

A Distribution-Free Phase II EWMA Control Chart for Joint Monitoring of Location and Scale^{*}

SONG Zhi^{1,2} LIU Yanchun³ TAO Guihong^{1*}

(¹*College of Science, Shenyang Agricultural University, Shenyang, 110866, China*)

(²*School of Mathematics, Liaoning University, Shenyang, 110036, China*)

(³*Business School, Liaoning University, Shenyang, 110036, China*)

Abstract: A single distribution-free (nonparametric) Phase II exponentially weighted moving average (EWMA) chart based on the Cucconi statistic, referred to as the EWMA-Cucconi (EC) chart, is considered here for simultaneously monitoring shifts in the unknown location and scale parameters of a univariate continuous process. A comparison with some other existing nonparametric EWMA charts is presented in terms of the average, the standard deviation and some percentiles of the run length distribution. Numerical results based on Monte Carlo analysis show that the EC chart provides quite a satisfactory performance. The effect of the Phase I (reference) sample size on the IC performance of the EC chart is studied in detail. The application of the EC chart is illustrated by two real data examples.

Keywords: Cucconi test; exponentially weighted moving average; nonparametric; run length distribution

2010 Mathematics Subject Classification: 62P30; 62G35

Citation: SONG Z, LIU Y C, TAO G H. A distribution-free Phase II EWMA control chart for joint monitoring of location and scale [J]. Chinese J Appl Probab Statist, 2019, 35(6): 639-653.

§1. Introduction

Distribution-free (nonparametric) control charts can be useful in practice when there is often a lack of enough knowledge about the process distribution. Although research on nonparametric control charts has increased in recent years, the majority of existing charts focus only on detecting process location shifts, such as mean and median. However,

^{*}The project was supported by the National Natural Science Foundation of China (Grant Nos. 11571191; 11431006; 11371202), the Scientific Research Fund of Liaoning Provincial Education Department of China (Grant No. LSNQN201912), the Science and Technology Project of Hebei Science and Technology Department (Grant No. 162176489) and the Education and Teaching Research Project of Shenyang Agricultural University (Grant No. 2018-84).

^{*}Corresponding author, E-mail: guihongtao@syau.edu.cn.

Received January 2, 2018. Revised August 8, 2018.

in practice, shifts in a process may happen to both the location and scale parameters simultaneously, hence schemes based on a single plotting statistic for joint monitoring of the location and scale parameters are often preferred. Marozzi^[1] comprehensively reviewed and compared several nonparametric tests for the jointly detection of location and scale changes. For the two-sample location-scale problem, the Lepage^[2] test which is a combination of the Wilcoxon test for location and the Ansari-Bradley test for scale has received more attention. Many nonparametric charts have been proposed based on Lepage-type tests, see [3–7].

Recently, Marozzi^[1,8] showed that the Cucconi^[9] test is slightly more powerful than the Lepage test for the two-sample location-scale problem in many cases, and the Cucconi statistic can be computed more simply. Chowdhury et al.^[10] proposed a nonparametric Shewhart chart based on the Cucconi statistic, referred to as Shewhart-Cucconi (SC) chart, for joint monitoring of location and scale parameters. Mukherjee^[11] presented an overview of the recent developments based on both Lepage and Cucconi statistics, including CUSUM schemes and EWMA schemes. Among them, a EWMA chart on the basis of the Cucconi statistic, referred to as EC chart, has no detailed results. Motivated by this, we present a detailed performance analysis for the EC chart from various angles. It is expected that the EC chart blends the advantages of a EWMA with that of the Cucconi test in detecting small to moderate shifts. The rest of the paper is organized as follows. In the next Section, the EC chart is presented. Section 3 is devoted to a comparison of the run length performances of the EC chart and some other EWMA nonparametric charts for a large class of location-scale models under varying magnitude of shifts in the process location and/or scale. In Section 4, we examine the effect of the reference sample size on the performance of the EC chart. The application of the EC chart is illustrated in Section 5 by two real data examples. Several remarks conclude this paper in Section 6.

§2. The Proposed EC Chart

2.1 Construction of EC Chart

Suppose that a reference sample $\mathbf{X} = (X_1, X_2, \dots, X_m)$ collected from an IC process with a continuous cumulative distribution function (cdf) $F(x)$. Let $\mathbf{Y}_j = (Y_{j1}, Y_{j2}, \dots, Y_{jn})$, $j = 1, 2, \dots$ be the j th Phase II (test) sample of size n mutually independent of the reference sample, from a cdf $G(x) = F((x - \theta)/\delta)$, $\theta \in \mathfrak{R}$, $\delta > 0$, where the constants θ

and δ represent the unknown location and scale parameter, respectively. The process is IC when $\theta = 0$, $\delta = 1$. Let w_k denotes the rank of Y_k ($k = 1, 2, \dots, n$) in the pooled sample $(X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n)$ of size $N (= m+n)$. Consider the Wilcoxon rank sum test statistic, say T , which is the sum of the ranks of the second sample Y_k 's in the combined sample. Similarly, S_1 represents the sum of squares of the ranks of Y_k 's in the combined sample and is given by $S_1 = \sum_{k=1}^n \omega_k^2$. Further, the sum of the squares of anti-ranks of Y_k 's in the combined sample, say S_2 , is given by

$$S_2 = \sum_{k=1}^n (N+1-w_k)^2 = n(N+1)^2 - 2(N+1)T + S_1.$$

Define the standardized statistics:

$$U = \frac{S_1 - \mu_1}{\sigma_1}, \quad V = \frac{S_2 - \mu_2}{\sigma_2} \quad \text{and} \quad \rho = \text{Corr}(U, V | IC),$$

where (μ_1, μ_2) and (σ_1, σ_2) are the respective means and standard deviations of S_1 and S_2 , ρ denotes the correlation coefficient between U and V , under the null hypothesis: $\theta = 0$ and $\delta = 1$. It is well known that

$$\mu_1 = \mu_2 = \frac{n(N+1)(2N+1)}{6}, \quad \sigma_1 = \sigma_2 = \sqrt{\frac{mn}{180}(N+1)(2N+1)(8N+11)},$$

and

$$\rho = \frac{2(N^2 - 4)}{(2N+1)(8N+11)} - 1.$$

Cucconi^[9] proposed a rank statistic to address the location-scale problem, based on

$$C = \frac{U^2 + V^2 - 2\rho UV}{2(1 - \rho^2)}.$$

As noted by Marozzi^[8] the C statistic corresponds to one half of the squared Mahalanobis distance between U and V . Let C_j denote the Cucconi statistic for the reference sample and the j th test sample, the plotting statistic of the EC control chart can be defined as

$$E_j = \lambda C_j + (1 - \lambda)E_{j-1}, \quad j = 1, 2, 3, \dots,$$

with the starting value $E_0 = 1$, as $E(C | IC) = 1$. Here, $0 < \lambda \leq 1$ is the smoothing parameter of the EWMA scheme. Note that, either in location or in scale or in both, larger values of E_j , may indicate a shift, that is an out-of-control (OOC) process. Consequently, the EC chart issues an OOC signal if $E_j > H$, where H is the upper control limit.

2.2 Performance Measures

Let RL be the random variable denoting the run length of the EC chart. The performance of a control chart in Phase II is traditionally measured by the distribution function of the run length (RL). Note that, for a given reference sample $\mathbf{X} = (X_1, X_2, \dots, X_m)$ with cdf $F(x)$, the conditional run length distribution can be written as $P(RL = t | \mathbf{X})$, where t is a realization of RL . Hence, all properties of the unconditional run length distribution can be obtained by integrating over the distribution of the reference sample

$$P(RL = t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} P(RL = t | \mathbf{X}) dF(x_1) \cdots dF(x_m).$$

Such a formulation is an m -dimensional integration and the exact conditional distribution of the Cucconi statistic is complicated with no clear explicit form. Hence, it is difficult to implement in practice. Alternatively, we employ Monte Carlo simulations to approximate the necessary quantities in this paper. Details of the simulations and the results are discussed in the subsequent sections.

The average run length (ARL) and the standard deviation of the run length ($SDRL$) are popular performance indicators, but since the run length distribution is right skewed, it is worthwhile to study various summary measures viz a number of percentiles including the 5th, 25th, 50th, 75th and the 95th.

It should be emphasized that several researchers have recommended other chart performance measures, such as the median run length (MRL), see [12, 13]. The MRL provides several advantages when the run length distribution is highly right skewed. The use of the MRL ensures a better control over the false alarm rate (FAR) in the sense that no more than 50% of the false alarms are guaranteed to be realised before the IC MRL (MRL_0). Therefore, in this paper, the H for the EC chart is given based on a target IC ARL (ARL_0) or a target MRL_0 in the next subsection.

2.3 Determination of H

A Monte Carlo simulation study is conducted in FORTRAN to determine H on the basis of 50,000 replications. Because of the distribution-free nature of the EC chart, we generate m observations from a standard normal distribution for the Phase I sample and n observations from the same distribution for each test sample. Table 1 lists some (λ, H) combinations for which the ARL_0 or the MRL_0 is equal to a nominal value, i.e. 250, 370,

500, for small to moderate reference sample sizes $m = 30, 50, 100, 150$ and test sample sizes $n = 5, 11$, respectively.

Table 1 Charting constant H for the EC chart, for various of m and n , and for some nominal values of ARL_0 and MRL_0

λ	m	n	ARL_0			MRL_0		
			250	370	500	250	370	500
0.05	30	5	1.128	1.159	1.187	1.284	1.319	1.343
	30	11	1.078	1.109	1.133	1.260	1.291	1.311
	50	5	1.185	1.217	1.244	1.325	1.353	1.382
	50	11	1.161	1.189	1.216	1.299	1.334	1.360
	100	5	1.244	1.280	1.308	1.346	1.389	1.422
	100	11	1.238	1.271	1.298	1.344	1.371	1.400
	150	5	1.270	1.308	1.336	1.357	1.398	1.431
	150	11	1.267	1.301	1.331	1.351	1.391	1.421
0.1	30	5	1.324	1.366	1.406	1.519	1.571	1.609
	30	11	1.274	1.319	1.356	1.492	1.536	1.569
	50	5	1.400	1.493	1.488	1.572	1.635	1.675
	50	11	1.374	1.418	1.459	1.546	1.591	1.629
	100	5	1.478	1.535	1.580	1.620	1.689	1.733
	100	11	1.474	1.520	1.561	1.600	1.652	1.695
	150	5	1.513	1.574	1.622	1.645	1.708	1.757
	150	11	1.507	1.562	1.604	1.623	1.681	1.724
0.2	30	5	1.654	1.721	1.779	1.936	2.014	2.082
	30	11	1.629	1.691	1.741	1.882	1.950	1.999
	50	5	1.773	1.854	1.919	2.026	2.121	2.195
	50	11	1.744	1.815	1.875	1.971	2.038	2.098
	100	5	1.898	1.996	2.081	2.122	2.223	2.307
	100	11	1.876	1.956	2.013	2.061	2.140	2.215
	150	5	1.970	2.059	2.144	2.156	2.251	2.336
	150	11	1.930	2.011	2.079	2.101	2.192	2.264

From Table 1, it is observed that the general pattern is quite similar for the control limit H for both the ARL and MRL, although the H value for a target MRL_0 is higher than that for the same ARL_0 value. For any fixed combination of (m, n, λ) values, the higher the nominal ARL_0 or MRL_0 values, the higher the values of H . Further, the H increases with m when both λ and n are fixed, but the H decreases with an increase in the test sample size n for fixed λ and m . Finally, H values become more stable when both

m and n get larger. Table 1 is useful for determining control limits H in many cases in practice. For other choices of parameters, a Fortran program for evaluating the control limits is available from the authors upon request.

§3. Numerical Results and Comparison

A control chart is considered to have better performance than its competitors if it has the smaller OOC ARL (ARL_1) value when ARL_0 is the same for all the charts. Next, we study the performance of the EC chart relative to a number of existing distribution-free EWMA charts, that is, the EWMA-WRS^[14], EWMA-EX^[15], EWMA-CvM^[16] and EL^[5] charts. In this comparison, we consider the normal distribution and some non-normal distributions from the location-scale family, such as, the Laplace distribution and the lognormal distribution. For the Laplace distribution (heavy-tailed symmetric), the IC sample is taken from a Laplace(0, 1) distribution which has a mean of 0 and a variance of 2, with the test samples coming from a Laplace(θ , δ) distribution where θ denotes the location and δ denotes the scale, and the process variance is $2\delta^2$. For the lognormal distribution (right skewed), the IC sample is from a lognormal(0, 1) distribution, but the test samples are from a lognormal(θ , δ) distribution with pdf $f(x) = (\delta x \sqrt{2\pi})^{-1} e^{-(\ln x - \theta)^2 / (2\delta^2)}$, $x \in (0, \infty)$. As we mostly consider small shifts, the smoothing parameter $\lambda = 0.05$ is used for fair comparisons. For the sake of brevity, we only consider the shifts of $\theta = 0, 0.25, 0.5, 1$ and $\delta = 1, 1.5, 2$ for $m = 100$, $n = 5$. The ARL_0 values of all the charts are fixed at or close to 500. The simulate results are shown in Table 2. The first row of each cell in Table 2 shows the ARL followed by the corresponding SDRL in parentheses, whereas the second row shows the values of the 5th, 25th, 50th, 75th and 95th percentiles (in this order).

Table 2 reveals that for the EC chart the OOC run length distribution is skewed to the right, the ARL and SDRL values all decrease sharply for an increase in the shift in both parameters but decrease at a faster rate for a shift in the scale parameter for all distributions under consideration. It is observed from Table 2 that for small shifts of $\theta = 0.25, 0.5$ in the location parameter when the scale is IC, the EWMA-WRS chart outperforms the competing charts for all distributions under consideration except for the shift of $\theta = 0.25$ in the Laplace distribution where the EWMA-EX chart performs the best, but these two charts based on the WRS and exceedance statistics respectively are not sensitive to the shifts in the process variance. Note that, when there are only shifts in scale, the ARL_1 values of the EWMA-EX chart are biased for all distributions under consideration. For moderately larger shifts in the location when the variability is under control, the EC chart

Table 2 Comparisons of various EWMA charts for $m = 100$, $n = 5$ and $\lambda = 0.05$

		N(0, 1)				
shift		EWMA	EWMA	EWMA	EWMA	EWMA
θ	δ	WRS	EXCEDANCE	CVM	LEPAGE	CUCCONI
0.00	1.0	499.66(779.56)	501.78(750.78)	503.37(770.95)	505.39(776.02)	508.05(891.34)
		23 69 197 566 2,055	24 72 198 592 2,031	18 84 230 578 1,948	21 85 230 572 1,968	12 56 173 512 2,309
0.25	1.0	75.00(233.14)	130.16(343.08)	161.32(353.50)	206.40(413.38)	173.12(415.62)
		10 18 28 53 234	12 22 38 85 543	8 24 58 149 634	11 33 80 200 797	6 21 54 147 686
0.50	1.0	13.24(7.18)	19.72(24.67)	21.35(32.19)	32.65(53.09)	24.08(54.43)
		9 9 12 15 25	8 11 15 22 42	4 8 13 24 61	5 11 19 35 102	3 7 14 26 73
1.00	1.0	5.93(1.32)	8.06(1.97)	4.15(2.21)	5.17(3.02)	4.00(2.53)
		4 5 6 7 8	6 7 8 9 12	2 3 4 5 8	2 3 5 7 11	1 2 3 5 9
0.00	1.5	217.07(252.62)	657.48(841.67)	33.86(32.24)	17.68(15.38)	9.97(8.54)
		20 57 129 277 719	36 123 338 846 2,403	5 13 25 43 93	4 8 14 22 45	2 5 8 13 25
0.25	1.5	61.71(90.61)	205.80(420.38)	25.49(23.85)	15.07(12.44)	8.93(7.35)
		11 20 35 65 198	18 37 74 180 832	4 11 19 32 69	4 7 12 19 37	2 4 7 11 22
0.50	1.5	17.74(11.45)	35.57(50.95)	14.46(11.76)	10.57(7.73)	6.71(4.97)
		7 11 15 21 38	11 18 25 39 89	3 7 11 19 36	3 5 9 13 25	2 3 5 9 16
1.00	1.5	7.58(2.38)	11.99(4.19)	5.36(3.20)	4.97(2.88)	3.56(2.22)
		5 6 7 9 12	7 9 11 14 20	2 3 5 7 11	2 3 4 6 10	1 2 3 5 8
0.00	2.0	159.96(164.17)	777.10(898.47)	14.03(9.65)	7.00(4.05)	4.19(2.58)
		18 50 106 212 485	45 175 445 1,029 2,681	3 7 12 18 33	2 4 6 9 15	1 2 4 5 9
0.25	2.0	66.117(76.959)	287.37(495.37)	12.94(8.74)	6.71(3.83)	4.05(2.49)
		11 23 41 79 203	23 56 120 292 1,143	3 7 11 17 30	2 4 6 9 14	1 2 4 5 9
0.50	2.0	23.17(16.79)	57.61(71.28)	10.32(6.88)	5.99(3.41)	3.73(2.26)
		8 13 19 28 53	14 24 38 65 158	2 6 9 13 24	2 4 5 8 12	1 2 3 5 8
1.00	2.0	9.54(3.74)	16.92(7.41)	5.83(3.49)	4.29(2.275)	2.85(1.61)
		5 7 9 11 17	9 12 15 20 31	2 3 5 8 12	2 3 4 5 9	1 2 3 4 6
Laplace(0, 1)						
0.00	1.0	505.30(795.43)	500.70(746.54)	491.29(757.48)	500.69(773.41)	501.64(884.18)
		24 69 194 572 2,081	25 71 198 592 2,066	18 82 222 558 1,903	22 87 226 566 1,945	12 56 171 502 2,281
0.25	1.0	107.25(321.37)	84.13(235.19)	194.16(429.35)	305.53(565.16)	254.49(560.03)
		11 20 33 69 383	12 20 33 61 269	8 26 65 177 775	12 42 115 310 1,241	8 27 74 224 1,083
0.50	1.0	16.82(20.74)	17.34(13.55)	29.62(81.39)	74.56(183.57)	52.65(143.20)
		7 10 14 19 35	8 11 15 20 33	4 9 16 30 88	6 14 28 65 276	4 10 20 45 184
1.00	1.0	7.06(1.93)	8.77(2.18)	5.16(3.11)	7.55(6.16)	6.07(4.88)
		5 6 7 8 11	6 7 8 10 13	2 3 4 6 11	2 4 6 9 17	2 3 5 8 14
0.00	1.5	252.55(313.71)	653.76(839.78)	51.45(58.04)	31.73(37.60)	18.10(20.56)
		22 60 141 321 862	35 121 331 838 2,394	6 17 34 64 153	5 12 21 38 93	3 7 12 22 52
0.25	1.5	88.52(156.46)	140.19(313.81)	37.84(42.68)	26.11(30.38)	15.37(17.25)
		12 23 41 86 323	17 32 58 120 504	5 13 25 46 113	5 10 18 31 74	2 6 11 19 43
0.50	1.5	22.06(20.20)	28.41(24.21)	19.18(19.29)	16.48(17.72)	10.60(10.34)
		8 12 17 25 51	11 16 23 33 62	3 8 14 24 52	3 7 12 20 44	2 5 8 13 28
1.00	1.5	8.81(3.12)	12.14(3.98)	6.41(4.16)	6.44(4.20)	4.77(3.36)
		5 7 8 10 15	7 9 11 14 20	2 4 5 8 14	2 4 5 8 14	1 3 4 6 11
0.00	2.0	187.17(204.17)	785.74(926.87)	21.17(17.35)	10.84(7.64)	6.56(4.75)
		19 53 117 245 588	45 170 439 1,039 2,772	4 10 17 27 54	3 6 9 14 25	2 3 5 8 16
0.25	2.0	85.74(117.82)	207.56(379.17)	18.81(15.22)	10.05(6.99)	6.17(4.36)
		12 25 47 97 289	21 46 92 203 759	4 9 15 24 47	3 5 8 13 23	2 3 5 8 14
0.50	2.0	27.52(25.86)	43.31(43.87)	13.81(10.73)	8.46(5.61)	5.42(3.69)
		8 14 21 32 68	13 22 33 51 106	3 7 11 18 33	3 5 7 11 19	1 3 5 7 12
1.00	2.0	10.74(4.57)	16.10(6.52)	6.77(4.26)	5.28(3.02)	3.69(2.27)
		5 8 10 13 19	8 12 15 19 28	2 4 6 9 15	2 3 5 7 11	1 2 3 5 8
Lognormal(0, 1)						
0.00	1.0	501.14(796.91)	501.15(748.59)	494.29(761.00)	507.09(772.12)	510.56(901.53)
		23 67 187 560 2,103	24 71 200 596 2,025	18 82 224 560 1,896	21 87 231 573 1,951	12 56 170 508 2,333
0.25	1.0	75.45(239.24)	125.11(335.42)	155.76(342.28)	201.53(398.01)	173.41(424.59)
		10 18 28 52 227	12 22 39 85 504	8 24 57 145 598	10 32 78 203 768	6 22 54 146 678
0.50	1.0	13.28(7.63)	19.49(21.60)	21.22(32.35)	32.81(64.82)	24.22(56.47)
		7 9 12 15 25	8 11 15 22 42	4 8 13 24 61	5 11 19 35 100	3 7 14 26 73
1.00	1.0	5.94(1.33)	8.07(1.95)	4.16(2.22)	5.17(3.04)	3.98(2.55)
		4 5 6 7 8	6 7 8 9 12	2 3 4 5 8	2 3 5 6 11	1 2 3 5 9
0.00	1.5	216.54(252.96)	661.71(845.46)	34.07(33.14)	17.52(14.98)	10.06(8.57)
		20 57 129 280 705	36 124 339 848 2,414	5 14 24 44 94	4 8 13 22 45	2 5 8 13 26
0.25	1.5	62.78(99.08)	205.46(422.85)	25.67(23.18)	15.11(12.32)	8.95(7.39)
		11 21 35 65 202	18 37 73 178 847	4 11 19 33 70	4 7 12 19 38	2 4 7 11 22
0.50	1.5	17.85(11.70)	35.27(46.77)	14.33(11.55)	10.52(7.58)	6.76(4.99)
		7 11 15 21 38	11 18 25 39 86	3 7 11 18 36	3 5 9 13 25	2 3 6 9 16
1.00	1.5	7.58(2.36)	11.98(4.17)	5.36(3.24)	5.00(2.90)	3.52(2.20)
		5 6 7 9 12	7 9 11 14 20	2 3 5 7 11	2 3 4 6 11	1 2 3 5 8
0.00	2.0	158.14(161.36)	797.50(932.99)	14.15(9.73)	7.01(4.05)	4.18(2.59)
		18 49 105 209 477	47 175 455 1,046 2,774	3 7 12 18 33	2 4 6 9 15	1 2 4 5 9
0.25	2.0	66.64(77.12)	284.68(484.27)	13.02(8.86)	6.69(3.85)	4.09(2.51)
		11 23 42 79 206	23 56 120 295 1,108	3 7 11 17 30	2 4 6 8 14	1 2 4 5 9
0.50	2.0	23.26(17.66)	57.77(77.74)	10.31(6.84)	5.99(3.36)	3.725(2.244)
		8 13 19 28 54	14 25 39 65 158	3 6 9 13 23	2 4 5 8 12	1 2 3 5 8
1.00	2.0	9.51(3.73)	16.91(7.46)	5.83(3.52)	4.27(2.25)	2.88(1.64)
		5 7 9 11 16	9 12 15 20 31	2 3 5 8 12	2 3 4 5 9	1 2 3 4 6

performs the best under the normal and lognormal distribution, whereas for the Laplace distribution, the EWMA-CvM chart performs slightly better. When there is a shift in scale and in both the location and scale, irrespective of the magnitude of the shifts in the process parameters and the distributions under consideration, it is once again the EC chart outperforms the other charts. For example, from Table 2, we see that for a 50% increase in the scale parameter (i.e. $\delta = 1.5$) when the location is IC, the ARL_1 values under the normal distribution are about 57%, 93%, 97% and 98% reduction in the EWMA-WRS, EWMA-CvM, EL and EC charts respectively, whereas for the EWMA-EX chart, the ARL_1 value in this case is 657.48 (biased). As a consequence, the EC chart is far superior than the competing charts for detecting a shift in scale with/without change in location.

§4. Effect of the Reference Sample

The proposed control chart is calibrated so that the unconditional ARL_0 is approximately equal to a target, for example, 500. However, practitioners may be more interested in the conditional ARL_0 depending on the Phase I sample they have. The attained (conditional) ARL_0 will vary significantly although the chart may have been designed for a nominal ARL_0 and the Phase I sample comes from an IC process. To investigate this question for our nonparametric chart, we consider the IC situation and suppose that the IC underly true process distribution is standard normal, we obtain 50,000 different Phase I samples each of size m from it, $m = 100, 150, 200, 300, 400, 500$ and 600 , respectively, with $n = 5$, and H determined as earlier to achieve the nominal ARL_0 of 500. From these, the conditional ARL_0 and SDRL of the EC chart with $\lambda = 0.05$ are calculated.

As expected, although the reference samples came from an IC process, they vary remarkably, and their differences are expressed in their means and variances. Following the same idea in the work of [10], we classify the values of the sample means and standard deviations into seven categories with respect to their sampling distributions and various percentiles respectively. The seven classifications are as follows: (i) large downward bias (less than or equal to 5th percentile); (ii) moderately large downward bias (5th–25th percentile); (iii) relatively small downward bias (25th–45th percentile); (iv) marginal or almost no bias (45th–55th percentile); (v) relatively small upward bias (55th–75th percentile); (vi) moderately large upward bias (75th–95th percentile); (vii) large upward bias (beyond the 95th percentile). Further, the 50,000 replications of the simulation study are categorized in a 7×7 way table depending on the values of the observed sample mean

and SD. The proportion of the observations along with the attained IC ARL and SDRL values are recorded in each of these cross classifications. For brevity, the results only for $m = 100$ and 150 are shown in Table 3. For each cell in Table 3, the first value shows the proportion, the middle value show the conditional ARL_0 and the last value (within the brackets) shows the conditional SDRL.

In general, we find that for a given range of values of the reference sample SD, as the reference sample mean increases, the conditional ARL_0 initially increases, reaches a peak somewhere between the 25th and the 75th percentiles of its distribution and then decreases. On the other hand, given a range of values of the reference sample mean, the conditional ARL_0 increases with an increase in its sample SD. From Table 3, it is observed that for $m = 100$, if the reference sample mean lies between the 5th and the 95th percentiles and the observed SD lies within the 45th and the 75th percentiles of their respective sampling distributions, the conditional ARL_0 belongs to $[250, 900]$, this closed interval can provide fair protection against the possibility of an early false alarm without much loss on the sensitivity of detecting out-of-control sustained shifts. If the reference sample mean lies below the 25th percentile or beyond the upper 5th percentile and the observed SD lies between the 75th and the 95th percentiles, then once again the conditional ARL_0 lies within $[250, 900]$. Moreover, for $m = 100$, the first three rows of Table 3 show that the conditional ARL_0 is less than 250, especially in the first row, that is, the observed SD lies below the 5th percentile irrespective of the values of the reference sample mean, the conditional ARL_0 is less than 100, this is a serious threat of an early false alarm. By contrast, the conditional ARL_0 values come from the last two rows of Table 3 for $m = 100$, most of them are greater than 900, even in three cases, the ARL_0 values exceed 2,000, this indicates that these combinations of sample mean and SD might lead to delayed or no detection of a true shift.

From Table 3, we see that for $m = 150$, there are expansions in the region of the conditional ARL_0 within $[250, 900]$, and the conditional ARL_0 values that fall below the target unconditional ARL_0 (i.e. 500) increases, whereas the conditional ARL_0 values that fall above the unconditional ARL_0 decreases with respect to the values in the corresponding cells for $m = 100$. These indicate that larger reference samples give a better protection to both high false alarm rate and delayed detection of a shift in practice. The attained ARL_0 values for $m = 100, 150, 200, 300, 400, 500$ and 600, are summarized in Figure 1 and Table 4 for a quick comparison over all the various classifications. The boxplots in Figure 1 show that the skewness of the conditional IC ARL distribution and how that is diminished for larger values of m . For instance, nearly 33%, 47%, 57%, 58%, 69%,

Table 3 Conditional ARL₀, SDRL and the proportion of samples belonging to various categories of the distribution of sample mean and variance defined by percentiles for $m = 100$ and $m = 150$

sample SD percentiles	sample mean percentiles							
	$m = 100$							
	0-5	5-25	25-45	45-55	55-75	75-95	95-100	
0-5	0.0026 30.89(29.96) 0.0099	0.0103 39.69(40.07) 0.0404	0.0101 48.35(46.41) 0.0399	0.0048 52.86(52.42) 0.0206	0.0098 49.72(50.48) 0.0404	0.0100 42.51(40.30) 0.0391	0.0024 30.47(26.57) 0.0098	
5-25	57.28(58.31) 0.0098	94.38(101.58) 0.0397	115.55(129.72) 0.0400	120.20(126.97) 0.0202	117.98(133.15) 0.0402	95.62(104.07) 0.0400	60.09(63.56) 0.0101	
25-45	99.36(104.08) 0.0048	178.80(194.24) 0.0194	246.89(287.69) 0.0194	235.40(278.21) 0.0102	231.94(264.43) 0.0203	189.74(209.34) 0.0206	102.06(114.48) 0.0053	
45-55	135.23(146.33) 0.0102	285.50(329.70) 0.0395	375.64(414.68) 0.0415	362.09(424.54) 0.0194	372.44(427.02) 0.0388	283.39(310.96) 0.0409	132.66(137.15) 0.0097	
55-75	176.67(196.49) 0.0103	446.18(538.94) 0.0400	604.45(712.44) 0.0397	643.66(781.20) 0.0196	606.59(723.08) 0.0405	442.26(543.68) 0.0395	197.39(213.17) 0.0105	
75-95	375.41(506.53) 0.0024	898.56(1,016.29) 0.0107	1,222.04(1,309.62) 0.0094	1,355.28(1,340.66) 0.0052	1,265.92(1,345.35) 0.0100	926.42(1,078.87) 0.0099	402.65(524.17) 0.0023	
95-100	951.03(1,146.04)	1,761.62(1,663.18)	2,419.92(1,870.04)	2,401.64(1,930.51)	2,413.16(1,811.35)	1,948.38(1,677.85)	852.67(1,150.96)	
$m = 150$								
0-5	0.0025 47.41(45.46) 0.0099	0.0099 64.79(75.05) 0.0396	0.0107 71.77(75.33) 0.0398	0.0053 66.76(74.07) 0.0196	0.0087 69.84(65.41) 0.0411	0.0103 57.79(52.82) 0.0403	0.0025 45.21(47.29) 0.0098	
5-25	86.40(92.37) 0.0095	126.94(136.50) 0.0396	139.74(154.36) 0.0397	151.99(161.21) 0.0203	153.26(170.53) 0.0402	129.62(135.50) 0.0407	84.30(87.09) 0.0101	
25-45	138.28(144.01) 0.0055	216.03(235.32) 0.0204	268.69(300.37) 0.0204	274.59(290.29) 0.0102	263.38(279.46) 0.0194	229.91(255.42) 0.0193	137.29(141.65) 0.0048	
45-55	182.53(186.17) 0.0099	287.84(315.42) 0.0399	397.93(429.35) 0.0403	386.57(413.01) 0.0207	358.16(395.16) 0.04	311.44(376.82) 0.0388	196.15(211.75) 0.0104	
55-75	261.56(303.68) 0.0102	450.44(489.22) 0.0404	577.82(644.64) 0.0395	610.20(681.13) 0.0195	585.22(637.91) 0.0405	444.10(497.19) 0.04	263.36(287.43) 0.0099	
75-95	438.55(467.87) 0.0025	867.70(975.44) 0.0101	1,099.23(1,188.08) 0.0097	1,089.46(1,170.96) 0.0044	1,103.33(1,180.25) 0.0101	899.90(987.88) 0.0105	471.62(552.84) 0.0026	
95-100	1,000.72(1,056.10)	1,840.03(1,595.97)	2,067.26(1,700.54)	2,282.99(1,710.92)	2,203.04(1,726.19)	1,668.97(1,601.48)	965.12(1,020.68)	

80% and 84% of the conditional ARL_0 values belong to the “safe” interval $[250, 900]$ for $m = 100, 150, 200, 300, 400, 500$ and 600 , respectively. Table 4 shows the summary statistics, it is also seen that the standard deviation of the conditional ARL_0 decreases for an increase in the reference sample size m , this indicates that the variation among the reference samples gets smaller as m increases.

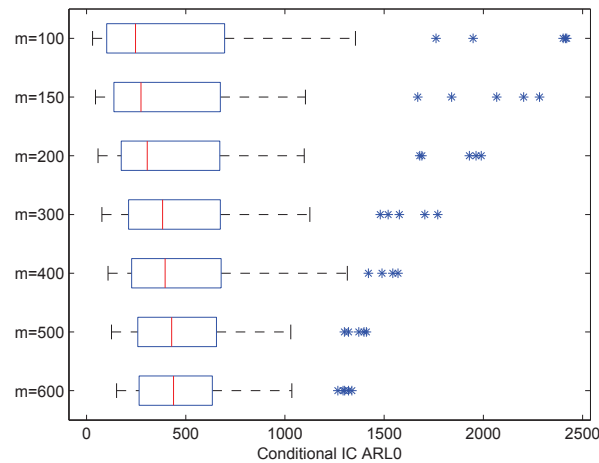


Figure 1 Boxplots for the attained IC run length distribution of the EC chart

Table 4 Summary statistics for the conditional IC ARL distribution of the EC chart for 50,000 different reference samples for $n = 5$

m	Mean	SD	Minimum	1st Quartile	Median	3rd Quartile	Maximum
100	540.79	663.14	30.47	102.06	246.89	643.66	2,419.92
150	533.37	592.79	45.21	138.28	274.59	610.20	2,282.99
200	536.80	532.21	58.12	177.08	305.89	608.97	1,988.95
300	534.59	456.75	77.69	212.38	382.91	621.46	1,770.23
400	521.57	400.40	108.17	228.93	396.06	642.02	1,569.65
500	526.87	364.32	126.01	258.19	429.03	620.35	1,410.42
600	524.74	342.92	151.42	264.67	438.58	597.87	1,337.11

As a consequence, a larger number of the reference sample is better in the sense that it produces less extreme runs for a given control limit. However, it could be difficult to obtain the Phase I data sufficiently due to practical limitations. At a minimum, 150–200 observations are recommended to reduce the between-practitioner (or sampling) variability. Note that, even with the use of $m = 600$, the standard deviation of the conditional ARL_0 is about 340, still reflecting a significant amount of variation, hence an alternative

method for controlling the chart performance is highly warranted, otherwise, there might be many false alarms or unnecessary delay in detecting the shifts in the mean and/or variance. Recently, a design procedure based on the bootstrap is proposed, which guarantees, with high probability, a certain conditional performance for control charts. This is currently being investigated and will be reported in a separate paper.

§5. Real Data Application

In this section, to illustrate the implementation of the EC chart, we use two examples from [17]. Our first example is monitoring the flow width on a hard-bake process. We consider another example based on the familiar piston ring data.

Example 1 We demonstrate the EC chart by a real data set from [17; Table 6.1 and 6.2] on a hard-bake process, which is used in conjunction with photolithography in semiconductor manufacturing. When using such a process, it is desirable to monitor for a change in the location or scale of the flow width attached to this process. Several wafers are taken from the process at one hour intervals and the flow width of each is recorded, measured in microns. Table 6.1 of [17] presents 25 samples, and each sample consists of 5 wafers. Pointed out by Montgomery^[17], the traditional Shewhart \bar{X} and R charts provide no indication of an OOC signal, so we may consider these 125 observations as a set of reference data, and hence the reference sample size is $m = 125$. Table 6.2 of [17] contains another 20 samples, and each sample consists of 5 flow width measurements, hence the test sample size is $n = 5$. In order to apply the EC chart for $m = 125$, $n = 5$, we find by simulation, the control limit H values to be 1.325, 1.605 and 2.119, respectively for $\lambda = 0.05, 0.1$ and 0.2 when $ARL_0 = 500$. The 20 plotting statistics are shown in Figure 2 for the three choices of λ . We can see that the EC charts for the three choice of λ all signal for the first time at test sample 16. In fact, the OOC signal persists in all the test samples from sample 16 onwards until sample 20, indicating possibly a shift in location, or scale, or both. Note that, it is observed from [17], only \bar{X} chart first signals at the 18th sample, hence the EC chart signals two samples earlier.

Example 2 The EC chart is applied with a real data set from [17; Table 6.3 and 6E.7] on the inside diameter measurements of forged automobile engine piston rings. The goal is to establish a statistical control of the inside diameters of the rings manufactured by this process. Table 6.3 of [17] presents 25 samples, and each sample consists of 5 piston rings. Pointed out by Montgomery^[17], the traditional Shewhart \bar{X} and R charts provide no indication of an OOC signal, so we may consider these 125 samples as a set of reference data,

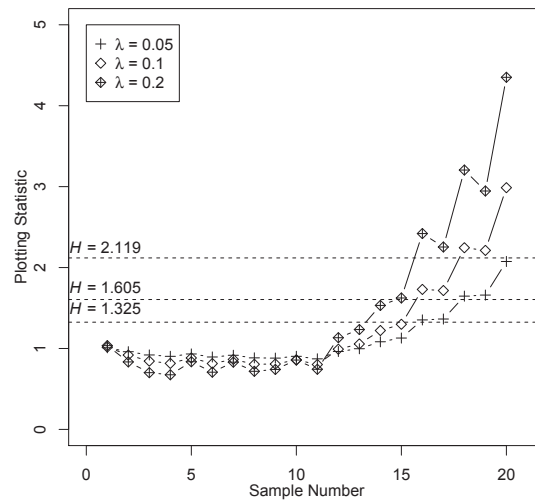


Figure 2 Phase II EC chart for the hard-bake process data

and hence the reference sample size is $m = 125$. Table 6E.7 of [17] contains another 15 samples, and each sample consists of 5 piston rings. Hence the test sample size is $n = 5$. For our EC chart, the parameters are $m = 125$, $n = 5$ and $ARL_0 = 500$. The control limits are searched by simulation and found to be 1.325, 1.605 and 2.119, respectively for $\lambda = 0.05, 0.1$ and 0.2. The 15 charting statistics are shown in Figure 3 for the three choices of λ . We can see that the EC charts for the three choice of λ all signal for the first time at test sample 12. Moreover, all the following EC statistics after the 12th are above the control limit line. Note that, Zhang et al. [16] showed that the EWMA-CvM chart signals at the 14th observation. This, again, shows that our EC chart is quite a useful tool for practitioners.

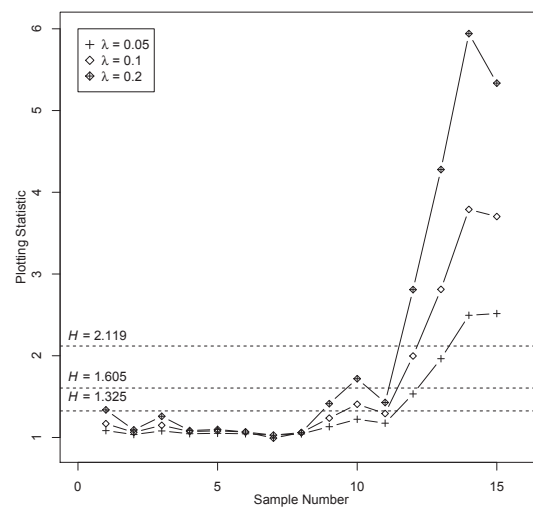


Figure 3 Phase II EC chart for the piston ring data

§6. Conclusions and Considerations

In this paper, we consider a single distribution-free control chart for monitoring of the location and scale parameters. The scheme integrates the EWMA procedure with the Cucconi statistic, and is referred as the EC chart. Being nonparametric, the IC performance of the chart remains the same for all univariate continuous distributions. A performance comparison of the EC chart is done with several existing EWMA distribution-free charts for a class of location-scale models by an extensive simulation study. Our results show that the EC chart performs better than its competitors for detecting shifts in many cases, particularly for the detection of the shifts in scale. Finally, according for the between-practitioner variability led to some quite different conclusions with respect to the chart performance, we study the effect of the Phase I sample on the attained IC run length distribution in terms of IC ARL and SDR. It is seen that a large amount of reference data are required in order to have consistent chart performance among practitioners, at least 150-200 observations are recommended. However, in most applications, it will not be realistic to obtain sufficient reference data from an IC process, thus an alternative design criterion to investigate this question warrants further research.

References

- [1] MAROZZI M. Nonparametric simultaneous tests for location and scale testing: a comparison of several methods [J]. *Comm Statist Simulation Comput*, 2013, **42(6)**: 1298–1317.
- [2] LEPAGE Y. A combination of Wilcoxon's and Ansari-Bradley's statistics [J]. *Biometrika*, 1971, **58(1)**: 213–217.
- [3] MUKHERJEE A, CHAKRABORTI S. A distribution-free control chart for the joint monitoring of location and scale [J]. *Qual Reliab Eng Int*, 2012, **28(3)**: 335–352.
- [4] CHOWDHURY S, MUKHERJEE A, CHAKRABORTI S. Distribution-free Phase II CUSUM control chart for joint monitoring of location and scale [J]. *Qual Reliab Eng Int*, 2015, **31(1)**: 135–151.
- [5] MUKHERJEE A. Distribution-free phase-II exponentially weighted moving average schemes for joint monitoring of location and scale based on subgroup samples [J]. *Int J Adv Manuf Tech*, 2017, **92(1-4)**: 101–116.
- [6] MUKHERJEE A, SEN R. Optimal design of Shewhart-Lepage type schemes and its application in monitoring service quality [J]. *European J Oper Res*, 2018, **266(1)**: 147–167.
- [7] SONG Z, MUKHERJEE A, LIU Y C, ZHANG J J. Optimizing joint location-scale monitoring – an adaptive distribution-free approach with minimal loss of information [J]. *European J Oper Res*, 2019, **274(3)**: 1019–1036.
- [8] MAROZZI M. The multisample Cucconi test [J]. *Stat Methods Appl*, 2014, **23(2)**: 209–227.
- [9] CUCCONI O. Un nuovo test non parametrico per il confronto tra due gruppi campionari [J]. *Giorn Econ*, 1968, **27**: 225–248. (in Italian)

- [10] CHOWDHURY S, MUKHERJEE A, CHAKRABORTI S. A new distribution-free control chart for joint monitoring of unknown location and scale parameters of continuous distributions [J]. *Qual Reliab Eng Int*, 2014, **30(2)**: 191–204.
- [11] MUKHERJEE A. Recent developments in Phase-II monitoring of location and scale – an overview and some new results [R]. Marrakesh, Morocco: 61st ISI World Statistics Congress, 2017.
- [12] GRAHAM M A, CHAKRABORTI S, MUKHERJEE A. Design and implementation of CUSUM exceedance control charts for unknown location [J]. *Int J Prod Res*, 2014, **52(18)**: 5546–5564.
- [13] GRAHAM M A, MUKHERJEE A, CHAKRABORTI S. Design and implementation issues for a class of distribution-free Phase II EWMA exceedance control charts [J]. *Int J Prod Res*, 2017, **55(8)**: 2397–2430.
- [14] LI S Y, TANG L C, NG S H. Nonparametric CUSUM and EWMA control charts for detecting mean shifts [J]. *J Qual Technol*, 2010, **42(2)**: 209–226.
- [15] GRAHAM M A, MUKHERJEE A, CHAKRABORTI S. Distribution-free exponentially weighted moving average control charts for monitoring unknown location [J]. *Comput Statist Data Anal*, 2012, **56(8)**: 2539–2561.
- [16] ZHANG J J, LI E J, LI Z H. A Cramér-von Mises test-based distribution-free control chart for joint monitoring of location and scale [J]. *Comput Ind Eng*, 2017, **110**: 484–497.
- [17] MONTGOMERY D C. *Introduction to Statistical Quality Control* [M]. 6th ed. New York: Wiley, 2009.

阶段 II 联合检测过程位置和尺度的非参数 EWMA 控制图

宋 贇^{1,2} 刘艳春³ 陶桂洪¹

(¹沈阳农业大学理学院, 辽宁, 110866; ²辽宁大学数学院, 辽宁, 110036)

(³辽宁大学商学院, 辽宁, 110036)

摘 要: 本文提出了一个基于 Cucconi 检验的非参数指数加权移动平均 (EWMA) 控制图 (简称为 EC 图) 来同时检测过程位置参数和尺度参数. 依据步长分布的均值、方差及分位数, 给出了 EC 图与其他一些现有的非参数 EWMA 控制图的模拟比较. 基于蒙特卡洛的模拟结果表明, EC 图具有很好的性能. 详细分析了阶段 I 中参考样本大小对 EC 图受控性能的影响. 最后用一个实例来说明 EC 图的实际应用.

关键词: Cucconi 检验; 指数加权移动平均; 非参数; 步长分布

中图分类号: O212.7