What happens to EWMA control charts when λ converges to zero?

Manuel Cabral Morais[1], Yarema Okhrin[2], and Wolfgang Schmid[3]

[1] CEMAT & Department of Mathematics, Instituto Superior Técnico,
Av. Rovisco Pais, 1049-001 Lisboa, Portugal
maj@math.ist.utl.pt

[2] Department of Economics, University of Bern,
Schanzeneckstr. 1, CH-3012 Bern, Switzerland
yarema.okhrin@vwi.unibe.ch

[3] Department of Statistics, European University Viadrina,
D-15207 Frankfurt(Oder), Germany
schmid@euv-frankfurt-o.de

Abstract. We provide a brief study on the behaviour of upper one-sided EWMA charts with exact control limits when its smoothing parameter λ converges to zero, and end up deriving an upper one-sided limit chart which only requires the use of an overall mean and time-varying control limits in a plot with a similar interpretation and operational aspects to those of EWMA and Shewhart charts. This limit chart also gives good protection against upward shifts in the process mean and outperforms several other charts with regard to the ARL and is still competitive when it comes to other performance criteria.

Keywords. exponentially weighted moving average charts, run length.

1 EWMA charts with exact control limits

The exponentially weighted moving average chart (Roberts, 1959) for monitoring shifts in the process mean μ of i.i.d. 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 μ_0; X_t is an estimator of μ, usually the sample mean at time t; and $\lambda \in (0, 1]$ is a smoothing parameter that corresponds to the weight given to the most recent sample. Moreover, we assume that

$$X_t \sim_{i.i.d.} \mathcal{N}(\mu = \mu_0 + a\,\sigma, \sigma^2),\ t = 1, 2, \ldots. \tag{2}$$

If $a > 0$ then a sustained upward shift in the process mean μ has occurred at time $t = 1^-$; needless to say, $\{X_t, t = 1, 2, \ldots\}$ is said to be in-control if $a = 0$, and to be out-of-control otherwise.

Upward shifts can be detected by upper one-sided EWMA charts with exact control limits that give a signal at the sampling period t if

$$Z_t > E_{a=0}(Z_t) + c\sqrt{\text{Var}_{a=0}(Z_t)} \tag{3}$$

where: $E_a(Z_t) = \mu_0 + a\,\sigma\left[1 - (1-\lambda)^t\right]$; $\text{Var}_a(Z_t) = \text{Var}_{a=0}(Z_t) = \frac{\lambda}{2-\lambda}\,\sigma^2\left[1 - (1-\lambda)^{2t}\right]$; and $c > 0$ is a pre-specified critical value that defines the range of the (time-varying) exact control limits.

We rely on the run length (RL),

$$N_a = \inf\{t : Z_t > E_0(Z_t) + c\sqrt{\text{Var}_0(Z_t)}\}, \tag{4}$$

to assess the performance of 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 λ and c, therefore it is going to be represented by $N_a(\lambda, c)$. According to Morais et al. (2008), its survival function can be written in terms of the c.d.f. of multivariate normal random vectors:

$$P[N_a(\lambda, c) > k] = F_{\mathcal{N}_k(\mathbf{0}_k, \mathbf{C}_k(\lambda))}(c - a \times g(\lambda, i),\ i = 1, \ldots, k),\ k = 1, 2, \ldots, \tag{5}$$

where: $\mathbf{0}_k$ is a column vector with k zeroes; $\mathbf{C}_k(\lambda)$ is a correlation matrix with entries

$$\text{Corr}_0(Z_i, Z_j) = (1-\lambda)^{j-i}\sqrt{[1-(1-\lambda)^{2i}]/[1-(1-\lambda)^{2j}]},\ 1 \leq i \leq j \leq k; \tag{6}$$

and

$$g(\lambda, i) = \sqrt{(2-\lambda) \times [1-(1-\lambda)^i]/\{\lambda \times [1+(1-\lambda)^i]\}}. \tag{7}$$

Unsurprisingly, problems arise when we try to obtain the ARL of this chart,

$$E[N_a(\lambda, c)] = 1 + \sum_{k=1}^{+\infty} P[N_a(\lambda, c) > k], \tag{8}$$

and have to depend on simulation to approximate this quantity.

2 On the stochastic monotone behaviour of the in-control run length

Even though we have to rely on simulation to obtain $E[N_a(\lambda, c)]$, 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 $\mathcal{F}(c)$.

If we assume that $X_t \sim_{i.i.d.} \mathcal{N}(\mu = \mu_0, \sigma^2)$, $t = 1, 2, \ldots$, then

$$P[N_0(\lambda_1, c) > k] \geq P[N_0(\lambda_2, c) > k],\ k = 1, 2, \ldots,\ 0 < \lambda_1 \leq \lambda_2 \leq 1, \tag{9}$$

i.e., $N_0(\lambda, c)$ stochastically decreases with λ in the usual sense. (See Chapter 1 of Shaked and Shanthikumar (1994) for this and other notions of stochastic ordering.)

The stochastic ordering result (9), proved by Morais et al. (2008), leads to several conclusions, namely that the in-control RL of the upper one-sided Shewhart control chart, $N_0(\lambda = 1, c)$, is the smallest in-control RL within $\mathcal{F}(c)$, stochastically speaking.

3 Choosing the critical value. Out-of-control run length

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 $\xi > 1$, that is, c is the solution of the equation $E[N_0(\lambda, c)] = \xi$. Thus, the critical value is a function of λ and ξ, say $c(\lambda, \xi)$.

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 λ.

By capitalizing on the previous stochastic ordering result, Morais et al. (2008) managed to prove that if we assume once again that the $X_t \sim_{i.i.d.} \mathcal{N}(\mu = \mu_0, \sigma^2)$, $t = 1, 2, \ldots$, and consider a fixed $\xi > 1$, then the critical value $c(\lambda, \xi)$ is an increasing function in $\lambda \in (0, 1]$.

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

By relying on the *regula falsi* method to find the root $c(\lambda, \xi)$, setting $\xi = 500$ samples, simulating the in-control process from a normal distribution with zero mean and unit variance and considering the number of replications equal to 10^7, we got the critical values $c(\lambda, \xi = 500)$, for different values of the smoothing parameter $\lambda = 10^{-i}$, $i = 0, 1, \ldots, 4$, and also the out-control ARL values for all these values of λ and several magnitude 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 ARL and $\xi = 500$ smaller than 0.2%. Moreover, any simulated RL is truncated if it is larger than 50000 samples. 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 $c(1, \xi) = \Phi^{-1}(1 - 1/\xi)$.

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 λ.

		a							
λ	$c(\lambda, 500)$	0	0.25	0.5	1.0	1.5	2.0	3.0	4.0
0.0001	0.346840	501.147487	5.433230	2.624336	1.459113	1.159824	1.054080	1.004007	1.000131
0.001	0.829928	499.611751	11.231302	4.428709	1.965261	1.364700	1.141339	1.015279	1.000776
0.01	1.654164	500.517635	30.814533	10.546469	3.651299	2.097137	1.502536	1.093122	1.009534
0.1	2.543225	499.745389	66.944150	21.634646	6.760731	3.539827	2.303960	1.367106	1.073346
1	2.878162	500.000000	232.970748	114.947864	33.135052	11.893905	5.265154	1.823199	1.150702

The simulation results in Table 1, not only illustrates the increasing behaviour of $c(\lambda, \xi)$ in λ, but also allows us to state that the upper one-sided Shewhart chart has the largest out-of-control ARL within the class of matched in-control EWMA charts.

These results, the additional ones in Morais et al. (2008), Figure 1 of Frisén and Sonesson (2006) and Figure 1 below suggest that λ should approach to zero to minimize the out-of-control ARL (for a fixed in-control ARL), regardless of the magnitude of the shift.

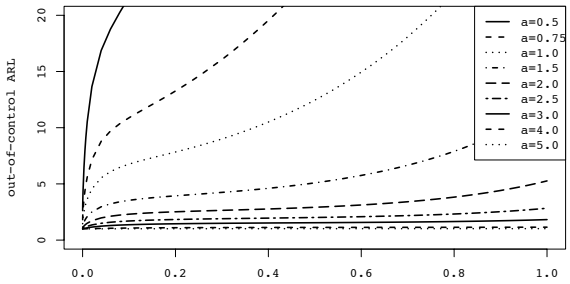

Fig. 1. Out-of-control ARL of matched in-control upper one-sided EWMA charts with exact control limits as a function of λ.

This by no means contradicts a well-known fact concerning EWMA charts with *asymptotic* control limits: Shewhart charts are faster than their matched in-control EWMA counterparts with asymptotic control limits when it comes to the detection of large shifts in the process mean (Lucas and Saccucci, 1990).

In the next section we present the upper one-sided limit chart, derived from the limiting distribution of $N_a(\lambda, c)$ as λ tends to 0^+. To obtain this limiting distribution it suffices to evaluate

$$\mathbf{C}_k(0) = \lim_{\lambda \to 0^+} \mathbf{C}_k(\lambda) = \left[\sqrt{\frac{\min\{i,j\}}{\max\{i,j\}}} \right]_{i,j=1,\ldots,k} \tag{10}$$

and $\lim_{\lambda \to 0^+} g(\lambda, i) = \sqrt{i}$.

4 The limit chart and its performance

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,

$$N_a(0, c) = \inf\{t \in I\!N : \bar{X}_t > E_0(\bar{X}_t) + c\sqrt{\text{Var}_0(\bar{X}_t)} = \mu_0 + c\sigma/\sqrt{t}\}, \tag{11}$$

has survival function defined by

$$\lim_{\lambda \to 0^+} P[N_a(\lambda, c) > k] = F_{\mathcal{N}_k(\mathbf{0}_k, \mathbf{C}_k(0))}\left(c - a\sqrt{i},\, i = 1, ..., k\right),\ k = 1, 2, \ldots. \tag{12}$$

Interestingly enough, this chart can be related to other ones.

1. Since $\bar{X}_t = (1 - \frac{1}{t})\bar{X}_{t-1} + \frac{1}{t}X_t$, the upper one-sided 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 $\mu_0 + c\sigma/\sqrt{t}$.

2. Using $\frac{1}{\sqrt{t}}\sum_{i=1}^{t}(X_i-\mu_0)$ instead of $\bar{X}_t$, leads to a sort of an upper one-sided CUSUM chart with a conveniently constant UCL, $c\sigma$.
3. Moreover, the upper one-sided limit chart is simpler than the generalized EWMA scheme proposed by Han and Tsung (2004) which uses a dimension-varying control statistic at time t and requires running a family of t EWMA charts at time point t, with smoothing parameters from the set $\{1, 1/2, 1/3, \ldots, 1/t\}$.

Now, it is time to assess how the (A) upper one-sided limit chart with $c(0, 500) = 0.164547$ compares to the following competing charts:

(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 $c = 2.532760$;
(D) upper one-sided Shewhart chart, $c(1, 500) = 2.878144$;
(E) upper one-sided version of the generalized EWMA scheme (Han and Tsung, 2004), $c^{(HT)} = 3.10$.

Table 2. 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).

	a							
Chart	0	0.25	0.5	1.0	1.5	2.0	3.0	4.0
A	499.520451	4.137894	2.188909	1.335961	1.112503	1.035934	1.002312	1.000064
B	499.745389	66.944150	21.634646	6.760731	3.539827	2.303960	1.367106	1.073346
C	500.289922	70.360046	24.726256	8.907849	5.389757	3.915183	2.604415	2.057725
D	500.000000	232.970748	114.947864	33.135052	11.893905	5.265154	1.823199	1.150702
E	499.3750	86.2551	29.5563	9.4000	4.8581	3.1067	1.7067	1.1909

From the results in Table 2 it is apparent that the upper one-sided limit chart outperforms all four other matched in-control charts.

This table certainly supports the well known fact that Shewhart charts are faster than the EWMA charts with asymptotic limits in the detection of very large shifts. However, when we compare their 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).

5 Further investigations

5.1 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

$$P_0[\bar{X}_1 = X_1 > \mu_0 + c(0,\xi)\,\sigma] = 1 - \Phi[c(0,\xi)]; \tag{13}$$

and since the critical value $c(0, \xi)$ is a small positive constant this probability is large, and the number of false alarms at sample 1 is 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

$$\bar{X}_t^- = \frac{HS^-}{t} + \frac{1}{t}\sum_{i=1}^{t} X_i. \tag{14}$$

HS^- should chosen in order to achieve a reasonably small value for the probability in (13). 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

$$HS^- = HS^-(1/\xi) = c(0,\xi) - \Phi^{-1}(1 - 1/\xi). \tag{15}$$

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 3 for the upper one-sided limit charts with critical value $c(0, \xi = 500) = 0.164547$ and

(A1) $HS^{-}(1/\xi) = c(0, \xi) - \Phi^{-1}(1 - 1/\xi) \simeq -2.7$
(A2) $HS^{-}(2/\xi) = c(0, \xi) - \Phi^{-1}(1 - 2/\xi) \simeq -2.5$.

Table 3. 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.

		a					
Chart	% of ones	0	0.05	0.1	0.15	0.2	0.25
A	0.434671	500.670590	22.616663	10.637120	6.920705	5.172090	4.133596
A1	0.002036	1933.249760	89.002448	41.717421	26.990211	19.911457	15.719583
A2	0.003826	1831.702782	83.899498	39.431086	25.524654	18.763227	14.861847
B	0.005457	500.303351	304.815952	195.265144	130.820231	91.864101	66.963427
C	0.000000	500.289922	307.293711	198.801400	134.479182	95.442338	70.429621
D	0.002000	500.000000	427.203031	365.848755	314.029065	270.170007	232.970748
E	0.000968	495.640	347.150	239.748	161.997	117.394	86.708

However, the upper one-sided limit charts with these two head starts (A1–A2) can still outperform the competing ones (B–F) in the detection of very small shifts (e.g. $a = 0.05, 0.1, 0.15, 0.2$), as shown by Table 3. 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 3.

5.2 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 that

$$X_t \sim_{i.i.d.} \begin{cases} \mathcal{N}(\mu = \mu_0, \sigma^2), & t < q \\ \mathcal{N}(\mu = \mu_0 + a\,\sigma, \sigma^2), & t \geq q, \end{cases} \tag{16}$$

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

Under (16) the ARL is no longer an adequate chart performance. If fact, the essential tools in the comparison of the performances of several competing charts under the distributional assumption (16) are the conditional average delay

$$CAD_a(q, \lambda, c) = E[N_a(\lambda, c) - q + 1 | N_a(\lambda, c) \geq q], \tag{17}$$

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,

$$MCAD_a(\lambda, c) = \max_{q=1,2,\ldots} CAD_a(q, \lambda, c). \tag{18}$$

We summarize some MCAD values in Table 4. These values refer to the charts A–D defined previously. (Scheme E was not included in this last simulation study because it was computationally unfeasible.)

It is apparent from Table 4 that the MCAD performance criterion favours the upper one-sided limit chart for small shifts, the upper one-sided CUSUM chart for moderate shifts and the upper one-sided Shewhart chart for large shifts.

Adding to this, charts A, A1 and A2 have very similar out-of-control performances regarding MCAD, suggesting that the adoption of the negative head starts $HS^{-}(1/\xi)$ and $HS^{-}(2/\xi)$ has a negligible effect on the out-control situation of the upper one-sided limit chart.

Table 4. Maximum conditional average delays of the upper one-sided limit chart (A) and other competing charts (A1, A2, B–D).

	a						
Chart	0.25	0.5	1.0	1.5	2.0	3.0	4.0
A	53.73451	26.74354	13.41873	9.06882	6.86498	6.70867	3.64195
A1	54.884611	27.201756	13.680594	9.218782	7.000753	4.802976	3.717786
A2	54.819058	27.129930	13.663792	9.194612	6.983917	6.983917	3.708094
B	70.52657	24.17937	8.90591	5.40826	3.94658	2.64587	2.05344
C	69.91940	24.15317	8.85420	5.39099	3.93217	2.63550	2.04735
D	232.97075	114.94786	33.13505	11.89391	5.26515	1.82320	1.15070

6 Conclusions

This paper essentially provides a study on the behaviour of the upper one-sided EWMA chart with exact control limits when the smoothing parameter λ converges to zero. Nevertheless, we ought to note that the resulting chart — the upper one-sided limit chart — combines two properties:

Simplicity — It requires only the use of an overall mean and time-varying control limits in a plot with a similar interpretation and operational aspects to those of EWMA and Shewhart charts.

Efficiency — It gives good protection against upward shifts in the process mean and outperforms several other charts with regard to the ARL and is still competitive when it comes to other performance criteria like the maximum conditional average delay.

On the downside, the upper one-sided limit chart triggers false alarms quite frequently in the first samples. However, the use of negative head starts dramatically improves the in-control behaviour and still provides good out-control ARL performance.

References

Frisén, M. and Sonesson, C. (2006). Optimal surveillance based on exponentially weighted moving averages. *Sequential Analysis*, **25**, 379–403.

Han, D. and Tsung, F. (2004). A generalized EWMA control chart and its comparison with the optimal EWMA, CUSUM, and GLR schemes. *Annals of Statistics*, **32**, 316–339.

Knoth, S. (2003). EWMA schemes with nonhomogenous transition kernels. *Sequential Analysis*, **22**, 241–255.

Lucas, J.M. and Saccucci, M.S. (1990). Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics*, **32**, 1–12.

Morais, M.C., Okhrin, Y. and Schmid, W. (2008). On the limiting behaviour of EWMA control charts with exact control limits. Submitted for publication.

Roberts, S.W. (1959). Control charts tests based on geometric moving averages. *Technometrics*, **1**, 239–250.

Shaked, M. and Shanthikumar, J.G. (1994). *Stochastic Orders and Their Applications*. Academic Press, Inc.