note that the value of $\lambda$ is in the interval $[0.05, 0.25]$, as recommended by Montgomery (Montgomery 2009, p. 423);
D. upper one-sided Shewhart chart—$c(1, 500) = 2.878162$;

**Table 3** In-control and out-of-control ARL of the upper one-sided limit chart (A) and four other matched in-control competing charts (B–E)

| Chart | $a$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0 | 0.25 | 0.5 | 1.0 | 1.5 | 2.0 | 3.0 | 4.0 |
| A | 501.1468 | 2.9139 | 1.7614 | 1.2122 | 1.0670 | 1.0197 | 1.0010 | 1.0000 |
| B | 499.7454 | 66.9442 | 21.6346 | 6.7607 | 3.5398 | 2.3040 | 1.3671 | 1.0733 |
| C | 500.2899 | 70.3600 | 24.7263 | 8.9078 | 5.3898 | 3.9152 | 2.6044 | 2.0577 |
| D | 500.0000 | 232.9707 | 114.9479 | 33.1351 | 11.8939 | 5.2652 | 1.8232 | 1.1507 |
| E | 500.4931 | 98.2612 | 30.8521 | 9.1548 | 5.1368 | 3.6029 | 2.3409 | 1.8456 |

E. upper one-sided CUSUM chart—$K = 0.5$ and $h = 4.38913$, as optimal reference value and upper control limit; recall that the control statistic of this scheme is $C_t^+ = \max\left\{0, C_{t-1}^+ + X_t - K\right\}$, with $C_0^+ = 0$, and where $K$ equals to half-distance between the target and the out-of-control process mean; settting $K = 0.5$ implies that we anticipate a shift from $\mu_0 = 0$ to $\mu_1 = 1$; and considering $h = 4.38913$ leads to an in-control ARL around $\xi = 500$ (according to (Montgomery 2009, Sect. 9.2.2), the upper one-sided CUSUM chart with this reference value and the upper one-sided EWMA chart with asymptotic control limits and $\lambda = 0.1$ have a similar behaviour when it comes to the ARL for $a = 1$).

It should be also noted that all simulated RL is truncated if it is larger than 50,000 samples, except for $\lambda = 0$, as in Sect. 2.3.

We ought to comment the negative character of the critical value of the upper one-sided limit chart—a surprising result thoroughly discussed by Morais et al. (2012), namely by its Theorem 3, stated below.

Assume that the random variables $\{Y_t\}$ are independent and identically distributed with mean $\mu$ and variance $\gamma_0$. (a) If $c \geq 0$ then $E[N_0(0, c)] = \infty$. (b) Suppose that $P[Y_t = 0] < 1$ and that the variables $\{Y_t\}$ are symmetric around $\mu$. If $c < 0$ then $E[N_0(0, c)] < \infty$ and $\text{Var}[N_0(0, c)] < \infty$.

This result is remarkable: it implies that the in-control ARL of the upper one-sided limit chart is equal to infinity if the control limit is nonnegative.

From Table 3 it is apparent that the upper one-sided limit chart outperforms all the four other matched in-control competing charts.

This table certainly supports the well known fact that Shewhart charts tend to be faster than the EWMA charts with asymptotic limits in the detection of very large shifts. However, when we compare their (zero-state) ARL values with the ones of the upper one-sided EWMA with exact control limits (B) or of the upper one-sided limit chart (A), we immediately recommend abandoning both the upper one-sided Shewhart chart (D) and the upper one-sided EWMA chart with asymptotic limits (C). The reader should recall that using the exact control limits for the EWMA chart is one of several possibilities to introduce fast initial response, as put by Steiner (1999) and Knoth (2005).

### 3.2 On the stochastic comparison of the run lengths of the limit and EWMA charts with exact control limits

It should be noted that we failed to establish a stochastic order relation between the RL of the (upper one-sided) limit chart, $N_a(0, c(0, \xi))$, and the RL of the matched in-control upper one-sided EWMA with exact control limits, $N_a(\lambda, c(\lambda, \xi))$, for any fixed values of $a$ ($a > 0$), $\xi$ ($\xi > 1$) and $\lambda$ ($\lambda \in (0, 1]$).

In fact, by capitalizing on the increasing behaviour of $c(\lambda, \xi)$ and the decreasing behaviour of $\sqrt{\frac{2-\lambda}{\lambda} \times \frac{1-(1-\lambda)^i}{1+(1-\lambda)^i}}$ in terms of $\lambda$, we can conclude that

$$\begin{aligned} P[N_a(0, c(0, \xi)) > k] &= F_{\mathcal{N}_k(\mathbf{0}_k, \mathbf{C}_k(0))}\left(c(0, \xi) - a\sqrt{i},\ i = 1, \ldots, k\right) \\ &\leq F_{\mathcal{N}_k(\mathbf{0}_k, \mathbf{C}_k(0))}\left(c(\lambda, \xi) - a\sqrt{\frac{2-\lambda}{\lambda} \times \frac{1-(1-\lambda)^i}{1+(1-\lambda)^i}},\ i = 1, \ldots, k\right), \end{aligned} \tag{31}$$

for $k = 1, 2, \ldots$. However, by using Theorem 5.1.7 by Tong (1990, p. 103), we get

$$\begin{aligned} &F_{\mathcal{N}_k(\mathbf{0}_k, \mathbf{C}_k(0))}\left(c(\lambda, \xi) - a\sqrt{\frac{2-\lambda}{\lambda} \times \frac{1-(1-\lambda)^i}{1+(1-\lambda)^i}},\ i = 1, \ldots, k\right) \\ &\geq F_{\mathcal{N}_k(\mathbf{0}_k, \mathbf{C}_k(\lambda))}\left(c(\lambda, \xi) - a\sqrt{\frac{2-\lambda}{\lambda} \times \frac{1-(1-\lambda)^i}{1+(1-\lambda)^i}},\ i = 1, \ldots, k\right) \\ &= P[N_a(\lambda, c(\lambda, \xi)) > k], \end{aligned} \tag{32}$$

for $k = 1, 2, \ldots$, thus failing to prove that $N_a(0, c(0, \xi)) \leq_{st} N_a(\lambda, c(\lambda, \xi))$, for any fixed values of $a$ ($a > 0$), $\xi$ ($\xi > 1$) and $\lambda$ ($\lambda \in (0, 1]$). Needless to say, that a similar difficulty arises when we try to prove that the RL $N_a(\lambda, c)$ stochastically decreases with $\lambda$ for a fixed and common critical value $c$.

Difficulties such as these do not come as a surprise, namely for the matched in-control run lengths $N_0(0, c(0, \xi))$ and $N_0(\lambda, c(\lambda, \xi))$, since Theorem 1.A.7 of (Shaked and Shanthikumar 1994, p. 8) can be restated as follows:

- if $N_0(0, c(0, \xi)) \leq_{st} N_0(\lambda, c(\lambda, \xi))$ and $E[N_0(0, c(0, \xi))] = E[N_0(\lambda, c(\lambda, \xi))]$ then $N_0(0, c(0, \xi)) =_{st} N_0(\lambda, c(\lambda, \xi))$.

But since we know that $N_0(0, c(0, \xi)) \neq_{st} N_0(\lambda, c(\lambda, \xi))$ and $E[N_0(0, c(0, \xi))] = E[N_0(\lambda, c(\lambda, \xi))]$ for matched in-control charts, we can conclude that these two RL cannot be compared in the usual sense, i.e. $N_0(0, c(0, \xi)) \not\leq_{st} N_0(\lambda, c(\lambda, \xi))$.

Curiously enough, if $a = 0$ and we consider a common critical value $c(\lambda, \xi)$, we can rewrite (32) as

$$\begin{aligned} P[N_0(0, c(\lambda, \xi)) > k] &= F_{\mathcal{N}_k(\mathbf{0}_k, \mathbf{C}_k(0))}\,(c(\lambda, \xi),\ i = 1, \ldots, k) \\ &\geq F_{\mathcal{N}_k(\mathbf{0}_k, \mathbf{C}_k(\lambda))}\,(c(\lambda, \xi),\ i = 1, \ldots, k) \\ &= P[N_0(\lambda, c(\lambda, \xi)) > k], \end{aligned} \tag{33}$$

for $k = 1, 2, \ldots$, that is,

$$N_0(0, c(\lambda, \xi)) \geq_{st} N_0(\lambda, c(\lambda, \xi)). \tag{34}$$

This result can be read as follows: the probability of no false alarms within the first $k$ samples is larger for the upper one-sided limit chart than for the upper one-sided EWMA chart with exact control limits if the same critical value $c(\lambda, \xi)$ is used in both charts.

Interestingly enough, if the upper one-sided limit chart is used with its own critical value $c(0, \xi)$, the probability of a false alarm at the first sample is high. The implications of this result and a solution will be further explored in Sect. 4.2.

Furthermore, since assuming that the shift occurs at time $t = 1$ is not very realistic, other performance measures than the ARL are also discussed in the next section.

## 4 Further investigations

### 4.1 Exact vs. asymptotic control limits

So far we have investigated the impact of $\lambda$ in the in-control and out-of-control ARL of upper one-sided EWMA charts with exact control limits, and in the range of the exact control limits of such charts with a common in-control ARL value. These investigations led to what we called the limit chart. Now it is time to examine how do the critical values and the out-of-control ARL change with $\lambda$ when we adopt asymptotic control limits and therefore the RL of the chart is defined by

$$\inf \left\{ t \in I\!N : Z_t > E_0(Z_t) + c \sqrt{\lim_{t \to +\infty} \mathrm{Var}_0(Z_t)} \right\}. \tag{35}$$

Our numerical investigations considering the same set of parameters as in Sect. 2.3 led to the obvious conclusion, made apparent from Fig. 2, that critical values associated to exact control limits are obviously larger than the ones associated to asymptotic control limits of matched in-control upper one-sided EWMA charts.

In fact, if we consider upper one-sided EWMA charts with exact and asymptotic control limits with the same critical value, say $c$, the chart with asymptotic control limits has a stochastically larger (in and out-of-control) RL since those limits are larger than the exact ones. This stochastic behaviour implies in turn the use of smaller critical value, say $c' < c$ in order to match in-control the upper one-sided EWMA chart with asymptotic control limits and the correspondent chart with exact control limits.

Moreover, these critical values seem to take quite similar values for large values of $\lambda$, namely larger than 0.2 in our numerical example. However, when asymptotic

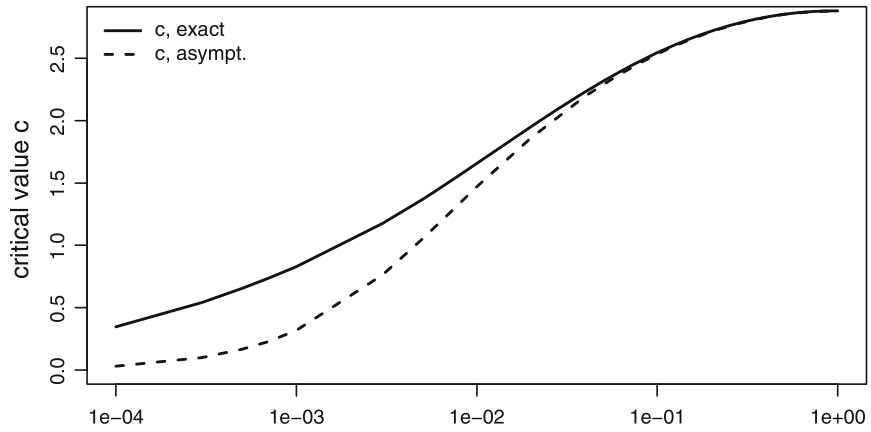


**Fig. 2** Critical values of matched in-control upper one-sided EWMA charts with exact and asymptotic control limits as a function of $\log(\lambda)$

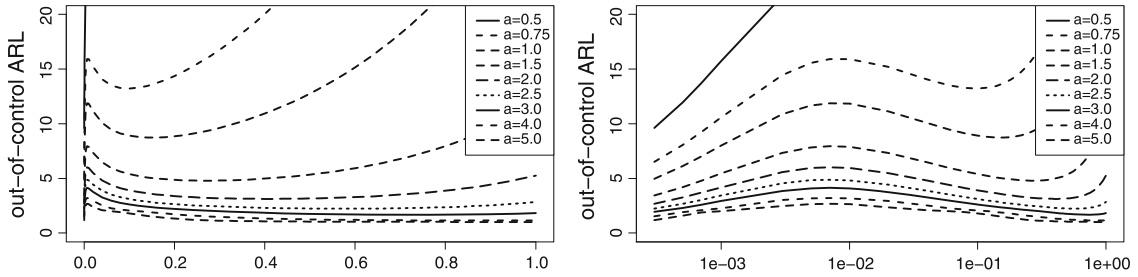


**Fig. 3** Out-of-control ARL of matched in-control upper one-sided EWMA charts with asymptotic control limits as a function of $\lambda$ (on the *left*) and of $\log(\lambda)$ (on the *right*)

control limits are at use, the out-of-control ARL is no longer an increasing function of $\lambda$, as shown by Fig. 3.

Deriving the survival function of the RL does not shed any light on this non monotone behaviour. In fact, this survival function is given by

$$F_{\mathcal{N}_k(\mathbf{0}_k,\mathbf{C}_k(\lambda))}\left(\frac{c}{\sqrt{1-(1-\lambda)^{2i}}} - a\sqrt{\frac{2-\lambda}{\lambda}\times\frac{1-(1-\lambda)^i}{1+(1-\lambda)^i}},\ i=1,\ldots,k\right), \tag{36}$$

for $k = 1, 2, \ldots$, which is quite similar to Equation (14): the critical value $c$ is different from the one of the in-control matched chart with exact control limits and it is now regretfully divided by an increasing function of $\lambda$.

Despite of this discouraging result, we are able to add that, when $a = 0$, the survival function in (36) decreases with $\lambda$, for any fixed $k = 1, 2, \ldots$ and any fixed critical value $c$. This means that the in-control RL of upper one-sided EWMA charts with both exact and asymptotic control limits stochastically decreases with $\lambda$ for any fixed critical value (please refer to Theorem 1).

It is also worth mentioning that, when asymptotic control limits are at use, the out-of-control ARL values in Fig. 3 are larger than the ones of the matched in-control upper one-sided EWMA chart with exact control limits in Table 1 and Fig. 1. Furthermore, if we superimpose Figs. 1 and 3 we conclude that: the out-of-control ARL values

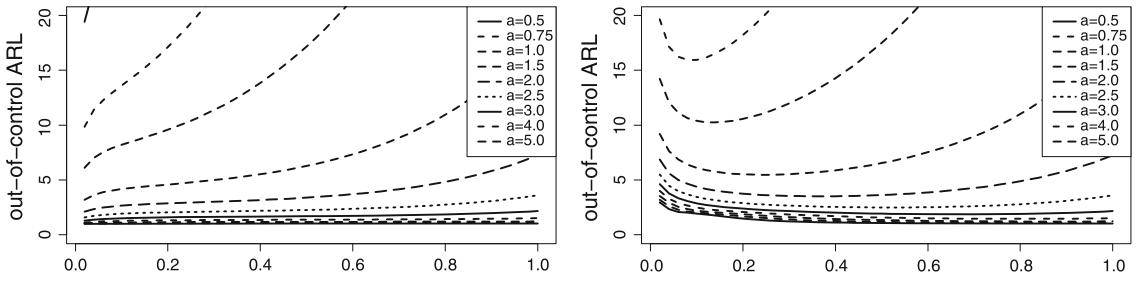


**Fig. 4** Out-of-control ARL of matched in-control two-sided EWMA charts—with exact control limits (on the *left*) and asymptotic ones (on the *right*)—as a function of $\lambda$

of the upper one-sided EWMA chart with asymptotic control limits converge to the ones associated to exact control limits, as $\lambda$ tend to one; the in-control ARL of the upper one-sided EWMA charts with exact or asymptotic limits seem to reach a global minimum value as $\lambda \to 0$.

Even though the detection of both increases and decreases in the process mean is beyond the scope of this paper, we briefly explore the impact of $\lambda$ when an EWMA chart with exact control limits is at use and when these limits are replaced by the asymptotic ones. Let us remind the reader that the two-sided EWMA chart with exact control limits has the following RL:

$$\inf\left\{ t \in I\!N : Z_t < E_0(Z_t) - c\sqrt{\mathrm{Var}_0(Z_t)} \text{ or } Z_t > E_0(Z_t) + c\sqrt{\mathrm{Var}_0(Z_t)} \right\}. \tag{37}$$

As shown by the two graphs of Fig. 4, the out-of-control ARL of the two-sided EWMA chart with exact control limits still increases with $\lambda$ but this monotonicity no longer holds when these limits are replaced by the asymptotic ones, as in the upper one-sided case.

## 4.2 On the in-control behaviour of the limit chart

A close inspection of the simulated values of the in-control RL of the upper one-sided limit chart made us realize that the probability of a false alarm at the first sample is very high. In fact, it is equal to

$$P_0\left[\bar{X}_1 = X_1 > \mu_0 + c(0,\xi)\,\sigma\right] = 1 - \Phi[c(0,\xi)]; \tag{38}$$

and since the absolute value of $c(0,\xi)$ is small this probability is large and the number of false alarms at sample 1 higher than the one of the competing charts even though they are matched in-control.

The impact of this undesired property can be reasonably minimized by adding a negative head start value, say $HS^-$, to $X_1$, therefore the control statistic now reads as

$$\bar{X}_t^- = \frac{HS^-}{t} + \frac{1}{t}\sum_{i=1}^{t} X_i. \tag{39}$$

**Table 4** Percentage of simulated in-control RL equal to one, in-control and out-of-control ARL of the upper one-sided limit charts A, A1 and A2, and the competing charts B–E

| Chart | % of RL equal to one | $a$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 0 | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 |
| A | 0.5336 | 501.1468 | 11.6335 | 6.2806 | 4.4263 | 3.4894 | 2.9139 |
| A1 | 0.0020 | 2843.5256 | 62.7517 | 33.0347 | 22.8104 | 17.5949 | 14.4308 |
| A2 | 0.0038 | 2593.7166 | 57.4177 | 30.2333 | 20.9299 | 16.1643 | 13.2812 |
| B | 0.0055 | 500.3034 | 304.8160 | 195.2651 | 130.8202 | 91.8641 | 66.9634 |
| C | 0.0000 | 500.2899 | 307.2937 | 198.8014 | 134.4792 | 95.4423 | 70.4296 |
| D | 0.0020 | 500.0000 | 427.2030 | 365.8488 | 314.0291 | 270.1700 | 232.9707 |
| E | 0.0000 | 500.4931 | 349.6541 | 247.8247 | 178.5377 | 131.2821 | 98.2469 |

$HS^-$ should chosen in order to achieve a reasonably small value for the probability in (38). For instance, the probability of triggering a false alarm at the first sample while using the upper one-sided limit chart could be taken as equal to the one of the upper one-sided Shewhart chart and therefore

$$HS^- = HS^-(1/\xi) = c(0, \xi) - \Phi^{-1}(1 - 1/\xi). \tag{40}$$

We ought to note that head start values are throughly recommended in the literature (e.g. Lucas and Crosier (1982)). Positive (resp. negative) head start values are used by quality control practitioners when dealing with upper one-sided (resp. lower one-sided) charts. The rationale is as follows: if the process is operating in control, the control statistic of the chart is soon brought to zero, so that the expected effect of the head start is minimal; otherwise, the operator is alerted to the out-of-control situation much sooner, which may prevent start-up problems.

In our specific case, the adoption of a negative head start while using an upper one-sided chart limit chart has a sort of opposed effect. As a matter of fact, there is a stochastic increase of RL and, thus, larger in-control and out-of-control ARL values, and we are no longer dealing with matched in-control charts, as illustrated by Table 4 for the upper one-sided limit charts with critical value $c(0, \xi = 500) = -0.084701$ and

A1. $HS^-(1/\xi) = c(0, \xi) - \Phi^{-1}(1 - 1/\xi) \simeq= -3$;
A2. $HS^-(2/\xi) = c(0, \xi) - \Phi^{-1}(1 - 2/\xi) \simeq= -2.7$.

However, the upper one-sided limit charts with these two head starts (A1–A2) can still outperform the competing ones (B–E) in the detection of very small shifts (e.g. $a = 0.05, 0.1, 0.15, 0.2$) in terms of the (zero-state) ARL, as shown by Table 4. Moreover, the adoption of head starts $HS^-(1/\xi)$ and $HS^-(2/\xi)$ substantially decreased the percentage of simulated in-control RL values equal to one of the upper one-sided limit chart, as we can see from Table 4.

It is worth noticing that there was no single false alarm at the first time point in these simulation results for the upper one-sided EWMA chart with asymptotic limits (C) and the upper one-sided CUSUM chart (E). This is essentially due to the fact that

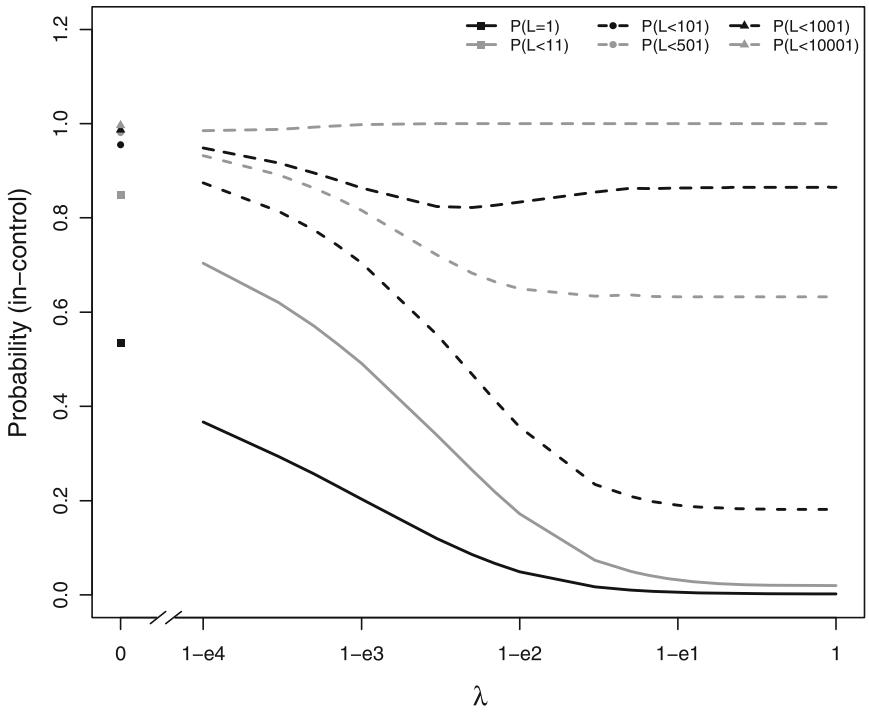


**Fig. 5** Upper one-sided EWMA with asymptotic limits—probabilities that the in-control RL is equal to one or is smaller than some prespecified value $l + 1$, $l = 10, 100, 500, 1000, 10000$, as a function of $\lambda$

the probabilities of these occurrences (pfa) are very small. In fact, in our examples, they are equal to $pfa_C = P(\lambda X_1 > c) = 1 - \Phi\left(2\sqrt{1/[\lambda(2-\lambda)]}\right) \simeq 3 \times 10^{-9}$ and $pfa_F = P(X_1 - K > h) = 1 - \Phi(K - h) \simeq 5 \times 10^{-5}$, respectively.

We ought to add that the upper one-sided EWMA charts with small values of the smoothing parameter are also associated to large values of the probability of a false alarm at the first sample, as illustrated by Fig. 5. This figure also includes the profiles of the probability of a false alarm before the collection of $l + 1$ samples ($l = 10, 100, 500, 1000, 10000$) and lead us to conclude that, despite of the in-control ARL of 500, the probability of a false alarm within, for instance, the first 500 samples is quite high for $\lambda < 0.1$, thus, suggesting that the upper one-sided EWMA charts with small $\lambda$ should be used with great care.

Finally, note that these new set of simulations (and the ones in the next subsection) were based on $10^6$ replications for charts A–E.

### 4.3 The (maximum) conditional average delay

Up to now we have assumed that the shift occurs at time $t = 1$ and this can be unrealistic. So we ought to investigate the performance of both the upper one-sided EWMA with exact control limits and the upper one-sided limit chart when the shift occurs at an arbitrary time, say $q$ ($q \geq 1$). In order to do that, let us consider the model

$$X_t = \begin{cases} Y_t, & t < q \\ Y_t + a, & t \geq q, \end{cases} \tag{41}$$

where $q \geq 1$ and is usually called the change point.

Under this model the ARL is no longer an adequate chart performance. In fact, the essential tools in the comparison of the performances of several competing charts under (41) are the conditional average delays

$$CAD_a(q, \lambda, c) = E[N_a(\lambda, c) - q + 1 | N_a(\lambda, c) \geq q], \tag{42}$$

which corresponds to the number of observations until the detection of a shift of magnitude $a$, conditionally on the fact that it occurred at time $q$ (see, for example, Knoth 2003), and the maximum conditional average delay,

$$MCAD_a(\lambda, c) = \max_{q=1,2,\ldots} CAD_a(q, \lambda, c). \tag{43}$$

We begin with an illustration with simulated conditional average delay values referring to matched in-control ($\xi = 500$) upper one-sided limit chart (A) and a few upper one-sided EWMA charts with exact control limits and different values of $\lambda$.

Firstly, let us remind the reader that $CAD_a(q, \lambda = 1, c) = E[N_a(\lambda = 1, c)]$, $q = 1, 2, \ldots$, for any (upper one-sided) Shewhart chart, after all the RL of this chart has a geometric distribution. Please also note that, under the change point model (41), we have, for $a > 0$:

- $E_a(Z_t) = \mu_0 + a\,\sigma\left[1 - (1-\lambda)^{t-q+1}\right]$, for $\lambda \in (0, 1]$ and considering $Z_0 = \mu_0$;
- $E_a(\frac{1}{t}\sum_{i=1}^{t} X_i) = \mu_0 + a\,\sigma\,\frac{t-q+1}{t}$.

Figure 6 provides the plots of CAD as a function of the change point $q$ and $\lambda$, where the horizontal line in each graph on the right hand side refers to $\lambda = 1$ (Shewhart chart).

It is apparent from these plots that the detection ability of upper one-sided EWMA charts with exact control limits and small values of $\lambda$, as well as the one of the upper one-sided limit chart, is heavily dependent on the change point $q$. In fact, $CAD_a(q, \lambda, c)$ is an nondecreasing function of $q$ in the absence of head-starts, reflecting the fact that the control statistics $Z_t$ and $\frac{1}{t}\sum_{i=1}^{t} X_i$ underestimate the out-of-control process mean $\mu_0 + a\sigma$, for $t \geq q$, and the absolute value of their biases, $|-a\,\sigma(1-\lambda)^{t-q+1}|$ and $|-a\,\sigma\,\frac{t-q+1}{t}|$, both increase with $q$.

Moreover, if the shift occurs quickly the charts with smaller values of $\lambda$ outperform charts with high smoothing parameters under the same criterion. However, if the shift occurs later the charts with weaker memory perform better. It is also worth adding that, for shifts of unit size, the control charts with $\lambda = 0.5$ or larger become almost insensitive to the change point $q$.

Unlike the out-of-control ARL, $CAD_a(q, \lambda, c)$ is not an increasing function of $\lambda$, as pictured by Fig. 7, for $q = 250, 500, 750, 1000$. As a consequence, the upper one-sided limit chart is not necessarily better than all of the matched in-control upper one-sided EWMA charts with exact control limits (or the Shewhart chart), under the CAD criterion. For instance, the optimal value for the smoothing parameter is around 0.1, for $a = 1$.

Unsurprisingly, the maximum conditional average delay (MCAD) is not an increasing function of $\lambda$ and therefore the upper one-sided limit chart can be outperformed by

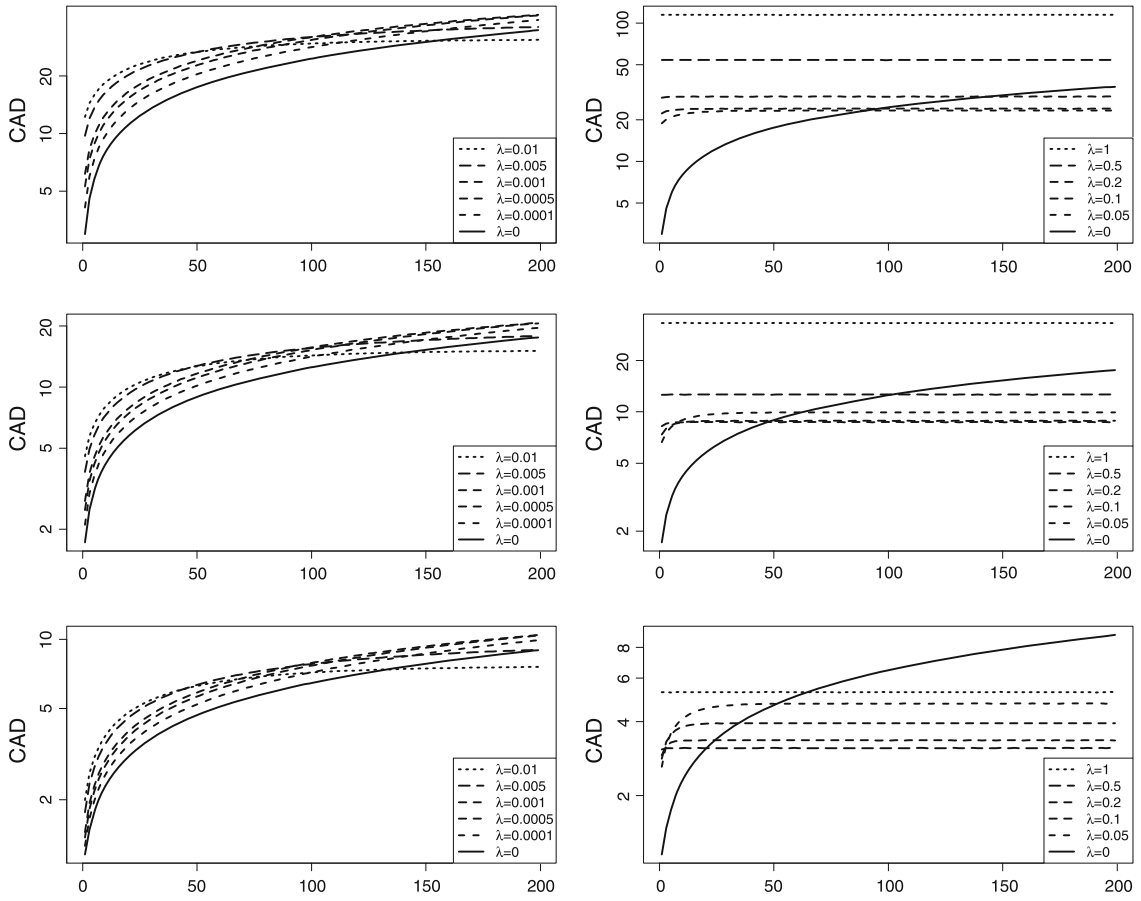


**Fig. 6** Conditional average delay as a function of $q$, for shifts $a = 0.5, 1, 2$ (from *top* to *bottom*) and different values of $\lambda$

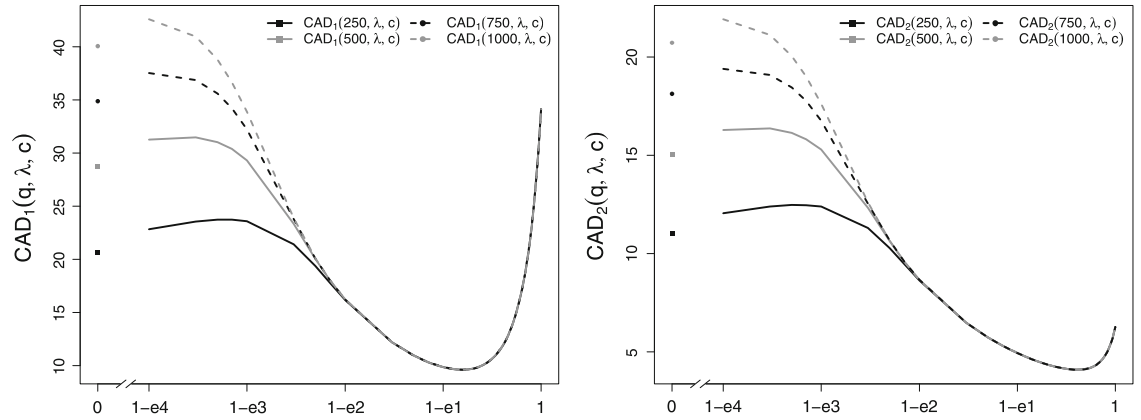


**Fig. 7** Conditional average delay for different values of the shift point $q$ as a function of $\lambda$, for shift of size $a = 1$ (*left*) and $a = 2$ (*right*)

some matched in-control upper one-sided EWMA with exact control limits, namely the ones with $\lambda$ in the interval [0.05, 0.1], as illustrated by Fig. 8.

Finally, we summarize some MCAD values in Table 5.These values refer to the charts A–E defined in Sect. 3.1, by considering the following approximation to MCAD in formula (43): $\max_{q=1,2,\ldots,q_{max}} CAD_a(q)$, where $q_{max} = 1000$.

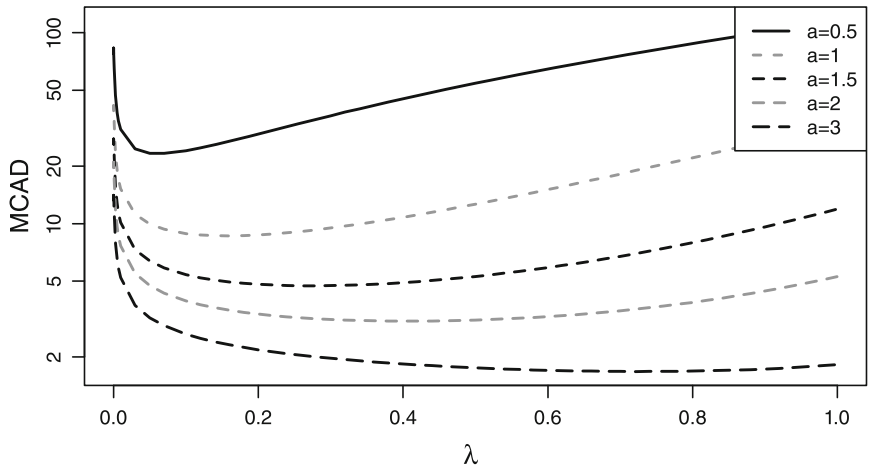


**Fig. 8** Maximum conditional average delay as a function of λ, for several shifts

**Table 5** Maximum conditional average delays of the upper one-sided limit chart (A) and five other matched in-control competing charts (B–E)

| Chart | $a$ | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0.25 | 0.5 | 1.0 | 1.5 | 2.0 | 3.0 | 4.0 |
| A | 156.0252 | 78.9033 | 40.1450 | 27.2221 | 20.7695 | 14.3244 | 11.1077 |
| B | 70.5266 | 24.1794 | 8.9059 | 5.4083 | 3.9467 | 2.6459 | 2.0534 |
| C | 69.9194 | 24.1532 | 8.8542 | 5.3910 | 3.9322 | 2.6355 | 2.0474 |
| D | 232.9708 | 114.9479 | 33.1351 | 11.8939 | 5.2652 | 1.8232 | 1.1507 |
| E | 97.2617 | 30.3813 | 8.8689 | 4.9589 | 3.4709 | 2.2635 | 1.7696 |

The reader should be also aware that: $\max_{q=1,2,\ldots,q_{max}} CAD_a(q)$ increases with $q_{max}$; this approximation to MCAD does not change if $q_{max}$ is increased beyond 100, while using the competing charts B–E.

It is apparent from Table 5 that, according to the MCAD performance criterion, the upper one-sided limit chart (A) is only able to outperform the upper one-sided Shewhart chart (D) for small shifts. Furthermore, this criterion favours the upper one-sided CUSUM chart (E) for moderate shifts and the upper one-sided Shewhart chart (D) for large shifts.

## 5 Conclusions

This paper essentially provides a thorough study on the behaviour of the upper one-sided EWMA chart with exact control limits when the smoothing parameter λ converges to zero. We ought to note that the resulting chart—the upper one-sided limit chart—has small out-of-control (zero-state) ARL values when compared to other matched in-control EWMA (and CUSUM) charts. However, we cannot advocate the use of the limit chart, namely because:

- it triggers false alarms quite frequently in the first samples;
- its conditional average delay profile is highly dependent on the change point $q$;

- its CAD and maximum conditional average delay are not necessarily better than the ones of the remaining matched in-control upper one-sided EWMA charts with exact control limits.

Finally, this study also brought to light a few useful results in the design of EWMA charts. For instance, we proved that:

- increasing $\lambda$ yields leads to a stochastically smaller in-control RL, within the family of upper one-sided EWMA charts with the same critical value;
- increasing $\lambda$ requires a larger critical value if we decide to deal with a matched in-control chart.

**Acknowledgments** The first author was partially supported by the *Fundação para a Ciência e a Tecnologia* (FCT) during the preparation of the final draft of this paper, while visiting the Department of Statistics of the European University Viadrina (Frankfurt–Oder). The remaining authors acknowledge the financial support given by the Center for Mathematics and its Applications (CEMAT) provided for visiting Instituto Superior Técnico during an early stage of the preparation of this paper. The authors are grateful to the Editor-in-Chief and the two anonymous Rewiewers for their extremely valuable suggestions and comments and for sharing their enthusiasm with us.

## References

Baxley RV Jr (1990) Discussion of Exponentially weighted moving average control schemes: properties and enhancements by J.M. Lucas and M.S. Saccucci. Technometrics 32:13–16

Conte SD, de Boor C (1980) Elementary numerical analysis: an algorithmic approach, 3rd edn. Mc. Graw-Hill, London

Crowder SV, Hamilton MD (1992) An EWMA for monitoring a process standard deviation. J Qual Technol 24:12–21

Frisén M, Sonesson C (2006) Optimal surveillance based on exponentially weighted moving averages. Seq Anal 25:379–403

Hunter JS (1986) The exponentially weighted moving average. J Qual Technol 18:203–210

Juran JM (1997) Early SQC: a historical supplement. Qual Prog 30:73–81

Knoth S (2003) EWMA schemes with nonhomogenous transition kernels. Seq Anal 22:241–255

Knoth S (2005) Fast initial response features for EWMA control charts. Stat Pap 46:47–64

Kramer H, Schmid W (1997) EWMA charts for multivariate time series. Seq Anal 16:131–154

Lucas JM, Saccucci MS (1990) Exponentially weighted moving average control schemes: properties and enhancements. Technometrics 32:1–12

Lucas JM, Crosier RB (1982) Fast initial response for CUSUM quality-control schemes: give your CUSUM a head start. Technometrics 24:199–205

Montgomery DC (2009) Introduction to statistical quality control, 6th edn. Wiley, New York

Morais MC, Okhrin Y, Schmid W (2012) Limit properties of EWMA charts for stationary processes. In: Lenz HJ, Schmid W, Wilrich PTh (eds) Frontiers of statistical quality control, vol 10. Springer, New York, pp 69–85

Roberts SW (1959) Control charts tests based on geometric moving averages. Technometrics 1:239–250

Schmid W (1997) On EWMA charts for time series. In: Lenz HJ, Wilrich P-Th (eds) Frontiers of statistical quality control, vol 5. Physica, Heidelberg, pp 115–137

Shaked M, Shanthikumar JG (1994) Stochastic orders and their applications. Academic Press Inc, New York

Shewhart WA (1931) Economic control of quality of manufactured product. Princeton, Van Notrand

Siegmund D (1977) Repeated significance tests for a normal mean. Biometrika 64:177–189

Steiner SH (1999) EWMA control charts with time-varying control limits and fast initial response. J Qual Technol 31:75–86

Stoumbos ZG, Reynolds MR Jr, Ryan TP, Woodall WH (2000) The state of statistical process control as we proceed into the 21st century. J Am Stat Assoc 95:992–998

Tong YL (1990) The multivariate normal distribution. Springer, New York