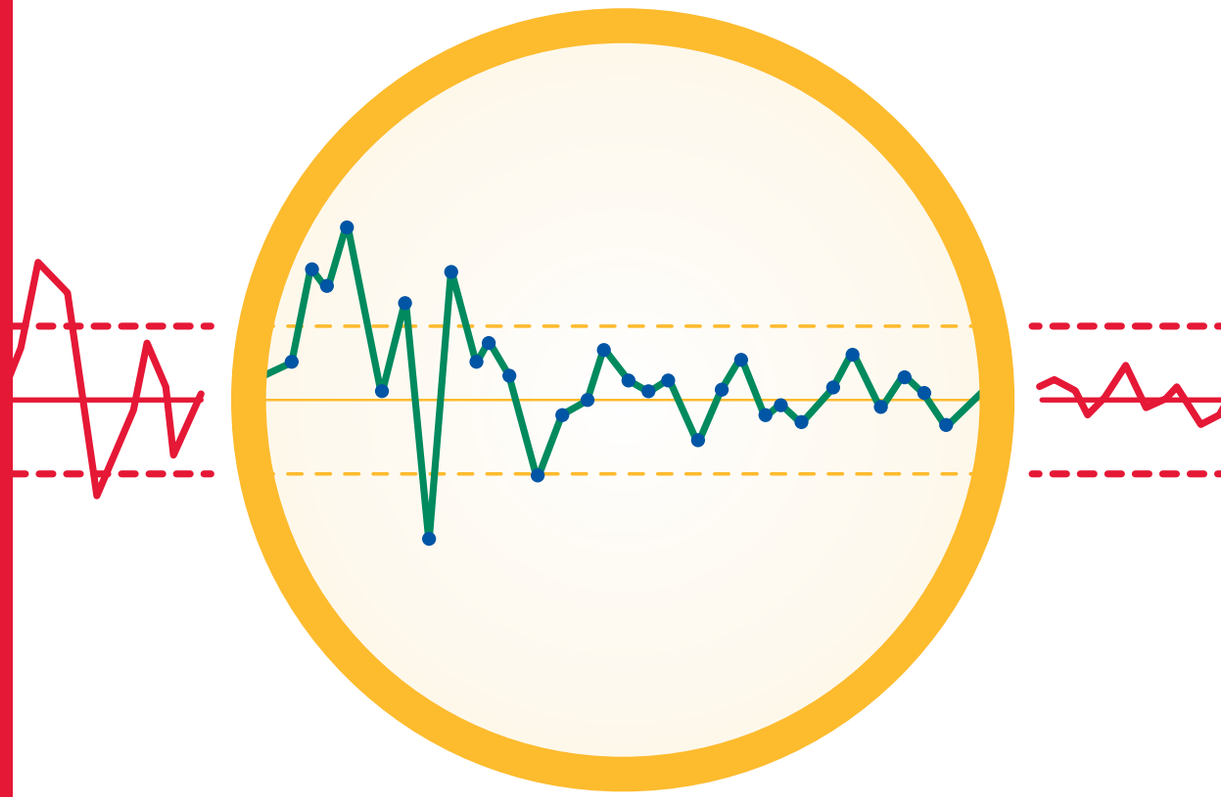
Design of control charts for statistical process monitoring

Diko, M.D.

Design of Control Charts for Statistical Process Monitoring

Mandla D. Diko

Design of Control Charts for Statistical Process Monitoring



Mandla D. Diko

Design of Control Charts for Statistical Process Monitoring

Mandla D. Diko

Instituut voor Bedrijfs- en Industriële Statistiek

Publisher IBIS UvA, Amsterdam

Printed by Gildeprint, Enschede

ISBN 978-94-6323-915-8

Design of Control Charts for Statistical Process Monitoring

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. ir. K.I.J. Maex

ten overstaan van een door het College voor Promoties ingestelde commissie,

in het openbaar te verdedigen in de Agnietenkapel

op dinsdag 3 december 2019, te 14:00 uur

door

Mandla Doctor Diko

geboren te Johannesburg

Promotiecommissie

Promotor

Prof. dr. R.J.M.M. Does	Universiteit van Amsterdam
Prof. dr. S. Chakraborti	University of Alabama

Overige leden

Prof. dr. C.A.J. Klaassen	Universiteit van Amsterdam
Prof. dr. P. Castagliola	Université de Nantes
Prof. dr. M.R.H. Mandjes	Universiteit van Amsterdam
Prof. dr. C.G.H. Diks	Universiteit van Amsterdam
Dr. M.A. Graham	University of Pretoria
Dr. I.M. Zwetsloot	City University of Hong Kong
Dr. M. Schoonhoven	Universiteit van Amsterdam

Faculteit

Economie en Bedrijfskunde

To my parents *Thembekile Diko* and

Nowelile Diko (Nontuli Garishe)

Contents

1. Introduction

This chapter provides an overview of one of the most important tools in Statistical Process Monitoring (SPM): the control chart. Important SPM terms like common cause variation, special cause variation, Phase I and Phase II analysis are explained. Basic notation is introduced. Some background information on the effects of parameter estimation and the design of control charts is provided, which leads to the motivation and outline of this thesis.

1.1 Statistical process monitoring

Everything that is required to transform an input into an output for a customer can be defined as a process. Processes exhibit variation. A process is improved by reducing process variation. Thus, the understanding of the process variation and its causes are important. Statistical Process Monitoring (SPM) provides statistical tools and techniques to understand process variation.

To clarify the issues, suppose a measurement *x* of some process characteristic X is obtained. We assume that X is continuous and its outcomes are independent. Also, we assume that the probability model for X belongs to a known family of distribution functions, $f(x, \theta)$, which depends on the parameter $\theta \in \Theta$, where Θ is the parameter space (a subset of the real line). The parameter θ can be a vector of parameters. Suppose that the desired value of θ is θ_0. A process that is operating at this value is said to be in-control (IC) or stable or predictable. For an IC process, the observed process variation is an acceptable part of the process. Since this variation cannot be attributed to any of an infinite number of likely causes, it is called unassignable cause variation or natural cause variation or common cause variation. An IC process can be improved, if necessary. Process improvement is done by reducing common cause variation. Common cause variation is reduced by redesigning the

process. This includes, changing job descriptions, training operators or introducing new equipment.

In addition to the inevitable common causes, other factors that are not part of the process will affect the process from time to time. The additional process variation that is introduced by these special factors is called special cause variation. The introduction of special cause variation may cause that the process parameter θ shifts from its IC value θ_0 to some other value in the Θ, say θ_1. A process operating at a shifted value of θ, is said to be out-of-control (OOC) or unpredictable or unstable. Before process improvement can begin, an OOC process has to be brought to an IC state. This is achieved by identifying and removing special cause variation. The presence of special cause variation can be identified by using control charts.

1.2 Control charts

Control charts are one of the most important statistical tools in the SPM tool kit, which are used to detect special cause variation. Like most other statistical methods, control chart designs assume a specific model, sampling strategy, data type and dimension. We will assume, throughout this thesis, that X is normally distributed with parameter vector $\underset{\sim}{\theta}_0 = (\mu_0, \sigma_0)$, where μ_0 and σ_0 are the parameters for the mean and standard deviation, respectively. Furthermore, we will assume that, at time i, a random sample $X_{i1}, X_{i2}, \dots, X_{in}$ of size *n* will be taken from the process. We also assume that the sample size and sampling intervals are constant over all i.

Control charts can be viewed as a series of hypothesis tests carried out over time with critical region A_i. The rejection region A_i is the set of all possible outcomes $\underset{\sim}{x}_i = (x_{i1}, x_{i2}, \dots, x_{in})$ of $\underset{\sim}{X}_i = (X_{i1}, X_{i2}, \dots, X_{in})$ that are not consistent with the monitored process parameter $\underset{\sim}{\theta}_0$. We would conclude that a process parameter has shifted (the process is OOC) if and only if $\underset{\sim}{x}_i \in A_i$. However, in-practice, this decision is not based on $\underset{\sim}{x}_i$ as such, but on some function $t\left(\underset{\sim}{x}_i\right)$, such that $T_i = t\left(\underset{\sim}{X}_i\right)$ is a random variable, which is called a charting statistic. Thus, we would conclude that $\underset{\sim}{\theta}_0$ has shifted if and only if

$\left\{\underset{\sim}{x}_i|(T_i \leq LCL) \cup (T_i \geq UCL)\right\} \in A_i$, where LCL and UCL stand for the lower control limit and upper control limit of a control chart, respectively. Note that, in making this OOC decision, we may be committing a *Type I error*. Hence, most control charts are designed to to keep $P(Type\ I\ error) = P(A_i|\theta_0) = \alpha$ small. This is done by specifying a very small value of α, like $\alpha_0 = 0.0027$, and finding LCL and UCL in such a way that

$$P(LCL < T < UCL|\theta_0) = 1 - \alpha. \tag{1.1}$$

Thus, for most control charts, LCL and UCL are functions of the $100(\alpha/2)^{th}$ and $100(1-\alpha/2)^{th}$ percentile of the *IC* sampling distribution of $T = t\left(\underset{\sim}{X}\right)$, respectively. In the sequel, we will sometimes call α and α_0 the false alarm rate (FAR) respectively the nominal FAR (FAR_0).

In this thesis, we consider three main classes of control charts, namely, the Shewhart charts, the cumulative sum (CUSUM) charts and the exponentially weighted moving average (EWMA) charts. Next, a brief description of each class is given.

1.2.1 The Shewhart control chart

Shewhart-type control charts were introduced by Walter A. Shewhart in 1924 (Shewhart (1926)). The charting statistic, $T = t\left(\underset{\sim}{X}\right)$, of a Shewhart chart is based on the information $\underset{\sim}{X}$ from the most recent i. Most Shewhart control chart designs assume that the IC sampling distribution of T is exactly or approximately normal with mean $\mu_T = E(T)$ and standard deviation $\sigma_T = \sqrt{Var(T)}$. In such cases, the control limits are written as

$$\begin{aligned} LCL &= \mu_T - c\sigma_T \\ UCL &= \mu_T + c\sigma_T \end{aligned} \tag{1.2}$$

where c is the charting constant or design constant to be found. The charting constant c is the $100(1-\alpha/2)^{th}$ percentile of a standard normal distribution, which can be found by specifying α. For example, if $\alpha = 0.0027$, then $c = 3$ and the resulting limits are called the 3-sigma limits. Note that, the Shewhart control limits in equation (1.2) apply to situations where T is normal or very close to normal.

Related to α, is the concept of the run length distribution and its properties. The run length is a random variable that counts the number of samples that must be taken before the control chart first gives an OOC signal. Since T_i is based only on current information and $\underset{\sim}{\theta}_0$ is known, the signaling events (A_i's) are independent. Hence $P\left(A_i|\underset{\sim}{\theta}_0\right) = \alpha$ is constant over all i and the IC run length distribution is geometric with parameter α. Under these conditions, a simple relationship exists between α and the IC average run length (ARL), denoted as ARL_{IN}. The ARL_{IN} is the reciprocal of α (also known as the false alarm rate (FAR)). Hence, control charts can also be designed by specifying a large value of ARL_{IN}, say ARL_0. Among charts with $ARL_{IN} = ARL_0$, the chart with the smallest OOC average run length (ARL_{ooc}) is preferred.

1.2.2 The EWMA chart

The exponentially weigthed moving average (EWMA) chart was introduced by Roberts (1959). The charting statistic of the EWMA chart is defined as

$$Y_i = \lambda T_i + (1-\lambda)Y_{i-1} \tag{1.3}$$

where $0 < \lambda \leq 1$ is the weight assigned to T_i. The chart's initial starting value Y_0 is usually set at μ_T. Setting $Y_0 \neq \mu_T$ gives the chart a head start (Lucas and Saccucci (1990)). The head start feature improves the OOC performance at the cost of a deterioration in the IC performance. The steady state EWMA control limits are defined as

$$\begin{aligned} LCL &= \mu_T - L\sigma_T\sqrt{\frac{\lambda}{(2-\lambda)}} \\ UCL &= \mu_T + L\sigma_T\sqrt{\frac{\lambda}{(2-\lambda)}} \end{aligned} \tag{1.4}$$

where the constant L is the charting constant to be found. The EWMA control chart provides a signal when $Y_i > \mu_T + L\sigma_T\sqrt{\lambda/(2-\lambda)}$ or $Y_i < \mu_T - L\sigma_T\sqrt{\lambda/(2-\lambda)}$. Note that, if $\lambda = 1$, the EWMA chart is equivalent to the Shewhart chart.

EWMA charts are designed by searching a pair of L and λ values that yields a specified ARL_0. Among these pairs, the pair that gives the smallest ARL_{ooc} is the optimal

pair. This search can be done using the R package *spc*. Note that, since Y_i is based on present and past information, the signaling events of the EWMA chart are not independent. As a result, the run length distribution is not geometric and the relationship $ARL_0 = 1/\alpha_0$ does not hold. Hence, the ARL of the EWMA chart has been calculated by simulations, Markov chains (Lucas and Saccucci (1990)) or integral equations (Crowder (1987)).

1.2.3 The CUSUM chart

CUSUM-type charts were introduced by Page (1954). Like the EWMA charts, the CUSUM charts are also based on past and current information. Due to this feature, CUSUM and EWMA charts are more efficient to detect small and moderate shifts than Shewhart charts (cf. Hawkins and Wu (2014)).

The charting statistics of the two-sided CUSUM control chart are given by

$$\begin{aligned} D_i^- &= min(0, D_{i-1}^- + T_i - \mu_T + k^-\sigma_T) \\ D_i^+ &= max(0, D_{i-1}^+ + T_i - \mu_T - k^+\sigma_T) \end{aligned} \tag{1.5}$$

where D_0^+ and D_0^- are the starting values, while k^+ and k^- are the reference values. The starting values are assumed to be equal to μ_T. Selecting $D_0^+ > \mu_T$ and/or $D_0^- < \mu_T$ gives the chart a fast initial response (FIR) or "head start", since it increases the charts ability to detect a process that is initially OOC (cf. Lucas and Crosier (1982)). We also assume that $k^+ = k^- = k$. The value k is usually chosen to be half the shift that is considered important enough to be detected. When the size of the shift is not known (e.g. with start-up processes), Sparks (2000) proposed using adaptive CUSUM (ACUSUM) charts, which employ current process data to estimate the shift and choose k in accordance to this estimate.

The control limits for the two-sided CUSUM chart are given as

$$\begin{aligned} LCL &= -h\sigma_T \\ UCL &= h\sigma_T \end{aligned} \tag{1.6}$$

where h is the charting constant to be found. For known parameters, the h values can be found in Qiu (2014) and also in the R package *spc* by specifying ARL_0 and k. Like EWMA charts, since the charting statistics D_i^+ and D_i^- are based on present and past information, the signaling events of the two-sided CUSUM chart are not independent. Hence, the ARL of

the two-sided CUSUM chart has been calculated by simulations, Markov chains (Brook and Evans (1972)) or integral equations (Page (1954)).

As mentioned before the control chart limits depend on the IC distribution of T_i that depends on $\underset{\sim}{\theta}_0$. However, in practice one or both parameters $\underset{\sim}{\theta}_0$ are not known and should be estimated from the IC process data. Next, we briefly look at the implications of parameter estimation.

1.3 Effects of parameter estimation

In most cases the parameters $\underset{\sim}{\theta}_0$ are not known and must be estimated from the IC process data. Thus, before estimation can begin, the IC state of the process has to be defined. The use of control charts to define the *IC* state of a process is called Phase I analysis (cf. Chakraborti et al. (2009) and Jones-Farmer et al. (2014)). Once the IC state of the process has been defined, the resulting IC data from Phase I is used to estimate the process parameters and control limits. The estimated control limits are used prospectively, in Phase II, to monitor the process for process parameter shifts. Hence, parameter estimation is an important step in setting up Phase II control limits. Unbiased minimum variance and robust estimators are generally preferred (cf. Mahmoud et al. (2010) and Schoonhoven et al. (2011a)).

Let $\underset{\sim}{\hat{\theta}}_0 = \left(\hat{\mu}_0 = t_1\left(\underset{\sim}{X}_i\right), \hat{\sigma}_0 = t_2\left(\underset{\sim}{X}_i\right)\right)$ denote a Phase I estimator for $\underset{\sim}{\theta}_0$, which has to be estimated based on m samples of size n, where $i = 1,2,\ldots,m$. To denote the dependence of a chart on Phase I parameter estimates, we write LCL and UCL as $\widehat{LCL}$ and $\widehat{UCL}$, respectively. To denote the dependence of the IC chart performance on Phase I parameter estimates, we write the FAR and ARL_{IN} as $CFAR$ and $CARL_{IN}$, respectively, where the prefix C stands for conditional. Thus, two or more practitioners that sample from the same process will get different Phase I data, which would lead to different parameter estimates, different charts and different chart performances. The observed variation between the results (observed values of $CFAR$ and $CARL_{IN}$) of the practitioners is termed "practitioner to practitioner variation". One of the serious implications of the practitioner to

practitioner variation are the radical differences between the observed $CFAR$ and $CARL_{IN}$ values and the nominally specified FAR_0 and ARL_0, respectively.

To study the effects of parameter estimation on the Phase II control chart performance, most studies choose to average the practitioner to practitioner variability and then compare the results to the corresponding parameters known ($Case\ K$) results. Specifically, the expected value of the $CFAR$ and the expected value of $CARL_{IN}$ have been calculated and then compared to the FAR_0 and ARL_0, respectively. This, the so-called unconditional or marginal perspective, for evaluating the effects of parameter estimation, has received a lot of attention in literature (cf. Quesenberry (1993), Chen (1997,1998), Chakraborti (2000,2006), Jones et al. (2004), Jensen et al. (2006), Psarakis et al. (2014), Goedhart et al. (2016), Jardim et al. (2019a, 2019b), Mehmood et al. (2018)). However, the unconditional approach ignores the random variability of the $CFAR$ or $CARL_{IN}$, the (i.e. practitioner to practitioner variability), which is inherent to parameter estimation.

In his work on prospective application of the Shewhart $\bar{X}$ chart, Chakraborti (2006) was among the first group of authors to emphasize how the $CFAR$ behaves as a random variable when parameters are estimated and used to construct Phase II charts. Inspired by this, for the Phase II Shewhart *S* and S^2 charts, Epprecht et al. (2015) examined the $CFAR$ distribution as a function of the Phase I sample size mn. They then made recommendations about the minimum size of the Phase I sample, which is required, to guarantee, with a high probability $1-p$, that the $CFAR$ will not exceed some specified nominal $CFAR$ value (FAR_0). This is the exceedance probability criterion (EPC) introduced by Albers and Kallenberg (2004a, 2004b) and Gandy and Kvaloy (2013) which sets an upper prediction bound to *CFAR*. In the same spirit, we examine the $CARL_{IN}$ distribution of the EWMA control chart for location (Chapter 3) and the CUSUM control chart for location (Chapter 4) as a function of m and n and set a lower prediction bound to $CARL_{IN}$. We then make recommendations about the value of *m*, which is required, to guarantee, with a specified high probability $1-p$, that the $CARL_{IN}$ will exceed a nominally specified value, ARL_0. This, the so called conditional approach, has been recommended in the recent literature to study the effects of parameter estimation as the $CARL_{IN}$ (also the $CFAR$) is a random variable with high variability which is the cause of practitioner to practitioner variation.

1.4 Design of control charts when parameters are not known

Recall that, if $\underset{\sim}{\theta}_0$ is known ($Case\ K$), a chart is designed by finding a charting constant $(c, L\ or\ h)$ in such a way that the IC performance of the chart is equal to some specified ARL_0/FAR_0. However, even when $\underset{\sim}{\theta}_0$ is not known, practitioners continue to use the parameters known charting constants to construct their charts. But, unless m is very large, the performance of the chart can be far from the nominally specified ARL_0/FAR_0 value. Thus, to compensate for small m and the effects of parameter estimation, the charting constants are adjusted or corrected. To do this, two approaches have been used, namely, the unconditional and the conditional approach.

The unconditional approach finds the adjusted charting constant, call it $(c^*, L^*or\ h^*)$, in such a way that the expected value of the random variable $CARL_{IN}/CFAR$ is equal to ARL_0/FAR_0. For Shewhart type charts, Chakraborti (2006) derived the exact integral equation for the $E(CARL_{IN})$ and evaluated the equation $E(CARL_{IN}) = ARL_0$ for c^*. Goedhart et al. (2016) used the two step Taylor series expansion on the $CFAR$ to derive an approximate formula for c^*. For EWMA type charts, Jones (2001) derived exact integral formulas for the $E(CARL_{IN})$. In addition, Jones et al. (2002) evaluated the integral equation $E(CARL_{IN}) = ARL_0$ for L^*, using Gaussian quadrature. For CUSUM charts, Jones et al. (2004) derived exact integral formulas for $E(CARL_{IN})$, however, $E(CARL_{IN}) = ARL_0$ has not been solved for h^*. Furthermore, in Diko et al. (2016), for a joint Shewhart location-spread charting scheme, they provide an exact integral formula for $E(CARL_{IN})$ and evaluate $E(CARL_{IN}) = ARL_0$ for c^*. Moreover, in Chapter 2, for various Shewhart dispersion charts, we provide an exact integral formula for $E(CARL_{IN})$ and evaluate $E(CARL_{IN}) = ARL_0$ for c^*.

However, even though the unconditionally c^* values result to an *IC* performance that is on average equal to ARL_0/FAR_0, they do not guarantee that small $CFAR$ values or large $CARL_{IN}$ values will not occur. To guarantee some minimum IC performance, the conditional approach has been proposed. The conditional approach finds $(c^*, L^*or\ h^*)$ in such a way that the probability that the $CFAR$ is less than FAR_0 or the probability that the $CARL_{IN}$ is more than ARL_0 is very high. This is the exceedance probability criterion (EPC) that we also use to study the effects of parameter estimation.

To apply the EPC to find $(c^*, L^* or\ h^*)$, Jones and Steiner (2012) and Gandy and Kvaloy (2013) have proposed the bootstrap method. Since then, many authors have worked in this area, including Saleh et al. (2015a, 2015b, 2016), Faraz et al. (2015, 2016), Aly et al. (2015) and Hu and Castagliola (2017). However, it has been shown (see, for example, Saleh et al. (2016)) that repeated application of the bootstrap approach results in different adjusted constants, which might lead to comparability issues. Also, for some practitioners, using the bootstrap method to apply the EPC might not guarantee the anticipated IC performance. Understanding these problems, for the Shewhart median chart, Hu and Castagliola (2017) found their charting constants by running the bootstrap approach a 100 times and averaging the results. For Shewhart charts, Goedhart et al. (2017a, 2017b, 2018) and Jardim et al. (2019a, 2019b) provided closed form analytic expressions to calculate the adjusted constants under the EPC and the normality assumption. Hence, the bootstrap method was no longer necessary. However, for the EWMA and CUSUM charts, such closed form analytical expressions are difficult to obtain, due to the dependence of signaling events. Consequently, we present different methods for adjusting the EWMA (Chapter 3) and CUSUM (Chapter 4) control chart limits, according to the EPC. The adjusted charting constants are presented graphically. Practitioners can use the graphs to find their own charting constants by interpolation.

1.5 Outline and scientific contribution

In this thesis, we make contributions to the design of some popular Phase II control charts. New methods for adjusting the control limits, according to some specified criteria, are proposed. Tables and graphs of the adjusted control limits or charting constants are given. Also, for applications where data is not scarce, Phase I sample sizes are recommended to design Phase II charts (without using the adjusted constants). In addition, we provide new simpler formulas that can be used to evaluate and design control charts.

Diko (2014) considered the joint monitoring of the process mean and standard deviation of a normally distributed process, using the Shewhart $(\bar{X}, R)$ scheme, where $\bar{X}$ denotes the sample mean and R the sample range. This scheme is implemented using the 3-sigma limits. There are three major issues with this. First, it is assumed that the charting

statistics are normally distributed. The second is multiple testing, since two charts are used at the same time to make decisions about the IC state of the process. The last issue is the effect of parameter estimation. We illustrate the severity of these issues on the IC performance of the chart and present a method to derive the adjusted charting constants taking proper account of these issues. Tables of the adjusted charting constants are given. Diko (2014) was my master's thesis. This thesis resulted in a paper entitled "Monitoring the process mean when standards are unknown: A classic problem revisited", which was published in *Quality and Reliability Engineering International* (cf. Diko et al. (2016)).

In Chapter 2, two methods are used to apply the unconditional approach to adjust the charting constants of Phase II Shewhart control charts for dispersion. The first method, the numerical approach, involves numerical integration. The second method, the analytic approach, is based on the first order Taylor approximation to the $CARL_{IN}$. Tables of the adjusted charting constants are given. This Chapter is based on a paper that was published in *Quality Engineering* under the title "Phase II control charts for monitoring dispersion when parameters are estimated" (cf. Diko et al. (2017)). The paper was a joint effort with dr. R. Goedhart, prof.dr. S. Chakraborti, prof.dr. R.J.M.M. Does and prof.dr. E.K. Epprecht.

In Chapter 3, a method for adjusting the charting constants of the Phase II EWMA control chart, according to the EPC, is presented. The method does not involve bootstrapping, but produces results that are comparable to existing results. Tables and graphs of the resulting adjusted constants are given. The chapter is based on a paper titled "Guaranteed in-control performance of the EWMA chart for monitoring the mean", which was published in *Quality and Reliability Engineering International* (cf. Diko et al. (2019a)). The paper was a joint effort with prof.dr. S. Chakraborti and prof.dr. R.J.M.M. Does.

In Chapter 4, simpler formulas to calculate the $CARL_{IN}$ values, the mean and standard deviation of the $CARL_{IN}$ distribution of the two-sided CUSUM chart are presented. The charting constants of a two-sided CUSUM chart are adjusted according to the EPC and without using bootstrapping. Graphs and tables of adjusted charting constants are given. These graphs are also shown to be useful in balancing the IC and OOC performance tradeoffs, which are associated with using the EPC adjusted charting constants. This Chapter is based on the paper called "An alternative design of the two-sided CUSUM chart

for monitoring the mean when parameters are estimated" that has been published in *Computers & Industrial Engineering* (cf. Diko et al. (2019b)). The paper was a joint effort with prof.dr. S. Chakraborti and prof.dr. R.J.M.M. Does.

In Chapter 5, for the classes of control charts that we considered in this thesis, we evaluate the OOC performance of their adjusted limits. To decide which class should be routinely used, a pairwise comparison of the control chart classes is done. This Chapter is based on the paper called "Comparison of the performance of the Shewhart, EWMA and CUSUM charts with estimated parameters" that has been accepted for publication in *Quality Engineering* (cf. Diko et al. (2019c)). The paper is written with dr. R. Goedhart and prof.dr. R.J.M.M. Does.

2. Control Charts for Monitoring Dispersion

2.1 Introduction

The two-sided Shewhart R (sample range) and S (sample standard deviation) control charts are widely used to monitor the process dispersion. In practice, the in-control standard deviation value is usually not known. Then these charts are applied with estimated control limits, where the parameter estimates are obtained from Phase I reference data. When applying these charts, it is common to use the 3-sigma limits, given in most textbooks (Montgomery (2013)), where a tabulation of the necessary chart constants can be found. The use of the standard 3-sigma limits is justified on the basis that the distribution of the charting statistic is normal or approximately normal. However, the charting statistics of the R and S charts are highly skewed. As a result, the performance of these charts can be quite questionable, particularly for smaller sample sizes typical in practice. To the best of our knowledge the performances of these charts have not been fully examined in the literature. But, these charts are critical in practical SPM applications since they are the most popular dispersion charts that are used in Phase II for keeping the process dispersion under control, before the location charts are constructed (which need an estimate of the process dispersion) and meaningfully interpreted (cf. Diko (2014) and Diko et al. (2016)). To look more closely into the issues we derive the expressions for the unconditional in-control average run length ($AARL_{IN}$) of the R and S charts (Montgomery (2013)), which are based on 3-sigma limits and the assumption of a normal distribution, using the conditioning-unconditioning method (Chen (1998) and Chakraborti (2000)) and evaluate them using the statistical software R. The evaluations are done for various values of the number of reference subgroups m, subgroup size $n = 5, 10$ and nominal in-control average run length (ARL_0) equal to 370. The results are shown in Table 2.1, where $PD = 100\left(\frac{AARL_{IN} - ARL_0}{ARL_0}\right)$ denotes the relative percentage difference between the $AARL_{IN}$ and the nominal in-control average run length (ARL_0), which is equal to 370.

Table 2.1: *The $AARL_{IN}$ and PD values for the estimated 3-sigma limits of the Phase II two-sided dispersion charts and three Phase I estimators; for $n = 5,10$; $ARL_0 = 370$, and various values of m.*

n	m	R chart with estimator $\bar{R}/d_2(n)$		S chart with estimator $\bar{S}/c_4(n)$		S chart with estimator S_p	
		$AARL_{IN}$	PD	$AARL_{IN}$	PD	$AARL_{IN}$	PD
5	5	4665	1161	745	101	668	80
	10	1000	170	570	54	523	41
	20	422	14	423	14	399	8
	30	332	-10	365	-2	349	-6
	50	278	-25	317	-14	308	-1
	100	245	-34	282	-24	278	-25
	500	222	-40	256	-31	256	-31
10	5	710	92	624	68	592	60
	10	478	29	551	49	530	43
	20	343	-8	469	27	456	23
	30	300	-19	429	16	421	14
	50	269	-27	393	6	388	5
	100	248	-33	363	-2	361	-3
	500	233	-37	339	-8	339	-9

Let us consider first the *R* chart. This chart uses the average range estimator $\bar{R}/d_2(n)$ for the unknown in-control standard deviation (σ_0), where $d_2(n)$ is the unbiasing constant assuming the normal distribution (Montgomery, 2013) and $\bar{R}$ is calculated from the m independent Phase I subgroup ranges $R_1, R_2, \ldots, R_m$. For $n = 5$, in Table 2.1, it can be

seen that the $AARL_{IN}$ values differ substantially from the nominal value, as the absolute PD values range from 10 (for $m = 30$) to 1161 (for $m = 5$). Note that a PD value greater (or smaller) than zero indicates that the $AARL_{IN}$ value is greater (or smaller) than the nominal value 370. Both cases are undesirable. It can also be observed that for $m \geq 30$, the *PD* values are negative, which means that increasing the number of reference subgroups *m* exacerbates the false alarm rate (*FAR*). Similar results are found for the *S* chart that uses the Phase I estimator $\bar{S}/c_4(n) = \sum_{i=1}^{m} S_i / mc_4(n)$, where $c_4(n)$ is the unbiasing constant assuming the normal distribution (Montgomery, 2013) and the S chart that uses the "pooled" estimator $S_p = \sqrt{\sum_{i=1}^{m} S_i^2 / m}$, respectively, where $S_1, S_2, \ldots, S_m$ denote the sample standard deviations of the *m* Phase I reference subgroups. Note that, the "pooled" estimator is not an unbiased estimator of σ_0. We use this estimator because the unbiasing constant $c_4(m(n-1)+1)$ is already 0.9876 in case $m = 5$ and $n = 5$, and it gets even closer to 1 as *m* and/or *n* increase (0.9975 for $m = 25$ and $n = 5$). Therefore, for all practical purposes this constant is indistinguishable from 1 and hence it is sufficient to use the estimator S_p. The reader is also referred to Mahmoud et al. (2010), where S_p and $S_p / c_4\left(m(n-1)+1\right)$ are compared and shown to be practically equal in terms of their probability distributions and mean squared error (*MSE*). For $n = 10$, in Table 2.1, the PD values are somewhat better than their counterparts for $n = 5$, but they are still unacceptable.

Based on these results, the standard estimated 3-sigma charts for dispersion available in the current literature cannot be recommended to monitor the dispersion in practice. This is an issue for anyone who uses these control limits available in most textbooks (Montgomery (2013)). In fact, most of the commercial software seem to use these same (incorrect) limits. An alternative approach is to use probability limits instead of the classical 3-sigma limits (Diko et al. (2016)). This mitigates this issue somewhat, but does not solve it entirely. Indeed, Montgomery (2013) mentions the use of probability limits and refers to some tables in Grant and Leavenworth (1986), but it is not clear whether or not these probability limits are commonly used in practice. Woodall (2017) advocates the use of probability limits for the dispersion control charts. For specified nominal FAR (denoted by α_0) such as 0.0027 or $ARL_0 = 370$, the probability limits may be constructed using the

exact distribution of the charting statistic. This will be discussed in more detail in Section 2.2. As an example, Table 2.2 shows the $AARL_{IN}$ and the PD values for the two sided R and the S charts with the estimated probability limits, for various values of $m, n = 5, 10$ and $ARL_0 = 370$.

Table 2.2 *The $AARL_{IN}$ and PD values for the Phase II two-sided R and S charts with estimated probability limits for various values of m; $n = 5, 10$;and $ARL_0 = 370$.*

n	m	R chart with estimator $\bar{R}/d_2(n)$		S chart with estimator $\bar{S}/c_4(n)$		S chart with estimator S_p	
		$AARL_{IN}$	PD	$AARL_{IN}$	PD	$AARL_{IN}$	PD
5	5	269	-27	270	-27	264	-29
	10	302	-18	302	-18	298	-19
	20	327	-11	328	-11	325	-12
	30	334	-10	339	-8	332	-10
	50	339	-8	350	-5	337	-9
	100	349	-5	359	-3	348	-6
	500	368	0	368	0	368	0
10	5	252	-32	252	-32	254	-31
	10	289	-22	289	-22	291	-21
	20	319	-14	319	-14	320	-14
	30	332	-10	332	-10	333	-10
	50	345	-7	345	-7	345	-7
	100	357	-4	357	-4	357	-4
	500	367	-1	367	-1	367	-1

It can be seen that now the PD values range from -29 to 0 and -32 to -1 for $n = 5$ and $n = 10$, respectively, and they approach zero as m increases, as one might expect (Chen, 1998). This means that the difference between the $AARL_{IN}$ values and the nominal $ARL_0 = 370$ value is not as bad as what was observed in Table 2.1. It also means that even though the situation has improved over using the 3-sigma limits, unless m is very large, the estimated probability limits may not lead to the desired ARL_0. The other thing to note is that the PD values are remarkably similar for all three estimators.

Thus, from a practical point of view, an important problem still persists. If the number of Phase I subgroups at hand m is small to moderate, even the estimated probability limits of the *R* and *S* charts do not quite maintain the advertised nominal in-control average run length. Hence, for a given nominal ARL_0 and a given amount of Phase I data, this chapter derives and tabulates new (correct/adjusted) charting constants, which account for the effects of parameter estimation. We achieve this by setting the $AARL_{IN}$ expression equal to some specified nominal value ARL_0 and then evaluating the resulting equation for $\alpha = \alpha(m,n)$. The in-control (IC) and out-of-control (OOC) average run length performance of the corrected probability limits charts are calculated and compared to the IC and OOC average run length performance of the uncorrected probability limits.

This chapter has been based on Diko et al. (2017) and is organized as follows. In Section 2.2 (the background) we describe the classical (uncorrected/unadjusted) 3-sigma limits and probability limits Shewhart control charts for dispersion. In Section 2.3 we derive new (corrected/adjusted) control limits based on a numerical and an analytical method. In Section 2.4, we evaluate the OOC behavior of the newly proposed probability limits. Finally, a summary and recommendations are provided in section 2.5.

2.2 Classical model for probability limits for the dispersion control charts

Suppose that m subgroups (samples) each of size n are available after a successful Phase I analysis to estimate the unknown parameters and set up the control limits that are to be used in prospective Phase II monitoring. Suppose that the data are from normal distributions and as before, let $R_1, R_2, \ldots, R_m$ denote the ranges and $S_1, S_2, \ldots, S_m$ denote the

standard deviations of the m Phase I subgroups. As noted earlier, the three commonly-used estimators of the unknown in-control process standard deviation σ_0 are (i) $\hat{\sigma}_{01} = \overline{R}/d_2(n)$, based on the average range, (ii) $\hat{\sigma}_{02} = \overline{S}/c_4(n)$, based on the average standard deviation and (iii) $\hat{\sigma}_{03} = S_p$, the pooled estimator.

Thus, using each of the three Phase I estimators above, the three most popular Phase II Shewhart standard deviation charts are (1) the *R* chart using the charting statistic $T_{i1} = R_i$ with the unbiased estimator $\hat{\sigma}_{01}$, (2) the *S* chart using the charting statistic $T_{i2} = S_i$ with the unbiased estimator $\hat{\sigma}_{02}$ and (3) the *S* chart using the charting statistic $T_{i3} = S_i$ with the estimator $\hat{\sigma}_{03}$, respectively. Note that for all three of the charts, we let $i = m+1, m+2, \ldots$ to emphasize that these are Phase II charts, where prospective monitoring starts from the $(m+1)^{\text{th}}$ sample having collected m Phase I samples. The subscript *j =1, 2, 3* is used to distinguish between the 3 charts. For chart *j*, we also write the unbiased Phase I estimator $\hat{\sigma}_{0j}$ as $\hat{\sigma}_{0j} = e_j/\epsilon_{0j}$, where e_j is a biased Phase I estimator based on the charting statistic (i.e. $\bar{R}$, $\bar{S}$, and S_p) and ϵ_{0j} is its corresponding unbiasing constant (with $\epsilon_{01} = d_2(n)$, $\epsilon_{02} = c_4(n)$, $\epsilon_{03} = c_4(m(n-1)+1) \approx 1$ shown in Appendix 2.A). Note also that even though Mahmoud et al. (2010) recommended the estimator S_p, we consider all three estimators here for completeness. It also allows us to contrast our results with those that are found in the current literature.

In general, the control limits of the j^{th} Phase II Shewhart chart for the process dispersion, with a charting statistic T_{ij}, can be written as

$$\begin{aligned} \widehat{UCL} &= G_{n,\alpha,j} e_j \\ \widehat{CL} &= e_j \\ \widehat{LCL} &= H_{n,\alpha,j} e_j \end{aligned} \tag{2.1}$$

where $G_{n,\alpha,j}$ and $H_{n,\alpha,j}$ are charting constants. Note that, for convenience, the constant ϵ_{0j}, that divides e_j to form the unbiased estimator, is taken to be a part of each of the charting constants G and H. The charting constants are based on the $100*\{1-\alpha/2\}^{th}$ and the $100*\{\alpha/2\}^{th}$ percentiles of the in-control distribution of the Phase II charting statistic T_{ij}, respectively, and are given in Appendix 2.A. Thus, these are probability limits. To this end,

note that (i) the in-control distribution of $T_{i1} = R_i$ is that of the random variable $W\sigma$, where *W* is the sample relative range, which has a well-known distribution for a normal population (cf. Gibbons and Chakraborti (2010)) (ii) the in-control distribution of T_{i2} and T_{i3} are both that of the random variable $\sqrt{\chi^2_{n-1}}/\sqrt{n-1}$, where χ^2_{n-1} is a chi-square variable with *n*-1 degrees of freedom.

Hence, for the *R* chart, the charting constants are given by $G_{n,\alpha,1} = \frac{F_{W_{n,1-\alpha/2}}}{d_2(n)}$ and $H_{n,\alpha,1} = \frac{F_{W_{n,\alpha/2}}}{d_2(n)}$, where $F_{W_{n,1-\alpha/2}}$ and $F_{W_{n,\alpha/2}}$ denote the $100*\{1-\alpha/2\}^{th}$ and the $100*\{\alpha/2\}^{th}$ percentiles of the in-control distribution of the sample relative range W, respectively. Given these charting constants plus $e_1 = \bar{R}$, we can find the control limits by substituting them into (2.1) . Similarly, the estimated probability limits for the *S* charts (i.e. S_2 and S_3) are obtained by substituting ($e_2 = \overline{S}$, $G_{n,\alpha,2} = \frac{\sqrt{\chi^2_{1-\alpha/2,n-1}}}{\sqrt{n-1}\,c_4(n)}$, $H_{n,\alpha,2} = \frac{\sqrt{\chi^2_{\alpha/2,n-1}}}{\sqrt{n-1}\,c_4(n)}$) and ($e_3 = S_p$, $G_{n,\alpha,3} = \frac{\sqrt{\chi^2_{1-\alpha/2,n-1}}}{\sqrt{n-1}}$, $H_{n,\alpha,3} = \frac{\sqrt{\chi^2_{\alpha/2,n-1}}}{\sqrt{n-1}}$) for S_2 and S_3, respectively, in (2.1). However, these charting constants were originally intended for use with the σ_0 known probability limits, and are thus incorporated using the nominal FAR ($\alpha = \alpha_0$). Furthermore, they only depend on the Phase II charting statistic, and not on the Phase I estimator or the Phase I sample size. Since they do not depend on m, these charting constants do not properly account for the effects of parameter estimation. Hence, they are *not* the appropriate constants in the case that σ_0 is unknown. Next, we correct them and so their control limits.

2.3 The *R* and *S* charts with estimated probability limits and corrected for the effects of parameter estimation

To properly account for the effects of parameter estimation arising in Phase I while using the Phase II charts, that is, to account for the effects of using m Phase I samples each of size n to estimate the in-control standard deviation σ_0, we propose to use the following probability limits

$$\widehat{UCL} = G_{n,\alpha(m,n),j}\, e_j \qquad (2.2)$$

$$\widehat{CL} = e_j$$

$$\widehat{LCL} = H_{n,\alpha(m,n),j} e_j$$

Note that, the above control limits are similar in form to those in (2.1) except that here we denote α as $\alpha(m,n)$ to emphasize that this probability should be a function of both *m* and *n*, to make the correct charting constants H and G depend on both of m and n, and thus account for parameter estimation.

In order to find the charting constants, we need to derive an expression for the unconditional in-control average run length ($AARL_{IN}$). This $AARL_{IN}$ depends on the in-control distributions of both the Phase I estimator (e_j) and the Phase II charting statistic (T_{ij}). In our derivations, we assume that e_j/σ follows a scaled chi-square distribution $\frac{\epsilon_{0j} a_{0j} \sqrt{U_{0j}}}{\sqrt{b_{0j}}}$, where U_{0j} denotes a chi-square random variable with b_{0j} degrees of freedom. Formulae and/or values for the constants a_{0j}, b_{0j} and ϵ_{0j} are given in Appendix 2.A, and are based on the well-known Patnaik (1950) approximation (see Chen (1998) for the explicit expressions). Note that conditional on the observed value of the Phase I estimator e_j (or equivalently, on the realization of U_{0j}) , the conditional in-control ($\sigma = \sigma_0$) Phase II run length distribution is geometric. The success probability of this distribution is equal to the conditional false alarm rate (denoted $CFAR$), which is defined as

$$\begin{aligned}
&CFAR_j \\
&= 1 - P\left(\widehat{LCL} < T_{ij} < \widehat{UCL} | \sigma = \sigma_0\right) \\
&= 1 - P\left(H_{n,\alpha,j} e_j < T_{ij} < G_{n,\alpha,j} e_j | \sigma = \sigma_0\right) \\
&= 1 - P\left(H_{n,\alpha,j} \frac{e_j}{\sigma} < \frac{T_{ij}}{\sigma} < G_{n,\alpha,j} \frac{e_j}{\sigma} | \sigma = \sigma_0\right) \qquad (2.3) \\
&= 1 - P\left(H_{n,\alpha,j} \frac{\epsilon_{0j} a_{0j} \sqrt{U_{0j}}}{\sqrt{b_{0j}}} < \frac{T_{ij}}{\sigma} < G_{n,\alpha,j} \frac{\epsilon_{0j} a_{0j} \sqrt{U_{0j}}}{\sqrt{b_{0j}}} | \sigma = \sigma_0\right) \\
&= CFAR_j\left(U_{0j}, m, n, \alpha\right)
\end{aligned}$$

Next, using the conditioning-unconditioning method in Diko (2014) and Diko et al. (2016), the unconditional ARL_{IN} of the j^{th} Phase II Shewhart dispersion chart can be obtained as

$$AARL_{IN,j}(m,n,\alpha) = \int_0^{\infty}\left[CFAR_j(u,m,n,\alpha)\right]^{-1} g_{\chi^2_{b_{0j}}}(u)du \qquad (2.4)$$

where $g_{\chi^2_{b_{0j}}}$ denotes the probability density function (pdf) of U_{0j}. We start with the numerical approach, which solves the equation $AARL_{IN,j}\big(m,n,\alpha(m,n)\big) = ARL_0$ numerically for $\alpha\big(m,n\big)$.

2.3.1 The numerical approach

The numerical approach finds $\alpha\big(m,n\big)$ numerically and uses it to correct the uncorrected constants in Montgomery (2013) as follows:

(i) specifies the values of *m, n* at hand and the desired nominal ARL_0;

(ii) uses the exact in control distributions of the charting statistics to (1) define the control limits, and (2) determine the expressions for the $CFAR_j\left(U_{0j},m,n,\alpha(m,n)\right)$ and the $AARL_{IN,j}\big(m,n,\alpha(m,n)\big)$;

(iii) numerically solves the equation $\int_0^{\infty}\left[CFAR_j\big(u,m,n,\alpha(m,n)\big)\right]^{-1} g_{\chi^2_{b_{0j}}}(u)du = ARL_0$ for the corresponding $\alpha\big(m,n\big)$ value; and

(iv) uses $\alpha\big(m,n\big)$ to correct the uncorrected charting constants in Montgomery (2013).

For example, for the R chart, recall that $H_{n,\alpha(m,n),j} = \frac{F_{W_{n,\alpha(m,n)/2}}}{d_2(n)}$ and $G_{n,\alpha(m,n),j} = \frac{F_{W_{n,1-\alpha(m,n)/2}}}{d_2(n)}$. Consequently, $CFAR_1$ becomes

$$\begin{aligned}
&CFAR_1\big(U_{01},m,n,\alpha(m,n)\big)\\
&= 1 - P\left(H_{n,\alpha(m,n),1}\frac{e_1}{\sigma} < \frac{R_i}{\sigma} < G_{n,\alpha(m,n),1}\frac{e_1}{\sigma}|\sigma=\sigma_0\right)\\
&= 1 - F_{W_n}\left(F_{W_{n,1-\alpha(m,n)/2}}\frac{a_{01}\sqrt{U_{01}}}{\sqrt{b_{01}}}\right) + F_{W_n}\left(F_{W_{n,\alpha(m,n)/2}}\frac{a_{01}\sqrt{U_{01}}}{\sqrt{b_{01}}}\right)
\end{aligned}$$

where F_{W_n} represents the cumulative distribution function (cdf) of the sample relative range. Using this equation to solve

$$\int_0^{\infty}\left[CFAR_j\left(u,m,n,\alpha(m,n)\right)\right]^{-1} g_{\chi^2_{b_{0j}}}(u)du = ARL_0 \tag{2.5}$$

will result in the required corrected charting constants. For example, with $m = 5, n = 5$, and $ARL_0 = 370$, the value $\alpha(m,n)$ that satisfies the above equation is 0.001949. This value is then used to correct the uncorrected charting constants for the estimated probability limits Phase II Shewhart R chart. The corrected charting constants for the Phase II R chart with $\hat{\sigma}_{01}$ as Phase I estimator, are given by

$$G_{5,\alpha(5,5),1} = \frac{F_{W_{5,1-\alpha(5,5)/2}}}{d_2(5)} = \frac{F_{W_{5,1-0.001949/2}}}{d_2(5)} = \frac{5.49281}{2.32593} = 2.3616$$

and

$$H_{5,\alpha(5,5),1} = \frac{F_{W_{5,\alpha(5,5)/2}}}{d_2(5)} = \frac{F_{W_{5,0.001949/2}}}{d_2(5)} = \frac{0.36499}{2.32593} = 0.1569,$$

respectively. For other values of n, m and ARL_0, the values of $\alpha(m,n)$, $G_{n,\alpha(m,n),1}$ and $H_{n,\alpha(m,n),1}$ are given in Table 2.3. The R codes for finding all these values are given in Diko et al. (2017).

Similarly, recall that for *j =2* we have $H_{n,\alpha(m,n),2} = \frac{\sqrt{\chi^2_{\alpha(m,n)/2,n-1}}}{c_4(n)\sqrt{n-1}}$ and $G_{n,\alpha(m,n),2} = \frac{\sqrt{\chi^2_{1-\alpha(m,n)/2,n-1}}}{c_4(n)\sqrt{n-1}}$, and that $\frac{S}{\sigma} \sim \frac{\sqrt{\chi^2_{n-1}}}{\sqrt{n-1}}$. This can be used to calculate $CFAR_2$ as

$$\begin{aligned}
&CFAR_2\left(U_{02},m,n,\alpha(m,n)\right)\\
&= 1 - P\left(H_{n,\alpha(m,n),2}\frac{e_2}{\sigma} < \frac{S_i}{\sigma} < G_{n,\alpha(m,n),2}\frac{e_2}{\sigma}|\sigma = \sigma_0\right)\\
&= 1 - F_{\chi^2_{n-1}}\left(\chi^2_{1-\alpha(m,n)/2,n-1}\frac{a_{02}^2U_{02}}{b_{02}}\right) + F_{\chi^2_{n-1}}\left(\chi^2_{\alpha(m,n)/2,n-1}\frac{a_{02}^2U_{02}}{b_{02}}\right)
\end{aligned}$$

which can in turn be used to determine the required charting constants.

Table 2.3: *New corrected charting constants for the Shewhart R and S charts when parameters are estimated from m Phase I subgroups, $m = 5,10,20,25,30,50,100,300$ each of subgroup size $n = 5, 10$ for a nominal in-control $ARL_0 = 370$ and 500.*

			Numerical Approach						Analytical Approach					
			ARL_0=370			ARL_0=500			ARL_0=370			ARL_0=500		
n	m	Chart	α	H	G	α	H	G	α	H	G	α	H	G
5	5	R - $\bar{R}$	0.001949	0.1569	2.3616	0.001433	0.1451	2.4074	0.001686	0.1513	2.3832	0.001229	0.1396	2.4300
		S - S_p	0.001908	0.1489	2.1547	0.001402	0.1377	2.1939	0.001582	0.1420	2.1786	0.001149	0.1309	2.2188
		S - $\bar{S}$	0.001954	0.1593	2.2890	0.001438	0.1474	2.3306	0.001672	0.1532	2.3103	0.001218	0.1414	2.3527
	10	R - $\bar{R}$	0.002194	0.1617	2.3436	0.001615	0.1496	2.3897	0.002091	0.1598	2.3509	0.001534	0.1477	2.3974
		S - S_p	0.002166	0.1538	2.1383	0.001594	0.1422	2.1777	0.002035	0.1513	2.1464	0.001490	0.1398	2.1862
		S - $\bar{S}$	0.002198	0.1642	2.2729	0.001617	0.1519	2.3148	0.002082	0.1619	2.2803	0.001527	0.1497	2.3225
	20	R - $\bar{R}$	0.002384	0.1652	2.3310	0.001757	0.1529	2.3771	0.002350	0.1646	2.3332	0.001730	0.1522	2.3795
		S - S_p	0.002368	0.1573	2.1268	0.001745	0.1455	2.1662	0.002320	0.1565	2.1294	0.001706	0.1447	2.1690
		S - $\bar{S}$	0.002387	0.1677	2.2614	0.001760	0.1552	2.3033	0.002345	0.1669	2.2639	0.001726	0.1544	2.3060
	25	R - $\bar{R}$	0.002434	0.1660	2.3278	0.001795	0.1537	2.3740	0.002410	0.1656	2.3294	0.001775	0.1532	2.3756
		S - S_p	0.002420	0.1581	2.1239	0.001783	0.1463	2.1634	0.002386	0.1576	2.1258	0.001756	0.1458	2.1653
		S - $\bar{S}$	0.002435	0.1685	2.2587	0.001797	0.1560	2.3005	0.002406	0.1680	2.2603	0.001772	0.1554	2.3024
	30	R - $\bar{R}$	0.002469	0.1666	2.3257	0.001821	0.1542	2.3718	0.002452	0.1664	2.3267	0.001807	0.1539	2.3729
		S - S_p	0.002457	0.1587	2.1219	0.001812	0.1469	2.1614	0.002432	0.1583	2.1233	0.001792	0.1465	2.1628
		S - $\bar{S}$	0.002470	0.1691	2.2566	0.001823	0.1566	2.2985	0.002449	0.1687	2.2579	0.001805	0.1562	2.2999
	50	R - $\bar{R}$	0.002549	0.1680	2.3208	0.001883	0.1555	2.3668	0.002543	0.1679	2.3211	0.001877	0.1554	2.3672
		S - S_p	0.002542	0.1601	2.1175	0.001876	0.1482	2.1569	0.002530	0.1599	2.1181	0.001867	0.1481	2.1575
		S - $\bar{S}$	0.002550	0.1705	2.2522	0.001883	0.1579	2.2941	0.002541	0.1703	2.2527	0.001875	0.1577	2.2947
	100	R - $\bar{R}$	0.002619	0.1692	2.3166	0.001936	0.1567	2.3626	0.002618	0.1692	2.3167	0.001935	0.1566	2.3627
		S - S_p	0.002615	0.1613	2.1137	0.001932	0.1494	2.1531	0.002612	0.1612	2.1139	0.001930	0.1493	2.1533
		S - $\bar{S}$	0.002621	0.1717	2.2484	0.001936	0.1590	2.2903	0.002617	0.1716	2.2486	0.001934	0.1589	2.2904
	300	R - $\bar{R}$	0.002674	0.1701	2.3134	0.001978	0.1575	2.3594	0.002673	0.1701	2.3135	0.001977	0.1575	2.3594
		S - S_p	0.002672	0.1622	2.1109	0.001976	0.1502	2.1502	0.002671	0.1622	2.1109	0.001976	0.1502	2.1502
		S - $\bar{S}$	0.002674	0.1726	2.2455	0.001977	0.1598	2.2874	0.002673	0.1725	2.2456	0.001977	0.1598	2.2874
10	5	R - $\bar{R}$	0.001813	0.3481	1.9514	0.001328	0.3349	1.9840	0.001467	0.3391	1.9737	0.001057	0.3256	2.0075
		S - S_p	0.001812	0.3534	1.7681	0.001328	0.3402	1.7931	0.001383	0.3419	1.7899	0.000992	0.3282	1.8162
		S - $\bar{S}$	0.001831	0.3638	1.8169	0.001342	0.3502	1.8427	0.001424	0.3527	1.8377	0.001024	0.3388	1.8647
	10	R - $\bar{R}$	0.002095	0.3545	1.9360	0.001538	0.3411	1.9687	0.001965	0.3517	1.9428	0.001435	0.3381	1.9759
		S - S_p	0.002095	0.3598	1.7561	0.001538	0.3463	1.7813	0.001921	0.3560	1.7633	0.001400	0.3424	1.7889
		S - $\bar{S}$	0.002107	0.3702	1.8050	0.001547	0.3564	1.8309	0.001942	0.3665	1.8119	0.001416	0.3525	1.8382
	20	R - $\bar{R}$	0.002319	0.3590	1.9251	0.001708	0.3455	1.9577	0.002280	0.3583	1.9270	0.001675	0.3447	1.9597
		S - S_p	0.002320	0.3644	1.7477	0.001708	0.3508	1.7729	0.002256	0.3632	1.7500	0.001656	0.3495	1.7754
		S - $\bar{S}$	0.002327	0.3748	1.7966	0.001713	0.3608	1.8224	0.002267	0.3736	1.7988	0.001664	0.3596	1.8248
	25	R - $\bar{R}$	0.002378	0.3602	1.9224	0.001752	0.3466	1.9550	0.002352	0.3597	1.9236	0.001730	0.3461	1.9563
		S - S_p	0.002377	0.3655	1.7457	0.001751	0.3519	1.7708	0.002333	0.3647	1.7472	0.001715	0.3510	1.7725
		S - $\bar{S}$	0.002384	0.3759	1.7945	0.001756	0.3619	1.8204	0.002341	0.3751	1.7960	0.001721	0.3611	1.8220
	30	R - $\bar{R}$	0.002420	0.3610	1.9206	0.001783	0.3474	1.9531	0.002403	0.3606	1.9213	0.001769	0.3470	1.9540
		S - S_p	0.002420	0.3663	1.7442	0.001783	0.3527	1.7694	0.002387	0.3657	1.7453	0.001756	0.3521	1.7706
		S - $\bar{S}$	0.002425	0.3767	1.7930	0.001787	0.3627	1.8189	0.002394	0.3761	1.7942	0.001762	0.3621	1.8201
	50	R - $\bar{R}$	0.002516	0.3627	1.9164	0.001857	0.3492	1.9488	0.002512	0.3627	1.9166	0.001852	0.3491	1.9491
		S - S_p	0.002516	0.3681	1.7410	0.001857	0.3545	1.7660	0.002502	0.3679	1.7414	0.001845	0.3542	1.7666
		S - $\bar{S}$	0.002520	0.3785	1.7898	0.001860	0.3645	1.8156	0.002506	0.3783	1.7902	0.001848	0.3642	1.8161
	100	R - $\bar{R}$	0.002602	0.3643	1.9127	0.001922	0.3507	1.9452	0.002602	0.3643	1.9127	0.001922	0.3507	1.9452
		S - S_p	0.002602	0.3697	1.7381	0.001921	0.3560	1.7633	0.002597	0.3696	1.7383	0.001918	0.3559	1.7634
		S - $\bar{S}$	0.002603	0.3801	1.7870	0.001924	0.3661	1.8127	0.002599	0.3800	1.7871	0.001920	0.3660	1.8129
	300	R - $\bar{R}$	0.002667	0.3654	1.9101	0.001973	0.3518	1.9424	0.002668	0.3654	1.9100	0.001973	0.3518	1.9424
		S - S_p	0.002668	0.3708	1.7361	0.001972	0.3571	1.7611	0.002666	0.3708	1.7361	0.001971	0.3571	1.7611
		S - $\bar{S}$	0.002668	0.3812	1.7849	0.001973	0.3672	1.8106	0.002667	0.3812	1.7849	0.001972	0.3672	1.8106

Finally, recall that for *j =3* we have $H_{n,\alpha(m,n),3} = \frac{\sqrt{\chi^2_{\alpha(m,n)/2,n-1}}}{\sqrt{n-1}}$ and $G_{n,\alpha,3} = \frac{\sqrt{\chi^2_{1-\alpha(m,n)/2,n-1}}}{\sqrt{n-1}}$, and that $\frac{S}{\sigma} \sim \frac{\sqrt{\chi^2_{n-1}}}{\sqrt{n-1}}$. This can be used to calculate $CFAR_3$ as

$$
\begin{aligned}
&CFAR_3\left(U_{03}, m, n, \alpha(m,n)\right) \\
&= 1 - P\left(H_{n,\alpha(m,n),3}\frac{e_3}{\sigma} < \frac{S_i}{\sigma} < G_{n,\alpha(m,n),3}\frac{e_3}{\sigma} | \sigma = \sigma_0\right) \\
&= 1 - F_{\chi^2_{n-1}}\left(\chi^2_{1-\alpha(m,n)/2,n-1}\frac{U_{03}}{m(n-1)}\right) + F_{\chi^2_{n-1}}\left(\chi^2_{\alpha(m,n)/2,n-1}\frac{U_{03}}{m(n-1)}\right)
\end{aligned}
$$

which can be used to determine the required charting constants for this case.

From Table 2.3, it is interesting to see that when m increases, as one might expect, the $\alpha(m,n)$, $G_{n,\alpha(m,n),j}$ and $H_{n,\alpha(m,n),j}$ values converge to their σ_0 known counterparts α_0, $G_{n,\alpha_0,j}$and $H_{n,\alpha_0,j}$, respectively.

2.3.2 The analytical approach

While the numerical solutions outlined above are useful, it is interesting to consider an approximation to the charting constants based on the recent work of Goedhart et al. (2016, 2017), which is based on a first order Taylor approximation of the $AARL_{IN}$. The numerical approach of finding $\alpha(m,n)$ involves numerical integration and solving some non-linear equations. However, it is also possible to find $\alpha(m,n)$ using a more easily implementable but approximate method. Our approach is to:

(i) specify the values of m, n at hand and the desired nominal ARL_0;

(ii) unify the control charts for dispersion under one chi-square framework, which assumes that the charting statistic T_{ij} is either exactly or approximately distributed as a scaled chi-square random variable $\frac{\epsilon_j a_j \sigma \sqrt{\chi^2_{b_j}}}{\sqrt{b_j}}$, where $\chi^2_{b_j}$ is a chi-square random variable with b_j degrees of freedom, $\epsilon_j$equals the expectation of T_{ij} , and a_j is some constant. Formulae and/or values for the constants a_j,b_j

and ϵ_j are given in Appendix 2.A, and are based on the Patnaik (1950) approximation, similar to e_j ;

(iii) use the above chi-square framework to (1) define the control limits, (2) determine the expressions for the $CFAR_j\left(U_{0j}, m, n, \alpha(m,n)\right)$ and the $AARL_{IN,j}\big(m, n, \alpha(m,n)\big)$;

(iv) obtain an analytical expression for $\alpha\big(m,n\big)$; and

(v) use the resulting value of $\alpha\big(m,n\big)$ to adjust the uncorrected charting constants.

Using the approximations in step (ii), we can write $CFAR_j$ more explicitly as

$$
\begin{aligned}
&CFAR_j\left(U_{0j}, m, n, \alpha(m,n)\right)\\
&= 1 - P\big(H_{n,\alpha,j}e_j < T_{ij} < G_{n,\alpha,j}e_j|\sigma = \sigma_0\big)\\
&= 1 - P\left(H_{n,\alpha,j}\frac{e_j}{\sigma} < \frac{T_{ij}}{\sigma} < G_{n,\alpha,j}\frac{e_j}{\sigma}|\sigma = \sigma_0\right)\\
&= 1 - P\left(H_{n,\alpha,j}\frac{\epsilon_{0j}a_{0j}\sqrt{U_{0j}}}{\sqrt{b_{0j}}} < \frac{\epsilon_j a_j\sqrt{\chi^2_{b_j}}}{\sqrt{b_j}} < G_{n,\alpha,j}\frac{\epsilon_{0j}a_{0j}\sqrt{U_{0j}}}{\sqrt{b_{0j}}}\right)\\
&= 1 - P\left(\left(\frac{a_j\sqrt{\chi^2_{\alpha/2,b_j}}}{\sqrt{b_j}}\right)^2\frac{a_{0j}^2b_jU_{0j}}{a_j^2b_{0j}} < \chi^2_{b_j} < \left(\frac{a_j\sqrt{\chi^2_{1-\alpha/2,b_j}}}{\sqrt{b_j}}\right)^2\frac{a_{0j}^2b_jU_{0j}}{a_j^2b_{0j}}\right)\\
&= 1 - P\left(\chi^2_{\alpha/2,b_j}\frac{a_{0j}^2U_{0j}}{b_{0j}} < \chi^2_{b_j} < \chi^2_{1-\alpha/2,b_j}\frac{a_{0j}^2U_{0j}}{b_{0j}}\right)\\
&= 1 - F_{\chi^2_{b_j}}\left(\chi^2_{1-\alpha/2,b_j}\frac{a_{0j}^2U_{0j}}{b_{0j}}\right) + F_{\chi^2_{b_j}}\left(\chi^2_{\alpha/2,b_j}\frac{a_{0j}^2U_{0j}}{b_{0j}}\right)
\end{aligned}
\tag{2.6}
$$

where $F_{\chi^2_{b_j}}$ represents the *cdf* of a chi-square variable with b_j degrees of freedom.

Consequently, the approximated $AARL_{IN,j}\big(m, n, \alpha(m,n)\big)$ can be calculated as

$$
\begin{aligned}
&AARL_{IN,j}\big(m, n, \alpha(m,n)\big)\\
&= \int_0^\infty\left[1 - F_{\chi^2_{b_j}}\left(\chi^2_{1-\alpha/2,b_j}\frac{a_{0j}^2u}{b_{0j}}\right) + F_{\chi^2_{b_j}}\left(\chi^2_{\alpha/2,b_j}\frac{a_{0j}^2u}{b_{0j}}\right)\right]^{-1} g_{\chi^2_{b_{0j}}}(u)du
\end{aligned}
\tag{2.7}
$$

where $g_{\chi^2_{b_{0j}}}$ represents the probability density function (*pdf*) of U_{0j} a chi-square variable with b_{0j} degrees of freedom.

The next step is to determine an analytical expression for $\alpha(m,n)$. In order to do this, we consider a first order Taylor approximation of $AARL_{IN,j}\big(m,n,\alpha(m,n)\big)$, around $\alpha_0 = \frac{1}{ARL_0}$, where α_0 is the nominal *FAR* as before. This gives the approximation

$$AARL_{IN,j}\big(m,n,\alpha(m,n)\big) = AARL_{IN,j}(m,n,\alpha_0) + (\alpha(m,n)-\alpha_0)\frac{dAARL_{IN,j}(m,n,\alpha=\alpha_0)}{d\alpha} \quad (2.8)$$

Since we want $AARL_{IN,j}\big(m,n,\alpha(m,n)\big) = ARL_0$, which equals $\frac{1}{\alpha_0}$, we solve

$$\frac{1}{\alpha_0} = AARL_{IN,j}(m,n,\alpha=\alpha_0) + (\alpha(m,n)-\alpha_0)\frac{dAARL_{IN,j}(m,n,\alpha=\alpha_0)}{d\alpha}$$

(2.9)

for $\alpha\big(m,n\big)$. This yields the approximation

$$\alpha(m,n) = \frac{1/\alpha_0 - AARL_{IN,j}(m,n,\alpha=\alpha_0)}{\frac{d\left[AARL_{IN,j}(m,n,\alpha=\alpha_0)\right]}{d\alpha}} + \alpha_0 \quad (2.10)$$

The next step is to determine $\frac{d\left[AARL_{IN,j}(m,n,\alpha)\right]}{d\alpha}$.

From the obtained equation for $AARL_{IN,j}(m,n,\alpha)$ it follows that, in order to find its derivative, we need the results

$$\frac{d\chi^2_{1-\frac{\alpha}{2},b}}{d\alpha} = -\left[2g_{\chi^2_b}\left(\chi^2_{1-\frac{\alpha}{2},b}\right)\right]^{-1} \text{ and } \frac{d\chi^2_{\frac{\alpha}{2},b}}{d\alpha} = -\left[2g_{\chi^2_b}\left(\chi^2_{\frac{\alpha}{2},b}\right)\right]^{-1}.$$

These are obtained using the fact that $[F^{-1}]'(x) = [F'(F^{-1}(x))]^{-1}$, where F and F' denote the *cdf* of a continuous random variable and its derivative (the *pdf*), respectively; F^{-1}denotes the inverse of the cdf F and $[F^{-1}]'$denotes the first derivative of F^{-1} (cf. Gibbons and Chakraborti (2010)). Thus we find

$$\frac{d\left[AARL_{IN,j}(m,n,\alpha)\right]}{d\alpha}$$
$$= \int_0^\infty -\left[1 - F_{\chi^2_{b_j}}\left(\chi^2_{1-\frac{\alpha}{2},b_j}\frac{a_{0j}^2 u}{b_{0j}}\right) + F_{\chi^2_{b_j}}\left(\chi^2_{\frac{\alpha}{2},b_j}\frac{a_{0j}^2 u}{b_{0j}}\right)\right]^{-2} A\, g_{\chi^2_{b_{0j}}}(u)du$$

where (2.11)

$$A = \left[\frac{g_{\chi^2_{b_j}}\left(\chi^2_{1-\alpha/2,b_j}\frac{a_{0j}^2 u}{b_{0j}}\right)}{2g_{\chi^2_{b_j}}\left(\chi^2_{1-\alpha/2,b_j}\right)} + \frac{g_{\chi^2_{b_j}}\left(\chi^2_{\alpha/2,b_j}\frac{a_{0j}^2 u}{b_{0j}}\right)}{2g_{\chi^2_{b_j}}\left(\chi^2_{\alpha/2,b_j}\right)} \right] \frac{a_{0j}^2 u}{b_{0j}}.$$

With this result we have all the pieces required to calculate an approximation to $\alpha(m,n)$ from equation (2.10). Once $\alpha(m,n)$ is found, we can again use it to correct the charting constants for the Montgomery (2013) probability limits given in Section 3.3.2 of this chapter. The approximate values of $\alpha(m,n)$, $G_{n,\alpha(m,n),j}$ and $H_{n,\alpha(m,n),j}$, for each chart ($j = 1,2,3$), for different combinations of values of m, n and $ARL_0 = 370$ and 500 values are tabulated in Table 2.3.

Note that this approximate result is more general than the provided numerical solutions. In fact, it can be generalized to any combination of Phase I and Phase II estimators. This can be done by determining the required constants a, b, a_0 and b_0 based on the Patnaik (1950) approximation, as described in steps (i) and (ii) of our approach. Moreover, any monotonic increasing function $g(\sigma)$ of σ can be considered, since in that case $P(\widehat{LCL} < T_j < \widehat{UCL})$ is equivalent to $P(g(\widehat{LCL}) < g(T_j) < g(\widehat{UCL}))$. Hence, our approach can also be applied to S^2 and log(S) charts.

Comparing the analytical solutions with the numerical solutions it is seen that the approximations from the analytical method are quite accurate and the accuracy increases for higher values of m, as is desirable.

In order to compare the charting constants obtained by the numerical and the analytical methods, we calculated the $AARL_{IN}$ values for each chart, and $ARL_0 = 370$; $n = 5$ and for various values of m. Table 2.4 shows the results including the PD values relative to 370. As expected, for the numerically calculated probability limits, the $AARL_{IN}$ values are exactly equal to the nominal value 370. On the other hand, it can be seen that for the analytically calculated probability limits, except for $m = 5$, the $AARL_{IN}$ values are not more than 6% above the nominal value 370. It can also be seen that as m increases, the $AARL_{IN}$ values corresponding to the analytical constants converge quickly to 370. This shows that the behavior of the numerically and analytically corrected probability limits is similar.

Table 2.4: *The $AARL_{IN}$ and the PD values for the R and S charts with the charting constants calculated analytically (ANA) and numerically (NUM) for $n = 5$, $ARL_0 = 370$ and various values of m.*

	R chart with $\bar{R}/d_2(n)$				S chart with $\bar{S}/c_4(n)$				S chart with S_p			
m	$AARL_{IN}$ **NUM**	**PD**	$AARL_{IN}$ **ANA**	**PD**	$AARL_{IN}$ **NUM**	**PD**	$AARL_{IN}$ **ANA**	**PD**	$AARL_{IN}$ **NUM**	**PD**	$AARL_{IN}$ **ANA**	**PD**
5	370	0	426	15	370	0	431	17	370	0	444	20
10	370	0	388	5	370	0	390	5	370	0	393	6
20	370	0	375	2	370	0	376	2	370	0	377	2
25	370	0	373	1	370	0	374	1	370	0	375	1
30	370	0	372	1	370	0	373	1	370	0	374	1
50	370	0	371	0	370	0	371	0	370	0	372	0
100	370	0	370	0	370	0	370	0	370	0	370	0
500	370	0	370	0	370	0	370	0	370	0	370	0

2.4 Out of control performance

The numerical control limits provided here guarantee that the in-control average run length of the charts is equal to the nominal value of 370 or 500. However, since the corrected limits are wider than the uncorrected probability limits, it is of interest to see whether the correction impacts the out-of-control performance. It may be noted at the outset that such a comparison is not really fair since the in-control performance of the uncorrected limits can be far worse than the nominal.

In order to make this comparison, we compute the ARL for the considered dispersion charts with $n = 5$, for a number of values of the ratio (κ) between the Phase II standard deviation (σ) and the in-control process standard deviation (σ_0), that is, for $\kappa = \sigma/\sigma_0$. In other words, we compute points of the ARL profiles of the charts in different cases, where $\kappa = 1$ corresponds to the $AARL_{IN}$, and $\kappa \neq 1$ corresponds to the out-of-

control ARL ($AARL_{OOC}$). This is done for several values of m. The results are given in Table 8 of Diko et al. (2017).

The ARL's with the uncorrected and the corrected limits could be easily computed from (2.3) by just replacing $CFAR_j$ by the general conditional probability of an alarm, CPA_j. Next, using $\kappa = \sigma/\sigma_0$, we calculate CPA_j for the corrected limits as

$$\begin{aligned} CPA_j &= 1 - P\left(H_{n,\alpha,j}\frac{e_j}{\sigma} < \frac{T_{ij}}{\sigma} < G_{n,\alpha,j}\frac{e_j}{\sigma}\right) \\ &= 1 - P\left(H_{n,\alpha,j}\frac{\epsilon_{0j}a_{0j}\sqrt{U_{0j}}}{\kappa\sqrt{b_{0j}}} < \frac{T_{ij}}{\sigma} < G_{n,\alpha,j}\frac{\epsilon_{0j}a_{0j}\sqrt{U_{0j}}}{\kappa\sqrt{b_{0j}}}\right) \\ &= CFAR_j(U_{0j}, m, n, \alpha, \kappa) \end{aligned}$$

Using the known distributions of T_{ij}/σ this gives

$$CPA_1(U_{01}, m, n, \alpha, \kappa) = 1 - F_{W_n}\left(G_{n,\alpha,1}\frac{e_1}{\kappa\sigma_0}\right) + F_{W_n}\left(H_{n,\alpha,1}\frac{e_1}{\kappa\sigma_0}\right)$$

$$= 1 - F_{W_n}\left(F_{W_{n,1-\alpha/2}}\frac{a_{01}\sqrt{U_{01}}}{\kappa\sqrt{b_{01}}}\right) + F_{W_n}\left(F_{W_{n,\alpha/2}}\frac{a_{01}\sqrt{U_{01}}}{\kappa\sqrt{b_{01}}}\right),$$

$$CPA_2(U_{02}, m, n, \alpha, \kappa) = 1 - F_{\chi^2_{n-1}}\left(\left(G_{n,\alpha,2}\frac{e_2\sqrt{n-1}}{\kappa\sigma_0}\right)^2\right) + F_{\chi^2_{n-1}}\left(\left(H_{n,\alpha,2}\frac{e_2\sqrt{n-1}}{\kappa\sigma_0}\right)^2\right)$$

$$= 1 - F_{\chi^2_{n-1}}\left(\chi^2_{1-\alpha/2,n-1}\frac{a_{02}^2U_{02}}{\kappa^2 b_{02}}\right) + F_{\chi^2_{n-1}}\left(\chi^2_{\alpha/2,n-1}\frac{a_{02}^2U_{02}}{\kappa^2 b_{02}}\right)$$

and

$$CPA_3(U_{03}, m, n, \alpha, \kappa) = 1 - F_{\chi^2_{n-1}}\left(\left(G_{n,\alpha,3}\frac{e_3\sqrt{n-1}}{\kappa\sigma_0}\right)^2\right) + F_{\chi^2_{n-1}}\left(\left(H_{n,\alpha,3}\frac{e_3\sqrt{n-1}}{\kappa\sigma_0}\right)^2\right)$$

$$= 1 - F_{\chi^2_{n-1}}\left(\chi^2_{1-\alpha/2,n-1}\frac{a_{03}^2U_{03}}{\kappa^2 b_{03}}\right) + F_{\chi^2_{n-1}}\left(\chi^2_{\alpha/2,n-1}\frac{a_{02}^2U_{03}}{\kappa^2 b_{03}}\right).$$

These values can in turn be used as described in the numerical approach in section 2.3.1, to determine the unconditional ARL as

$$AARL_j(m, n, \alpha, \kappa) = \int_0^\infty \left[CPA_j(u, m, n, \alpha, \kappa)\right]^{-1} g_{\chi^2_{b_{0j}}}(u)dx$$

where again u is the value of the random variable U_{0j}. Formulae for the CPA_j and $AARL_j$ of the uncorrected limits are the same as above, except that α_0 is used instead of $\alpha(m,n)$.

Table 8 in Diko et al. (2017) shows the $AARL_{IN}$ $(\kappa = 1)$, the $AARL_{OOC}$ $(\kappa \neq 1)$ and the PD values associated with the uncorrected and the corrected estimated probability limits based *R* and *S* charts, for $n = 5, 10$ and various values of λ and m. The PD values measure the percentage difference between the unconditional ARL values for the estimated σ_0 case and the nominal unconditional ARL values. Note that the results for the uncorrected estimated probability limits have been thoroughly discussed by Chen (1998). From Table 8 in Diko et al. (2017) it can be seen that when the process is IC and σ_0 is estimated, using the uncorrected charting constants to construct the uncorrected probability limits gives unconditional ARL values that are up to 29% lower than the nominal 370 (corresponding to the σ_0 known case) for $n = 5$ and 32% lower than the nominal 370 (corresponding to the σ_0 known case) for $n = 10$, respectively. This means a lot of false alarms. Using the corrected (new/adjusted) charting constants to construct the probability limits yields the nominal value 370, which is desirable. However, this also leads to larger unconditional ARL values for the corrected charts compared to the uncorrected charts when the process is OOC. Interestingly, this difference is smaller for decreases in variability $(\kappa < 1)$ than for increases $(\kappa > 1)$. In general, both the corrected and uncorrected charts have more difficulty in detecting decreases in variability than increases. It can also be seen that the effect of using either the corrected or uncorrected estimated probability limits is a function of m. In general, increasing m diminishes the effects of parameter estimation on both the IC and OOC unconditional ARL performance for both uncorrected and corrected probability limits, as expected.

To summarize, the corrected estimated probability limits provide a much better IC performance than the uncorrected limits, as it yields on average the nominally specified ARL_{IN} performance. However, this generally comes with a deterioration of the ARL_{OOC} performance relative to the uncorrected limits. Note that this tradeoff between IC and OOC performance can be altered by adjusting the value of ARL_0 .

2.5 Summary and conclusions

Shewhart control charts are often used to monitor process dispersion. However, the standard versions of these charts assume known in-control parameters, which is typically not the case in practice. When the parameters are estimated to set up the control limits, both the IC and OOC performance of the control charts are affected (Chen, 1998). In this chapter, based on the unconditional approach, we have corrected the control limits of the R and S charts to account for the effects of parameter estimation. Two methods are used to find the corrected charting constants. The first method, the numerical approach, involves numerical integration and solving non-linear equations. The second method, the analytical approach is based on a first order Taylor approximation to the $AARL_{IN}$. Differences in the values obtained with these two methods are small, indicating that the analytical approximations are quite accurate. However, the analytical approach is more general in the sense that it can be applied to any desired estimator. Extensions to other functions of S, such as S^2 or log-S are straightforward.

The tabulated constants provided here ensure that the $AARL_{IN}$ is equal to a pre-specified desired value, taking into account the estimators that are used, the number of Phase I subgroups (m) and the subgroup size (n). However, this IC robustness is achieved at the price of a deterioration (increase) in the $AARL_{OOC}$. This deterioration, due to the use of the corrected limits, is negligible for large values of m or large changes of variability.

In conclusion, this chapter provides the correct charting constants for the popular dispersion charts, for independent and identically distributed normal data, properly accounting for the effects of parameter estimation, in terms of a specified nominal value of the unconditional in-control average run length. Note that in practice, it is possible that the data do not follow a normal distribution. How these corrected limits perform for other distributions and their required modifications requires further investigation, which will be considered elsewhere. A similar study, for the important case when $n = 1$, is required.

Appendix 2.A. *Formulae for the various constants that are associated with the three spread charts that are considered in this paper.*

	R chart with $\overline{R}/d_2(n)$ (*j =1*)	S chart with $\overline{S}/c_4(n)$ (*j =2*)	S chart with S_p (*j =3*)
Values of the biased estimator e_j	$e_1 = \overline{R}$	$e_2 = \overline{S}$	$e_3 = S_p$
Values of the unbiasing constants ϵ_{0j} **and** ϵ_j	$\epsilon_{01} = d_2(n)$	$\epsilon_{02} = c_4(n)$	$\epsilon_{03} = c_4(m(n-1)+1) \approx 1$
	$\epsilon_1 = d_2(n)$	$\epsilon_2 = c_4(n)$	$\epsilon_3 = c_4(n)$
Values of the unbiased estimator $\hat{\sigma}_{0j}$	$\frac{e_1}{\epsilon_{01}}$	$\frac{e_2}{\epsilon_{02}}$	$\frac{e_3}{\epsilon_{03}}$
Charting statistic T_{ij}	$R \sim W\sigma$	$S \sim \frac{\sqrt{\chi^2_{n-1}}}{\sqrt{n-1}}\sigma$	$S \sim \frac{\sqrt{\chi^2_{n-1}}}{\sqrt{n-1}}\sigma$
Uncorrected probability limits	$H_{n,\alpha,1} = \frac{F_{W_{\alpha/2}}}{\epsilon_{01}}$	$H_{n,\alpha,2} = \frac{\sqrt{\chi^2_{\alpha/2,n-1}}}{\epsilon_{02}\sqrt{n-1}}$	$H_{n,\alpha,3} = \frac{\sqrt{\chi^2_{\alpha/2,n-1}}}{\epsilon_{03}\sqrt{n-1}}$
	$G_{n,\alpha,1} = \frac{F_{W_{1-\alpha/2}}}{\epsilon_{01}}$	$G_{n,\alpha,2} = \frac{\sqrt{\chi^2_{1-\alpha/2,n-1}}}{\epsilon_{02}\sqrt{n-1}}$	$G_{n,\alpha,3} = \frac{\sqrt{\chi^2_{1-\alpha/2,n-1}}}{\epsilon_{03}\sqrt{n-1}}$
3-sigma limits (cf. Montgomery (2013))	$H_{n,\alpha,1} = \left(1 - 3\frac{d_3(n)}{\epsilon_1}\right)$	$H_{n,\alpha,2} = \left(1 - 3\frac{\sqrt{1-c_4^2(n)}}{\epsilon_2}\right)$	$H_{n,\alpha,3} = \left(1 - 3\frac{\sqrt{1-c_4^2(n)}}{\epsilon_3}\right)$
	$G_{n,\alpha,1} = \left(1 + 3\frac{d_3(n)}{\epsilon_1}\right)$	$G_{n,\alpha,2} = \left(1 + 3\frac{\sqrt{1-c_4^2(n)}}{\epsilon_2}\right)$	$G_{n,\alpha,3} = \left(1 + 3\frac{\sqrt{1-c_4^2(n)}}{\epsilon_3}\right)$
Chi approximation of T_{ij}	$R \sim \frac{\epsilon_1 a_1\sqrt{\chi^2_{b_1}}}{\sqrt{b_1}}\sigma$	$S \sim \frac{\epsilon_2 a_2\sqrt{\chi^2_{b_2}}}{\sqrt{b_2}}\sigma$	$S \sim \frac{\epsilon_3 a_3\sqrt{\chi^2_{b_3}}}{\sqrt{b_3}}\sigma$
Chartings Constants $H_{n,\alpha,j}$; $G_{n,\alpha,j}$ **Based on the chi approximation of** T_{ij}	$H_{n,\alpha,1} = \frac{\epsilon_1 a_1\sqrt{\chi^2_{\alpha/2,b_1}}}{\epsilon_{01}\sqrt{b_1}}$	$H_{n,\alpha,2} = \frac{\epsilon_2 a_2\sqrt{\chi^2_{\alpha/2,b_2}}}{\epsilon_{02}\sqrt{b_2}}$	$H_{n,\alpha,3} = \frac{\epsilon_3 a_3\sqrt{\chi^2_{\alpha/2,b_3}}}{\epsilon_{03}\sqrt{b_3}}$
	$G_{n,\alpha,1} = \frac{\epsilon_1 a_1\sqrt{\chi^2_{1-\alpha/2,b_1}}}{\epsilon_{01}\sqrt{b_1}}$	$G_{n,\alpha,2} = \frac{\epsilon_2 a_2\sqrt{\chi^2_{1-\alpha/2,b_2}}}{\epsilon_{02}\sqrt{b_2}}$	$G_{n,\alpha,3} = \frac{\epsilon_3 a_3\sqrt{\chi^2_{1-\alpha/2,b_3}}}{\epsilon_{03}\sqrt{b_3}}$
Variance (V_{0j}) **of standardized unbiased estimator** $\hat{\sigma}_{0j}/\sigma$	$V_{01} = \mathrm{Var}\left(\frac{\overline{R}}{d_2(n)\sigma}\right) = \frac{d_3^2(n)}{md_2^2(n)}$	$V_{02} = \mathrm{Var}\left(\frac{\bar{s}}{c_4(n)\sigma}\right) = \frac{1-c_4^2(n)}{mc_4^2(n)}$	$V_{03} \approx \mathrm{Var}\left(\frac{S_p}{\sigma}\right) = 1 - c_4^2(m(n-1)+1)$

Patnaik and Chen r_{0j} and t_{0j}	$r_{01} = \left(-2 + 2\sqrt{1 + 2V_{01}}\right)^{-1}$ $t_{01} = V_1 + \frac{1}{16r_{01}^3}$	$r_{02} = \left(-2 + 2\sqrt{1 + 2V_{02}}\right)^{-1}$ $t_{01} = V_2 + \frac{1}{16r_{02}^3}$	Not required (distribution is exact)
Value of b_{0j}	$b_{01} = \left(-2 + 2\sqrt{1 + 2V_{01}}\right)^{-1}$	$b_{02} = \left(-2 + 2\sqrt{1 + 2V_{02}}\right)^{-1}$	$b_{03} = m(n-1)$
Value of a_{0j}	$a_{01} = 1 + \frac{1}{4b_{01}} + \frac{1}{32b_{01}^2} - \frac{5}{128b_{01}^3}$	$a_{02} = 1 + \frac{1}{4b_{02}} + \frac{1}{32b_{02}^2} - \frac{5}{128b_{02}^3}$	$a_{03} = 1$
Variance (V_j) of standardized charting statistic $T_j/(\epsilon_j\sigma)$	$V_1 = \mathrm{Var}\left(\frac{R}{d_2(n)\sigma}\right) = \frac{d_3^2}{d_2^2}$	$V_2 = \mathrm{Var}\left(\frac{S}{c_4(n)\sigma}\right) = \frac{1 - c_4^2(n)}{c_4^2(n)}$	$V_3 = \mathrm{Var}\left(\frac{S}{c_4(n)\sigma}\right) = \frac{1 - c_4^2(n)}{c_4^2(n)}$
Patnaik and Chen r_j and t_j	$r_1 = \left(-2 + 2\sqrt{1 + 2V_1}\right)^{-1}$ $t_1 = V_1 + \frac{1}{16r_1^3}$	Not required (distribution is exact)	Not required (distribution is exact)
Value of b_j	$b_1 = \left(-2 + 2\sqrt{1 + 2V_1}\right)^{-1}$	$b_2 = n - 1$	$b_3 = n - 1$
Value of a_j	$a_1 = 1 + \frac{1}{4b_1} + \frac{1}{32b_1^2} - \frac{5}{128b_1^3}$	$a_2 = 1/c_4(n)$	$a_3 = 1/c_4(n)$

$$c_4(n) = \left(\frac{2}{n-1}\right)^{1/2} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}$$

Let $W = \frac{R_i}{\sigma}$ and F_{W_n} denote the sample relative range and its distribution function, respectively, then

$$d_2(n) = E(W) = \int_{-\infty}^{\infty} \left(1 - F_{W_n}(w)\right) dw$$

$$E\left(W^2\right)=\int_0^\infty\left(1-F_{W_n}\left(w\right)\right)dw^2=2\int_0^\infty w\left(1-F_{W_n}\left(w\right)\right)dw$$

$$d_3(n)=\sqrt{Var(W)}=\sqrt{E\left(W^2\right)-d_2^2(n)}$$

3. Guaranteed In-Control Performance of the EWMA Chart for monitoring the Mean

3.1 Introduction

Jones et al. (2001) studied the conditional and the unconditional run length distribution of the EWMA chart with estimated parameters in both the in-control (IC) and the out-of-control (OOC) cases. Based on the percentage increase in the false alarm rate (FAR), they concluded that when parameters are estimated and the smoothing constant λ is small, larger Phase I sample sizes are needed, to design charts with acceptable FAR performance. However, their study did not take into account the random variability of the FAR, the so-called "practitioner to practitioner" variability, which is inherent to parameter estimation. Motivated by this, Saleh et al. (2015) examined the conditional in-control average run length ($CARL_{IN}$) distribution of the EWMA chart as a function of the number of Phase I subgroups (m), subgroup size (n) and λ. Based on the standard deviation $CARL_{IN}$ (denoted by $SDCARL_{IN}$), they concluded, contrary to Jones et al. (2001), that, much larger Phase I sample sizes are required to design Phase II EWMA charts with larger λ than with smaller λ.

In his work on prospective application of the Phase II $\overline{X}$ chart, Chakraborti (2006) highlight the variation present in the conditional run length distribution and hence the importance of examining the practitioner to practitioner variability via the conditional run length distribution. He emphasized how the conditional false alarm rate ($CFAR$) behaves as a random variable when parameters are estimated and used to construct Phase II charts. Inspired by this, for the Phase II S and S^2 charts, Epprecht et al. (2015) examined the $CFAR$ distribution as a function of the Phase I sample size mn. They then made recommendations about the minimum size of the Phase I sample, which is required, to guarantee, with a high probability $1-p$, that the $CFAR$ will not exceed some specified nominal $CFAR$ value (denoted $CFAR_0$). This is the exceedance probability criterion (EPC) that was introduced by Albers and Kallenberg (2004a,2004b) and Gandy and Kvaloy (2013), which sets an upper prediction bound to $CFAR$. In the same spirit, we examine the $CARL_{IN}$ distribution of the

EWMA chart as a function of λ, m, n, and set a lower prediction bound to $CARL_{IN}$. We then make recommendations about the value of m, which is required, to guarantee, with a specified high probability $1-p$, that the $CARL_{IN}$ will exceed a nominally specified value, denoted ARL_0. Our results reveal that in-order for the EWMA chart to meet the EPC specification, even more Phase I data are needed than was previously recommended by Saleh et al. (2015) and Jones et al. (2001). Moreover, consistently with Jones et al. (2001) but contrary to Saleh et al. (2015) it will be seen that small values of λ require larger Phase I sample sizes than large values of λ

However, in practice, it may be difficult and expensive to get such huge amounts of Phase I data. Hence, control limits are adjusted (corrected) as a function of the amount of data available at hand. Jones (2002) adjusted the control limits of the Phase II EWMA chart using the unconditional approach. Values of the charting constant (L) were given graphically for different values of ARL_0, m, n and choice of λ ranging from 0.02 to 1. Saleh et al. (2015) used the EPC and bootstrapping (i.e. the conditional approach) to design the EWMA chart when parameters are estimated. However, bootstrapping is computer intensive and may be somewhat difficult to apply in practice. This is also exacerbated by the fact that, even though the underlying problem and the chart performance specifications may be the same, repeated applications of the bootstrap approach would almost surely result in different adjusted limits and can lead to comparability issues. Hence, it is not surprising that, with the exception of Faraz et al. (2016) and Hu and Castagliola (2017) , the authors who have used the bootstrap approach did not provide or show tables of their new charting constants. Each of the Hu and Castagliola (2017) charting constants, was found by running the bootstrap approach 100 times and averaging the results. On an average computer, this takes a lot of time. Hence, without these tables, coming up with the charting constant can be frustrating for a practitioner.

Under the assumption that the process output is normally distributed, bootstrapping is not necessary to apply the EPC. For example, Goedhart et al. (2017a, 2017b, 2018) provided analytical results in the form of numerical solutions and approximations for the Shewhart charts for the mean, and provided tables for the charting constants. But, for the EWMA chart, such analytical approximations are difficult to obtain because the charting statistics are dependent. Consequently, in this chapter, we present a different method of

adjusting the Phase II control limits according to the EPC, that guarantees, with a specified high confidence, that the $CARL_{IN}$ of the EWMA chart exceeds a nominal ARL_0. Our approach is based on the simple idea of approximating the $CARL_{IN}$ distribution by an empirical distribution, which is obtained by generating many Phase I subgroups, and using the Markov Chain to calculate the corresponding $CARL_{IN}$ values. It will be seen that this approach requires less computational effort than the bootstrap approach, yet it produces results that are as accurate as some known analytical results. Thus, tables and graphs of the required charting constants are provided to help practitioners implement the EWMA chart with estimated parameters easily in practice.

This chapter has been based on Diko et al. (2019a) and is organized as follows. Section 3.2 introduces some notation and terminology used, gives an overview of the EWMA chart, the Markov Chain technique and presents the estimators that are used to estimate the unknown process parameters. Section 3.3 evaluates the traditional EWMA chart in terms of the EPC and provides rough guidelines on the number of Phase I subgroups required to achieve a certain high proportion of high $CARL_{IN}$ values relative to a reasonable nominal value. Section 3.4 presents the new charting constants (adjusted control limits) so that the Phase II EWMA chart has a guaranteed nominal IC performance according to the EPC. Section 3.5 gives a detailed evaluation of the IC and OOC performance of the new constants (the EPC adjusted limits based Phase II EWMA chart) and compares it with the performance of the traditional Phase II EWMA chart with unadjusted limits (limits calculated for Case K) according to the EPC. Finally, a summary and some conclusions are given.

3.2 EWMA chart with estimated parameters

Let $X_{ij}, i = 1,2,...,m$ and $j = 1,2,...,n$ denote the *IC* Phase I data from a normal distribution with an unknown mean μ_0 and an unknown standard deviation σ_0. For a smoothing constant $0 < \lambda \leq 1$, starting at sampling stage $i = m+1, m+2,...$, the standardized plotting statistic for the Phase II EWMA chart with the estimated parameters is given by

$$Y_i = \lambda B_i + (1-\lambda)Y_{i-1} \tag{3.1}$$

Where $Y_m = 0$, $B_i = \frac{\bar{X}_i - \hat{\mu}_0}{\frac{\hat{\sigma}_0}{\sqrt{n}}}$, $\bar{X}_i$ is the i^{th} Phase II sample mean, and $\hat{\mu}_0$ and $\hat{\sigma}_0$ are the Phase I estimators of the unknown parameters μ_0 and σ_0, respectively. It is also assumed that the Phase II data are normally distributed, and for generality, let μ and σ denote the mean and the standard deviation, respectively, of this distribution. In this chapter we use the estimators $\hat{\mu}_0 = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}$, the grand sample mean (see Schoonhoven et al. (2011)), and $\hat{\sigma}_0 = \sqrt{\frac{1}{m}\sum_{i=1}^{m} S_i^2} = S_p$, the pooled standard deviation estimator, where S_i^2 denotes the variance of the i^{th} Phase I sample. Among the commonly used estimators for σ_0, the pooled standard deviation estimator provides the lowest values of the mean squared error (Mahmoud et al. (2010)). In addition, as noted in chapter 2 and in Diko et al. (2017), the corresponding unbiased version $\hat{\sigma}_0 = S_p / c_4\left(m(n-1)+1\right)$ (see Montgomery (2013), Schoonhoven et al. (2009, 2011)) is equivalent, because, $m\left(n-1\right)$ is typically quite large in our applications and hence the constant $c_4\left(m\left(n-1\right)+1\right)$ is almost indistinguishable from 1.

We write the statistic B_i in its canonical form

$$B_i = \frac{1}{Q}\left(\kappa T_i + \frac{\sqrt{n}}{\sigma_0}\delta - \frac{Z}{\sqrt{m}}\right) \tag{3.2}$$

where $T_i = \frac{\bar{X}_i - \mu}{\sigma/\sqrt{n}}$, $Q = \frac{S_p}{\sigma_0}$, $Z = \frac{\hat{\mu}_0 - \mu_0}{\sigma_0/\sqrt{mn}}$, $\kappa = \frac{\sigma}{\sigma_0}$ and $\delta = \frac{\mu - \mu_0}{\sigma_0}$. Note that the random variables T_i and Z are independent standard normal variables that are mutually independent and are also independent of Q. Since $m(n-1)S_p^2 / \sigma_0^2 \sim \chi^2_{m(n-1)}$, Q is distributed as $\sqrt{\frac{\chi^2_{m(n-1)}}{m\left(n-1\right)}}$. Furthermore, like in Saleh et al. (2015a), we assume that $\sigma_0/\sqrt{n} = 1$. For simplicity, we use the asymptotic (steady state) control limits

$$h = h\left(n, \lambda, L\right) = \pm L\sqrt{\frac{\lambda}{\left(2-\lambda\right)}}. \tag{3.3}$$

where L is the charting constant to be found. For example, for a given value of λ and a nominal ARL_0, assuming that the parameters are known, the L values can be found in Crowder (1989) or by using the function *xewma.crit(λ,ARL_0,sided="two")* from the R-package *spc*. Often, these parameters known L values are used to construct the Phase II EWMA chart when estimated parameters are used in the control limits. It is recognized in the literature (cf. Jensen et al. (2016)) that this is a problem in the sense of getting many more false alarms than nominally expected, particularly when the amount of Phase I data is small to moderately large. We provide some solutions for correcting this problem.

The conditional run length distribution and the $CARL$ of an EWMA chart may be calculated (approximated) using the Markov Chain method, see Brook and Evans (1972) and Lucas and Saccucci (1990), among others. Applying the Markov Chain method, conditionally on Q and Z, the $CARL$ of the Phase II EWMA chart can be conveniently written as

$$\begin{aligned} & CARL \\ & = \underset{\sim}{v}'\left(I-P\right)^{-1}\underset{\sim}{u} \\ & = CARL\left(\delta,Q,Z,m,n,\lambda,L,t\right), \end{aligned} \tag{3.4}$$

where t (which is generally taken as an odd integer) represents the number of transient states in the state space of a Markov Chain, $\underset{\sim}{v}'$ is the $1\times t$ row vector with one in the middle position (for an odd integer t, the middle position is unique) and 0 elsewhere, $\underset{\sim}{u}$ is a $t\times 1$ column vector of ones, I is the $t\times t$ identity matrix, $P=\left[p_{lk}\right]$ is the $t\times t$ "essential" (conditional) transition probability matrix and $l,k=-\frac{t-1}{2},...,0,...,\frac{t-1}{2}$.

The transition probabilities of the essential conditional transition probability matrix, p_{lk}, are calculated, under normality and conditional on Q and Z, as follows.

$$\begin{aligned} p_{lk} = {} & \Phi\left(Q\left(\frac{S_k+w/2-(1-\lambda)S_l}{\lambda}\right)-\delta+\frac{Z}{\sqrt{m}}\right) \\ & -\Phi\left(Q\left(\frac{S_k-w/2-(1-\lambda)S_l}{\lambda}\right)-\delta+\frac{Z}{\sqrt{m}}\right) \\ & = p_{lk}(\delta,Q,Z,m.n,\lambda,L,t) \end{aligned} \tag{3.5}$$

where Φ denotes the cumulative distribution function of a standard normal variable, $w = \frac{2h}{t} = w(n, \lambda, L, t)$, $S_f = -\frac{w}{2} + \left(2\left(\frac{t-1}{2} + f\right) + 1\right)\frac{w}{2} = S_f(n, \lambda, L, t)$ and $f = l, k$. More information on the derivation of result (3.5) can be found in Saleh et al. (2013).

From equation (3.4), for fixed m, n, δ, λ and L, it is clearly seen that the $CARL$ depends on the random variables Q and Z and hence the $CARL$ is a random variable. Saleh et al. (2015) studied the effect of m and Phase I estimates on the distribution of $CARL_{IN}$ (the $CARL$ when $\sigma = 0$). They found that unless the Phase I parameter estimates are "close" to the true but unknown parameter values, the $CARL_{IN}$ values can vary widely and deviate substantially from the nominal ARL_0. However, a practitioner will almost never know where his/her estimates are in relation to the unknown process parameters. Thus, when parameters are estimated, using the charting constants for Case K to design Phase II EWMA charts is a risky proposition, since it can result in very low $CARL_{IN}$ values which will almost surely call into question the process monitoring regime. This risk can be somewhat reduced by increasing m. However, as will be seen in the next section, the value of m that is required to reduce the probability of low $CARL_{IN}$ values can be very large. Hence, many control charts in the recent literature with estimated parameters are now designed such that

$$P(CARL_{IN} > ARL_0) = 1 - p. \tag{3.6}$$

It follows that the ARL_0 is the $100p^{th}$ percentile value of the distribution of $CARL_{IN}$. This is the *EPC* approach that we use to evaluate and design the EWMA chart in the following sections.

3.3 Performance assessment of a standard Phase II EWMA chart using the *EPC*

Consider again the *EPC* given in equation (3.6), which can be re-written as

$$P(CARL_{IN}(Q, Z, m, n, \lambda, L, t) \leq ARL_0) = p$$

Thus for a given $p \in (0,1)$ and m, n, λ, L, t, we want to find the 100 p^{th} percentile, $CARL_{IN,p}$, of the distribution of $CARL_{IN}(Q, Z, m, n, \lambda, L, t)$. Once found, $CARL_{IN,p}$ is compared to ARL_0, which is the theoretical value that must be exceeded, in an application, with a high probability $1-p$. The comparison between the $CARL_{IN,p}$ and ARL_0 will be based on the percentage difference (PD), which we define as $PD = \frac{CARL_{IN,p} - ARL_0}{ARL_0} \times 100$. The algorithm for the evaluation of the traditional Phase II EWMA using the EPC is given in Diko et al. (2019a).

Table 1 of Diko et al. (2019a) shows the $CARL_{IN,p}$ values of the standard Phase II EWMA charts for $ARL_0 = 100, 200, 370, 500$; $n = 5$ and different combinations of λ, m and p. From this Table 1, it can be seen that when m is small, the *PD* values are very high in absolute values. For example, for $\lambda = 0.1, m = 30, p = 0.05$ and $ARL_0 = 500$, we have $CARL_{IN,p} = 50$, which is 90% below ($PD = -90\%$) the nominal $ARL_0 = 500$. Thus, in this case, we expect the $CARL_{IN}$ of the chart to be at least 50 with 95% probability (and conversely, the $CARL_{IN}$ of the chart to be at most 50 with 5% probability). Ideally, we would like the chart to deliver at least a large $CARL_{IN}$ value (say $ARL_0 = 500$) with 95% certainty. The value $CARL_{IN,p} = 50$ is too low and the risk of getting a number that low is very high. It can also be seen from Table 1 in Diko et al. (2019a) that when *m* increases, the $CARL_{IN,p}$ values increase to within *6%* less than the nominal ARL_0 values. The convergence is faster for $\lambda = 0.5$ than for $\lambda = 0.1$. Furthermore, it can be seen that larger values of p or/and λ are associated with larger $CARL_{IN,p}$ values, improving the results slightly. Thus, when parameters are estimated, small $CARL_{IN}$ values (i.e. the $CARL_{IN}$ values that are less than the ARL_0) occur more often than desired. This is not acceptable. Table 1 from Diko et al. (2019a) also allows us to make rough recommendations about the number of Phase I subgroups m required to achieve adequate Phase II EPC performance. These recommendations are summarized in Table 3.1 and they are compared to the Jones et al. (2001) recommendations (that were based on the ARL_{IN} criteria) and Saleh et al. (2015) recommendations (that were based on $SDCARL_{IN}$ criteria) in Table 3.2.

Table 3.1 shows the number of subgroups m required to guarantee that the $CARL_{IN}$ exceeds $CARL_{IN,p}$ by a certain specified high probability $(1-p)$. Mathematically this is written as

$$P\left(CARL_{IN} > CARL_{IN,p}\right) \geq 1 - p$$
$$P\left(CARL_{IN} > ARL_0(1-\varepsilon)\right) \geq 1 - p \quad (3.7)$$

where $\varepsilon \geq 0\%$ is a nominally specified PD value. Note that $\varepsilon \geq 0$ because in general $CARL_{IN,p} < ARL_0$ (see Table 3.1). Note also that if $\varepsilon = 0\%$, then $CARL_{IN,p} = ARL_0$ and therefore equation (3.7) reduces to equation (3.6).

Table 3.1 *Minimum m required for* $CARL_{IN,p}$ *to be* $\varepsilon = 0\%, 10\%, 20\%$ *below the nominally* $ARL_0 = 100, 200, 370, 500$ *for* $n = 5$; $\lambda = 0.1, 0.5$ *and* $p = 0.05, 0.10$

		$ARL_0 = 100$		$ARL_0 = 200$		$ARL_0 = 370$		$ARL_0 = 500$	
	ε	p=0.05	p=0.10	p=0.05	p=0.10	p=0.05	p=0.10	p=0.05	p=0.10
$\lambda = 0.1$	0%	>10000	>10000	>10000	>10000	>10000	>10000	>10000	>10000
	10%	10000	6000	6000	4000	6000	4000	6000	4000
	20%	900	600	1000	900	1500	900	1500	1000
$\lambda = 0.5$	0%	>10000	>10000	>10000	>10000	>10000	>10000	>10000	>10000
	10%	2000	1500	4000	2000	4000	2000	4000	2000
	20%	500	400	600	400	900	500	900	600

Looking at Table 3.1; for fixed ARL_0, λ and p; it can be seen that decreasing ε from 20% to 0% increases the number of Phase I subgroups m required to achieve adequate IC EPC performance. It can also be seen that; for fixed ARL_0, λ and ε; decreasing p from 0.10 to 0.05 increases the value of m. Thus, decreasing ε or p or both improves the IC chart performance, while increasing ε or p or both degrades the IC chart performance. This also shows the flexibility of the EPC formulation (equation (3.7)), which can be used to improve the IC chart performance or to balance it with the OOC chart performance by manipulating ε or p or both. Later, we will provide an example of how this balance can be achieved. Table 3.2 compares our recommendations with the Jones et al. (2001) and Saleh et al. (2015) recommendations.

From Table 3.2; it can be seen that for $p = 0.05,0.10$; $\varepsilon = 0\%$, $n = 5$ and all λ it will take more than $10{,}000$ Phase I subgroups to guarantee (with a high probability) that the nominal ARL_0 value will be exceeded. Thus, based on the EPC, it is seen that significantly more Phase I data are required than previously recommended by both Jones et

al. (2001) and Saleh et al. (2015). Furthermore; for the EPC approach; it can be seen that when $CARL_{IN,p}$ is $\varepsilon = 10\%$ or $\varepsilon = 20\%$ below the ARL_0; a large number of subgroups is still required to guarantee with high certainty that $CARL_{IN} > CARL_{IN,p}$. Moreover, small λ values require more data than larger λ values. This agrees with the findings of Jones et al. (2001), but it is in contrast with the findings of Saleh et al. (2015).

Table 3.2 *Recommended minimum number of Phase I subgroups when* $n = 5$ *and* $ARL_0 = 200$

	$\boldsymbol{\lambda = 0.1}$	$\boldsymbol{\lambda = 0.2}$	$\boldsymbol{\lambda = 0.5}$	$\boldsymbol{\lambda = 1}$
Jones et al. (2001) (marginal ARL criteria)	400	300	200	100
Saleh et al. (2015) $SDCARL_{IN}$ criteria	600	700	900	1000
This chapter EPC criteria with $p = 0.10$ and $\varepsilon = 20\%$	900	600	400	370
This chapter EPC criteria with $p = 0.05$ and $\varepsilon = 20\%$	1000	700	600	560
This chapter EPC criteria with $p = 0.10$ and $\varepsilon = 10\%$	4000	3000	2000	1500
This chapter EPC criteria with $p = 0.05$ and $\varepsilon = 10\%$	6000	5000	4000	2500
This chapter EPC criteria with $p = 0.05, 0.10$ and $\varepsilon = 0\%$	>10000	>10000	>10000	>10000

3.4 Adjustment of the standard Phase II EWMA chart limits for guaranteed conditional performance.

We have seen that, to achieve adequate EPC performance, a very high number of Phase I subgroups is required when using the standard Phase II EWMA chart limits. In practice it may be difficult and expensive to come up with these high Phase I subgroup numbers. Thus, for a given amount of Phase I data (number of Phase I subgroups, with a fixed sample size), the control limits need to be adjusted.

Consider again the *EPC*: $P\left(CARL_{IN}(Q,Z,m,n,\lambda,L,t) > ARL_0(1-\varepsilon)\right) \geq 1-p$, which is equivalent to stating that the *cdf* of $CARL_{IN}\left(Q,Z,m,n,\lambda,L,t\right)$ at ARL_0 must be less than or equal to p. Then given $\varepsilon, p, ARL_0, m, n, \lambda$ and t we want to solve this equation for L. Since a closed form analytical expression for the *cdf* of $CARL_{IN}$ is not available, but a formula to calculate the $CARL_{IN}$ is. Our approach is to generate the empirical distribution of $CARL_{IN}$ using different values of L in the interval $[\,L,\infty)$, starting from the from the L value that corresponds to Case K and going towards infinity. For each empirical distribution, the $CARL_{IN,p}$ value is calculated. The first value of L for which $CARL_{IN,p} > ARL_0(1-\varepsilon)$ is chosen to be the solution. A step by step algorithm for finding L is given in Diko et al. (2019a). Like the other algorithms by Diko et al. (2019a) this algorithm requires an approximation of the $CARL_{IN}$ distribution, via the empirical distribution. In our view this is what gives it an edge over the bootstrap algorithm used in Saleh et al. (2015) and others, which requires more computational effort.

Table 3.3 gives the L values that guarantee, with $(1-p)\%$ probability, that the $CARL_{IN}$ will exceed a specified lower bound ARL_0. Looking at Table 3.3 for $ARL_0 = 370, \lambda = 1$ and $m = 50,100,300,1000$ it can be seen that our constants $L = 3.24,3.16,3.09,3.05$ are exactly equal to those in Goedhart et al. (2017a,2018). The constants in Goedhart et al. (2017a,2018) were obtained analytically and are regarded as an improvement to the computationally intensive bootstrap approach. This validates our method. In addition to Table 3.3 for $ARL_0 = 100,200,370,500$, we have generated 4 figures in which the practitioner may find his/her constant L given its own m and λ by means of interpolation. These figures may be found in Diko et al (2019a).

Table 3.3 *L values that guarantee that $P(CARL_{IN} > ARL_0) = 0.90$ for the EWMA $\bar{X}$ chart for n=5; m=30,50,100,300,1000; $\lambda = 0.1,0.2,0.5,1$; $\varepsilon = 0\%$ and $ARL_0 = 100,200,370,500$*

		p=0.10			
$\boldsymbol{ARL_0}$	***m***	$\boldsymbol{\lambda = 0.1}$	$\boldsymbol{\lambda = 0.2}$	$\boldsymbol{\lambda = 0.5}$	$\boldsymbol{\lambda = 1}$
100	**30**	3.09	3.01	2.92	2.88
	50	2.79	2.79	2.79	2.79
	100	2.50	2.62	2.71	2.72
	300	2.32	2.48	2.62	2.66
	1000	2.23	2.42	2.58	2.62
	Known parameter	**2.148**	**2.360**	**2.534**	**2.576**
200	**30**	3.49	3.34	3.20	3.13
	50	3.16	3.12	3.08	3.03
	100	2.86	2.92	2.96	2.96
	300	2.63	2.77	2.87	2.89
	1000	2.53	2.70	2.83	2.85
	Known parameter	**2.454**	**2.636**	**2.777**	**2.807**
370	**30**	3.78	3.59	3.43	3.34
	50	3.46	3.38	3.30	3.24
	100	3.16	3.16	3.16	3.16
	300	2.89	2.99	3.09	3.09
	1000	2.78	2.92	3.04	3.05
	Known parameter	**2.702**	**2.859**	**2.978**	**3.000**
500	**30**	3.92	3.70	3.54	3.44
	50	3.59	3.49	3.40	3.34
	100	3.29	3.28	3.26	3.26
	300	3.02	3.10	3.18	3.18
	1000	2.90	3.03	3.13	3.14
	Known parameter	**2.815**	**2.962**	**3.071**	**3.090**

Looking at Figures 1 to 4 of Diko et al (2019a), for any given ARL_0, m and λ values , it can be seen that the adjusted L values are all greater than the corresponding parameters known L values. It can also be seen that, for a given ARL_0 and λ, the adjusted L values decrease as *m* increases and converge to the known parameter (unadjusted/standard) L value. Consequently, Phase II EWMA charts that are designed using the new

(corrected/adjusted) L values will have wider control limits and this will lead to an improved IC performance than the charts whose design uses the parameters known L values. This improved IC performance, that is widening the limits, can lead to some deterioration of the *OOC* chart performance. This has been noted in the literature (cf. Goedhart et al. (2017b)) as the price to pay for satisfactory nominal IC chart performance with a high probability. However, it is possible with our approach to relax the IC behavior of the EWMA chart. This can be done by increasing ε or *p* or both in equation (3.7). As will be seen, in the next section, the result will be less wider adjusted limits, which will improve the *OOC* performance.

3.5 IC and OOC performance of the adjusted and the unadjusted limits

Using the bootstrap approach, Saleh et al. (2015) came up with an EWMA chart such that $P\left(CARL_{IN} > 200\right) = 0.90$ for $\lambda = 0.1, m = 50$ and $n = 5$. Following this, they evaluated the IC and OOC conditional performance of this chart. However, their performance evaluations were done only for this chart and were limited only to $\delta = 0$ and $\delta = 1$. In this section we make a much more detailed evaluation and comparison between the performance of the EWMA charts with the proposed adjusted limits and that of the standard (unadjusted) limits chart, for various shifts δ. We also compare our results with the performance results of the bootstrap based (adjusted limits) EWMA chart in Saleh et al. (2015). Furthermore, we use the flexibility of the EPC formulation in equation (3.7) to adjust the trade-off between the IC and OOC performance of the EWMA chart. By trade-off we mean, for example, sacrificing a little IC performance for a better OOC performance.

Table 3.4a *The percentiles (perc) of the $CARL$ distribution of the EWMA $\overline{X}$ chart with the adjusted $(ARL_0 = 200, p = 10\%)$ and unadjusted $(ARL_0 = 200)$ limits for $n = 5$; $m = 50\ 50,100$; $\lambda = 0,0.1,0.2,0.5,1$; $\delta = 0,0.25,0.5,1,1.5$*

	m = 50								
δ	perc	$\boldsymbol{\lambda = 0.1}$		$\boldsymbol{\lambda = 0.2}$		$\boldsymbol{\lambda = 0.5}$		$\boldsymbol{\lambda = 1}$	
		Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
0	0.05	48	139	60	147	79	166	89	169
	0.10	**62**	**202**	**75**	**198**	**92**	**205**	**104**	**205**
	0.25	92	374	105	322	121	283	134	269
	0.50	133	696	146	518	166	411	178	369
	0.75	177	1138	197	786	221	581	246	520
	0.90	220	1694	252	1140	295	801	327	726
	0.95	251	2142	293	1453	346	981	392	889
0.25	0.05	27	50	29	57	43	82	70	122
	0.10	31	63	35	73	54	103	82	144
	0.25	43	102	50	115	77	161	108	199
	0.50	**66**	**198**	**81**	**212**	**112**	**254**	**149**	**285**
	0.75	108	452	127	401	164	399	207	415
	0.90	163	916	184	704	226	596	278	586
	0.95	196	1295	222	928	267	776	333	717
0.5	0.05	16	23	15	24	21	35	42	69
	0.10	17	27	17	27	24	42	48	82
	0.25	21	34	22	37	33	58	64	115
	0.50	**26**	**48**	**29**	**55**	**47**	**90**	**89**	**167**
	0.75	36	76	41	89	69	143	126	245
	0.90	51	125	61	148	102	228	174	341
	0.95	65	182	79	214	125	302	209	425
1	0.05	8	11	7	9	8	10	15	22
	0.10	9	12	8	10	8	12	17	26
	0.25	10	13	9	11	10	14	21	34
	0.50	**11**	**15**	**10**	**14**	**12**	**19**	**28**	**46**
	0.75	12	18	12	17	16	25	38	66
	0.90	14	21	14	20	20	33	51	90
	0.95	15	23	15	23	23	40	62	109
1.5	0.05	6	7	5	6	4	5	6	9
	0.10	6	8	5	6	4	6	7	10
	0.25	6	8	5	7	5	6	8	12
	0.50	**7**	**9**	**6**	**7**	**6**	**7**	**10**	**16**
	0.75	7	10	6	8	7	9	13	21
	0.90	8	11	7	9	8	10	17	28
	0.95	8	11	7	10	8	12	19	33

Table 3.4b *The percentiles (perc) of the $CARL$ distribution of the EWMA $\overline{X}$ chart with the adjusted $(ARL_0 = 200, p = 10\%)$ and unadjusted $(ARL_0 = 200)$ limits for $n = 5$; $m = 100$; $\lambda = 0,0.1,0.2,0.5,1$; $\delta = 0,0.25,0.5,1,1.5$*

	m = 100								
δ	perc	$\boldsymbol{\lambda = 0.1}$		$\boldsymbol{\lambda = 0.2}$		$\boldsymbol{\lambda = 0.5}$		$\boldsymbol{\lambda = 1}$	
		Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
0	0.05	75	152	90	173	107	171	117	177
	0.10	**92**	**200**	**106**	**208**	**121**	**196**	**131**	**199**
	0.25	123	290	131	267	145	242	157	243
	0.50	155	399	163	344	178	307	192	306
	0.75	187	523	202	443	219	390	237	385
	0.90	221	650	239	547	268	486	287	472
	0.95	240	734	264	619	303	548	326	547
0.25	0.05	33	50	38	57	58	86	86	133
	0.10	38	58	43	68	66	100	98	152
	0.25	49	79	56	94	86	133	122	189
	0.50	**67**	**119**	**79**	**138**	**113**	**182**	**151**	**242**
	0.75	95	188	111	213	148	248	193	315
	0.90	134	302	149	304	189	328	239	398
	0.95	162	387	174	374	217	388	267	453
0.5	0.05	18	23	18	24	26	36	51	74
	0.10	19	25	20	26	30	41	57	85
	0.25	22	30	24	32	37	53	71	106
	0.50	**26**	**37**	**30**	**42**	**47**	**70**	**90**	**137**
	0.75	32	48	38	56	61	96	115	178
	0.90	40	63	49	74	78	127	144	225
	0.95	46	76	58	90	93	151	165	262
1	0.05	9	11	8	9	7	11	17	23
	0.10	9	11	8	10	9	11	19	26
	0.25	10	12	9	11	11	13	23	32
	0.50	**11**	**13**	**10**	**12**	**12**	**16**	**28**	**40**
	0.75	12	15	11	13	15	19	35	50
	0.90	13	16	12	15	17	23	43	62
	0.95	14	17	13	176	19	25	47	70
1.5	0.05	6	7	5	6	5	5	7	9
	0.10	6	7	5	6	5	5	8	10
	0.25	6	8	5	6	5	6	9	12
	0.50	**7**	**8**	**6**	**7**	**6**	**7**	**10**	**14**
	0.75	7	9	6	7	6	7	12	17
	0.90	8	9	7	8	7	8	14	20
	0.95	8	9	7	8	7	9	16	22

Table 3.4a and b shows some percentiles (perc) of the $CARL$ distribution for various combinations of λ, δ and m for both adjusted and unadjusted limits. Again, for given λ and $\delta = 0$, the adjusted limits were obtained such that $P(CARL_{IN} > 200) = 1 - p = 0.90$ (so that the $10th$ percentiles of the $CARL_{IN}$ distribution should be close to 200) while the unadjusted limits were obtained using the function *xewma.crit(λ,ARL_0,sided="two")* from the R-package *spc*, such that $ARL_0 = 200$, in the known parameters case. Looking at Table 3.4, for $\delta = 0$ and all λ, it can be seen that the *IC* performance of the chart with the adjusted limits is as specified. For example, for $m = 50, p = perc = 10\%$ and $\lambda = 0.1,0.2,0.5,1$; it can be seen that $CARL_{IN,p} = 202,198,205,205$; respectively, and for $m = 100$ we have $CARL_{IN,p} = 200,208,196,199$. All of these $CARL_{IN,p}$ are very close to the nominal $ARL_0 = 200$. Besides, for $\delta = 0$ and all $\lambda, perc, m$ values, the $CARL$ values for the adjusted limits are always higher than those for the corresponding unadjusted limits. So, for all percentiles (*perc*), the adjusted limits charts always guarantee, with high probability (close to the nominal), larger $CARL_{IN}$ values compared to the unadjusted limits charts. Thus, the good IC performance of the adjusted limits charts is not only limited to the design percentile ($perc = p = 10\%$ in this case), but extends over the entire range of $perc's$.

However, as mentioned before, because the adjusted limits are wider, they can be insensitive to true process shifts compared to the unadjusted limits. We explore this for the cases when $\delta = 0.25,0.5,1,1.5$. Looking at Table 3.4 for $m = 50$, small shifts $\delta = 0.25,0.5$ and all λ values, it can be seen that the medians of the $CARL$ distributions for the unadjusted and the adjusted limits charts are radically different. The largest difference occurs at $\lambda = 0.1$, while the smallest difference occurs at $\lambda = 1$. Increasing *m* reduces the differences slightly, but the pattern remains the same. Thus, for a small shift $\delta \leq 0.50$, the OOC chart performance of the adjusted limits EWMA chart is not as good as that of the unadjusted limits charts, particularly when $\lambda = 0.1$. But of course the point is that the IC performance of the unadjusted limits based chart is a much bigger problem. However, for larger shifts, such as $\delta = 1$ or $\delta = 2$ and λ values, the $CARL$ percentiles for the unadjusted and the adjusted limit charts are quite close. This is even more so when $m = 100$. Thus, for moderate to large values of δ, the OOC chart performance of the adjusted limits EWMA chart is comparable to that of the EWMA chart with the unadjusted limits or the limits for the known parameter case.

Note that, in the literature (e.g. Saleh et al. (2015)), authors who use the bootstrap approach often compare the OOC behavior of the unadjusted and the adjusted limits charts solely on the basis of a shift of size $\delta = 1$. Figure 3.1 shows the boxplots for the $CARL$ distributions of the EWMA chart with the adjusted and the unadjusted limits for $\lambda = 0.1, \delta = 1$, $m = 50$ and $n = 5$. Based on Figure 3.1, it has always been concluded that the OOC performance of the bootstrap adjusted limits is not radically different from that of the unadjusted limits. However, we have shown through the $CARL$ values, in Table 3.4, that this only occurs when δ is moderate to large. Therefore, widening the control limits by the EPC criterion, makes them a little insensitive to small process shifts but guarantees a nominal performance with high probability. This may be the trade-off one has to accept. However, it is possible to adjust this trade-off to get a better OOC performance. This can be done by sacrificing a bit of IC performance. For fixed p, the IC performance can be sacrificed by increasing ε in equation (3.7), while for fixed ε , it can be reduced by increasing p in equation (3.7). To illustrate the former, Figures 3.2 and 3.3 show the boxplots for the $CARL_{IN}$ and $CARL_{OOC}$ distributions of the EWMA chart; respectively; for $\lambda = 0.1; ARL_0 = 200$; $\varepsilon = 0\%, 35\%, 69\%$; $p = 0.10; m = 50$ and $n = 5$.

From Figure 3.2, it can be seen that increasing ε from 0% to 35% leads to a slight loss of the IC performance. For example when $\varepsilon = 35\%$ the proportion of $CARL_{IN}$ values that are less than 200 is 20%. But, this is still way better than the 85% that occurs when the unadjusted limits $(\varepsilon = 69\%)$ are used. From Figure 3.3, it can also be seen that increasing ε from 0% to 35% leads to an improved OOC performance in the sense that the median for the $CARL_{OOC}$ distribution of $\varepsilon = 35\%$ is closer to the median (the dotted vertical line) for the $CARL_{OOC}$ distribution of $\varepsilon = 69\%$. Thus, by sacrificing a bit of the IC performance, it is possible to improve the EWMA charts ability to detect small shifts.

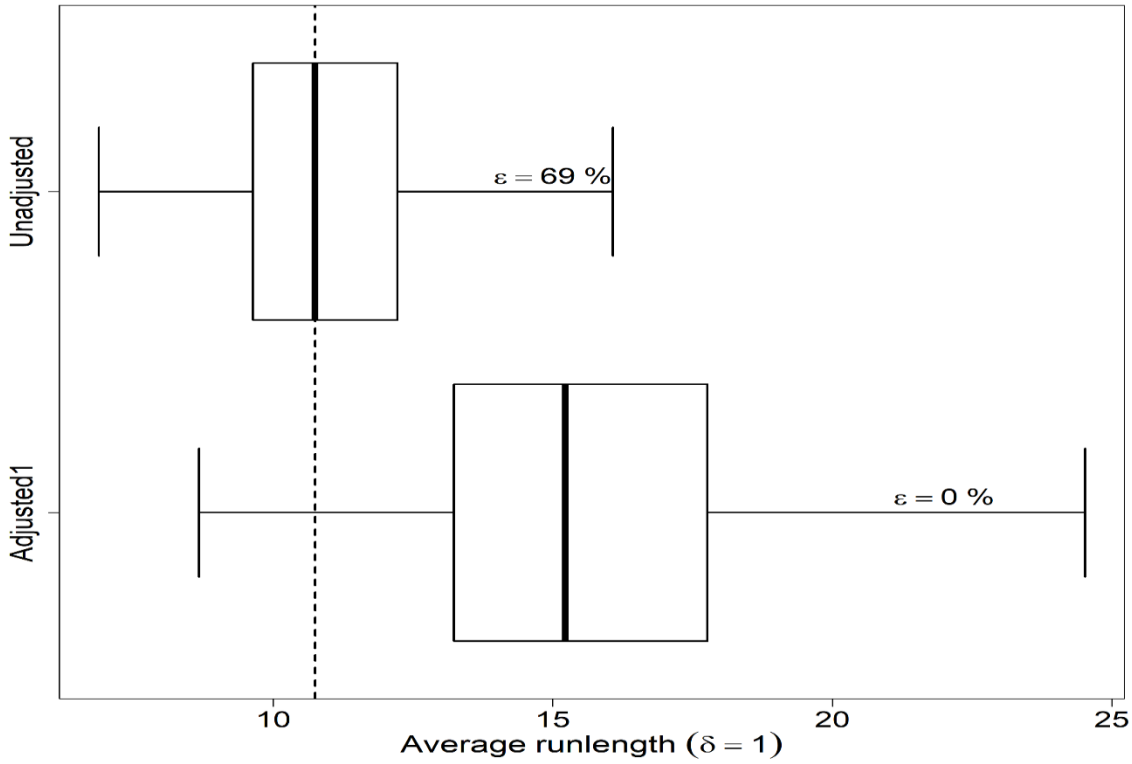


Figure 3.1 *Boxplots of the CARL distribution when* $\delta = 1, \lambda = 0.1, p = 10\%, ARL_0 = 200, m = 50$ *and* $n = 5$

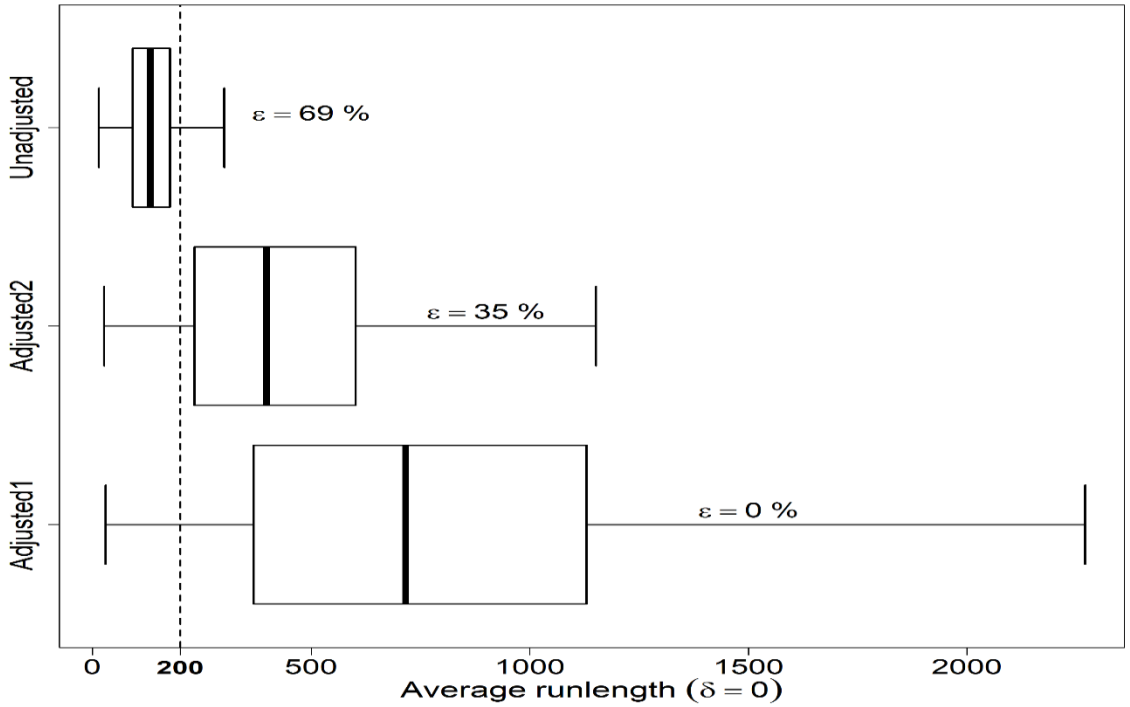


Figure 3.2 *Boxplots of the* $CARL$ *distribution when* $\delta = 0, \lambda = 0.1, p = 10\%, ARL_0 = 200, m = 50$ *and n=5*

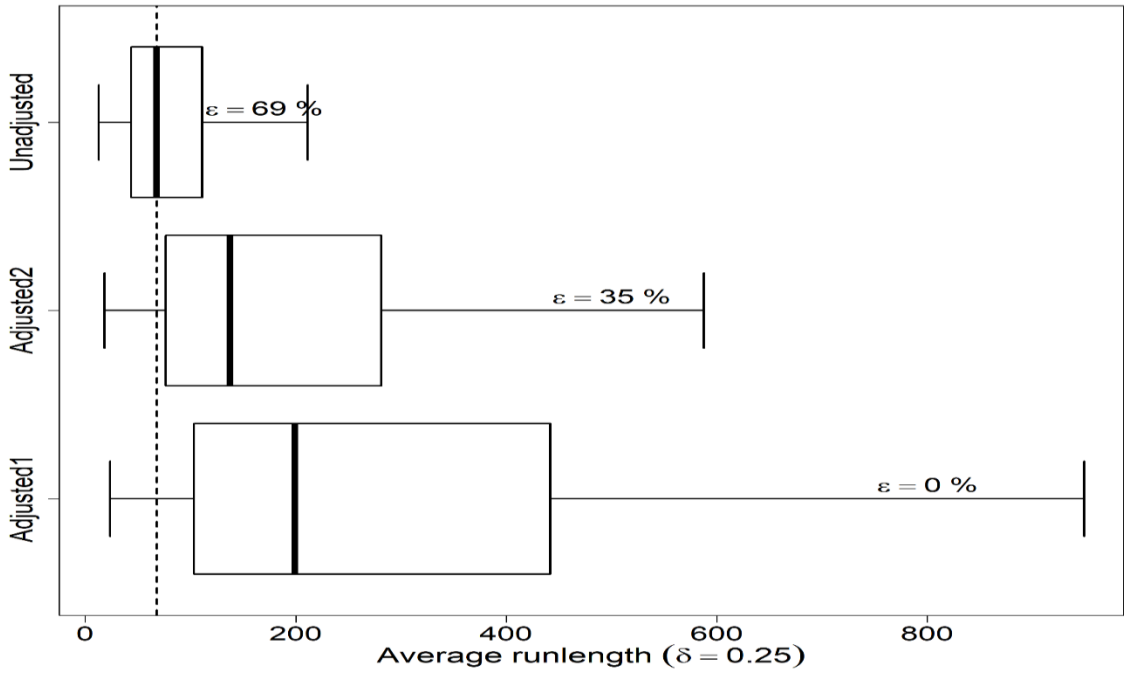


Figure 3.3 *Boxplots of the* $CARL$ *distribution when* $\delta = 0.25, \lambda = 0.1, p = 10\%, ARL_0 = 200, m = 50$ *and n=5.*

3.6 Summary and conclusions

We study the impact of practitioner to practitioner variability on the performance of the Phase II EWMA chart. As in Epprecht et al. (2015) we use the EPC criterion to evaluate the performance of a Phase II EWMA chart with limits for the known parameter case and give recommendations about the required number of Phase I subgroups to achieve nominal performance. Our results show that in order to attain or exceed a specified lower bound of $CARL_{IN}$ (given by ARL_0) with a specified high probability, more Phase I data are required than previously recommended by Saleh et al. (2015) and Jones et al. (2001). Moreover, consistently with Jones et al. (2001) but contrary to Saleh et al. (2015), our results also show that smaller values of λ may require a larger number of Phase I subgroups, that is, more Phase I data.

Since it is expensive and sometimes impractical to get such large amount of Phase I data to estimate the process parameters and construct Phase II charts that guarantee a high probability of high $CARL_{IN}$'s under the EPC, the control limits are adjusted as a function of the available Phase I data. In this regard, where analytical methods could not be conveniently used e.g. for the *EWMA* chart or where normality cannot be assumed, the bootstrap approach has been an attractive choice. However, many SPM practitioners and researchers have felt that the bootstrap approach may be somewhat complex and have looked for an alternative. In this chapter, we presented an alternative method that can be used instead of the bootstrap approach. Our method produces the same results as the bootstrap approach, but it is faster. Based on the new method, tables and the graphs of the adjusted charting constants are provided to help practitioners implement the Phase II *EWMA* chart with estimated parameters more easily in practice. The new charting constants are larger than the traditional ones commonly used for Case K. Thus, the *EWMA* charts constructed using these new constants have wider limits, particularly for small λ and/or m.

Adjusting the limits of the *EWMA* chart, using our new constants, guarantees with high probability that the $CARL_{IN}$ performance will be as nominally specified. However, there is some concern about the deterioration in the $OOC\ CARL$ performance relative to using the unadjusted limits which are wider. This is of course true for all types of control charts with estimated parameters and has been observed, for example, for the Shewhart charts (see Goedhart et al. (2017b)). The extent of the deterioration depends on the size of the shift δ and m. For moderate to large shifts (say $\delta = 1$ and more), the difference in the $OOC\ CARL$ performance between the adjusted and unadjusted limits is negligible. However for small shifts (say $\delta = 0.25, 0.50$) and small m, the difference is not negligible. Thus, adjusting the control limits can make the chart somewhat insensitive to detecting small shifts. The insensitivity to small shifts may be improved by sacrificing some IC chart performance as illustrated in Figures 3.2 and 3.3. Nonetheless, it is important to keep in mind that the IC chart performance is perhaps the most important to have higher confidence in, so sacrificing some OOC performance may be the price one has to pay when a given amount of Phase I data are used to estimate the parameters to construct a control chart.

4. An alternative design of the CUSUM chart for monitoring the mean

4.1 Introduction

Page (1954) introduced the CUSUM control chart for monitoring the process mean. The tabular form of the CUSUM chart was developed by Ewan and Kemp (1960). The upper one-sided CUSUM chart is used to monitor upward shifts, while the lower one-sided CUSUM chart is used to monitor downward shifts. If we are interested in monitoring both the upward and downward shifts of the process mean, we have to use a two-sided CUSUM control chart. The two-sided CUSUM chart of Crosier (1986) uses one statistic to detect both shifts. In this chapter, we study the two-sided CUSUM chart, which uses two charting statistics to monitor upwards and downwards shifts, respectively. We assume that the process is normal and the process parameters are unknown.

Since the $CARL_{IN}$ depends on the Phase I parameter estimates, it is a random variable with its own probability distribution. It has been shown (see Chakraborti (2006) and Saleh et al. (2015a, 2015b, 2016)) that these distributions have large variability and therefore the values of the $CARL_{IN}$ can be quite different from the nominally specified average run length (ARL_0) . In this context, the expected value of the $CARL_{IN}$, denoted as $AARL_{IN}$, has been used to evaluate and design Phase II control charts. This approach, the so-called unconditional perspective, has received a lot of attention in the literature. See, for example, Quesenberry (1993), Chen (1997, 1998), Chakraborti (2000, 2006), Jones et al. (2004), Jensen et al. (2006), Abbasi et al. (2012), Psarakis et al. (2014), Diko et al. (2016, 2017), Goedhart et al. (2016), Sanusi et al. (2017), Jardim et al. (2019a,2019b) and Mehmood et al. (2018)). However, the unconditional perspective does not show individual chart performance, which is known to vary from practitioner to practitioner. Accordingly, another approach, the so-called conditional perspective, has been suggested. Under this perspective, a number of criteria has been used. Jones et al. (2004) used the mean, the standard deviation and the $10^{th}, 50^{th}$, and 90^{th} percentiles of the conditional run length distribution to study the performance of the one-sided and two-sided CUSUM location

charts. Furthermore, Jeske (2016) derived Phase I sample size requirements to ensure probabilistic control of the relative error of the $CARL_{IN}$. Moreover, Saleh et al. (2016) used the standard deviation of $CARL_{IN}$, denoted as $SDCARL_{IN}$, to quantify the amount of variation in the in-control chart performance corresponding to different amounts of Phase I data. For the EWMA control chart for location, Diko et al. (2019a) considered a one-sided prediction interval for $SDCARL_{IN}$ and, for several sample sizes, obtained the minimum number of Phase I sample observations that guarantee, with a specified high probability, that the $CARL_{IN}$ will exceed a nominal ARL_0 by more than a specified percentage. This is called the exceedance probability criterion (EPC).

The EPC was originally proposed by Albers and Kallenberg (2004a,2004b) to adjust Phase II charting constants to compensate for the effects of parameter estimation when *m* is small. To adjust the Phase II charting constants according to the EPC, Jones and Steiner (2012) and Gandy and Kvaloy (2013) proposed the bootstrap method. Since then, many authors have used the bootstrap method to adjust the Phase II charting constants according to the EPC. For example, Saleh et al. (2015a, 2015b, 2016), Faraz et al. (2015, 2016), Aly et al. (2015), and Hu and Castagliola (2017). However, it is known that repeated application of the bootstrap approach would most likely result in different solutions (cf. Saleh et al. (2016)). Thus, the bootstrap based EPC adjusted charting constants might not guarantee the anticipated IC performance. To overcome this limitation, Hu and Castagliola (2017) found for a Shewhart median chart the EPC adjusted charting constants by running the bootstrap method a 100 times and averaging the results. Goedhart et al. (2017a, 2017b, 2018) and Jardim et al. (2019a, 2019b) provided closed form analytical expressions to calculate the EPC adjusted constants, for Shewhart charts, when the process distribution is assumed normal. Hence in this situation bootstrapping was not necessary. Moreover, Diko et al. (2019a) proposed an alternative method for finding the EPC adjusted constants for the EWMA chart. They concluded that the method can be used, instead of the bootstrap method, whenever analytical methods cannot be conveniently applied.

In this chapter, we take another look at the two-sided CUSUM control charts for the mean with estimated parameters. In the first place, we extend Jeske's (2016) modified Siegmund (1985) formula for the one-sided CUSUM chart to the two-sided case, to calculate the values of the $CARL_{IN}$ distribution. It will be shown that the modified Siegmund formula

is much easier to use compared with the Markov Chain method for obtaining the $CARL_{IN}$ distribution. Next, we study the effect of the amount of Phase I data on the Phase II performance of the two-sided CUSUM control chart, based on the EPC. Finally, we use the EPC to adjust the CUSUM charting constants. We also provide graphs of the adjusted charting constants. Practitioners can use these graphs to get the charting constants for their own particular situation by interpolation. Note that unlike Saleh et al. (2016), we do not use the bootstrap method to apply the EPC to adjust the charting constants. We use the method that was proposed by Diko et al. (2019a), as it is more accurate.

This chapter is based on Diko et al. (2019b) and it is structured as follows. In Section 4.2 we give an overview of the two-sided Phase II CUSUM control chart for monitoring the mean of a normal process based on some popular estimators for the unknown process mean and variance. In Section 4.3, we provide the modified Siegmund formula to calculate the $CARL_{IN}$ of the two-sided Phase II CUSUM control chart and compare it with the well-known Markov Chain method. In Section 4.4, we introduce the EPC to obtain, for the two-sided CUSUM control charts, the $CARL_{IN}$ prediction bounds. In Section 4.5, we find the minimum m required to design a two-sided Phase II CUSUM control chart. Results are compared with the available results in the literature. In Section 4.6, the charting constants are adjusted and presented graphically. In Section 4.7, the application of the adjusted and unadjusted control charts is illustrated using real life data from Montgomery (2013). The adjusted and unadjusted limits charts are compared in terms of their IC and out-of-control (OOC) performance. Possible trade-offs to balance the IC and OOC performance of the adjusted limits charts are suggested. Finally, in Section 4.8, a summary and conclusions are offered.

4.2 Two-sided CUSUM charts with estimated parameters

Let X_{ij}, $i = 1,2,\dots,m$ and $j = 1,2,\dots,n$, denote the IC Phase I data from a normal distribution with an unknown mean μ_0 and unknown standard deviation σ_0. To estimate μ_0, we use the estimator $\hat{\mu}_0 = \sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}/mn$. To estimate σ_0, for $n > 1$, we use the pooled standard deviation estimator

$$\hat{\sigma}_{01} = S_p/c_4(m(n-1)+1) = \frac{1}{c_4(m(n-1)+1)}\sqrt{\frac{1}{m}\sum_{i=1}^{m} S_i^2} \tag{4.1}$$

where c_4 is the unbiasing constant for the estimator S_p assuming the normal distribution (Montgomery(2013)) and S_i is the standard deviation of the i^{th} Phase I sample. Note that, for $n > 1$, there are several other estimators of σ_0 that can be used. Popular among these are the mean range and the mean standard deviation estimators. However, the results we obtained using these estimators marginally differ from the results we obtained using $\hat{\sigma}_{01}$. Hence, we only focus on $\hat{\sigma}_{01}$. Furthermore, when the number of Phase I subgroups m is moderately large, the constant c_4 becomes indistinguishable from 1 (Diko et al. (2017), Mahmoud et al. (2010)), hence, we may use $\hat{\sigma}_{01} = S_p$. For $n = 1$, we use the average moving ranges as estimator for σ_0

$$\hat{\sigma}_{02} = MR/d_2(2)(m-1) = \sum_{i=2}^{m} |X_i - X_{i-1}|/1.128(m-1) \tag{4.2}$$

where $d_2(2)$ is an unbiasing constant (cf. Montgomery (2013)).

Assuming that the Phase II data are normally distributed, with mean μ and standard deviation σ, the standardized charting statistics of the two-sided CUSUM control chart for the mean are given by

$$C_i^+ = max(0, C_{i-1}^+ + B_i - k)$$
$$\text{and} \tag{4.3}$$
$$C_i^- = min(0, C_{i-1}^- + B_i + k), \quad i = m+1, m+2, \ldots$$

where $B_i = \frac{\bar{X}_i - \hat{\mu}_0}{\hat{\sigma}_{0g}/\sqrt{n}} = \left[\kappa T_i + \sqrt{n}\delta - \frac{Z}{\sqrt{m}}\right] Q_g^{-1}$, $Q_g = \frac{\hat{\sigma}_{0g}}{\sigma_0} = a_g\sqrt{\frac{Y}{b_g}}$, $\kappa = \frac{\sigma}{\sigma_0}$, $\delta = \frac{\mu - \mu_0}{\sigma_0}$,

$T_i = \frac{\sqrt{n}(\bar{X}_i - \mu)}{\sigma}$, $Z = \frac{\sqrt{mn}(\hat{\mu}_0 - \mu_0)}{\sigma_0}$, $\bar{X}_i$ is the *i th* sample mean, Y is a chi-square variable with b_g degrees of freedom (cf. Jones, Champ, and Rigdon (2001)). We use the subscript $g = 1,2$ to distinguish between the two unbiased estimators of σ_0 that we employ in this chapter. The constants b_g and a_g $(g = 1,2)$ are functions of m and n. If $g = 1$ then $b_1 = m(n-1)$ and $a_1 = 1$. If $g = 2$ then we calculate b_2 and a_2 using the Patnaik (1950) approximation as described in Saleh et al. (2016). The calculation of b_2 and a_2 can be obtained with the R codes that we provide in Diko et al. (2019b) and in chapter 2. We assume that the starting

values (C_m^+ and C_m^-) are both equal to zero, which means that the process is initially *IC*. Note that, selecting $C_m^+ > 0$ and/or $C_m^- < 0$ gives the chart a fast initial response or "head start", since it increases the charts ability to detect a process that is initially OOC (see Lucas and Crosier (1982)). We also assume that the reference values ($k^+$and k^-) are both equal to *k*. Note that, the reference value k is usually chosen to be half the shift (δ) that is considered important enough to be detected. When the size of the shift is not known (e.g. with start-up processes), Sparks (2000) proposed using adaptive CUSUM charts, which employ current process data to estimate the shift and choose *k* in accordance to this estimate. We do not consider these charts here.

The standardized two-sided Phase II CUSUM control chart gives an OOC signal when

$$C_i^+ \geq h^+ \text{ or } C_i^- \leq -h^- \tag{4.4}$$

where $h^+ = h^- = h$ is the charting limit to be found. For the parameters known case , the h values can be found in the literature; see for example, Qiu (2014) or the R package *spc* by specifying values of the nominal average run length ARL_0 and k. As an example, for some chosen ARL_0 and k, the h values in Table 4.1 were obtained using the function xcusum.crit(k, ARL_0,*sided="two"*), from the package *spc*, in R.

Table 4.1. *Values of k and their corresponding values of h for the two-sided CUSUM control chart and $ARL_0 = 100, 200, 370$.*

ARL_0	*k*	*h*
100	0.12	7.968
	0.25	5.597
	0.50	3.502
200	0.12	9.998
	0.25	6.854
	0.50	4.172
370	0.12	12.083
	0.25	8.008
	0.50	4.774

Note however, that the *h* values in Table 4.1 are for the parameters known case and hence, they do not account for the effects of parameter estimation from Phase I data. Thus,

unless the number of Phase I subgroups m is very large, they should not be used to construct Phase II control charts when estimated parameters are used in the control chart. However, most practitioners continue to use these Table 4.1 constants to construct their Phase II CUSUM charts regardless of the value of m. This can be attributed to the fact that adjusting these charting limits for the effects of parameter estimation is computer intensive. Moreover, no tables, graphs or software packages for the adjusted h values are conveniently available to date to help practitioners to correctly implement their Phase II CUSUM control charts with estimated parameters. Hence, in this chapter, we adjust the charting constants for parameter estimation and present the adjusted constants graphically, so that they can be found without much effort and the CUSUM chart with estimated parameters can be implemented more easily in practice.

Next, we present simple formulas for approximating the $CARL_{IN}$ distribution, its expected value denoted by $AARL_{IN}$ and standard deviation denoted by $SDCARL_{IN}$.

4.3 The modified Siegmund formula

Traditionally, the $CARL_{IN}$ values of CUSUM charts have been calculated by simulations (cf. Nazir et al. (2013)) and Markov Chains (cf. Brook and Evans (1972), Crosier (1982), Woodall (1984)). Recently, Jeske (2016) presented a modified Siegmund formula for approximating the $CARL_{IN}$ of the upper one-sided CUSUM chart. We first extend this formula to the two-sided CUSUM chart and compare the results with available results from the well-known Markov Chain method.

The original Siegmund approximation (Siegmund (1985)) for the IC average run length (ARL_{IN}) of the upper or the lower one-sided CUSUM with reference value (k) and control limit (h), when the parameters are known, is given by

$$ARL_{IN}(k,h) = \frac{\exp(2\mathrm{k}(h+1.166)) - 2\mathrm{k}(h+1.166) - 1}{2k^2}. \tag{4.5}$$

For the upper one-sided CUSUM chart, when parameters are estimated, Jeske (2016) showed that the signaling event $C_i^+ \geq h$ is equivalent to

$$max\left(0, C_{i-1}^{+} + \kappa T_i + \sqrt{n}\delta - (U + kV)\right) \geq hV \tag{4.6}$$

where $U = \frac{Z}{\sqrt{m}}$, $V = Q_g = a_g\sqrt{\frac{Y}{b_g}}$ and in which one can identify the modified "reference value" as $U + kV$ and the modified "control limit" as hV, respectively. By substituting the modified "reference value" and the modified "control limit" into equation (4.6), Jeske (2016) proposed the modified Siegmund formula for approximating the $CARL_{IN}$ of the upper (C_i^+) one-sided CUSUM chart as

$$CARL_{IN}^{+}(Z, Y, k, h, m, n)$$

$$= \frac{\exp(2(U + kV)(hV + 1.166)) - 2(U + kV)(hV + 1.166) - 1}{2(U + kV)^2}$$

$$= \frac{exp\left[2\left(\frac{Z}{\sqrt{m}} + ka_g\sqrt{\frac{Y}{b_g}}\right)\left(ha_g\sqrt{\frac{Y}{b_g}} + 1.166\right)\right] - 2\left(\frac{Z}{\sqrt{m}} + ka_g\sqrt{\frac{Y}{b_g}}\right)\left(ha_g\sqrt{\frac{Y}{b_g}} + 1.166\right) - 1}{2\left(\frac{Z}{\sqrt{m}} + ka_g\sqrt{\frac{Y}{b_g}}\right)^2}. \tag{4.7}$$

Note that the $CARL_{IN}^{+}$ is a random variable, being a function of the two random variables Z and Y. Likewise, for the lower (C_i^-) one-sided CUSUM charting statistic one can rewrite the signaling event $C_i^- \leq -h$. Then simplify in a similar manner with the "reference value" $-U + kV = -\frac{Z}{\sqrt{m}} + ka_g\sqrt{\frac{Y}{b_g}}$ and the modified "control limit" $hV = ha_g\sqrt{\frac{Y}{b_g}}$, respectively, and express the modified Siegmund formula for approximating the $CARL_{IN}$ of the lower (C_i^-) one-sided CUSUM control chart as

$$CARL_{IN}^{-}(Z, Y, k, h, m, n)$$

$$= \frac{exp\left[2\left(-\frac{Z}{\sqrt{m}} + ka_g\sqrt{\frac{Y}{b_g}}\right)\left(ha_g\sqrt{\frac{Y}{b_g}} + 1.166\right)\right] - 2\left(-\frac{Z}{\sqrt{m}} + ka_g\sqrt{\frac{Y}{b_g}}\right)\left(ha_g\sqrt{\frac{Y}{b_g}} + 1.166\right) - 1}{2\left(-\frac{Z}{\sqrt{m}} + ka_g\sqrt{\frac{Y}{b_g}}\right)^2}. \tag{4.8}$$

Now, using the Van Dobben de Bruyn (1968) formulation for calculating the ARL of the two-sided tabular CUSUM chart, the modified Siegmund formula for approximating the $CARL_{IN}$ of the two-sided CUSUM control chart can be expressed as

$$CARL_{IN}(Z,Y,k,h,m,n) = \left[\frac{1}{CARL^{+}_{IN}(Z,Y,k,h,m,n)} + \frac{1}{CARL^{-}_{IN}(Z,Y,k,h,m,n)}\right]^{-1}, \quad (4.9)$$

where $CARL^{+}_{IN}(Z,Y,k,h,m,n)$ and $CARL^{-}_{IN}(Z,Y,k,h,m,n)$ are given in equation (4.7) and equation (4.8), respectively.

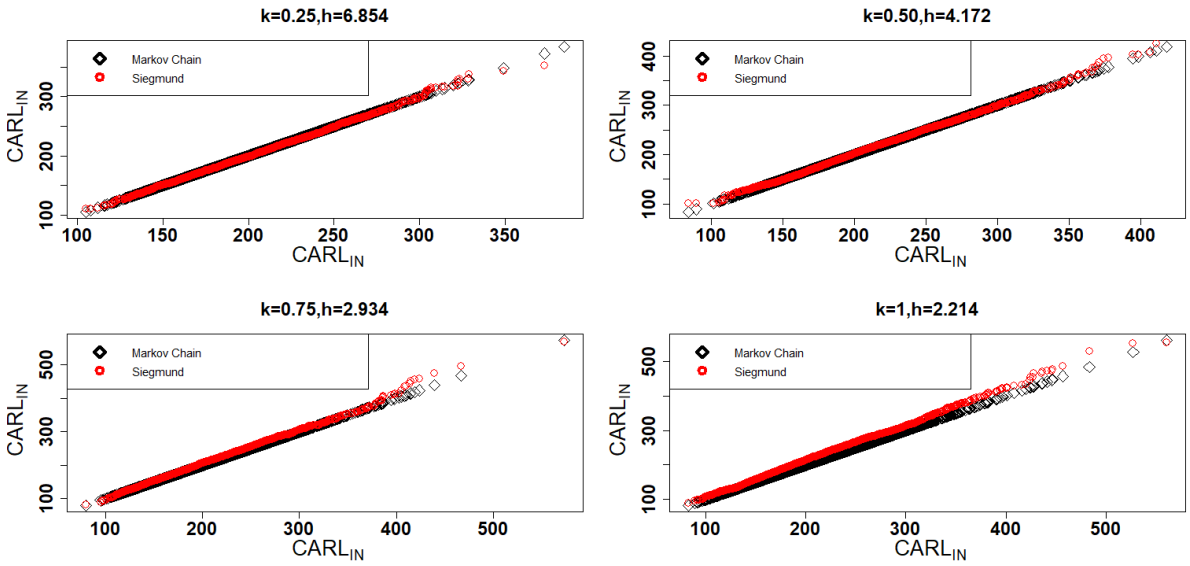


Figure 4.1. *QQ-plots for the empirical $CARL_{IN}$ distributions of two-sided CUSUM charts, calculated using the Markov Chain and the modified Siegmund approximation methods for $m = 1000, n = 1$ and $ARL_0 = 200$.*

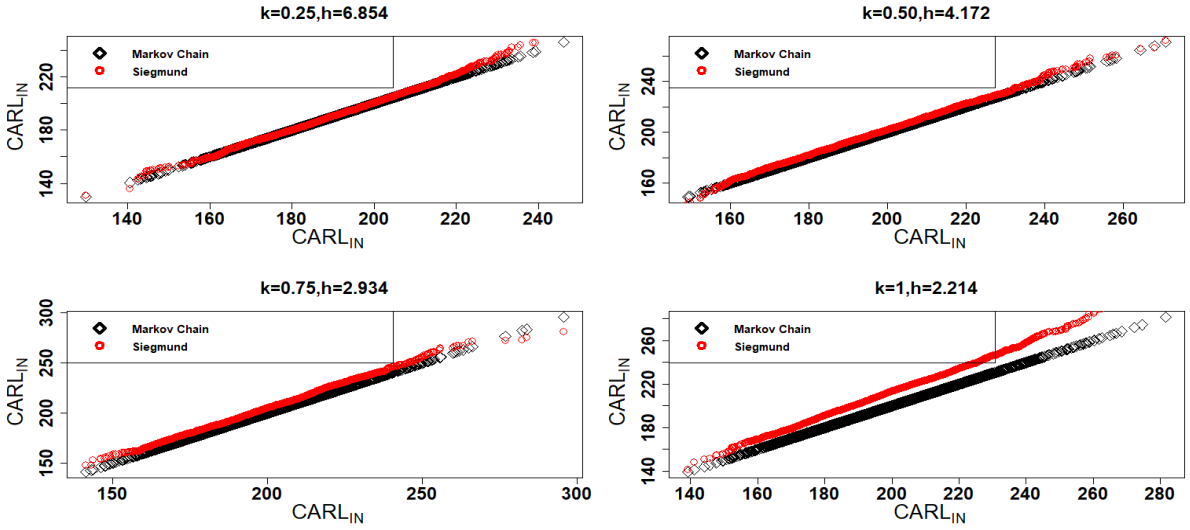


Figure 4.2. *QQ-plots for the empirical $CARL_{IN}$ distributions of two-sided CUSUM charts, calculated using the Markov Chain and the modified Siegmund approximation methods for $m = 1000, n = 5$ and $ARL_0 = 200$.*

In Figures 4.1 and 4.2, Q-Q plots are used to illustrate the quality of the modified Siegmund approximation (equation (4.9)) relative to the Markov Chain approximation for the $CARL_{IN}$ for $n = 1, 5, m = 1000$, $ARL_0 = 200$ and $k = 0.25, 0.50, 0.75, 1$. The annotated R codes for applying the modified Siegmund formula and Markov Chain approximation to calculate the empirical $CARL_{IN}$ distributions are given in Diko et al (2019b).

Looking at Figures 4.1 and 4.2, it can be seen that for *k*=0.25, 0.50 and 0.75, the modified Siegmund approximation based $CARL_{IN}(z, y, k, h, m, n)$ values are similar to their corresponding Markov Chain values. But, for $k = 1$, they are consistently higher than their Markov Chain counterparts. The latter is more pronounced when *n=5* than *n=1*. Note that in practice we use $k \leq 0.75$, since smaller shifts are of more interest while using a CUSUM chart. Thus, there is no practical need for using the Markov Chain method to calculate the $CARL_{IN}$ of the two-sided Phase II CUSUM control chart. We recommend in these cases equation (4.9) because it is simple and more practical.

The modified Siegmund approximation formula to the $CARL_{IN}$ given in equation (4.9) is also convenient for approximating some important CUSUM control chart performance metrics. For example, the mean of the $CARL_{IN}$, denoted as $AARL_{IN}$ and the standard deviation of the $CARL_{IN}$, denoted as $SDCARL_{IN}$, which can be obtained by

$$\begin{aligned} AARL_{IN}(k, h, m, n) &= E\big(CARL_{IN}(Z, Y, k, h, m, n)\big) \\ &= \int_{-\infty}^{\infty}\int_{0}^{\infty} CARL_{IN}(z, y, k, h, m, n)\, \phi(z) q(y) dy dz \end{aligned} \tag{4.10}$$

respectively

$$\begin{aligned} SDCARL_{IN}(k, h, m, n) &= \sqrt{Var\big(CARL_{IN}(Z, Y, k, h, m, n)\big)} \\ &= \sqrt{E\big(CARL_{IN}(Z, Y, k, h, m, n)\big)^2 - \Big(E\big(CARL_{IN}(Z, Y, k, h, m, n)\big)\Big)^2} \\ &= \left[\int_{-\infty}^{\infty}\int_{0}^{\infty} \big(CARL_{IN}(z, y, k, h, m, n)\big)^2 \phi(z) q(y) dy dz \right. \\ &\quad \left. - \left(\int_{-\infty}^{\infty}\int_{0}^{\infty} CARL_{IN}(z, y, k, h, m, n)\, \phi(z) q(y) dy dz\right)^2\right]^{\frac{1}{2}} \end{aligned} \tag{4.11}$$

where the $CARL_{IN}(z, y, k, h, m, n)$ is given in (9), q is the pdf of a chi-square distribution with b_g degrees of freedom $(g = 1,2)$ and ϕ is the pdf of the standard normal distribution. In Figures 4.3, we compare the Siegmund approximation based $AARL_{IN}$ and $SDCARL_{IN}$ values from equations (4.10) and (4.11) with those from the Markov Chain method as a function of m. The horizontal line at $CARL_{IN} = 200$ represents the nominal ARL_0. The two horizontal lines, above and below $ARL_0 = 200$, are drawn at one nominal $SDCARL_{IN}$ (say $SDARL_0$) away from $ARL_0 = 200$. Like Saleh et al. (2016), we chose $SDARL_0$ to be 10% of the nominal ARL_0. The function "adaptintegrate" in the package "cubuture" of the software R was used to evaluate the integrals in equations (10) and (11) (see also Diko et al. (2019b)).

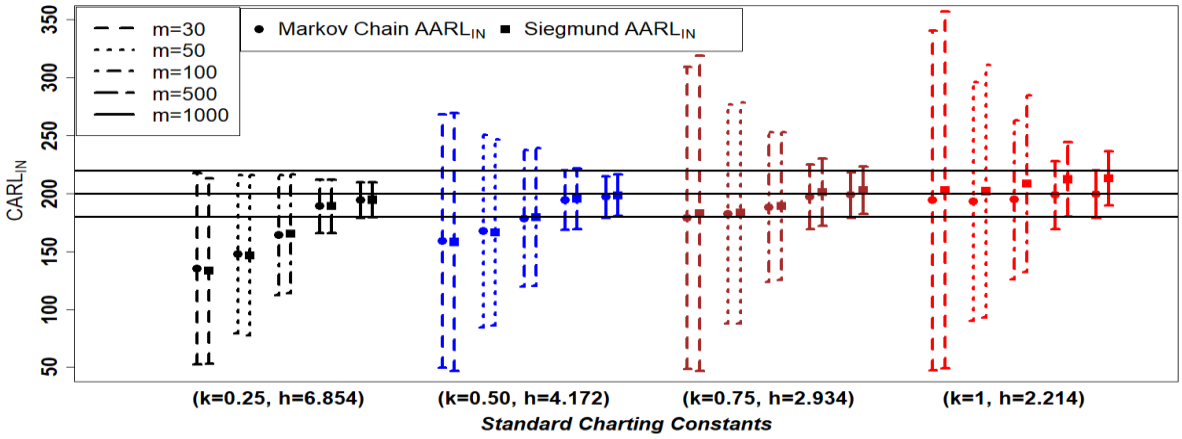


Figure 4.3. *$AARL_{IN}$ and $SDCARL_{IN}$ for a two-sided CUSUM chart for $k = 0.25, 0.50, 0.75, 1$; $n = 5$ and $ARL_0 = 200$.*

Looking at Figure 4.3, it can be seen that for $k \leq 0.75$, even when m decreases from 1000 to 30, the $CARL_{IN}$ values, obtained by both methods, are almost the same. Thus, changes in m do not alter the quality of the modified Siegmund approximations relative to the Markov Chain approximations for the two-sided CUSUM chart with estimated parameters. However, for all k, m affects the $AARL_{IN}$ and $SDCARL_{IN}$ values. For example, for $k = 0.25$, for both the Siegmund and Markov Chain methods, it can be seen that increasing m to 500 has an impact of reducing the $SDCARL_{IN}$ to within 10% of the $ARL_0 = 200$. Furthermore, increasing m to 1000 increases the $AARL_{IN}$ to $ARL_0 = 200$. These results agree with those of Saleh et al. (2016) who used the Markov Chain method.

We conclude that the accuracy of the modified Siegmund formula is not dependent on m and that the modified Siegmund formula leads to the same conclusions that have been reached by others such as Saleh et al. (2016), who used the Markov Chain method. However, the modified Siegmund formula is much simpler and more convenient to use in practice.

Next, the modified Siegmund formula is used to find the $CARL_{IN}$ prediction bounds based on the EPC, to study the performance of the two-sided CUSUM control chart with estimated parameters.

4.4 Prediction bounds

Saleh et al. (2016) used the $SDCARL_{IN}$ to study the effects of parameter estimation on the Phase II CUSUM control chart. The $SDCARL_{IN}$ indicates the amount of spread of the $CARL_{IN}$ values around the $AARL_{IN}$. However, it does not distinguish between low $CARL_{IN}$'s and high $CARL_{IN}$'s. Thus, it cannot explicitly indicate how much of the $CARL_{IN}$ variability is due to the low $CARL_{IN}$'s and how much is due to the high $CARL_{IN}$'s, which is very important to know, because the goal is to minimize the low $CARL_{IN}$'s and maximize the high $CARL_{IN}$'s. On the other hand, the $CARL_{IN}$ prediction bounds can be used to provide boundaries to separate low $CARL_{IN}$*'s* from high $CARL_{IN}$*'s* values and to indicate the proportion of each. Hence, to study the effects of parameter estimation, we advocate using the EPC approach over the $SDCARL_{IN}$.

The EPC approach involves setting up the prediction bounds (or one-sided prediction intervals). As in Epprecht et al. (2015), an upper one-sided prediction bound for the $CARL_{IN}$ can be defined as

$$P(CARL_{IN}(Z, Y, k, h, m, n) \geq ARL_0) = 1 - p, \tag{4.12}$$

where ARL_0 is the specified lower prediction bound and p is a probability. A small value of p such as 0.05 is desirable, since it means that there is a 95% probability that the $CARL_{IN}$ will be greater than ARL_0. Thus, there will be a small proportion (5%) of low *CARL*$_{IN}$'s *(CARL*$_{IN}$'s that are less than ARL_0) and hence a large proportion (95%) of $CARL_{IN}$'s greater than ARL_0 This is desirable from a practical point of view.

To apply the EPC the distribution of the $CARL_{IN}$ has to be found or at least approximated. For time-weighted charts, such as the EWMA and CUSUM charts, the exact distribution is not available. Diko et al. (2019) approximated the $CARL_{IN}$ distribution of the two-sided EWMA chart by an empirical distribution (F_N). Their F_N was obtained by generating many Phase I subgroups, calculating the corresponding $CARL_{IN}$ values by using the Markov Chain method and then ordering them in ascending order. We will adopt this approach. However, instead of using the Markov Chain method to calculate the $CARL_{IN}$ values, we use equation (4.9). Once F_N has been found, we calculate its 100^{pth} percentile, denoted by $CARL_{IN,p}$. The $CARL_{IN,p}$, which is the observed (or the estimated) $CARL_{IN}$ upper prediction bound, is compared with the nominal ARL_0. The ARL_0 is the theoretical value that must be exceeded with high probability $1-p$. The comparison between $CARL_{IN,p}$ and ARL_0 is based on the percentage difference, given by

$$PD = \frac{CARL_{IN,p} - ARL_0}{ARL_0} \times 100. \tag{4.13}$$

The R code for generating F_N and the algorithm for finding PD are given in Diko et al. (2019b) and Appendix B, respectively, while the results are given in Table 4.2.

Table 4.2 shows the $CARL_{IN,p}$'s and PD's of the standard Phase II CUSUM control chart for $k = 0.25, 0.50$ and $ARL_0 = 200$ for different combinations of m, n and p. From Table 4.2, it can be seen that, for all n, nearly all the $CARL_{IN,p}$'s are less than or equal to ARL_0. Hence, nearly all of the PD's are negative. Thus, in addition to high $CARL_{IN}$'s, the observed $CARL_{IN}$ prediction bounds include low $CARL_{IN}$'s. For this reason, the Phase II CUSUM control charts, designed using the unadjusted h values (which are shown in Table 4.1), will have a lower than nominally expected $CARL_{IN}$, which means a deterioration in chart performance. It can also be seen that, for all n and small m, the $CARL_{IN,p}$ and PD are far less than ARL_0 and 0, respectively. Moreover, as m increases, $CARL_{IN,p}$ and PD converge to ARL_0 and 0, respectively.

Table 4.2. *$CARL_{IN,p}$ and PD values of a two-sided CUSUM chart as a function of m, n for $k = 0.25, 0.50$; $p = 0.05, 0.10$; and $ARL_0 = 200$.*

			M									
p	***n***	***k***	***30***	***50***	***200***	***500***	***750***	***1000***	***3000***	***5000***	***10000***	***50000***
0.05	***1***	***0.25***	28.78	41.96	92.09	125.68	138.25	145.98	168.29	175.42	182.71	192.52
			-85.61	-79.02	-53.95	-37.16	-30.87	-27.01	-15.86	-12.29	-8.65	-3.74
		0.50	24.66	37.55	85.39	117.92	130.25	138.37	162.73	170.89	179.48	191.52
			-87.67	-81.23	-57.31	-41.04	-34.88	-30.81	-18.64	-14.56	-10.26	-4.24
	5	***0.25***	37.31	53.46	114.68	150.72	162.31	168.98	184.84	188.97	192.75	197.27
			-81.34	-73.27	-42.66	-24.64	-18.85	-15.51	-7.58	-5.52	-3.63	-1.37
		0.50	44.42	64.69	126.02	155.64	164.79	170.20	184.30	188.37	192.40	197.65
			-77.79	-67.66	-36.99	-22.18	-17.60	-14.91	-7.85	-5.81	-3.80	-1.17
	10	***0.25***	38.27	54.57	117.23	154.31	166.27	172.95	188.23	191.82	194.90	198.29
			-80.87	-72.72	-41.38	-22.85	-16.87	-13.53	-5.88	-4.09	-2.55	-0.86
		0.50	47.61	69.24	134.80	164.19	172.76	177.52	189.16	192.30	195.28	198.97
			-76.19	-65.38	-32.56	-17.91	-13.62	-11.24	-5.42	-3.85	-2.36	-0.52
	20	***0.25***	38.56	55.17	118.20	155.95	168.04	174.71	190.05	193.45	196.18	198.93
			-80.72	-72.42	-40.90	-22.03	-15.98	-12.65	-4.98	-3.27	-1.91	-0.54
		0.50	48.91	71.24	138.59	168.75	177.18	181.75	192.15	194.75	197.07	199.80
			-75.55	-64.38	-30.71	-15.62	-11.41	-9.13	-3.92	-2.62	-1.46	-0.10
0.10	***1***	***0.25***	36.67	52.24	105.45	136.96	148.29	155.27	174.38	180.35	186.34	194.21
			-81.67	-73.88	-47.28	-31.52	-25.86	-22.37	-12.81	-9.83	-6.83	-2.89
		0.50	33.04	48.52	100.05	130.91	142.13	149.55	170.18	177.04	184.07	193.69
			-83.48	-75.74	-49.97	-34.54	-28.93	-25.23	-14.91	-11.48	-7.97	-3.15
	5	***0.25***	47.51	66.73	129.25	160.37	169.71	175.04	187.89	191.21	194.33	197.97
			-76.25	-66.64	-35.38	-19.81	-15.15	-12.48	-6.06	-4.40	-2.84	-1.02
		0.50	57.04	79.47	138.28	163.70	171.39	175.88	187.72	191.08	194.34	198.52
			-71.48	-60.27	-30.86	-18.15	-14.31	-12.06	-6.14	-4.46	-2.83	-0.74
	10	***0.25***	48.95	68.39	132.59	164.58	173.89	179.00	190.70	193.56	196.04	198.76
			-75.53	-65.81	-33.71	-17.71	-13.06	-10.50	-4.65	-3.22	-1.98	-0.62
		0.50	61.72	85.82	147.17	171.30	178.26	182.09	191.64	194.21	196.61	199.57
			-69.14	-57.09	-26.42	-14.35	-10.87	-8.95	-4.18	-2.90	-1.70	-0.21
	20	***0.25***	49.31	69.21	133.96	166.34	175.87	180.98	192.38	194.93	197.07	199.25
			-75.34	-65.40	-33.02	-16.83	-12.07	-9.51	-3.81	-2.54	-1.46	-0.37
		0.50	63.61	88.59	151.83	175.76	182.29	185.81	194.04	196.13	198.02	200.21
			-68.19	-55.71	-24.09	-12.12	-8.85	-7.10	-2.98	-1.93	-0.99	0.11

Thus, for all n, the effects of parameter estimation are more severe when m is small. Also, note that, for all n, the convergence to ARL_0 and zero occurs faster when $p = 0.10$ than when $p = 0.05$. Thus, the effects of parameter estimation are more evident when p is small (higher probability of exceedance) than when it is large (lower probability of exceedance). Using the same convergence argument, for $n = 1$, the effects of parameter estimation are even more pronounced for larger k values than for smaller k values. In

contrast, for $n > 1$, the opposite is true. Note that these conclusions are consistent with those of Saleh et al. (2016), who used the $SDCARL_{IN}$ metric. Thus, even though our conditional analysis has used a criterion that is different from the criterion that was used by Saleh et al. (2016), we arrive at the same general conclusions on this point.

Based on these results, we can already infer that huge numbers of Phase I subgroups are required to make $CARL_{IN,p}$ close enough to ARL_0 or PD close to 0. Next, the question of how large m should be, is investigated in more detail.

4.5 Required number of Phase I samples.

For the upper one-sided CUSUM control chart, Jeske (2016) determined the minimum *m* required to guarantee with $1-\alpha$ probability that the relative error (RE) will be different from 0 by at most ε, tha is,

$$P(|RE| < \varepsilon) = 1 - \alpha$$

$$P\left(-\varepsilon < \frac{CARL_{IN}(Z,Y,k,h,m,n) - ARL_0}{ARL_0} < \varepsilon\right) = 1 - \alpha \qquad (4.14)$$

$$P\big(ARL_0(1-\varepsilon) < CARL_{IN}(Z,Y,k,h,m,n) < ARL_0(1+\varepsilon)\big) = 1 - \alpha.$$

Note that this is not equal to the EPC. In this situation, the quantity α is the sum of the lower and higher tail probabilities of the distribution of $CARL_{IN}$. Thus, the m-values that are recommended on the basis of equation (4.14) do not explicitly protect against high proportions of low $CARL_{IN}$'s.

In Saleh et al. (2016) the number of samples (m) recommendations were made to keep the $SDCARL_{IN}$ down at some acceptable nominal level, $SDARL_0 = ARL_0 \times \varepsilon$, where $\varepsilon = 0.1$. Note that, the $SDCARL_{IN}$ is calculated from the deviations of $CARL_{IN}(Z,Y,k,h,m,n)$ from $AARL_{IN}$ and the $AARL_{IN}$ is not necessarily equal to ARL_0, unless m is very large. If m is large, then $CARL_{IN}(Z,Y,k,h,m,n) \approx Normal(AARL_{IN}, SDARL_{IN})$, because of the central limit theorem. Consequently, since the area under the normal curve, within one standard deviation of the mean, is approximately 68%, it can be said that Saleh et al. (2016) m recommendations were made to guarantee with 0.68 probability that the difference

between $CARL_{IN}(Z, Y, k, h, m, n)$ and $AARL_{IN}$ (or ARL_0, when m is very large) will be no more than $ARL_0 \times \varepsilon$. This makes the $SDCARL_{IN}$ criterion a special case of equation (4.14) with $1 - \alpha = 0.68$, which implies a lower (and an upper) tail probability ($p = \alpha/2$) of approximately 0.16. To illustrate this in more detail, for the charts ($n = 1, k = 0.25, h = 6.854$), ($n = 1, k = 0.50, h = 4.172$), ($n = 5, k = 0.25, h = 6.854$) and ($n = 5, k = 0.50, h = 4.172$), Saleh et al. (2016)'s recommended $m = 3000, m = 5000, m = 600 \; and \; m = 800$, respectively. Based on these sample sizes, we calculated the $AARL_{IN}$, $SDCARL_{IN}$, $P\big(CARL_{IN} < AARL_{IN}(1-\varepsilon)\big)$ and $P\big(CARL_{IN} < ARL_0(1-\varepsilon)\big)$ of the four CUSUM charts. The results are shown in Table 4.3.

Table 4.3. Four Phase II CUSUM control charts for $ARL_0 = 200$. Each constructed using the parameters known case limits with m values that satisfy $ARL_0 \times \varepsilon$, where $\varepsilon = 0.1$.

	(n=1,m=3000,k=0.25,h=6.854)	***(n=1,m=5000,k=0.50,h=4.172)***
$\boldsymbol{AARL_{IN}}$	199.3	202
$\boldsymbol{SDCARL_{IN}}$	20.2	20.3
$\boldsymbol{P\big(CARL_{IN} < AARL_{IN}(1-\varepsilon)\big)}$	0.16	0.16
$\boldsymbol{P\big(CARL_{IN} < ARL_0(1-\varepsilon)\big)}$	0.17	0.13
	(n=5,m=600,k=0.25,h=6.854)	***(n=5,m=800,k=0.50,h=4.172)***
$\boldsymbol{AARL_{IN}}$	190.9	197.6
$\boldsymbol{SDCARL_{IN}}$	20.8	20.2
$\boldsymbol{P\big(CARL_{IN} < AARL_{IN}(1-\varepsilon)\big)}$	0.16	0.16
$\boldsymbol{P\big(CARL_{IN} < ARL_0(1-\varepsilon)\big)}$	0.29	0.19

From Table 4.3, it can be seen that the four CUSUM charts are performing as nominally specified, in terms of the $SDCARL_{IN}$ being $\varepsilon = 10\%$ of $ARL_0 = 200$. Except for the chart with ($n = 5, m = 600, k = 0.25, h = 6.854$), all the other charts have $AARL_{IN}$'s that are very close to $ARL_0 = 200$. Furthermore, for all charts, it can be observed that $P\big(CARL_{IN} < AARL_{IN}(1-\varepsilon)\big) = 0.16$, so the central limit theorem does apply. Moreover, since $AARL_{IN} \cong ARL_0$, it can be seen that $P\big(CARL_{IN} < ARL_0(1-\varepsilon)\big) \cong 0.16$, which confirms our explanation. Therefore, even though the probability p is not explicitly mentioned in Saleh et al. (2016) , their "$SDCARL_{IN} = ARL_0 \times \varepsilon$" based m values do offer some protection against the frequent occurrence of low $CARL_{IN}$'s. But, this protection

might not be enough since the probability of an unsatisfactory $CARL_{IN}$ is usually much higher than what might be deemed acceptable. When ε is made smaller, for example, for the $(n = 1, m = 5000, k = 0.50, h = 4.172)$ chart, and $\varepsilon = 0.05$ then $p = 37\%$. And for $\varepsilon = 0$ then $p = 57\%$. This suggests that in practice, one should better consider the EPC instead of the $SDCARL_{IN}$ as a chart performance criterion.

Given ε, most practitioners would be interested in higher levels of protection, such as $p = 0.05, 0.10$, (so that the exceedance probability in (12) is high, such as 90% or 95%) and would naturally want to know about the minimum m required to achieve these levels. In technical terms, the practitioner may want to know the minimum m required to guarantee with probability $1 - p$ that the RE will be less than 0 by at most ε, where p is an explicitly specified small proportion, such as *0.05 or 0.10*. To answer this, a method that explicitly controls the lower tail area, p, of the $CARL_{IN}$ distribution is needed. This can be derived from a modification of equation (4.14) by simply considering only the lower prediction bound (and not the interval)

$$P(RE > -\varepsilon) = 1 - p$$

$$P\left(\frac{CARL_{IN}(Z, Y, k, h, m, n) - ARL_0}{ARL_0} > -\varepsilon\right) = 1 - p \qquad (4.15)$$

so that $$P\big(CARL_{IN}(Z, Y, k, h, m, n) > ARL_0(1 - \varepsilon)\big) = 1 - p.$$

This, indeed, is the EPC that was used by Epprecht et al. (2015), Loureiro et al. (2018) and Diko et al. (2019a) to recommend Phase I subgroup numbers for Phase II Shewhart and EWMA control charts. Thus, for a given set of k, h, n, p, ε, and ARL_0 values, we solve equation (4.15) for m. The algorithm is given in Diko et al. (2019b). The results are shown in Table 4.4.

As in Loureiro et al. (2018), Table 4.4 reveals that, for a given ARL_0, m varies substantially with n, ε and p, and m decreases with an increase of any of these parameters. For example, for *n = 1*, $ARL_0 = 200, k = 0.25, \varepsilon = 0.10$, m varies from about 7600 (when $p = 0.05$) down to 4800 (when $p = 0.10$). Thus, more Phase I data are required to make p smaller than 0.16. It can also be seen that, for $n = 1$, large k values go together with the larger m values and vice versa, while, for $n > 1$, large values of k are associated with small m and vice versa. These results are consistent with the results in Table 4.2. Thus, from a

practical point of view, an adjustment to the CUSUM control limit h is necessary when parameters are estimated with reasonable amounts of Phase I data, in order to avoid a large number of short in-control run lengths.

Table 4.4 *Required minimum number of Phase I subgroups, m, as a function of ε, p, k, n and $ARL_0 = 100, 200, 370$.*

			ε = 10%		ε = 20%		ε = 30%	
ARL_0	n	k	p = 0.05	p = 0.10	p = 0.05	p = 0.10	p = 0.05	p = 0.10
100	**1**	**0.25**	5300	3350	1300	900	550	400
		0.50	7550	4700	1850	1200	800	500
	5	**0.25**	1300	900	450	300	250	200
		0.50	1440	1000	400	300	250	200
	10	**0.25**	1050	700	350	260	200	150
		0.50	850	600	300	233	200	150
	20	**0.25**	900	650	350	260	200	150
		0.50	650	450	250	180	200	150
200	**1**	**0.25**	7600	4800	1850	1250	800	550
		0.50	10500	6500	2600	1650	1100	700
	5	**0.25**	1950	1400	700	500	350	250
		0.50	2000	1350	600	450	300	200
	10	**0.25**	1500	1100	600	450	350	250
		0.50	1200	850	450	300	250	150
	20	**0.25**	1300	950	550	400	350	250
		0.50	900	650	350	250	200	150
370	**1**	**0.25**	10650	6850	2650	1800	1150	800
		0.50	14050	8750	3450	2200	1450	950
	5	**0.25**	2900	2050	1050	750	550	400
		0.50	2700	1800	800	600	400	300
	10	**0.25**	2250	1600	950	650	550	400
		0.50	1650	1150	600	450	350	250
	20	**0.25**	2050	1450	900	600	500	350
		0.50	1200	900	500	350	300	200

The take home message is that the required number of Phase I subgroups given in Table 4.4 provide higher levels of protection than their $SDCARL_{IN}$ based counterparts given in Saleh et al. (2016) (see also Table 4.3). But these are huge numbers. Practitioners with such large amounts of data at their disposal may not need to adjust the h-values for the effects of parameter estimation. They can simply use Table 4.2 and the parameter known case charting constants to design Phase II CUSUM control charts with guaranteed conditional performance. However, in applications, data may be limited or scarce and in such applications, protection against low $CARL_{IN}$'s can be achieved by adjusting the control

limits. Motivated by this, we adjust the CUSUM control charts limits according to the EPC for small to moderate values of *m*, and we present the results graphically.

4.6 Adjusting the CUSUM limits

To reduce the problem of low $CARL_{IN}$, we adjust the control limits of the two-sided Phase II CUSUM control chart according to the EPC. We present the adjusted control limits, graphically, for $n = 5, m = 30, 50, 100, 200$, $0.12 \leq k \leq 0.60$, $ARL_0 = 100, 200, 370$ $\varepsilon = 0, 0.20$, and $p = 0.05;\ 0.10$ for convenience of the users since not all values can be tabulated and the graphs can be used to approximate the adjusted charting constants in cases not covered. Note that, our method of applying the EPC does not involve bootstrapping. Also, the method assumes that the form of the underlying process distribution is known, e.g. normal. The algorithm for finding the adjusted charting constants using our method and, for comparison reasons, the parametric bootstrap approach, are given in Diko et al. (2019b). It is seen that our method is more logical and less arduous when compared to parametric bootstrapping. Moreover, as was observed for the one-sided CUSUM (cf. Saleh et al. (2016)), we also observe that the bootstrap based adjusted limits vary. Whereas, our adjusted limits (see, for example, Figures 4.4 to 4.6) are virtually constant. Hence, where analytical methods cannot be found, our method is a better alternative than parametric bootstrapping.

Figures 4.4 to 4.6 present the adjusted charting limits, calculated according to (Diko et al 2019b). As expected, the adjusted limits are greater than their corresponding unadjusted limits. Furthermore, for a given ARL_0, increases in ε, p and m, cause the adjusted limits to converge to the unadjusted limits. Also, for given values of ε, p, m, decreasing ARL_0 causes the adjusted limits to converge to the unadjusted limits. Note also that, for larger k, the adjusted charting limits are much closer to the unadjusted limits compared to smaller k. Next, the application of the adjusted and unadjusted control charts is illustrated using some data from Montgomery (2013). The in-control (IC) and out-of-control (OOC) performance trade-offs between the adjusted and unadjusted limits are illustrated. A way to balance them is suggested.

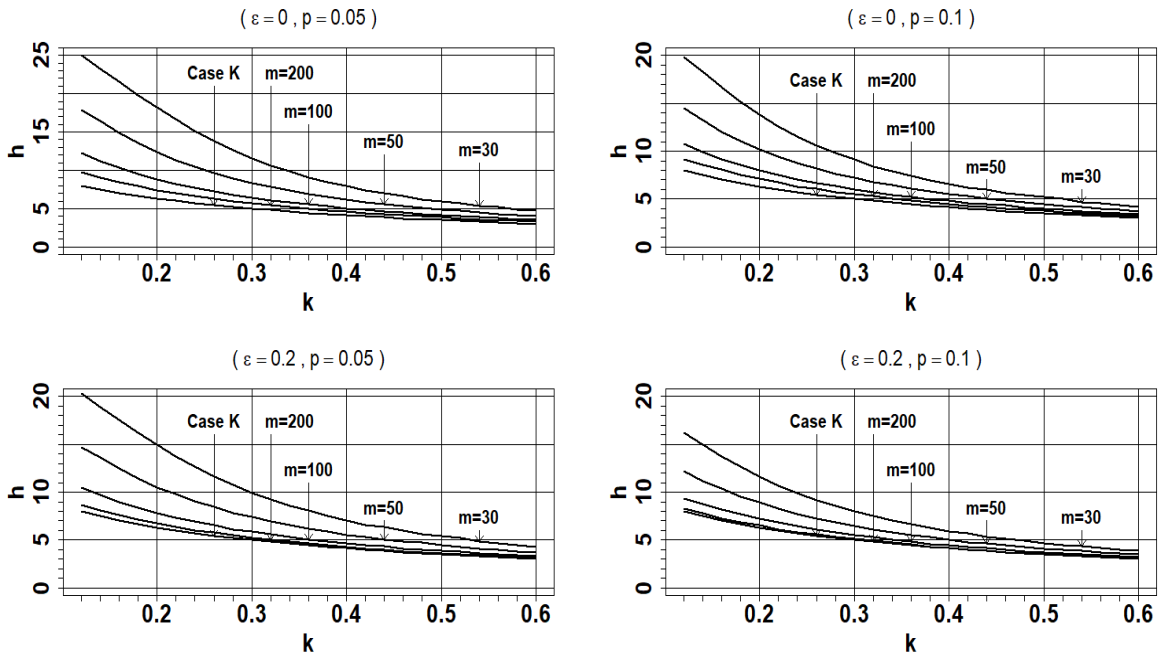


Figure 4.4. *Graphs of the unadjusted and adjusted h values for* $0.12 \leq k \leq 0.60$; $ARL_0 = 100$ *and* $n = 5$. *The adjusted L were generated to guarantee* $\left(CARL_{IN} > ARL_0(1-\varepsilon)\right) = 1 - p$.

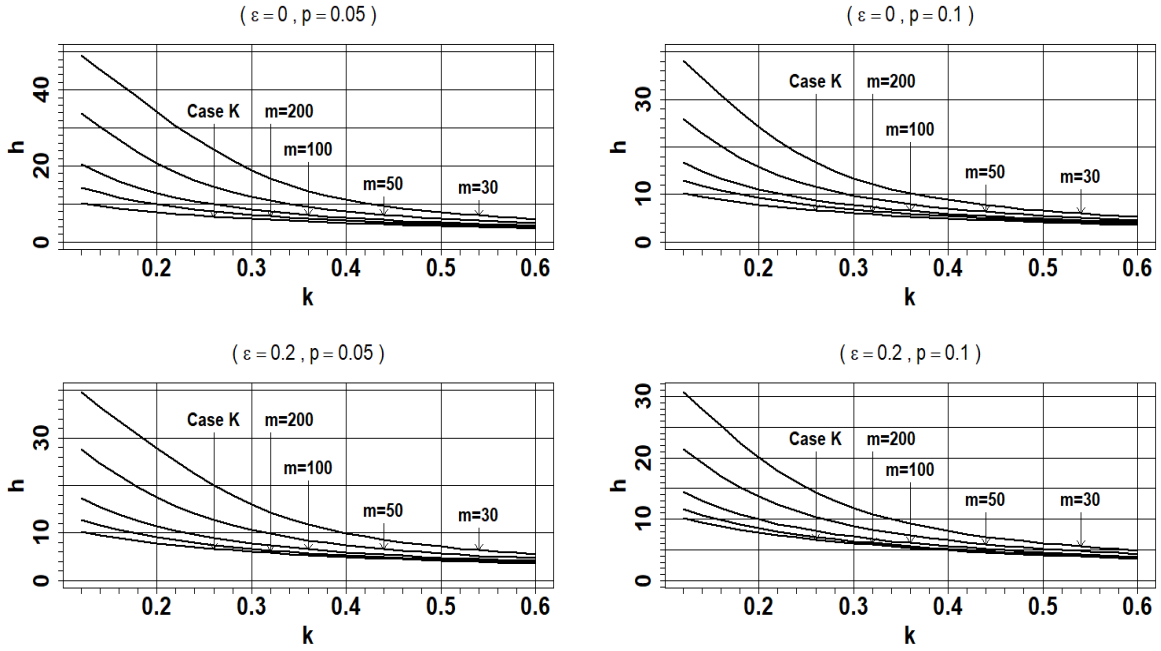


Figure 4.5. *Graphs of the unadjusted and adjusted h values for* $0.12 \leq k \leq 0.60$; $ARL_0 = 200$ *and* $n = 5$. *The adjusted L were generated to guarantee* $P\left(CARL_{IN} > ARL_0(1-\varepsilon)\right) = 1 - p$.

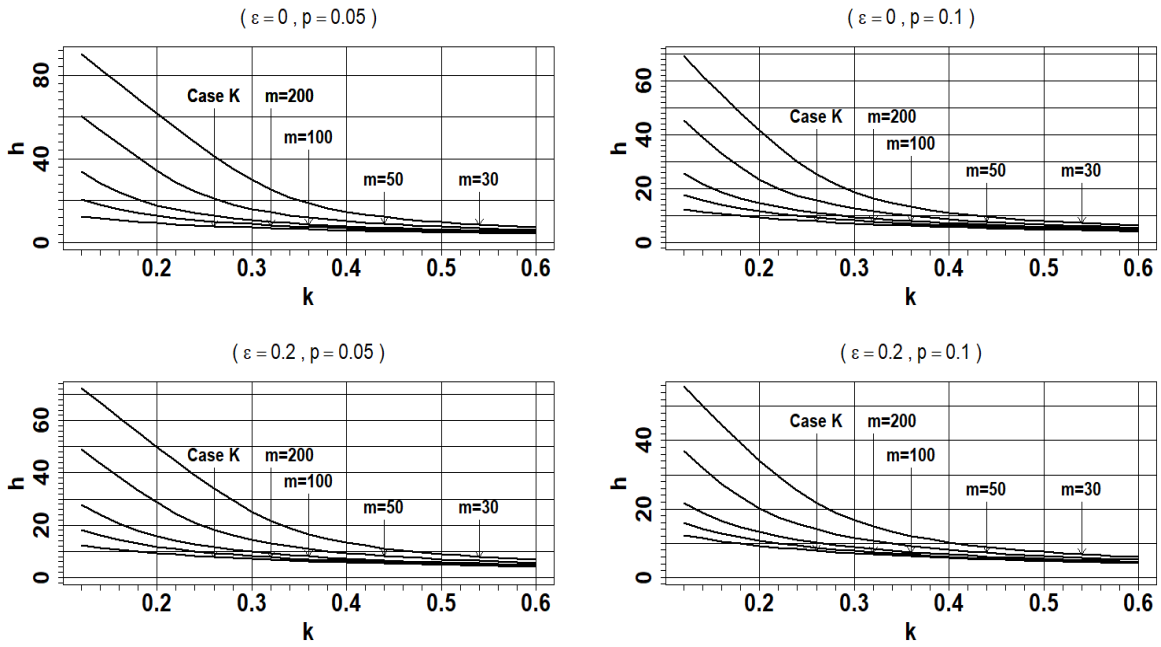


Figure 4.6. *Graphs of the unadjusted and adjusted h-values for* $0.12 \leq k \leq 0.60$; $ARL_0 = 370$ *and* $n = 5$. *The adjusted L were generated to guarantee* $\left(CARL_{IN} > ARL_0(1-\varepsilon)\right) = 1 - p$

4.7 An illustration, with some data, of the performance of using the adjusted and unadjusted limits

For our illustration, we consider the hard bake process data in Montgomery (2013). The data consist of 45 samples of size $n = 5$. We use the first 30 *IC* samples as our Phase I data to calculate the mean and standard deviation estimates. We found the Phase I mean and standard deviation estimates to be 1.504 and 0.138, respectively. Table 4.5 shows the sample means of the remaining 15 Phase II samples. Table 4.5 also uses these samples to calculate the standardized Phase II charting statistics. We have plotted the standardized Phase II charting constants in Figure 4.7. Figure 4.7 shows the standardized two-sided Phase II CUSUM control chart, with two sets of control limits, the adjusted limits and the unadjusted limits. The adjusted limits were taken from Figure 4.5 for $\varepsilon = 0, p = 0.1, k = 0.50$ and $m = 30$. The unadjusted limits were obtained from Table 4.1, for $ARL_0 = 200$ and $k = 0.50$.

Table 4.5. *Hard bake process data for Phase II and the calculations for the standardized CUSUM charting statistics.*

Phase II Sample #	$\bar{X}$	W	k	C^-	C^+
31	1,472	-1,24707	0,5	-1,84094	0
32	1,5292	1,012808	0,5	-1,32813	1,512808
33	1,5317	1,112275	0,5	-0,71585	3,125083
34	1,57934	3,007709	0,5	0	6,632792
35	1,4279	-3,01758	0,5	-3,51758	4,115216
36	1,48238	-0,85	0,5	-4,86758	3,765214
37	1,49098	-0,50784	0,5	-5,87541	3,757378
38	1,61278	4,338173	0,5	-2,03724	8,595551
39	1,65598	6,056955	0,5	0	15,15251
40	1,64202	5,501534	0,5	0	21,15404
41	1,67156	6,676831	0,5	0	28,33087
42	1,62516	4,830732	0,5	0	33,6616
43	1,69696	7,687411	0,5	0	41,84901
44	1,63214	5,108442	0,5	0	47,45746
45	1,77	10,59343	0,5	0	58,55088

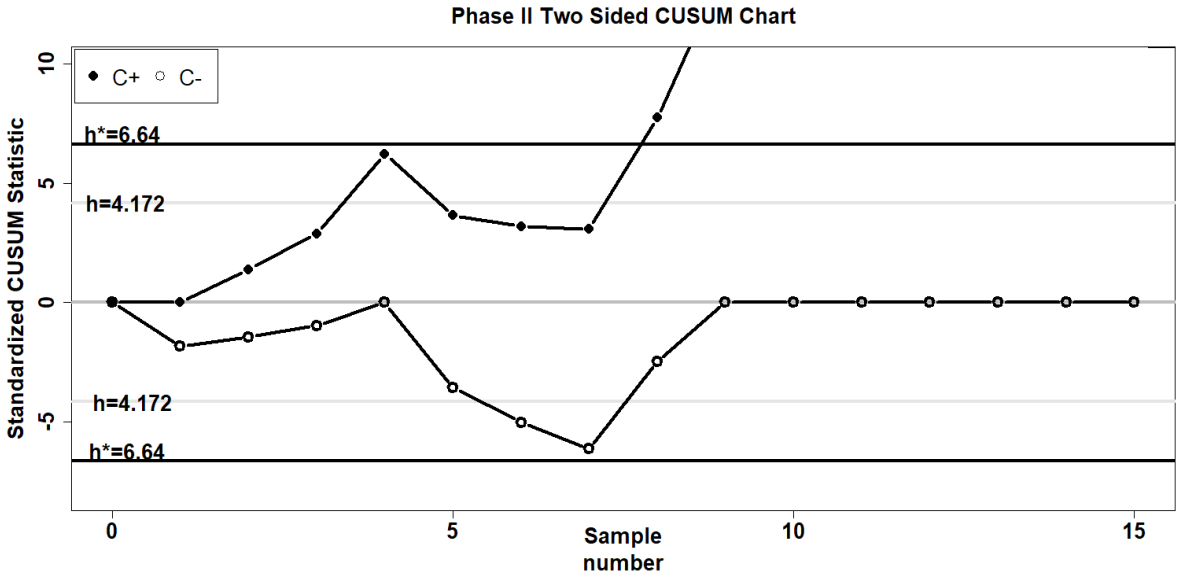


Figure 4.7. *Phase II CUSUM Control chart to monitor the mean of the hard bake process data with adjusted (h^*) and unadjusted limits (h) when $m = 30, n = 5$, $ARL_0 = 200$ and $p = 0.10$.*

From Figure 4.7, it can be seen that, since the adjusted limits are wider than the unadjusted limits, the unadjusted limits signal three times (at sample numbers 4, 6 and 7) before the adjusted limits give their first signal on sample number 8. Thus, if the process is IC then the adjusted limits would have an advantage over the unadjusted limits in terms of

false alarms. However, if the process is OOC then the unadjusted limits would have an advantage.

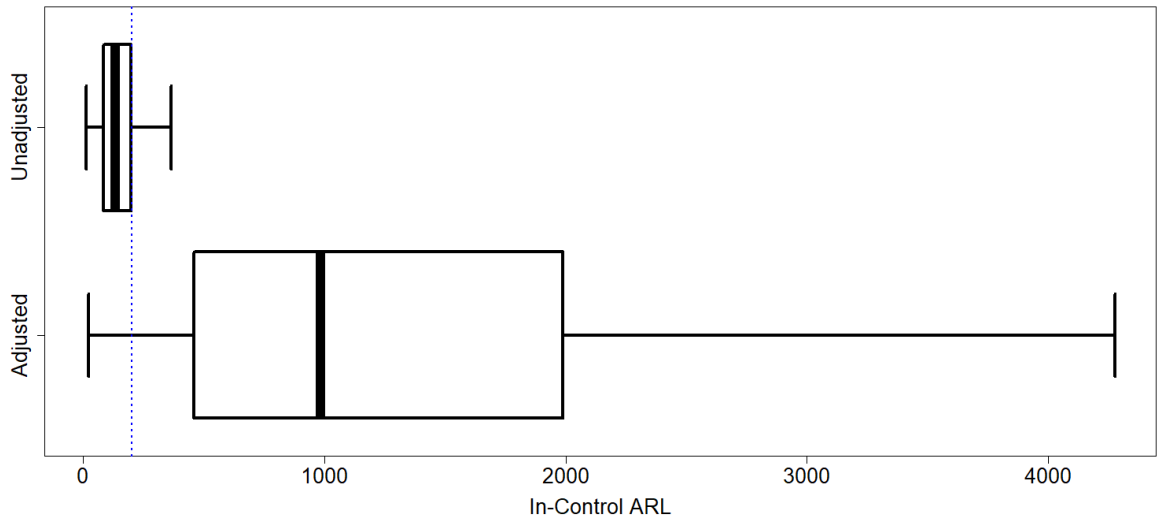


Figure 4.8. *Boxplots of the $CARL_{IN}$ distribution for the adjusted and unadjusted limits two-sided Phase II CUSUM chart for the mean of the hard bake process data.*

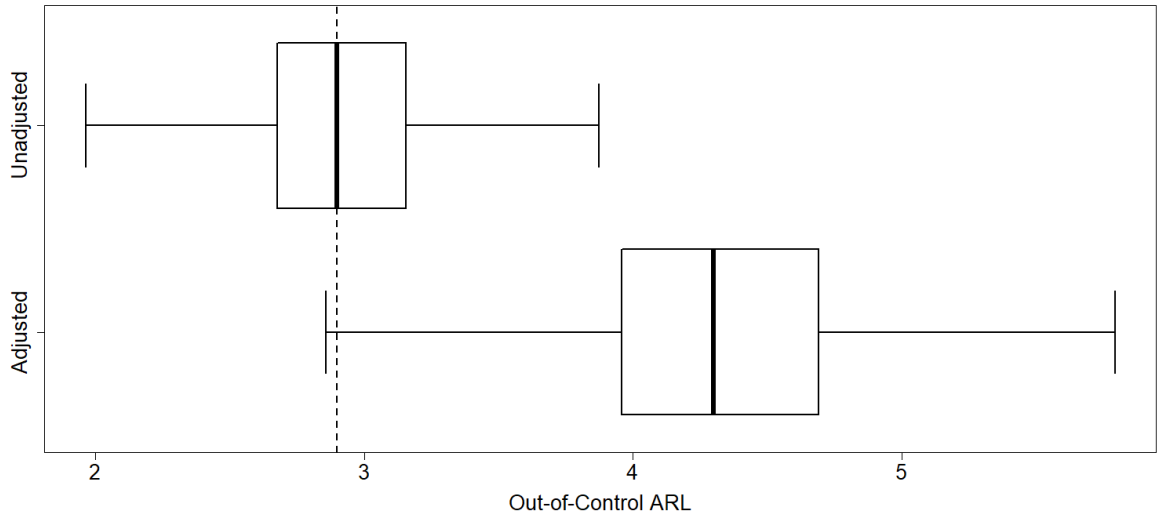


Figure 4.9. *Boxplots of the $CARL_{OOC}$ distribution for the adjusted and unadjusted limits two-sided Phase II CUSUM chart for the mean of the hard bake process data.*

Figures 4.8 and 4.9 show the boxplots of the $CARL_{IN}$ and $CARL_{OOC}$ distribution, respectively, of the adjusted and unadjusted limits two-sided Phase II CUSUM chart for the

mean of the hard bake process data. The vertical line in Figure 4.8 is drawn at $CARL_{IN} = 200$. From Figure 4.8, it can be seen that the unadjusted control limits chart have a very high proportion of low $CARL_{IN}$ values (values less than 200). On the other hand, the adjusted control limits chart have a very low proportion of low $CARL_{IN}$. Thus, with regards to the IC performance, the adjusted limits have an advantage over the unadjusted limits. Now, looking at Figure 4.9, where a shift of 1 standard deviation in the mean is introduced ($\delta = 1$), it can be seen that the median of the $CARL_{OOC}$ distribution of the unadjusted limits is 2.9, while that of the $CARL_{OOC}$ distribution of the adjusted limits is 4.3. Thus, the adjusted limits have a slightly worse performance. Even though the shift from 2.9 to 4.3 is not so dramatic, it is possible to bring 4.3 closer to 2.9. For a given ARL_0 and k, this can be done by using smaller adjusted limits.

There are two ways to find the smaller adjusted limits. First, for given m, smaller adjusted limits can be found by increasing ε and/or *p.* In our case, since we are using $(\varepsilon = 0, p = 0.10, m = 30, ARL_0 = 200)$, this means that we can use an adjusted charting limit that corresponds to $(\varepsilon = 0.20, p = 0.10, m = 30, ARL_0 = 200)$. This will improve the OOC performance of the adjusted limits, but the IC performance will be lower than the nominally specified. However, it is still way better than the IC performance with the unadjusted limits. Secondly, for given constants ε and p, smaller adjusted limits can also be found by increasing m. In our case, since we are using $(\varepsilon = 0, p = 0.10, m = 30, ARL_0 = 200)$, this means using the adjusted limits that correspond to $(\varepsilon = 0, p = 0.10, m = 50, ARL_0 = 200)$ or $(\varepsilon = 0, p = 0.10, m = 100, ARL_0 = 200)$ or $(\varepsilon = 0, p = 0.10, m = 200, ARL_0 = 200)$. This will improve the OOC performance drastically and, unlike the previous method above, the *IC* performance remains as nominally specified. To see how drastic the OOC can improve, notice how close the $(\varepsilon = 0, p = 0.10, m = 100, ARL_0 = 200)$ and $(\varepsilon = 0, p = 0.10, m = 200, ARL_0 = 200)$ are to the parameters known case line in Figure 4.5,.

The issue is of course how to attain $m = 50$ or even $m = 200$ when we have only $m = 30$ Phase I subgroups. To solve this problem, we suggest first using the $(\varepsilon = 0, p = 0.10, m = 30, ARL_0 = 200, k = 0.50)$ adjusted charting limit, as in Figure 4.7. Charting should continue until twenty subgroups plot within the adjusted control limits. These 20 additional subgroups should be in-control and can be combined with the *m=30* Phase I

subgroups. The resulting 50 subgroups can then be used to update the Phase I parameter estimates, which should now be used with the $(\varepsilon = 0, p = 0.10, m = 50, ARL_0 = 200, k = 0.50)$ adjusted charting limit. This updating process can be repeated until 200 IC subgroups have been collected. At this point the OOC performance of the adjusted limits will be nearly equal to that of the unadjusted limits and the IC performance will be as nominally specified.

4.8 Summary and conclusions

In this chapter, we extend the work of Jeske (2016), by deriving the modified Siegmund formula for the two-sided Phase II CUSUM control chart for the mean assuming normality with unknown mean and standard deviation. We show the simplicity, accuracy and the versatility of this formula for calculating various popular CUSUM control chart performance measures such as the $AARL_{IN}$ and $SDCARL_{IN}$. We argue in favor of the usefulness of the EPC based $CARL_{IN}$ prediction bounds to study the effects of parameter estimation. Based on these prediction bounds, it was seen that even more Phase I data are required than previously recommended using the $SDCARL_{IN}$ criteria. Then we have adjusted the control chart limits of the two-sided CUSUM control chart according to the EPC. Like the bootstrap method, our method involves, in a different way, simulations and a search algorithm. However, unlike the bootstrap method, the method has a search algorithm that has a known lower bound. The solution we seek is very close to this bound. This makes the method faster than bootstrapping. In addition, the different way in which the simulations are done makes the method more accurate and consistent than the bootstrap method. Using the new method, we adjusted the charting limits according to the EPC. Graphs of the adjusted limits were provided for an extensive range of chart parameters. This will help practitioners to implement the two-sided Phase II CUSUM chart more easily. The graphs proved useful in making decisions on how to balance the IC and OOC performance of the adjusted control limits. In order to balance the trade-off between the IC and OOC performance of the adjusted limits we recommend increasing m and updating the control limits accordingly (using Figures 4.4 – 4.6).

5. A head-to-head comparison of the out-of-control performance of control charts adjusted for parameter estimation.

5.1 Introduction

Control charts are an important tool to aid in the detection of process changes. Three commonly-used types of control charts are the Shewhart, cumulative sum (CUSUM), and exponentially weighted moving average (EWMA) charts (see Chapters 2 through 4). Each of these charts have their own characteristics which make them more or less applicable to detect certain types of shifts. For example, the Shewhart control chart is better suited to detect large shifts, while the CUSUM and EWMA yield better detection capabilities against small sustained shifts. These aspects increase the demand for comparative studies, where the control chart capabilities are evaluated based on different scenarios.

To this end, Hawkins and Wu (2014) performed a comparative study of the CUSUM and EWMA control charts. They found that the CUSUM outperforms the EWMA in the case that the actual shift to be expected is (approximately) known in size. However, their comparison is based on the assumption of known in-control process parameters. In practice, these parameters have to be estimated using a Phase I reference sample. Since different practitioners obtain different Phase I samples their parameter estimates will differ, leading to different estimated control limits. The performance of the control chart is then conditional on these obtained estimates. This effect of Phase I estimation has received much attention in recent literature. Zwetsloot and Woodall (2017) perform a comparative study on the conditional performance of the Shewhart, CUSUM, and EWMA control charts, where they compare the effect of estimation error across these charts. They conclude that the Shewhart chart is most affected by estimation error, and that the EWMA and CUSUM charts behave quite differently when evaluated on conditional performance. They advise to consider the conditional performance for an appropriate comparison. This is the main topic of this chapter. Additionally, it is possible to adjust the control chart design to match a certain performance criterion when parameters are estimated. The performance is generally measured in terms of the false alarm rate (FAR) or the average run length (ARL). This

approach has been proposed by various researchers mentioned in earlier chapters. Albers and Kallenberg (2004a, 2004b) introduced the *exceedance probability criterion*, which aims to provide a specified minimum in-control performance with a specified probability. This criterion has gained increasing attention in the more recent literature on statistical process monitoring (SPM), and was recommended by Jones and Steiner (2012), Gandy and Kvaløy (2013), and Saleh et al (2015a, 2015b), amongst others. Moreover, bootstrap, analytical, and numerical (approximation) methods have been applied to determine the required control limits for Shewhart, CUSUM and EWMA control charts. In Chapters 3 and 4, this is done for the CUSUM and EWMA control charts for location.

With these new conditional control chart designs for location, the question arises how it influences the performance of the Shewhart, CUSUM and EWMA control charts in various scenarios. In this chapter we therefore perform a comparative study between these control charts with estimated parameters. Because the exceedance probability criterion provides a specified in-control performance, we focus mainly on the out-of-control detection capability of the control charts.

This Chapter has been based on Diko et al. (2019c) and is organized as follows. In Section 5.2 we define once more the conditional Shewhart, CUSUM and EWMA control chart designs, the estimators used in Phase I estimation, and our evaluation criteria. In Section 5.3 we explain our simulation procedures and discuss the corresponding results. Finally, in Section 5.4 we provide some concluding remarks.

5.2 Shewhart, CUSUM and EWMA control charts

In this chapter we consider the Shewhart, CUSUM and EWMA control charts for location. For each of these charts, the first step (Phase I) is to estimate the in-control behavior of the underlying process, before one can start the monitoring phase (Phase II). Suppose that in Phase I, an in-control sample consisting of *m* subgroups of size *n* is available to estimate the in-control behavior of the process. Let X_{ij} (for *i* = 1,...,*m* and j = *1,…,n*) be the *j*-th observation in the *i*-th subgroup in Phase I, where X_{ij} follows a normal distribution with mean μ_0 and standard deviation σ_0 (i.e., $N(\mu_0, \sigma_0)$). In Phase II ($i = m+1, m+2, \ldots$), we assume that there may be a shift in the mean of δ standard deviations, such that the observations are drawn from a $N(\mu_0 + \delta\sigma_0, \sigma_0)$ distribution. We denote the Phase II mean

as $\mu = \mu_0 + \delta\sigma_0$. Note that $\delta = 0$ corresponds to the in-control situation. This setup is similar to that of the one used in the comparison papers of Hawkins and Wu (2014) and Zwetsloot and Woodall (2017).

Since the values of μ_0 and σ_0 are generally unknown, they have to be estimated. As an estimator for μ_0 we use the overall sample mean:

$$\hat{\mu}_0 = \sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}/mn \tag{5.1}$$

As an estimator for σ_0 we use the pooled sample standard deviation, defined as

$$\hat{\sigma}_0 = \sqrt{\frac{1}{m}\sum_{i=1}^{m} S_i^2} \tag{5.2}$$

where S_i^2 is the *i*-th subgroup sample variance, i.e. $\frac{1}{n-1}\sum_{i=1}^{n}\left(X_{ij} - \bar{X}_i\right)^2$. Note that the pooled standard deviation is only applicable when $n > 1$. When $n = 1$, one has to resort to other estimators of σ_0 such as the moving range. However, we do not consider this situation explicitly in this chapter.

In Phase II we monitor the process characteristic

$$B_i = \frac{\bar{X}_i - \hat{\mu}_0}{\hat{\sigma}_0/\sqrt{n}} \tag{5.3}$$

and we want to detect changes in the location parameter μ_0. Similarly to Jones, Champ, and Rigdon (2001), we expand this equation for comparison purposes into

$$B_i = \frac{\bar{X}_i - \hat{\mu}_0}{\hat{\sigma}_0/\sqrt{n}} = \left[T_i + \delta\sqrt{n} - \frac{Z}{\sqrt{m}}\right] Q^{-1}, \tag{5.4}$$

where $T_i = \frac{(\bar{X}_i - \mu)}{\sigma_0/\sqrt{n}}$, $\delta = \frac{\mu - \mu_0}{\sigma_0}$, $Z = \frac{\sqrt{mn}(\hat{\mu}_0 - \mu_0)}{\sigma_0}$, and $Q = \frac{\hat{\sigma}_0}{\sigma_0}$. Note that T_i represents the standardized difference between the mean of Phase II subgroup i and the Phase II population mean, δ represents the size of the Phase II mean shift, Z represents the estimation error of the mean in Phase I, and Q represents the Phase I estimation error of the standard deviation. Note also that, for the chosen estimators as in (5.1) and (5.2), T_i and Z are standard normal variables, and that $m(n-1)Q^2$ follows a chi-square distribution with $m(n-1)$ degrees of freedom.

5.2.1 Shewhart control chart

The Shewhart control chart consists of a lower control limit (LCL) and an upper control limit (UCL), and a charting statistic (cf. Montgomery (2013)). The charting statistic in Phase II at time period $i = m+1, m+2, \ldots$, is equal to the i-th subgroup average $\bar{X}_i$. The chart provides a signal when either $\bar{X}_i > \widehat{UCL}$ or $\bar{X}_i < \widehat{LCL}$, where $\widehat{LCL}$ and $\widehat{UCL}$ are the estimated control limits defined as

$$\begin{aligned} \widehat{LCL} &= \hat{\mu}_0 - c\,\frac{\hat{\sigma}_0}{\sqrt{n}} \\ \widehat{UCL} &= \hat{\mu}_0 + c\,\frac{\hat{\sigma}_0}{\sqrt{n}} \end{aligned} \tag{5.5}$$

Note that this is equivalent to a signal when $B_i > c$ or $B_i < -c$. Here c is some constant determined such that the control chart yields a specified in-control performance. For example, when parameters are known and would not require estimation, a value of $c = 3$ would provide a false alarm rate of 0.0027, or equivalently an in-control average run length (ARL_0) of 370.4. When parameters are estimated the conditional probability of an alarm (CPA), conditional on the estimates Q and Z, and given the size of shift δ, is equal to

$$\begin{aligned} CPA &= P(B_i > c \mid Q, Z, \delta) + P(B_i < -c \mid Q, Z, \delta) \qquad (5.6) \\ &= P\left(\left[T_i + \delta\sqrt{n} - \frac{Z}{\sqrt{m}}\right]Q^{-1} > c \mid Q, Z, \delta\right) + P\left(\left[T_i + \delta\sqrt{n} - \frac{Z}{\sqrt{m}}\right]Q^{-1} < -c \mid Q, Z, \delta\right) \\ &= P\big(T_i > cQ + Z/\sqrt{m} - \delta\sqrt{n} \mid Q, Z, \delta\big) + P\big(T_i < -cQ + Z/\sqrt{m} - \delta\sqrt{n} \mid Q, Z, \delta\big) \\ &= 1 - \Phi\big(cQ + Z/\sqrt{m} - \delta\sqrt{n}\big) + \Phi\big(-cQ + Z/\sqrt{m} - \delta\sqrt{n}\big), \end{aligned}$$

where Φ represents the standard normal cdf. The conditional run length distribution of this control chart is then equal to a geometric distribution with parameter CPA, so that the conditional average run length ($CARL$) is equal to $1/CPA$. Note that, because Z and Q are in practice unknown, the CPA and $CARL$ are not known to the practitioners, but can only be used for a general performance evaluation as in this chapter.

5.2.2 CUSUM control chart

Like in Chapter 4, we consider the two-sided CUSUM chart of Page (1954). This CUSUM control chart plots two charting statistics at the same time: $C_i^+ = \max(0, C_{i-1}^+ + W_i - k)$ and $C_i^- = \min(0, C_{i-1}^- + W_i + k)$. As starting values we choose $C_0^+ = C_0^- = 0$. The

value k is usually chosen to be half the shift that is considered important enough to be detected, measured as number of standard deviations from the mean. The CUSUM control chart provides a signal when either $C_i^+ > h$ or $C_i^- < -h$, where the constant h is determined such that the control chart yields a specified in-control performance.

In this chapter we consider CUSUM charts designed to detect shifts in the mean equal to 0.5 (small), 1 (medium), and 1.5 (large) standard deviations of the monitoring statistic, which means the corresponding values of k are equal to $0.25, 0.5$ and 0.75 respectively. When parameters are known, the corresponding values of h for an in-control $ARL_0 = 370$ are obtained using the function *xcusum.crit(k, ARL_0, sided="two")* from the R-package *spc*.

5.2.3 EWMA control chart

For the EWMA control chart, the charting statistic at time i is defined as $Y_i = \lambda T_i + (1-\lambda)Y_{i-1}$, where λ is a smoothing constant that depends on the size of the shift that is desired to be detected quickly. As a starting value we set $Y_0 = 0$. The EWMA control chart provides a signal when $Y_i > L\sqrt{\lambda/(2-\lambda)}$ or $Y_i < -L\sqrt{\lambda/(2-\lambda)}$ where the constant L is determined such that the control chart yields a specified in-control performance.

There are many combinations of λ and L that achieve the same in-control performance. Among these, the chart with the lowest out-of-control ARL at a specified shift of interest, is considered to be the best. To find such a chart, we used a sequence of λ values from 0.01 to 1 using steps of 0.01. For each λ value, the corresponding L value was found such that $ARL_0 = 370$ when parameters are known. This was done by using the function *xewma.crit(λ,ARL_0,sided="two")* from the R-package *spc*. For each obtained combination of λ and L, and for a given shift δ, the out-of-control ARL was calculated using the function *xewma.arl(λ,L,δ,sided="two")* from the same R-package. We consider EWMA charts which are designed to detect shifts in the mean equal to 0.5 (small), 1 (medium), and 1.5 (large) standard deviations. The corresponding best charts (i.e. lowest out-of-control ARL) for these shifts were found to be (λ, L) equal to $(0.05,2.490)$, $(0.14,2.785)$ and $(0.25,2.898)$, respectively.

5.2.4 Adjusted control limits

When parameters are estimated, control chart properties such as the average run length become random variables. This leads to a large variation in control chart performance across practitioners. In order to prevent low in-control *ARL*'s, many researchers recently suggested the use of the *exceedance probability criterion* (*EPC*), where a specified minimum in-control performance ARL_0 is obtained with a specified probability $1-p$. See for example Albers and Kallenberg (2004a, 2004b), Jones and Steiner (2012), Gandy and Kvaloy (2013), Saleh et al. (2015a,b), Goedhart et al. (2017a,2017b, 2018), Huberts et al. (2018), and Diko et al. (2019a, 2019b). More specifically, if we define $CARL_{IN}$ as the conditional in-control ARL conditional on the Phase I parameter estimates, then the EPC can formally be written as

$$P(CARL_{IN} \geq ARL_0) = 1 - p, \tag{5.7}$$

Choosing a small value of p such as 0.05 or 0.10 ensures that there is only a small probability of obtaining a $CARL_{IN}$ value that is below ARL_0.

For each of the discussed control charts, $CARL_{IN}$ depends on the Phase I parameters estimates as well as the chosen control limit constants c, h, and L. For a given Phase I sample (size), these constants can be adjusted to match the EPC. We consider adjustments of c, h, and L as provided in Goedhart et al. (2018), Chapter 3, and Chapter 4, respectively.

Table 5.1 *Adjusted and unadjusted limits charts designed for* $ARL_0 = 370$ *and various process shifts*

Chart	**δ (shift)**	***K***	***h***	***Adjusted h***
CUSUM	0.5 (small)	0.25	8.01	16.46
	1 (medium)	0.50	4.774	6.68
	1.5(large)	0.75	3.339	4.25
EWMA	**δ (shift)**	***λ***	***L***	***Adjusted L***
	0.5 (small)	0.05	2.490	3.60
	1 (medium)	0.14	2.785	3.42
	1.5(large)	0.25	2.898	3.35
Shewhart	**δ (shift)**	***λ***	***c***	***Adjusted c***
	3 (for all shifts)	1	3	3.24

In Table 5.1 we provide an overview of the required values for the Shewhart control chart, and for the EWMA and CUSUM control charts designed to detect shifts of δ equal to 0.5 (small), 1 (medium) and 1.5 (large). For all charts, we used a Phase I sample of $m = 50$ subgroups of size $n = 5$, $ARL_0 = 370$ and $p = 0.1$ in this chapter. We have considered other values as well, but they did not provide any additional insights for the comparison.

5.3 Evaluation

In this section we evaluate the performance of the Shewhart, CUSUM and EWMA control charts with estimated parameters. We evaluate a wide range of combinations for estimation errors and shift sizes, and compare the performance of the charts based on these results.

The performance of a control chart is conditional on its parameter estimates, through Q and Z. Therefore, we consider various percentiles of Q and Z as well as various possible shift sizes δ to evaluate the control chart performances. For the Shewhart control chart, all properties of the conditional run length distribution can be determined from a geometric distribution with parameter CPA according to (5.6). For the CUSUM control chart we have calculated the *CARL* values using a modified Siegmund formula approximation as described in Chapter 4, and for the EWMA we use a Markov-Chain approximation such as in Chapter 3. We have also evaluated other properties of the conditional run length distributions such as the standard deviation and several percentiles obtained by simulations. However, as these did not provide any additional insights, we do not discuss them further in this chapter.

After obtaining all the results, we have determined, for each pair of control charts, the ratio between the average run length for any considered combination of Q, Z, and δ. For Q, we consider the $25^{th}, 50^{th}$ and 75^{th} percentile of its corresponding distribution. These correspond to an underestimation, good estimate, and overestimation of the process standard deviation respectively. For Z, we consider the $5^{th}, 25^{th}, 50^{th}, 75^{th}$ and 95^{th} percentile of its corresponding distribution. Here we included the 5th and 95th percentile in order to display the results for more substantial under- and overestimations respectively. For Q, the 5th and 95th percentiles did not contribute further to the comparison, so we have omitted them for a better overview. For δ, we consider the range of 0 to 1 in steps of 0.25.

As described earlier, we consider CUSUM and EWMA charts for 0.5 (small), 1 (medium) and 1.5 (large) shifts in terms of standard deviation. However, one should note that these designs relate to the shift in the monitoring statistic, and not of the original (individual data). The standard deviation of $\bar{X}_i$ equals $\sigma_0/\sqrt{n}$. This means that for $n = 5$, the value of δ corresponding to the design shift size of the EWMA and CUSUM is closest to $0.25, 0.5$ and 0.75 for small, medium and large shifts, respectively.

5.3.1 Results

5.3.1.1 General performance

The $CARL$ values for the Shewhart, CUSUM and EWMA control charts are given in Tables 5.2, 5.3 and 5.4, respectively. For all three charts, we observe that the in-control ($\delta = 0$) $CARL$ values are in general larger when the mean is estimated more accurately (i.e., Z closer to the 50^{th} percentile). When the mean is accurately estimated, these $CARL$ values are substantially larger than ARL_0 of 370 for the values of Q considered here. This is a natural consequence of the control limit adjustments in order to guarantee a minimum in-control ARL with a specified (large) probability. Note that large in-control $CARL$ values are actually favourable, while for the out-of-control situation, smaller values mean quicker detection on average. Another general observation is that in the out-of-control situation ($\delta > 0$), because we only consider positive shifts, the $CARL$ values increase with the percentiles of Z. Obviously, if the mean is overestimated in Phase I, detection of increases in the process mean become more difficult to detect. If we would consider negative shifts, this would also hold the other way around.

When considering the results for the Shewhart chart only (Table 5.2), we observe that the estimate of the mean has substantially less impact on the $CARL$ than the estimate of standard deviation. This can be seen by first considering only the values for which Q is at the 50^{th} percentile (good estimate), and evaluating the impact of changes in Z. Comparing the in-control situation for the different percentiles, we observe some, but not too much change in the $CARL$ values. However, if we consider a good estimate of the mean (50^{th} percentile of Z), we observe that the impact of estimation of standard deviation contributes to larger differences in the $CARL$ values. This result was also obtained in Zwetsloot and Woodall (2017). Note that an overestimation of the mean will increase both the UCL and

the LCL, which will increase the likeliness of a signal below the LCL, but decrease that of a signal above the UCL. On the other hand, overestimation of the standard deviation increases the UCL, and decreases the LCL, decreasing the alarm rate on both sides of the control chart.

Table 5.2 *Shewhart $CARL$ values*

Z	Q / δ	25th	50th	75th
5th	0	435	623	906
	0.25	102	137	187
	0.50	26	34	43
	0.75	9	11	13
	1	4	5	5
25th	0	538	780	1149
	0.25	147	200	277
	0.50	36	46	60
	0.75	11	14	17
	1	5	5	6
50th	0	564	821	1213
	0.25	191	263	368
	0.50	45	58	77
	0.75	14	17	21
	1	5	6	7
75th	0	538	780	1149
	0.25	247	344	487
	0.50	56	74	99
	0.75	16	20	26
	1	6	7	9
95th	0	435	623	906
	0.25	353	499	718
	0.50	80	106	143
	0.75	22	27	35
	1	8	9	11

The results of the CUSUM and EWMA are quite similar. Compared to the Shewhart chart, these charts are much more affected by estimation error in the mean. This is especially so for the charts designed to detect small shifts. Without adjustments, small estimation errors can lead to low in-control $CARL$ values (see for example Saleh et al 2015b). In order to correct for this, substantial control limit adjustments are required. When

designed to detect small shifts, this required correction is additionally large, as small estimation errors would otherwise quickly be detected as process changes. For CUSUM and EWMA charts designed for larger shifts, the impact of estimation error of the mean is more similar to that of the Shewhart charts. Note also that the Shewhart chart is actually a special case of the EWMA when $\lambda = 1$. The impact of estimation error of the standard deviation on the CUSUM and EWMA control charts is quite similar to that of the Shewhart control chart. Overestimation of the standard deviation leads to larger $CARL$ values, both in-control and out-of-control.

Table 5.3 *CUSUM CARL values*

		Small *k=0.25, h=16.46*			Medium *k=0.5, h=16.46*			Large *k=0.75, h=16.46*		
Z	**Q / δ**	**25th**	**50th**	**75th**	**25th**	**50th**	**75th**	**25th**	**50th**	**75th**
5th	**0**	321	382	460	319	420	564	343	476	673
	0.25	29	31	32	19	21	23	22	25	29
	0.50	15	16	16	8	9	9	7	8	8
	0.75	10	10	11	5	5	6	4	4	5
	1	8	8	8	4	4	4	3	3	3
25th	**0**	3226	4622	6785	1058	1529	2254	765	1128	1697
	0.25	38	40	43	29	32	35	36	42	51
	0.50	17	18	18	9	10	11	9	9	10
	0.75	11	11	12	6	6	6	5	5	5
	1	8	8	9	4	4	4	3	3	3
50th	**0**	15977	26180	44042	1651	2494	3849	971	1465	2257
	0.25	49	52	55	41	46	52	54	66	81
	0.50	19	20	20	11	11	12	10	11	12
	0.75	12	12	13	6	6	7	5	5	6
	1	8	9	9	4	4	5	3	3	4
75th	**0**	3226	4622	6785	1058	1529	2254	765	1128	1697
	0.25	67	71	76	64	75	88	87	109	140
	0.50	21	22	23	12	13	14	12	13	15
	0.75	12	13	13	7	7	7	5	6	6
	1	9	9	10	4	5	5	3	4	4
95th	**0**	321	382	460	319	420	564	343	476	673
	0.25	135	150	168	151	189	239	190	253	344
	0.50	25	26	28	16	17	18	17	19	21
	0.75	14	14	15	7	8	8	6	7	7
	1	10	10	10	5	5	5	4	4	4

Table 5.4 *EWMA $CARL$ values*

		Small λ=0.05,L=3.60			Medium λ=0.14,L=3.42			Large λ=0.25,L=3.35		
Z	**Q / δ**	**25th**	**50th**	**75th**	**25th**	**50th**	**75th**	**25th**	**50th**	**75th**
5th	**0**	330	407	510	334	437	583	341	460	632
	0.25	29	30	32	23	25	28	25	29	33
	0.50	15	15	16	10	10	11	9	9	10
	0.75	10	10	11	6	7	7	5	5	6
	1	8	8	8	5	5	5	4	4	4
25th	**0**	1979	2749	3894	1084	1530	2198	797	1137	1651
	0.25	38	41	43	34	38	43	39	46	55
	0.50	17	17	18	11	12	13	10	11	12
	0.75	11	11	12	7	7	8	6	6	6
	1	8	8	9	5	5	5	4	4	4
50th	**0**	5833	8800	13549	1715	2517	3763	1043	1523	2263
	0.25	50	53	58	48	55	64	57	70	85
	0.50	18	19	20	13	14	15	12	13	14
	0.75	12	12	12	7	8	8	6	7	7
	1	8	9	9	5	5	6	4	4	5
75th	**0**	1908	2671	3810	1062	1506	2172	788	1127	1639
	0.25	70	77	85	74	87	105	89	111	141
	0.50	21	21	22	15	16	17	15	16	18
	0.75	12	13	13	8	8	9	7	7	8
	1	9	9	10	6	6	6	4	5	5
95th	**0**	287	362	462	314	416	560	329	448	619
	0.25	148	171	201	166	207	264	189	247	328
	0.50	25	26	27	19	20	22	20	22	25
	0.75	14	14	15	9	9	10	8	8	9
	1	10	10	10	6	6	6	5	5	5

5.3.1.2 Pairwise comparison

For an easier comparison, we have determined the $CARL$ ratios for the considered values of Q, Z and δ for each of the charts. In Tables 5.5, 5.6 and 5.7 we have listed the results for the pairwise results between the Shewhart, CUSUM and EWMA when the latter two are designed for small, medium, or large shift detection respectively. Note that a value larger than 1 means that the chart considered in the numerator has a larger $CARL$ value

than the one in the denominator for that particular combination of Q, Z and δ. A ratio of 1 means that the $CARL$ values are equal.

Note also that for $\delta = 0$, a ratio above 1 is advantageous for the chart numerator, while for $\delta > 0$ it is better for the chart in the numerator to have a ratio below 1. Since the control limit adjustments considered in this chapter provide a minimum in-control performance, we consider the out-of-control performance to be most important in the comparison.

From Table 5.5 we observe that the out-of-control performance of the EWMA and CUSUM charts designed to detect small shifts is very similar, with many ratios close or equal to 1. The CUSUM performs slightly better (ratios of 0.84 to 0.91) for detecting shifts of $\delta = 0.25$ when the mean is substantially overestimated (95^{th} percentile of Z). The in-control values differ the most when the mean is estimated more accurately (closer to 50^{th}percentile of Z).

The Shewhart control chart is only able to outperform the CUSUM and EWMA schemes for the detection of large shifts ($\delta = 1$). For other sizes of shifts, the out-of-control $CARL$ values are larger than those of the CUSUM and EWMA scheme for almost any combination of Q and Z.

From Tables 5.6 and 5.7, it can be seen that the CUSUM chart outperforms the EWMA and Shewhart in terms of out-of-control performance on almost any combination and shift size. For the in-control situation the Shewhart chart has larger in-control $CARL$ values for more severe over- and underestimation of the mean (5^{th} and 95^{th} percentile of Z). However, as stated earlier, the control limit adjustments considered here already provide a good in-control performance, such that the faster out-of-control detection is probably more valuable. From Tables 5.6 and 5.7, it can also be seen that the Shewhart control chart is far behind on out-of-control detection speed compared to both CUSUM and EWMA schemes.

Table 5.5 *Pairwise $CARL$ ratios (CUSUM and EWMA designed for small shift)*

		CUSUM/EWMA			CUSUM/SHEWHART			EWMA/SHEWHART		
Z	Q / δ	25th	50th	75th	25th	50th	75th	25th	50th	75th
5th	0	0.97	0.94	0.90	0.74	0.61	0.51	0.76	0.65	0.56
	0.25	1.00	1.03	1.00	0.28	0.23	0.17	0.28	0.22	0.17
	0.50	1.00	1.07	1.00	0.58	0.47	0.37	0.58	0.44	0.37
	0.75	1.00	1.00	1.00	1.11	0.91	0.85	1.11	0.91	0.85
	1	1.00	1.00	1.00	2.00	1.60	1.60	2.00	1.60	1.60
25th	0	1.63	1.68	1.74	6.00	5.93	5.91	3.68	3.52	3.39
	0.25	1.00	0.98	1.00	0.26	0.20	0.16	0.26	0.21	0.16
	0.50	1.00	1.06	1.00	0.47	0.39	0.30	0.47	0.37	0.30
	0.75	1.00	1.00	1.00	1.00	0.79	0.71	1.00	0.79	0.71
	1	1.00	1.00	1.00	1.60	1.60	1.50	1.60	1.60	1.50
50th	0	2.74	2.98	3.25	28.33	31.89	36.31	10.34	10.72	11.17
	0.25	0.98	0.98	0.95	0.26	0.20	0.15	0.26	0.20	0.16
	0.50	1.06	1.05	1.00	0.42	0.34	0.26	0.40	0.33	0.26
	0.75	1.00	1.00	1.08	0.86	0.71	0.62	0.86	0.71	0.57
	1	1.00	1.00	1.00	1.60	1.50	1.29	1.60	1.50	1.29
75th	0	1.69	1.73	1.78	6.00	5.93	5.91	3.55	3.42	3.32
	0.25	0.96	0.92	0.89	0.27	0.21	0.16	0.28	0.22	0.17
	0.50	1.00	1.05	1.05	0.38	0.30	0.23	0.38	0.28	0.22
	0.75	1.00	1.00	1.00	0.75	0.65	0.50	0.75	0.65	0.50
	1	1.00	1.00	1.00	1.50	1.29	1.11	1.50	1.29	1.11
95th	0	1.12	1.06	1.00	0.74	0.61	0.51	0.66	0.58	0.51
	0.25	0.91	0.88	0.84	0.38	0.30	0.23	0.42	0.34	0.28
	0.50	1.00	1.00	1.04	0.31	0.25	0.20	0.31	0.25	0.19
	0.75	1.00	1.00	1.00	0.64	0.52	0.43	0.64	0.52	0.43
	1	1.00	1.00	1.00	1.25	1.11	0.91	1.25	1.11	0.91

Table 5.6 *Pairwise $CARL$ ratios (CUSUM and EWMA designed for medium shift)*

		CUSUM/EWMA			CUSUM/SHEWHART			EWMA/SHEWHART		
Z	Q / δ	25th	50th	75th	25th	50th	75th	25th	50th	75th
5th	0	0.96	0.96	0.97	0.73	0.67	0.62	0.77	0.70	0.64
	0.25	0.83	0.84	0.82	0.19	0.15	0.12	0.23	0.18	0.15
	0.50	0.80	0.90	0.82	0.31	0.26	0.21	0.38	0.29	0.26
	0.75	0.83	0.71	0.86	0.56	0.45	0.46	0.67	0.64	0.54
	1	0.80	0.80	0.80	1.00	0.80	0.80	1.25	1.00	1.00
25th	0	0.98	1.00	1.03	1.97	1.96	1.96	2.01	1.96	1.91
	0.25	0.85	0.84	0.81	0.20	0.16	0.13	0.23	0.19	0.16
	0.50	0.82	0.83	0.85	0.25	0.22	0.18	0.31	0.26	0.22
	0.75	0.86	0.86	0.75	0.55	0.43	0.35	0.64	0.50	0.47
	1	0.80	0.80	0.80	0.80	0.80	0.67	1.00	1.00	0.83
50th	0	0.96	0.99	1.02	2.93	3.04	3.17	3.04	3.07	3.10
	0.25	0.85	0.84	0.81	0.21	0.17	0.14	0.25	0.21	0.17
	0.50	0.85	0.79	0.80	0.24	0.19	0.16	0.29	0.24	0.19
	0.75	0.86	0.75	0.88	0.43	0.35	0.33	0.50	0.47	0.38
	1	0.80	0.80	0.83	0.80	0.67	0.71	1.00	0.83	0.86
75th	0	1.00	1.02	1.04	1.97	1.96	1.96	1.97	1.93	1.89
	0.25	0.86	0.86	0.84	0.26	0.22	0.18	0.30	0.25	0.22
	0.50	0.80	0.81	0.82	0.21	0.18	0.14	0.27	0.22	0.17
	0.75	0.88	0.88	0.78	0.44	0.35	0.27	0.50	0.40	0.35
	1	0.67	0.83	0.83	0.67	0.71	0.56	1.00	0.86	0.67
95th	0	1.02	1.01	1.01	0.73	0.67	0.62	0.72	0.67	0.62
	0.25	0.91	0.91	0.91	0.43	0.38	0.33	0.47	0.41	0.37
	0.50	0.84	0.85	0.82	0.20	0.16	0.13	0.24	0.19	0.15
	0.75	0.78	0.89	0.80	0.32	0.30	0.23	0.41	0.33	0.29
	1	0.83	0.83	0.83	0.63	0.56	0.45	0.75	0.67	0.55

Tables 5.7 *Pairwise $CARL$ ratios (CUSUM and EWMA designed for large shift)*

		CUSUM/EWMA			CUSUM/SHEWHART			EWMA/SHEWHART		
Z	Q / δ	25th	50th	75th	25th	50th	75th	25th	50th	75th
5th	0	1.01	1.03	1.06	0.79	0.76	0.74	0.78	0.74	0.70
	0.25	0.88	0.86	0.88	0.22	0.18	0.16	0.25	0.21	0.18
	0.50	0.78	0.89	0.80	0.27	0.24	0.19	0.35	0.26	0.23
	0.75	0.80	0.80	0.83	0.44	0.36	0.38	0.56	0.45	0.46
	1	0.75	0.75	0.75	0.75	0.60	0.60	1.00	0.80	0.80
25th	0	0.96	0.99	1.03	1.42	1.45	1.48	1.48	1.46	1.44
	0.25	0.92	0.91	0.93	0.24	0.21	0.18	0.27	0.23	0.20
	0.50	0.90	0.82	0.83	0.25	0.20	0.17	0.28	0.24	0.20
	0.75	0.83	0.83	0.83	0.45	0.36	0.29	0.55	0.43	0.35
	1	0.75	0.75	0.75	0.60	0.60	0.50	0.80	0.80	0.67
50th	0	0.93	0.96	1.00	1.72	1.78	1.86	1.85	1.86	1.87
	0.25	0.95	0.94	0.95	0.28	0.25	0.22	0.30	0.27	0.23
	0.50	0.83	0.85	0.86	0.22	0.19	0.16	0.27	0.22	0.18
	0.75	0.83	0.71	0.86	0.36	0.29	0.29	0.43	0.41	0.33
	1	0.75	0.75	0.80	0.60	0.50	0.57	0.80	0.67	0.71
75th	0	0.97	1.00	1.04	1.42	1.45	1.48	1.46	1.44	1.43
	0.25	0.98	0.98	0.99	0.35	0.32	0.29	0.36	0.32	0.29
	0.50	0.80	0.81	0.83	0.21	0.18	0.15	0.27	0.22	0.18
	0.75	0.71	0.86	0.75	0.31	0.30	0.23	0.44	0.35	0.31
	1	0.75	0.80	0.80	0.50	0.57	0.44	0.67	0.71	0.56
95th	0	1.04	1.06	1.09	0.79	0.76	0.74	0.76	0.72	0.68
	0.25	1.01	1.02	1.05	0.54	0.51	0.48	0.54	0.49	0.46
	0.50	0.85	0.86	0.84	0.21	0.18	0.15	0.25	0.21	0.17
	0.75	0.75	0.88	0.78	0.27	0.26	0.20	0.36	0.30	0.26
	1	0.80	0.80	0.80	0.50	0.44	0.36	0.63	0.56	0.45

5.3.2 Discussion

We find that CUSUM charts yield faster detection of sustained shifts than Shewhart and EWMA charts when their designs are adjusted conform the exceedance probability criterion. This result holds regardless of whether the CUSUM and EWMA are designed for small, medium or large shifts. Zwetsloot and Woodall (2017) found that, without adjustment, the EWMA yields lower in-control $CARL$ values than the CUSUM. Because of these lower $CARL$ values, a larger adjustment is required for the EWMA chart to provide the desired in-control performance. This leads to larger out-of-control $CARL$ values for the

adjusted charts. Since the in-control performance is already sufficient due to the adjustments according to the exceedance probability criterion, this out-of-control performance is most important for the comparison.

The differences between the CUSUM and EWMA control charts are not very large. On the other hand, the Shewhart control chart is not able to compete with these two schemes in general, except for especially large shifts when the CUSUM and EWMA are designed for small shifts. This has to do with the fact that we are considering sustained shifts in this chapter. The EWMA and CUSUM are especially suitable to detect these kind of shifts, while the Shewhart chart is memoryless (i.e., does not use information from previous Phase II observations). Of course, countless other types of issues are thinkable in practice, but in this chapter we focus on the detection of sustained shifts.

5.4 Concluding remarks

In this chapter we have compared the performance of Shewhart, CUSUM and EWMA control charts when they are adjusted for parameter estimation. The adjustments are made to provide a specified minimum in-control performance with a specified probability. This is in line with adjustments advocated recently by many authors, such as Jones and Steiner (2012), Gandy and Kvaløy (2013), and Saleh et al. (2015a, 2015b), in order to reduce the risk of large false alarm rates when parameters are estimated.

We find that CUSUM control charts provide the fastest out-of-control detection of sustained shifts for almost all shift sizes and estimation errors. Since the control limit adjustments are designed to control the in-control performance, the out-of-control detection speed is considered the most important aspect in the comparison. The EWMA control chart is only slightly behind the CUSUM, while the Shewhart control chart is only able to compete with the other two schemes for very large shifts.

In this chapter we have focused on the detection of sustained shifts in normally distributed process data for control charts with estimated parameters. How our findings generalize towards other types of shifts or specific designs such as nonparametric Shewhart, EWMA and CUSUM schemes are interesting topics for future research.

References

Abbasi, S.A., Riaz, M., & Miller, A. (2012). Enhancing the performance of CUSUM scale chart. *Computers & Industrial Engineering* 63(2): 400-409.

Albers, W. & Kallenberg, W. C. M. (2004a). Are estimated control charts in control?. *Statistics* 38(1):67-79.

Albers, W. & Kallenberg, W.C.M. (2004b). Estimation in Shewhart control charts: Effects and corrections. *Metrika* 59(3):207-234.

Aly, A.A., Saleh, N.A., Mahmoud, M.A. & Woodall, W.H. (2015). A reevaluation of the adaptive exponentially weighted moving average control chart when parameters are estimated. *Quality and Reliability Engineering International* 31(8):1611-1622.

Brook, D. & Evans, D.A. (1972). An approach to the probability distribution of CUSUM run length. *Biometrika*, 59(3):539-549.

Chakraborti, S. (2000). Run length, average run length and false alarm rate of Shewhart chart: exact derivations by conditioning. *Communications in Statistics: Simulation and Computation* 29(1):61-81.

Chakraborti, S. (2006). Parameter estimation and design considerations in prospective applications of the chart. *Journal of Applied Statistics* 33(4):439-459.

Chakraborti, S., Graham, M.A. & Human, S.W. (2009). Phase I statistical process control charts: an overview and some results. *Quality Engineering* 21(1):52-62.

Chen, G. (1997). The mean and standard deviation of the run length distributions of $\bar{X}$ charts when control limits are estimated. *Statistica Sinica* 7(3):789-798.

Chen, G. (1998). The run length distribution of the R, S and S^2 charts when σ is estimated. *The Canadian Journal of Statistics* 26(2):311-322.

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

Crowder, S. V. (1987). A simple method for studying the run length distributions of exponentially weighted moving average charts. *Technometrics*, 29(4):401–407.

Crowder, S.V. (1989). Design of Exponentially Weighted Moving Average Schemes. *Journal of Quality Technology* 21(3):155-162.

Diko, M.D. (2014). Some contributions to joint monitoring of mean and variance of normal populations. *MSc dissertation*, Department of Statistics, University of Pretoria: South Africa.

Diko, M.D., Chakraborti, S. & Graham, M.A. (2016). Monitoring the process mean when standards are unknown: A classic problem revisited. *Quality and Reliability Engineering International* 32(2):609-622.

Diko, M.D., Goedhart, R., Chakraborti, S., Epprecht, E.K. & Does, R.J.M.M. (2017). Phase II control charts for monitoring dispersion when parameters are estimated. *Quality Engineering* 29(4):605-622.

Diko, M.D., Chakraborti, S. & Does, R.J.M.M. (2019a). Guaranteed in-control performance of the EWMA chart for monitoring the mean. *Quality and Reliability Engineering International* 35(4):1144:1160.

Diko, M.D., Chakraborti, S. & Does, R.J.M.M. (2019b). An alternative design of the two-sided CUSUM chart for monitoring the mean when parameters are estimated. *Computers & Industrial Engineering* 137(106042).

Diko, M.D., Goedhart, R. & Does, R.J.M.M. (2019c). A head-to-head comparison of the out-of-control performance of control charts adjusted for parameter estimation. Accepted for publication *Quality Engineering*.

Epprecht, E.K., Loureiro, L.D. & Chakraborti, S. (2015). Effect of the amount of phase I data on the phase II performance of S^2 and S control charts. *Journal of Quality Technology* 47(2):139-155.

Ewan W.D. & Kemp, K.W. (1960). Sampling inspection of continuous processes with no autocorrelation between successive results. *Biometrika* 47(3-4):363-380.

Faraz, A., Woodall, W.H. & Heuchenne, C. (2015). Guaranteed conditional performance of the S^2 control chart with estimated parameters. *International Journal of Production Research* 53(14):4405-4413.

Faraz, A., Heuchenne, C. & Saniga E. (2016). The np chart with guaranteed In-control average run lengths. *Quality and Reliability Engineering International* 33(5):1057-1066.

Gandy, A. & Kvaloy , J.T. (2013). Guaranteed conditional performance of control charts via bootstrap methods. *Scandanavian Journal of Statistics* 40(4):647-668.

Gibbons, J. D. & Chakraborti, S. (2010). *Nonparametric Statistical Inference*. Chapman & Hall, Boca Raton, FL, 5th edition.

Goedhart, R., Schoonhoven, M. & Does, R.J.M.M. (2016). Correction factors for Shewhart X and X-bar control charts to achieve desired unconditional ARL. *International Journal of Production Research* 54(24):7464-7479.

Goedhart, R., Schoonhoven, M. & Does, R.J.M.M. (2017a). Guaranteed in-control performance for the Shewhart X and Xbar control charts. *Journal of Quality Technology* 49(2):155-171.

Goedhart, R., da Silva, M.M., Schoonhoven, M., Epprecht, E.K., Chakraborti, S., Does, R.J.M.M. & Veiga, A. (2017b). Shewhart control charts for dispersion adjusted for parameter estimation. *IISE Transactions* 49(8):838-848.

Goedhart, R., Schoonhoven, M. & Does, R.J.M.M. (2018). Letter to the editor: On guaranteed in-control performance for the Shewhart X and Xbar control charts. *Journal of Quality Technology* 50(1):130-132.

Grant, E. L. & Leavenworth, R.S. (1986). *Statistical Quality Control*. McGraw-Hill, New York, 5th edition.

Hawkins, D. M. & Wu, Q. (2014). The CUSUM and the EWMA head-to-head. *Quality Engineering* 26(2):215-222.

Hu, X. & Castagliola, P. (2017). Guaranteed conditional design of the median chart with estimated parameters. *Quality and Reliability Engineering International* 33(8):1873-1884.

Huberts, L.C.E., Schoonhoven, M., Goedhart, R., Diko, M.D. & Does, R.J.M.M. (2018). The performance of $\bar{X}$ control charts for large non-normally distributed datasets. *Quality and Reliability Engineering International* 34(6):979-996.

Jardim, F.S., Chakraborti, S. & Epprecht, E.K. (2019a). Two perspectives for designing a phase II control chart with estimated parameters: The case of the Shewhart $\bar{X}$ chart. *Journal of Quality Technology*, DOI: 10.1080/00224065.2019.1571345.

Jardim, F.S., Chakraborti, S. & Epprecht, E.K. (2019b). $\bar{X}$ chart with estimated parameters: The conditional ARL distribution and new insights. *Production and Operations Management* 28(6), 1545:1557.

Jensen, W.A., Jones-Farmer, L.A., Champ, C.W. & Woodall, W.H. (2006). Effects of parameter estimation on control chart properties: a literature review. *Journal of Quality Technology* 38(4):349-364.

Jeske, D.R. (2016). Determining the reference sample size needed to control the accuracy of the conditional in-control ARL of a normal theory CUSUM. *Quality and Reliability Engineering International* 32(7):2499-2504.

Jones, L.A., Champ, C.W. & Rigdon, S.E. (2001). The Performance of Exponentially Weighted Moving Average Charts with Estimated Parameters. *Technometrics* 43(2),156–167.

Jones, L.A. (2002). The statistical Design of EWMA Control Charts with Estimated Parameters. *Journal of Quality Technology* 34(3):277-288.

Jones, L.A., Champ, C.W. & Rigdon, S.E. (2004). The run length distribution of the CUSUM with estimated parameters. *Journal of Quality Technology* 36(1):95-108.

Jones, M.A. & Steiner S.H. (2012). Assessing the effect of estimation error on the risk-adjusted CUSUM chart performance. *International Journal for Quality in Health Care* 24(2):176-181.

Jones-Farmer, L.A., Woodall, W.H., Steiner, S.H. & Champ, C.W. (2014). An overview of phase I analysis for process improvement and monitoring. *Journal of Quality Technology* 46(3):265-280.

Loureiro, L.D., Epprecht, E.K., Chakraborti, S. & Jardim, F.S. (2018). In-control performance of the joint Phase II Xbar-S control charts when parameters are estimated. *Quality Engineering* 30(2):253-267.

Lucas, J.M. & Crosier, R.B. (1982). Fast initial response for CUSUM quality control schemes. *Technometrics* 24(3):199-205.

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

Mahmoud, M.A., Henderson, G.R., Epprecht, E.K. & Woodall, W.H. (2010). Estimating the standard deviation in quality-control applications. *Journal of Quality Technology* 42(4):348-357.

Mehmood, R., Qazi, M.S. & Riaz, M. (2018). On the performance of $\bar{X}$ control chart for known and unknown parameters supplemented with runs rules under different probability distributions. *Journal of Statistical Computation and Simulation* 88(4):675-711.

Montgomery, D.C. (2013). *Introduction to Statistical Quality Control: A modern Introduction*. John Wiley & Sons, Hoboken, 7th edition.

Page, E.S. (1954). Continuous inspection schemes. *Biometrika* 41(1-2):100-114.

Patnaik, P.B. (1950). The use of mean range as an estimator of variance in statistical tests. *Biometrika* 37(1-2):78-87.

Psarakis, S., Vyniou, A.K. & Castagliola, P. (2014). Some recent developments on the effects of parameter estimation on control charts. *Quality and Reliability Engineering International* 30(8):1113-1129.

Qiu, P. (2014). *Introduction to Statistical Process Control*. CRC Press, Boca Raton, FL.

Quesenberry, C.P. (1993). The effect of sample size on the estimated limits of $\bar{X}$ and X control charts. *Journal of Quality Technology* 25(4):237-247.

Roberts, S.W. (1959). Control chart tests based on geometric moving averages. *Technometrics* 1(3):239-250.

Saleh, N.A., Mahmoud, M.A., Jones-Farmer, L.A., Zwetsloot, I.M. & Woodall, W.H. (2015a). Another look at the EWMA control chart with estimated parameters. *Journal of Quality Technology* 47(4):363-382.

Saleh, N.A., Mahmoud, M.A., Keefe, M.J. & Woodall, W.H. (2015b). The difficulty in designing Shewhart X-bar and X control charts with estimated parameters. *Journal of Quality Technology* 47(2):127-138.

Saleh, N.A., Zwetsloot, I.M., Mahmoud, M.A. & Woodall, W.H. (2016). CUSUM charts with controlled conditional performance under estimated parameters. *Quality Engineering* 28(4):402-415.

Sanusi, R.A., Abujiya, M.R., Riaz, M. & Abbas, N. (2017). Combined Shewhart CUSUM charts using auxiliary variable. *Computers & Industrial Engineering* 105(2017):329-337.

Schoonhoven, M., Nazir, H.Z., Riaz, M. & Does, R.J.M.M. (2011a). Robust location estimators for the $\bar{X}$ control chart. *Journal of Quality Technology* 43(4):363-379.

Schoonhoven, M., Riaz, M. & Does, R.J.M.M. (2011b). Design and analysis of control charts for standard deviation with estimated parameters. *Journal of Quality Technology* 43(4):307-333.

Schoonhoven, M., Riaz, M. & Does, R.J.M.M. (2009). Design schemes for the $\bar{X}$ control chart. *Quality and Reliability Engineering International* 25(5):581-594.

Shewhart, W.A. (1926). Quality control charts. *Bell System Technical Journal* 5(4):593-603.

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Berlin:Springer.

Sparks, R.S. (2000). CUSUM charts for signaling varying location shifts. *Journal of Quality Technology* 32(2):157-171.

Van Dobben de Bruyn, C.S. (1968). *Cumulative Sum Tests: Theory and Practice*. London: Griffin.

Woodall, W.H. (1984). On the Markov Chain approach to the two-sided CUSUM procedure. *Technometrics* 26(1):41-46.

Woodall, W.H. (2017). Bridging the gap between theory and practice in basic statistical process monitoring. *Quality Engineering* 29(1):2-15.

Zwetsloot, I. M. & Woodall, W. H. (2017). A head-to-head comparative study of the conditional performance of control charts based on estimated parameters. *Quality Engineering* 29(2):244-253.

Summary

Process variation is divided into common cause and special cause variation. A process operating under common cause variation is said to be in-control (*IC*), while a process operating under both common cause and special cause variation is said to be out-of-control (*OOC*). Control charts are graphical tools that help to distinguish common causes from special causes. There are different types of control charts. Among these, the most used are Cumulative Sum (CUSUM), Exponentially Weighted Moving Average (EWMA) and Shewhart control charts. CUSUM and EWMA control charts are designed to detect small process parameter shifts, while Shewhart charts are excellent to detect large process parameter shifts.

The run length distribution and its properties are used to evaluate control charts. The run length is a random variable that counts the number of samples that must be taken before a control chart gives an *OOC* signal. If the process is *IC*, the probability that the chart gives an *OOC* signal is the false alarm rate (*FAR*). For the Shewhart control chart, when process parameters are known (Case K), the signaling events are independent and the *IC* run length has a geometric distribution with parameter *FAR*. In this case, the *FAR* is the reciprocal of the *IC* average run length (ARL_{IN}). On the other hand, for CUSUM and EWMA charts, since the charting statistics are based on past and present data, the signaling events are not independent. As a result, the run length distribution is not geometric, so the *FAR* is not the reciprocal of the ARL_{IN}. Case K control charts are designed by finding a charting constant that corresponds to a specified nominal ARL_{IN} value, say ARL_0 , or a nominal *FAR* value, say FAR_0. For the Shewhart chart, the charting constant is usually the FAR_0^{th} percentile of the *IC* sampling distribution of the charting statistic. However, for EWMA and CUSUM charts, since the signaling events are not independent, the charting constant is found by numerical methods.

Control chart limits are functions of the charting statistic and the assumed process parameters. When parameters are unknown, control charts are used retrospectively to bring an unruly process under control. This is called Phase I analysis. The resulting *IC*

retrospective data is used to estimate the unknown process parameters and the chart control limits for process monitoring in Phase II. Because of sampling variation between and within practitioners, practitioners get different estimates, different control limits and thus different chart performance levels. In SPM parlance, this sampling variation is known as practitioner to practitioner variation.

Consequently, the *ARL* is a random variable with a distribution. As a random variable, the *ARL* is called the conditional *ARL* and is denoted by *CARL*. Ideally, the in-control *CARL* (denoted as $CARL_{IN}$) distribution should be centered around some nominally specified *ARL* (ARL_0). Hence, control charts have been evaluated by calculating the expected value of the $CARL_{IN}$ distribution and comparing this to ARL_0. Moreover, they have been designed such that the $E(CARL_{IN})=ARL_0$. This is called the unconditional perspective/approach. In reality, however, a practitioner will only observe a single sample or a single chart. The $CARL_{IN}$ value that corresponds to this sample or chart is almost always different from the ARL_0. Thus, the unconditional perspective is not very meaningful. To get some sense of the closeness of the $CARL_{IN}$ values to the ARL_0, the standard deviation of the distribution of the $CARL_{IN}$ (denoted by $SDCARL_{IN}$) is often calculated. In reality, however, the $SDCARL_{IN}$ does not measure the closeness of the $CARL_{IN}$ values to the ARL_0, it measures the closeness of the $CARL_{IN}$ values to the $E(CARL_{IN})$. Therefore, unless $E(CARL_{IN})=ARL_0$, the $SDCARL_{IN}$ is also not very meaningful. Note that, in practice, the proximity of the $CARL_{IN}$ values to the ARL_0 is less important than their direction relative to the ARL_0. For example, a practitioner is always happy to have a $CARL_{IN}$ value that is larger than the ARL_0, irrespective of its proximity to the ARL_0. For this reason, control charts are now evaluated and designed through exceedance probabilities. An exceedance probability, denoted as *1-p*, is the probability that the $CARL_{IN}$ exceeds the ARL_0. Ideally, the exceedance probability should be high, say $1 - p \geq 0.90$. In other words, the exceedance probability criterion (*EPC)* states that the ARL_0 should preferably be near the *pth* percentile of the $CARL_{IN}$ distribution, where *p* is ideally less than or equal to 10%. In contrast, the unconditional approach requires that the ARL_0 be near the $E(CARL_{IN})$.

A deterioration in the *OOC* performance is the price that must be paid for adjusting the control limits for the effects of parameter estimation. However, it is possible to balance the improvement of the *IC* performance with the deterioration of the *OOC* performance. For unconditionally designed charts, this can be done by manipulating ARL_0. For conditionally

adjusted limits, it can be done by manipulating p and $\varepsilon = \frac{CARL_{IN,p} - ARL_0}{ARL_0}$. In this regard, the conditional approach is more flexible.

For the last 2 decades, *SPM* researchers have been trying to find out: which estimators should be used to estimate the unknown process parameters? How bad or how good will a chart with estimated parameters perform in comparison to a chart with known parameters? Which chart is the best for detecting a specified sustained parameter shift? How large should the Phase I sample size be in order for a chart with estimated parameters to perform like a chart with known parameters? or How large should the Phase I sample size be for a chart to meet a specified performance criteria? How the control limits should be adjusted to compensate for the effects of parameter estimation as a function of the available Phase I sample size? For adjusted control, how to balance the improvement of the *IC* performance with the deterioration of the *OOC* performance? In this thesis, although we make meaningful contributions to all the above questions, our main focus is on the latter three questions, which deal with the design of control charts. We consider univariate Shewhart, EWMA and CUSUM charts. We assume that the process data is independent and normally distributed. Contributions and limitations of this thesis are explained in more detail below.

Using the unconditional perspective, in Diko (2014) and Diko et al. (2016) we showed the deleterious effects of (i) parameter estimation (ii) the use of 3-sigma limits (iii) multiple testing on the *IC* properties of the $(\bar{X}, R)$ charting scheme. We presented formulas for finding the correct charting constants, taking proper account of the above issues. For some commonly used ARL_0, we provided the tables of the corrected charting constants. The corrected constants ensure that the $CARL_{IN}$ values will be centred around ARL_0. However, these corrected constants give no absolute guarantee that a practitioner will not get a low $CARL_{IN}$ value. The *OOC* performance evaluations have not been done. The evaluation of the conditional performance of the corrected charting constants, through for example the $SDCARL_{IN}$ and the *EPC*, still needs to be done. In fact, we believe that using the $SDCARL_{IN}$ will be meaningful, in this case, because $E(CARL_{IN})=ARL_0$. Our ideas can be extended to other two-chart monitoring schemes, including the CUSUM and EWMA for the mean and the variance of a normal process.

Under the unconditional perspective, for the Phase II Shewhart dispersion (e.g. range and standard deviation) control charts we derived and tabulated new corrected charting constants, in Chapter 2, that should be used to construct the estimated probability limits. To apply the unconditional perspective, we used a numerical and an analytic approach. The two approaches gave similar results. Unlike the unadjusted (Case K) constants, the corrected constants give the $CARL_{IN}$ performance that is centred around ARL_0. For this reason, when parameters are estimated, the adjusted constants are better than the unadjusted constants (Case K constants). However, these adjusted constants do not guarantee the non-occurrence of low $CARL_{IN}$ values. The seriousness of this limitation still needs to be thoroughly studied by, for example, $SDCARL_{IN}$ and the *EPC*. The important case, when $n = 1$, has also not been done.

Using the *EPC*, under the conditional perspective, we studied the performance and the design of the two-sided Phase II EWMA and CUSUM charts for the location, in Chapter 3 and Chapter 4, respectively. We showed that in-order to design charts that guarantee a specified *EPC*, more Phase I data are needed than previously recommended in the literature. We presented a new method to adjust the control chart limits according to the *EPC*. We see this method as a better alternative to parametric bootstrapping. Like the bootstrap method, our method involves, in a different way, simulations that are accompanied by a search algorithm. However, unlike the bootstrap method, the search algorithm is bounded. This makes the method faster than parametric bootstrapping. Moreover, because of a different way that we do the simulations, our method produces more accurate and reliable results compared to the bootstrap method. Based on our new method, tables and graphs of the adjusted constants were provided over a wide range of chart parameters. This will help with the implementation of the *EPC* adjusted charts, in practice, since these constants are currently unavailable in SPM textbooks or in computer software's. An in-control and out-of-control performance evaluation of the adjusted limits was presented. It was seen that the adjusted constants guarantee (with high probability) a specified minimum $CARL_{IN}$ performance at a marginal cost of a lower *OOC* performance. We showed how the *IC* and *OOC* performance trade-offs can be done. Moreover, since the numerical methods for approximating the *CARL* distribution of the two-sided CUSUM chart are demanding and sometimes impossible, in Chapter 4, we presented alternative formulas to approximate the

CARL distribution. We showed their simplicity, accuracy, versatility and highly recommend them for day to day use. The unconditional analysis of the conditionally adjusted limits of CUSUM and EWMA charts still needs to be done.

In Chapter 5, we compared the conditionally adjusted Shewhart, CUSUM and EWMA control charts with respect to detecting sustained shifts in the process mean. We found that the CUSUM control chart has a faster detection of sustained shifts. This finding generalized to almost all shift sizes and estimation errors considered in this paper. The performance of the EWMA was not far worse than that of the CUSUM, but the Shewhart control chart was much slower in detecting sustained shifts in the mean compared to these two charts.

In this thesis we have assumed that the process data are normally distributed. How our control charts perform under non-normal distributions is an interesting topic for future research.

Abbreviations

Abbreviation	Description
ARL_{IN}	in-control average run length
ARL_{OOC}	out-of-control average run length
ARL_0	nominal in-control average run length
CARL	conditional average run length
$CARL_{IN}$	conditional in-control average run length
$CARL_{OOC}$	conditional out-of-control average run length
$AARL_{IN}$	unconditional in-control average run length
$AARL_{OOC}$	unconditional out-of-control average run length
$SDCARL_{IN}$	standard deviation of the conditional in-control average run length
$SDARL_0$	nominal standard deviation of the conditional in-control average run length
$CARL_{IN,p}$	$100p^{th}$ percentile of the conditional in-control average run length
CPA	conditional probability of an alarm
FAR	false alarm rate
FAR_0	nominal false alarm rate
CFAR	conditional false alarm rate
RE	relative error(difference)
PD	percentage difference (RE*100)

Samenvatting

Variatie in processen wordt verdeeld in variatie door gewone oorzaken (common cause variation) en variatie door speciale oorzaken (special cause variation). Wanneer een proces uitsluitend variatie door gewone oorzaken bevat wordt een proces verondersteld beheerst te zijn, terwijl van een proces met variatie door zowel gewone als speciale oorzaken wordt gezegd dat het niet-beheerst is. We noemen beheerst ook wel in-control (IC) en niet-beheerst out-of-control (OOC). OCC regelkaarten zijn grafische hulpmiddelen waarmee een onderscheid gemaakt kan worden tussen gewone oorzaken en speciale oorzaken van variatie. Er zijn verschillende typen regelkaarten, waaronder de meest gebruikte de *Cumulative Sum* (CUSUM), *Exponentially Weighted Moving Average* (EWMA) en de *Shewhart* regelkaart zijn. CUSUM en EWMA regelkaarten zijn ontworpen om kleine verschuivingen in procesparameters te detecteren. Shewhart regelkaarten daarentegen zijn uitermate geschikt om grote verschuivingen in procesparameters waar te nemen.

De kansverdeling van de run lengte en zijn eigenschappen wordt gebruikt om regelkaarten te evalueren. De run lengte is een stochastische variabele die het aantal steekproeven telt dat genomen moet worden voordat een regelkaart een OOC signaal geeft dat het proces niet beheerst meer is. Als het proces beheerst (IC) is, is de kans dat de kaart zo'n OOC signaal geeft de *false alarm rate* (*FAR*). Wanneer voor de Shewhart regelkaart de procesparameters bekend zijn, zijn deze signalen onafhankelijk en heeft de IC run lengte een geometrische verdeling met parameter *FAR*. In dat geval is de *FAR* de reciproque van de gemiddelde IC run lengte, die we noteren als ARL_{IN}. Echter, voor de CUSUM en EWMA regelkaarten zijn de signaalmomenten niet onafhankelijk, aangezien de corresponderende karakteristieken waarop die regelkaarten gebaseerd zijn, gebruik maken van huidige en oude data. Dit heeft als gevolg dat de verdeling van de run lengte niet meer geometrisch is, en dus is de ARL_{IN} niet de reciproque van de FAR. Regelkaarten, gebaseerd op procesparameters die bekend zijn, zijn ontworpen door het vinden van regelgrenzen met een constante die correspondeert met een specifieke nominale ARL_{IN} waarde, zeg ARL_0 , of een nominale *FAR* waarde, zeg FAR_0. De regelkaartconstante voor de Shewhart kaart is

meestal het FAR_0^{th} percentiel van de *IC* steekproefverdeling van de stochastische grootheid waarop de regelkaart is gebaseerd. Echter, voor de EWMA en CUSUM regelkaarten worden de regelkaartconstanten gevonden via een numerieke methode, aangezien de signaalmomenten niet onafhankelijk zijn.

Regelgrenzen zijn functies van de constante en de veronderstelde procesparameters. Wanneer parameters onbekend zijn, worden regelkaarten retrospectief gebruikt om de procesparameters te kunnen schatten. Hiervoor hebben we data nodig waarvan vastgesteld wordt dat deze beheerst (IC) zijn. Dit wordt *Phase I* analyse genoemd. Het resultaat van de *IC* retrospectieve data wordt gebruikt om de onbekende procesparameters en de regelgrenzen te schatten voor het monitoren van het proces in de volgende fase, die Phase II heet. Vanwege de variatie in steekproeven door gebruikers krijgt men verschillende schatters, verschillen in regelgrenzen en daardoor verschillen in prestatie niveaus van de regelkaart. In de literatuur staat deze steekproefvariatie bekend als *van gebruiker tot gebruiker variatie.*

Derhalve is de *ARL* een stochastische variabele met een kansverdeling. Als stochastische variabele wordt de *ARL* de voorwaardelijke *ARL* genoemd en aangeduid als *CARL*. Ideaal zou de in-control *CARL* (aangeduid als $CARL_{IN}$) verdeling gecentreerd rond een nominaal gespecificeerde *ARL* (ARL_0) moeten zijn. Vandaar dat regelkaarten worden geëvalueerd door de verwachte waarde van de $CARL_{IN}$ verdeling te vergelijken met de ARL_0. Bovendien zijn ze zo ontworpen dat de $E(CARL_{IN})=ARL_0$. In werkelijkheid, echter, zal een gebruiker alleen een enkele steekproef of een enkele regelkaart gebruiken. De $CARL_{IN}$ waarde die correspondeert met deze steekproef verschilt bijna altijd van de ARL_0. Derhalve is de verwachte waarde niet de juiste maat. Om inzicht te krijgen hoe dicht de $CARL_{IN}$ waarden bij de ARL_0 waarde liggen, wordt vaak de standaarddeviatie van de verdeling van de $CARL_{IN}$ (aangeduid door $SDCARL_{IN}$) berekend. Echter, in werkelijkheid berekent de $SDCARL_{IN}$ niet hoe dicht de $CARL_{IN}$ waarden bij ARL_0 liggen, maar hoe dicht de $CARL_{IN}$ waarden bij $E(CARL_{IN})$ liggen. Daarom is de $SDCARL_{IN}$ ook niet de juiste maat, tenzij $E(CARL_{IN})=ARL_0$. Merk op dat in de praktijk de nabijheid van de $CARL_{IN}$ waarden tot ARL_0 minder belangrijk is dan hun afwijking in relatie tot de ARL_0. Bijvoorbeeld, een gebruiker is blijer met een $CARL_{IN}$ waarde die groter is dan de ARL_0, ongeacht de nabijheid tot de ARL_0. Daarom worden regelkaarten tegenwoordig beoordeeld en ontworpen door middel van

overschrijdingskansen. Zo'n kans (in het Engels *exceedance probability)*, aangeduid als *1-p*, is de kans dat $CARL_{IN}$ de ARL_0 overschrijdt. Ideaal zou de overschrijdingskans hoog moeten zijn, laten we zeggen $1 - p \geq 0.90$. Oftewel, deze zogenaamde *exceedance probability criterion* (*EPC)* geeft aan dat de ARL_0 bij voorkeur dichtbij het *pth* percentiel van de $CARL_{IN}$ verdeling moet liggen, waarbij *p* idealiter minder dan of gelijk aan 10% is. Daartegenover staat een verslechtering van de *OOC p*restatie. Het is de prijs die betaald moet worden voor het aanpassen van de regelgrenzen voor de effecten van de parameter schatting. Echter, het is mogelijk de verbetering van de *IC* prestatie te compenseren met de achteruitgang van de *OOC* prestatie.

De laatste twee decennia hebben onderzoekers geprobeerd uit te vinden welke schatters gebruikt zouden moeten worden om de onbekende procesparameters te schatten. Hoe goed of hoe slecht zal een regelkaart met geschatte parameters presteren in vergelijking tot een kaart met bekende parameters? Welke regelkaart is het beste om een specifieke verschuiving te detecteren? Hoe groot moet de steekproef in Phase I zijn zodat een regelkaart met geschatte parameters ongeveer hetzelfde presteert als een regelkaart met bekende parameters? Of hoe groot moet een steekproef in Phase I zijn zodat een regelkaart voldoet aan gespecificeerde prestatie criteria? Hoe moeten regelgrenzen worden aangepast om het effect van geschatte parameters als functie van de beschikbare steekproef voor Phase I te compenseren? Met betrekking tot aangepaste regelgrenzen: hoe brengen we de verbetering van de *IC* prestatie in balans met de verslechtering van de *OOC* prestatie? Alhoewel we belangrijke bijdragen leveren aan alle bovengenoemde vragen, ligt de focus in dit proefschrift op de laatste drie vragen, welke betrekking hebben op het ontwerpen van regelkaarten. We bestuderen Shewhart, EWMA en CUSUM regelkaarten. We nemen aan dat de procesdata onafhankelijk en normaal verdeeld zijn. Bijdragen aan en beperkingen met betrekking tot dit proefschrift worden hieronder meer in detail verklaard.

Bij gebruik van de standaard aanpak in Diko (2014) en Diko et al. (2016), constateerden we de nadelige effecten van (i) parameter schatting, (ii) het gebruik van 3-sigma regelgrenzen en (iii) meervoudig toetsen van de *IC* waarden van de $(\bar{X}, R)$ regelkaart. We leiden formules af om de juiste regelkaartconstanten te vinden, met inachtneming van de bovengenoemde punten. We presenteerden tabellen voor de gecorrigeerde regelkaartconstanten voor een aantal veel gebruikte ARL_0 Waarden. De gecorrigeerde

regelkaartconstanten garanderen dat de $CARL_{IN}$ waarden gecentreerd zijn rondom ARL_0. Echter, deze gecorrigeerde regelkaartconstanten bieden geen garantie dat een gebruiker geen lage $CARL_{IN}$ waarde krijgt. Bij de standaard aanpak voor de Phase II Shewhart regelkaarten voor de spreiding hebben we nieuwe gecorrigeerde regelkaartconstanten afgeleid en getabelleerd in Hoofdstuk 2. Deze kunnen gebruikt worden om de geschatte regelgrenzen te bepalen. Om de standaard aanpak toe te passen hebben we gebruik gemaakt van numerieke en analytische methoden. De twee methoden gaven een vergelijkbaar resultaat. In tegenstelling tot de ongecorrigeerde regelkaartconstanten gaven de gecorrigeerde constanten de $CARL_{IN}$ prestatie gecentreerd rond ARL_0. Daarom zijn de gecorrigeerde constanten beter dan de ongecorrigeerde constanten wanneer parameters worden geschat. Echter, er is geen garantie dat bij deze aangepaste constanten lage $CARL_{IN}$ waarden niet voorkomen. In respectievelijk hoofdstuk 3 en hoofdstuk 4 bestuderen we de prestatie en het ontwerp van de tweezijdige Phase II EWMA en CUSUM regelkaarten voor de locatie, gebruikmakend van de *EPC*. We laten zien dat meer Phase I data nodig zijn dan in de literatuur worden aanbevolen om regelkaarten te ontwerpen die een specifieke EPC garanderen. We presenteren een nieuwe methode voor het aanpassen van de regelkaartgrenzen volgens de *EPC*. Deze methode is een beter alternatief voor parametrische *bootstrapping.* Net zoals de *bootstrap* methode biedt onze methode een andere manier voor simulaties die gepaard gaan met een zoekalgoritme. Onze methode is sneller dan parametrische *bootstrapping*. Bovendien geeft onze methode meer accurate en betrouwbare resultaten dan de *bootstrap* methode, omdat we de simulaties op een andere manier uitvoeren. Gebaseerd op deze nieuwe methode worden tabellen en grafieken van de aangepaste regelkaartconstanten bepaald voor een uitgebreid scala van de regelkaart parameters. In de praktijk zal dit helpen bij de implementatie van de door *EPC* gecorrigeerde regelkaartcontanten, aangezien deze constanten momenteel niet beschikbaar zijn in studieboeken of in computer software. Een in-control (IC) en out-of-control (*OOC*) prestatie evaluatie van de aangepaste regelgrenzen wordt ook gepresenteerd. We zien dat de aangepaste regelgrenzen een gespecificeerde minimum $CARL_{IN}$ prestatie garanderen (met grote waarschijnlijkheid) tegen de marginale verslechtering van de *OOC* prestatie. We geven aan hoe een balans gevonden kan worden voor de *IC* en *OOC* prestatie. Aangezien de numerieke methoden om de *CARL* verdeling te benaderen veeleisend en soms onmogelijk zijn, presenteren we in hoofdstuk 4 bovendien alternatieve formules om de *CARL* verdeling

te benaderen. We laten hun eenvoud, accuratesse en veelzijdigheid zien en bevelen ze ten zeerste aan voor dagelijks gebruik.

In Hoofdstuk 5 vergelijken we de gecorrigeerde Shewhart, CUSUM en EWMA regelgrenzen in het geval we te maken hebben met verschuivingen in het proces gemiddelde. We vinden dat de CUSUM regelkaart verschuivingen sneller detecteert. Deze constatering geldt in het algemeen voor zowel grote als kleine verschuivingen. De prestaties van de EWMA kaart zijn niet veel slechter dan die van de CUSUM kaart, maar de Shewhart regelkaart was veel langzamer in het detecteren van verschuivingen in het gemiddelde in vergelijking tot deze twee regelkaarten.

In dit proefschrift gaan we ervan uit dat de procesgegevens normaal verdeeld zijn. Hoe deze regelkaarten zich gedragen als het proces niet-normaal verdeeld is kan een interessant onderwerp zijn voor toekomstig onderzoek.

Acknowledgements

This PhD thesis is dedicated to my parents. I thank them for their unconditional love and support. Moreover, I thank them for emphasizing to me, above all other virtues, authenticity.

I am very grateful to my PhD promoter, *Ronald Does,* for choosing me to be his student. Ronald has won a lot of accolades in industry and academia, for example, the Shewhart and the Box medals. He radiates experience, humility, professionalism and knowledge. Anyone next to him is bound to catch his positive aura. The three years I spent under his supervision have been the most efficient and productive years in my academic career. During this time I managed to write 5 peer reviewed articles (4 as the first author and 1 as a co-author). Through him, I also travelled the world and met great people.

I am also grateful to my second PhD promoter, *Subha Chakraborti*. I met him 10 years ago at the University of Pretoria, where he was a SARCHi chair in non-parametric statistics and statistical process control. Since then, even though he was based in the USA, we never lost contact and we wrote great papers together. *Subha* introduced me to *Ronald Does*. *Subha* was also my MSc promoter.

Nkosinathi Tyolo, Mpho Moletsane, Dan Modigoe, Xolani Notshe, Goodfriday and *Nonezile Kom*, you came in at very critical moments of my teenage years and motivated me to be the best I can be. I thank you for your love and unwavering trust in me. *Mncedisi Michael Willie, Sandile Dube, Thokozani Gamede, Thulani Sibiya* and *Bonani Mkoko*, without your friendship, I would not have made it this far. Kwanga kunganda apho nithatha khona.

Many thanks to my work colleagues *Rob Goedhart, Leo, Marit, Atie, Reza, Bart, Janusz, Stefan, Ujjwal, Julien, Alex, Rob Lee, Jieyi, Yannick* and *Antonio* for the wonderful research environment and all the fun times we had in our department, in particular, the plop moments. I am also thankful to have stayed in Hotel Jansen, it was a home away from home and a place where I met friends like *Slavo, Emma, Dragos, Elif, Miodrag, Marco, Juliana, Lilian, Antonella, Simonetta, Ron, Klaas, Klaartje, Fransesco, Giuliano* and *Konstantin*. I am so glad I met you comrades.

Umzingisi akanashwa!

Mandla Diko

Amsterdam, October 2019

Curriculum Vitae

Mandla Diko was born on the 26th of June 1978 in Alexandra Township, north of Johannesburg, South Africa. He did his primary education at Zenzeleni, Ithute and Majali primary schools. Mandla started his secondary education at Siyazama high school in Peelton and completed it at Thubalethu high school in Fort Beaufort. In 2005, he graduated at the University of Fort Hare in Alice with a BSc in Mathematical Statistics and a BSc Hons in Applied Statistics. In 2014, he completed his MSc in Applied Statistics from the University of Pretoria. Mandla was a Statistician for the Socio-Economic Surveys section at the Human Sciences Research Council (from 2006 to 2009), the Academic Development and Support unit at the University of Johannesburg (from 2011 to 2013), the Centre for Post Graduate Studies at the Cape Peninsula University of Technology (in 2016) and the Research Ethics Committee of the Council for Scientific and Industrial Research (in 2016). In 2015, Mandla was a lecturer of business statistics at Walter Sisulu University at Ibika in Butterworth.

From 2017 to 2019, he was working as a PhD student at the University of Amsterdam. Mandla likes to play chess, pool and football. He enjoys listening to music and likes reading poetry and parables.