CUSUM, EWMA, and Shiryaev-Roberts under drift

Sven Knoth

Department of Economics and Social Sciences,
Helmut Schmidt University, University of the Federal Armed Forces,
Postfach 700822, 22008 Hamburg, Germany
Sven.Knoth@hsu-hh.de

Abstract. There is much less literature on drift detection than on step change monitoring. Fortunately, the procedures constructed for the step change behave quite well also for identifying drifts. The calculation of the usual performance measures is more demanding under drift. Moreover, there are many types of special drift detection schemes. The talk will provide numerical results and present various schemes.

Keywords. ARL, CUSUM, drift, EWMA, SPC.

1 Introduction

Typically, change point detection schemes (control charts) such as Page's CUSUM, Roberts' EWMA, and the Shiryaev-Roberts procedures are setup to detect a step change as quickly as possible while maintaining a low false alarm rate. In practical applications, a change may happen gradually. Drifts linear in time or in observation number are the most simple model. As usual, these models would already cover a wide range of applications. Nevertheless, their numerical treatment, looking for optimal or nearly optimal schemes etc. is less intensively studied than for the popular step change model. While designing a control chart it is not clear, whether a specific drift detection scheme should be set up or the usual step change chart is needed anyway. Thus, one should not only construct new schemes, but also analyze the performance of the classics under drift. Note that a very simple approach for constructing a linear drift detection scheme could rely on the differences of the observed data sequence, which is something like a signed Moving Range statistic. Davis and Woodall (1988) evaluated popular chart examples from the quality circle world – the so-called trend rules. However, they conclude that they are ineffective in detecting drifts. Most of the papers deploy Monte Carlo studies to get performance measures such as the Average Run Length. The most accurate numerical algorithm was developped by Gan (1991, 1992). It is surprising that nearly no other author is using his method. Here, his approach and one already used in Knoth (2003) are utilized for one-sided EWMA, CUSUM and Shiryaev-Roberts schemes. Contrary to the step change performance it is quite difficult to evaluate the drift performance of two-sided CUSUM and Shiryaev-Roberts schemes. For all other schemes it is even more complicated so that Monte Carlo studies dominate. The aim of this paper is twofold. First, it illustrates that on the current level of knowledge, the classics could be used for drift detection. Second, it should inspire the SPC community to do more in order to enhance the knowledge about drift detection.

2 Status quo drift detection with attention to numerical algorithms

One of the first papers is Bissell (1984) who already treated CUSUM charts under drift. His numerical algorithm, a modified Markov chain approach, did not work well (also noted by himself in a correction) and was refined in Asbagh (1985). Bissell's conclusions regarding the choice of an appropriate scheme are therefore not reliable. Then, in Davis and Woodall (1988) the Shewhart chart with trend rules was analyzed. The authors clearly discourage from their usage. A more elaborated study was done in Aerne et al. (1991), where based on Markov chain and Monte Carlo methods Shewhart charts with and without runs rules, CUSUM and EWMA were considered. Basically, they finally suggest to apply the competing charts as in the step change case. That is, for small drift coefficients deploy CUSUM or EWMA, otherwise Shewhart charts. In the same year, Gan (1991) published the first time his algorithm to attain high accuracy in calculating Average Run Lengths of classical control charts under drift. His first paper discusses two-sided EWMA charts. In Gan (1992) one-sided CUSUM charts were analyzed. His idea is simple and impressive. It should be sketched here for the two-sided EWMA chart. Beforehand, the general change point model utilized in this paper is described.

Let $X_1, X_2, \ldots$ be a sequence of independent normal random variables with variance 1. Their mean is under risk to change. Specifically,

$$E(X_t) = \begin{cases} \mu_0 = 0 & , t < \tau \\ \mu_{t-\tau} = (t - \tau + 1)\Delta & , t \geq \tau \end{cases} .$$

Note that Gan preferred $\mu_{t-\tau} = (t - \tau)\Delta$ with the special feature $E(X_\tau) = 0 = \mu_0$. Here, the above model is chosen. The parameter Δ is the drift coefficient. Gan considered the case $\tau = \infty$ as in-control scenario and $\tau = 1$, as usual, as representative out-of-control pattern. Imagine now that the change, step or drift, should be detected by an EWMA control chart. Thus, apply Roberts' (1959) classic EWMA:

$$\begin{aligned} Z_0 &= z_0 = \mu_0 = 0\,, \\ Z_t &= (1 - \lambda) Z_{t-1} + \lambda X_t \quad , \ \lambda \in (0, 1]\,, \\ L &= \inf \left\{ t \in I\!N : |Z_t| > c \sqrt{\lambda / (2 - \lambda)} =: C \right\} . \end{aligned}$$

The most popular performance measure is the zero-state Average Run Length (ARL). It is defined as

$$ARL = \begin{cases} E_\infty(L) & \text{, in-control case: no change at all.} \\ E_1(L) & \text{, out-of-control case: change already at the beginning.} \end{cases}$$

Is is a simplified measure and already often criticized. In the drift detection area, it is the dominating one. Now Gan denoted with

$$\mathcal{L}_j(y, \mu_j)$$

the ARL for an EWMA chart starting at $z_0 = y$ and mean sequence $\mu_j = j\Delta, \mu_{j+1}, \ldots, \mu_m, \mu_m, \ldots$ He showed in Gan (1991) that for $y \in [-C, C]$

$$\begin{aligned} \mathcal{L}_j(y, \mu_j) &= 1 + \int_{-C}^{C} \mathcal{L}_{j+1}(y, \mu_{j+1}) \frac{1}{\lambda} f_{\mu_j} \left(\frac{x - (1 - \lambda) y}{\lambda} \right) dx \ , \ j = 0, 1, \ldots, m - 1\,, \\ \mathcal{L}_m(y, \mu_m) &= 1 + \int_{-C}^{C} \mathcal{L}_m(y, \mu_m) \frac{1}{\lambda} f_{\mu_m} \left(\frac{x - (1 - \lambda) y}{\lambda} \right) dx\,. \end{aligned}$$

For fixed m, the latter integral equation can be solved as usual with the Nyström method utilizing a reasonable quadrature method such as Gauss-Legendre. Afterwards, the first iteration sequence is executed by plugging in the same quadrature as for the integral equation. The final approximation of $\mathcal{L}_0(y = z_0, \mu_1)$ (Gan took $\mathcal{L}_0(y = z_0, \mu_0)$) provides the ARL $E_1(L)$. It remains the choice of the index m. A simple idea is to increase m until $\mathcal{L}_0(y = z_0, \mu_0)$ remains constant. It depends heavily on the drift rate Δ. Gan reported for $\Delta = 1, .1, .01, .001$ and $.0001$ the following values for m: 6, 30, 150, 700 and 2500, respectively. For implementing Gan's algorithm one should hide the search for sufficiently large m. The resulting procedure is slower than the method described in Knoth (2003). This algorithm could also be used to calculate the ARL for drift. Additionally, it could be extended easily for the steady-state ARL and ARL vehicles for $1 < \tau < \infty$.

For CUSUM, in Gan (1996) the Algorithm AS 305 was published. In this paper Gan also mentioned that the performance of the usual two-sided CUSUM chart – simultaneously running a lower and an upper chart – could not not be treated under drift like under a step change. The problem occurs, e. g., by calculating the ARL of the lower chart, if an upwards drift is present. The above approach – increase m until the resulting ARL does not change anymore – fails because $\mathcal{L}_m(y, \mu_m)^{\text{lower}}$ becomes larger for increasing m. Or to put it in other words, if the lower chart does not signal already during the first drifted values, then it becomes quite unlikely that it will ever signal. To sum up, except Monte Carlo methods no other method is currently available for calculating reliably the ARL of two-sided CUSUM schemes under drift.

Already before Gan (1991), Sweet (1988) published an algorithm for monitoring data with potential drifts. His coupled EWMA charts directly monitor the slope of the data:

$$
\begin{aligned}
S_0 &= \mu_0 = 0\,,\\
S_t &= (1-\lambda_S)(S_{t-1}+B_{t-1}) + \lambda_S X_t\,,\\
B_0 &= \Delta_0 = 0\,,\\
B_t &= (1-\lambda_B)B_{t-1} + \lambda_B(S_t - S_{t-1})\,.
\end{aligned}
$$

The sequence B_t is built for monitoring the trend slope. S_t is, more or less, the usual mean monitor. Sweet's framework allows also to monitor changes from one drift coefficent to another. The whole design stems from the forecasting literature (Holt, 1960) and seems to be rarely discussed in SPC literature. Sweet provides some guide lines to design the two control charts. Thereby, he tries to link distributional properties of the sequences to the standard Shewhart $3\,\sigma$ rule. He did not calculate performance measures like the ARL. It is not that surprinsing, because the design of these coupled EWMA charts is not easily treated numerically. Thus, only Monte Carlo studies could be done.

A further contribution to both the modelling of the change point for potentially drifting data and new monitoring procedures is Chang and Fricker Jr. (1999). They look at a broader class of drift patterns: monotonically increasing means. However, they have a maximal value for the mean in mind. The aim of their scheme is not to detect the drift itself, but the exceeding of that maximal value. This sounds reasonable for some practical applications. Chang and Fricker Jr. evaluate one-sided CUSUM and EWMA charts, and a GLR-like scheme. The latter is custom-built for detecting the exceedance of their threshold mean under their more general mean drift model. The basic idea of GLR (generalized likelihood ratio) in change detection is to dispense the explicite knowledge of the post-change parameters. Here, the additional estimation task (under the monotonicity restriction) leads to isotonic regressions. Of course, the complicate implementation of GLR schemes in general and specifically for the model here hampers their application in practice. As side effect it offers a nice on-line estimate of the mean. However, Chang and Fricker Jr. conclude that the classics are not outperformed by the specialized GLR procedure. Thus, there is no reason to switch from the classics in the considered drift change point model. Note that the authors calculated a kind of steady-state ARL with Monte Carlo studies. Thus, it is difficult to compare it to any other analysis done for drift monitoring.

More recently, Reynolds Jr. and Stoumbos (2001), studied drift behavior in simultaneous monitoring schemes (mean and variance). They allowed both drifts in mean and variance. All was evaluated with Monte Carlo studies. It was concluded, not surprisingly, that EWMA outperforms the Shewhart counterparts for small and moderate drift cofficients. Note that separate analysis of variance control charts under drift would be a interesting and useful task.

Fahmy and Elsayed (2006) introduced a drift detection scheme based on slope estimates on rolling windows. Afterwards, they compared their new chart with one-sided Shewhart, CUSUM and EWMA charts, and drift GLR scheme under drift. They obtained their numbers by Monte Carlo simulation. They concluded that their scheme is better than the classics under drift. Some more reflections about their results will be given in the next section. Given the only slight performance advantages and the higher complexity of the rolling window scheme, in application the classics would outlast.

The most recent paper is Zou et al. (2008). It gives a thorough introduction to the subject and a large comparison study for one-sided EWMA, CUSUM, GEWMA (the smoothing constant λ is optimized in a certain way), and GLR schemes for the step change and drift. All schemes are analyzed for drift coefficients from very small up to large ones. They deploy only Monte Carlo studies. Based on ther measure RMI (measures the performance over the whole range of considered drift coefficients) they conclude that the more sophisticated schemes outperform the classics. This will be reviewed in the next section.

3 New and old Results

3.1 Fahmy and Elsayed (2006)

Start with a comparison of two-sided EWMA and CUSUM results given in Fahmy and Elsayed (2006) and new ones based on a much more extensive Monte Carlo study and, for EWMA, on the Gan algorithm.

	CUSUM		EWMA		
Δ	FE (10^4)	here (10^7)	FE (10^4)	her (10^7)	here (Gan)
0	368.333±3.549	368.251±0.111	365.749±3.598	369.021±0.114	368.994
0.10	13.986±0.026	14.086±0.001	12.971±0.029	12.986±0.001	12.986
0.25	8.560±0.014	8.656±0.000	7.738±0.015	7.758±0.000	7.758
0.50	5.946±0.008	6.033±0.000	5.312±0.009	5.318±0.000	5.318
0.75	4.827±0.007	4.898±0.000	4.279±0.007	4.286±0.000	4.285
1.00	4.156±0.006	4.224±0.000	3.680±0.006	3.688±0.000	3.688
2.00	2.950±0.003	2.989±0.000	2.598±0.005	2.616±0.000	2.616

Table 1. Comparison of ARL values for two-sided CUSUM and EWMA charts under drift. FE denotes the numbers from Fahmy and Elsayed (2006) based on a MC runs with 10^4 replicates. The remaining numbers stem from MC runs with 10^7 replicates and Gan's procedure.

Note that the EWMA results of Fahmy and Elsayed (2006) match quite well to the more recent ones, while there are some slight differences for CUSUM. Fahmy and Elsayed concluded that their rolling regression chart (the resulting statistic is χ^2-distributed) outperform the rest. This is mainly because of their specific setup of the EWMA and CUSUM competitors. Taking different λ values changes the picture, see Table 2:

	FE χ^2			EWMA		
Δ	$w^* = 3$	$w^* = 5$	$w^* = 20$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.5$
0	379.138±3.790	370.048±3.682	373.458±3.546	370	370	370
0.10	17.445±0.056	16.047±0.049	12.860±0.035	**12.747**	13.041	14.136
0.25	8.537±0.032	8.127±0.023	7.623±0.018	7.304	**7.231**	7.497
0.50	5.027±0.021	4.869±0.013	5.260±0.012	4.881	4.722	**4.706**
0.75	3.672±0.017	3.673±0.009	4.250±0.009	3.886	3.715	**3.620**
1.00	**2.939**±0.014	3.055±0.007	3.660±0.008	3.318	3.149	3.023
2.00	**1.816**±0.003	2.042±0.000	2.579±0.000	2.254	2.124	2.005

Table 2. Comparison of the ARL values of Fahmy and Elsayed (2006) favorite schemes (χ^2) and additional EWMA charts under drift. The bold values mark the smallest ARL values.

Because the considered drift coefficients are quite large, one should apply λ values larger than 0.1. Additionally, in other papers also much smaller drift coefficients are evaluated (see also next section). Summing up, the rolling window procedure needs more computational efforts and exhibits not better performance than classical control charts under drift.

3.2 Zou et al. (2008)

Zou et al. evaluated one-sided control charts. Among them are the classical EWMA and CUSUM chart, the more sophisticated GEWMA (the smoothing constant of the EWMA sequence is permanently adjusted), and two GLR charts designed for step changes and drifts. All results were calculated by Monte Carlo studies. Based on the measure RMI, which evaluates the robustness of the charts to various magnitudes of drifts, they conclude that the GEWMA and the drift GLR outperform the rest. Nevertheless, on smaller ranges of possible drift coefficients, the classics keep up with the newer ones or even beat them. It seems so that mainly the specific dynamic adaption of one of the chart parameters gives the more recent schemes some advance. In Table 3 the EWMA and CUSUM (zero-state) ARL results are given now together with Shiryaev-Roberts ARL values under drift. To perform the comparison as in Zou et al. (2008) the Monte Carlo results from their Table 1 are added.

	EWMA			CUSUM			GRSR			GEWMA	GLR-S	GLR-L
Δ	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$			
0	1750	1747	1733	1741	1742	1735	1730	1730	1730	-	-	-
0.0005	**318**	378	437	345	412	468	337	399	448	375	381	368
0.001	**215**	254	295	231	276	316	227	267	301	252	257	249
0.005	**83.5**	92.2	106	86.7	98.3	112	85.8	95.7	107	96.2	97.8	95.4
0.01	**55.7**	58.7	66.3	57.0	61.9	69.4	56.6	60.4	66.6	62.1	63.3	62.0
0.05	22.6	**21.1**	22.0	22.6	21.6	22.6	22.7	21.4	22.1	22.4	22.7	22.5
0.1	15.5	**13.9**	**13.9**	15.4	14.0	14.2	15.7	14.1	14.0	14.4	14.6	14.5
0.5	6.65	5.56	**5.09**	6.60	5.54	5.16	6.84	5.76	5.32	5.10	5.23	5.18
1.0	4.67	3.83	3.43	4.63	3.80	3.45	4.86	4.03	3.66	**3.26**	3.38	3.31
2.0	3.21	2.74	2.32	3.17	2.67	2.32	3.42	2.91	2.59	**2.09**	2.16	2.12
3.0	2.86	2.06	1.98	2.79	2.04	1.96	2.97	2.20	2.02	**1.69**	1.75	1.72
4.0	2.14	2.00	1.83	2.10	1.98	1.74	2.39	2.20	1.97	**1.31**	1.37	1.34

Table 3. Zero-state ARL values of one-sided EWMA, CUSUM and Shiryaev-Roberts (GRSR) procedures. The schemes are optimized for step changes of size δ. The corresponding λ values are 0.03479, 0.11125, and 0.23052. Additionally, the Monte Carlo results for GEWMA, GLR-S, and GLR-L are taken from Table 1 in Zou et al. (2008). The smallest values are boldly written.

Roughly speaking, the classical schemes do not differ heavily under drift in terms of the zero-state ARL. Moreover, the EWMA (in it's classical dress and as GEWMA) exhibits the best zero-state ARL behavior. Note that the EWMA type schemes start at their mean level, while the other three schemes are evaluated from their worst level.

Therefore, the next table shows all the related steady-state ARL values (but not for the sophisticated schemes). They are calculated based on the algorithms described in Knoth (2003). Note that it could be done also with Gan's procedure together with the approximated left eigenfunction of the chart transition kernel in the in-control case.

	EWMA			CUSUM			GRSR		
Δ	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$
0.0005	**314**	376	436	340	410	467	333	397	446
0.001	**213**	253	295	228	275	315	224	266	301
0.005	**82.6**	91.8	106	85.4	97.9	112	84.2	95.1	107
0.01	**55.1**	58.4	66.2	55.9	61.6	69.2	55.3	60.0	66.4
0.05	22.3	**20.9**	21.9	21.8	21.4	22.6	21.6	21.1	21.9
0.1	15.4	13.8	13.8	14.8	13.8	14.1	14.7	**13.7**	13.8
0.5	6.59	5.50	**5.05**	6.17	5.36	5.08	6.18	5.38	5.08
1.0	4.62	3.79	**3.40**	4.30	3.65	3.37	4.31	3.68	**3.40**
2.0	3.27	2.66	2.33	2.98	2.53	2.26	3.00	2.58	**2.30**
3.0	2.68	2.13	1.91	2.50	1.99	**1.90**	2.52	2.01	1.92
4.0	2.32	1.90	1.73	2.01	1.89	**1.66**	2.04	1.90	1.73

Table 4. Steady-state ARL values of one-sided EWMA, CUSUM and Shiryaev-Roberts (GRSR) procedures under drift. The schemes are optimized for step changes of size δ. The corresponding λ values are 0.03479, 0.11125, and 0.23052. The smallest values are boldly written.

There are only slight differences between the three schemes in both the zero-state and steady-state ARL under drift. EWMA remains, however, a good candidate for detecting also drifts.

4 Conclusions

Generally speaking, the schemes specifically designed for detecting drifts (instead of a step change) are not really worth the effort. The classical charts as CUSUM, EWMA, and Shiryaev-Roberts for mean surveillance are sufficiently sensitive to detect also drifts, even small ones. It is open how the ARL of two-sided CUSUM charts under drift could be numerically calculated. For the Shiryaev-Roberts scheme it is not clear, how the two-sided chart looks like – there are several ideas to create one. Summarizing, statistical drift monitoring is just at it's beginning.

References

Aerne, L. A., Champ, C. W. and Rigdon, S. E. (1991). Evaluation of control charts under linear trend. *Commun. Stat., Theory Methods*, **20**(10), 3341–3349.

Asbagh, N. A. (1985). Performance of CUSUM and combined Shewhart-CUSUM charts under linear trend. *Unpublished MSc Thesis*, Department of Statistics, University of Southwestern Louisiana, Lafayette.

Bissell, A. F. (1984). The performance of control charts and Cusums under linear trend. *Applied Statistics*, **33**(2), 145–151.

Chang, J. T. and Fricker Jr., R. D. (1999). Detecting when a monotonically increasing mean has crossed a threshold. *Journal of Quality Technology*, **31**(2), 217–234.

Davis, R. B. and Krehbiel, T. C. (2002). Shewhart and zone control chart performance under linear trend. *Commun. Stat. Simula. Comput.*, **31**(1), 91–96.

Davis, R. B. and Woodall, W. H. (1988). Performance of the control chart trend rule under linear shift. *Journal of Quality Technology*, **20**, 260–262.

Fahmy, H. M. and Elsayed, E. A. (2006). Detection of linear trends in process mean. *International Journal of Production Research*, **44**(3), 487–504.

Gan, F. F. (1991). EWMA control chart under linear drift. *J. Stat. Comput. Simulation* **38**(1-4), 181–200.

Gan, F. F. (1992). CUSUM control charts under linear drift. *Statistician*, **41**(1), 71–84.

Gan, F. F. (1996). Average Run Lengths for Cumulative Sum control chart under linear trend. *Applied Statistics,* **45**(4), 505–512.

Holt, C. C. (1960). *Planning Production, Inventories, and Work Force.* Prentice-Hall, Englewood Cliffs, New Jersey, Chapter 14.

Knoth, S. (2003). EWMA schemes with non-homogeneous transition kernels. *Sequential Analysis*, **22**(3), 241–255.

Reynolds Jr., M. R. and Stoumbos, Z. G. (2001). Individuals control schemes for monitoring the mean and variance of processes subject to drifts. *Stochastic Anal. Appl.*, **19**(5), 863–892.

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

Sweet, A. L. (1988). Using coupled EWMA control charts for monitoring processes with linear trends. *IIE Transactions*, **20**, 404–408.

Zou, C., Liu, Y., and Wang, Z. (2008). Comparisons of control schemes for monitoring the means of processes subject to drifts. *Metrika – Published online*, DOI 10.1007/s00184–008–0183–6.