Statistical Process Control

Sven Knoth[1]

Statistical Process Control (SPC) is the misleading title of the area of statistics which is concerned with the statistical monitoring of sequentially observed data. Together with the theory of sampling plans, capability analysis and similar topics it forms the field of Statistical Quality Control. SPC started in the 1930s with the pioneering work of Shewhart (1931). Then, SPC became very popular with the introduction of new quality policies in the industries of Japan and of the USA. Nowadays, SPC methods are considered not only in industrial statistics. In finance, medicine, environmental statistics, and in other fields of applications practitioners and statisticians use and investigate SPC methods.

A SPC scheme – in industry mostly called control chart – is a sequential scheme for detecting the so called change point in the sequence of observed data. Here, we consider the most simple case. All observations $X_1, X_2, \dots$ are independent, normally distributed with known variance σ^2. Up to an unknown time point $m-1$ the expectation of the X_i is equal to μ_0, starting with the change point m the expectation is switched to $\mu_1 \neq \mu_0$. While both expectation values are known, the change point m is unknown. Now, based on the sequentially observed data the SPC scheme has to detect whether a change occurred.

SPC schemes can be described by a stopping time L – known as run length – which is adapted to the sequence of sigma algebras $\mathcal{F}_n = \mathcal{F}(X_1, X_2, \dots, X_n)$. The performance or power of these schemes is usually measured by the Average Run Length (ARL), the expectation of L. The ARL denotes the average number of observations until the SPC scheme signals. We distinguish false alarms – the scheme signals before m, i. e. before the change actually took place – and right ones. A suitable scheme provides large ARLs for $m = \infty$ and small ARLs for $m = 1$. In case of $1 < m < \infty$ one has to consider further performance measures. In the case of the oldest schemes – the Shewhart charts – the typical inference characteristics like the error probabilities were firstly used.

The chapter is organized as follows. In Section 1 the charts in consideration are introduced and their graphical representation is demonstrated. In the Section 2 the most popular chart characteristics are described. First, the characteristics as the ARL and the Average Delay (AD) are defined. These performance measures are used for the setup of the applied SPC scheme. Then, the three subsections of Section 2 are concerned with the usage of the SPC routines for determination of the ARL, the AD, and the probability mass function (PMF) of the run length. In Section 3 some results of two papers are reproduced with the corresponding XploRe quantlets.

1 Control Charts

Recall that the data $X_1, X_2, \dots$ follow the change point model

$$\begin{cases} X_t \sim N(\mu_0, \sigma^2) & , t = 1, 2, \dots, m-1 \\ X_t \sim N(\mu_1 \neq \mu_0, \sigma^2) & , t = m, m+1, \dots \end{cases} . \tag{1}$$

[1]European University Viadrina Frankfurt (Oder)

The observations are independent and the time point m is unknown. The control chart (the SPC scheme) corresponds to a stopping time L. Here we consider three different schemes – the Shewhart chart, EWMA and CUSUM schemes. There are one- and two-sided versions. The related stopping times in the one-sided upper versions are:

1. The Shewhart chart introduced by Shewhart (1931)

$$L^{\text{Shewhart}} = \inf\left\{ t \in I\!N : Z_t = \frac{X_t - \mu_0}{\sigma} > c_1 \right\} \tag{2}$$

   with the design parameter c_1 called critical value.

2. The EWMA scheme (exponentially weighted moving average) initially presented by Roberts (1959)

$$\begin{aligned} L^{\text{EWMA}} &= \inf\left\{ t \in I\!N : Z_t^{\text{EWMA}} > c_2 \sqrt{\lambda/(2-\lambda)} \right\} , & (3) \\ Z_0^{\text{EWMA}} &= z_0 = 0 , & \\ Z_t^{\text{EWMA}} &= (1-\lambda)\, Z_{t-1}^{\text{EWMA}} + \lambda\, \frac{X_t - \mu_0}{\sigma} , \; t = 1, 2, \ldots & (4) \end{aligned}$$

   with the smoothing value λ and the critical value c_2. The smaller λ the faster EWMA detects small $\mu_1 - \mu_0 > 0$.

3. The CUSUM scheme (cumulative sum) introduced by Page (1954)

$$\begin{aligned} L^{\text{CUSUM}} &= \inf\left\{ t \in I\!N : Z_t^{\text{CUSUM}} > c_3 \right\} , & (5) \\ Z_0^{\text{CUSUM}} &= z_0 = 0 , & \\ Z_t^{\text{CUSUM}} &= \max\left\{ 0, Z_{t-1}^{\text{CUSUM}} + \frac{X_t - \mu_0}{\sigma} - k \right\} , \; t = 1, 2, \ldots & (6) \end{aligned}$$

   with the reference value k and the critical value c_3 (known as decision interval). For fastest detection of $\mu_1 - \mu_0$ CUSUM has to be set up with $k = (\mu_1 + \mu_0)/(2\,\sigma)$.

The above notation uses normalized data. Thus, it is not important whether X_t is a single observation or a sample statistic as the empirical mean.

Remark, that for using one-sided lower schemes one has to apply the upper schemes to the data multiplied with -1. A slight modification of one-sided Shewhart and EWMA charts leads to their two-sided versions. One has to replace in the comparison of chart statistic and threshold the original statistic Z_t and Z_t^{EWMA} by their absolute value. The two-sided versions of these schemes are more popular than the one-sided ones. For two-sided CUSUM schemes we consider a combination of two one-sided schemes, Lucas (1976) or Lucas and Crosier (1982), and a scheme based on Crosier (1986). Note, that in some recent papers the same concept of combination of two one-sided schemes is used for EWMA charts.

Recall, that Shewhart charts are a special case of EWMA schemes ($\lambda = 1$). Therefore, we distinguish 5 SPC schemes – `ewma1`, `ewma2`, `cusum1`, `cusum2` (two one-sided schemes), and `cusumC` (Crosier's scheme). For the two-sided EWMA charts the following quantlets are provided in the XploRe quantlib `spc`.

SPC quantlets for two-sided EWMA scheme	
spcewma2	– produces chart figure
spcewma2arl	– returns ARL
spcewma2c	– returns critical value c_2
spcewma2ad	– returns AD (steady-state ARL)
spcewma2pmf	– returns probability mass and distribution function of the run-length for single time points
spcewma2pmfm	– the same up to a given time point

By replacing `ewma2` by one of the remaining four scheme titles the related characteristics can be computed.

The quantlets `spcewma1`,...,`spccusumC` generate the chart figure. Here, we apply the 5 charts to artificial data. 100 pseudo random values from a normal distribution are generated. The first 80 values have expectation 0, the next 20 values have expectation 1, i. e. model (1) with $\mu_0 = 0$, $\mu_1 = 1$, and $m = 81$. We start with the two-sided EWMA scheme and set $\lambda = 0.1$, i. e. the chart is

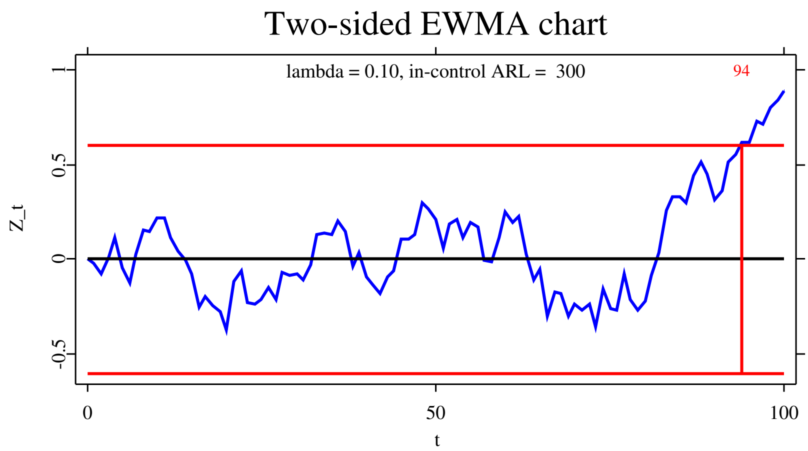


Figure 1: Two-sided EWMA chart   `XFGewma2fig.xpl`

very sensitive to small changes. The critical value c_2 (see (3)) is computed to provide an in-control ARL of 300 (see Section 2). Thus, the scheme leads in average after 300 observations to a false alarm.

In Figure 1 the graph of Z_t^{EWMA} is plotted against time $t = 1, 2, \ldots, 100$. Further, the design parameter λ, the in-control ARL, and the time of alarm (if there is one) are printed. One can see, that the above EWMA scheme detects the change point $m = 81$ at time point 94, i. e. the delay is equal to 14. The related average values, i. e. ARL and Average Delay (AD), for $\mu_1 = 1$ are 9.33 and 9.13, respectively. Thus, the scheme needs here about 5 observations more than average.

In the same way the remaining four SPC schemes can be plotted. Remark, that in case of `ewma1` one further parameter has to be set. In order to obtain a suitable figure and an appropriate scheme the EWMA statistic Z_t^{EWMA} (see

(4)) is reflected at a pre-specified border `zreflect` $\leq 0\,(= \mu_0)$, i. e.

$$Z_t^{\text{EWMA}} = \max\{\texttt{zreflect}, Z_t^{\text{EWMA}}\} \quad , \; t = 1, 2, \ldots$$

for an upper EWMA scheme. Otherwise, the statistic is unbounded, which leads to schemes with poor worst case performance. Further, the methods used

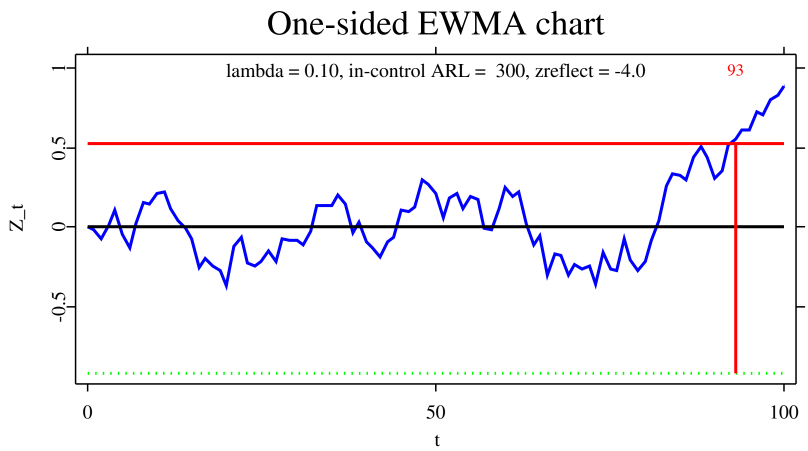


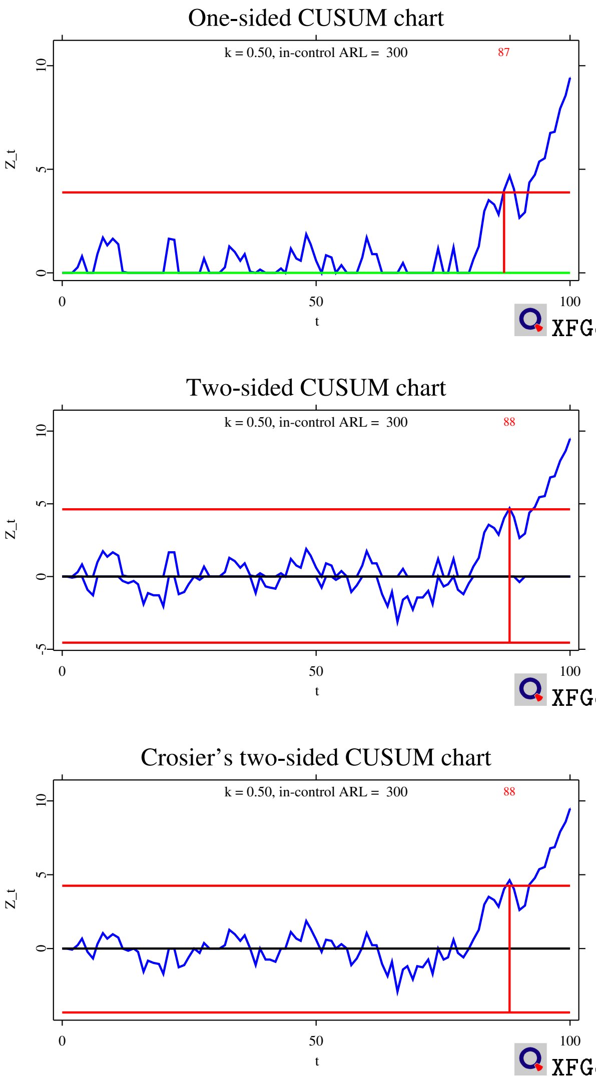
Figure 2: One-sided EWMA chart `XFGewma1fig.xpl`

in Section 2 for computing the chart characteristics use bounded continuation regions of the chart. If `zreflect` is small enough, then the ARL and the AD (which are not worst case criterions) of the reflected scheme are the same as of the unbounded scheme. Applying the quantlet `XFGewma1fig.xpl` with `zreflect`$= -4$ leads to Figure 2. Thereby, `zreflect` has the same normalization factor $\sqrt{\lambda/(2-\lambda)}$ like the critical value c_2 (see 2.). The corresponding normalized border is printed as dotted line (see Figure 2). The chart signals one observation earlier than the two-sided version in Figure 1. The related ARL and AD values for $\mu_1 = 1$ are now 7.88 and 7.87, respectively.

In Figure 3 the three different CUSUM charts with $k = 0.5$ are presented. They signal at the time points 87, 88, and 88 for `cusum1`, `cusum2`, and `cusumC`, respectively. For the considered dataset the CUSUM charts are faster because of their better worst case performance. The observations right before the change point at $m = 81$ are smaller than average. Therefore, the EWMA charts need more time to react to the increased average. The related average values of the run length, i. e. ARL and AD, are 8.17 and 7.52, 9.52 and 8.82, 9.03 and 8.79 for `cusum1`, `cusum2`, and `cusumC`, respectively.

2 Chart characteristics

Consider the change point model (1). For fixed m denote $P_m(\cdot)$ and $E_m(\cdot)$ the corresponding probability measure and expectation, respectively. Hereby,



XFGcusum1fig.xpl

XFGcusum2fig.xpl

XFGcusumCfig.xpl

Figure 3: CUSUM charts: one-sided, two-sided, Crosier's two-sided

$m = \infty$ stands for the case of no change, i. e. the so called in-control case. Then the Average Run Length (ARL) (expectation of the run length L) is defined as

$$\mathcal{L}_\mu = \begin{cases} E_\infty(L) & , \ \mu = \mu_0 \\ E_1(L) & , \ \mu \neq \mu_0 \end{cases} . \tag{7}$$

Thus, the ARL denotes the average number of observations until signal for a sequence with constant expectation. $\mu = \mu_0$ or $m = \infty$ stands for no change, $\mu \neq \mu_0$ and $m = 1$ mark, that just at the first time point (or earlier) a change takes place from μ_0 to μ. Therefore, the ARL evaluates only the special scenario of $m = 1$ of the SPC scheme. Other measures, which take into account that usually $1 < m < \infty$, were introduced by Lorden (1971) and Pollak and Siegmund (1975), Pollak and Siegmund (1975). Here, we use a performance measure which was firstly proposed by Roberts (1959). The so called (conditional) Average Delay (AD, also known as steady-state ARL) is defined as

$$\begin{aligned} \mathcal{D}_\mu &= \lim_{m\to\infty} \mathcal{D}_\mu^{(m)} , \\ \mathcal{D}_\mu^{(m)} &= E_m\big(L - m + 1 | L \geq m\big) , \end{aligned} \tag{8}$$

where μ is the value of μ_1 in (1), i.e. the expectation after the change. While $\mathcal{L}_\mu$ measures the delay for the case $m = 1$, $\mathcal{D}_\mu$ determines the delay for a SPC scheme which ran a long time without signal. Usually, the convergence in (8) is very fast. For quite small m the difference between $\mathcal{D}_\mu^{(m)}$ and $\mathcal{D}_\mu$ is very small already. $\mathcal{L}_\mu$ and $\mathcal{D}_\mu$ are average values for the random variable L. Unfortunately, L is characterized by a large standard deviation. Therefore, one might be interested in the whole distribution of L. Again, we restrict on the special cases $m = 1$ and $m = \infty$. We consider the probability mass function $P_\mu(L = n)$ (PMF) and the cumulative distribution function $P_\mu(L \leq n)$ (CDF). Based on the CDF, one is able to compute quantiles of the run length L.

For normally distributed random variables it is not possible to derive exact solutions for the above characteristics. There are a couple of approximation techniques. Besides very rough approximations based on the Wald approximation known from sequential analysis, Wiener process approximations and similar methods, three main methods can be distinguished:

1. Markov chain approach due to Brook and Evans (1972): Replacement of the continuous statistic Z_t by a discrete one
2. Quadrature of integral equations which are derived for the ARL, Vance (1986) and Crowder (1986) and for some eigenfunctions which lead to the AD
3. Waldmann (1986) approach: Iterative computation of $P(L = n)$ by using quadrature and exploiting of monotone bounds for the considered characteristics

Here we use the first approach, which has the advantage, that all considered characteristics can be presented in a straightforward way. Next, the Markov chain approach is briefly described. Roughly speaking, the continuous statistic Z_t is approximated by a discrete Markov chain M_t. The transition $Z_{t-1} = x \to Z_t = y$ is approximated by the transition $M_{t-1} = i\,w \to M_t = j\,w$ with $x \in [i\,w - w/2, i\,w + w/2]$ and $y \in [j\,w - w/2, j\,w + w/2]$. That is, given an integer r the continuation region of the scheme $[-c, c]$, $[\texttt{zreflect}, c]$, or $[0, c]$ is separated into $2\,r + 1$ or $r + 1$ intervals of the kind $[i\,w - w/2, i\,w + w/2]$ (one exception is $[0, w/2]$ as the first subinterval of $[0, c]$). Then, the transition kernel f of Z_t is approximated by the discrete kernel of M_t, i.e.

$$f(x, y) \approx P(i\,w \to [j\,w - w/2, j\,w + w/2])/w$$

for all $x \in [i\,w - w/2, i\,w + w/2]$ and $y \in [j\,w - w/2, j\,w + w/2]$. Eventually, we obtain a Markov chain $\{M_t\}$ with $2\,r + 1$ or $r + 1$ transient states and one absorbing state. The last one corresponds to the alarm (signal) of the scheme.

Denote by $Q = (q_{ij})$ the matrix of transition probabilities of the Markov chain $\{M_t\}$ on the transient states, $\underline{1}$ a vector of ones, and $\underline{L} = (L_i)$ the ARL vector. L_i stands for the ARL of a SPC scheme which starts in point $i\,w$ (corresponds to z_0). In the case of a one-sided CUSUM scheme with $z_0 = 0 \ni [0, w/2]$ the value L_0 approximates the original ARL. By using $\underline{L}$ we generalize the original schemes to schemes with possibly different starting values z_0. Now, the following linear equation system is valid, Brook and Evans (1972):

$$(I - Q)\,\underline{L} = \underline{1}\,, \tag{9}$$

where I denotes the identity matrix. By solving this equation system we get the ARL vector $\underline{L}$ and an approximation of the ARL of the considered SPC scheme. Remark that the larger r the better is the approximation. In the days of Brook and Evans (1972) the maximal matrix dimension $r + 1$ (they considered `cusum1`) was 15 because of the restrictions of the available computing facilities. Nowadays, one can use dimensions larger than some hundreds. By looking at different r one can find a suitable value. The quantlet `XFGrarl.xpl` demonstrates this effect for the Brook and Evans (1972) example. 9 different values of r from 5 to 500 are used to approximate the in-control ARL of a one-sided CUSUM chart with $k = 0.5$ and $c_3 = 3$ (variance $\sigma^2 = 1$). We get

r	5	10	20	30	40	50	100	200	500
$\mathcal{L}_0$	113.47	116.63	117.36	117.49	117.54	117.56	117.59	117.59	117.60

`XFGrarl.xpl`

The true value is 117.59570 (obtainable via a very large r or by using the quadrature methods with a suitable large number of abscissas). The computation of the average delay (AD) requires more extensive calculations. For details see, e. g., Knoth (1998) on CUSUM for Erlang distributed data. Here we apply the Markov chain approach again, Crosier (1986). Given one of the considered schemes and normally distributed data, the matrix Q is primitive, i. e. there exists a power of Q which is positive. Then Q has one single eigenvalue which is larger in magnitude than the remaining eigenvalues. Denote this eigenvalue by ϱ. The corresponding left eigenvector $\underline{\psi}$ is strictly positive, i. e.

$$\underline{\psi}\,Q = \varrho\,\underline{\psi}\,,\ \underline{\psi} > 0\,. \tag{10}$$

It can be shown, Knoth (1998), that the conditional density $f(\cdot|L \geq m)$ of both the continuous statistic Z_t and the Markov chain M_t tends for $m \to \infty$ to the normalized left eigenfunction and eigenvector, respectively, which correspond to the dominant eigenvalue ϱ. Therefore, the approximation of $\mathcal{D} = \lim_{m\to\infty} E_m(L - m + 1|L \geq m)$ can be constructed by

$$D = (\underline{\psi}^T \underline{L})/(\underline{\psi}^T \underline{1})\,.$$

Note, that the left eigenvector $\underline{\psi}$ is computed for the in-control mean μ_0, while the ARL vector $\underline{L}$ is computed for a specific out-of-control mean or μ_0 again.

If we replace in the above quantlet (`XFGrarl.xpl`) the phrase `arl` by `ad`, then we obtain the following output which demonstrates the effect of the parameter r again.

r	5	10	20	30	40	50	100	200	500
$\mathcal{D}_0$	110.87	114.00	114.72	114.85	114.90	114.92	114.94	114.95	114.95

XFGrad.xpl

Fortunately, for smaller values of r than in the ARL case we get good accuracy already. Note, that in case of cusum2 the value r has to be smaller (less than 30) than for the other charts, since it is based on the computation of the dominant eigenvalue of a very large matrix. The approximation in case of combination of two one-sided schemes needs a twodimensional approximating Markov chain. For the ARL only exists a more suitable approach. As, e. g., Lucas and Crosier (1982) shown it is possible to use the following relation between the ARLs of the one- and the two-sided schemes. Here, the two-sided scheme is a combination of two symmetric one-sided schemes which both start at $z_0 = 0$. Therefore, we get a very simple formula for the ARL $\mathcal{L}$ of the two-sided scheme and the ARLs $\mathcal{L}_{upper}$ and $\mathcal{L}_{lower}$ of the upper and lower one-sided CUSUM scheme

$$\mathcal{L} = \frac{\mathcal{L}_{upper} \cdot \mathcal{L}_{lower}}{\mathcal{L}_{upper} + \mathcal{L}_{lower}} . \tag{11}$$

Eventually, we consider the distribution function of the run length L itself. By using the Markov chain approach and denoting with p_i^n the approximated probability of $(L > n)$ for a SPC scheme started in $i\,w$, such that $\underline{p}^n = (p_i^n)$, we obtain

$$\underline{p}^n = \underline{p}^{n-1}\, Q = \underline{p}^0\, Q^n . \tag{12}$$

The vector $\underline{p}^0$ is initialized with $p_i^0 = 1$ for the starting point $z_0 \in [i\,w - w/2, i\,w + w/2]$ and $p_j^0 = 0$ otherwise. For large n we can replace the above equation by

$$p_i^n \approx g_i\, \varrho^n . \tag{13}$$

The constant g_i is defined as

$$g_i = \phi_i / (\underline{\phi}^T \underline{\psi}) ,$$

where $\underline{\phi}$ denotes the right eigenvector of Q, i. e. $Q\,\underline{\phi} = \varrho\,\underline{\phi}$. Based on (12) and (13) the probability mass and the cumulative distribution function of the run length L can be approximated. (12) is used up to a certain n. If the difference between (12) and (13) is smaller than 10^{-9}, then exclusively (13) is exploited. Remark, that the same is valid as for the AD. For the two-sided CUSUM scheme (cusum2) the parameter r has to be small (≤ 30).

2.1 Average Run Length and Critical Values

The spc quantlib provides the quantlets spcewma1arl,...,spccusumCarl for computing the ARL of the corresponding SPC scheme. All routines need the actual value of μ as a scalar or as a vector of several μ, two scheme parameters, and the integer r (see the beginning of the section). The XploRe example XFGarl.xpl demonstrates all ...arl routines for $k = 0.5$, $\lambda = 0.1$, zreflect$= -4$, $r = 50$, $c = 3$, in-control and out-of-control means $\mu_0 = 0$ and $\mu_1 = 1$, respectively. The next table summarizes the ARL results

chart	ewma1	ewma2	cusum1	cusum2	cusumC
$\mathcal{L}_0$	1694.0	838.30	117.56	58.780	76.748
$\mathcal{L}_1$	11.386	11.386	6.4044	6.4036	6.4716

XFGarl.xpl

Remember that the ARL of the two-sided CUSUM (cusum2) scheme is based on the one-sided one, i. e. $58.78 = 117.56/2$ and $6.4036 = (6.4044 \cdot 49716)/(6.4044 + 49716)$ with $49716 = \mathcal{L}_{-1}$.

For the setup of the SPC scheme it is usual to give the design parameter λ and k for EWMA and CUSUM, respectively, and a value ξ for the in-control ARL. Then, the critical value c (c_2 or c_3) is the solution of the equation $\mathcal{L}_{\mu_0}(c) = \xi$. Here, the regula falsi is used with an accuracy of $|\mathcal{L}_{\mu_0}(c) - \xi| < 0.001$. The quantlet XFGc.xpl demonstrates the computation of the critical values for SPC schemes with in-control ARLs of $\xi = 300$, reference value $k = 0.5$ (CUSUM), smoothing parameter $\lambda = 0.1$ (EWMA), `zreflect`$= -4$, and the Markov chain parameter $r = 50$.

chart	ewma1	ewma2	cusum1	cusum2	cusumC
c	2.3081	2.6203	3.8929	4.5695	4.288

XFGc.xpl

The parameter $r = 50$ guarantees fast computation and suitable accuracy. Depending on the power of the computer one can try values of r up to 1000 or larger (see XFGrarl.xpl in the beginning of the section).

2.2 Average Delay

The usage of the routines for computing the Average Delay (AD) is similar to the ARL routines. Replace only the code `arl` by `ad`. Be aware that the computing time is larger than in case of the ARL, because of the computation of the dominant eigenvalue. It would be better to choose smaller r, especially in the case of the two-sided CUSUM. Unfortunately, there is no relation between the one- and two-sided schemes as for the ARL in (11). Therefore, the library computes the AD for the two-sided CUSUM based on a twodimensional Markov chain with dimension $(r+1)^2 \times (r+1)^2$. Thus with values of r larger than 30, the computing time becomes quite large. Here the results follow for the above quantlet XFGrarl.xpl with `ad` instead of `arl` and $r = 30$ for `spccusum2ad`:

chart	ewma1	ewma2	cusum1	cusum2	cusumC
$\mathcal{D}_0$	1685.8	829.83	114.92	56.047	74.495
$\mathcal{D}_1$	11.204	11.168	5.8533	5.8346	6.2858

XFGad.xpl

2.3 Probability Mass and Cumulative Distribution Function

The computation of the probability mass function (PMF) and of the cumulative distribution function (CDF) is implemented in two different types of routines. The first one with the syntax `spc`*chart*`pmf` returns the values of the PMF $P(L = n)$ and CDF $P(L \le n)$ at given single points of n, where *chart* has to be replaced by `ewma1`, ..., `cusumC`. The second one written as `spc`*chart*`pmfm` computes the whole vectors of the PMF and of the CDF up to a given point n, i.e. $\big(P(L = 1), P(L = 2), \ldots, P(L = n)\big)$ and the similar one of the CDF.

Note, that the same is valid as for the Average Delay (AD). In case of the two-sided CUSUM scheme the computations are based on a twodimensional Markov chain. A value of parameter r less than 30 would be computing time friendly.

With the quantlet `XFGpmf1.xpl` the 5 different schemes ($r = 50$, for `cusum2` $r = 25$) are compared according their in-control PMF and CDF ($\mu = \mu_0 = 0$) at the positions n in $\{1, 10, 20, 30, 50, 100, 200, 300\}$. Remark, that the in-control ARL of all schemes is chosen as 300.

chart	ewma1	ewma2	cusum1	cusum2	cusumC
$P(L = 1)$	$6 \cdot 10^{-8}$	$2 \cdot 10^{-9}$	$6 \cdot 10^{-6}$	$4 \cdot 10^{-7}$	$2 \cdot 10^{-6}$
$P(L = 10)$	0.00318	0.00272	0.00321	0.00307	0.00320
$P(L = 20)$	0.00332	0.00324	0.00321	0.00325	0.00322
$P(L = 30)$	0.00315	0.00316	0.00310	0.00314	0.00311
$P(L = 50)$	0.00292	0.00296	0.00290	0.00294	0.00290
$P(L = 100)$	0.00246	0.00249	0.00245	0.00248	0.00245
$P(L = 200)$	0.00175	0.00177	0.00175	0.00176	0.00175
$P(L = 300)$	0.00125	0.00126	0.00124	0.00125	0.00125
$P(L = 1)$	$6 \cdot 10^{-8}$	$2 \cdot 10^{-9}$	$6 \cdot 10^{-6}$	$4 \cdot 10^{-7}$	$2 \cdot 10^{-6}$
$P(L \le 10)$	0.01663	0.01233	0.02012	0.01675	0.01958
$P(L \le 20)$	0.05005	0.04372	0.05254	0.04916	0.05202
$P(L \le 30)$	0.08228	0.07576	0.08407	0.08109	0.08358
$P(L \le 50)$	0.14269	0.13683	0.14402	0.14179	0.14360
$P(L \le 100)$	0.27642	0.27242	0.27728	0.27658	0.27700
$P(L \le 200)$	0.48452	0.48306	0.48480	0.48597	0.48470
$P(L \le 300)$	0.63277	0.63272	0.63272	0.63476	0.63273

`XFGpmf1.xpl`

A more appropriate, graphical representation provides the quantlet `XFGpmf2.xpl`. Figure 4 shows the corresponding graphs.

3 Comparison with existing methods

3.1 Two-sided EWMA and Lucas/Saccucci

Here, we compare the ARL and AD computations of Lucas and Saccucci (1990) with XploRe results. In their paper they use as in-control ARL $\xi = 500$. Then

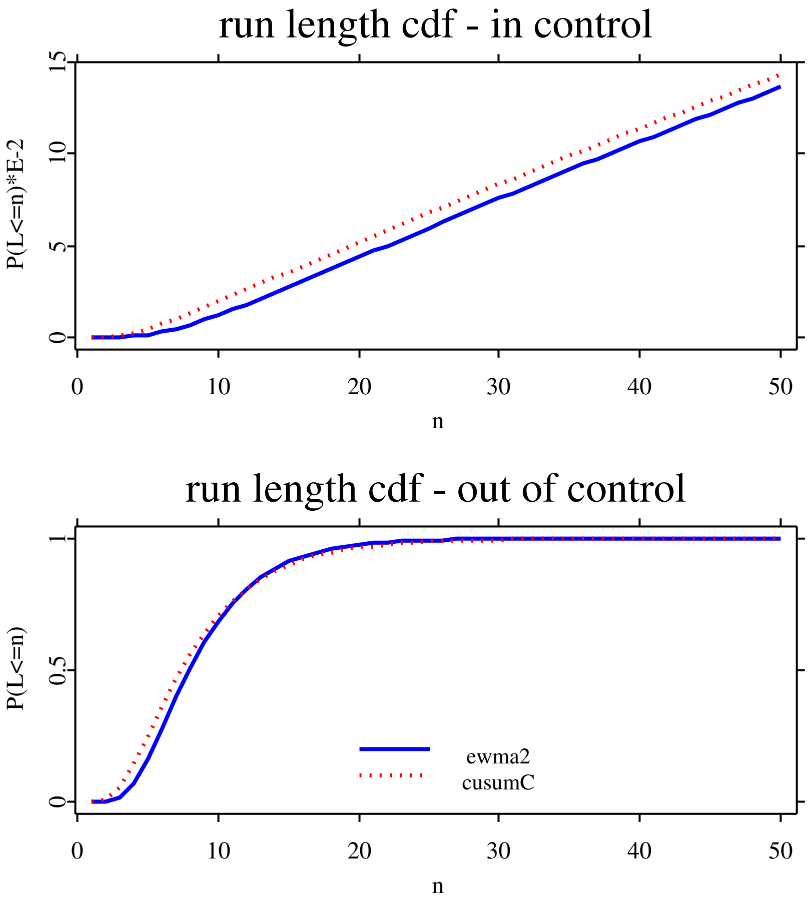


Figure 4: CDF for two-sided EWMA and Crosier's CUSUM for $\mu = 0$ (in control) and $\mu = 1$ (out of control)

`XFGpmf2.xpl`

for, e. g., $\lambda = 0.5$ and $\lambda = 0.1$ the critical values are 3.071 and 2.814, respectively. By using XploRe the related values are 3.0712 and 2.8144, respectively. It is known, that the smaller λ the worse the accuracy of the Markov chain approach. Therefore, r is set greater for $\lambda = 0.1$ ($r = 200$) than for $\lambda = 0.5$ ($r = 50$). Table 1 shows some results of Lucas and Saccucci (1990) on ARLs and ADs. Their results are based on the Markov chain approach as well. However, they used some smaller matrix dimension and fitted a regression model on r (see Subsection 3.2). The corresponding XploRe results by using the quantlet `XFGlucsac.xpl` coincide with the values of Lucas and Saccucci (1990).

`XFGlucsac.xpl`

μ	0	0.25	0.5	0.75	1	1.5	2	3	4	5
					$\lambda = 0.5$					
$\mathcal{L}_\mu$	500	255	88.8	35.9	17.5	6.53	3.63	1.93	1.34	1.07
$\mathcal{D}_\mu$	499	254	88.4	35.7	17.3	6.44	3.58	1.91	1.36	1.10
					$\lambda = 0.1$					
$\mathcal{L}_\mu$	500	106	31.3	15.9	10.3	6.09	4.36	2.87	2.19	1.94
$\mathcal{D}_\mu$	492	104	30.6	15.5	10.1	5.99	4.31	2.85	2.20	1.83

Table 1: ARL and AD values from Table 3 of Lucas and Saccucci (1990)

3.2 Two-sided CUSUM and Crosier

Crosier (1986) derived a new two-sided CUSUM scheme and compared it with the established combination of two one-sided schemes. Recall Table 3 of Crosier (1986), where the ARLs of the new and the old scheme were presented. The reference value k is equal to 0.5. First, we compare the critical values. By

μ	0	0.25	0.5	0.75	1	1.5	2	2.5	3	4	5
					old scheme, $h = 4$						
$\mathcal{L}_\mu$	168	74.2	26.6	13.3	8.38	4.74	3.34	2.62	2.19	1.71	1.31
					new scheme, $h = 3.73$						
$\mathcal{L}_\mu$	168	70.7	25.1	12.5	7.92	4.49	3.17	2.49	2.09	1.60	1.22
					old scheme, $h = 5$						
$\mathcal{L}_\mu$	465	139	38.0	17.0	10.4	5.75	4.01	3.11	2.57	2.01	1.69
					new scheme, $h = 4.713$						
$\mathcal{L}_\mu$	465	132	35.9	16.2	9.87	5.47	3.82	2.97	2.46	1.94	1.59

Table 2: ARLs from Table 3 of Crosier (1986)

using XploRe (`XFGcrosc.xpl`) with $r = 100$ one gets $c = 4.0021$ (4), 3.7304 (3.73), 4.9997 (5), 4.7133 (4.713), respectively – the original values of Crosier are written in parentheses. By comparing the results of Table 2 with the results obtainable by the quantlet `XFGcrosarl.xpl` ($r = 100$) it turns out, that again the ARL values coincide with one exception only, namely $\mathcal{L}_{1.5} = 4.75$ for the old scheme with $h = 4$.

`XFGcrosarl.xpl`

Further, we want to compare the results for the Average Delay (AD), which is called Steady-State ARL in Crosier (1986). In Table 5 of Crosier we find the related results. A slight modification of the above quantlet `XFGcrosarl.xpl` allows to compute the ADs. Remember, that the computation of the AD for the two-sided CUSUM scheme is based on a twodimensional Markov chain. Therefore the parameter r is set to 25 for the scheme called old scheme by Crosier. The results are summarized in Table 4.

While the ARL values in the paper and computed by XploRe coincide, those for the AD differ slightly. The most prominent deviation (459 vs. 455) one

μ	0	0.25	0.5	0.75	1	1.5	2	2.5	3	4	5
					old scheme, $h = 4$						
$\mathcal{L}_\mu$	163	71.6	25.2	12.3	7.68	4.31	3.03	2.38	2.00	1.55	1.22
					new scheme, $h = 3.73$						
$\mathcal{L}_\mu$	164	69.0	24.3	12.1	7.69	4.39	3.12	2.46	2.07	1.60	1.29
					old scheme, $h = 5$						
$\mathcal{L}_\mu$	459	136	36.4	16.0	9.62	5.28	3.68	2.86	2.38	1.86	1.53
					new scheme, $h = 4.713$						
$\mathcal{L}_\mu$	460	130	35.1	15.8	9.62	5.36	3.77	2.95	2.45	1.91	1.57

Table 3: ADs (steady-state ARLs) from Table 5 of Crosier (1986)

μ	0	0.25	0.5	0.75	1	1.5	2	2.5	3	4	5
					old scheme, $h = 4$						
$\mathcal{L}_\mu$	163	71.6	25.2	*12.4*	*7.72*	*4.33*	*3.05*	*2.39*	*2.01*	1.55	1.22
					new scheme, $h = 3.73$						
$\mathcal{L}_\mu$	*165*	*69.1*	*24.4*	*12.2*	*7.70*	*4.40*	3.12	*2.47*	2.07	1.60	1.29
					old scheme, $h = 5$						
$\mathcal{L}_\mu$	*455*	136	36.4	16.0	*9.65*	*5.30*	*3.69*	*2.87*	2.38	1.86	*1.54*
					new scheme, $h = 4.713$						
$\mathcal{L}_\mu$	460	130	35.1	15.8	*9.63*	*5.37*	3.77	2.95	2.45	1.91	1.57

Table 4: ADs (steady-state ARLs) computed by XploRe, different values to Table 3 are printed as italics

XFGcrosad.xpl

observes for the old scheme with $h = 5$. One further in-control ARL difference one notices for the new scheme with $h = 3.73$. All other differences are small.

There are different sources for the deviations:

1. Crosier computed $D^{(32)} = (\underline{p}^{32\,T}\underline{L})/(\underline{p}^{32\,T}\underline{1})$ and not the actual limit D (see 8, 10, and 12).

2. Crosier used $ARL(r) = ARL_\infty + B/r^2 + C/r^4$ and fitted this model for $r = 8, 9, 10, 12, 15$. Then, ARL_∞ is used as final approximation. In order to get the above $D^{(32)}$ one needs the whole vector $\underline{L}$, such that this approach might be more sensitive to approximation errors than in the single ARL case.

4 Real data example – monitoring CAPM

There are different ways of applying SPC to financial data. Here, we use a twosided EWMA chart for monitoring the Deutsche Bank (DBK) share. More precisely, a capital asset pricing model (CAPM) is fitted for DBK and the DAX which is used as proxy of the efficient market portfolio. That is, denoting with $r_{\text{DAX},t}$ and $r_{\text{DBK},t}$ the log returns of the DAX and the DBK, respectively, one assumes that the following regression model is valid:

$$r_{\text{DBK},t} = \alpha + \beta\, r_{\text{DAX},t} + \varepsilon_t \tag{14}$$

Usually, the parameters of the model are estimated by the ordinary least squares method. The parameter β is a very popular measure in applied finance, Elton and Gruber (1991). In order to construct a real portfolio, the β coefficient is frequently taken into account. Research has therefore concentrated on the appropriate estimation of constant and time changing β. In the context of SPC it is therefore useful to construct monitoring rules which signal changes in β. Contrary to standard SPC application in industry there is no obvious state of the process which one can call "in-control", i. e. there is no target process. Therefore, pre-run time series of both quotes (DBK, DAX) are exploited for building the in-control state. The daily quotes and log returns, respectively, from january, 6th, 1995 to march, 18th, 1997 (about 450 observations) are used for fitting (14):

```
A  N  O  V  A                    SS      df      MSS       F-test   P-value
---------------------------------------------------------------------------
Regression                     0.025      1     0.025     445.686   0.0000
Residuals                      0.025    448     0.000
Total Variation                0.050    449     0.000

Multiple R       = 0.70619
R^2              = 0.49871
Adjusted R^2     = 0.49759
Standard Error   = 0.00746

PARAMETERS         Beta         SE          StandB         t-test   P-value
---------------------------------------------------------------------------
b[ 0,]=         -0.0003      0.0004      -0.0000         -0.789   0.4307
b[ 1,]=          0.8838      0.0419       0.7062         21.111   0.0000
```

With `b[1,]` $= \beta = 0.8838$ a typical value has been obtained. The $R^2 = 0.49871$ is not very large. However, the simple linear regression is considered in the sequel. The (empirical) residuals of the above model are correlated (see Figure 5). The SPC application should therefore be performed with the (standardized) residuals of an AR(1) fit to the regression residuals. For an application of the XploRe quantlet `armacls` (quantlib `times`) the regression residuals were standardized. By using the conditional least squares method an estimate of $\hat{\varrho} = 0.20103$ for the AR(1) model

$$\varepsilon_t = \varrho\, \varepsilon_{t-1} + \eta_t \tag{15}$$

has been obtained. Eventually, by plugging in the estimates of α, β and ϱ, and standardizing with the sample standard deviation of the pre-run residuals series (see (15)) one gets a series of uncorrelated data with expectation 0 and variance 1, i. e. our in-control state. If the fitted model (CAPM with AR(1) noise) remains valid after the pre-run, the related standardized residuals behave like in the in-control state. Now, the application of SPC, more precisely of a twosided EWMA chart, allows to monitor the series in order to get signals, if the original model was changed. Changes in α or β in (14) or in ϱ in (15) or in the residual variance of both models lead to shifts or scale changes in the empirical residuals series. Hence, the probability of an alarm signaled by the

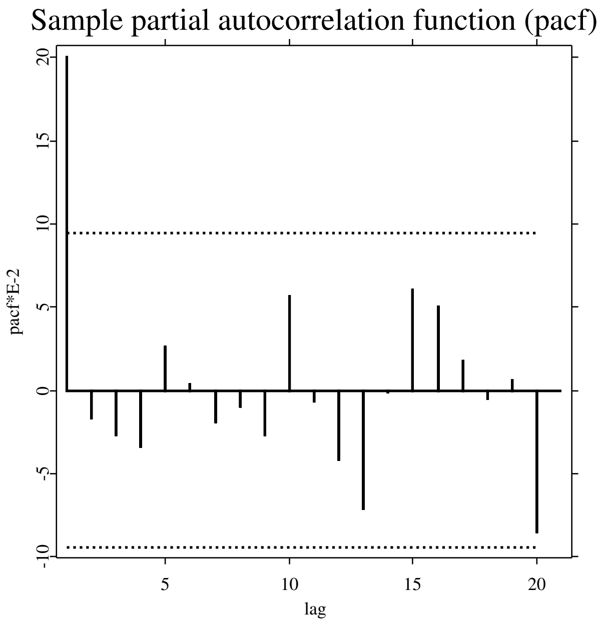


Figure 5: Partial autocorrelation function of CAPM regression residuals

EWMA chart increases (with one exception only – decreased variances). In this way a possible user of SPC in finance is able to monitor an estimated and presumed CAPM.

In our example we use the parameter $\lambda = 0.2$ and an in-control ARL of 500, such that the critical value is equal to $c = 2.9623$ (the Markov chain parameter r was set to 100). Remark, that the computation of c is based on the normality assumption, which is seldom fulfilled for financial data. In our example the hypothesis of normality is rejected as well with a very small p value (Jarque-Bera test with quantlet `jarber`). The estimates of skewness 0.136805 and kurtosis 6.64844 contradict normality too. The fat tails of the distribution are a typical pattern of financial data. Usually, the fat tails lead to a higher false alarm rate. However, it would be much more complicated to fit an appropriate distribution to the residuals and use these results for the "correct" critical value.

The Figures 6 and 7 present the EWMA graphs of the pre-run and the monitoring period (from march, 19th, 1997 to april, 16th, 1999). In the pre-run period the EWMA chart signals 4 times. The first 3 alarms seem to be outliers, while the last points on a longer change. Nevertheless, the chart performs quite typical for the pre-run period. The first signal in the monitoring period was obtained at the 64th observation (i. e. 06/24/97). Then, we observe more frequently signals than in the pre-run period, the changes are more persistent and so one has to assume, that the pre-run model is no longer valid. A new CAPM has therefore to be fitted and, if necessary, the considered portfolio has to be reweighted. Naturally, a new pre-run can be used for the new monitoring period.

`XFGcapmar1.xpl`

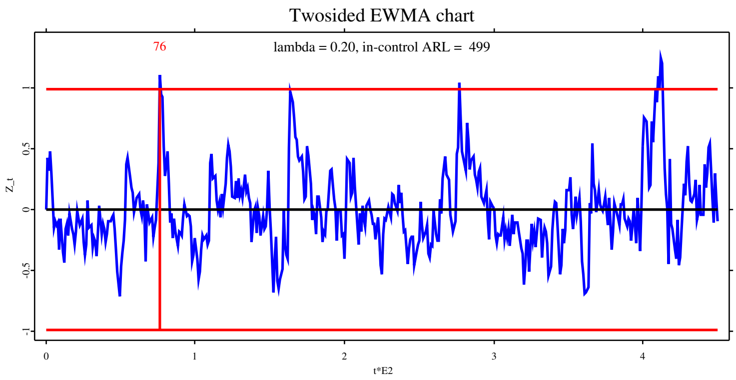


Figure 6: Twosided EWMA chart of the standardized CAPM-AR(1) residuals for the pre-run period (06/01/95 - 03/18/97)

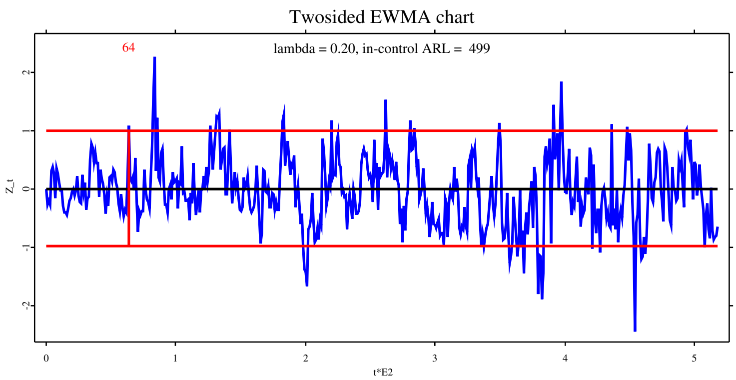


Figure 7: Twosided EWMA chart of the standardized CAPM-AR(1) residuals for the monitoring period (03/19/97 - 04/16/99)

References

Brook, D. and Evans, D. A. (1972). An approach to the probability distribution of cusum run length, *Biometrika* **59**: 539–548.

Crosier, R. B. (1986). A new two-sided cumulative quality control scheme, *Technometrics* **28**: 187–194.

Crowder, S. V. (1986). A simple method for studying run-length distributions of exponentially weighted moving average charts, *Technometrics* **29**: 401–407.

Elton, EJ. and Gruber, MJ. (1991). *Modern portfolio theory and investment analysis*, Wiley, 4. edition.

Knoth, S. (1998). Quasi-stationarity of CUSUM schemes for Erlang distributions, *Metrika* **48**: 31–48.

Lorden, G. (1971). Procedures for reacting to a change in distribution, *Annals of Mathematical Statistics* **42**: 1897–1908.

Lucas, J. M. (1976). The design and use of V-mask control schemes, *Journal of Quality Technology* **8**: 1–12.

Lucas, J. M. and Crosier, R. B. (1982). Fast initial response for cusum quality-control schemes: Give your cusum a headstart, *Technometrics* **24**: 199–205.

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

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

Pollak, M. and Siegmund, D. (1975). Approximations to the expected sample size of certain sequential tests, *Annals of Statistics* **3**: 1267–1282.

Pollak, M. and Siegmund, D. (1985). A diffusion process and its applications to detection a change in the drift of brownian motion, *Biometrika* **72**: 267–280.

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

Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product*, D. van Nostrand Company, Inc., Toronto, Canada.

Vance, L. (1986). ARL of cumulative sum control chart for controlling normal mean, *Journal of Quality Technology* **18**: 189–193.

Waldmann, K.-H. (1986). Bounds to the distribution of the run length in general quality–control schemes, *Statistische Hefte* **27**: 37–56.