A Cramér-von Mises Test-based Distribution-free Control Chart for Joint Monitoring of Location and Scale

Jiujun Zhang [a], Erjie Li [a], Zhonghua Li [b,*]

[a] *Department of Mathematics, Liaoning University, Shenyang 110036, P.R.China*

[b] *Institute of Statistics and LPMC, Nankai University, Tianjin 300071, P.R.China*

Abstract

This paper proposes a new distribution-free control chart by integrating the powerful nonparametric two-sample Cramér-von Mises test and the exponentially weighted moving average control scheme to on-line monitoring. The proposed control chart can be used to monitor the location and the scale parameters of a univariate continuous distribution, simultaneously. The control limits based on Monte-Carlo simulation are provided in a table. The sensitivity analysis of effect of the number of reference samples on the control chart is studied in detail. Comparison results based on Monte-Carlo simulation show that the proposed chart is quite robust to non-normally distributed data, and moreover, it shows satisfactory performance in detecting various process shifts in terms of the average run length and standard deviation of run length. The application of our proposed chart is illustrated by a real data example for automobile engine piston rings.

Key words: Exponentially weighted moving average; Empirical cumulative distribution function; Cramér-von Mises test; Nonparametric; Statistical process control.

1 Introduction

Statistical process control (SPC) charts are one of the widely used tools for keeping the quality of products in a stable level and preventing lots of products

* Corresponding author.
Email address: zli@nankai.edu.cn Tel:86-22-23503423 Fax: 86-22-23501103 (Zhonghua Li).

with low quality by monitoring and analyzing the production process, controlling the variability of the process and detecting abnormal causes in time. As the rapid development of economy world widely, SPC charts have got great advancement and been widely applied in many areas (Megahed et al., 2011; Woodall, 2006). The study of SPC charts originates from Shewhart (1925), who proposes the Shewhart-$\bar{X}$ control chart based on 3σ-rule. It is known in the literature that Shewhart control charts are more effective for monitoring large shifts, as they only make use of the current observations and totally ignore the historical samples. To monitor small to moderate shifts efficiently, the cumulative sum (CUSUM) charts (Page, 1954) and the exponentially weighted moving average (EWMA) charts (Roberts, 1959) are proposed. The study of SPC charts and their applications is quite plentiful now, such as text books of Montgomery (2013) and Qiu (2014). One widely accepted criterion for evaluating the performance of control charts is the average run length (ARL), which is the average number of observations needed for the procedure to signal a change in the measurement distribution (Montgomery, 2013). The ARL value of control charts when the process is in-control (IC), denoted as ARL_0, is often controlled at some specific level. Then, the control charts perform better if the out-of-control (OC) ARL, denoted as ARL_1, is shorter, when detecting a given distributional change.

When using control charts, practitioners have to distinguish between Phase I and Phase II techniques and applications (Chakraborti et al., 2008). In Phase I, practitioners use historical samples to evaluate the process stability and to monitor and estimate the IC model parameters. Samples that show any deviations from the IC model should be determined and removed if corresponding assignable causes can be identified, and then, the IC model parameters can be estimated. The causes of the deviation for these samples should be investigated to be avoided in the future. In Phase II, on-line data and samples are used to quickly determine any sustained or non-sustained shifts from the IC model estimated in Phase I. Practitioners use different types of control charts and measures of performance for the two phases. In this paper, we propose a new chart for Phase II monitoring.

Many conventional control charts are based on the assumption that the observations of a process follow a specific probability distribution. In practice, the process observations may, however, not follow the specified distributions. The performance of these control charts based on the specified assumption is often not robust to other distributions and would be highly deteriorated. Moreover, some parameters, such as the mean and variance of the process, need to be estimated before monitoring. Generally, conventional control charts need relatively more reference samples to estimate the IC distribution or the unknown parameters of the distribution. Jensen et al. (2006) give a detailed overview on the effect of parameter estimation on control chart properties. Maravelakis and Castagliola (2009) show that the chart with estimated parameters

needs more reference data on average in order to detect an OC situation than the one with known parameters. The number of reference samples, in many cases, is too small or even none such that it is improbable to obtain accurate estimation of the IC distribution or the unknown parameters. It is urgently necessary to develop appropriate control procedures that do not require the specified distribution assumption in cases when the process distribution is any kind of continuous distribution. To this end, a number of distribution-free or nonparametric SPC (NSPC) charts have been developed.

NSPC charts do not assume a specific distribution of the process, or only assume some general conditions, such as continuous distribution or symmetric distribution. As the distribution is not specified, NSPC charts could maintain the ARL_0 at the nominal level, whatever the IC distribution is. Due to this advantage, NSPC charts show satisfactory robust performances and gain broad applications. For different purposes, there were developed corresponding NSPC charts, such as those for location (Chakraborti and Eryilmaz, 2007; Chakraborti and van de Wiel, 2008; Chakraborti et al., 2009; Hawkins and Deng, 2010; Li et al., 2010; Qiu and Li, 2011a,b; Graham et al., 2011a,b, 2012), those for scale (Jones and Champ, 2010), and those for distribution (Ross and Adams, 2012). For some recent developments, to name a few, Zhou et al. (2014), Ambartsoumian and Jeske (2015), Lu (2015), Liu et al. (2015), Li et al. (2016), Mukherjee (2016), Zhou et al. (2016), Li et al. (2017) and Abid et al. (2017a,b). Chakraborti et al. (2001), Chakraborti and Graham (2007) and Chakraborti et al. (2011) give thorough overviews on existing research in the area of univariate NSPC. Woodall and Montgomery (2014), Chakraborti et al. (2015) and Capizzi (2015) also encourage research of nonparametric methods.

It is worth pointing out that most of the control charts above are designed to detect a single shift in the location or scale of the process only. However, the process location and scale may change simultaneously during the monitoring period. Therefore, it is desirable to construct a control chart that can detect both the changes in the process location and in the process variability, simultaneously. Though it is direct to monitor the shifts of location and variability simultaneously by two separate control charts, it would be quite complicated if shifts of one parameter would affect those of the other parameter. Cheng and Thaga (2006), McCracken and Chakraborti (2013) and Celano et al. (2016) give overviews of the control charts to use only one chart to simultaneously monitor both process location and scale. There exist some attempts for single charts in the literature. Among them, the rank based method for jointly monitoring location and scale parameters proposed by Lepage (1971) is well applied, which integrates the Wilcoxon rank sum test for location and Ansari-Bradley test for scale (Gibbons and Chakraborti, 2011). Mukherjee and Chakraborti (2012) propose a single distribution-free Shewhart Lepage-type chart (denoted as SL). Chowdhury et al. (2014) propose a distribution-free Shewhart Cucconi-

type chart (denoted as SC). Recently, Chong et al. (2017) extend Lepage-type charts to premier charts, and Mukherjee and Marozzi (2017) extend Lepage-type charts to circular-grid charts, respectively. The research, however, is still lacking for nonparametric single charts.

The objective of this paper is to propose a single NSPC chart that can monitor the location and the scale parameters of a univariate continuous distribution, simultaneously. Motivated by the powerful performance of nonparametric two-sample Cramér-von Mises (CvM) test, we propose a single NSPC procedure by integrating EWMA and CvM test (denoted as ECvM). The proposed procedure is compared with some competing methods in the literature, including the SL chart of Mukherjee and Chakraborti (2012) and the SC chart of Chowdhury et al. (2014).

The rest of the paper is organized as follows. Our proposed ECvM control chart is described in Section 2. Then some implementation issues are described in detail in Section 3, including the search algorithm for finding control limits and sensitivity analysis of the effect of number of reference samples on the performance. Studies based on Monte-Carlo simulation to evaluate the performance in comparison with some competing procedures are presented in Section 4. An application on forged automobile engine piston rings is illustrated in Section 5 to demonstrate the application of the ECvM chart. Some remarks conclude this paper in Section 6. The derivation of the formula for ARL is deferred to the Appendix.

2 Proposed NSPC ECvM Chart

Motivated by the discussions above, a new NSPC chart is described here. We first give brief review of the CvM test in Section 2.1, and then present the methodology for constructing our NSPC ECvM chart in Section 2.2.

2.1 Brief review of CvM test

Assume that $\{X_1, X_2, \ldots, X_n\}$ and $\{Y_1, Y_2, \ldots, Y_m\}$ are independent samples from continuous cumulative distribution functions (CDF) $F(x)$ and $G(x) = F(\frac{x-\theta}{\delta})(\theta \in \Theta, \delta > 0)$, respectively, where $F(x)$ is unknown, θ and δ are unknown location parameter and scale parameter, respectively, and Θ is the parameter space. In terms of hypothesis testing, the null hypothesis is that both of the two samples are from $F(x)$, and the alternative hypothesis is that the samples $\{X_1, X_2, \ldots, X_n\}$ are from $F(x)$, while the samples $\{Y_1, Y_2, \ldots, Y_m\}$

are from $G(x)$, i.e., for $i = 1, 2, \ldots, n, j = 1, 2, \ldots, m$

$$H_0 : X_i, Y_j \sim F \leftrightarrow H_1 : X_i \sim F, Y_j \sim G. \tag{1}$$

To make use of CvM test, the empirical CDF (ECDF) are defined as

$$\hat{F}_1(t) = \frac{1}{n}\sum_{j=1}^{n} I(x_j \le t), \quad \hat{F}_2(t) = \frac{1}{m}\sum_{j=1}^{m} I(y_j \le t),$$

where $I(A)$ is the indicator function being 1 if A is true and 0 otherwise.

The two-sample CvM test statistic in Anderson (1962) is defined as the average square of difference of two ECDF based on the two samples, that is

$$W_{m,n} = \frac{mn}{m+n}\int_{-\infty}^{\infty} |\hat{F}_1(x) - \hat{F}_2(x)|^2 dF_{m,n}(x), \tag{2}$$

where $F_{m,n}(x)$ is the ECDF of the joint samples. The CvM test statistic can be computed as

$$W_{m,n} = \frac{mn}{(m+n)^2}\Big(\sum_{i=1}^{n} |\hat{F}_1(x_i) - \hat{F}_2(x_i)|^2 + \sum_{j=1}^{m} |\hat{F}_1(y_j) - \hat{F}_2(y_j)|^2\Big). \tag{3}$$

From the definition (2) or computation (3), larger $W_{m,n}$ leads to rejecting the null hypothesis H_0 in (1).

It is shown by Anderson (1962) that the mean and variance of $W_{m,n}$ are respectively

$$\mu_{W_{m,n}} = \frac{m+n+1}{6(m+n)},$$

and

$$\sigma^2_{W_{m,n}} = \frac{(m+n+1)[(1-\frac{3}{4n})(m+n)^2 + (1-n)(m+n) - n]}{45(m+n)^2 m}.$$

2.2 Construction of ECvM control chart

To construct a control chart based on CvM test, assume that $\{X_1, X_2, \ldots, X_n\}$ are a sequence of independent observations obtained when the process is IC and $Y_i = \{Y_{i1}, Y_{i2}, \ldots, Y_{im}\}$ are the ith sample with sample size m, which are a sequence of independent observations obtained during process monitoring. At time i, we construct our ECvM control chart by the CvM statistic $W_{m,n}$ in (3). It is well known that the EWMA chart (Roberts, 1959) is more effective than

the traditional Shewhart chart for small changes of a process (Graham et al., 2011a, 2012; Li et al., 2010; Zou and Tsung, 2010). We suggest integrating EWMA procedure with CvM statistic at time point i to construct our ECvM control chart.

The procedure steps are summarized as follows.

(1) Collect sample $X = \{X_1, X_2, \ldots, X_n\}$ with sample size n when the process is IC (Phase I).
(2) At time i, obtain the ith sample $Y_i = \{Y_{i1}, Y_{i2} \ldots Y_{im}\}$ with sample size m.
(3) Compute the CvM test statistic $W_{m,n,i}$ based on X and Y_i. Standardize $W_{m,n,i}$ by

$$U_i = \frac{W_{m,n,i} - \mu_{W_{m,n}}}{\sigma_{W_{m,n}}}. \tag{4}$$

(4) Construct EWMA statistic

$$E_i = \lambda U_i + (1 - \lambda) E_{i-1}, \tag{5}$$

where $E_0 = 0$ and the smoothing constant $\lambda \in (0, 1]$.
(5) ECvM control chart issues an OC signal if $E_i > h$, where h is the control limit for a prespecified ARL_0.

Lucas and Saccucci (1990) provide guidelines for selecting λ and it is shown that smaller values of λ should be used to detect smaller shifts and larger values should be used to detect larger shifts. In general, values of λ in the interval $0.05 \leq \lambda \leq 0.25$ work well in practice.

Note that as some existing NSPC procedures (Ross and Adams, 2012; Li et al., 2010), we assume the IC CDF $F(\cdot)$ is unknown. As a matter of fact, we assume that an IC dataset of size n has been collected when the process is IC, and it can be used for estimating certain IC parameters or the IC distribution $F(\cdot)$. The set of IC sample $\{X_1, X_2, \ldots, X_n\}$ is used to represent the IC process distribution $F(\cdot)$. Note that it might be a significant issue to do Phase I analysis efficiently in cases when $F(\cdot)$ is nonparametric and unknown, although in such cases it is not the focus of this paper. Jones et al. (2009) propose a distribution-free method for identifying an IC reference sample and defining the IC state. And it will also be shown in Section 3.2 that the size n of IC sample has great effect on the IC performance on our proposed ECvM chart.

3 Implementation Issues

We describe some implementation issues of our proposed ECvM procedure. In Section 3.1, we use a bisection search algorithm to obtain the control limit h based on Monte-Carlo simulation. Then, in Section 3.2, we study sensitivity analysis of the effect of IC sample size n on the IC performance.

3.1 Finding control limit h

To find the control limit, we derive the formula for ARL (see the Appendix). It is direct to obtain h by solving (A.3), it is, however, quite complicated and inconvenient for practical usage, as it involves solving nonlinear integration equation. Li et al. (2014) give an overview of methods for computing ARL and show that Monte-Carlo method is widely accepted. Next we obtain the control limit h by Monte-Carlo simulation.

The IC sample X with sample size n can be generated from standard normal distribution $N(0,1)$, as well as the sample Y_i with sample size m. Note that our ECvM control chart is based on the CvM statistic, which distribution independence is proved by Anderson (1962) and Ross and Adams (2012). We also assume the independence of samples $\{X_1, X_2, \ldots, X_n\}$ and $Y_i = \{Y_{i1}, Y_{i2}, \ldots, Y_{im}\}$. These ensure that our ECvM control chart is distribution-free, though the EWMA procedure in (5) makes the decisions whether to signal or not dependent. Hence the ARL values are nearly the same for other non-normal continuous distributions. In addition, we have also verified this by simulation results. Numerical computations based on 50,000 runs are used to determine h. The results are displayed in Table 1. As shown in Graham et al. (2014), the run-length distribution is highly right skewed, which is consistent with our findings in Section 4.1. Important information about the performance of a control chart may be overlooked if only ARL is considered. Some researchers have suggested that the chart should be designed so that a nominal median run-length (MRL) is attained. So, in this paper, we also present the control limits of some target MRL in Table 1. We have chosen $\lambda = 0.1$, $n = 30, 50, 100, 150$ and $m = 5, 10, 15, 25$. The h values in Table 1 are for $ARL_0 = 200, 370$ and 500 and $MRL_0 = 200, 370$ and 500, respectively.

From Table 1, for a combination of (n, m), a control limit h can be found such that the ARL_0 or MRL_0 equals to the prespecified one. For example, when 30 reference observations and 10 monitoring observations are available and an ARL_0 of 500 is desired, the h for the ECvM chart is given by 0.415. From Table 1, we see that for any fixed combination of (n, m) values, the higher the nominal ARL_0 or MRL_0 values, the higher the values of h. Further, for fixed

Table 1. The control limits h for various combinations of n, m, ARL_0 and MRL_0.

n	m	ARL_0=200	ARL_0=370	ARL_0=500	MRL_0=200	MRL_0=370	MRL_0=500
30	5	0.391	0.468	0.504	0.576	0.669	0.705
30	10	0.310	0.378	0.415	0.522	0.615	0.645
30	15	0.225	0.287	0.318	0.462	0.543	0.579
30	25	0.080	0.128	0.157	0.373	0.438	0.471
50	5	0.460	0.547	0.587	0.601	0.695	0.746
50	10	0.415	0.499	0.534	0.571	0.672	0.719
50	15	0.352	0.432	0.472	0.542	0.631	0.668
50	25	0.243	0.310	0.343	0.461	0.541	0.597
100	5	0.516	0.613	0.658	0.622	0.727	0.768
100	10	0.510	0.598	0.643	0.615	0.716	0.756
100	15	0.475	0.562	0.607	0.598	0.699	0.737
100	25	0.411	0.496	0.536	0.577	0.652	0.719
150	5	0.538	0.635	0.679	0.643	0.738	0.781
150	10	0.533	0.628	0.675	0.628	0.732	0.772
150	15	0.514	0.610	0.658	0.619	0.721	0.768
150	25	0.475	0.568	0.610	0.596	0.691	0.742

n, the h decreases with the increase in m, and for fixed m, the h increases with the increase in n. Our simulation results (not shown here) show that for given combination of (n, m, ARL_0) or (n, m, MRL_0), nearly the same h values can be obtained if the normal distribution is replaced by other continuous distributions. For example, when $n = 30, m = 5$ and $\lambda = 0.1$, the ARL_0 is 500 with control limit $h = 0.504$ when the distribution is normal. When the distribution is highly non-normal, i.e., $\chi^2(1)$ distribution, the simulated ARL value is 502.3 based on 50,000 Monte Carlo simulations.

For given n, we also study ARL_0 values for different m and h. The results for $n = 30, 50$ and $m = 5, 10, 15, 25$ are shown in Figure 1. From Figure 1, it is easy to find the h values for prespecified ARL_0, such as for $ARL_0 = 250, 370, 500$. It is as expected that ARL_0 values increase as h increases, but the rate of increasing shows different patterns. For large h, ARL_0 values increase as m increases, and the increasing rate is larger if m is larger, especially when $n = 30, m = 25$. For small h, the ARL_0 values will be nearly stable if m is small.

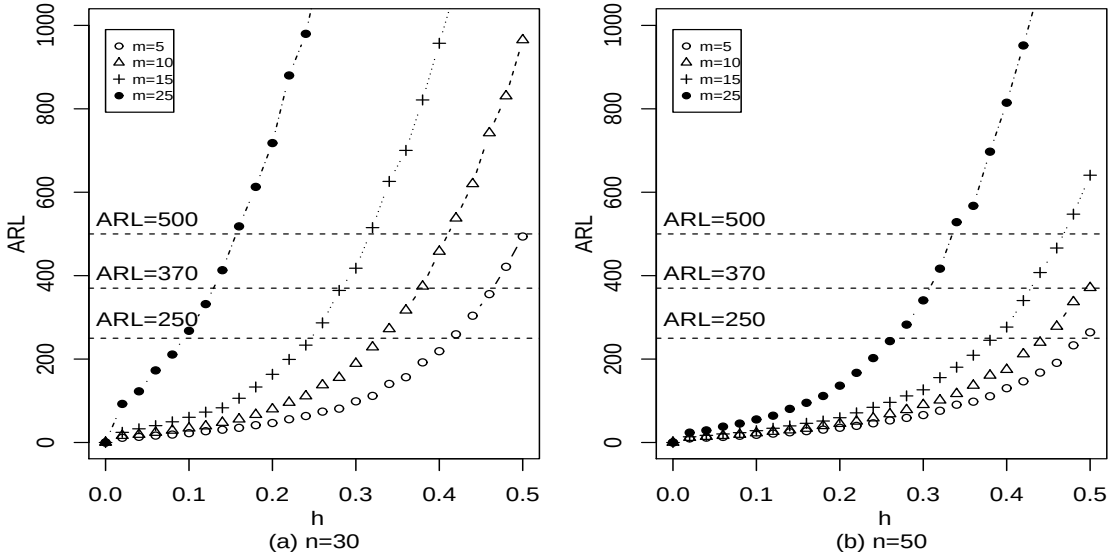

Fig. 1. ARL_0 values for different m and h for given n.

3.2 Sensitivity analysis

As an on-line monitoring NSPC procedure, it is assumed for our proposed ECvM chart that the IC process is represented by an IC reference sample $X = \{X_1, X_2, \ldots, X_n\}$ of size n. The size n would inevitably affect how well the IC process is estimated and how well the ECvM chart performs during monitoring. In practice, we may adjust the control limits such that the ARL_0 fits our target, such as 500. It should be pointed out that when n is not large,

there would be considerable uncertainty in the estimation of the process, which in turn would distort the IC run length distribution of the ECvM control chart.

We study the effect of n by finding the ARL and standard deviation of run length (SDRL) values for given IC reference sample $X = \{X_1, X_2, \ldots, X_n\}$ of size n. Note the standard error of the ARL is another measure on the stability of the estimation of ARL, and it equals to the SDRL divided by the square root of simulation times. Now that the simulation times are fixed in this paper, we only use SDRL instead of the standard error of the ARL. For this purpose, n is set to be 100 or 150, m is set to be 5, λ is set to be 0.1 and ARL_0 is set to be 500, respectively. For these settings, the control limit h can be found from Table 1. Then n IC observations are generated from standard normal distribution $N(0, 1)$, the average μ and standard deviation σ are computed, and ARL_0 values are simulated with 50,000 runs based on these IC samples, rather than the IC distribution. Note that though the IC samples are generated from the IC distribution, there would be deviation of the (μ, σ) values from the target values $(0, 1)$, which makes the ARL_0 values deviate from the target 500. For clear description, we categorize different deviations of the (μ, σ) values into four different cases by direction: (1) μ larger and σ smaller (denoted as $\mu \uparrow \sigma \downarrow$); (2) both μ and σ smaller (denoted as $\mu \downarrow \sigma \downarrow$); (3) both μ and σ larger (denoted as $\mu \uparrow \sigma \uparrow$); (4) μ smaller and σ larger (denoted as $\mu \downarrow \sigma \uparrow$). We also categorize different deviations of the (μ, σ) values into three different cases by magnitude: (1) small deviation $(0\% - 5\%)$; (2) moderate deviation $(5\% - 10\%)$; (3) large deviation $(\geq 10\%)$ and another special case, i.e., small deviation of μ $(< 1\%)$ and nearly none deviation of σ $(< 0.01\%)$.

The estimated μ, σ, ARL and SDRL values are shown in Table 2 for $n = 100$ and in Table 3 for $n = 150$, respectively. From Table 2, 21% of the ARL values are smaller than 250, 13% are larger than 750, 28% are outside the range of 200-1000, 6% are outside the range of 100-2000, 5 cases are less than 100 and 2 cases are larger than 2000. When the deviations of both μ and σ are small, the ARL values will be in the range of 220-750, which indicates that our proposed ECvM chart can control well the error rate when the IC samples are close to the IC distribution. When the deviation of μ is moderate and the deviation of σ is large, the ARL is larger than 1000 only in one case. When the deviation of either is moderate and the other is large, there are 2 cases with ARL larger than 2000 and 7 cases with ARL larger than 1000. The number of cases with ARL smaller than 200 is 23, when the deviations of both μ and σ are larger than 5%. From Table 3, 15% of the ARL values are smaller than 250, 1% are larger than 750, 14% are outside the range of 200-1000, 3 cases are less than 100 and none are larger than 2000. In summary, the deviation effect of σ is larger than μ and the effect is larger when n is smaller.

In some of our numerical examples, some deviations of μ and σ cause the ARL value to increase, sometimes higher than 2000 when ARL_0 is 500. In

Table 2. Effect of IC sample size on ARL and SDRL ($n = 100, m = 5, ARL_0 = 500$).

deviation		$\mu\uparrow$	$\sigma\downarrow$		$\mu\downarrow$	$\sigma\downarrow$		$\mu\uparrow$	$\sigma\uparrow$		$\mu\downarrow$	$\sigma\uparrow$
$\mu = 0$ $\sigma = 1$	μ	σ	ARL (SDRL)	μ	σ	ARL (SDRL)	μ	σ	ARL (SDRL)	μ	σ	ARL (SDRL)
μ(0%-5%)	0.001	0.999	764.93	-0.005	0.998	761.14						
σ(<0.01%)			(758.58)			(761.80)						
μ(0%-5%)	0.003	0.952	219.11									
σ(0%-5%)			(217.49)									
	0.024	0.966	558.72	-0.043	0.987	763.51	0.004	1.025	859.46	-0.048	1.030	895.52
			(558.02)			(773.69)			(856.38)			(913.35)
	0.047	0.987	579.63	-0.016	0.951	558.32	0.029	1.019	621.28	-0.042	1.039	300.23
			(579.20)			(552.03)			(626.92)			(300.90)
	0.021	0.977	473.80	-0.019	0.961	474.97				-0.046	1.029	366.19
			(469.68)			(476.71)						(358.66)
	0.047	0.952	196.53	-0.022	0.988	742.30				-0.390	1.013	256.47
			(192.66)			(740.14)						(252.81)
				-0.021	0.959	215.69						
						(210.21)						
				-0.021	0.975	240.66						
						(243.36)						
				-0.012	0.969	374.89						
						(376.61)						
μ(0%-5%)	0.005	0.938	351.46	-0.014	0.916	453.80	0.007	1.052	680.25	-0.025	1.038	1406.75
σ(5%-10%)			(352.39)			(440.83)			(678.24)			(1396.54)
	0.023	0.935	303.58	-0.027	0.912	194.84						
			(299.07)			(193.49)						
	0.048	0.936	275.44	-0.006	0.901	192.82						
			(272.26)			(186.31)						
				-0.030	0.947	369.98						
						(364.56)						
μ(0%-5%)	0.024	0.888	156.30	-0.034	0.848	161.14				-0.026	1.107	2370.68
σ($\geq$10%)			(152.45)			(153.74)						(2137.34)
	0.042	0.898	104.98	-0.034	0.835	122.03						
			(101.65)			(116.59)						
μ(5%-10%)	0.095	0.990	435.23	-0.052	0.978	397.35	0.085	1.016	300.72	-0.066	1.018	516.66
σ(0%-5%)			(426.65)			(400.80)			(290.30)			(509.40)
	0.052	0.993	526.57	-0.081	1.000	396.12	0.093	1.039	1052.17	-0.057	1.015	287.21
			(519.24)			(395.03)			(1031.2)			(282.45)
	0.072	0.955	298.88	-0.081	0.992	582.08	0.080	1.001	600.83	-0.080	1.008	602.55
			(294.06)			(583.87)			(597.14)			(600.07)
	0.073	0.964	658.21	-0.057	0.973	258.95	0.061	1.079	1863.67	-0.076	1.027	515.73
			(658.56)			(254.82)			(1768.03)			(506.097)
	0.059	0.980	3887.63	-0.051	0.970	463.25	0.052	1.029	540.63			
			(390.13)			(462.78)			(540.75)			
							0.093	1.022	362.24			
									(357.49)			
μ(5%-10%)	0.063	0.934	193.36	-0.058	0.913	216.72	0.065	1.096	761.38	-0.079	1.083	1298.40
σ(5%-10%)			(190.24)			(213.06)			(753.81)			(1309.88)
				-0.100	0.938	193.09				-0.066	1.057	551.36
						(189.60)						(548.36)
										-0.100	1.061	806.66
												(799.72)
μ(5%-10%)	0.066	0.899	251.77	-0.078	0.882	146.17				-0.075	1.101	2302.83
σ($\geq$10%)			(245.16)			(143.17)						(2089.93)

Table 2. Continued.

deviation	$\mu\uparrow$	$\sigma\downarrow$		$\mu\downarrow$	$\sigma\downarrow$		$\mu\uparrow$	$\sigma\uparrow$		$\mu\downarrow$	$\sigma\uparrow$	
$\mu=0$ $\sigma=1$	μ	σ	ARL (SDRL)	μ	σ	ARL (SDRL)	μ	σ	ARL (SDRL)	μ	σ	ARL (SDRL)
$\mu(\geq 10\%)$	0.138	0.973	282.81	-0.136	0.965	132.12	0.185	1.018	255.40	-0.101	1.017	429.89
$\sigma(0\%\text{-}5\%)$			(279.14)			(128.15)			(250.69)			(434.04)
	0.267	0.958	57.24	-0.123	0.956	279.14	0.121	1.026	520.15	-0.137	1.010	706.26
			(52.51)			(279.59)			(520.89)			(702.53)
	0.114	0.978	527.71	-0.135	0.987	383.82	0.108	1.022	497.17			
			(526.79)			(381.93)			(486.52)			
	0.140	0.978	397.58	-0.117	0.971	383.43						
			(394.62)			(384.30)						
	0.139	0.976	373.17	-0.147	0.967	283.61						
			(376.12)			(274.46)						
	0.242	0.996	102.47	-0.127	0.988	533.42						
			(95.36)			(531.53)						
	0.139	0.984	271.27	-0.115	0.994	263.94						
			(269.58)			(260.21)						
	0.105	0.991	647.53	-0.168	0.971	347.36						
			(657.03)			(340.73)						
				-0.132	0.961	435.51						
						(430.08)						
$\mu(\geq 10\%)$	0.150	0.903	101.70	-0.100	0.938	193.09	0.298	1.056	105.44	-0.105	1.067	1368.56
$\sigma(5\%\text{-}10\%)$			(99.33)			(189.60)			(100.25)			(1357.55)
	0.111	0.941	74.13	-0.144	0.921	193.62	0.216	1.089	234.52	-0.105	1.062	720.75
			(68.72)			(187.49)			(232.66)			(719.233)
							0.133	1.098	367.84			
									(368.25)			
$\mu(\geq 10\%)$	0.110	0.893	81.21	-0.213	0.831	60.63						
$\sigma(\geq 10\%)$			(77.94)			(55.51)						
	0.126	0.846	117.77	-0.166	0.875	108.82						
			(114.35)			(103.58)						
				-0.125	0.852	59.17						
						(55.27)						

Table 3. Effect of IC sample size on ARL and SDRL ($n = 150, m = 5, ARL_0 = 500$).

deviation		$\mu\uparrow$	$\sigma\downarrow$		$\mu\downarrow$	$\sigma\downarrow$		$\mu\uparrow$	$\sigma\uparrow$		$\mu\downarrow$	$\sigma\uparrow$	
$\mu = 0$ $\sigma = 1$	μ	σ	ARL (SDRL)	μ	σ	ARL (SDRL)	μ	σ	ARL (SDRL)	μ	σ	ARL (SDRL)
μ(0%-5%)	0.02	1.000	611.22				0.032	1.001	312.35			
σ(<0.01%)			(622.10)						(313.29)			
μ(0%-5%)	0.022	0.956	279.26	-0.015	0.979	773.73	0.050	1.036	758.78	-0.024	1.035	726.64
σ(0%-5%)			(274.28)			(774.12)			(756.12)			(734.15)
	0.010	0.960	378.47	-0.011	0.973	479.36	0.004	1.006	741.12	-0.044	1.031	651.35
			(376.47)			(476.46)			(736.43)			(646.93)
	0.026	0.989	268.75	-0.018	0.988	447.83	0.021	1.034	668.97	-0.017	1.032	518.70
			(261.47)			(443.63)			(652.81)			(515.51)
	0.014	0.995	482.67	-0.048	0.990	377.04	0.050	1.025	414.75	-0.019	1.042	1214.35
			(477.79)			(371.61)			(417.25)			(1217.18)
				-0.042	0.985	621.04	0.041	1.040	923.12	-0.044	1.025	875.05
						(616.11)			(915.70)			(883.72)
				-0.011	0.987	665.18	0.032	1.026	414.97	-0.002	1.035	1120.63
						(665.52)			(421.42)			(1103.08)
				-0.040	0.974	399.81	0.018	1.048	1472.82	-0.010	1.009	624.55
						(392.69)			(1431.63)			(615.92)
				-0.012	0.967	382.12				-0.010	1.033	560.15
						(378.17)						(563.70)
μ(0%-5%)				-0.066	0.905	305.99	0.033	1.093	1050.35	-0.018	1.068	650.06
σ(5%-10%)						(308.02)			(1036.25)			(642.56)
				-0.027	0.923	234.34	0.025	1.052	1164.98			
						(228.37)			(1163.06)			
				-0.038	0.948	542.41	0.007	1.086	1165.26			
						(544.48)			(1161.0)			
				-0.017	0.948	252.08	0.021	1.051	609.50			
						(250.57)			(06.23)			
				-0.003	0.936	537.72						
						(536.26)						
μ(0%-5%)	0.041	0.895	337.86	-0.043	0.872	165.23	0.010	1.157	1685.02			
σ(≥10%)			(333.51)			(161.89)			(1617.81)			
μ(5%-10%)	0.089	0.986	398.04	-0.095	0.958	317.96	0.087	1.013	459.87	-0.076	1.010	371.21
σ(0%-5%)			(400.21)			(315.37)			(455.32)			(370.92)
	0.058	0.974	269.79	-0.071	0.953	282.72	0.086	1.026	411.05	-0.072	1.038	820.79
			(267.88)			(279.88)			(408.13)			(801.60)
	0.058	0.959	348.80	-0.098	0.997	458.88	0.068	1.023	178.33	-0.092	1.007	293.25
			(340.96)			(452.14)			(174.92)			(294.77)
				-0.056	0.971	447.88	0.082	1.044	528.44	-0.058	1.031	684.14
						(445.54)			(519.31)			(675.16)
				-0.079	0.987	486.60	0.058	1.016	836.93	-0.070	1.021	712.17
						(483.19)			(828.03)			(712.37
							0.076	1.013	1061.91			
									(1052.95)			
							0.079	1.028	507.13			
									(499.56)			
μ(5%-10%)	0.067	0.937	128.19	-0.079	0.922	226.60	0.088	1.051	350.96	-0.057	1.064	1117.65
σ(5%-10%)			(126.26)			(220.24)			(346.66)			(1096.14)
	0.078	0.950	295.48	-0.072	0.911	214.18				-0.056	1.070	413.36
			(295.64)			(211.07)						(408.00)
	0.063	0.940	167.18									
			(163.30)									
μ(5%-10%)				-0.079	0.888	147.31						
σ(≥10%)						(141.62)						

Table 3. Continued.

deviation	$\mu\uparrow$	$\sigma\downarrow$		$\mu\downarrow$	$\sigma\downarrow$		$\mu\uparrow$	$\sigma\uparrow$		$\mu\downarrow$	$\sigma\uparrow$	
$\mu = 0$ $\sigma = 1$	μ	σ	ARL (SDRL)	μ	σ	ARL (SDRL)	μ	σ	ARL (SDRL)	μ	σ	ARL (SDRL)
$\mu(\geq 10\%)$	0.137	0.955	174.65	-0.132	0.981	468.46	0.116	1.022	350.78	-0.105	1.034	618.28
$\sigma(0\%\text{-}5\%)$			(168.16)			(470.99)			(351.47)			(624.73)
	0.127	0.972	333.60	-0.110	0.995	531.81	0.183	1.002	93.53	-0.127	1.010	211.33
			(333.76			(522.07)			(87.894)			(204.48)
	0.111	0.960	234.19	-0.104	0.999	503.61	0.142	1.039	243.48	-0.114	1.050	479.40
			(227.43)			(511.23)			(238.57)			(476.80)
	0.112	0.918	255.95	-0.181	0.953	139.50	0.121	1.040	435.68	-0.125	1.007	180.56
			(253.13)			(135.59)			(439.61)			(173.30)
	0.110	0.982	138.35									
			(133.97)									
	0.115	0.990	237.37									
			(235.73)									
$\mu(\geq 10\%)$	0.137	0.967	254.15	-0.168	0.943	112.30	0.120	1.087	1028.50	-0.127	1.073	755.25
$\sigma(5\%\text{-}10\%)$			(250.29)			(107.33)			(1005.08)			(749.22)
	0.273	0.930	57.39	-0.127	0.928	291.79	0.177	1.066	212.45			
			(52.79)			(215.89)			(209.85)			
				-0.137	0.948	208.92						
						(208.44)						
$\mu(\geq 10\%)$	0.237	0.897	59.56	-0.103	0.892	117.91	0.142	1.119	606.35			
$\sigma(\geq 10\%)$			(54.40)			(113.02)			(603.29)			

such cases, it will take a long time to detect such shifts. Comparing Table 2 and Table 3, when n changes from 100 to 150, the number of the cases with $ARL > 2000$ reduced from 2 to 0, which implies that practitioners might need collect at least 150 observations to avoid too long waiting.

When sufficiently large number of IC samples are unavailable, one alternative method is the bootstrap method, such as Chatterjee and Qiu (2009) and Gandy and Kvaløy (2013), which is beyond the scope of this paper.

4 Comparison Studies

We present some simulation results in this section regarding the performance of our proposed ECvM control chart and compare it with some competing methods in the literature. In Section 4.1, we study the performance of our ECvM chart when the process is IC. Then, in Section 4.2, we compare our ECvM chart with some competing methods under different location and/or scale shifts settings. Finally, in Section 4.3, we further compare our ECvM chart with the competing methods under more general shifts settings.

4.1 Performance for IC process

We set $n = 30, 50, 100, 150$, $m = 5, 10, 15, 25$, $\lambda = 0.1$ and $ARL_0 = 500$. The control limits can be found from Table 1. Note our ECvM chart is distribution-free, hence the run length distribution is nearly the same for other non-normal continuous distributions due to simulation variation. Numerical computations based on standard normal samples with 50,000 runs are used to determine the ARL, SDRL and some percentiles of run length. The results are displayed in Table 4.

From Table 4, it can be seen that for fixed m, SDRL and the 95% percentile values decrease as n increases, while, all the other percentiles increase as n increases. On the contrary, for fixed n, the 95% percentiles increase as m increases, while except the 75% percentiles, all other percentiles decrease as m increases. For $n = 30$, SDRL values increase as m increases, while for other values of n, SDRL values also increase as m increases only when $m > 5$. As for the percentiles, for all combinations of (n, m), the medians of the run lengths are all less than the prespecified target $ARL_0 = 500$. For example, when $n = 100$ or 150, the median values are nearly half of the target, and when both n and m are even smaller, the median values are farther from target. For these listed percentiles, the 75% percentiles are the most close to the target when $n \geq 50$. For example, the 75% percentile is 499 when $n = 50$

Table 4. Run length performance of ECvM chart with $ARL_0 = 500$.

n	m	h	ARL	SDRL	5^{th}	25^{th}	$Median$	75^{th}	95^{th}
30	5	0.504	499.41	1124.42	7	37	123	411	2294
30	10	0.415	504.92	1140.95	4	24	100	417	2515
30	15	0.318	500.71	1189.12	3	14	69	369	2540
30	25	0.157	498.71	1312.87	1	5	30	258	2940
50	5	0.587	500.47	973.37	11	58	175	494	2028
50	10	0.534	499.03	965.96	8	43	156	499	2146
50	15	0.472	503.31	1019.03	6	31	129	478	2332
50	25	0.343	502.49	1135.90	3	17	84	421	2571
100	5	0.658	506.26	783.22	18	92	247	595	1824
100	10	0.643	506.92	779.29	16	79	229	605	1932
100	15	0.607	500.88	820.89	11	65	205	578	1986
100	25	0.536	500.45	892.35	8	45	172	552	2131
150	5	0.679	501.45	693.54	20	107	271	616	1723
150	10	0.675	499.53	674.05	18	95	259	626	1798
150	15	0.658	498.40	721.95	17	90	247	604	1824
150	25	0.610	504.55	776.82	12	66	216	630	1987

and $m = 10$. The 95% percentiles are almost 3 to 5 times of the target. In summary, the run length distribution is right skewed, which is consistent with the literature, such as Teoh et al. (2014) and Graham et al. (2014), who show that the run length distribution is highly right skewed, especially when the shift is small.

4.2 Performance for location and/or scale shifts

The SL chart of Mukherjee and Chakraborti (2012) and the SC chart of Chowdhury et al. (2014) are the existing two competing nonparametric control charts, which are designed for monitoring the location and scale, simultaneously. So, they are chosen as the benchmark charts in this paper. Now the proposed ECvM control chart is compared with the SL chart of Mukherjee and Chakraborti (2012) and the SC chart of Chowdhury et al. (2014). As the number and variety of OC settings are too large to allow a comprehensive comparison and our goal is to show the effectiveness and sensitivity of the ECvM chart, we only choose certain representative models for illustration and consider the following four distributions: normal distribution $N(\theta, \delta)$, chi-squared distribution $\chi^2(1)$ with degree-of-freedom 1, Laplace distribution $Laplace(\theta, \delta)$ and lognormal distribution $\mathrm{LN}(\theta, \delta)$. The parameters are set as $n = 30$, $m = 5$, $\lambda = 0.1$ and ARL_0=500. When the process is IC, the location and scale parameters are set as $\theta_0 = 0$ and $\delta_0 = 1$. When the process is OC, the location parameter is assumed to shift from $\theta_0 = 0$ to θ, and the scale parameter is assumed to shift from $\delta_0 = 1$ to δ. We consider 24 combinations of (θ,δ) when θ=0.0, 0.25, 0.5, 1.0, 1.5, 2.0 and δ=1.0, 1.25, 1.5, 1.75, 2.0. The ARL and SDRL results are shown in Table 5 for normal distribution and chi-squared distribution and in Table 6 for Laplace distribution and lognormal distribution, respectively.

From Table 5 and Table 6, several conclusions can be drawn as follows.

(1) For the selected n, m and ARL_0, the ARL_1 values decrease as the location parameter θ and/or the scale parameter δ increases, which indicates that it is effective for our ECvM chart to jointly monitor location and scale. As for the same magnitude, the ECvM chart detects scale shifts faster than location shifts. For example, for normal distribution without shifts in δ, the ARL_1 value is about 42% less when θ_0 increases by 25%, while without shifts in θ, the ARL_1 value is about 77% less when δ_0 increases by 25%. SDRL values show similar findings.

(2) For one thing, when there are only shifts in θ, the ARL_1 values of ECvM chart are smaller than those of SL and SC charts, indicating better performance of ECvM chart. Take the normal distribution as an example, when $\theta = 0.5$, δ=1.00, the ARL_1 value of ECvM chart is 60.49, which

Table 5. ARL(SDRL) values of ECvM, SL and SC charts for $N(0,1)$ and $\chi^2(1)$ distributions when $n = 30$, $m = 5$ and $ARL_0 = 500$.

		$N(0,1)$			$\chi^2(1)$		
θ	δ	ECvM	SL	SC	ECvM	SL	SC
0.00	1.00	499.41	499.92	498.45	500.52	499.37	500.54
		(1124.42)	(946.58)	(1201.4)	(1131.61)	(937.22)	(1117.06)
0.25	1.00	282.20	347.93	338.74	190.99	308.90	1198.64
		(861.97)	(746.76)	(932.54)	(843.46)	(881.83)	(2185.52)
0.50	1.00	60.49	139.36	123.36	13.68	254.96	571.30
		(323.14)	(397.76)	(457.16)	(167.18)	(868.49)	(1518.79)
1.00	1.00	4.13	13.24	11.18	1.87	111.78	104.08
		(4.10)	(51.20)	(38.32)	(1.07)	(592.68)	(571.56)
1.50	1.00	1.92	2.74	2.37	1.29	17.97	17.74
		(0.961)	(4.81)	(3.33)	(0.47)	(216.23)	(191.23)
2.00	1.00	1.33	1.33	1.26	1.08	4.48	3.42
		(0.51)	(0.79)	(0.66)	(0.26)	(97.96)	(40.27)
0.00	1.25	108.47	115.98	71.02	11.67	10.06	6.21
		(247.50)	(216.29)	(165.75)	(9.61)	(10.12)	(6.08)
0.00	1.50	39.39	39.60	21.92	6.59	5.10	3.30
		(67.37)	(63.86)	(33.61)	(4.61)	(4.76)	(2.87)
0.00	1.75	21.74	19.38	10.44	5.11	3.63	2.44
		(25.20)	(25.24)	(14.10)	(3.29)	(3.17)	(1.97)
0.00	2.00	14.52	11.65	6.33	4.39	2.91	2.03
		(13.95)	(14.07)	(7.23)	(2.67)	(2.43)	(1.50)
0.25	1.25	71.73	91.96	55.10	377.10	479.87	649.42
		(168.56)	(177.14)	(127.79)	(1079.46)	(1036.66)	(1453.18)
0.50	1.25	26.03	46.85	30.63	34.65	126.92	333.02
		(68.79)	(102.94)	(94.31)	(279.47)	(379.17)	(990.57)
1.00	1.25	4.78	9.17	7.19	2.21	60.41	60.19
		(5.22)	(17.74)	(11.33)	(5.42)	(307.96)	(339.23)
1.50	1.25	2.28	2.92	2.51	1.39	12.92	10.46
		(1.27)	(3.38)	(2.65)	(0.53)	(145.07)	(101.40)
2.00	1.25	1.55	1.54	1.40	1.11	2.77	2.56
		(0.68)	(1.03)	(0.86)	(0.31)	(43.74)	(29.72)
0.25	1.50	30.68	33.84	19.05	23.63	20.39	10.48
		(49.92)	(54.83)	(29.93)	(24.46)	(22.39)	(11.62)
0.50	1.50	16.40	22.22	13.34	81.12	169.61	203.53
		(22.66)	(35.82)	(20.23)	(394.87)	(441.53)	(622.55)
1.00	1.50	5.16	7.42	5.35	2.71	39.88	43.72
		(4.47)	(10.08)	(6.49)	(2.81)	(159.09)	(200.59)
1.50	1.50	2.62	3.02	2.57	1.53	9.97	8.79
		(1.57)	(3.01)	(2.46)	(0.61)	(53.91)	(53.25)
2.00	1.50	1.77	1.73	1.56	1.16	2.52	2.33
		(0.85)	(1.33)	(1.01)	(0.37)	(16.05)	(16.39)
0.25	2.00	13.32	10.87	6.05	6.53	4.50	2.80
		(13.15)	(12.5)	(6.74)	(4.42)	(4.17)	(2.43)
1.00	2.00	5.43	5.25	3.63	7.00	35.24	32.27
		(4.17)	(5.54)	(3.59)	(24.82)	(86.01)	(92.57)
1.50	2.00	5.43	3.04	2.38	1.95	10.54	9.05
		(4.17)	(2.76)	(2.00)	(0.95)	(27.62)	(44.04)
2.00	2.00	2.23	1.99	1.69	1.32	2.96	2.73
		(1.21)	(1.52)	(1.15)	(0.49)	(9.38)	(10.49)

Table 6. ARL(SDRL) values of ECvM, SL and SC charts for *Laplace* and LN(0,1) distributions when $n = 30$, $m = 5$ and $ARL_0 = 500$.

		Laplace			LN(0,1)		
θ	δ	ECvM	SL	SC	ECvM	SL	SC
0.00	1.00	499.84	503.01	501.10	496.36	498.56	498.56
		(1127.27)	(996.73)	(1124.78)	(1129.80)	(964.78)	(1126.42)
0.25	1.00	229.25	393.73	377.21	203.99	405.75	1057.49
		(755.45)	(865.84)	(969.75)	(949.90)	(981.59)	(2078.40)
0.50	1.00	30.02	200.27	182.11	5.36	258.29	438.22
		(212.58)	(644.24)	(663.26)	(71.66)	(865.90)	(1346.31)
1.00	1.00	2.76	21.62	20.33	1.49	63.00	67.92
		(2.32)	(162.59)	(188.87)	(0.60)	(461.63)	(498.68)
1.50	1.00	1.59	2.76	2.79	1.11	10.28	12.98
		(0.73)	(23.69)	(14.51)	(0.31)	(175.21)	(213.80)
2.00	1.00	1.24	1.24	1.19	1.02	3.39	3.26
		(0.45)	(6.72)	(2.01)	(0.14)	(94.45)	(74.63)
0.00	1.25	153.54	177.51	123.22	24.40	22.26	13.36
		(381.92)	(364.80)	(326.84)	(28.30)	(26.19)	(15.51)
0.00	1.50	64.79	76.64	44.13	9.37	7.62	4.98
		(124.17)	(152.09)	(95.03)	(7.61)	(7.74)	(4.90)
0.00	1.75	35.32	40.02	21.99	6.28	4.67	3.13
		(53.36)	(68.65)	(36.69)	(4.46)	(4.40)	(2.76)
0.00	2.00	24.28	23.44	13.21	5.12	3.48	2.44
		(30.76)	(34.36)	(19.42)	(3.30)	(3.03)	(2.00)
0.25	1.25	84.96	140.74	102.76	321.44	418.97	557.54
		(279.08)	(329.52)	(303.75)	(1000.75)	(954.24)	(1355.11)
0.50	1.25	21.32	74.05	56.67	19.51	155.48	288.25
		(81.25)	(230.66)	(190.61)	(219.63)	(481.70)	(979.85)
1.00	1.25	3.23	11.30	9.43	1.68	48.65	41.62
		(2.68)	(73.32)	(47.23)	(0.69)	(372.54)	(323.78)
1.50	1.25	1.83	2.44	2.21	1.15	7.52	6.82
		(0.93)	(6.05)	(10.41)	(0.36)	(141.15)	(97.28)
2.00	1.25	1.38	1.33	1.26	1.03	3.16	2.60
		(0.56)	(1.12)	(0.99)	(0.16)	(100.50)	(68.72)
0.25	1.50	43.29	61.35	37.13	66.27	69.49	39.14
		(111.69)	(129.35)	(90.57)	(135.16)	(106.91)	(61.62)
0.50	1.50	15.77	36.54	23.98	48.27	157.38	198.25
		(36.51)	(88.67)	(65.96)	(339.73)	(486.25)	(696.83)
1.00	1.50	3.61	8.03	6.45	2.00	32.46	32.92
		(2.91)	(25.36)	(27.58)	(2.29)	(192.87)	(214.48)
1.50	1.50	2.03	2.44	2.20	1.23	6.47	6.06
		(1.09)	(4.77)	(11.78)	(0.43)	(104.99)	(72.72)
2.00	1.50	1.52	1.41	1.33	1.04	1.80	1.89
		(0.68)	(1.02)	(0.85)	(0.20)	(25.31)	(19.42)
0.25	2.00	18.97	20.89	12.11	10.06	7.27	4.43
		(26.08)	(33.04)	(17.71)	(7.97)	(7.36)	(4.38)
1.00	2.00	4.13	5.74	4.14	3.68	28.11	26.02
		(3.16)	(9.41)	(7.52)	(7.62)	(117.32)	(148.17)
1.50	2.00	2.43	2.51	2.12	1.49	6.32	5.84
		(1.41)	(2.88)	(2.39)	(0.59)	(43.39)	(60.16)
2.00	2.00	1.80	1.57	1.42	1.10	1.67	1.85
		(0.88)	(1.16)	(0.96)	(0.30)	(11.48)	(10.13)

is about 56.5% less than 139.36 of SL chart, and about 51% less than 123.36 of SC chart. When θ is larger, the performance of the three charts are similar. For another thing, when there are only shifts in δ, the ARL_1 values of SC chart are smaller than those of ECvM and SL charts. Moreover, if both θ and δ shift, ECvM chart has better performance for most cases, and nearly comparable for the rest cases. Results for chi-squared distribution, Laplace distribution and lognormal distribution show similar findings, except it is worth pointing out that the ARL_1 values of SC chart are biased, that is, ARL_1 is even larger than ARL_0, such as for chi-squared distribution and lognormal distribution when $\theta = 0.25$ and $\delta = 1.00$.

(3) Our ECvM chart is distribution-free in the sense that the ARL_0 values are close to 500 for all these four continuous distributions. It is interesting to study the performance of our ECvM chart when the process is OC under different distributions. Note that normal distribution and Laplace distribution represent light tailed and heavy tailed distributions, respectively, and chi-squared distribution and lognormal distribution represent right skewed distributions. When there are only location shifts, the ARL_1 values for normal distribution are the largest; when there are only scale shifts, the ARL_1 values for Laplace distribution are the largest; when both location and scale shift, in most cases, the ARL_1 performances are comparable, and the ARL_1 values for normal distribution are the smallest when both shifts are small, such as when $(\theta, \delta) = (0.25, 1.25)$. SDRL values show similar findings.

To see more clearly the performance comparison, Figures 2-4 show, respectively, the ARL values of ECvM, SL and SC charts when only the location parameter shifts, when only the scale parameter shifts and when the location parameter shifts with scale parameter shifting to 1.2. From Figure 2, when only the location parameter shifts, for normal distribution $N(\theta, \delta)$, the ARL_1 values of ECvM, SL and SC charts are all smaller than ARL_0=500, and directly at the beginning and for $\theta \geq 1$, the values are nearly the same, while for other θ, ECvM performs better than the other two charts. For chi-squared distribution $\chi^2(1)$, the ARL_1 values of ECvM, SL and SC charts are all larger than ARL_0=500 when θ is small, and the deviation is the largest for SC chart. When $0.1 < \theta < 1.5$, ECvM performs better than the other two charts and when $\theta > 1.5$, three charts are similar. For Laplace distribution, ECvM performs better than the other two charts and three charts have similar performances when $\theta > 1$. For lognormal distribution, the ARL_1 values of the three charts are all larger than ARL_0=500 when $\theta < 0.2$, but for small magnitude shifts, ECvM performs better than the other two charts. From Figure 3, when only the scale parameter shifts, the three charts have similar performances. From Figure 4, when the location parameter shifts with scale parameter shifting to 1.2, for normal distribution and Laplace distribution, SC chart is better when θ is small, while ECvM is better when θ is moderate

to large. For chi-squared distribution and lognormal distribution, SL and SC charts are still biased when θ is small. When θ gets larger, ECvM performs better than the other two charts.

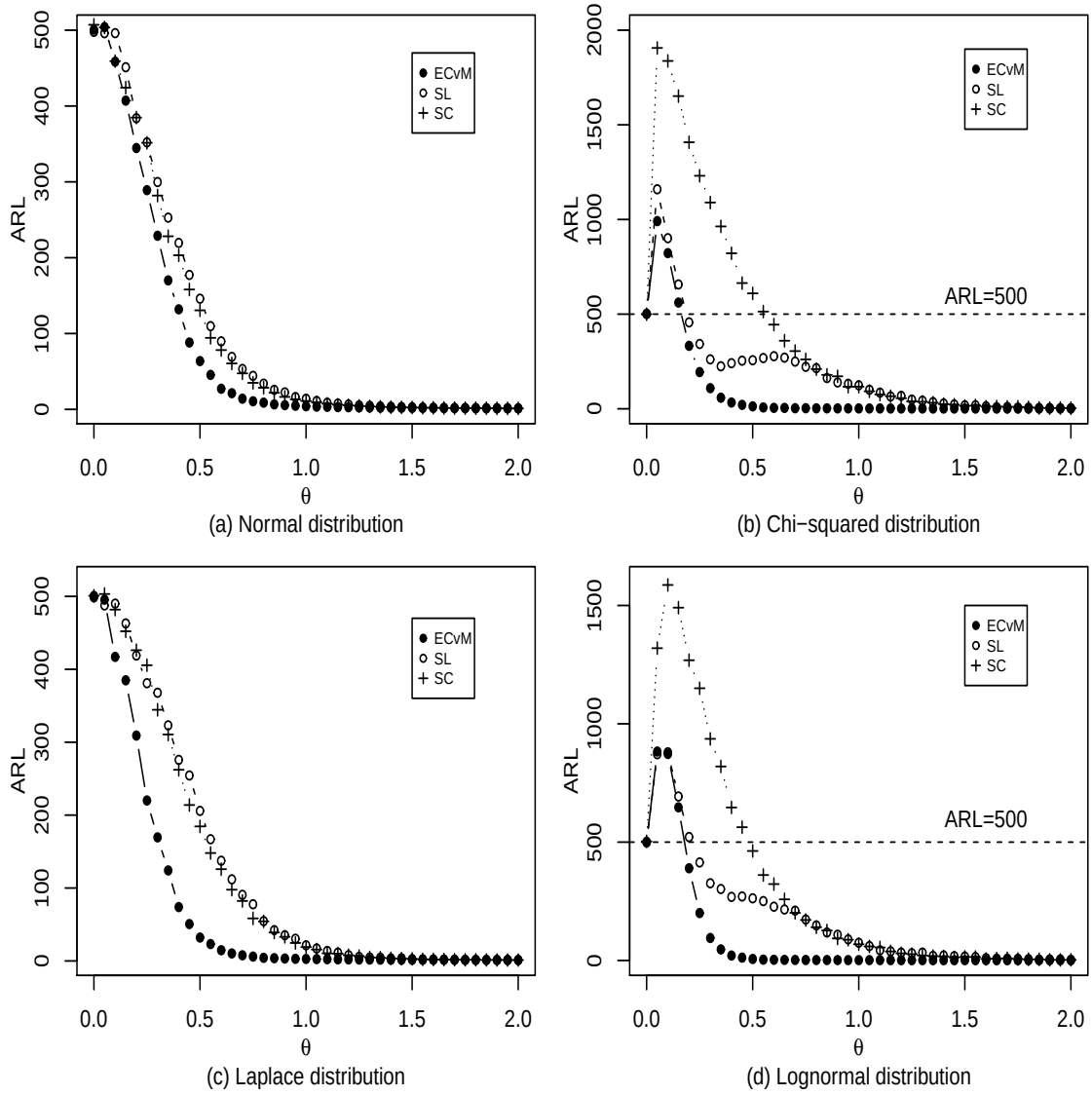

Fig. 2. ARL values of ECvM, SL and SC charts when only the location parameter shifts ($n = 30$, $m = 5$ and $ARL_0 = 500$).

4.3 Performance for general shifts

From the CvM test statistic defined in (2) or (3), which involves ECDF based on the two samples, our proposed ECvM chart is not restricted to detecting location and/or scale shifts, and it is expected to be able to detect more general patterns of changes. We investigate this and consider changes in the parameters of several distributions. The choice of which distributions to con-

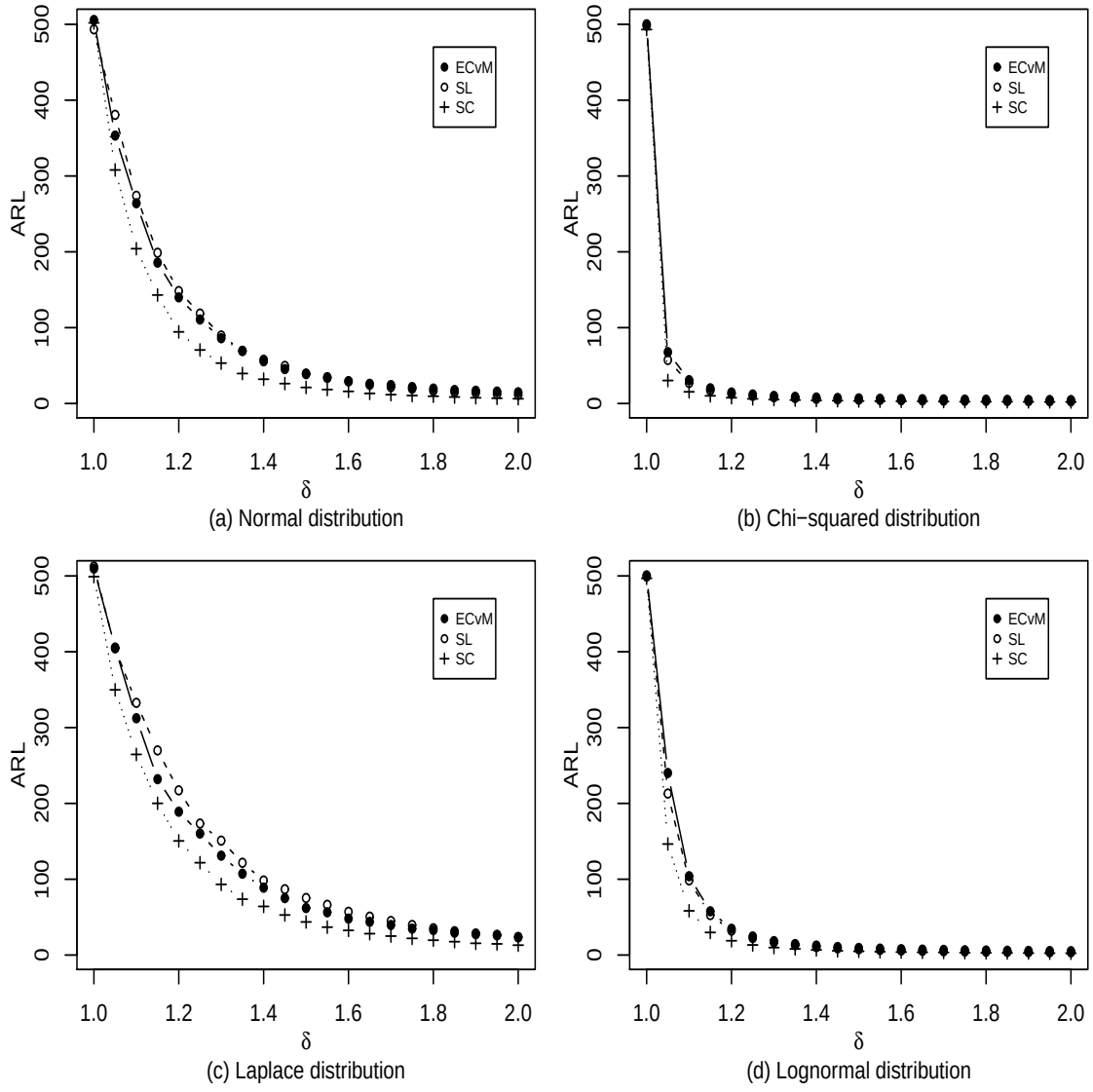

Fig. 3. ARL values of ECvM, SL and SC charts when only the scale parameter shifts ($n = 30$, $m = 5$ and $ARL_0 = 500$).

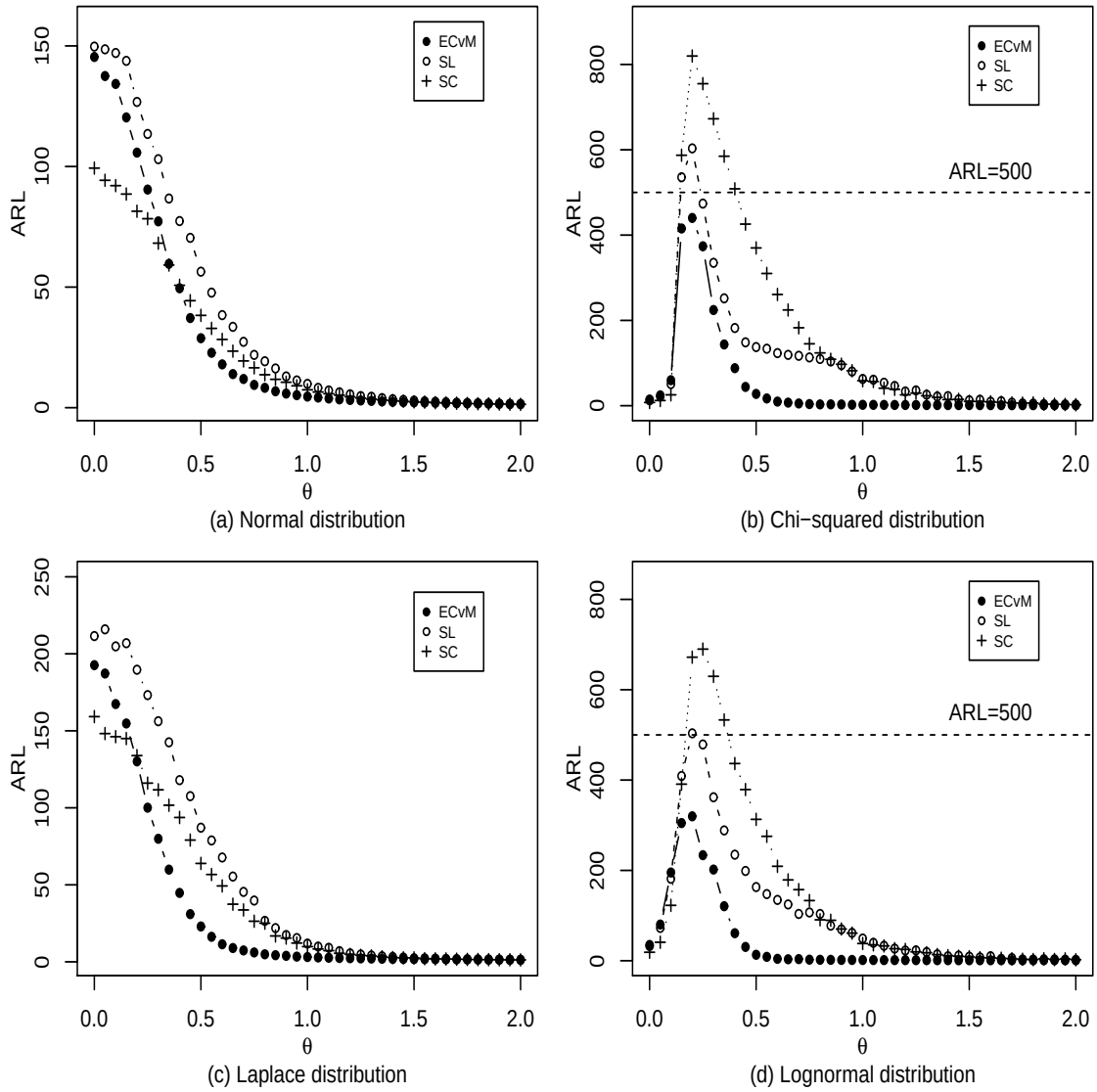

Fig. 4. ARL values of ECvM, SL and SC charts when the location parameter shifts with scale parameter shifting to 1.2 ($n = 30$, $m = 5$ and $ARL_0 = 500$).

sider is somewhat arbitrary because there is an infinite number of potential change patterns. We consider changes in the following four commonly used distributions.

(1) A change in the rate parameter of an exponential distribution, i.e., change from $Exp(1)$ to $Exp(3)$, and the reversed case.
(2) A change in the shape parameter of a gamma distribution, i.e., change from $Gamma(2, 2)$ to $Gamma(3, 2)$, and the reversed case.
(3) A change in the shape parameter of a Weibull distribution, i.e., change from $Weibull(1, 1)$ to $Weibull(3, 1)$, and the reversed case.
(4) A change in the degree-of-freedom parameter of a chi-squared distribution, i.e., change from $\chi^2(1)$ to $\chi^2(3)$, and the reversed case.

We set $n = 50$, $m = 5, 10$ and $ARL_0 = 500$. ARL values of ECvM, SL and SC charts based on 50,000 runs are listed in Table 7. From Table 7, when $m = 5$, for exponential distribution, if the rate parameter changes from 1 to 3, which indicates the location and scale parameters both have down-sided shifts, the ECvM, SL and SC chart have comparable performance; while for the reversed case, i.e., the rate parameter changes from 3 to 1, ECvM chart is much better than SL and SC charts. For gamma distribution, whatever up-sided or down-sided shifts, ECvM chart is better than SL and SC charts. For Weibull distribution, if shape parameter changes from 1 to 3, all the three charts are biased, but the ARL_1 values of the ECvM chart are still smaller than the other two charts, while for the reversed case, SC chart is the best. For chi-squared distribution, ECvM chart is better than SL and SC charts. When $m = 10$, the results show similar findings, except that all these charts are not biased any more for Weibull distribution.

Considering the significant ARL reduction for location and/or scale shifts and general patterns shifts, our proposed ECvM chart should be a good alternative to those used in the comparison.

5 Real Application Example

We illustrate the application of the proposed ECvM control chart using a well-known dataset for forged automobile engine piston rings, which can be found in Table 6.3 and Table 6E.8 of Montgomery (2013). The dataset contains the inside diameter measurements of the forged automobile engine piston rings. The goal of this study is to establish statistical control of the inside diameter of the piston rings in a forging process. Table 6.3 of Montgomery (2013) contains 25 samples, and each sample consists of 5 piston rings. Pointed out by Montgomery (2013), the traditional Shewhart $\overline{X}$ and R charts provide no indication of an OC signals, so these 125 samples are considered to be from IC

Table 7. ARL values of ECvM, SL and SC charts with $n = 50$ and $ARL_0 = 500$.

m	change pattern	ECvM	SL	SC
m=5	Exp(1)→Exp(3)	3.95	4.87	4.27
	Exp(3)→Exp(1)	4.92	42.37	47.47
	Gamma(2,2)→Gamma(3,2)	12.34	49.02	54.09
	Gamma(3,2)→Gamma(2,2)	9.50	24.41	21.24
	Weibull(1,1)→Weibull(3,1)	1528.05	2599.84	9947.73
	Weibull(3,1)→Weibull(1,1)	7.32	4.68	2.92
	$\chi^2(1)\rightarrow\chi^2(3)$	2.56	5.69	5.73
	$\chi^2(3)\rightarrow\chi^2(1)$	2.49	2.80	2.55
m=10	Exp(1)→Exp(3)	2.23	2.41	2.04
	Exp(3)→Exp(1)	2.37	12.54	10.65
	Gamma(2,2)→Gamma(3,2)	4.47	26.63	21.99
	Gamma(3,2)→Gamma(2,2)	5.43	12.39	9.86
	Weibull(1,1)→Weibull(3,1)	14.91	89.83	186.93
	Weibull(3,1)→Weibull(1,1)	4.43	2.23	1.53
	$\chi^2(1)\rightarrow\chi^2(3)$	1.44	2.06	1.91
	$\chi^2(3)\rightarrow\chi^2(1)$	1.48	1.47	1.35

process, and hence the IC sample size is $n = 125$. Table 6E.8 of Montgomery (2013) contains another 15 samples, and each sample consists of 5 piston rings. Hence the monitoring sample size is $m = 5$.

To establish statistical control of the inside diameter of the piston rings, we first checked the normality assumption of the IC data. The Shapiro-Wilk test for checking the normality of the IC data gives p-value less than 10^{-4}, which implies that the IC data are significantly non-normal. In this case, control charts constructed on the normal assumption are no longer applicable, as there would be many false alarms even when the process is IC. Our objective is to monitor the process and detect whether the process is IC by our ECvM chart, which does not rely on the normality assumption. This dataset are also studied by many researchers, such as Chakraborti et al. (2004), Chakraborti and Eryilmaz (2007), Chakraborti and van de Wiel (2008), Chakraborti et al. (2009), Mukherjee et al. (2013), and etc.

For our ECvM control chart, the parameters are $n = 125$, $m = 5$ and $ARL_0 = 500$. The control limit is searched by simulation and found to be $h = 0.668$. The charting statistics are shown in Figure 5. From Figure 5, our ECvM chart gives OC signal at the 14th observation, which is consistent with the result of the precedence chart of Chakraborti et al. (2004), that also signals at the 14th observation. However, with runs-type signaling rules based on two consecutive points, Chakraborti et al. (2009) show that precedence charts signal at the 10th observation, which implies that the runs-type signaling rules would greatly enhance the performance of precedence charts. It will be interesting to note that the signed-rank-based chart with runs-type signaling rules of Chakraborti and Eryilmaz (2007) signals at the 13rd observation, the Mann-Whitney-based chart of Chakraborti and van de Wiel (2008) signals at the 12th observation, and the exceedance CUSUM chart of Mukherjee et al. (2013) signals at the 13rd observation, respectively. However, their ARL_0 values are 271, 400 and 370, respectively, which are all far from our nominal 500. This implies that although one or two observations earlier of these three control charts give OC signals, the false alarm rate of these three charts are also higher when the process is IC.

6 Concluding Remarks

We developed a new distribution-free control chart, called ECvM, for jointly monitoring location and/or scale shifts and some general patterns of shifts of univariate continuous process in cases where the IC process distribution or the IC process parameters cannot be specified or estimated accurately enough. It combines the nonparametric two-sample CvM test and the EWMA model. Compared to some existing control procedures, including the SL and

SC charts, the proposed ECvM chart is robust to non-normally distributed data due to the IC distribution independence, and efficient in detecting various process shifts, including location and/or scale shifts, as well as general patterns of changes. A real data example for forged automobile engine piston rings is used to illustrate the application of our ECvM chart. Due to the satisfactory performance in the simulation comparisons and real example application, we recommend to use the ECvM chart in practice.

In SPC, it is important to detect abnormal changes as quickly as possible, and it is also important to diagnose the factors that lead to the changes, especially for a single control chart. Take the application of forged automobile engine piston rings as an example, it would be helpful, and sometimes, necessary, to determine which factors cause the change of the quality. Therefore, it warrants further research to isolate the causes of our proposed ECvM chart after it triggers an OC signal.

Acknowledgements

The authors are grateful to the editor, the associate editor and three anonymous referees for their valuable comments that have vastly improved this paper. This paper is supported by the National Natural Science Foundation of China Grants 11571191, 11431006 and 11371202.

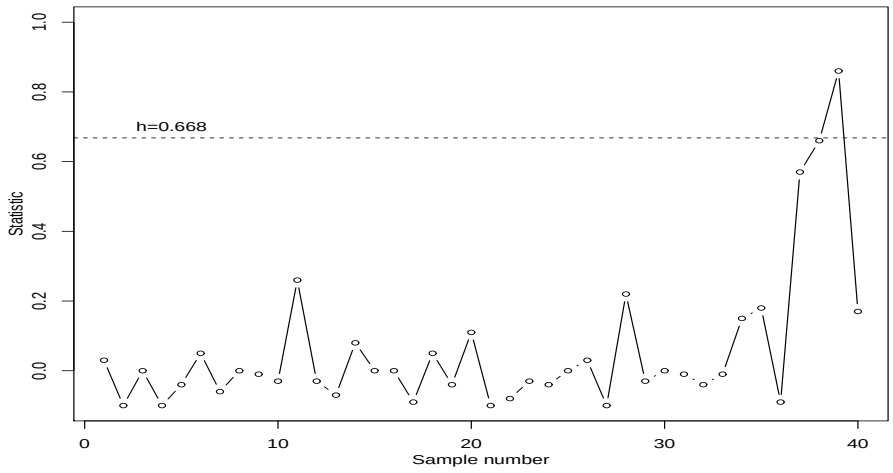

Fig. 5. Monitoring statistics of ECvM chart for the automobile engine piston rings (horizontal dashed lines indicating corresponding control limit).

Appendix

A Formula for ARL

Let R denote the run length of ECvM control chart. Note that the statistics $U_i, i = 1, 2, \ldots$ in (4), are dependent, they are, however, independent given the IC sample $X = \{X_1, X_2, \ldots, X_n\}$. Therefore, the conditional distribution of the run length is a geometric distribution with probability of success $Prob[U_i > h|X]$, i.e., the probability of ECvM control chart issuing an OC signal. Let $P(X|h) = Prob[U_i < h|X], i = 1, 2, \ldots$. Thus, conditional ARL (CARL) is $\frac{1}{1-P(X|h)}$ and unconditional ARL (UARL) can be obtained by averaging the information of X as $E[\frac{1}{1-P(X|h)}]$. Both CARL and UARL would provide valuable information for evaluating the performance of our ECvM control chart.

Note the probability of unconditional run length (URL) is

$$\begin{aligned} Prob(RL = r) &= E\{[P(X|h)]^{r-1}[1 - P(X|h)]\} \\ &= E[P(X|h)]^{r-1} - E[P(X|h)]^r, \qquad r = 1, 2, \ldots. \end{aligned}$$

Then the formula for ARL can be derived as

$$ARL = E\Big(\frac{1}{1 - P(X|h)}\Big) = \int_X \frac{1}{1 - P(X|h)} dF(X). \tag{A.1}$$

From (A.1), the formula for ARL_0 when $F = G$ can be derived as

$$ARL_0 = E_{IC}\Big(\frac{1}{1 - P(X|h)}\Big) = \int_X \frac{1}{1 - P_{F=G}(X|h)} dF(X). \tag{A.2}$$

From (A.2), for a prespecified ARL_0^*, the control limit h can be obtained by solving the equation

$$ARL_0^* = \int_X \frac{1}{1 - P_{F=G}(X|h)} dF(X). \tag{A.3}$$

References

Abid, M., Nazir, H. Z., Riaz, M. and Lin, Z. (2017a), "Investigating the Impact of Ranked Set Sampling in Nonparametric CUSUM Control Charts," *Quality and Reliability Engineering International*, 33, 203–214.

Abid, M., Nazir, H. Z., Riaz, M. and Lin, Z. (2017b), "An Efficient Non-

parametric EWMA Wilcoxon Signed-Rank Chart for Monitoring Location," *Quality and Reliability Engineering International*, 33, 669–685.

Ambartsoumian, T. and Jeske, D. R. (2015), "Nonparametric CUSUM Control Charts and Their Use in Two-Stage SPC Applications," *Journal of Quality Technology*, 47, 264–277.

Anderson, T. W. (1962), "On the Distribution of the Two-sample Cramer-von Mises Criterion," *The Annals of Mathematical Statistics*, 33, 1148–1159.

Capizzi, G. (2015), "Recent Advances in Process Monitoring: Nonparametric and Variable-Selection Methods for Phase I and Phase II," *Quality Engineering*, 27 (1), 44–67.

Celano, G., Castagliola, P. and Chakraborti, S. (2016), "Joint Shewhart Control Charts for Location and Scale Monitoring in Finite Horizon Processes," *Computers & Industrial Engineering*, 101, 427–439.

Chakraborti, S. and Eryilmaz, S. (2007), "A Nonparametric Shewhart-Type Signed-Rank Control Chart Based on Runs," *Communications in Statistics-Simulation and Computation*, 36, 335–356.

Chakraborti, S., Eryilmaz, S. and Human, S. W. (2009), "A Phase II Nonparametric Control Chart Based on Precedence Statistics with Runs-Type Signaling Rules," *Computational Statistics and Data Analysis*, 53, 1054–1065.

Chakraborti, S. and Graham, M. A. (2007), "Nonparametric Control Charts" *Encyclopedia of Statistics in Quality and Reliability*, 1, 415–429, John Wiley, New York.

Chakraborti, S., Human, S. W. and Graham, M. A. (2008), "Phase I Statistical Process Control Charts: An Overview and Some Results," *Quality Engineering*, 21 (1), 52–62.

Chakraborti, S, Human, S. W. and Graham, M. A. (2011), "Nonparametric (Distribution-Free) Quality Control Charts," *In Handbook of Methods and Applications of Statistics: Engineering, Quality Control, and Physical Sciences*, N Balakrishnan, Ed, 298–329, John Wiley & Sons, New York.

Chakraborti, S., Qiu, P. and Mukherjee, A. (2015), "Editorial to the Special Issue: Nonparametric Statistical Process Control Charts," *Quality and Reliability Engineering International*, 31, 1–2.

Chakraborti, S. and van de Wiel, M. A. (2008), "A Nonparametric Control Chart Based on the Mann-Whitney Statistic," *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, 1, 156–172.

Chakraborti, S., van der Laan, P. and Bakir, S. T. (2001), "Nonparametric Control Charts: An Overview and Some Results," *Journal of Quality Technology*, 33, 304–315.

Chakraborti, S., van der Laan, P. and van der Wiel, M. A. (2004), "A Class of Distribution-Free Control Charts," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53, 443–462.

Chatterjee, S. and Qiu, P. (2009), "Distribution Free Cumulative Sum Control Charts Using Bootstrap-Based Control Limits," *Annals of Applied Statis-*

tics, 3, 349–369.

Cheng, S. W. and Thaga, K. (2006), "Single Variables Control Charts: An Overview," *Quality and Reliability Engineering International*, 22, 811–820.

Chong, Z. L., Mukherjee, A. and Khoo, M. B. C. (2017), "Distribution-free Shewhart-Lepage Type Premier Control Schemes for Simultaneous Monitoring of Location and Scale," *Computers & Industrial Engineering*, 104, 201–215.

Chowdhury, S., Mukherjee, A. and Chakraborti, S. (2014), "A New Distribution-free Control Chart for Joint Monitoring of Unknown Location and Scale Parameters of Continuous Distributions," *Quality and Reliability Engineering International*, 30, 191–204.

Gandy, A. and Kvaløy, J. T. (2013), "Guaranteed Conditional Performance of Control Charts via Bootstrap Methods," *Scandinavian Journal of Statistics*, 40, 647–668.

Gibbons, J. D. and Chakraborti, S. (2011), *Nonparametric Statistical Inference*, Boca Raton: Chapman and Hall/CRC.

Graham, M. A., Chakraborti, S. and Human, S. W. (2011a), "A Nonparametric Exponentially Weighted Moving Average Signed-Rank Chart for Monitoring Location," *Computational Statistics and Data Analysis*, 55, 2490–2503.

Graham, M. A., Chakraborti, S. and Human, S. W. (2011b), "A Nonparametric EWMA Sign Chart for Location Based on Individual Measurements," *Quality Engineering*, 23 (3), 227–241.

Graham, M. A., Chakraborti, S. and Mukherjee, A. (2014), "Design and Implementation of CUSUM Exceedance Control Charts for Unknown Location," *International Journal of Production Research*, 52 (18), 5546–5564.

Graham, M. A., Mukherjee, A., and Chakraborti, S. (2012), "Distribution-free Exponentially Weighted Moving Average Control Charts for Monitoring Unknown Location," *Computational Statistics and Data Analysis*, 56, 2539–2561.

Hawkins, D. M. and Deng, Q. (2010), "A Nonparametric Change-Point Control Chart," *Journal of Quality Technology*, 42, 165–173.

Jensen, W. A., Jones-farmer, L. A., Champ, C. W. and Woodall, W. H. (2006), "Effects of Parameter Estimation on Control Chart Properties : A Literature Review," *Journal of Quality Technology*, 38, 349–364.

Jones, L. A. and Champ, C. W. (2010), "A Distribution-Free Phase I Control Chart for Subgroup Scale," *Journal of Quality Technology*, 42, 373–387.

Jones, L. A., Jordan, V. and Champ, C. W. (2009), "Distribution-free Phase I Control Charts for Subgroup Location," *Journal of Quality Technology*, 41, 304–316.

Lepage, Y. (1971), "A Combination of Wilcoxon's and Ansari-Bradley's Statistics," *Biometrika*, 58, 213–217.

Li, C., Mukherjee, A., Su, Q. and Xie, M. (2016), "Optimal Design of a Distribution-free Quality Control Scheme for Cost-efficient Monitoring of Unknown Location," *International Journal of Production Research*, 54,

7259–7273.

Li, S.-Y., Tang, L. C. and Ng, S. H. (2010), "Nonparametric CUSUM and EWMA Control Charts for Detecting Mean Shifts," *Journal of Quality Technology*, 42, 209–226.

Li, Z., Xie, M. and Zhou, M. (2017), "Rank-based EWMA Procedure for Sequentially Detecting Changes of Process Location and Variability," *Quality Technology and Quantitative Management*, in press, doi: 10.1080/16843703.2016.1208941.

Li, Z., Zou, C., Gong, Z. and Wang, Z. (2014), "The Computation of Average Run Length and Average Time to Signal: An Overview," *Journal of Statistical Computation and Simulation*, 84, 1779–1802.

Liu, L., Chen, B., Zhang, J. and Z, X. (2015), "Adaptive Phase II Nonparametric EWMA Control Chart with Variable Sampling Interval," *Quality and Reliability Engineering International*, 31, 15–26.

Lu, S. L. (2015), "An Extended Nonparametric Exponentially Weighted Moving Average Sign Control Chart," *Quality and Reliability Engineering International*, 31, 3–13.

Lucas, J. M. and Saccucci, M. S. (1990), "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements," *Technometrics*, 32, 1–12.

Maravelakis, P. E. and Castagliola, P. (2009), "An EWMA chart for Monitoring the Process Standard Deviation When Parameters are Estimated," *Computational Statistics and Data Analysis*, 53, 2653–2664.

McCracken, A. K. and Chakraborti, S. (2013), "Control Charts for Joint Monitoring of Mean and Variance: An Overview," *Quality Technology and Quantitative Management*, 10, 17–36.

Megahed, F. M., Woodall, W. H. and Camelio, J. A. (2011), "A Review and Perspective on Control Charting with Image Data," *Journal of Quality Technology*, 43, 83–98.

Montgomery, D. C. (2013), *Introduction to Statistical Quality Control*, New York: John Wiley and Sons, 7th ed.

Mukherjee, A. and Chakraborti, S. (2012), "A Distribution-free Control Chart for the Joint Monitoring of Location and Scale," *Quality and Reliability Engineering International*, 28, 335–352.

Mukherjee, A., Graham, M. A. and Chakraborti, S. (2013), "Distribution-free Exceedance CUSUM Control Charts for Location," *Communications in Statistics-Simulation and Computation*, 42, 1153–1187.

Mukherjee, A. and Marozzi, M. (2017), "Distribution-free Lepage Type Circular-grid Charts for Joint Monitoring of Location and Scale Parameters of a Process," *Quality and Reliability Engineering International*, 33, 241–274.

Mukherjee, P. S. (2016), "On Phase II Monitoring of the Probability Distributions of Univariate Continuous Processes," *Statistical Papers*, 28, 57(2),539–562.

Page, E. S. (1954), "Continuous Inspection Schemes," *Biometrika*, 41, 100–

115.

Qiu, P. (2014), *Introduction to Statistical Process Control*, Boca Raton: Chapman and Hall/CRC.

Qiu, P. and Li, Z. (2011a), "Distribution-free Monitoring of Univariate Processes," *Statistics and Probability Letters*, 81, 1833–1840.

Qiu, P. and Li, Z. (2011b), "On Nonparametric Statistical Process Control of Univariate Processes," *Technometrics*, 53, 390–405.

Roberts, S. W. (1959), "Control Chart Tests Based on Geometric Moving Averages," *Technometrics*, 1, 239–250.

Ross, G. J. and Adams, N. M. (2012), "Two Nonparametric Control Charts for Detecting Arbitrary Distribution Changes," *Journal of Quality Technology*, 44, 102–116.

Shewhart, W. A. (1925), "The Application of Statistics as an Aid in Maintaining Quality of a Manufactured Product," *Journal of the American Statistical Association*, 20, 546–548.

Teoh, W. L., Khoo, M. B.C., Castagliola, P. and Chakraborti, S. (2014), "Optimal Design of the Double Sampling $\overline{X}$ Chart with Estimated Parameters Based on Median Run Length," *Computers & Industrial Engineering*, 67, 104–115.

Woodall, W. H. (2006), "The Use of Control Charts in Health-Care and Public-Health Surveillance," *Journal of Quality Technology*, 38, 89–104.

Woodall, W. H. and Montgomery, D. C. (2014), "Some Current Directions in the Theory and Application of Statistical Process Monitoring," *Journal of Quality Technology*, 46, 78–94.

Zhou, M., Geng, W. and Wang, Z. (2014), "Likelihood Ratio-Based Distribution-Free Sequential Change-Point Detection," *Journal of Statistical Computation and Simulation*, 84, 2748–2758.

Zhou, M., Zhou, Q. and Geng, W. (2016), "A New Nonparametric Control Chart for Monitoring Variability," *Quality and Reliability Engineering International*, 32, 2471–2479.

Zou, C. and Tsung, F. (2010), "Likelihood Ratio-Based Distribution-Free EWMA Control Charts," *Journal of Quality Technology*, 42, 174–196.