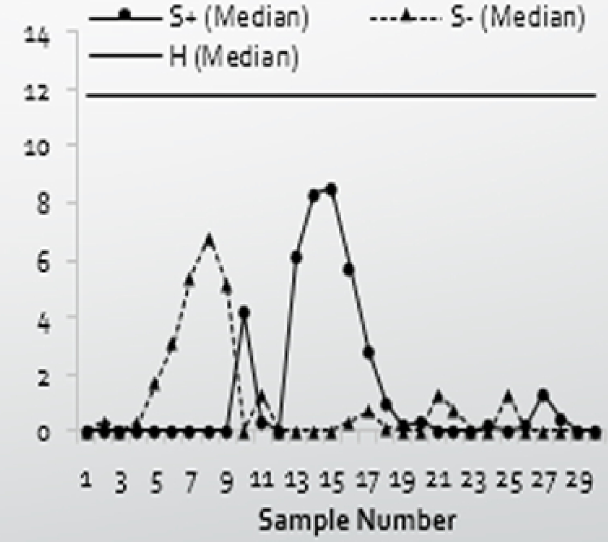
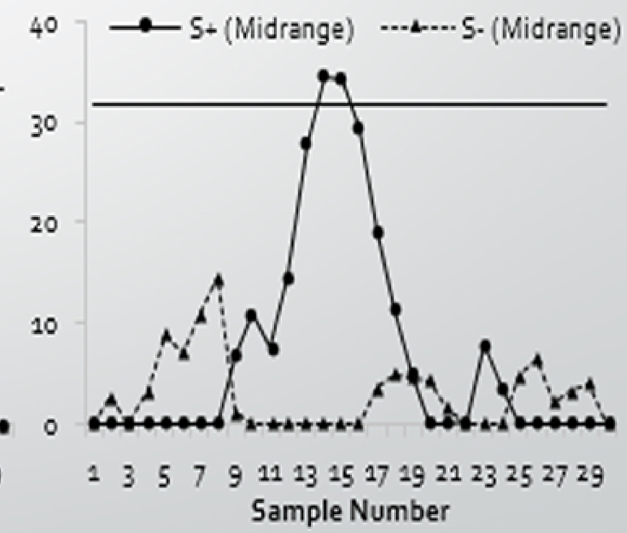
Robust Control Charts in Statistical Process Control

$$S_i^{+} = \max\left[0, +\left(\hat{\theta}_i - \mu_0\right) - K_{\hat{\theta}} + S_{i-1}^{+}\right]$$
$$S_i^{-} = \max\left[0, -\left(\hat{\theta}_i - \mu_0\right) - K_{\hat{\theta}} + S_{i-1}^{-}\right]$$





Hafiz Zafar Nazir

Robust Control Charts in Statistical Process Control
Hafiz Zafar Nazir

Robust Control Charts in Statistical Process Control

Dit proefschrift is mede mogelijk gemaakt door een financiële bijdrage van het Instituut voor Bedrijfs- en Industriële Statistiek van de Universiteit van Amsterdam

Omslagontwerp: Tahir Mahmood

MIX
Paper from responsible sources
FSC® C012854

ISBN: 9789461087140

Robust Control Charts in Statistical Process Control

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties ingestelde
commissie, in het openbaar te verdedigen in de Agnietenkapel
op donderdag 11 september 2014, te 10.00 uur

door

Hafiz Zafar Nazir

geboren te Okara, Punjab, Pakistan

Promotiecommissie:

Promotor: Prof.dr. R.J.M.M. Does

Co-Promotors: Dr. M. Riaz
Dr. M. Schoonhoven

Overige leden: Prof.dr. N.M. van Dijk
Prof.dr. C.G.H. Diks
Prof.dr. M.R.H. Mandjes
Prof.dr. J. de Mast
Prof.dr. K.C.B. Roes
Prof.dr. M. Salomon
Prof.dr. J.E. Wieringa

Faculteit Economie en Bedrijfskunde

To my family

Preface

I am thankful to Almighty Allah, My Creator and Saviour, for the talents and grace bestowed on me. All respect to His Prophet (Peace be upon him) for enlightening the essence of faith in Allah, converting all his kindness and mercies upon me.

First of all I would like to sincerely thank, Prof. dr. Ronald J. M. M. Does, my promoter, for his invaluable advices throughout my Ph.D. research. His professional knowledge, constructive guidance and years long experience were indispensable during the entire process of doing my research and preparing this thesis.

I would like to express my wholehearted gratitude to my co-promoter, Dr. Muhammad Riaz. His contributions in this thesis are not negligible. He really helped me by providing me with many useful suggestions and comments. I would like to say here without any hesitation "Thank you Sir" for your passion, for your constructive criticism, and for being such a nice friendly supporter. I really enjoyed working with you, and I am proud to be your student.

I would also like to thank my co-promoter, Dr. Marit Schoonhoven, whose creative ideas inspired me a lot and who collaborated with me on several articles.

Thanks also to all of my teachers who always guided me throughout my academic career. I am also thankful to Ms. Atie Buisman, Office Manager at the Institute for Business and Industrial Statistics, for her cooperation in all kind of administrative matters and issues. Also thanks go to my colleagues, especially, Ms. Noureen Akhtar and Ms. Sadia Qamar, and Dr. Nasir Abbas, at University of Sargodha, Sargodha, Pakistan, for supporting me in different capacities.

This Ph.D. thesis is based on final versions of some published, accepted and submitted articles. I would like to say thanks to the referees and editors for their helpful comments which always helped a lot to improve the initial versions of these articles and guided me to learn many new things of which I was otherwise ignorant.

At last but not least, thanks are due to my family for their support during all the years of my study. I especially thank my father for always believing in me and my mother for her non-ending prayers for me. Writing this thesis without my family's support was really unthinkable for me.

Hafiz Zafar Nazir

CONTENTS

Chapter 1

Introduction

The chapter provides a brief description of sources of variation in a process, and the most important tool, from the statistical process control tool kit, which is used to eliminate and control sources of variation. Furthermore, an outline of the thesis is given.

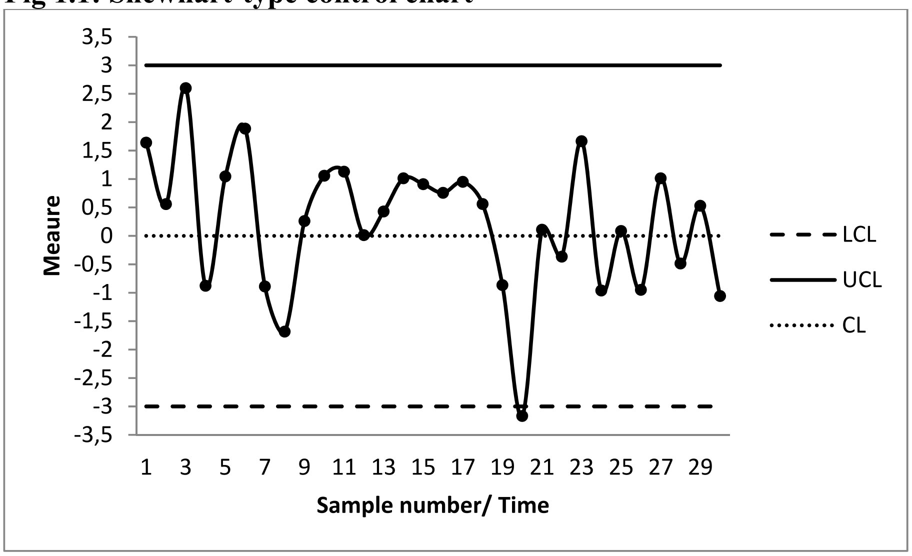
1.1 STATISTICAL PROCESS CONTROL

There exists always a certain amount of variation in the output of the process. Variation due to common causes refers to the inherent or natural variability of the process and is unavoidable and uncontrollable. In the presence of only natural variability, the process is functioning normal and is called statistically in-control. On the contrary, special causes are those sources of variability that arises from improperly adjusted or controlled machines, operator errors, or defective raw material, among others. These special causes are not part of the common causes and therefore directly affect the quality of the output of the process. A process that operates in the presence of special causes of variability is said to be statistically out-of-control. Investigation and corrective actions are required to identify and eliminate special causes so that the process can be brought back to the state of in-control.

Statistical process control (SPC) refers to the kit of statistical procedures and problem solving tools used to control and monitor the quality of the output of any process (cf. Montgomery (2009)). SPC aims in achieving process stability and improving capability by detecting and eliminating, or at least reducing, unwanted sources of variability in the output of the process.

The most valuable tool of the SPC kit is the control chart, which is employed to study the process behavior. The simplest and most widely used control chart is the Shewhart-type chart. This chart is named after the father of quality control, i.e. Walter A. Shewhart, who developed the chart in 1920's at Bell Telephone Laboratories and originally published in 1924. A typical example of a Shewhart-type control chart is given in Figure 1.1.

Fig 1.1: Shewhart-type control chart



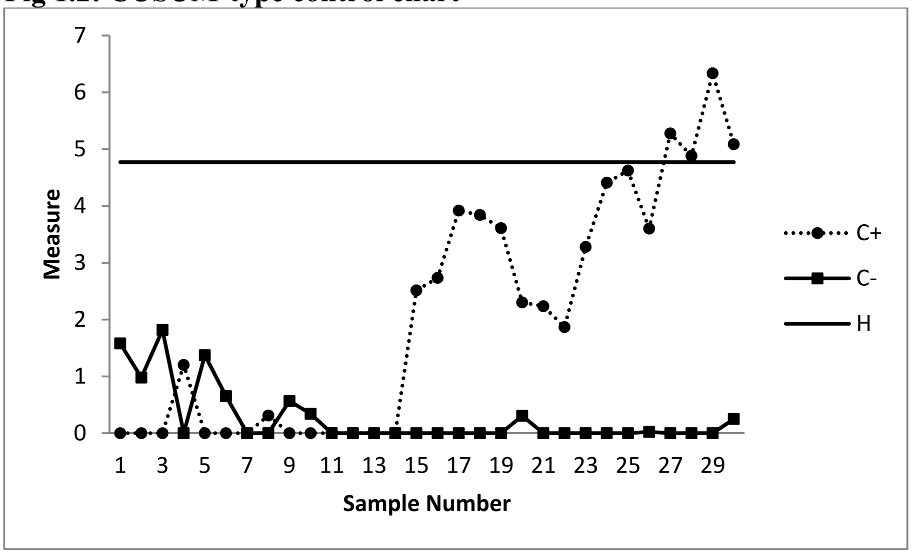
A control chart is a graphical trend chart of a quality characteristic of the process under study. The structure of control chart includes three lines named as Upper Control Limit (UCL), Centre Line (CL) and Lower Control Limit (LCL). These lines are also called the parameters of control chart and the process is declared in-control as long as the plotting statistic(s) (summary measure(s) of the quality characteristic of the process) remain inside these limits (i.e. UCL and LCL). Once the plotting statistic goes beyond UCL or LCL the process is deemed out-of-control. The parameters of the control chart are chosen such that under an in-control situation, there is very small probability of the plotting statistic going beyond the control limits. This probability of getting an out-of-control signal when the process is actually under in-control, is called as false alarm rate (FAR) and is usually denoted by α. This means that when a process is in-control, it is expected that the chart does not signal too often. On the other hand, the probability of getting an out-of-control signal when the process is actually out-of-control is known as the power of the control chart. This means that when a process is out-of-control, it is required that the chart signals as soon as possible. The performance of a chart is evaluated using different measures. Commonly used and the most known measure is the average run length (ARL), which is based on the run length (RL) distribution, where RL is the number of samples needed until the first plotting statistic falls beyond the limits. Along with ARL, it is recommended to report the standard deviation of the run length ($SDRL$) and percentiles of the run length distribution.

The control charting system is normally practiced in two distinct stages: Phase I (so called retrospective phase) and Phase II (the prospective phase). In Phase I, the key concern is to understand the process and to access process stability, making sure that the process is operating at the intended target under some natural causes of variation. Phase I also involves the estimation of the parameters as well as setting up or estimating the control limits. In Phase II, the control chart is used to monitor the process on line in order to detect special causes, like shifts and trends occurring in the process so that any corrective actions can be taken quickly. Phase II focuses on the performance of the control, i.e. how efficient the chart is to detect changes.

The choice of the control charts to be used depends on the characteristics to be measured in the process and what type of amount of change / shift has to be determined. Control charts are classified into two categories, namely memoryless control charts and memory control charts. Shewhart-type control charts are termed as memoryless control charts and their main deficiency is that they are less sensitive to small and moderate shifts in the parameters (location and dispersion). The commonly used memory control charts in the literature include Cumulative Sum (CUSUM) control charts (cf. Page (1954)) and

Exponentially Weighted Moving Average (EWMA) control charts (cf. Roberts (1959)). These memory control charts are designed such that they use the past information along with the current information which makes them very sensitive to small and moderate shifts in the process parameters. An example of a CUSUM-type chart is shown in figure 1.2.

Fig 1.2: CUSUM-type control chart



Most of the evaluations of the existing control charts depend on the assumptions of normality, no contaminations, no outliers and no measurement errors in Phase I for the quality characteristic of interest. In case of violation of these assumptions, the design structures of the charts lose their performance ability and hence are of less practical use. There are a lot of practical situations where non-normality is more common (see for example Janacek and Miekle (1997)). Based on experience, it is common that processes have occasionally outliers in their outputs (cf. Yang et al. (2010)). One of the solutions to deal with this is to use control charts which are robust against violations of the basic assumptions, like normality.

To date, different authors developed control chart schemes based on robust statistics. Langenberg and Iglewicz (1986) suggested using the trimmed mean $\bar{X}$ and R charts. Rocke (1992) proposed the plotting of $\bar{X}$ and R charts with limits determined by the mean of the subgroup interquartile ranges and showed that this method resulted in easier detection of outliers and greater sensitivity to other forms of out-of-control behavior when outliers are present. Tatum (1997) suggested an interesting method for robust estimation of the process standard deviation for control charts. Moustafa and Mokhtar (1999) proposed a robust control chart for location which uses the Hodges-Lehmann and the Shamos-Bickel-Lehmann estimators as estimates of location and scale parameters respectively. This chart provides protection against data errors and gives better performance than the traditional charts if the underlying distribution is not normal. Wu et al. (2002) studied the median absolute deviations based estimators and their application to $\bar{X}$ chart. Moustafa (2009) modified the Shewhart chart by introducing the median as a robust estimator for location and absolute deviations to median as a robust estimator for dispersion and discussed further the different properties under different parent environments. Recently, properties and effects of violations of ideal assumptions (e.g. normality, outlier's free environment, special causes etc.) on Shewhart-type control charts have been studied in detail by Riaz (2008), Schoonhoven, Nazir, et al. (2011) and Schoonhoven and Does (2012). Hawkins (1993) proposed a robust CUSUM chart based on winsorization.

Jensen et al. (2006) emphasized the concerns for future research that: *"The effect of using robust or other alternative estimators has not been studied thoroughly. Most evaluations of performance have considered standard estimators based on the sample mean and the standard deviation and have used the same estimators for both Phase I and Phase II. However, in Phase I applications it seems more appropriate to use an estimator that will be robust to outliers, step changes and other data anomalies. Examples of robust estimation methods in Phase I control charts include Rocke (1989), Rocke (1992), Tatum (1997), Vargas (2003) and Davis and Adams (2005). The effect of using these robust estimators on Phase II performance is not clear, but it is likely to be inferior to the use of standard estimates* *because robust estimators are generally not as efficient"*. Schoonhoven for her Ph.D. thesis worked on the recommendations of Jensen et al. (2006) in context of only Shewhart-type control charts (cf. Schoonhoven (2011)).

This thesis is also planned to a similar subject but in context of Shewhart-type as well as CUSUM-type and other memory control charts. Particularly, this thesis will employ highly robust estimators in Phase I and will study the impact of these estimators on the performance of Phase II control charts.

1.2 LAYOUT OF THIS THESIS

The presence of outliers and contaminations in the output of the process highly affects the performance of the design structures of commonly used control charts and hence makes them of less practical use. One of the solutions to deal with this problem is to use control charts which are robust against violations of the basic assumptions, like normality. The effect of using robust or other alternative estimators has not been investigated thoroughly in perspective of control charting literature. Some authors made their contributions in this foregoing concern. This thesis contributes to the same subject and develops different robust Shewhart-type and CUSUM-type control charts for process spread and process location.

The process dispersion parameter is controlled first, followed by the location parameter. Chapter 2 focuses on monitoring the process dispersion parameter with Shewhart-type control charts by using different Phase I robust estimators, which are not common in literature and evaluates their effect on the Phase II performance of the standard deviation control chart. Along with, a procedure for practitioners is also proposed that is robust against disturbances and is practically more attractive and simple in use. The work of Chapter 2 has been published in *Quality Engineering* (Nazir et al. (2014a)).

Chapter 3 concentrates on an estimation method for Phase I analysis. For this purpose, several robust location estimators are considered. The Phase I method and estimators are evaluated in terms of their mean squared errors and their effect on the $\overline{X}$ control charts used for real-time process monitoring is assessed. It turns out that the Phase I control chart based on the trimmed trimean outperforms the existing estimation methods. The findings in Chapter 3 have been published in *Quality Engineering* (Nazir et al. (2014b)) and the *Journal of Quality Technology* (Schoonhoven, Nazir, et al. (2011)).

The process dispersion parameter is monitored in chapter 4 through Phase II robust Cumulative Sum control charting as this scheme is very sensitive to small shifts in the process parameters. Properties, of the CUSUM design structures based on some already existing estimators as well as some new robust dispersion estimators, are investigated. By evaluating the performance of these estimators based CUSUM control charts in terms of average run length, charts are identified which are more capable to make a good compromise between the efficiency and resistance against unusual situations in terms of statistical and practical needs. The result of Chapter 4 has been accepted for publication in *Quality and Reliability Engineering International* (Nazir et al. (2013)).

The focus of the chapter 5 is to control the location parameter using a CUSUM structure and the major concern is to identify those CUSUM control charts which are of more practical value under different normal, non-normal, contaminated normal parent scenarios. In this chapter, we propose and compare the performance of different CUSUM control charts, for Phase II monitoring of location, based on the mean, median, midrange, Hodges-Lehmann and trimean statistics. An article based on Chapter 5 has been published in *Quality Engineering* (Nazir, Riaz, Does and Abbas (2013)).

Chapter 6 provides a comparison of three memory-types control charts for monitoring process location parameter under normal, contaminated normal and non-normal environments. Performance measures, like the average run length and the extra quadratic loss, are used for comparison purposes. Different location estimators are studied with the structures of the three memory charts and their robustness and efficiency properties are examined. The outcome of Chapter 6 has been submitted for possible publication in *Quality Technology and Quality Management* (Nazir et al. (2014))

Finally, the thesis ends with some concluding remarks, a list of references, a summary and curriculum vitae.

Chapter 2

Robust Shewhart-type Control Charts for Process Dispersion

Commonly used dispersion Shewhart-type control charts are based on the assumption that the unknown process parameters are estimated with clean Phase I data. There are many practical situations where this assumption is violated. This chapter concentrates on the use of high breakdown robust dispersion point estimators in Phase I to control for violation of this assumption and the design of control charts by determining the factors for the control limits. These estimators are evaluated under uncontaminated and contaminated normal environments. Simulations show that control charts based on robust estimators perform better than the other competing control charts when contaminations are present. The chapter ends giving practitioners a stepwise Phase I and Phase II procedure on how to set up a robust Shewhart-type dispersion control chart. This chapter is partly based on Nazir et al. (2014a) (cf. section 2.5).

2.1 INTRODUCTION

Control charts are used to monitor the location and dispersion parameters of the process. The dispersion parameter of the process is controlled first, followed by the location parameter. The concern of this chapter is to monitor the process standard deviation.

Let $Y_{ij}, i = 1,2,\ldots n,$ and $j = 1,2,\ldots,$ indicate the Phase II data taken from the process to be monitored and assume to be distributed as $N(\mu, \delta\theta)$, where δ is a constant. Process dispersion is deemed in-control if $\delta = 1$, otherwise the process standard deviation has changed. When the in-control θ is known, the process dispersion can be monitored through Shewhart-type charts by plotting a summary measure ($\hat{\theta}_j$, an estimate of $\delta\theta$ based on j^{th} sample) against the process control limits:

$$LCL = L_n\theta \quad \text{and} \quad UCL = U_n\theta \tag{2.1}$$

Where L_n and U_n are chosen in such a way that $P\left(LCL \leq \hat{\theta}_j \leq UCL\right) = 1 - P_0$, where P_0 is the probability of false alarms. If the summary measure $\hat{\theta}_j$ falls beyond the control limits, the process is out-of-control and need to be adjusted.

In practice, process parameters are unknown and they need to be estimated from samples. These samples are taken when the process is deemed to be in control. The concern of this chapter is to monitor the process dispersion parameter, where we assume that the in-control process dispersion is unknown and need to be estimated. Let $X_{ij}, i = 1,2,\ldots,n$, and $j = 1,2,\ldots,k$ denote the Phase I data. We assume that X_{ij} are independent and identically distributed with location parameter μ and dispersion parameter θ. Define $\hat{\theta}$ be an unbiased estimator of θ based on Phase I data. Then the control limits in equation (2.1) should be replaced by

$$\widehat{LCL} = L_n\hat{\theta} \quad \text{and} \quad \widehat{UCL} = U_n\hat{\theta} \tag{2.2}$$

The factors L_n and U_n in equation (2.2) may not be same as in equation (2.1) although the probability of false alarms is the same. Hence the properties of the chart's structures in equation (2.1) and equation (2.2) are different and must be interpreted carefully.

Most of the existing Shewhart-type charts depend on the assumptions of normality and an outlier free environment in the quality characteristic of interest. Normality seems more of a theoretical value and it is hard to find practical situations where the normality assumption is easily fulfilled. There are many practical situations where non-normality is more common (cf. Janacek and Miekle (1997)). Based on experience, it is observed that some processes have occasionally outliers in their outputs (cf. Yang et al. (2010)). In case of violation of the normality assumption and the presence of outliers, these commonly used charts lose their efficiency and the performance ability and hence are of less practical use. Generally, control charts are preferred and are of more practical use, whose design structure is not affected by the violation of above mentioned ideal assumptions.

Monitoring of the process dispersion parameter through robust control charting has not received a great deal of attention. Rocke (1989) proposed robust control charts based on different trimmed estimators of dispersion and introduced a two phase procedure which excludes the subgroups that seems to be out-of-control. Tatum (1997) proposed a variant of the biweight A estimator for monitoring the process dispersion parameter and proved this estimator is robust against diffuse and localized disturbances. However, the use of this biweight A estimator makes this charting procedure relatively complicated. Schoonhoven, Riaz and Does (2011) proposed a robust estimation method based on the mean absolute deviation from the median supplemented with a simple screening method. It turns out that this approach is efficient under normality and performs substantially better than the traditional estimators and several robust proposals when contaminations are present. Abbasi and Miller (2012) addressed the issue regarding the proper choice of a control chart for efficient monitoring of process variability. However, they took into account the situation where a large number of samples are available and did not consider the Phase I situation that may involve contaminations. Schoonhoven and Does (2012) proposed algorithms based on a procedure that also screens for individual outliers. Their algorithms remove the excessive variation between subgroups so that only the variation within subgroups is measured. They have shown that these algorithms are very effective when there are diffuse disturbances and when there might also be localized disturbances, the method can be combined with subgroup screening based on the inter quartile range. They observed that the latter procedure reveals a performance very similar to the robust estimator for the standard deviation control chart proposed by Tatum (1997). Nazir, Riaz and Does (2013) investigated the properties of Phase II robust cumulative sum control charts for the process dispersion parameter but they did not consider the estimation effects on the chart's performance.

The chapter focuses, on Shewhart-type control charts for monitoring the process dispersion parameter using a number of robust estimators that are simple in their use and are not common in control charts literature, particularly their design structures and performances under different parent environments and in the presence of special causes in dispersion parameter of the process. The motivation and inspiration of this is taken from Schoonhoven, Riaz and Does (2011), Abbasi and Miller (2012) and Schoonhoven and Does (2012). The next section provides the description of the estimators of dispersion parameter θ.

2.2 PHASE I ESTIMATORS OF DISPERSION AND THEIR EFFICIENCY

David (1998) provided a brief history of standard deviation estimators. The traditional estimators are the pooled sample standard deviation, the mean sample standard deviations and the mean of the sample ranges. Mahmoud et al. (2010) studied the relative efficiencies of these estimators for different sample sizes n and number of samples k under ideal parent environment, i.e. normality. Schoonhoven, Riaz and Does (2011) considered different estimators of the population standard deviation and provided a comprehensive analysis on their efficiency and use in Phase I and Phase II. In deriving the estimates of the process dispersion parameter, we will look at the usual estimator as well as some robust estimators which are not common in the control charts literature.

Let θ be the process dispersion parameter of the quality characteristic which needs to be monitored through control charting and $\hat{\theta}$ be its estimator based on Phase I data, i.e. k

subsamples each of size n. There may be many choices for $\hat{\theta}$ out of which we consider here the following estimators: Pooled sample standard deviation (S_p), Bickel and Lehmann's estimator (cf. Bickel and Lehmann (1979)), and robust estimators proposed by Rousseeuw and Croux (cf. Croux and Rousseeuw (1992) and Rousseeuw and Croux (1993)).

Description of the Proposed Phase I Estimators: The first estimator of the dispersion θ is based on the pooled sample standard deviation

$S_p = \left(\frac{1}{k}\sum_{j=1}^{k} S_j^2\right)^{1/2}$, where S_j is defined by:

$$S_j = \left(\frac{1}{n-1}\sum_{j=1}^{k}(X_{ij} - \bar{X}_j)^2\right)^{1/2} \qquad (2.3)$$

The unbiased estimator is given by $S_p/c_4(k(n+1)-1)$, where $c_4(f)$ is defined by

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

This estimator is very proficient under normality but has a zero breakdown point (proportion of the outlying observations that an estimator can handle before giving arbitrary results) which makes it non robust to outliers and is only included for comparison purposes.

We also evaluate an estimator of the population dispersion mentioned by Shamos (1976) and Bickel and Lehmann (1979). This estimator is obtained replacing pair wise averages by pair wise distances, and is defined as

$$B_{n_j} = 1.0483 * median\{|X_{ij} - X_{lj}|; i < l\} \qquad (2.4)$$

The unbiased estimator is $\bar{B}_n = \frac{1}{b(n)k}\sum_{j=1}^{k} B_{n_j}$ with $b(n)$ as an unbiasing factor. Some values of $b(n)$ may be found in Table 2.1. This robust estimator has a 86% efficiency under normality and has only a 29% breakdown point.

Croux and Rousseeuw (1992) and Rousseeuw and Croux (1993) proposed different robust estimators of the population dispersion parameter θ, which are highly robust against outliers and their efficiency under normality is more as compared to the estimator proposed by Hampel (1974).

The first estimator of Rousseeuw and Croux (1993) is defined as:

$$Q_{n_j} = 2.2219 * \{|X_{ij} - X_{lj}|; i < l\}_{(p)} \tag{2.5}$$

with $p = \frac{h!}{2!(h-2)!}$ where $h = [n/2] + 1$, is roughly half the number of observations (the symbol [.] represents the integer part of a fraction). That is, we take the pth order statistic of the $\binom{n}{2}$ interpoint distances. $\bar{Q}_n = \frac{1}{q(n)k}\sum_{j=1}^{k} Q_{n_j}$, is the unbiased estimator of the population standard deviation, where the values of the unbiasing constant $q(n)$ can be found in Croux and Rousseeuw (1992). This estimator is very robust in presence of outliers in the data and its efficiency at normal distribution is 82% and having breakdown point 50%.

The next estimator proposed by Rousseeuw and Croux (1993) is defined as:

$$S_{n_j} = 1.1926 * median_i\{median_l|X_{ij} - X_{lj}|; i \neq l\} \tag{2.6}$$

This estimator has low efficiency under normality (52%) as compared to the Q_n estimator but has high efficiency as compared to estimator proposed by Hampel (1974). This estimator is also very robust against outliers having a 50% breakdown point. An unbiased estimator of population dispersion is given as $\bar{S}_n = \frac{1}{s(n)k}\sum_{j=1}^{k} S_{n_j}$. Values of the unbiasing constant $s(n)$ can be found in Croux and Rousseeuw (1992).

The third estimator of Rousseeuw and Croux (1993) included in this chapter is

$$T_{n_j} = 1.38 * \frac{1}{h}\sum_{d=1}^{h}\{median|X_{ij} - X_{lj}|; i \neq l\}_{(d)} \tag{2.7}$$

This reads as follows: for each i we compute the median of $|X_{ij} - X_{lj}|, l = 1,2, \dots, n$. This yields n values and the average of the first h-order statistics gives the final estimate T_{n_j}. The breakdown point of the T_{n_j} estimator is 50% and is very robust against outlying observations and its efficiency under normality is only 52%. $\bar{T}_n = \frac{1}{t(n)k}\sum_{j=1}^{k} T_{n_j}$ is an unbiased estimator of population dispersion parameter where $t(n)$ is an unbiasing factor. Some values of $t(n)$ may be found in Table 2.1.

The last estimator we use in this chapter is based on order statistics of certain subranges proposed by Croux and Rousseeuw (1992) which has a breakdown point of 50% and is defined as:

$$R_{nj} = 1.4826 * \left|X_{(i+[0.25n]+1)j} - X_{(i)j}\right|_{\left(\left[\frac{n}{2}\right]-[0.25n]\right)}, i = 1,2, \dots \dots . n. \quad (2.8)$$

Estimator R_{nj} is a very robust estimator in the presence of outliers and its different features can be found in Croux and Rousseeuw (1992). Its efficiency under normality is only 37%. However, it is more efficient than that of the median absolute deviation from median estimator for small samples. Using unbiasing factor $r(n)$ as a function of n, the unbiased estimator is $\bar{R}_n = \frac{1}{r(n)k}\sum_{j=1}^{k} R_{n\,j}$. The unbiasing factors of the Phase I estimators are evaluated through simulation and are given in Table 2.1. It is gratifying to note that the values of $c_4(f)$, $q(n)$ and $s(n)$ are equal to ones from the literature.

Table 2.1: Unbiasing factors for the Phase I estimators

n	$c_4(f)$		$b(n)$	$q(n)$	$s(n)$	$t(n)$	$r(n)$	$c_4(n)$
	$k = 30$	$k = 75$						
5	0.9979	0.9992	1.1848	0.7413	1.0999	1.1281	0.959	0.9400
9	0.9990	0.9996	1.1451	0.8824	1.053	1.0391	1.0004	0.9693

Efficiency of the Phase I Estimators: To understand the usefulness and properties of estimators, Tatum (1997) suggested the use of diffuse and localized disturbances. Diffuse disturbances are those which are equally likely to perturb any observation, whereas a localized disturbance will have an impact on all observations of a particular subsample(s). Standardized variances are used to evaluate the accuracy of the dispersion estimators used in this study, as suggested by Rousseeuw and Croux (1993) and Abbasi and Miller (2012). The standardized variance (SV) of dispersion estimator $\hat{\theta}$ is calculated as:

$$SV_{\hat{\theta}} = \frac{nVAR(\hat{\theta})}{\left[E(\hat{\theta})\right]^2} \tag{2.9}$$

The denominator of $SV_{\hat{\theta}}$ is needed to obtain a natural measure of the accuracy of a scale estimator (cf. Bickel and Lehmann (1976)). $SV_{\hat{\theta}}$ is simulated by generating 10^4 subsamples with $k = 30$ and $k = 75$, each of size $n = 9$ under the following environments:

1. A model (say uncontaminated case) in which all observations of a subsample are form $N(0,1)$.
2. A model for diffuse symmetric variance disturbances in which each observation of a sample has a 95% probability of being drawn from $N(0,1)$ distribution and a 5% probability of being drawn from $N(0, a)$ distribution with $a = 1.5, 2, \dots \dots .5.5, 6.0$.
3. A model for diffuse asymmetric variance disturbances in which each observation in a sample is drawn from $N(0,1)$ and has a 5% probability of having a multiple of a χ_1^2 variable added to it, with the multiplier equal to $a = 1.5, 2, \dots \dots .5.5, 6.0$.
4. A model for localized variance disturbances in which observations in three (when $k = 30$) or six (when $k = 75$) samples are drawn from $N(0, a)$ distribution with $a = 1.5, 2, \dots \dots .5.5, 6.0$.
5. A model for diffuse mean disturbances in which each observation has a 95% probability of being drawn from $N(0, 1)$ distribution and a 5% probability of being drawn from the $N(b, 1)$ distribution setting $b = 1, 1.5, 2, \dots \dots .5.5, 6.0$.

Results obtained from above environments are given in Tables 2.2-2.5, and the following observations from these tables are made:

- When a Phase I dataset contains no contaminations, the pooled standard deviation S_p has the lowest SV following the estimators $\bar{B}_n$ and $\bar{Q}_n$. The other estimators have somewhat larger SV. However, when the magnitude of diffuse symmetric variances increases, the SV of S_p increases quickly and hence making this estimator non-robust to outliers as S_p has a zero breakdown point. The $\bar{T}_n$ estimator (having a 50% breakdown point) is not disturbed by such type of disturbances. The other estimators are robust to such types of contaminations and differences among their SVs more of less the same values (cf. Table 2.2).

- Reading Table 2.3 of asymmetric diffuse disturbances, it can be seen that $\bar{B}_n$ performs best having minimum SV followed by $\bar{T}_n$. Commonly used pooled standard deviation S_p in this situation performs badly. The efficiency of other estimators is almost similar. Increasing the subgroup size has the same general results of the estimators.
- In case of localized variance disturbances, the pooled standard deviation S_p does not perform well for large values of the localized variance disturbance as it is non-robust to outliers. Over all, the estimator $\bar{B}_n$ is effective for such type of contaminations. The performance of the other estimators is similar (cf. Table 2.4).
- When the Phase I dataset contains small sizes of diffuse mean disturbances, i.e. ($b = 0.5$ to $b = 3$, the estimator that has the lowest SV is the pooled standard deviation S_p, followed by $\bar{B}_n$ (cf. Table 2.5). Increase in the magnitude of diffuse means ($b > 3$) results in a poor performance of S_p. The performance of the estimators $\bar{Q}_n$ and $\bar{T}_n$ appears the same for large magnitudes of diffuse mean disturbances.

Table 2.2: *SV* values When Phase I data contains Diffuse Symmetric Variance Disturbances

a	$n = 9$ and $k = 30$						a	$n = 9$ and $k = 75$					
	S_p	$\bar{B}_n$	$\bar{Q}_n$	$\bar{S}_n$	$\bar{T}_n$	$\bar{R}_n$		S_p	$\bar{B}_n$	$\bar{Q}_n$	$\bar{S}_n$	$\bar{T}_n$	$\bar{R}_n$
1	0.0186	0.0246	0.0352	0.0389	0.0354	0.0446	1	0.0076	0.0102	0.0139	0.0155	0.0141	0.0177
1.5	0.0205	0.0256	0.0346	0.0382	0.0368	0.0454	1.5	0.0082	0.0103	0.0143	0.0157	0.0146	0.0179
2	0.0262	0.0259	0.0356	0.0411	0.0362	0.0449	2	0.0110	0.0106	0.0147	0.0164	0.0146	0.0183
2.5	0.0384	0.0284	0.0375	0.0416	0.0376	0.0470	2.5	0.0159	0.0114	0.0154	0.0166	0.0153	0.0188
3	0.0580	0.0307	0.0393	0.0420	0.0389	0.0461	3	0.0229	0.0118	0.0157	0.0171	0.0153	0.0193
3.5	0.0768	0.0318	0.0407	0.0438	0.0392	0.0482	3.5	0.0323	0.0128	0.0163	0.0175	0.0155	0.0192
4	0.1023	0.0335	0.0416	0.0452	0.0400	0.0485	4	0.0435	0.0136	0.0166	0.0176	0.0158	0.0198
4.5	0.1326	0.0346	0.0422	0.0447	0.0397	0.0499	4.5	0.0519	0.0140	0.0173	0.0179	0.0159	0.0194
5	0.1549	0.0379	0.0432	0.0440	0.0418	0.0497	5	0.0640	0.0148	0.0172	0.0178	0.0160	0.0195
5.5	0.1838	0.0395	0.0434	0.0455	0.0404	0.0501	5.5	0.0731	0.0147	0.0173	0.0183	0.0161	0.0200
6	0.2079	0.0418	0.0451	0.0462	0.0424	0.0507	6	0.0832	0.0160	0.0174	0.0189	0.0163	0.0203

Table 2.3: *SV* values When Phase I data contains Diffuse Asymmetric Variance Disturbances

a	$n = 9$ and $k = 30$						a	$n = 9$ and $k = 75$					
	S_p	$\bar{B}_n$	$\bar{Q}_n$	$\bar{S}_n$	$\bar{T}_n$	$\bar{R}_n$		S_p	$\bar{B}_n$	$\bar{Q}_n$	$\bar{S}_n$	$\bar{T}_n$	$\bar{R}_n$
1.5	0.1200	0.0269	0.0376	0.0400	0.0377	0.0462	1.5	0.0514	0.0111	0.0147	0.0160	0.0146	0.0186
2	0.2296	0.0284	0.0380	0.0421	0.0375	0.0461	2	0.1020	0.0117	0.0153	0.0167	0.0152	0.0186
2.5	0.3657	0.0304	0.0388	0.0425	0.0381	0.0468	2.5	0.1704	0.0120	0.0157	0.0167	0.0152	0.0182
3	0.5106	0.0306	0.0398	0.0424	0.0389	0.0473	3	0.2276	0.0124	0.0159	0.0170	0.0155	0.0193
3.5	0.6421	0.0322	0.0405	0.0432	0.0387	0.0484	3.5	0.2775	0.0129	0.0161	0.0171	0.0153	0.0191
4	0.7732	0.0333	0.0406	0.0433	0.0399	0.0487	4	0.3489	0.0134	0.0163	0.0173	0.0154	0.0194
4.5	0.8830	0.0350	0.0406	0.0444	0.0385	0.0480	4.5	0.3942	0.0141	0.0166	0.0181	0.0162	0.0191
5	1.0101	0.0359	0.0413	0.0438	0.0396	0.0490	5	0.4422	0.0145	0.0168	0.0178	0.0160	0.0189
5.5	1.0468	0.0366	0.0409	0.0446	0.0392	0.0470	5.5	0.4670	0.0148	0.0167	0.0178	0.0158	0.0197
6	1.1336	0.0385	0.0419	0.0456	0.0395	0.0484	6	0.5188	0.0152	0.0167	0.0175	0.0156	0.0195

Table 2.4: SV values When Phase I data contains Localized Variance Disturbances

a	$n = 9$ and $k = 30$						a	$n = 9$ and $k = 75$					
	S_p	$\bar{B}_n$	$\bar{Q}_n$	$\bar{S}_n$	$\bar{T}_n$	$\bar{R}_n$		S_p	$\bar{B}_n$	$\bar{Q}_n$	$\bar{S}_n$	$\bar{T}_n$	$\bar{R}_n$
1.5	0.0206	0.0258	0.0347	0.0392	0.0363	0.0453	1.5	0.0083	0.0099	0.0141	0.0156	0.0145	0.0182
2	0.0280	0.0268	0.0365	0.0402	0.0383	0.0465	2	0.0107	0.0105	0.0149	0.0164	0.0152	0.0190
2.5	0.0385	0.0280	0.0399	0.0452	0.0413	0.0518	2.5	0.0152	0.0111	0.0157	0.0175	0.0161	0.0204
3	0.0520	0.0304	0.0436	0.0473	0.0443	0.0548	3	0.0204	0.0120	0.0167	0.0189	0.0174	0.0216
3.5	0.0659	0.0328	0.0465	0.0512	0.0485	0.0590	3.5	0.0263	0.0134	0.0178	0.0211	0.0192	0.0243
4	0.0802	0.0363	0.0514	0.0564	0.0526	0.0656	4	0.0328	0.0141	0.0198	0.0220	0.0202	0.0252
4.5	0.0894	0.0398	0.0545	0.0608	0.0567	0.0698	4.5	0.0391	0.0154	0.0209	0.0242	0.0222	0.0277
5	0.1023	0.0429	0.0596	0.0670	0.0604	0.0775	5	0.0447	0.0167	0.0234	0.0260	0.0240	0.0295
5.5	0.1131	0.0452	0.0663	0.0726	0.0663	0.0840	5.5	0.0495	0.0174	0.0254	0.0275	0.0252	0.0321
6	0.1207	0.0491	0.0698	0.0769	0.0719	0.0890	6	0.0553	0.0191	0.0267	0.0302	0.0277	0.0350

Table 2.5: SV values When Phase I data contains Diffuse Mean Disturbances

b	$n = 9$ and $k = 30$						b	$n = 9$ and $k = 75$					
	S_p	$\bar{B}_n$	$\bar{Q}_n$	$\bar{S}_n$	$\bar{T}_n$	$\bar{R}_n$		S_p	$\bar{B}_n$	$\bar{Q}_n$	$\bar{S}_n$	$\bar{T}_n$	$\bar{R}_n$
0.5	0.0187	0.0247	0.0345	0.0385	0.0358	0.0454	0.5	0.0076	0.0097	0.0138	0.0155	0.0147	0.0178
1	0.0191	0.0251	0.0353	0.0396	0.0365	0.0433	1	0.0077	0.0100	0.0139	0.0157	0.0141	0.0179
1.5	0.0192	0.0246	0.0364	0.0386	0.0355	0.0450	1.5	0.0081	0.0102	0.0141	0.0158	0.0145	0.0181
2	0.0219	0.0267	0.0372	0.0407	0.0372	0.0468	2	0.0090	0.0109	0.0147	0.0166	0.0151	0.0190
2.5	0.0260	0.0288	0.0388	0.0422	0.0386	0.0496	2.5	0.0102	0.0114	0.0152	0.0169	0.0154	0.0192
3	0.0300	0.0311	0.0404	0.0447	0.0404	0.0498	3	0.0121	0.0124	0.0165	0.0178	0.0162	0.0200
3.5	0.0353	0.0340	0.0425	0.0471	0.0423	0.0503	3.5	0.0142	0.0135	0.0170	0.0186	0.0165	0.0202
4	0.0432	0.0365	0.0443	0.0479	0.0428	0.0515	4	0.0167	0.0144	0.0179	0.0191	0.0170	0.0208
4.5	0.0497	0.0410	0.0468	0.0513	0.0459	0.0529	4.5	0.0196	0.0156	0.0177	0.0196	0.0178	0.0210
5	0.0569	0.0429	0.0465	0.0506	0.0449	0.0534	5	0.0220	0.0165	0.0180	0.0201	0.0179	0.0210
5.5	0.0650	0.0459	0.0475	0.0521	0.0469	0.0533	5.5	0.0250	0.0175	0.0185	0.0202	0.0184	0.0213
6	0.0703	0.0472	0.0475	0.0519	0.0469	0.0527	6	0.0273	0.0190	0.0190	0.0207	0.0187	0.0214

To summarize, the most effective estimators are $\bar{B}_n$ and $\bar{T}_n$ with respect to all types of disturbances. The $\bar{Q}_n$ estimator is also a potential candidate for diffuse symmetric variances disturbances. The pooled standard deviation S_p is only recommended for an uncontaminated environment.

2.3 DERIVATION OF PHASE II CONTROL LIMITS

The construction of Phase II control limits requires a derivation of the factors L_n and U_n in equation (2.2). We derive the factors L_n and U_n to obtain the desired probability of false alarms (P_0). These factors depend on n, k and P_0. We employ the same summary measure $S/c_4(n)$ as Phase II plotting statistic so that charts differences are due to Phase I estimators. The factors L_n and U_n are hard to find in analytic form. Therefore they are derived by means of 5×10^4 simulation runs. The chosen probability of false alarms is set equal to 0.0027. The resulting L_n and U_n factors for $n = 5, 9$ and $k = 30, 75$ are given in Table 2.6.

Table 2.6: Factors L_n and U_n for Phase II Control Limits

n	$\hat{\theta}$	$k = 20$		$k = 30$		$k = 75$	
		L_n	U_n	L_n	U_n	L_n	U_n
5	S_p	0.1710	2.3520	0.1720	2.3150	0.1730	2.2720
	$\bar{B}_n$	0.1754	2.4000	0.1733	2.3564	0.1710	2.2855
	$\bar{Q}_n$	0.1749	2.4846	0.1690	2.4168	0.1681	2.3000
	$\bar{S}_n$	0.1741	2.5228	0.1703	2.4247	0.1690	2.3100
	$\bar{T}_n$	0.1761	2.4085	0.1711	2.3605	0.1712	2.2922
	$\bar{R}_n$	0.1672	2.5493	0.1710	2.4300	0.1710	2.3234
9	S_p	0.3490	1.8900	0.3500	1.8720	0.3510	1.8510
	$\bar{B}_n$	0.3510	1.8918	0.3520	1.8907	0.3543	1.8576
	$\bar{Q}_n$	0.3523	1.9334	0.3465	1.8948	0.3508	1.8651
	$\bar{S}_n$	0.3486	1.9471	0.3470	1.9190	0.3536	1.8680
	$\bar{T}_n$	0.3514	1.9376	0.3480	1.9060	0.3524	1.8687
	$\bar{R}_n$	0.3467	1.9735	0.3480	1.9200	0.3440	1.8600

To judge the accuracy of the derived factors, marginal probabilities of false alarms are evaluated under a normal environment. These probabilities are assessed by use of simulation and are provided in Table 2.7. We note that the outcomes are satisfactory.

Table 2.7: False Alarms Probability of In-Control Control Limits

n	$\hat{\theta}$	$k = 30$		$k = 75$	
		L_n	U_n	L_n	U_n
5	S_p	0.00135	0.00135	0.00136	0.00134
	$\bar{B}_n$	0.00140	0.00126	0.00135	0.00136
	$\bar{Q}_n$	0.00131	0.00136	0.00132	0.00139
	$\bar{S}_n$	0.00134	0.00131	0.00136	0.00136
	$\bar{T}_n$	0.00134	0.00135	0.00133	0.00134
	$\bar{R}_n$	0.00138	0.00138	0.00133	0.00130
9	S_p	0.00137	0.00132	0.00136	0.00133
	$\bar{B}_n$	0.00145	0.00130	0.00147	0.00128
	$\bar{Q}_n$	0.00132	0.00142	0.00138	0.00128
	$\bar{S}_n$	0.00135	0.00121	0.00147	0.00128
	$\bar{T}_n$	0.00138	0.00130	0.00143	0.00131
	$\bar{R}_n$	0.00140	0.00134	0.00131	0.00145

2.4 PERFORMANCE EVALUTIONS AND COMPARISONS

Proposed schemes are assessed under similar environments (uncontaminated and contaminated) as considered in section 2.2. We set the values of a, b and the multiple equal to 4. The Phase II performance of the design structures is evaluated in terms of the unconditional average run length (ARL). When the process is in-control, an efficient control chart has the probability of false alarms very low and the ARL should be as high as possible. On contrary, for an out-of-control process, we want to achieve the opposite in magnitudes. From a robustness point of view, a chart has its in-control ARL equal to intended ARL value, i.e. 370, corresponding to unconditional probability of false signals set to choose Phase II control limits. For comparison purposes the in-control ARL is a measure for robustness and out-of-control ARL is a measure of efficiency. These performance measures are computed considering different shifts in the process dispersion $\delta\theta$ setting δ equal to 0.5, 1, 1.5 and 2 and are obtained by simulation. The following steps are used for simulation:

(i) For the Phase I dataset $X_{ij}, i = 1,2, \ldots, n,$ and $j = 1,2, \ldots, k,$ under parent environments discussed in section 2, we estimate the process dispersion parameter and construct the control limits $\widehat{LCL}$ and $\widehat{UCL}$ using Table 2.6 of the control factors.

(ii) The phase II sample $Y_{ij}, i = 1,2, \dots n$ is generated from the uncontaminated normal environment and the plotting statistic $S_j/c_4(n)$ is calculated.

(iii) If $\widehat{LCL} < \frac{S_j}{c_4(n)} < \widehat{UCL}$, then the run length encounter is incremented.

(iv) Steps (ii) and (iii) are repeated until $S_j/c_4(n) \geq \widehat{UCL}$ or $S_j/c_4(n) \leq \widehat{LCL}$. When this occurs, a signal is given and the run length at time j is recorded. This run length is one value of the conditional run length.

To obtain unconditional run lengths, steps (i)-(iv) are repeated 5×10^4 times. By averaging these 50,000 values, we obtain the unconditional ARL. The results of the simulations are given in Tables 2.8-2.12. We provide below a comparative analysis among the proposals and other existing control charts under different process scenarios.

2.4.1 COMPARISONS AMONG PROPOSALS

Uncontaminated Case: Table 2.8 shows the unconditional ARL for the situation when the Phase I data are uncontaminated and normally distributed. The chart based on S_p is most powerful and effective in spotting out-of-control states followed by the charts based on the $\bar{B}_n$ and $\bar{T}_n$ estimators. The performance across the charts based on estimators $\bar{Q}_n$ and $\bar{S}_n$ looks similar.

Table 2.8: ARL values When the Phase I data is Uncontaminated

$n = 5$						$n = 9$					
k	$\hat{\theta}$	δ				k	$\hat{\theta}$	δ			
		0.5	1	1.5	2			0.5	1	1.5	2
30	S_p	54.01	424.32	14.54	3.28	30	S_p	8.92	402.16	6.54	1.74
	$\bar{B}_n$	53.31	437.97	17.99	3.57		$\bar{B}_n$	8.75	410.60	7.12	1.80
	$\bar{Q}_n$	60.22	497.21	24.92	4.09		$\bar{Q}_n$	9.90	427.45	7.55	1.83
	$\bar{S}_n$	59.12	490.71	26.25	4.10		$\bar{S}_n$	9.78	460.41	8.64	1.91
	$\bar{T}_n$	56.53	449.64	18.59	3.63		$\bar{T}_n$	9.70	436.89	7.87	1.88
	$\bar{R}_n$	58.87	471.26	28.16	4.26		$\bar{R}_n$	9.61	435.72	8.69	1.90
75	S_p	52.30	389.12	11.80	3.01	75	S_p	8.69	380.40	5.78	1.69
	$\bar{B}_n$	55.41	401.09	12.59	3.10		$\bar{B}_n$	8.21	372.61	6.08	1.70
	$\bar{Q}_n$	59.37	431.59	13.78	3.18		$\bar{Q}_n$	8.80	395.91	6.16	1.72
	$\bar{S}_n$	58.24	435.07	14.21	3.26		$\bar{S}_n$	8.37	378.67	6.29	1.73
	$\bar{T}_n$	54.34	414.96	12.97	3.15		$\bar{T}_n$	8.53	396.93	6.35	1.74
	$\bar{R}_n$	55.74	426.91	14.97	3.40		$\bar{R}_n$	9.78	414.13	6.24	1.72

Diffuse Symmetric Variance Disturbances: When the Phase I data contains such type of contaminations, the $\bar{T}_n$ based chart performs the best. The detection ability of other estimators is somewhat less as compare to $\bar{T}_n$ estimator for small sample sizes. The chart based on S_p estimator loses its performance badly under this contamination. For larger samples the efficiency of other estimators is relatively equal (cf. Table 2.9).

Diffuse Asymmetric Variance Disturbances: Observing the results given in Table 2.10 indicates that the disturbances in the form of diffuse asymmetric variance disturbances degrade the efficiency of the pooled standard deviation S_p based chart badly. The $\bar{T}_n$ estimator based chart is the most effective one under such contaminations. Other estimators based charts work equally well.

Localized Variance Disturbances: Under such contaminations, the chart that performs proficiently over all is based on the $\bar{B}_n$ estimator followed by the estimator $\bar{T}_n$ for small sample sizes. With an increase in sample size from $n = 5$ to $n = 9$, the chart based on $\bar{Q}_n$ gets edge in performing better as compared to other estimators. The chart based on the pooled standard deviation S_p estimator becomes ARL biased when the Phase I data contains localized variance contaminations (cf. Table 2.11).

Table 2.9: ARL values When Phase I data contains Diffuse Symmetric Variance Disturbances

$n = 5$						$n = 9$					
k	$\hat{\theta}$	δ				k	$\hat{\theta}$	δ			
		0.5	1	1.5	2			0.5	1	1.5	2
30	S_p	23.47	294.02	194.53	24.06	30	S_p	2.89	145.91	153.76	8.48
	$\bar{B}_n$	37.12	436.23	50.10	5.95		$\bar{B}_n$	5.62	362.47	17.05	2.47
	$\bar{Q}_n$	44.37	485.59	63.75	6.61		$\bar{Q}_n$	6.47	403.15	17.07	2.42
	$\bar{S}_n$	43.76	475.55	65.25	6.40		$\bar{S}_n$	6.87	427.51	17.59	2.45
	$\bar{T}_n$	41.29	457.70	43.33	5.45		$\bar{T}_n$	6.80	411.37	15.14	2.33
	$\bar{R}_n$	44.34	468.10	65.55	6.46		$\bar{R}_n$	6.94	418.30	17.10	2.43
75	S_p	20.31	265.63	159.08	13.07	75	S_p	2.62	118.87	125.34	6.07
	$\bar{B}_n$	37.55	454.32	29.72	4.73		$\bar{B}_n$	5.27	340.42	12.71	2.23
	$\bar{Q}_n$	42.11	478.41	30.26	4.67		$\bar{Q}_n$	5.78	380.02	12.35	2.24
	$\bar{S}_n$	42.28	476.89	30.16	4.64		$\bar{S}_n$	5.80	369.51	11.44	2.14
	$\bar{T}_n$	39.27	460.03	26.43	4.38		$\bar{T}_n$	6.08	387.71	10.89	2.09
	$\bar{R}_n$	41.37	471.07	30.54	4.59		$\bar{R}_n$	6.99	433.93	10.53	2.09

Table 2.10: ARL values When Phase I data contains Diffuse Asymmetric Variance Disturbances

$n = 5$						$n = 9$					
k	$\hat{\theta}$	δ				k	$\hat{\theta}$	δ			
		0.5	1	1.5	2			0.5	1	1.5	2
30	S_p	16.17	191.68	212.00	139.75	30	S_p	1.91	67.99	174.39	120.45
	$\bar{B}_n$	39.29	434.78	48.62	5.98		$\bar{B}_n$	6.06	378.30	14.61	2.33
	$\bar{Q}_n$	46.72	489.51	57.36	6.10		$\bar{Q}_n$	6.99	412.92	14.35	2.30
	$\bar{S}_n$	46.10	484.72	55.47	5.97		$\bar{S}_n$	7.32	441.36	14.96	2.32
	$\bar{T}_n$	43.07	461.31	39.73	5.17		$\bar{T}_n$	7.25	432.04	13.32	2.25
	$\bar{R}_n$	47.26	478.92	54.50	6.02		$\bar{R}_n$	7.54	430.90	14.80	2.29
75	S_p	10.52	128.26	259.25	165.79	75	S_p	1.43	32.90	206.05	146.09
	$\bar{B}_n$	38.86	461.97	28.34	4.45		$\bar{B}_n$	5.58	364.80	11.15	2.13
	$\bar{Q}_n$	44.95	481.93	26.04	4.36		$\bar{Q}_n$	6.26	401.69	10.72	2.11
	$\bar{S}_n$	45.04	482.38	25.68	4.40		$\bar{S}_n$	6.28	385.02	10.29	2.08
	$\bar{T}_n$	41.34	464.69	24.05	4.21		$\bar{T}_n$	6.48	405.22	9.57	2.03
	$\bar{R}_n$	44.25	464.87	26.25	4.31		$\bar{R}_n$	7.50	447.27	9.45	2.02

Table 2.11: ARL values When Phase I data contains Localized Variance Disturbances

$n = 5$						$n = 9$					
k	$\hat{\theta}$	δ				k	$\hat{\theta}$	δ			
		0.5	1	1.5	2			0.5	1	1.5	2
30	S_p	12.20	153.47	370.75	95.64	30	S_p	1.61	42.36	345.32	49.85
	$\bar{B}_n$	21.14	286.54	205.66	16.24		$\bar{B}_n$	2.70	129.78	135.09	6.23
	$\bar{Q}_n$	24.34	314.00	267.92	23.66		$\bar{Q}_n$	2.92	149.91	151.40	6.75
	$\bar{S}_n$	23.63	312.96	274.96	25.54		$\bar{S}_n$	2.95	149.73	178.08	7.56
	$\bar{T}_n$	22.43	302.93	218.22	16.68		$\bar{T}_n$	2.91	146.31	160.77	7.20
	$\bar{R}_n$	23.92	307.97	274.56	27.70		$\bar{R}_n$	2.96	151.60	180.85	7.81
75	S_p	13.44	174.07	326.40	32.49	75	S_p	1.78	54.61	322.05	15.53
	$\bar{B}_n$	24.74	337.09	85.59	8.28		$\bar{B}_n$	3.06	159.96	48.04	3.93
	$\bar{Q}_n$	26.89	362.59	101.95	8.95		$\bar{Q}_n$	3.26	172.63	54.59	4.04
	$\bar{S}_n$	26.49	353.34	108.14	9.31		$\bar{S}_n$	3.14	165.65	57.31	4.10
	$\bar{T}_n$	24.95	340.21	90.06	8.24		$\bar{T}_n$	3.16	168.16	55.26	4.06
	$\bar{R}_n$	25.87	345.00	112.65	9.36		$\bar{R}_n$	3.51	203.89	55.26	4.07

Table 2.12: ARL values When Phase I data contains Diffuse Mean Disturbances

$n = 5$						$n = 9$					
k	$\hat{\theta}$	δ				k	$\hat{\theta}$	δ			
		0.5	1	1.5	2			0.5	1	1.5	2
30	S_p	20.37	276.65	204.82	15.82	30	S_p	2.65	119.02	151.56	6.77
	$\bar{B}_n$	30.39	389.38	84.73	7.82		$\bar{B}_n$	4.46	279.55	29.76	3.10
	$\bar{Q}_n$	36.22	443.25	102.04	8.79		$\bar{Q}_n$	5.27	339.88	27.94	2.92
	$\bar{S}_n$	38.10	436.58	102.59	8.81		$\bar{S}_n$	5.68	374.44	28.19	2.93
	$\bar{T}_n$	34.61	424.41	72.42	7.03		$\bar{T}_n$	5.73	367.08	22.31	2.73
	$\bar{R}_n$	38.15	433.84	97.15	8.72		$\bar{R}_n$	5.91	370.67	25.32	2.76
75	S_p	19.12	256.25	154.69	12.11	75	S_p	2.49	108.91	112.82	5.79
	$\bar{B}_n$	30.66	409.36	50.20	6.14		$\bar{B}_n$	4.16	255.68	20.52	2.69
	$\bar{Q}_n$	35.31	449.10	46.81	5.91		$\bar{Q}_n$	4.70	305.68	18.92	2.61
	$\bar{S}_n$	36.05	448.71	46.06	5.75		$\bar{S}_n$	4.92	315.19	16.11	2.47
	$\bar{T}_n$	33.09	427.26	41.42	5.55		$\bar{T}_n$	5.14	336.33	15.25	2.40
	$\bar{R}_n$	35.55	444.38	46.30	5.64		$\bar{R}_n$	5.87	393.93	14.39	2.32

Diffuse Mean Disturbances: Reading Table 2.12 it can be observed that the chart based on the $\bar{T}_n$ estimator is most efficient and effective under diffuse mean disturbances and the chart based on $\bar{R}_n$ performs good with larger sample. The performance of the charts based on the $\bar{Q}_n$ and $\bar{S}_n$ estimators is relatively equal and well.

Overall, the chart based on the estimator $\bar{T}_n$ outperforms all charts based on the other estimators in the presence of any type of contaminations and this chart has a similar performance as compared to chart based on the pooled standard deviation under an uncontaminated normal environment.

2.4.2 COMPARISONS WITH SOME EXISTING COUNTERPARTS

The proposals of Schoonhoven, Riaz and Does (2011) and Schoonhoven and Does (2012) are designed to filter out the control charts that are robust against disturbances and are efficient to detect changes in the process dispersion. Abbasi and Miller (2012) also addressed issues regarding the proper choice of a control chart for efficient monitoring of the process variability and investigated the performance of the control charts based on different estimators of process standard deviation under normal and non-normal parent environments. As this article also follows a similar spirit, a comparative discussion is made with the proposed charts of this study and the charts developed by Schoonhoven, Riaz and Does (2011) and Schoonhoven and Does (2012) and the proposals of Abbasi and Miller (2012).

Schoonhoven, Riaz and Does (2011) recommended the use of charts based on the estimators $D7$ (a variant of biweight A estimator proposed by Tatum (1997)) and $\overline{ADM}'$(a screening method based on the mean absolute deviation of the median) under different types of contaminations. By looking at Table 2.13 (the results of $D7$ and $\overline{ADM}'$ are taken from Schoonhoven, Riaz and Does (2011), an interesting feature may be observed that the proposed charts slightly outperforms the charts recommended by Schoonhoven, Riaz and Does (2011) for $\delta < 1$ in the presence of Phase I contaminations, but that the in-control ARL of this control chart is worse as well as its behavior for $\delta > 1$.

Table 2.13: ARL when contaminations are in Phase I for $n = 5$ and $k = 30$

Environments	Chart	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$
Uncontaminated Case	$\overline{ADM'}$	56.5	434	15.7
	$D7$	55.1	427	15.7
	$\bar{T}_n$	56.53	449.64	18.59
Diffuse Symmetric Variance	$\overline{ADM'}$	47.2	450	27
	$D7$	44.1	452	27.8
	$\bar{T}_n$	41.29	457.7	43.33
Diffuse Asymmetric Variance	$\overline{ADM'}$	51.1	448	21
	$D7$	47.5	452	22.6
	$\bar{T}_n$	43.07	461.31	39.73
Localized Variance	$\overline{ADM'}$	55.1	433	17.3
	$D7$	43.6	454	27.9
	$\bar{T}_n$	22.43	302.93	218.22
Diffuse Mean	$\overline{ADM'}$	38.2	403	67.7
	$D7$	36.8	422	52.1
	$\bar{T}_n$	33.09	427.26	41.42

The algorithms developed by Schoonhoven and Does (2012) are also compared with the proposed chart based on the $\bar{T}_n$ estimator. Comparable results of $\bar{T}_n$ chart are given in Table 2.14. Other results in Table 2.14 are taken from Schoonhoven and Does (2012). Outcomes of this table indicate the effectiveness of our proposed chart based on $\bar{T}_n$ estimator for $\delta < 1$ as it also slightly outperforms the algorithms given by Schoonhoven and Does (2012). But the same restrictions mentioned just before still holds.

Abbasi and Miller (2012) compared different Shewhart-type dispersion charts under different parent environments setting the probability of false alarms equal to 0.002 and assessed the performance by using power curves of these charts. We evaluate and compare the performances of those charts with our proposed charts by ARL. The in-control ARL is set equal to 500 so that valid comparisons are made.

Table 2.14: ARL when contaminations are in Phase I for $n = 5$ and $k = 50$

Environments	Chart	$\delta = 0.6$	$\delta = 1$	$\delta = 1.2$
Uncontaminated Case	$\tilde{S}$	131	378	69.5
	$D7$	135	373	69.6
	$\bar{R}^s$	132	369	68.9
	$\overline{MD}^s$	135	375	69.7
	$\overline{MD}^i$	143	371	70
	$\overline{MD}^{i,s}$	143	371	69.9
	$\bar{T}_n$	132.05	370.87	76.91
Diffuse Symmetric Variance	$\tilde{S}$	43.9	297	425
	$D7$	102	484	159
	$\bar{R}^s$	92.2	464	206
	$\overline{MD}^s$	101	474	171
	$\overline{MD}^i$	123	443	114
	$\overline{MD}^{i,s}$	122	446	114
	$\bar{T}_n$	97.28	482.40	187.34
Diffuse Asymmetric Variance	$\tilde{S}$	23.2	149	231
	$D7$	112	461	121
	$\bar{R}^s$	108	450	133
	$\overline{MD}^s$	115	449	119
	$\overline{MD}^i$	130	419	92.7
	$\overline{MD}^{i,s}$	130	422	95
	$\bar{T}_n$	101.00	468.94	166.64
Localized Variance	$\tilde{S}$	42.8	293	436
	$D7$	118	442	103
	$\bar{R}^s$	129	384	77.1
	$\overline{MD}^s$	131	391	78.5
	$\overline{MD}^i$	127	428	100
	$\overline{MD}^{i,s}$	135	404	85.9
	$\bar{T}_n$	58.31	398.40	438.10
Diffuse Mean	$\tilde{S}$	40.2	280	470
	$D7$	80.3	471	286
	$\bar{R}^s$	54.1	358	431
	$\overline{MD}^s$	65	409	394
	$\overline{MD}^i$	115	447	152
	$\overline{MD}^{i,s}$	115	449	152
	$\bar{T}_n$	80.98	467.81	276.26

The proposed chart based on the estimator B_n (obtained by pair wise distances) in this study outperforms the proposals of Abbasi and Miller (2012) in almost all non-normal environments and the Q_n estimator works well for small shifts under skewed distributions (cf. Table 2.15).

Table 2.15: ARL under Normal and Non-Normal Parent Environments for $n = 10$

Environment	δ	0	0.316	0.632	0.947	1.263	1.579	1.895	2.211	2.526	2.842	3.158	3.474
$N(0,1)$	S	500	16.3	3.3	1.7	1.3	1.1	1.1	1.0	1.0	1.0	1.0	1.0
	B_n	500	24.3	4.7	2.2	1.5	1.2	1.1	1.1	1.0	1.0	1.0	1.0
	T_n	500	39.6	7.6	3.3	2.1	1.6	1.4	1.2	1.2	1.1	1.1	1.1
	Q_n	500	32.6	6.2	2.7	1.8	1.4	1.2	1.1	1.1	1.1	1.0	1.0
	S_n	500	45.9	8.7	3.6	2.2	1.7	1.4	1.3	1.2	1.1	1.1	1.1
	R_n	500	31.1	6.5	3.0	2.0	1.5	1.3	1.2	1.1	1.1	1.1	1.1
T_5	S	500	168.9	48.2	16.9	7.4	4.1	2.7	2.0	1.6	1.4	1.3	1.2
	B_n	500	58.7	11.2	4.3	2.4	1.7	1.4	1.2	1.2	1.1	1.1	1.0
	T_n	500	51.3	11.0	4.6	2.7	2.0	1.6	1.4	1.3	1.2	1.1	1.1
	Q_n	500	56.0	11.5	4.6	2.6	1.9	1.5	1.3	1.2	1.1	1.1	1.1
	S_n	500	50.3	10.7	4.5	2.7	1.9	1.6	1.4	1.3	1.2	1.1	1.1
	R_n	500	49.7	10.9	4.6	2.7	2.0	1.6	1.4	1.3	1.2	1.1	1.1
$exp(1)$	S	500	98.8	23.9	9.6	5.3	3.4	2.5	2.0	1.7	1.5	1.4	1.3
	B_n	500	91.7	21.3	8.3	4.5	3.0	2.2	1.8	1.6	1.4	1.3	1.2
	T_n	500	103.0	26.1	10.8	5.9	3.9	2.8	2.3	1.9	1.7	1.5	1.4
	Q_n	500	73.0	17.5	7.4	4.1	2.8	2.1	1.8	1.5	1.4	1.3	1.2
	S_n	500	93.4	24.5	10.2	5.7	3.8	2.8	2.2	1.9	1.7	1.5	1.4
	R_n	500	86.6	22.0	9.4	5.3	3.5	2.6	2.1	1.8	1.6	1.5	1.4
χ^2_5	S	500	65.1	13.6	5.2	2.8	1.9	1.5	1.3	1.2	1.1	1.1	1.1
	B_n	500	51.9	10.3	4.1	2.3	1.7	1.4	1.2	1.1	1.1	1.1	1.0
	T_n	500	69.3	14.9	5.9	3.3	2.2	1.8	1.5	1.3	1.2	1.2	1.1
	Q_n	500	45.8	9.5	3.9	2.3	1.7	1.4	1.3	1.2	1.1	1.1	1.1
	S_n	500	63.3	14.0	5.6	3.2	2.2	1.7	1.5	1.3	1.2	1.2	1.1
	R_n	500	54.6	11.6	4.8	2.8	2.0	1.6	1.4	1.3	1.2	1.1	1.1
$Logist(0,1)$	S	500	39.0	7.4	3.0	1.9	1.4	1.2	1.1	1.1	1.0	1.0	1.0
	B_n	500	43.0	8.1	3.3	2.0	1.5	1.3	1.2	1.1	1.1	1.0	1.0
	T_n	500	54.4	10.9	4.4	2.6	1.9	1.5	1.4	1.2	1.2	1.1	1.1
	Q_n	500	42.8	8.8	3.7	2.2	1.7	1.4	1.2	1.2	1.1	1.1	1.1
	S_n	500	47.7	10.2	4.3	2.6	1.9	1.5	1.3	1.2	1.2	1.1	1.1
	R_n	500	48.1	10.4	4.3	2.6	1.9	1.6	1.4	1.2	1.2	1.1	1.1
$G(2,1)$	S	500	66.5	15.3	5.8	3.2	2.2	1.7	1.4	1.3	1.2	1.1	1.1
	B_n	500	62.7	12.2	4.7	2.7	1.9	1.5	1.3	1.2	1.1	1.1	1.1
	T_n	500	75.2	16.6	6.6	3.7	2.5	1.9	1.6	1.4	1.3	1.2	1.2
	Q_n	500	54.4	11.1	4.4	2.6	1.8	1.5	1.3	1.2	1.1	1.1	1.1
	S_n	500	70.5	16.0	6.4	3.6	2.5	1.9	1.6	1.4	1.3	1.2	1.2
	R_n	500	60.8	13.1	5.4	3.1	2.2	1.7	1.5	1.3	1.2	1.2	1.1

2.5 STEPWISE PROCEDURE FOR PRACTITIONERS

The main objective of this section is to give practitioners a stepwise procedure on how to set up a robust Shewhart-type dispersion chart. In the control chart design, it is standard practice to separate Phase I from Phase II (cf. Vining 2009). During Phase I, control charts are used retrospectively to study historical data samples. Once representative samples are established, the parameters are estimated and control limits are determined and used for the monitoring in Phase II. The following two sections describe these phases for the standard deviation control chart. This section is based on Nazir et al. (2014a). In this paper also a real life example is given which illustrates the procedure.

Phase I Procedure: The UCL and LCL of the Shewhart standard deviation control chart are given in (2.2) with $\hat{\theta}$ the estimated in-control standard deviation and U_n and L_n the constants such that the desired in-control performance is obtained. Usually, these constants are chosen such that the false alarm probability is sufficiently small, namely, 0.0027. Note that for normally distributed random variables the expectation and standard deviation of the sample standard deviation are linear functions of θ and hence the formula (2.2) makes sense. Recall that formula (2.2) can be used for control charts in Phase I as well as in Phase II. For the sake of clarity, we shall add subscripts I and II to $\widehat{UCL}$, $\widehat{LCL}, U_n$ and L_n to indicate the phase to which we refer.

Schoonhoven and Does (2013) analyzed the performance of the Phase I control charts. They showed that the type of estimator used to construct these charts is important. A robust estimator should be selected first because then the Phase I limits are not affected by disturbances and therefore the correct data samples from which θ is estimated are retained. However, an efficient estimator of θ should be used to obtain the final estimate in order to ensure efficiency under normality. Below, we describe a practical step-by-step approach that meets these requirements.

Step 1: Select Phase I Data: We draw k samples of size n from the process when the process is assumed to be in control and we denote again these samples by $X_{ij}, i = 1,2, \dots, n$

and $j = 1,2, \dots, k$. The subgroups (i.e., samples) should reflect random, short-term rather than special cause variation. To ensure this, items within a subgroup should be produced under conditions in which only random effects are responsible for the observed variation. Additional variability due to potential special causes such as a change in materials or personnel will then occur only between subgroups. Furthermore, the subgroup should not be selected over an interval that is too short because measurements may then be highly correlated and not represent just short-term variation.

However, in practice the k samples of size n may contain outliers, shifts, or other contaminations. These can be filtered out by following the next few steps.

Step 2: Construct a Phase I Standard Deviation Control Chart: We start with a robust estimator of the standard deviation based on the trimmed mean of the sample interquartile ranges. Estimate the 10% trimmed mean of the sample interquartile ranges, defined by

$$\overline{IQR}_{10} = \frac{1}{k-2(\lceil k/10 \rceil - 1)} \times \left[\sum_{v=\lceil k/10 \rceil}^{k-\lceil k/10 \rceil + 1} IQR_{(v)} \right] \quad (2.10)$$

where $\lceil z \rceil$ denotes the ceiling function (i.e., the smallest integer not less than z) and $IQR_{(v)}$ is the vth-ordered value of the sample interquartile ranges. The interquartile range of sample j is defined by $IQR_j = Q_{j,3} - Q_{j,1}$, where $Q_{j,1} = X_{j,(a)}$ and $Q_{j,3} = X_{j,(b)}$, with $a = \lceil n/4 \rceil, b = n - a + 1$, and $X_{j,(v)}$ the vth-order statistic in sample j.

To obtain an unbiased estimate of σ from $\overline{IQR}_{10}$ we divide this quantity by $d_{\overline{IQR}_{10}}$, which is a normalizing constant. Values for this constant for various sample sizes are given in Table 2.16.

Table 2.16: Constant for Phase I Procedure

n	$d_{\overline{IQR}_{10}}$	U_I	L_I	d_{IQR}	$d_{\bar{\bar{S}}'}$
3	1.644	2.923	0.042	1.692	0.998
4	2.020	2.525	0.108	2.060	0.997
5	0.951	3.220	0.035	0.990	0.980
6	1.253	2.688	0.093	1.284	0.983
7	1.490	2.403	0.154	1.514	0.985
8	1.683	2.225	0.208	1.704	0.986
9	1.122	2.474	0.146	1.144	0.984
10	1.293	2.281	0.198	1.312	0.985

The Phase I standard deviation control chart limits are derived from $\widehat{UCL}_I = U_I\overline{IQR}_{10}/d_{\overline{IQR}_{10}}$, $\widehat{LCL}_I = L_I\overline{IQR}_{10}/d_{\overline{IQR}_{10}}$,

With U_I and L_I the 0.99865 and 0.00135 quantiles of the distribution of IQR/d_{IQR} (see Table 2.16).

Step 3: Screen for Sample Shifts: Plot the IQR/d_{IQR} 's of the Phase I samples on the standard deviation control chart generated in step 2 (charting the IQR instead of the sample standard deviation or the sample range ensures that localized variance disturbances are identified and samples that contain only one single outlier are retained).

Exclude from the Phase I data set all samples whose IQR/d_{IQR} falls outside the control limits.

Step 4: Construct a Phase I Individuals Chart: Update the spread estimate according to the formula

$$\overline{IQR}' = \frac{1}{k'}\sum_{j\in K} IQR_j \times 1_{\widehat{LCL}_I \le IQR_j/d_{IQR} \le \widehat{UCL}_I}\left(IQR_j\right), \qquad (2.11)$$

with $1_D(x)$ the indicator function, K is the set of samples that are not excluded in step 3, and k' is the number of non-excluded samples.

The next steps should be applied if individual outliers are likely. First, we construct the limits of the Phase I individuals control chart from

$$\widehat{UCL}_{ind} = 3\overline{IQR}'/d_{IQR}\ , \quad \widehat{LCL}_{ind} = -3\overline{IQR}'/d_{IQR}\ . \qquad (2.12)$$

Step 5: Screen for Individual Outliers: Determine the residuals in each sample by subtracting the trimean from each observation in the corresponding sample (to filter out changes representing shifts in the standard deviation only): $resid_{ij} = X_{ij} - TM_j$ with

$$TM_j = \left(Q_{j,1} + 2Q_{j,2} + Q_{j,3}\right)/4$$

where $Q_{j,2}$ is the median of sample j. Note that the trimean is a robust estimate of the location (cf. Tukey (1997)).

Plot the residuals on the individuals chart derived in step 4. Remove the observations from the Phase I data set corresponding to the residuals that fall outside the limits.

Step 6: Obtain the final Estimate of the Standard Deviation: Obtain a new estimate of the standard deviation from the mean of the sample standard deviations

$$\bar{S}' = \frac{1}{k'} \sum_{j \in K} S_j'/c_4\left(n_j'\right), \qquad (2.13)$$

with K the set of samples that are not excluded, k' is the number of non-excluded samples, n_j' is the number of non-excluded observations in sample j, S_j' is the standard deviation derived from the remaining n_j' observations

$$S_j' = \left(\frac{1}{n_j'-1} \sum_{i: X_{ij} \in K_j} \left(X_{ij} - \bar{X}_j\right)^2\right)^{1/2}$$

(2.14)

with K_j the set of retained observations in sample j and $c_4\left(n_j'\right)$ defined by

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

Divide the result by $d_{\bar{S}'}$ (see Table 2.16) to obtain an overall unbiased estimate.

Phase II Steps: Once the in-control reference data set is established, the standard deviation can be estimated from the Phase I data and control limits determined for use in Phases II. Recall that the Phase II control limits are given by (2.2) and $\hat{\theta}$ is determined in step 6 of the

Phase I procedure described above. What remains is the determination of the factors U_{II} and L_{II} in order to obtain the desirable in-control performance.

Schoonhoven, Riaz and Does (2011) presented a formula for U_{II} and L_{II} of the Phase II standard deviation control chart based on the pooled mean of the sample standard deviation $\tilde{S}$. They tested this formula for charts derived from a broad range of Phase I estimators and concluded that the formula is suitable when the variance of the estimator is close to the variance of $\tilde{S}$. Subsequently, Schoonhoven and Does (2013) showed that the variance of the estimator given in (2.11) is close to $\tilde{S}$. Hence, the formula presented by Schoonhoven, Riaz and Does (2011) can also be applied in the procedure discussed here. This is a nice result because the formula is a plug-in so no simulations are required to obtain the constants.

In the next part, we continue with the steps in the approach applicable to Phase II.

Steps 7: Construct a Phase II Standard Deviation Chart: Obtain U_{II} and L_{II} for the Phase II control limits from

$$U_{II} = \frac{\sqrt{F_{n-1,k(n-1)}(1-\alpha/2)}c_4(k(n-1)+1)}{c_4(n)} \tag{2.15}$$

and

$$L_{II} = \frac{\sqrt{F_{n-1,k(n-1)}(\alpha/2)}c_4(k(n-1)+1)}{c_4(n)}, \tag{2.16}$$

with $F_{v,w}$ denoting an F distribution with v numerator degrees of freedom and w denominator degrees of freedom and α the desired false alarm probability (usually 0.0027). Values for $U_{II}$and L_{II} when $n = 3, \ldots, 10$ and $k = 20, 50$ are provided in Table 2.17.

Table 2.17: Constants for Phase II Procedure

n	$c_4(n)$	$k = 20$		$k = 50$	
		U_{II}	L_{II}	U_{II}	L_{II}
3	0.886	3.138	0.041	2.992	0.041
4	0.921	2.625	0.107	2.535	0.108
5	0.940	2.352	0.171	2.286	0.172
6	0.952	2.178	0.227	2.126	0.228
7	0.959	2.055	0.274	2.012	0.276
8	0.965	1.963	0.314	1.926	0.312
9	0.963	1.890	0.349	1.858	0.351
10	0.973	1.832	0.378	1.803	0.380

Substitute the values for $\hat{\theta}$, $U_{II}$and L_{II} into the formula for the standard deviation control limits given by (2.2).

Step 8: Use the Phase II Chart for Online Monitoring: Periodically, collect newly available data (Y_{ij} with $j = 1, 2, 3, \ldots$ and $i = 1, 2, \ldots, n$) and calculate $S_j/c_4(n)$ according to (2.12) (use S in Phase II because this estimator is efficient under normality and sensitive to disturbances). When $S_j/c_4(n)$ falls outside the control limits, look for the cause of the out-of-control signal.

2.6 CONCLUSIONS

Commonly used Shewhart-type control charts for dispersion are based on the assumption that the unknown process parameters are estimated with clean Phase I data. There are many practical situations where this assumption is violated. The charts having a robust design structure are required to cope with such environments. This chapter presents different robust estimators, which have a high breakdown point, to estimate the process dispersion in Phase I. We compare their efficiency by using standardized variances under different contaminated environments. Moreover, we have investigated the effect of estimating the process dispersion by robust estimators and their impact on the performance of Phase II charts. The charts performance is evaluated under different types of disturbances. We have found that the proposed charts of this chapter have very attractive properties relative to the other existing competing charts. Moreover, the chart based on the B_n estimator is effective overall for non-normal environments and the chart based on $\bar{T}_n$ outperforms all other charts (investigated in the section 2.4.1) and the charts developed by (Schoonhoven, Riaz and Does (2011)) and Schoonhoven and Does (2012)), are even more robust, in the presence of any type of contaminations (see section 2.4.2). Their procedures are robust against both diffuse and localized disturbances so that Phase I limits are not affected by these disturbances. One of their robust control charts for dispersion is described step by step, thus making it easy for practitioners to implement and apply.

Chapter 3

Robust Shewhart-type Control Charts for Process Location

This chapter studies estimation methods for the location parameter. Estimation method based on a Phase I analysis as well as several robust location estimators are considered. The Phase I method and estimators are evaluated in terms of their mean squared errors and their effect on the $\bar{X}$ control charts used for real-time process monitoring is assessed. It turns out that the Phase I control chart based on the trimmed trimean far outperforms the existing estimation methods. This method has therefore proven to be very suitable for determining $\bar{X}$ Phase II control chart limits. This chapter is based on the papers Schoonhoven, Nazir, et al. (2011) and Nazir et al. (2014b).

3.1 INTRODUCTION

Shewhart control charts are extensively used in practice to monitor changes in process characteristics from the in-control state. There are many types of control charts applications, ranging from charts that monitor industrial process characteristics to charts that monitor the number of chronic diseases within healthcare.

There is a broad range of control charts available: there are control charts for monitoring individual measurements of a process characteristic as well as subgroup control charts suitable for monitoring process from which samples are taken periodically. Within the category of subgroup control charts, there is a distinction between control charts for the location and control charts for the spread.

A further distinction is made based on the size of the shift to be detected: when the application requires that smaller shifts have to be detected (small means $\delta/(\sigma/\sqrt{n}) < 2$, with δ the shift, σ the in-control standard deviation and n the sample size), control charts that take into account the history of observations are recommended. Examples include the Cumulative Sum (CUSUM) and Exponentially Weighted Moving Average (EWMA) control charts (see Page (1954), respectively, Roberts (1959)). For this chapter we focus on the Shewhart control chart for location.

Control charts were originally developed and tested assuming that the process characteristics are known. In real-world applications, however, the mean and standard deviation of the process characteristic are not known and have to be estimated before the control chart can be developed. To derive such estimates, it is common practice to separate Phase I from Phase II. During Phase I, control charts are used retrospectively to study historical data samples. Once representative samples are established, the parameters are estimated and control limits are determined and used for online monitoring in Phase II. Jensen et al. (2006) conducted a literature survey of the effects of parameter estimation on control chart properties and identified the following issue for future research: *"The effect of using robust or other alternative estimators has not been studied thoroughly. Most evaluations of performance have considered standard estimators based on the sample mean and the standard deviation and have used the same estimators for both Phase I and Phase II. However, in Phase I applications it seems more appropriate to use an estimator that will be robust to outliers, step changes and other data anomalies. Examples of robust estimation methods in Phase I control charts include Rocke (1989), Rocke (1992), Tatum (1997), Vargas (2003) and Davis and Adams (2005). The effect of using these robust estimators on Phase II performance is not clear, but it is likely to be inferior to the use of standard estimates because robust estimators are generally not as efficient"* (Jensen et al. 2006, p.360).

Schoonhoven, Nazir et al. (2011) and Schoonhoven and Does (2013) performed an extensive study of robust estimation in the context of Shewhart location and control charts. Nazir et al. (2014b) summarized these findings in a practical procedure that enables practitioners to create and implement a robust control chart for location without too much difficulty. The main purpose is to give practitioners a stepwise procedure on how to set up a robust location control chart.

The following two sections describe the Phase I and II stages for the Shewhart location control chart.

3.2 PHASE I PROCEDURE

The UCL and LCL of the Shewhart location control chart are given by

$$\widehat{UCL} = \hat{\mu} + C\,\hat{\sigma}/\sqrt{n}\,,\ \widehat{LCL} = \hat{\mu} - C\,\hat{\sigma}/\sqrt{n}\,, \tag{3.1}$$

with $\hat{\mu}$ and $\hat{\sigma}$ the estimates of the in-control mean μ and standard deviation σ, respectively, and C the constant chosen such that the desired in-control performance is obtained. Usually, C is chosen such that the false alarm probability is sufficiently small, namely 0.0027. Recall that formula (3.1) can be used for control charts in Phase I as well as in Phase II. For the sake of clarity, we shall and subscripts I and II to $\widehat{UCL}$, $\widehat{LCL}$ and C to indicate the phase to which we refer.

Schoonhoven, Nazir et al. (2011) analysed the performance of the Phase I location control charts. They showed that the type of location estimator used to construct these charts is important. A robust estimator should be selected first because then the Phase I limits are not affected by disturbances and therefore the correct data samples from which μ is estimated are retained. However, an efficient estimator should be used to obtain the final estimates in order to ensure efficiency under normality. Below, we describe a practical step-by-step approach which meets these requirements.

Step 1: Select Phase I Data: We draw k samples of size n from the process when the process is assumed to be in control and we denote these samples by X_{ij}, $i = 1,2, \ldots, n$, and $j = 1,2, \ldots, k$. The samples should reflect random, short-term rather than special-cause variation. To ensure this, items within a sample should be protected under conditions is which only random effects are responsible for the observed variations. Additional variability due to potential special cause such as a change in material or personnel will then occur only between samples. Furthermore, the sample should not be selected over an interval that is too short because measurements may then be highly correlated and not represent just short-term variation. In practice the k samples of size n may contain outliers, sample shifts or other contaminations. These can be filtered out in Phase I.

Step 2: Obtain a Robust Estimate of the Standard Deviation: The second step is to estimate the standard deviation using the stepwise procedure given by Nazir et al. (2014a) (cf. Chapter 2). The resulting estimate is denoted by $\hat{\sigma}$.

Step 3: Construct a Phase I Location Control Chart: We estimate the mean with a robust estimator, namely the 10% trimmed mean of the sample trimeans (cf. Tukey (1977)), defined by:

$$\overline{TM}_{10} = \frac{1}{k-2\lceil k/10\rceil} \times \left[\sum_{v=\lceil k/10\rceil+1}^{k-\lceil k/10\rceil} TM_{(v)}\right], \tag{3.2}$$

where $\lceil z \rceil$denotes the ceiling function (i.e. the smallest integer not less than z) and $TM_{(v)}$ denotes the v-th ordered value of the sample trimeans. The trimean of sample j is defined by

$$TM_j = \left(Q_{j,1} + 2Q_{j,2} + Q_{j,3}\right)/4, \tag{3.3}$$

where $Q_{j,2}$is the median and $Q_{j,1} = X_{j,(a)}$ and $Q_{j,3} = X_{j,(b)}$ the first and third quartiles with $X_{j,(v)}$the v-th order statistic in sample j and $a = \lceil n/4 \rceil$, $b = n - a + 1$.

The Phase I location control chart limits are derived from

$\widehat{UCL}_I = \overline{TM}_{10} + 3\hat{\sigma}/\sqrt{n}$, $\widehat{LCL}_I = \overline{TM}_{10} - 3\hat{\sigma}/\sqrt{n}$.

Step 4: Screen for Sample Shifts: We plot the TM_j*'s* of the Phase I samples (cf. (3.3)) on the location control chart generated in step 3 (charting the TM_j*'s* instead of the sample means ensures that localized mean disturbances are identified and samples that contain only one single outlier are retained).

We exclude from the Phase I data set all samples whose TM_j falls outside the control limits ($j = 1,2,\dots,k$).

Step 5: Construct a Phase I Individuals Chart: The mean estimate is updated according to formula

$$\overline{TM}' = \frac{1}{k'}\sum_{j\in K'} TM_j \times I_{\widehat{LCL}_I \le TM_j \le \widehat{UCL}_I}\left(TM_j\right) \tag{3.4}$$

with $I_D(x)$ the indicator function, K' the set of samples which are not excluded in step 4 and k' the number of non-excluded samples.

The next steps should be applied if individual outliers are likely. First, we construct the limits of the Phase I individuals control chart from

$$\widehat{UCL}_{ind} = \overline{TM}' + 3\hat{\sigma} , \widehat{LCL}_{ind} = \overline{TM}' - 3\hat{\sigma} . \tag{3.5}$$

Step 6: Screen for Individual Outliers: We plot the individual observations of the sample remaining from step 4 on the individuals chart derived in step 5 and remove from the Phase I data set the observations that fall outside the limits.

Step 7: Obtain the Final Estimate of the Mean: We now obtain a new estimate of the mean from the mean of the sample means

$$\bar{\bar{X}}'' = \frac{1}{k''}\sum_{j\in K''} \frac{1}{n_j'}\sum_{i\in N_j'} X_{ij} \times I_{\widehat{LCL}_{ind} \le X_{ij} \le \widehat{UCL}_{ind}}\left(X_{ij}\right) , \tag{3.6}$$

with K'' the set of samples which are not excluded in steps 4 and 6, k'' the number of non-excluded samples, N_j' the set of observations that are not excluded in sample j and n_j' the number of non-excluded observations in sample j.

3.3 PHASE II STEPS

In Phase I, we have established the reference data set. From this data set, the in-control mean is estimated. We then derive the control limits for Phase II, i.e. the online monitoring stage. Recall that the formula for the Phase II control limits is given by (3.1), $\hat{\sigma}$ is calculated with the procedure described in Nazir et al. (2014a) and $\hat{\mu}$ is calculated in step 7 of the Phase I procedure described in the previous section. What remains is the determination of the factor C_{II} for the control limits.

Schoonhoven et al. (2009) presented a formula to calculate C_{II} for the $\bar{X}$ control chart based on the pooled mean of the sample standard deviations, $\tilde{S}$. They tested this formula for $\bar{X}$ chars derived from a broad range of standard deviation estimators and concluded that the formula is suitable when the variance of the estimator is close to the variance of $\tilde{S}$. Since the variance of the standard deviation estimator described by Nazir et al. (2014a) is close to the variance of $\tilde{S}$ (see Schoonhoven and Does (2013)), the formula presented by Schoonhoven et al. (2009) can also be applied in the procedure discussed here. This is a nice result because the formula is a plug-in so no simulations are required to obtain the constants.

In the next part, we continue with the steps in the approach applicable to Phase I.

Steps 8: Construct a Phase II Location Chart: We obtain C_{II} for the Phase II control limits from the equation

$$C_{II} = c_4(k(n-1)+1)\sqrt{k+1}\,t_{k(n-1)}\,(1-\alpha/2)/\sqrt{k}\,, \tag{3.7}$$

where $c_4(m)$ is defined by

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

and $t_{k(n-1)}(1-\alpha/2)$ denotes the $(1-\alpha/2)$-th percentile of a *t*-distribution with $k(n-1)$ degrees of freedom and α the desired false alarm probability (usually 0.0027). Values for C_{II} and $c_4(m)$ with $m = k(n-1)+1$ for $n = 3, \ldots, 10$, $k = 20,50$ and $\alpha = 0.0027$ are provided in Table 3.1.

The values for $\hat{\sigma}$, $\hat{\mu}$ and C_{II} are substituted into the formula for $\bar{X}$ control limits given by (3.1).

Table 3.1: Constants for Phase II procedure

n	$k = 20$		$k = 50$	
	$c_4(m)$	C_{II}	$c_4(m)$	C_{II}
3	0.994	3.257	0.998	3.100
4	0.996	3.194	0.998	3.076
5	0.997	3.163	0.999	3.064
6	0.998	3.145	0.999	3.057
7	0.998	3.133	0.999	3.053
8	0.998	3.124	0.999	3.049
9	0.998	3.118	0.999	3.047
10	0.999	3.113	0.999	3.045

Step 9: Use the Phase II Location Chart for Online Monitoring: Newly available data (Y_{ij} with $i = 1,2,\ldots.,n$ and $j = 1,2,3,\ldots,$) are collected periodically and used to calculate $\bar{Y}_j = \frac{1}{n}\sum_{i=1}^{n} Y_{ij}$ (the mean in Phase II is used as plotting statistic because the mean is efficient under normality and sensitive to disturbances). When $\bar{Y}_j$ falls outside the control limits, the cause of this out-of-control signal should be investigated.

In Nazir et al. (2104b) a real life example is given to illustrate the practical use of this procedure.

3.4 ROBUST LOCATION ESTIMATORS FOR THE $\bar{X}$ CONTROL CHART

Monitoring a process with respect to its parameters (location and dispersion), Shewhart introduced the idea of control charts in the 1920s. The forthcoming sections focus on Phase I location estimation methods for constructing the location control chart.

Let Y_{ij}, $i = 1,2,\ldots,n$ and $j = 1,2,3,\ldots$, denote Phase II samples of size n taken in sequence of the process variable to be monitored. We assume the Y_{ij}'s to be independent and $N(\mu + \delta\sigma, \sigma^2)$ distributed, where δ is a constant. When $\delta = 0$, the mean of the process is in control; otherwise, the process mean has changed. Let $\bar{Y}_j = \sum_{j=1}^{n} Y_{ij}/n$ be an estimate of $\mu + \delta\sigma$ based on the j^{th} sample Y_{ij}, $i = 1, 2 \ldots n$. When the in-control μ and σ are known, the process mean can be monitored by plotting $\bar{Y}_j$ on a control chart with respective upper and lower control limits

$$UCL = \mu + C_n\sigma/\sqrt{n} \text{ and } LCL = \mu - C_n\sigma/\sqrt{n} \qquad (3.8)$$

where C_n is the factor such that, for a chosen type-I error probability p, we have

$P\left(LCL \leq \bar{Y}_j \leq UCL\right) = 1 - p,$

When $\bar{Y}_j$ falls within the control limits, the process is deemed to be in control. We define E_j as the event that $\bar{Y}_j$ falls beyond the limits, $P(E_j)$ as the probability that sample j falls beyond the limits and RL as the run length, i.e., the number of samples until the first $\bar{Y}_j$ falls beyond the limits. When μ and σ are known, the events E_j are independent, and therefore RL is geometrically distributed with parameter $p = P(E_j)$. It follows that the average run length (ARL) is given by $1/p$ and that the standard deviation of the run length (SDRL) is given by $\sqrt{1-p}/p$.

In practice, the process parameters μ and σ are usually unknown. Therefore, they must be estimated from samples taken when the process is assumed to be in control. This stage in the control charting process is denoted as Phase I (cf. Woodall and Montgomery

(1999), Vining (2009)). The resulting estimates determine the control limits that are used to monitor the location of the process in Phase II. Define $\hat{\mu}$ and $\hat{\sigma}$ as unbiased estimates of μ and σ, respectively, based on k Phase I samples of size n, which are denoted by X_{ij}, $i = 1, 2 \ldots, n$ and $j = 1, 2 \ldots, k$. The control limits can be estimated by

$$\widehat{UCL} = \hat{\mu} + C_n \hat{\sigma}/\sqrt{n} \quad \text{and} \quad \widehat{LCL} = \hat{\mu} - C_n \hat{\sigma}/\sqrt{n} \tag{3.9}$$

Let F_j denote the event that $\overline{Y}_j$ is above $\widehat{UCL}$ or below $\widehat{LCL}$. We define $P(F_j|\hat{\mu}, \hat{\sigma})$ as the probability that sample j generates a signal given $\hat{\mu}$ and $\hat{\sigma}$, i.e.,

$$P(F_j|\,\hat{\mu}, \hat{\sigma}) = P(\overline{Y}_j < \widehat{LCL} \text{ or } \overline{Y}_j > \widehat{UCL} \mid \hat{\mu}, \hat{\sigma}) \tag{3.10}$$

Given $\hat{\mu}$ and $\hat{\sigma}$, the distribution of the run length is geometric with parameter $P(F_j|\hat{\mu}, \hat{\sigma})$. Consequently, the conditional ARL is given by

$$E(RL|\hat{\mu}, \hat{\sigma}) = \frac{1}{P(F_j|\hat{\mu}, \hat{\sigma})} \tag{3.11}$$

In contrast with the conditional RL distribution, the unconditional RL distribution takes into account the random variability introduced into the charting procedure through parameter estimation. It can be obtained by averaging the conditional RL distribution over all possible values of the parameter estimates. The unconditional p is

$$p = E\left(P\left(F_j|\hat{\mu}, \hat{\sigma}\right)\right) \tag{3.12}$$

and the unconditional average run length is

$$ARL = E\left(\frac{1}{P(F_j|\,\hat{\mu}, \hat{\sigma})}\right) \tag{3.13}$$

Quesenberry (1993) showed, for the $\bar{X}$ control chart, that the unconditional in-control and out-of-control ARL values are higher than in the case where the process parameters are known. Furthermore, a higher in-control ARL is not necessarily better because the RL distribution will reflect an increased number of short RLs as well as an increased number of long RLs. He concluded that, if limits are to behave like known limits, the number of samples in Phase I should be at least $400/(n-1)$. Jensen et al. (2006) conducted a literature survey of the effects of parameter estimation on control chart properties and identified some issues for future research. One of their main recommendations is to study robust or alternative estimators for μ and σ (e.g., Rocke (1989, 1992), Tatum (1997), Vargas (2003), Davis and Adams (2005)). The effect of using these robust estimators on Phase II should also be assessed (Jensen et al. (2006, p.360)). These recommendations are the subject of the said sections, i.e. we will examine alternative location estimation methods as well as the impact of these estimators on the Phase II performance of the $\bar{X}$ control chart.

So far the literature has proposed several alternative robust location estimators. Rocke (1989) proposed the 25% trimmed mean of the sample means, the median of the sample means, and the mean of the sample medians. Rocke (1992) gave the practical details for the construction of the charts based on these estimators. Alloway and Raghavachari (1991) constructed a control chart based on the Hodges–Lehmann estimator. Tukey (1997) and Wang et al. (2007) developed the trimean estimator, which is defined as the weighted average of the median and the two other quartiles. Finally, Jones-Farmer et al. (2009) proposed a rank-based Phase I control chart. Based on this Phase I control chart, they define the in-control state of a process and identify an in-control reference sample. The resultant reference sample can be used to estimate the location parameter.

We compare existing and new methods for estimating the in-control μ. The collection of methods includes robust sample statistics for location and estimation methods based on Phase I control charts (cf. Jones-Farmer et al. (2009)). The methods are evaluated in terms of their mean-squared errors (MSE) and their effect on the Phase II $\bar{X}$ control chart performance. We consider situations where the Phase I data are uncontaminated and normally distributed, as well as various types of contaminated Phase I situations.

The rest of this chapter is as follows. First, we present several Phase I sample statistics for the process location and assess the MSE of these estimators. Then we present a new algorithm for Phase I analysis and evaluate its efficiency. Following that, we present the design schemes for the $\bar{X}$ Phase II control chart and derive the control limits. Next, we describe the simulation procedure and present the effect of the proposed methods on the Phase II performance.

3.5 PROPOSED LOCATION ESTIMATORS AND THEIR EFFICIENCY

To understand the behaviour of the estimators it is useful to distinguish two groups of disturbances, namely, diffuse and localized (cf. Tatum (1997)). Diffuse disturbances are outliers that are spread over all of the samples whereas localized disturbances affect all observations in one sample. We include various types of estimators (both robust estimators and several estimation methods based on the principle of control charting) and compare them under various types of disturbances. The first subsection introduces the estimators, while the second subsection presents the MSE of the estimators.

Location Estimators: Recall that X_{ij}, $i = 1,2,\ldots,n,$ and $j = 1,2,\ldots,k$, denote the Phase I data. The X_{ij}'s are assumed to be independent and $N(\mu,\sigma^2)$ distributed. We denote by $X_{j(v)}$, $v = 1,2,\ldots,n$, the v^{th} order statistic in sample j.

The first estimator that we consider is the mean of the sample means,

$$\bar{\bar{X}} = \frac{1}{k}\sum_{j=1}^{k}\bar{X}_j = \frac{1}{k}\sum_{j=1}^{k}\left(\frac{1}{n}\sum_{i=1}^{n}X_{ij}\right) \tag{3.14}$$

This estimator is included to provide a basis for comparison, as it is the most efficient estimator for normally distributed data. However, it is well known that this estimator is not robust against outliers.

We also consider three robust estimators proposed earlier by Rocke (1989): the median of the sample means,

$$M(\bar{X}) = \text{median}(\bar{X}_1,\ \bar{X}_2,\ldots,\bar{X}_k) \tag{3.15}$$

And the mean of the sample medians,

$$\bar{M} = \frac{1}{k}\sum_{j=1}^{k} M_j \tag{3.16}$$

with M_j the median of sample j; and the trimmed mean of the sample means,

$$\bar{\bar{X}}_\alpha = \frac{1}{k-2\lceil k\alpha\rceil} \times \left(\sum_{v=\lceil k\alpha\rceil+1}^{k-\lceil k\alpha\rceil} \bar{X}_{(v)}\right) \tag{3.17}$$

where α denotes the percentage of samples to be trimmed at one side, $\lceil z\rceil$ denotes the ceiling function, i.e., the smallest integer not less than z, and $\bar{X}_{(v)}$ denotes the v^{th} ordered value of the sample means. Here, we consider the 20% trimmed mean, which trims the six smallest and the six largest sample means when $k = 30$. Of course, other trimming percentages could have been used. In fact, we have also used 10% and 25%, but the results with 20% are representative for this estimator.

Furthermore, our analysis includes the Hodges–Lehmann estimator (Hodges and Lehmann (1963)), an estimator based on the so-called Walsh averages. The $n(n+1)/2$ Walsh averages of sample j are:

$$W_{j,k,l} = (X_{j,k} + X_{j,l})/2$$

for $k = 1,2,\ldots,n$, $l = 1,2,\ldots,n$, and $k \leq l$. The Hodges–Lehmann estimate for sample j, denoted by HL_j, is defined as the median of the Walsh averages. Alloway and Raghavachari (1991) conducted a Monte Carlo simulation to determine whether the mean or the median of the sample Hodges–Lehmann estimates should be used to determine the final location estimate. They concluded that the mean of the sample values should be used,

$$\overline{HL} = \frac{1}{k}\sum_{j=1}^{k} HL_j \tag{3.18}$$

and that the resulting estimate is unbiased.

We also include the trimean statistic. The trimean of sample j is the weighted average of the sample median and the two other quartiles,

$$TM_j = \left(Q_{j,1} + 2Q_{j,2} + Q_{j,3}\right)/4,$$

where $Q_{j,q}$ is the q^{th} quartile of sample j, $q = 1,2,3$ (cf., Tukey(1997), Wang et al.(2007)). It also equals the average of the median and the midhinge $(1/2)[Q_{j,2} + (Q_{j,1} + Q_{j,3})/2]$ (cf. Weisberg (1992)). We use the following definitions for the quartiles: $Q_{j,1} = X_{j,(a)}$ and $Q_{j,3} = X_{j,(b)}$with $a = \lceil n/4 \rceil$ and $b = n - a + 1$. This means that $Q_{j,1}$ and $Q_{j,3}$ are defined as the second smallest and the second largest observations, respectively, for $4 \leq n \leq 7$, and as the third smallest and the third largest values, respectively, for $8 \leq n \leq 11$. Like the median and the midhinge, but unlike the sample mean, the trimean is a statistically resistant L-estimator (a linear combination of order statistics), with a breakdown point of 25% (see Wang et al. (2007)). According to Tukey (1977), using the trimean instead of the median gives a more useful assessment of location or centering. According to Weisberg (1992), the "statistical resistance" benefit of the trimean as a measure of the centre of a distribution is that it combines the median's emphasis on centre values with the midhinge's attention to the extremes. The trimean is almost as resistant to extreme scores as the median and is less subject to sampling fluctuations than the arithmetic mean in extremely skewed distributions. Asymptotic distributional results of the trimean can be found in Wang et al. (2007). The location estimate analysed below is the mean of the sample trimeans, i.e.,

$$\overline{TM}_j = \frac{1}{k} \sum_{j=1}^{k} TM_j \tag{3.19}$$

Finally, we consider a statistic that is expected to be robust against both diffuse and localized disturbances, namely, the trimmed mean of the sample trimeans, defined by

$$\overline{\overline{TM}}_{(\alpha)} = \frac{1}{k - 2\lceil k\alpha \rceil} \times \left(\sum_{v=\lceil k\alpha \rceil+1}^{k-\lceil k\alpha \rceil} \overline{TM}_{(v)} \right) \tag{3.20}$$

where $TM_{(v)}$ denotes the v^{th} ordered value of the sample trimeans. We consider the 20% trimmed trimean, which trims the six smallest and the six largest sample trimeans when $k = 30$.

The estimators outlined above are summarized in Table 3.2.

Table 3.2: Proposed Location Estimators

Estimator	Notation
Mean of sample means	$\bar{\bar{X}}$
Median of sample means	$M(\bar{X})$
Mean of sample medians	$\bar{M}$
20% trimmed mean of sample means	$\bar{\bar{X}}_{20}$
Mean of sample Hodges-Lehmann	$\overline{HL}$
Mean of sample trimeans	$\overline{TM}$
20% trimmed mean of sample trimeans	$\overline{TM}_{20}$

Efficiency of the Proposed Estimators: The efficiency of control charting procedures is often evaluated by comparing the variance of the respective location estimators. We use a procedure similar to what was adopted by Tatum (1997) and consider the MSE of the estimators. The MSE will be estimated as

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{\mu}^{i}-\mu}{\sigma}\right)^{2},$$

where $\hat{\mu}^i$ is the value of the unbiased estimate in the i^{th} simulation run and N is the number of simulation runs. We include the uncontaminated case, i.e. the situation where all X_{ij} are from the $N(0,1)$ distribution, as well as five types of disturbances (cf. Tatum (1997)):

1. A model for diffuse symmetric variance disturbances in which each observation has a 95% probability of being drawn from the $N(0,1)$ distribution and a 5% probability of being drawn from the $N(0,a)$ distribution, with $a = 1.5, 2.0, \ldots, 5.5, 6.0$.
2. A model for diffuse a symmetric variance disturbances in which each observation is drawn from the $N(0,1)$ distribution and has a 5% probability of having a multiple of a χ_1^2 variable added to it, with the multiplier equal to $0.5, 1.0, \ldots, 4.5, 5.0$.
3. A model for localized variance disturbances in which observations in 3 out of 30 samples are drawn from the $N(0,a)$ distribution, with $a = 1.5, 2.0, \ldots, 5.5, 6.0$.

4. A model for diffuse mean disturbances in which each observation has a 95% probability of being drawn from the $N(0,1)$ distribution and a 5% probability of being drawn from the $N(b,1)$ distribution, with $b = 0.5, 1.0, \ldots, 9.0, 9.5$.
5. A model for localized mean disturbances in which observations in 3 out of 30 samples are drawn from the $N(a,1)$ distribution, with $a = 0.5, 1.0, \ldots, 5.5, 6.0$.

The MSE of the estimators in form of figures are not given here and can be seen in Schoonhoven, Nazir et al. (2011). In Summary, the following findings from MSE can be observed. The most efficient estimator under normality is $\bar{\bar{X}}$, as was to be expected. The estimators $\overline{HL}$, $\bar{\bar{X}}_{20}$ and $\overline{TM}$ are slightly less efficient under normality, followed by $\overline{TM}_{20}$, $\bar{M}$, and $M(\bar{X})$. The reason behind the last finding is that they use less information. $\bar{M}$, $\overline{TM}$, and $\overline{TM}_{20}$ have the lowest MSE when there are diffuse disturbances. $\bar{M}$ and $\overline{TM}$ lose their efficiency advantage when contaminations take the form of localized mean or variance disturbances. In such situations, $\bar{\bar{X}}$, $\bar{\bar{X}}_{20}$, and $\overline{TM}_{20}$, which involve trimming the sample means, perform relatively well. It is worth noting that $\overline{TM}_{20}$ has the best performance over all because it is reasonably robust against all types of contaminations.

3.6 PROPOSED PHASE I CONTROL CHART LOCATION ESTIMATOR

In-control process parameters can be obtained not only via robust statistics but also via Phase I control charting. In this section, we consider the Phase I analysis, based on the principle of control charting in order to generate robust estimates of the process location. The outline of the method is given in section 3.2 and will be denoted as $ATM_{\overline{TM}'}$ (cf. Schoonhoven, Nazir et al. (2011). The next section presents the MSE of the proposed estimation method.

Efficiency of the Proposed Phase I Control Charts: To determine the efficiency of the proposed Phase I control chart, five types of contaminations defined in our MSE study of the statistics are considered and presented in the previous section. The MSE results for the Phase I control chart are given in Schoonhoven, Nazir et al. (2011). To facilitate comparison, we have also included the MSE of the estimators $\bar{\bar{X}}$ and $\overline{TM}_{20}$. The Phase I method, $ATM_{\overline{TM}'}$, which first screens for localized disturbances and then for occasional outliers, outperforms all estimation methods. The method is particularly powerful in the presence of diffuse disturbances because its use of an individual control chart in Phase I to identify single outliers. It increases the probability that such disturbances will be detected.

3.7 DERIVATION OF THE PHASE II CONTROL LIMITS

We now turn to the effect of the proposed location estimators on the $\bar{X}$ control chart performance in Phase II. The formulas for the $\bar{X}$ control limits with estimated limits are given by Equation (3.9). For the Phase II control limits, we only estimate the in control mean μ; we treat the in-control standard deviation σ as known because we want to isolate the effect of estimating the location parameter. The factor C_n that is used to obtain accurate control limits when the process parameters are estimated is derived such that the probability of a false signal equals the desired probability of a false signal. Except for the estimator $\bar{\bar{X}}$, C_n cannot be obtained easily in analytic form and is therefore obtained by means of simulation. The factors are chosen such that p is equal to 0.0027 under normality. Fifty thousand simulation runs are used. For $k = 30, n = 5$, and $n = 9$, the resulting factors are equal to 3.05 for $\bar{\bar{X}}$, and $ATM_{\overline{TM}'}$; 3.06 for $\bar{\bar{X}}_{20}$ and $\overline{TM}$ and 3.07 for $M(\bar{X})$, $\bar{M}$, $\overline{HL}$, and $\overline{TM}_{20}$.

3.8 CONTROL CHART PERFORMANCE

The effect on $\bar{X}$ Phase II performance of the proposed location statistics and estimation method based on Phase I control charting is evaluated. We consider the same

Phase I situations as those used to assess the MSE with a, b and the multiplier equal to 4 to simulate the contaminated case (cf. section 3.5).

Following Jensen et al. (2006), we use the unconditional run length distribution to assess the performance. Specifically, we look at several characteristics of that distribution, namely the average run length (ARL) and the standard deviation of the run length ($SDRL$). In addition, we also give the probability that one sample gives a signal (p). We compute these characteristics in an in-control and several out-of-control situations. We consider different shifts of size $\delta\sigma$ in the mean, setting δ equal to $0, 0.5, 1$ and 2. The performance characteristics are obtained by simulation.

Simulation Procedure: The performance characteristics p and ARL for estimated control limits are determined by averaging the conditional characteristics, i.e., the characteristics for a given set of estimated control limits, overall possible values of the control limits. Recall the definitions (when σ known) of $P(F_j|\hat{\mu})$ from Equation (3.10), $E(RL|\hat{\mu})$ from (3.11), $p = E(P(F_j|\hat{\mu})$ from (3.12), and $ARL = E(\frac{1}{P(F_j|\hat{\mu})})$ from (3.13). These expectations will be obtained by simulation: numerous datasets are generated and, for each dataset, $P(F_j|\hat{\mu})$ and $E(RL|\hat{\mu})$ are computed. By averaging these values, we obtain the unconditional values. The unconditional standard deviation is determined by

$$\mathrm{SDRL} = \sqrt{Var(RL)} = \sqrt{E\big(Var(RL|\hat{\mu})\big) + Var\big(E(RL|\hat{\mu})\big)}$$

$$= \left(2E(\frac{1}{P(F_j|\hat{\mu})})^2 - \left(E\frac{1}{P(F_j|\hat{\mu})}\right)^2 - E\frac{1}{P(F_j|\hat{\mu})}\right)^{1/2}$$

Enough replications of the above procedure were performed to obtain sufficiently small relative estimated standard errors for p and ARL. The relative estimated standard error is the estimated standard error of the estimate relative to the estimate. The relative standard error of the estimates is never higher than 0.60%.

Results: First, we consider the situation where the process follows a normal distribution and the Phase I data are not contaminated. We investigate the impact of the estimator used to estimate μ in Phase I. Table 3.3 presents the probability of one sample showing a signal (p) and the average run length (ARL) when the process mean equals $\mu + \delta\sigma$. When $\delta = 0$, the process is in control, so we want p to be as low as possible and ARL to be as high as possible. When $\delta \neq 0$, i.e. in the out-of-control situation, we want to achieve the opposite. We can see that, in the absence of any contamination (cf. Table 3.3), the efficiency of the estimators is very similar. We can therefore conclude that using a more robust location estimator does not have a substantial impact on the control chart performance in the uncontaminated situation.

Table 3.3: p, ARL and (in Parentheses) $SDRL$ of Corrected Limits under Normality $k = 30$

n	$\hat{\mu}$	p				$ARL(SDRL)$			
		$\delta = 0$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$\delta = 0$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$
5	$\bar{\bar{X}}$	0.0027	0.029	0.21	0.92	384(392)	41.7(49.4)	5.03(4.90)	1.09(0.32)
	$M(\bar{X})$	0.0027	0.028	0.21	0.91	390(406)	46.2(59.9)	5.31(5.43)	1.10(0.33)
	$\bar{M}$	0.0027	0.028	0.21	0.91	392(407)	45.9(59.0)	5.29(5.37)	1.10(0.33)
	$\bar{\bar{X}}_{20}$	0.0027	0.028	0.21	0.92	391(401)	43.3(52.4)	5.14(5.08)	1.09(0.32)
	$\overline{HL}$	0.0027	0.029	0.21	0.92	380(389)	42.0(50.4)	5.05(4.94)	1.09(0.32)
	$\overline{TM}$	0.0027	0.028	0.21	0.92	390(400)	43.4(53.0)	5.14(5.09)	1.09(0.32)
	$\overline{TM}_{20}$	0.0027	0.028	0.21	0.92	396(410)	45.3(56.9)	5.26(5.29)	1.09(0.33)
	$ATM_{\overline{TM}'}$	0.0027	0.029	0.21	0.92	381(390)	42.0(50.3)	5.06(4.96)	1.09(0.32)
9	$\bar{\bar{X}}$	0.0027	0.064	0.48	1.00	384(393)	17.9(20.0)	2.13(1.62)	1.00(0.043)
	$M(\bar{X})$	0.0027	0.063	0.47	1.00	390(405)	19.5(23.5)	2.19(1.74)	1.00(0.046)
	$\bar{M}$	0.0027	0.063	0.47	1.00	390(405)	19.5(23.6)	2.19(1.74)	1.00(0.046)
	$\bar{\bar{X}}_{20}$	0.0027	0.063	0.48	1.00	391(401)	18.5(21.1)	2.15(1.66)	1.00(0.044)
	$\overline{HL}$	0.0027	0.064	0.48	1.00	380(389)	18.0(20.4)	2.13(1.63)	1.00(0.043)
	$\overline{TM}$	0.0027	0.063	0.48	1.00	390(400)	18.6(21.4)	2.16(1.67)	1.00(0.045)
	$\overline{TM}_{20}$	0.0027	0.062	0.47	1.00	395(409)	19.3(22.7)	2.18(1.71)	1.00(0.046)
	$ATM_{\overline{TM}'}$	0.0027	0.064	0.48	1.00	380(390)	18.0(20.5)	2.13(1.64)	1.00(0.043)

The Phase II control charts based on the estimators $M(\bar{X})$, $\bar{\bar{X}}_{20}$ and $\overline{TM}_{20}$, perform relatively well when localized disturbances are present, while the charts based on $\bar{M}$, $\overline{HL}$, $\overline{TM}$, and $\overline{TM}_{20}$ perform relatively well when diffuse disturbances are present (cf. Tables 3.4–

3.8). The Phase II chart based on $ATM_{\overline{TM}'}$ performs best: this chart is as efficient as $\bar{\bar{X}}$ in the uncontaminated normal situation and its performance does not change much when there are contaminations. Moreover, the chart outperforms the other methods in all situations because it successfully filters out both diffuse and localized disturbances. In the presence of asymmetric disturbances, in particular, the added value of this estimation method is substantial.

Table 3.4: $\boldsymbol{p}$, $\boldsymbol{ARL}$ and (in Parentheses) $\boldsymbol{SDRL}$ under Diffuse Symmetric Variance $\boldsymbol{k = 30}$

n	$\hat{\mu}$	p				$ARL(SDRL)$			
		$\delta = 0$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$\delta = 0$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$
5	$\bar{\bar{X}}$	0.0030	0.030	0.21	0.92	358(375)	45.0(60.9)	5.21(5.44)	1.09(0.33)
	$M(\bar{X})$	0.0029	0.029	0.21	0.91	375(395)	48.4(67.7)	5.41(5.77)	1.10(0.34)
	$\bar{M}$	0.0028	0.029	0.21	0.91	387(403)	46.5(61.5)	5.34(5.52)	1.10(0.33)
	$\bar{\bar{X}}_{20}$	0.0028	0.029	0.21	0.92	376(392)	45.1(58.8)	5.23(5.35)	1.09(0.33)
	$\overline{HL}$	0.0029	0.029	0.21	0.92	370(383)	43.2(54.5)	5.13(5.15)	1.09(0.32)
	$\overline{TM}$	0.0027	0.029	0.21	0.92	384(396)	44.2(55.6)	5.19(5.21)	1.09(0.32)
	$\overline{TM}_{20}$	0.0027	0.028	0.21	0.91	390(405)	46.0(59.2)	5.29(5.38)	1.10(0.33)
	$ATM_{\overline{TM}'}$	0.0028	0.029	0.21	0.92	375(386)	42.7(52.8)	5.09(5.06)	1.09(0.32)
9	$\bar{\bar{X}}$	0.0030	0.066	0.48	1.00	358(375)	19.1(24.0)	2.17(1.73)	1.00(0.045)
	$M(\bar{X})$	0.0030	0.065	0.47	1.00	371(393)	20.6(27.5)	2.23(1.83)	1.00(0.048)
	$\bar{M}$	0.0028	0.063	0.47	1.00	386(403)	19.8(24.5)	2.20(1.75)	1.00(0.047)
	$\bar{\bar{X}}_{20}$	0.0029	0.064	0.48	1.00	373(389)	19.4(24.1)	2.18(1.74)	1.00(0.046)
	$\overline{HL}$	0.0028	0.064	0.48	1.00	373(385)	18.4(21.5)	2.14(1.66)	1.00(0.044)
	$\overline{TM}$	0.0027	0.063	0.48	1.00	384(397)	18.8(22.1)	2.16(1.69)	1.00(0.045)
	$\overline{TM}_{20}$	0.0027	0.063	0.47	1.00	391(406)	19.4(23.3)	2.19(1.74)	1.00(0.046)
		0.0028	0.064	0.48	1.00	375(386)	18.3(21.3)	2.14(1.65)	1.00(0.044)

When localized mean disturbances are present, we see a strange phenomenon for the $\bar{X}$, $\bar{M}$, $\overline{HL}$, and $\overline{TM}$ charts: the in-control ARL is lower than the out of-control ARL for $\delta = 0.5$. In other words, these charts are more likely to give a signal in the in-control situation than in the out-of-control situation for $\delta = 0.5$ and hence, in the presence of disturbances, are highly ARL-biased (cf. Jensen et al. (2006)).

Table 3.5: p, ARL and (in Parentheses) $SDRL$ under Diffuse Asymmetric Variance $k = 30$

n	$\hat{\mu}$	p				$ARL(SDRL)$			
		$\delta = 0$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$\delta = 0$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$
5	$\bar{\bar{X}}$	0.0076	0.012	0.12	0.82	233(295)	143(210)	12.8(24.0)	1.23(0.60)
	$M(\bar{X})$	0.0034	0.019	0.16	0.88	347(378)	77.0(110)	7.34(8.26)	1.14(0.41)
	$\bar{M}$	0.0029	0.022	0.18	0.9	374(395)	61.6(82.4)	6.32(6.73)	1.12(0.37)
	$\bar{\bar{X}}_{20}$	0.0034	0.018	0.16	0.88	337(366)	75.9(103)	7.22(7.91)	1.14(0.41)
	$\overline{HL}$	0.0033	0.02	0.17	0.89	340(363)	67.3(90.4)	6.73(7.36)	1.13(0.39)
	$\overline{TM}$	0.003	0.021	0.17	0.89	365(384)	61.2(69.2)	6.33(6.61)	1.12(0.37)
	$\overline{TM}_{20}$	0.0029	0.022	0.18	0.9	379(398)	60.4(78.7)	6.26(6.56)	1.12(0.37)
	$ATM_{\overline{TM}'}$	0.0028	0.025	0.2	0.91	373(385)	48.9(60.2)	5.53(5.55)	1.10(0.34)
9	$\bar{\bar{X}}$	0.011	0.02	0.27	0.99	175(239)	89.5(148)	4.58(6.42)	1.01(0.12)
	$M(\bar{X})$	0.0044	0.035	0.36	0.99	299(346)	42.1(63.4)	3.04(2.89)	1.01(0.077)
	$\bar{M}$	0.0031	0.048	0.42	1.00	366(389)	26.8(34.1)	2.51(2.12)	1.00(0.058)
	$\bar{\bar{X}}_{20}$	0.0047	0.033	0.35	0.99	280(324)	42.7(61.0)	3.09(2.89)	1.01(0.078)
	$\overline{HL}$	0.0033	0.044	0.41	1.00	336(358)	27.9(33.9)	2.57(2.15)	1.00(0.059)
	$\overline{TM}$	0.0031	0.046	0.42	1.00	356(378)	26.4(31.9)	2.51(2.08)	1.00(0.057)
	$\overline{TM}_{20}$	0.003	0.047	0.42	1.00	368(391)	26.8(33.2)	2.52(2.11)	1.00(0.058)
	$ATM_{\overline{TM}'}$	0.0028	0.055	0.45	1.00	370(383)	21.4(24.9)	2.30(1.83)	1.00(0.050)

Table 3.6: p, ARL and (in Parentheses) $SDRL$ under Diffuse Localized Variance $k = 30$

n	$\hat{\mu}$	p				$ARL(SDRL)$			
		$\delta = 0$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$\delta = 0$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$
5	$\bar{\bar{X}}$	0.0034	0.032	0.22	0.91	337(361)	48.7(73.0)	5.42(6.09)	1.10(0.34)
	$M(\bar{X})$	0.0028	0.029	0.21	0.91	382(400)	47.4(64.2)	5.40(5.66)	1.10(0.33)
	$\bar{M}$	0.0037	0.033	0.22	0.91	335(372)	57.2(98.5)	5.90(7.45)	1.11(0.36)
	$\bar{\bar{X}}_{20}$	0.0028	0.029	0.21	0.92	382(395)	44.3(55.8)	5.21(5.25)	1.09(0.32)
	$\overline{HL}$	0.0035	0.032	0.22	0.91	332(358)	50.0(76.8)	5.46(6.22)	1.10(0.34)
	$\overline{TM}$	0.0035	0.032	0.22	0.91	338(368)	52.3(83.4)	5.62(6.63)	1.10(0.35)
	$\overline{TM}_{20}$	0.0028	0.029	0.21	0.91	387(403)	46.6(61.7)	5.36(5.55)	1.10(0.33)
	$ATM_{\overline{TM}'}$	0.0028	0.029	0.21	0.92	372(384)	43.0(53.7)	5.11(5.10)	1.09(0.32)
9	$\bar{\bar{X}}$	0.0034	0.068	0.48	1.00	337(362)	20.4(28.6)	2.21(1.84)	1.00(0.048)
	$M(\bar{X})$	0.0028	0.064	0.47	1.00	381(400)	19.9(25.0)	2.21(1.78)	1.00(0.047)
	$\bar{M}$	0.0038	0.069	0.47	1.00	332(370)	24.0(40.4)	2.32(2.12)	1.00(0.054)
	$\bar{\bar{X}}_{20}$	0.0028	0.064	0.48	1.00	382(395)	18.9(22.3)	2.17(1.69)	1.00(0.045)
	$\overline{HL}$	0.0035	0.069	0.48	1.00	333(359)	20.8(30.1)	2.22(1.88)	1.00(0.049)
	$\overline{TM}$	0.0035	0.068	0.48	1.00	338(368)	21.8(32.7)	2.25(1.94)	1.00(0.051)
	$\overline{TM}_{20}$	0.0028	0.063	0.47	1.00	385(402)	19.7(24.2)	2.20(1.76)	1.00(0.047)
	$ATM_{\overline{TM}'}$	0.0029	0.065	0.48	1.00	368(381)	18.6(22.2)	2.15(1.68)	1.00(0.045)

Table 3.7: p, ARL and (in Parentheses) $SDRL$ under Diffuse Mean $k = 30$

n	$\hat{\mu}$	p				$ARL(SDRL)$			
		$\delta = 0$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$\delta = 0$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$
5	$\bar{\bar{X}}$	0.0061	0.011	0.11	0.83	224(271)	137(182)	10.9(12.9)	1.22(0.53)
	$M(\bar{X})$	0.0048	0.014	0.13	0.85	289(340)	115(168)	9.57(12.2)	1.19(0.49)
	$\bar{M}$	0.0033	0.019	0.16	0.88	351(380)	74.6(103)	7.12(7.89)	1.14(0.41)
	$\bar{\bar{X}}_{20}$	0.0049	0.013	0.13	0.85	271(316)	115(158)	9.52(11.2)	1.19(0.49)
	$\overline{HL}$	0.0042	0.015	0.14	0.86	290(326)	93.0(126)	8.22(9.18)	1.16(0.45)
	$\overline{TM}$	0.0035	0.017	0.15	0.88	333(361)	77.8(103)	7.33(7.95)	1.14(0.41)
	$\overline{TM}_{20}$	0.0032	0.019	0.16	0.88	356(383)	72.5(97.0)	7.00(7.57)	1.13(0.40)
	$ATM_{\overline{TM}'}$	0.0031	0.024	0.19	0.90	356(374)	57.0(78.1)	6.01(6.39)	1.11(0.36)
9	$\bar{\bar{X}}$	0.0089	0.018	0.26	0.99	161(208)	77.3(107)	4.17(4.16)	1.01(0.11)
	$M(\bar{X})$	0.0074	0.023	0.29	0.99	212(274)	71.7(111)	3.96(4.17)	1.01(0.10)
	$\bar{M}$	0.0035	0.041	0.39	1.00	339(371)	32.4(42.7)	2.72(2.39)	1.00(0.065)
	$\bar{\bar{X}}_{20}$	0.0075	0.021	0.28	0.99	193(246)	68.9(98.5)	3.91(3.90)	1.01(0.10)
	$\overline{HL}$	0.0046	0.033	0.35	0.99	272(310)	39.6(51.4)	3.00(2.69)	1.01(0.074)
	$\overline{TM}$	0.0037	0.038	0.38	1.00	317(349)	33.6(42.7)	2.79(2.43)	1.00(0.067)
	$\overline{TM}_{20}$	0.0035	0.039	0.38	1.00	336(368)	32.9(42.0)	2.75(2.40)	1.00(0.066)
	$ATM_{\overline{TM}'}$	0.0031	0.051	0.43	1.00	352(371)	24.5(30.7)	2.42(2.00)	1.00(0.054)

Table 3.8: p, ARL and (in Parentheses) $SDRL$ under Localized Mean $k = 30$

n	$\hat{\mu}$	p				$ARL(SDRL)$			
		$\delta = 0$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$\delta = 0$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$
5	$\bar{\bar{X}}$	0.017	0.0034	0.046	0.70	72.3(87.5)	329(351)	25.0(28.8)	1.45(0.83)
	$M(\bar{X})$	0.0031	0.021	0.17	0.89	366(389)	66.0(89.3)	6.62(7.15)	1.13(0.38)
	$\bar{M}$	0.017	0.0033	0.045	0.69	80.1(105)	343(372)	27.3(33.6)	1.47(0.86)
	$\bar{\bar{X}}_{20}$	0.003	0.02	0.17	0.89	360(379)	64.8(81.8)	6.57(6.81)	1.13(0.38)
	$\overline{HL}$	0.017	0.0034	0.047	0.70	72.8(89.3)	327(350)	25.2(29.2)	1.45(0.83)
	$\overline{TM}$	0.017	0.0033	0.046	0.69	75.6(94.4)	337(362)	26.0(30.7)	1.46(0.84)
	$\overline{TM}_{20}$	0.0031	0.019	0.16	0.89	360(385)	70.4(92.0)	6.89(7.32)	1.13(0.40)
	$ATM_{\overline{TM}'}$	0.0028	0.029	0.21	0.92	375(386)	43.4(53.4)	5.14(5.11)	1.09(0.32)
9	$\bar{\bar{X}}$	0.035	0.0039	0.11	0.96	34.3(30.7)	293(321)	10.1(10.7)	1.04(0.22)
	$M(\bar{X})$	0.0031	0.048	0.42	1.00	366(389)	26.8(34.3)	2.51(2.12)	1.00(0.058)
	$\bar{M}$	0.034	0.0039	0.11	0.95	37.9(48.5)	308(346)	10.8(12.3)	1.05(0.23)
	$\bar{\bar{X}}_{20}$	0.003	0.046	0.41	1.00	359(379)	26.3(31.5)	2.51(2.07)	1.00(0.057)
	$\overline{HL}$	0.035	0.0039	0.11	0.96	34.5(40.8)	293(322)	10.1(10.9)	1.05(0.22)
	$\overline{TM}$	0.034	0.0039	0.11	0.96	35.7(43.1)	301(333)	10.4(11.4)	1.05(0.22)
	$\overline{TM}_{20}$	0.0031	0.044	0.41	1.00	361(385)	28.4(35.4)	2.58(2.18)	1.00(0.060)
	$ATM_{\overline{TM}'}$	0.0028	0.063	0.48	1.00	376(386)	18.6(21.5)	2.16(1.67)	1.00(0.044)

3.9 A DISCUSSION ON ROBUST LOCATION AND DISPERSION CHARTS

Grove (2013) pointed out that the estimation methods can further be improved by tightening the Phase I limits developed by Schoonhoven, Nazir et al. (2011) for the location control charts and by Schoonhoven and Does (2011) for dispersion control charts. He suggested the use of an estimate of σ based on the control limit corresponding to a cumulative probability of the null distribution of S/c_4 and also he indicated the application of the median of subgroup means as the centre line of $\bar{X}$ chart to get a more robust version of this procedure. Schoonhoven and Does (2013) agreed on the points mentioned by Grove (2013) and concluded that setting Phase I limits too tight would result in a decrease of efficiency for situations when the data are clean because then too many observations are deleted. Currently, there is no algorithm enabling practitioners to set the limits in this way and they recommended providing some guidance for practitioners that are needed on how to set these limits without too much effort.

In Chapter 2 (dispersion control charts) and this chapter (location control charts), robust techniques are applied to obtain robust Phase I limits that are in line with the points mentioned by Schoonhoven, Nazir et al. (2011) and Grove (2013). These chapters develop robust control charts along with stepwise approaches for setting up control charts for practitioners to monitor process parameters.

3.10 CONCLUDING REMARKS

A practical procedure is given to obtain a robust estimate for the mean, based on a Phase I analysis. In Phase I, the initial estimate of μ is based on the trimmed mean of the sample trimeans. This estimator is robust against both localized and diffuse disturbances so that the limits of the Phase I control chart are not affected by potential disturbances. Several Phase I estimators of the location parameter are considered in establishing Phase II control limits of $\bar{\bar{X}}$ charts. The MSE of the estimators has been assessed under various circumstances: the uncontaminated situation and various situations contaminated with diffuse symmetric and asymmetric variance disturbances, localized variance disturbances, diffuse mean disturbances, and localized mean disturbances. Moreover, the effect of the location estimator is investigated on the $\bar{X}$ Phase II control chart performance when the methods are used to determine the Phase II limits. The standard methods suffer from a number of problems. Estimators that are based on the principle of trimming individual observations (e.g. $\overline{M}$ and $\overline{TM}$) perform reasonably well when there are diffuse disturbances but not when localized disturbances are present. In the latter situation, estimators that are based on the principle of trimming samples (e.g., $M(\bar{X})$ and $\bar{\bar{X}}_{20}$) are efficient. All of these methods are biased when there are asymmetric disturbances, as the trimming principle does not take into account the asymmetry of the disturbance. The proposed new type of Phase I analysis addresses the problems encountered in the standard Phase I analysis. The initial estimate of μ for the Phase I control chart is based on a trimmed version of the trimean, namely, $\overline{TM}_{20}$, and a subsequent procedure for both sample screening and outlier screening (resulting in $ATM_{\overline{TM}'}$). The proposed method is efficient under normality and outperforms the existing methods when disturbances are present. Consequently, $ATM_{\overline{TM}'}$ is a very effective method for estimating μ for the $\bar{X}$ Phase II chart limits.

Chapter 4

Robust CUSUM Control Charts for Spread

Process monitoring through control charts is a quite popular practice in statistical process control. From a statistical point of view a superior control chart is one which has an efficient design structure, but having resistance against unusual situations is of more practical importance. In order to have a compromise between the statistical and practical purposes a natural desire is to have a control chart which can serve both purposes simultaneously in a good capacity. This chapter is planned for the same objective focusing on monitoring the dispersion parameter by using a Cumulative Sum (CUSUM) control chart scheme. We investigate the properties of the design structure of different control charts based on some already existing estimators as well as some new robust dispersion estimators. By evaluating the performance of these estimators based CUSUM control charts in terms of average run length (ARL), we identify those charts which are more capable to make a good compromise between the above mentioned purposes in terms of statistical and practical needs. This chapter is based on the papers of Nazir, Riaz and Does (2013) and Abbasi, Riaz et al. (2014).

4.1 INTRODUCTION

Existing control charts are usually designed under the assumptions of normality and outlier free environments in the quality characteristic of concern. Normality seems more of a theoretical value and it is generally hard to find practical situations where the normality assumption is easily fulfilled. In case of violation of the normality assumption and the presence of outliers, these commonly used charts lose their efficiency and performance ability and hence are of less practical use. In general, robust control charts are preferred and are of more practical use when the design structure is not affected by the violation of above mentioned ideal assumptions.

The choice of the control charts to be used depends on the characteristics to be measured in the process and what type of amount of change / shift has to be determined. Control charts are classified into two categories, namely memoryless control charts and memory control charts. Shewhart-type control charts are termed as memoryless control charts and their main deficiency is that they are less sensitive to small and moderate shifts in the parameters (location and dispersion). The commonly used memory control charts in the literature include Cumulative Sum (CUSUM) control charts (cf. Page (1954)) and Exponentially Weighted Moving Average (EWMA) control charts (cf. Roberts (1959)). These memory control charts are designed such that they use the past information along with the current information which makes them very sensitive to small and moderate shifts in the process parameters.

Jensen et al. (2006) suggested that more data in Phase I are needed than typically is recommended in order to achieve a performance comparable with the known parameters cases. In particular, for the CUSUM chart, the number of preliminary samples should be in the hundred scales rather than the dozen scales as used by the Shewhart chart (cf. Hawkins and Olwell (1998)). For example, Quesenberry (1993) recommended that at least 100 samples of size five should be used in Phase I for the CUSUM chart. That is because the CUSUM chart is sensitive to small shifts and any random error in the estimated parameter will tend to cause deviated in-control and out-of-control performance.

In this chapter we concentrate on robust control charts for the process dispersion parameter in Phase II. For the related problem for the location parameter the reader is referred to chapter 2 for Shewhart-type charts, chapter 5 for the CUSUM charts and Zwetsloot, et al.

(2014a) for the EWMA charts. For the dispersion parameter the reader is referred to Schoonhoven and Does (2012) for Shewhart-type charts.

The CUSUM control chart has received a great deal of attention in the quality control literature due to its simplicity and efficiency. It has been primarily used as a tool for monitoring process mean levels. The theoretical properties of the CUSUM chart for monitoring the process mean have been thoroughly investigated (cf. Siegmund (1985), and Moustakides (1986)). In contrast, the CUSUM chart as a tool for monitoring process variability has received less attention and investigation. Some published properties are found in Page (1963), Chang and Gan (1995), Hawkins and Olwell (1998), Acosta-Mejia (1998), Acosta-Mejia et al. (1999), and Acosta-Mejia and Pignatiello (2000). Note that the corresponding EWMA control chart for dispersion is one of the subjects of a PhD project at the University of Amsterdam (cf. Zwetsloot et al. (2014b)).

In this chapter, we compare a number of estimators that have been presented in the literature. Some of them are not common in the control charts literature. We derive the charts factors that determine the control limits. The performance of the charts based on these estimators is evaluated by assessing the average run length under normality and in the presence of various types of contaminations by means of simulation.

Thus, the present chapter focuses on robust CUSUM control charts for monitoring the process dispersion parameter. Particularly, their design structures and performances are studied under different parent environments and in the presence of special causes in the dispersion parameter of the process. The motivation and inspiration of this study is taken from Schoonhoven, Nazir et al. (2011) and Abbasi and Miller (2012). Before moving on towards the basic structure of the CUSUM chart, we provide a description of the robust estimators in the next section. Then we evaluate the performance of the different CUSUM charts by means of the average run length. Finally, we describe our main conclusions.

4.2 DESCRIPTION OF ESTIMATORS OF PROCESS DISPERSION

Let θ be the process dispersion parameter which needs to be monitored through control charting and $\hat{\theta}$ be its estimator based on a sample of size n. There are many choices for $\hat{\theta}$. David (1998) provided a brief history of standard deviation estimators. The traditional estimators are the pooled sample standard deviation, the mean of the sample standard deviations, and the mean of the sample ranges. Mahmoud et al. (2010) studied the relative efficiencies of these estimators for different sample sizes n and number of samples k. Schoonhoven, Nazir, et al. (2011) considered different estimators of the population standard deviation and provided a comprehensive analysis on their efficiency and use in control charts for different phases.

In deriving the estimates of the population dispersion parameter, we will look at some of the estimators discussed in Schoonhoven, Nazir, et al. (2011) as well as some other robust estimators which are not common in the control charts literature. Below we give a short description of the estimators used in this chapter.

The first estimator of the population dispersion θ (which will also be used as a reference estimator throughout the rest of article) is the sample standard deviation S which is defined by:

$$S = \left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^{1/2} \tag{4.1}$$

where X_i denotes the i^{th} observation in a sample of size n and $\bar{X}$ denotes the corresponding mean of the sample. The sample standard deviation S is the most efficient estimator in normally distributed environments, but studies have shown that it is highly affected by the presence of outliers and special causes. The breakdown point (proportion of the outlying observations that an estimator can cope) of the sample standard deviation is zero.

The next estimator is based on sample interquartile range (IQR) and is defined as:

$$IQR = \frac{Q_3 - Q_1}{1.34898}, \tag{4.2}$$

where Q_3 and Q_1 are the third and the first quartiles of the sample, respectively. Different properties of the sample interquartile range related to efficiency can be seen from Riaz (2008). The sample interquartile range is more robust to departures from normality and outliers than the sample standard deviation (cf. Riaz (2008)). The breakdown point of IQR is 25%.

We also take into account an estimator proposed by Gini (1912), which is known as Gini's mean differences G and can be written as

$$G = \frac{4}{n-1} \sum_{i=1}^{n} \left(\frac{2i-n-1}{2n} \right) X_{(i)} \tag{4.3}$$

, where $X_{(i)}$ is the i^{th} order statistic of the sample. Gini's estimator is highly efficient and is more robust to outliers than the estimators based on the range and standard deviation (cf. David (1968) and Montanari and Monari (2005)). Two similar estimators named as Downton's estimator ($D = \frac{2\sqrt{\pi}}{n(n-1)} \sum_{i=1}^{n} \left(i - \frac{n+1}{2} \right) X_{(i)}$) and the probability weighted moments based estimator ($S_{pw} = \frac{\sqrt{\pi}}{n^2} \sum_{i=1}^{n} (2i - n - 1) X_{(i)}$) are used by Khoo (2004) and Muhammad and Riaz (2006), respectively, with the control structure of Shewhart's charts. The properties of those estimators are found to be similar to Gini's estimator, because the three estimators are proportional to each other.

We also consider a robust estimator proposed by Hampel (1974). His robust estimator is based on the median of the absolute deviations from the median defined as:

$$MADM = 1.4826 * median \left| X_i - \tilde{X} \right| \tag{4.4}$$

where $\tilde{X}$ is the sample median. This estimator is very robust against outliers but its efficiency under normality is very low (i.e. only 37%). There are some more estimators based on the absolute deviations, i.e. the mean of the absolute deviations from the mean ($MD = \frac{\sum_{i=1}^{n}|X_i - \bar{X}|/n}{1.2533}$), the median of the absolute deviations from the mean ($MAD = \text{median}|X_i - \bar{X}|$) and the mean of the absolute deviations from the median ($AADM = \sum_{i=1}^{n}|X_i - \tilde{X}|/n$). Wu et al. (2002) showed that the $MADM$ estimator performs the best, as compared to the other 3 estimators based on absolute deviations, in case of contaminated environments. The breakdown point of $MADM$ estimator is 50%. The gross error sensitivity (which measures the worst influence on the value of the estimator that a small amount of contamination of fixed size can have) of $MADM$ is equal to 1.167, which is the smallest value that one can obtain for any scale estimator in the case of the normal distribution.

Rousseeuw and Croux (1993) proposed different robust estimators of the population dispersion parameter θ, which are highly robust against outliers and their efficiencies under normality is higher compared to the estimator proposed by Hampel (1974). One of the estimators of Rousseeuw and Croux (1993) is defined as:

$$T_n = 1.38 * \frac{1}{h}\sum_{k=1}^{h}\{median|X_i - X_l|; i \neq l\}_{(k)} \tag{4.5}$$

This reads as follows: for each i we compute the median of $|X_i - X_l|, l = 1,2, \dots, n$. This yields n values, the average of first h order statistics gives the final estimate T_n , where $h = [n/2] + 1$, is roughly half the number of observations (the symbol [.] represents the integer part of a fraction). Rousseeuw and Croux (1993) also proposed two other robust estimators (similar to T_n), i.e. $S_n = 1.1926 * median_i\{median_l|X_i - X_l|; i \neq l\}$ and $Q_n = 2.2219 * \{|X_i - X_l|; i < l\}_{(p)}$ with $p = \frac{h!}{2!(h-2)!}$. The breakdown point of the T_n estimator is 50%. T_n is very robust against outliers and its efficiency under normality is 52%. The gross error sensitivity of T_n is 1.4688. The reason of choosing T_n is its low gross error sensitivity as compared to S_n and Q_n.

We also evaluate an estimator of the population dispersion parameter mentioned by Shamos (1976) and Bickel and Lehmann (1979). This estimator is obtained by replacing pairwise averages by pairwise distances, and is defined as

$$B_n = 1.0483 * median\{|X_i - X_l|; i < l\} \tag{4.6}$$

This robust estimator has an efficiency of 86% under normality, but is less robust compared to the estimators proposed by Rousseeuw and Croux (1993) and its breakdown point is only 29%.

The last estimator we use in this study is based on order statistics of certain subranges proposed by Croux and Rousseeuw (1992) having a breakdown point of 50% and is defined as:

$$S_r = 1.4826 * \left|X_{(i+[0.25n]+1)} - X_{(i)}\right|_{\left(\left[\frac{n}{2}\right]-[0.25n]\right)} \tag{4.7}$$

S_r is a very robust estimator in the presence of outliers and its different features can be found in Croux and Rousseeuw (1992). Its efficiency under normality is only 37%, however, it is more efficient than $MADM$ for small samples.

Further description and different properties (e.g. efficiency and robustness) of these estimators can be seen from Rousseeuw and Croux (1993), David (1998), Mahmoud et al. (2010), Schoonhoven, Nazir et al. (2011), and Abbasi and Miller (2012).

It is anticipated that some robust estimators will perform well in an uncontaminated environment as well as in the presence of outliers and under special cause environments as the aim of these robust estimators is to estimate the population dispersion parameter efficiently and provide resistance against outliers and special cause environments.

Efficiency of the Estimators: For comparison purposes and to evaluate the accuracy of the dispersion estimators used in this study, we compute the standardized variances of the estimators as suggested by Rousseeuw and Croux (1993) and relative efficiencies of the estimators as used by Abbasi and Miller (2012).

Table 4.1: Standardized variance of Dispersion Estimators under Different Environments

Environments	Sample size	Estimators						
		G	IQR	S	$MADM$	B_n	T_n	S_r
$N(0,1)$	4	0.7250	0.7250	0.7171	1.3092	0.7724	1.2804	1.3092
	5	0.6702	0.7028	0.6598	1.7131	0.8956	1.0250	1.6461
	9	0.5902	0.9259	0.5799	1.5378	0.7408	1.0647	1.3332
$1\%CNormal$	4	0.8257	0.8257	0.8372	1.3358	0.8640	1.3077	1.3358
	5	0.7771	0.7857	0.8007	1.7377	0.9407	1.0575	1.6749
	9	0.6902	0.9308	0.7521	1.5468	0.7681	1.0791	1.3491
$5\%CNormal$	4	1.1174	1.1174	1.1784	1.4268	1.1292	1.3998	1.4268
	5	1.0659	1.0154	1.1651	1.8343	1.0834	1.1510	1.7882
	9	0.9958	1.0024	1.2246	1.5910	0.8961	1.1503	1.4190
$10\%CNormal$	4	1.3096	1.3096	1.3876	1.5758	1.3099	1.5500	1.5758
	5	1.2622	1.1855	1.3872	1.9584	1.2583	1.2867	1.9254
	9	1.1986	1.1338	1.4710	1.6683	1.0544	1.2391	1.5196
$G(1,1)$	4	1.5654	1.5654	1.6818	2.2602	1.6045	2.3068	2.2602
	5	1.5031	1.4771	1.6719	2.6164	1.8703	1.8482	2.5363
	9	1.4142	1.7224	1.7064	2.3855	1.6212	1.9844	2.2373
T_4	4	1.3564	1.3564	1.4559	1.6076	1.3670	1.5731	1.6076
	5	1.2898	1.2133	1.4401	2.0256	1.2645	1.3157	1.9658
	9	1.2494	1.2102	1.6207	1.7941	1.1040	1.3254	1.6316
$Logis(0,1)$	4	0.8972	0.8972	0.9018	1.4320	0.9178	1.3880	1.4320
	5	0.8604	0.8636	0.8735	1.8761	1.0491	1.1583	1.8221
	9	0.7786	1.0588	0.8129	1.6788	0.9028	1.1995	1.4934

The standardized variance (SV) of dispersion estimator $\hat{\theta}$ is calculated as:

$$SV_{\hat{\theta}} = \frac{nVAR(\hat{\theta})}{[E(\hat{\theta})]^2} \tag{4.8}$$

The denominator of $SV_{\hat{\theta}}$ is needed to obtain a natural measure of the accuracy of a scale estimator (cf. Bickel and Lehmann (1976)). The relative efficiency (RE) of the estimator is computed as:

$$RE_{\hat{\theta}} = \frac{min(SV_{\hat{\theta}})}{SV_{\hat{\theta}}} \tag{4.9}$$

SV and RE are computed by generating 10^5 samples of sizes $n = 4, 5$ and 9 under the following environments: uncontaminated normal, contaminated normal, gamma, Student's t and logistic environments.

Table 4.2: Relative Efficiencies of Dispersion Estimators under Different Environments

Environments	Sample size	Estimators						
		G	IQR	S	$MADM$	B_n	T_n	S_r
$N(0,1)$	4	98.9165	98.9165	100.0000	54.7773	92.8511	56.0095	54.7773
	5	98.4566	93.8921	100.0000	38.5162	73.6775	64.3705	40.0835
	9	98.2451	62.6240	100.0000	37.7069	78.2786	54.4628	43.4951
$1\% CNormal$	4	100.0000	100.0000	98.6331	61.8148	95.5718	63.1449	61.8148
	5	100.0000	98.9111	97.0528	44.7188	82.6053	73.4821	46.3963
	9	100.0000	74.1547	91.7771	44.6220	89.8658	63.9606	51.1607
$5\% CNormal$	4	100.0000	100.0000	94.8265	78.3189	98.9554	79.8274	78.3189
	5	95.2590	100.0000	87.1455	55.3539	93.7169	88.2139	56.7802
	9	89.9808	89.3916	73.1731	56.3196	100.0000	77.8978	63.1453
$10\% CNormal$	4	100.0000	100.0000	94.3793	83.1068	99.9806	84.4896	83.1068
	5	93.9279	100.0000	85.4598	60.5369	94.2173	92.1382	61.5738
	9	87.9711	92.9959	71.6800	63.2022	100.0000	85.0965	69.3864
$G(1,1)$	4	100.0000	100.0000	93.0762	69.2590	97.5629	67.8590	69.2590
	5	98.2688	100.0000	88.3484	56.4567	78.9771	79.9242	58.2383
	9	100.0000	82.1044	82.8764	59.2821	87.2285	71.2664	63.2101
T_4	4	100.0000	100.0000	93.1645	84.3743	99.2279	86.2255	84.3743
	5	94.0692	100.0000	84.2510	59.8976	95.9523	92.2180	61.7203
	9	88.3615	91.2249	68.1185	61.5369	100.0000	83.2988	67.6655
$Logis(0,1)$	4	100.0000	100.0000	99.4937	62.6551	97.7565	64.6437	62.6551
	5	100.0000	99.6309	98.5057	45.8615	82.0180	74.2831	47.2222
	9	100.0000	73.5346	95.7754	46.3757	86.2379	64.9107	52.1343

The description of these environments is given in section 4.4. SV and RE are given in Tables 4.1 and 4.2 which read for example under uncontaminated normal environment as: the dispersion estimator S as was expected has the lowest SV and the $MADM$ has the largest SV. The efficiency of the other estimators falls within these two (S and $MADM$) estimators. Under 5% and 10% symmetric variance contaminated normal models, B_n has the lowest SV and for small 1% contamination, G obtains the lowest SV. For most of the non-normal environments, G and IQR are efficient estimators as compared to all other estimators. The dispersion estimator S is extremely affected by contaminations and non-normal environments.

4.3 THE PROPOSED CUSUM CHARTS SCHEME FOR PROCESS DISPERSION

For the CUSUM procedures, we assume that we want to detect an increase in the process dispersion parameter θ. Let $\hat{\theta}$ be any estimator from Section 4.2 of the process dispersion parameter θ from a random sample of size n which are taken from a continuous production process at regular intervals. The rule for the CUSUM - $\hat{\theta}$charts is as follows:

$$Z_t = \max\left[0,\ \left(\hat{\theta} - K_{\hat{\theta}}\right) + Z_{t-1}\right], \quad t = 1, 2, 3, \ldots, \tag{4.10}$$

where $Z_0 = 0$ according to Tuprah and Ncube (1987)) and $K_{\hat{\theta}}$ is the reference value for the scheme. Z_t is plotted against the sample number t. If $Z_t > H_{\hat{\theta}}$ (where $H_{\hat{\theta}}$ determines the decision interval) for any value of t, the process is deemed to be out of control and it is concluded that the process dispersion has increased. The sample number at which $Z_t > H_{\hat{\theta}}$ is the run length of the process and the expected value of the random variable run length is the average run length (ARL) of the scheme. The values of $K_{\hat{\theta}}$ are chosen in such a way that a shift in the process dispersion parameter is detected quickly. The values of $H_{\hat{\theta}}$ are chosen for a fixed value of ARL along with value of $K_{\hat{\theta}}$, when the process is in control under all environments considered in this study and it is expected that the ARL will be small, when the process is out of control. The reference value $K_{\hat{\theta}}$ will be based on Ewan and Kemp (1960), Page (1963) and Tuprah and Ncube (1987), so the value $K_{\hat{\theta}}$ is taken to be half of expected values of $\hat{\theta}$ given $\theta_0 = 1$ and expected values of $\hat{\theta}$ given $\theta_1 = 1.4$, where θ_0 is the target value and θ_1 is the value of process dispersion that needs to be detected quickly. Page (1963) provided in his Table 1, the reference values to notice a shift (that is $\theta_1 = 1.40$ to $\theta_1 = 2.23$) quickly in the process dispersion by using the sample range.

Accordingly, $K_{\hat{\theta}} = [E(\hat{\theta}|\theta_0) + E(\hat{\theta}|\theta_1)]/2$. As it is hard to find the analytical values of $E(\hat{\theta}|.)$, simulation is used for this purpose and 50,000 random samples are generated from a normal distribution with mean $\theta_0 = 1$, respectively, $\theta_1 = 1.40$ and variance equal to 1 and the said expected value is evaluated. Table 4.3 gives the values of $K_{\hat{\theta}}$ for samples of sizes $n = 5$ and $n = 9$ to detect a shift of size $\theta_1 = 1.4$.

Table 4.3: Values of $K_{\hat{\theta}}$ for CUSUM- $\hat{\theta}$ charts under Normal distribution

n	Estimator						
	S	IQR	G	$MADM$	B_n	T_n	S_r
5	1.13	1.47	1.35	0.98	1.32	1.35	1.15
9	1.16	1.34	1.35	1.09	1.26	1.25	1.2

For the values of $H_{\hat{\theta}}$ under the different environments (normal and non-normal), we have searched out the values of $H_{\hat{\theta}}$ by drawing random samples from the mentioned environments separately, and we have used an iterative method until the value of $H_{\hat{\theta}}$ in each case is obtained that fixes an intended ARL along with reference value $K_{\hat{\theta}}$. Values of $H_{\hat{\theta}}$ with ARL_0=500 are given in Table 4.4. Alternative values of $H_{\hat{\theta}}$ can be found in the same way for other choices of ARL_0.

Table 4.4: Values of $H_{\hat{\theta}}$ for CUSUM- $\hat{\theta}$ charts for different environments with $ARL_0 = 500$

Estimator	$N(0,1)$		$G(1,1)$	T_4	$Logis(0,1)$
	$n = 5$	$n = 9$	$n = 5$	$n = 5$	$n = 5$
S	1.531	0.816	2.954	3.12	1.971
IQR	2.203	1.478	2.9	3.06	2.412
G	1.91	0.973	2.812	3.195	2.294
$MADM$	3.31	1.951	1.959	2.24	2.888
B_n	2.46	1.161	2.623	2.312	2.433
T_n	2.877	1.57	2.25	2.22	2.62
S_r	3.641	1.86	2	2.596	3.309

The values of $K_{\hat{\theta}}$ and $H_{\hat{\theta}}$ have to be chosen carefully because the ARL performance of the CUSUM chart is sensitive to these values.

4.4 PERFORMANCE EVALUATION OF CUSUM - $\hat{\theta}$ CHARTS

To assess the performance of the proposed CUSUM - $\hat{\theta}$ charts, the average run length (ARL) is used as a performance measure. Monte Carlo simulation is used to determine the ARL of in-control and out-of-control processes. The simulation details are: we have generated 10^5 random samples of size n from the parent environments (i.e. normal, contaminated normal or non-normal) and the concerned dispersion statistic (i.e. S, $IQR, G\ MADM$, B_n, T_n or S_r) is calculated. The corresponding control limits of the chart are developed using Tables 4.3 and 4.4. Then, the sample number at which the plotting statistic Z_t falls outside the control limits is noted. This noted sample number is called the run length and it is a random variable. The same procedure is repeated 10^4 times to obtain the distribution of the run lengths. The mean of the run length distribution is represented by ARL and the standard deviation of run length distribution is represented by $SDRL$.

Inspired by Tatum (1997) and Schoonhoven, Nazir et al (2011), the performance of the CUSUM -$\hat{\theta}$ charts is evaluated under the following parent environments:

1) A model (say uncontaminated case) in which all observation are from $N(0,1)$.

2) A model for symmetric variance disturbances in which each observation has 99% probability of being drawn from $N(0,1)$ distribution and a 1% probability of being drawn from $N(0,9)$.

3) A model for asymmetric variance disturbances in which each observation is drawn from a $N(0,1)$ and has a 1% probability of having a multiple of a χ_1^2 variable added to it, with the multiplier equal to 4.

4) A model for mean disturbances in which each observation has a 99% probability of being drawn from $N(0, 1)$ distribution and a 1% probability of being drawn from the $N(4, 1)$ distribution.

5) To investigate the effect of using non-normal distributions we consider two cases: one by disturbing the kurtosis and the other by disturbing the symmetry of the distribution. For the case of disturbing the kurtosis we use Student's t distribution with 4 degrees of freedom (T_4) and the logistic distribution ($logis(0,1)$), and for the disturbance in symmetry we use the gamma distribution ($G(1,1)$).

Discussion of Results: Above mentioned environments are used and in each case the ARL values of the CUSUM - $\hat{\theta}$ charts are determined. We have considered shifts in terms of θ (i.e. $\delta\theta$), which means that the shifted dispersion parameter say $\acute{\theta}$ is defined as: $\acute{\theta} = \delta\theta$. Here $\delta = 1$ means no shift in θ, and the process dispersion is stable, and $\delta > 1$ means that the process θ has increased.

Uncontaminated Case: This environment is the basic assumption of the design structure of each chart. This provides a basis for comparisons for different control charts structures and hence for the proposed CUSUM - $\hat{\theta}$ charts. The following results can be observed from Table 4.5.

When no contaminations are present, the CUSUM - $\hat{\theta}$ chart based on the sample standard deviation S performs the best as was to be expected, followed by the chart based on G. The IQR based CUSUM chart works well as compared to charts based on B_n and T_n. The other CUSUM - $\hat{\theta}$ charts based on the remaining estimators ($MADM$ and S_r) are somewhat less efficient.

Table 4.5: ARL values of CUSUM- $\hat{\theta}$ charts under uncontaminated ($N\ (0, 1)$) environment when $ARL_0 = 500$

n	Estimator	δ								
		1	1.1	1.15	1.2	1.25	1.5	1.75	2	3
5	S	501.95	74.11	40.19	25.07	17.93	7.16	4.8	3.84	2.57
	IQR	499.29	78.63	42.79	26.81	18.85	7.6	5.11	4.04	2.67
	G	500.26	75.62	40.72	25.11	18.27	7.3	4.88	3.92	2.61
	$MADM$	498.8	117.58	69.97	47.43	34.31	14.25	9.04	6.94	4.15
	B_n	498.34	90.86	49.86	32.39	22.43	8.97	5.95	4.67	2.97
	T_n	502.76	92.92	53.11	34.72	24.7	9.98	6.51	5.05	3.16
	S_r	498.13	111.97	69.05	45.6	33.42	13.52	8.81	6.67	3.97
9	S	500.54	54.86	26.61	16.04	11.13	4.58	3.26	2.71	2.1
	IQR	498.78	75.21	39.65	24.68	17.01	6.66	4.5	3.61	2.45
	G	499.58	56.07	27.43	16.3	11.36	4.66	3.3	2.75	2.11
	$MADM$	504.81	89.09	48.58	31.23	22.23	8.81	5.87	4.54	2.91
	B_n	503.25	62.87	32.31	19.37	13.32	5.42	3.75	3.07	2.22
	T_n	498.83	73.99	39.29	24.58	17.21	6.76	4.59	3.66	2.5
	S_r	503.4	79.68	43.75	27.74	19.34	7.81	5.2	4.14	2.72

Increasing the sample size from $n = 5$ to $n = 9$ results that the B_n and T_n based charts perform better as compared to IQR for $\delta < 1.25$, but for $\delta > 1.25$ the IQR chart works well. (cf. Table 4.5).

To further explain the run length distribution under uncontaminated environment, we also report the standard deviations of the run length ($SDRL$) of the CUSUM - $\hat{\theta}$ charts to quantify the behavior of run length distribution as suggested by Antzoulakos and Rakitzis (2008). These are given in Table 4.6. When the process is in control, we want the $SDRL$ to be close to its intended value, namely 500. Table 4.6 reads that $SDRL$ is slightly lower to its intended value for some CUSUM - $\hat{\theta}$ charts and $SDRL$ decreases as δ increases for all charts.

Table 4.6: $SDRL$ values of CUSUM- $\hat{\theta}$ charts under uncontaminated $(N\ (\mathbf{0}, \mathbf{1}))$ environment when $ARL_0 = 500$

n	Estimator	δ								
		1	1.1	1.15	1.2	1.25	1.5	1.75	2	3
5	S	489.16	70.57	35.38	20.73	13.32	3.69	1.98	1.38	0.69
	IQR	498.66	72.48	36.65	21.67	13.90	3.89	2.12	1.48	0.74
	G	483.77	69.34	35.19	20.21	13.49	3.73	2.03	1.41	0.71
	$MADM$	484.19	108.42	61.49	38.89	26.35	8.47	4.52	3.17	1.57
	B_n	488.85	84.24	43.78	26.34	16.83	4.88	2.61	1.81	0.90
	T_n	491.10	86.20	46.83	28.31	18.86	5.39	3.00	2.04	1.01
	S_r	489.99	102.42	60.96	37.09	25.50	8.01	4.49	3.08	1.49
9	S	486.06	51.89	22.69	12.50	7.73	2.00	1.11	0.77	0.31
	IQR	495.56	65.46	32.31	18.19	11.44	3.12	1.66	1.14	0.55
	G	492.33	51.72	23.96	12.52	7.99	2.08	1.12	0.79	0.32
	$MADM$	495.82	82.93	42.48	25.65	16.82	4.78	2.61	1.77	0.88
	B_n	499.08	58.66	27.94	15.14	9.33	2.52	1.33	0.95	0.44
	T_n	494.86	68.97	34.22	19.75	12.45	3.40	1.85	1.25	0.64
	S_r	494.06	73.88	37.86	22.67	14.34	4.07	2.22	1.54	0.77

Symmetric Variance Case: When symmetric disturbances are present, the best performing CUSUM charts are based on the $MADM$, S_r, T_n followed by B_n. These estimators are very robust to outliers as these deviate less from the in-control ARL. The other CUSUM charts are very poor in the in-control situation as their $ARLs$ deviate very much from the intended ARL (cf. Table 4.7). The performance of the CUSUM charts based on S, G and B_n become even worse when the sample size increases. However, the CUSUM chart based on IQR performs better than the one based on T_n for $n = 9$.

Asymmetric Variance Case: When asymmetric variance disturbances are present, the CUSUM charts based on the estimators $MADM$, S_r, T_n, and B_n show good resistance against such disturbances and perform efficiently in detecting small shifts in the process dispersion parameters. The other charts perform very badly in maintaining the in-control ARL. The CUSUM chart based on IQR recovers quite substantially as the sample size gets larger but the performance of S and G is even worse (cf. Table 4.8).

Table 4.7: ARL values of CUSUM- $\hat{\theta}$ charts under symmetric variance contaminated environment when $ARL_0 = 500$

n	Estimator	δ								
		1	1.1	1.15	1.2	1.25	1.5	1.75	2	3
5	S	123.33	43.29	28.4	19.98	14.82	6.76	4.67	3.77	2.55
	IQR	177.72	49.83	31.64	21.8	16.14	7.19	4.9	3.93	2.64
	G	145.58	45.37	29.23	20.42	15.28	6.86	4.72	3.83	2.59
	$MADM$	388.63	95.4	61.11	42.5	31.28	13.65	8.89	6.75	4.08
	B_n	291.45	67.63	40.72	27.95	20.55	8.65	5.72	4.57	2.93
	T_n	343.45	75.44	45.33	30.49	22.44	9.46	6.33	4.96	3.13
	S_r	362.72	93.62	58.73	40.65	29.99	12.97	8.47	6.55	3.92
9	S	74.89	27.21	17.66	12.06	9.31	4.33	3.16	2.65	2.1
	IQR	355.39	58	31.95	20.41	14.78	6.01	4.16	3.36	2.35
	G	106.7	30.84	18.61	12.76	9.64	4.42	3.2	2.68	2.11
	$MADM$	392.09	76.71	43.75	28.73	20.62	8.52	5.65	4.48	2.89
	B_n	266.37	46.75	25.17	16.54	11.86	5.24	3.67	3.03	2.21
	T_n	342.77	59.86	33.9	21.96	16.11	6.59	4.51	3.62	2.48
	S_r	357.75	66.48	37.39	25.33	17.81	7.53	5.11	4.06	2.69

Table 4.8: ARL values of CUSUM-$\hat{\theta}$ charts under asymmetric variance environment when $ARL_0 = 500$

n	Estimator	δ								
		1	1.1	1.15	1.2	1.25	1.5	1.75	2	3
5	S	65	37.27	26.11	19.57	14.86	6.8	4.7	3.77	2.58
	IQR	78.47	40.41	28.21	20.41	15.75	7.25	4.99	3.98	2.65
	G	70.16	38.21	26.66	19.85	15.13	6.91	4.8	3.86	2.59
	$MADM$	376.45	97.42	61.68	42.95	31.99	13.71	9.04	6.88	4.11
	B_n	251.93	65.73	40.43	27.31	20.29	8.7	5.82	4.58	2.97
	T_n	318.11	75.17	46.13	31.12	22.73	9.64	6.41	5.02	3.15
	S_r	359.13	95.05	60.27	41.39	30.77	13.22	8.6	6.68	3.94
9	S	37.94	23.06	16.54	11.94	9.09	4.36	3.21	2.69	2.1
	IQR	328.43	57.56	32.12	20.86	14.8	6.1	4.21	3.4	2.37
	G	44.47	24.61	17.01	12.34	9.45	4.47	3.23	2.72	2.11
	$MADM$	390.99	78.62	44.7	29.54	21.35	8.61	5.74	4.52	2.9
	B_n	237.47	45.36	26.17	17.03	12.14	5.25	3.74	3.04	2.22
	T_n	337.68	61.14	34.64	22.36	16.18	6.63	4.54	3.63	2.49
	S_r	360.59	66.87	38.89	25.29	18.27	7.58	5.14	4.12	2.71

Location Contaminated Case: When disturbances are present in form of introducing outliers in the location of the process, then the CUSUM chart based on $MADM$ performs well as it maintains more or less the in-control intended ARL, followed by charts based on S_r and T_n. All other charts perform very poor as they are unable to maintain the in-control properties to the target ARL (cf. Table 4.9). Increasing the sample size results in an even better performance of the CUSUM chart based on $MADM$. The CUSUM chart based on IQR performs much better for a larger sample size but the CUSUM charts based on S, G and B_n deviate even more from their intended ARL in these circumstances.

Table 4.9: ARL values of CUSUM- $\hat{\theta}$ charts under location contaminated environment when $ARL_0 = 500$

n	Estimator	δ								
		1	1.1	1.15	1.2	1.25	1.5	1.75	2	3
5	S	64.86	31.68	23.03	17.24	13.61	6.56	4.65	3.75	2.56
	IQR	102.66	38.58	26.13	19.2	14.88	7.01	4.93	3.95	2.65
	G	80.54	34.5	23.79	17.6	14.06	6.61	4.73	3.8	2.6
	$MADM$	285.24	84.42	55.24	39.23	30.04	13.29	8.85	6.82	4.07
	B_n	177.21	53.92	34.92	24.65	18.74	8.26	5.68	4.55	2.93
	T_n	234.58	63.94	39.65	27.93	20.82	9.18	6.29	4.96	3.14
	S_r	267.55	79.41	52.26	36.86	28.22	12.73	8.47	6.53	3.94
9	S	36.91	19.34	13.92	10.44	8.27	4.22	3.15	2.65	2.1
	IQR	234.5	49.35	28.35	18.56	13.72	5.9	4.12	3.34	2.36
	G	56.12	22.65	15.05	11.18	8.67	4.3	3.17	2.7	2.1
	$MADM$	325.3	68.19	40.11	26.75	19.68	8.44	5.61	4.46	2.89
	B_n	152.15	36.07	21.83	14.77	10.96	5.1	3.64	3.01	2.21
	T_n	252.78	52.25	30.34	20.34	14.79	6.43	4.45	3.59	2.48
	S_r	273.73	56.56	33.75	23.19	16.91	7.36	5.07	4.03	2.71

Breakdown Points and Robustness of the Charts: Under an uncontaminated environment as was expected no other CUSUM chart can perform better than the one based on S. It can be seen from Table 4.6 that this chart outperforms all, followed by the CUSUM charts based on G and IQR for small samples and for large samples the B_n and T_n based CUSUM charts work well. $MADM$ based CUSUM chart performs the worst from all as its efficiency under normality is very low (i.e. only 37%). But when there is contamination in the data one can read from Tables 4.7 and 4.8 that the $MADM$ based CUSUM chart (with 50% breakdown point and with low gross error sensitivity) maintains its in control properties well, followed by the other robust estimators based charts S_r and T_n (both estimators have 50% breakdown points). With an increase in the sample sizes the same output is observed but the in control performance of IQR based CUSUM chart (having 25% breakdown point) improves substantially. Under contaminations the S and G based CUSUM charts work poorly as both are based on non-robust estimators (e.g. S has a zero breakdown point).

Non-Normal Cases: Without loss of generality, the drawn samples are transformed in such a way that the resulting sample has a mean equal to zero and a variance equal to one. For this purpose, the mean of concerned environment is subtracted from every drawn sample and then divided by the standard deviation of the concerned environment, so that valid and comparable results are evaluated.

Table 4.10: ARL values of CUSUM- $\hat{\theta}$ charts under $G(1,1)$ Environment when $ARL_0 = 500$

n	Estimator	δ								
		1	1.1	1.15	1.2	1.25	1.5	1.75	2	3
5	S	502.91	158.82	102.39	67.93	48.92	16.55	9.77	7.05	3.97
	IQR	501.02	157.33	97.27	64.05	46.22	14.96	8.54	6.3	3.58
	G	503.32	159.14	100.77	66.16	46.57	15.35	8.89	6.45	3.6
	$MADM$	504.29	219.76	152.19	110.23	81.42	27.41	14.49	9.54	4.69
	B_n	498.86	201.46	135.6	95.3	68.38	21.41	11.42	7.84	4.04
	T_n	504.79	196.02	127.77	90.87	64.85	21.05	11.1	7.51	3.84
	S_r	504.78	222.3	151.62	110.93	81.52	28.29	14.62	9.56	4.55

Under the gamma distribution, the CUSUM - $\hat{\theta}$ chart based on IQR performs best followed by the charts based on S and G. The charts based on the other estimators detect somewhat less efficiently shifts in the process (see Table 4.10). It can be observed from Table 4.11 that the CUSUM - $\hat{\theta}$ charts based on T_n outperforms all other charts under the T_4 distribution followed by the charts based on S_r , B_n and $MADM$. The CUSUM charts based on the G estimator performs efficiently followed by the charts based on IQR under the logistic distribution (cf. Table 4.12). Other charts relatively work well under this environment.

Table 4.11: ARL values of CUSUM- $\hat{\theta}$ charts under T_4Environment when $ARL_0 = 500$

n	Estimator	δ								
		1	1.1	1.15	1.2	1.25	1.5	1.75	2	3
5	S	499.31	209.97	135.56	87.64	60.88	17.61	9.95	7.18	3.94
	IQR	503.76	189.22	118.32	77.49	51.75	15.06	8.42	6.12	3.48
	G	505.82	194.44	122.35	80.49	54.45	15.9	8.91	6.49	3.65
	$MADM$	503.52	175.9	114.06	76.57	56.27	18.6	10.4	7.49	4.07
	B_n	500.6	172.75	111.75	72.34	49.35	14.52	8.12	5.89	3.35
	T_n	498.75	166.88	102.52	69.47	47.9	14.47	8.17	5.82	3.33
	S_r	500.47	169.8	109.51	74.32	53.85	17.73	10.31	7.32	4

Table 4.12: ARL values of CUSUM- $\hat{\theta}$ charts under $Logist(0,1)$Environment when $ARL_0 = 500$

n	Estimator	δ								
		1	1.1	1.15	1.2	1.25	1.5	1.75	2	3
5	S	501.32	103.18	58.28	36.67	25.63	9.32	5.99	4.63	2.93
	IQR	500.76	100.59	56.41	35.88	25.02	9.12	5.85	4.54	2.87
	G	499.43	100.33	56.25	35.47	25.02	9.13	5.86	4.6	2.89
	$MADM$	506.61	137.01	83.99	56.96	41.17	15.38	9.47	7.07	4.09
	B_n	498.67	113.42	64.18	42.11	28.92	10.51	6.56	5.02	3.08
	T_n	498.65	117.24	66.95	43.99	30.27	11.04	6.9	5.22	3.21
	S_r	504.4	135.15	83.04	55.21	38.58	14.97	9.28	6.91	4.02

4.5 EWMA DISPERSION CONTROL CHARTS

The EWMA control charts, originally proposed by Roberts (1959), are capable to quickly detect small shifts in the process parameters such as locations and dispersion. The assumption of normality is generally required for the commonly used EWMA structures. In practice, one may not frequently come across normally distributed process scenarios and hence we have to deal with non-normal situations (cf. Bissell (1994), Levinson and Polny (1999)). The monitoring of process dispersion using EWMA setup has attracted the attention of many researchers. Maravelakis et al. (2005) showed that the run length behavior of the EWMA dispersion charts is seriously affected due to non-normality. Some important contributions in this concern are Shu and Jiang (2008), Abbasi and Miller (2013) and the references therein.

Recently, Zwetsloot et al. (2014b) proposed a new Phase I estimation method for the dispersion parameter. They studied the contamination effects during phase I and evaluated the performance of the Phase II EWMA control chart based on the various dispersion estimators. For non-normality concerns, we have investigated the EWMA dispersion charting structures under normal and a variety of non-normal distributions. We have considered a wide range of dispersion estimators as discussed in Section 4.2. It is observed that under the ideal assumption of normality, the best performance is shown by the S chart, followed by D and $AADM$ charts. For most of the non-normal distributions, the best performance is generally offered by D and $AADM$ charts. It is to be mentioned that for extremely skewed Log-normal distribution, Q_n chart appears as a better choice. The relative performance of R and IQR charts reduces with an increase in n, whereas for the Q_n chart, this phenomenon is opposite. For more details on the non-normal investigation with EWMA dispersion charts, one may see Abbasi, Riaz, Miller, Shabbir and Nazir (2014).

4.6 CONCLUSION

In this chapter, we have considered several estimators of the dispersion parameter for the use in establishing phase II control limits. These estimators comprise some commonly used as well as robust estimators which are not common in the control charts literature. A CUSUM chart scheme has been used to monitor the dispersion parameter using these estimators. The performance of these estimators has been assessed under various situations: the uncontaminated situation and various situations contaminated with symmetric and asymmetric variance disturbances, location disturbances and non-normal environments.

Under the uncontaminated situation all charts perform well but the CUSUM chart based on the sample standard deviation S outperforms all, as was expected under normality. When there are symmetric and asymmetric variance disturbances the CUSUM charts based on T_n, $MADM$ and S_r perform satisfactory and the performance of the other charts is (very) poor. The CUSUM chart based on the IQR estimator performs well for the gamma distribution, the G estimator performs well under the logistic distribution and the T_n estimator performs well under the t-distribution. However, the differences between the estimators are not very substantial. In short, the dispersion CUSUM charts based on robust estimators (T_n, $MADM$ and S_r) behave well in all types of environments (uncontaminated, contaminated and non-normal).

Chapter 5

Robust CUSUM Control Charts for Location

Cumulative sum (CUSUM) control charts are very effective in detecting special causes. In general the underlying distribution is supposed to be normal for efficient working of the said charts. In designing a CUSUM chart, it is important to know how the chart will respond to disturbances of normality. The focus of this chapter is to control the location parameter using a CUSUM structure and the major concern is to identify the CUSUM control charts which are of more practical value under different normal, non-normal, contaminated normal and special cause contaminated parent scenarios. In this chapter, we propose and compare the performance of different CUSUM control charts, for Phase II monitoring of location, based on mean, median, midrange, Hodges-Lehmann and trimean statistics. The average run length is used as the performance measure of the CUSUM control charts. The results of this chapter are published in Nazir, Riaz, Does and Abbas (2013). In that paper also a numerical example is given.

5.1 INTRODUCTION

Control charts, one of the statistical process control tools, help differentiating between the common cause and the special cause variations in the response of a quality characteristic of interest in a process. Some major classifications of the control charts are: memory-less (Shewhart's type) and memory (like cumulative sum (CUSUM) and exponentially weighted moving range (EWMA)) charts. The commonly used Shewhart's variables control charts are the mean ($\bar{X}$), median and midrange charts for monitoring the process location and the range (R), standard deviation (S) and variance (S^2) charts for monitoring the process variability (cf.

Montgomery (2009)). The main deficiency of Shewhart-type control charts is that they are less sensitive to small and moderate shifts in the process parameter(s).

Another approach to address the detection of small shifts is to use memory control charts. CUSUM control charts proposed by Page (1954) and EWMA control charts proposed by Roberts (1959) are the names of two commonly used memory-type control charts. These charts are designed such that they use the past information along with the current information, which make them very sensitive to small and moderate shifts in the process parameters.

In order to get efficient control limits for Phase II monitoring, it is generally assumed that the process has a stable behavior when the data is collected during Phase I (cf. Woodall and Montgomery (1999) and Vining (2009)). Most of the evaluations of existing control charts depend on the assumptions of normality, no contaminations, no outliers and no measurement errors in Phase I for the quality characteristic of interest. In case of violation of these assumptions, the design structures of the charts lose their performance ability and hence are of less practical use. There are a lot of practical situations where non-normality is more common (see for example Janacek and Miekle (1997)). One of the solutions to deal with this is to use control charts which are robust against violations of the basic assumptions, like normality.

Some authors have discussed on the robustness of CUSUM chart to the situations where the underlying assumptions are not fulfilled. Lucas and Crosier (1982) studied the robustness of the standard CUSUM chart and proposed four methods to reduce the effect of outliers on the ARL performance. McDonald (1990) proposed the use of the CUSUM chart which is based on a non-parametric statistic. He used the idea of ranking the observations first and then using those ranks in the CUSUM structure. Hawkins (1993) proposed a robust CUSUM chart based on winsorization. MacEachern et al. (1997) proposed a robust CUSUM chart based on the likelihood of the variate and named their newly proposed chart as RLCUSUM chart. Li et al. (2010) proposed a non-parametric CUSUM chart based on Wilcoxon rank-sum test. Reynolds Jr and Stoumbos (2010) considered the robustness of the CUSUM chart for monitoring the process location and dispersion simultaneously. Midi and Shabbak (2011) studied the robust CUSUM control charting for the multivariate case. Lee (2011) proposed the economic design of the CUSUM chart for non-normally serially correlated data. Yang and Cheng (2011) proposed a new non-parametric CUSUM chart which is based on a non-parametric test named as sign test. They studied the ARL performance of the proposed chart for monitoring different location parameters. Similarly, a lot of work has been done in the direction of robust EWMA control charting e.g. see Amin and Searcy (1991), Yang et al. (2011), Graham et al. (2012) and Zwetsloot et al. (2014a).

Most of the CUSUM control charting techniques discussed above are based on, first transforming the observed data into a non-parametric statistic and then applying the CUSUM chart on that transformed statistic. Unlike these approaches, Yang et al. (2010) proposed the use of robust location estimator (i.e. the sample median) with the CUSUM control structure. Extending their approach, we present in this chapter, the robust CUSUM chart which is based on 5 different estimators for monitoring the process location of Phase II samples. The performance of the CUSUM chart with different robust estimators is studied in the presence of disturbances to normality, contaminations, outliers and special causes in the process of interest. Before moving on towards the robust estimators, we provide the basic structure of the CUSUM chart in the next section.

5.2 THE CLASSICAL MEAN CUSUM CONTROL CHART

The mean CUSUM control chart proposed by Page (1954) has become one of the most popular methods to monitor processes. For a two sided CUSUM chart, we plot the two statistics C_i^+ and C_i^- against a single control limit H. These two plotting statistics are defined as:

$$\left.\begin{aligned} C_i^+ &= \max[0, (\bar{X}_i - \mu_0) - K + C_{i-1}^+] \\ C_i^- &= \max[0, -(\bar{X}_i - \mu_0) - K + C_{i-1}^-] \end{aligned}\right\} \tag{5.1}$$

where i is the subgroup number, $\bar{X}_i$ is the mean of study variable X, μ_0 is the target mean of the study variable X and K is the reference value of the CUSUM scheme (cf. Montgomery (2009)). The starting value for both plotting statistics is usually taken equal to zero, i.e. $C_0^+ = C_0^- = 0$. Next, we plot these two statistics together with the control limit H. It is concluded that the process mean has moved upward if $C_i^+ > H$ for any value of i, whereas the process mean is said to be shifted downwards if $C_i^- > H$ for any value of i. Thus the CUSUM chart is defined by two parameters, i.e. K and H, which have to be chosen carefully, because the ARL performance of the CUSUM chart is sensitive to these parameters. These two parameters are used in the standardized manner (cf. Montgomery (2009)) and are given as:

$$K = k\sigma_{\bar{X}}, \qquad H = h\sigma_{\bar{X}} \tag{5.2}$$

,where $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}$, σ_X is the standard deviation of the study variable X and n is the sample size. The CUSUM statistic in (5.1) is given for the mean of samples. In the next section, we provide some robust estimators which are used with the CUSUM charts to make its structure robust.

5.3 DESCRIPTION OF THE PROPOSED ESTIMATORS AND THEIR CORRESPONDING CUSUM CHARTS

Let θ be the population location parameter which will be monitored through control charting and $\hat{\theta}$ be its estimator based on a subgroup of size n. There are many choices for $\hat{\theta}$ out of which we consider here the following: mean, median, midrange, Hodges-Lehmann estimator and trimean.

Based on a random sample $X_1, X_2, \ldots \ldots, X_n$ of size n, these estimators are defined as:

Mean: This estimator is included in this chapter to provide a basis for comparison, as it is the most efficient estimator for normally distributed data. The sample mean $\bar{X}$ is defined as $\bar{X} = \frac{\sum_{j=1}^{n} X_j}{n}$, is a linear function of data and is widely used to estimate the population location parameter. According to Rousseeuw (1991), the sample mean $\bar{X}$ performs well under normality assumption but it is highly affected due to non-normality, is sensitive to outliers and has zero breakdown point, which means that even a single inconsistent observation can change its value.

Median: The sample median $\tilde{X}$ is defined as the middle order statistic for odd sample sizes and the average of the two middle order statistics in case of even sample sizes. The median is a robust estimator so it is less affected by non-normality. Dixon and Massey (1969) showed that the efficiency of the median with respect to the mean decreases with an increase in sample size and the efficiency approaches 0.637 for large sample sizes.

Midrange: The midrange is defined as $MR = \frac{X_{(1)} + X_{(n)}}{2}$ where $X_{(1)}$ and $X_{(n)}$ are the lowest and highest order statistics in a random sample of size n. It is highly sensitive to outliers as its design structure is based on only extreme values of data and therefore it has a zero breakdown point. Although its use is not very common due to its low efficiency for mesokurtic distributions, its design structure has the ability to perform good in the case of small samples from platykurtic distributions. Ferrell (1953) presented a comparison of both the efficiency of estimation and the detection of disturbances by medians and midranges. He indicated an advantage of these less conventional measures in detecting outliers, while not losing much efficiency in detecting a change in the central value.

Hodges-Lehmann estimator: This estimator is defined as the median of the pairwise Walsh averages that is: $HL = \text{median}\big((X_j + X_k)/2\,, 1 \le j \le k \le n\big)$. The main advantage of the

HL estimator is that it is robust against outliers in a sample. It has a breakdown point of 0.29 (i.e. 29% is the least portion of data contamination needed to derive the estimate beyond all bounds (cf. Hettmansperger and McKean (1998))). If the underlying distribution for the data is normal, then the asymptotic relative efficiency (ARE) of the HL estimator relative to the sample mean is 0.955; otherwise it is often greater than unity (cf. Lehmann (1983)).

Trimean: The TriMean (TM) of sample is the weighted average of the sample median and two quartiles: $TM = \frac{Q_1+2Q_2+Q_3}{4}$ (cf. Tukey (1977) and Wang et al. (2007)). It also equals the average of the median and the midhinge: $TM = \frac{1}{2}\left(Q_2 + \frac{Q_1+Q_3}{2}\right)$,where Q_p is the p^{th} quartile of the sample (cf. Weisberg (1992)). Like the median and the midhinge, but unlike the sample mean, TM is a statistically resistant L-estimator (a linear combination of order statistics) having a breakdown point of 25%. According to Tukey (1977), using TM instead of the median gives a more useful assessment of the location parameter. According to Weisberg (1992) the "statistical resistance" benefit of TM as a measure of the center (of a distribution) is that it combines the median's emphasis on center values with the midhinge's attention to the extremes. If the underlying distribution for the data is normal then the ARE of the TM relative to the sample mean is 0.83 (cf. Wang et al. (2007)).

The performance of these 5 estimators for Phase I analysis could be seen in Schoonhoven, Nazir et al. (2011). In this chapter, we study the performance of these estimators with the CUSUM structure for the Phase II analysis. Next we present the description of the proposed CUSUM control charts. Following (5.1), the $\hat{\theta}$ −CUSUM control chart statistics S_i^+ and S_i^- can be represented as:

$$\left.\begin{aligned} S_i^+ &= \max\left[0, \left(\hat{\theta}_i - \mu_0\right) - K_{\hat{\theta}} + S_{i-1}^+\right] \\ S_i^- &= \max\left[0, -\left(\hat{\theta}_i - \mu_0\right) - K_{\hat{\theta}} + S_{i-1}^-\right] \end{aligned}\right\} \tag{5.3}$$

In (5.3) $\hat{\theta}$ may be any choice of the estimators mentioned above. Initial values for the statistics given in (5.3) are taken equal to zero, i.e. $S_0^+ = S_0^- = 0$. The decision rule for the proposed chart(s), following the spirit of (5.2), is given in the form of $K_{\hat{\theta}}$ and $H_{\hat{\theta}}$ defined as:

$$K_{\hat{\theta}} = k_{\hat{\theta}}\sigma_{\hat{\theta}}, \quad H_{\hat{\theta}} = h_{\hat{\theta}}\sigma_{\hat{\theta}} \tag{5.4}$$

5.4 PERFORMANCE OF THE PROPOSED CUSUM CONTROL CHARTS

The performance of the proposed CUSUM control charting design structures given in (5.3) and (5.4) is evaluated in this section. The performance is measured in terms of their average run lengths (ARLs), which is the expected number of subgroups before a shift is detected. The in-control ARL is denoted by ARL_0 and an out-of-control ARL is denoted by ARL_1.

A code developed in R language is used to simulate ARL. For the said purpose, first we have simulated the standard errors for all the estimators, i.e. 5×10^5 samples of size n are generated and for each sample, the value of estimator is determined. The standard deviation of those 5×10^5 values of estimator is represented by $\sigma_{\hat{\theta}}$. Then the Phase II samples of size n from different environments (explained later in this section) are generated under an in-control situation and the proposed charts are applied on those samples until an out-of-control signal is detected. The respective sample number (when the shift is detected by the chart) is noted which is the in-control run length. This process is repeated $10{,}000$ times so that we can average out these run lengths to get the ARL_0. Finally, a shift in the form of δ is introduced in the process to evaluate the ARL_1 performance of the proposed charts, where δ is the difference between μ_0 and μ_1, and μ_1 is the shifted mean. We have chosen to have $ARL_0 \cong 370$.

Uncontaminated normal environment: An uncontaminated normal distribution means that all observations come from $N(\mu_0, \sigma_0^2)$. The $ARLs$ of the CUSUM charts using different estimators under the uncontaminated environment are given in Tables 5.1 and 5.2.

Table 5.1: ARL values for the CUSUM chart based on different estimators under uncontaminated normal distribution with $k_{\hat{\theta}} = 0.25$ and $h_{\hat{\theta}} = 8.03$

Subgroup size	Estimator	δ						
		0	0.25	0.5	0.75	1	1.5	2
$n = 5$	**Mean**	370.096	24.592	10.006	6.353	4.645	3.151	2.395
	Median	372.498	31.594	12.377	7.702	5.600	3.734	2.850
	Midrange	370.754	29.815	11.677	7.309	5.352	3.545	2.729
	HL	373.121	25.834	10.438	6.545	4.806	3.245	2.476
	TM	373.932	27.589	10.918	6.900	5.066	3.393	2.599
$n = 10$	**Mean**	372.356	15.352	6.722	4.391	3.322	2.251	1.976
	Median	372.746	18.954	8.053	5.194	3.877	2.653	2.057
	Midrange	373.508	23.407	9.578	6.043	4.503	3.039	2.292
	HL	370.992	15.988	6.963	4.539	3.420	2.315	1.990
	TM	375.278	16.773	7.284	4.707	3.526	2.411	2.001

Tables 5.1 and 5.2 indicate that all estimators have almost the same in-control ARL, i.e. $ARL_0 \cong 370$. For the out-of-control case, the $ARLs$ of the CUSUM based on the mean are clearly smaller as compared to the other estimators. Here the performance of the HL estimator is comparable to the mean estimator as the ARL performance of both is almost the same, whereas the other estimators have larger ARL_1 values. Moreover, the effect of change in subgroup size and $k_{\hat{\theta}}$ is identical for all estimators.

For describing more about the run length distribution, we also report the standard deviations of the run length ($SDRL$) and some percentile points of the run length distribution in Tables 5.3 and 5.4, respectively.

Table 5.2: ARL values for the CUSUM chart based on different estimators under uncontaminated normal distribution with $k_{\hat{\theta}} = 0.5$ and $h_{\hat{\theta}} = 4.774$

Subgroup size	Estimator	δ						
		0	0.25	0.5	0.75	1	1.5	2
$n = 5$	**Mean**	370.469	28.338	8.304	4.788	3.383	2.242	1.803
	Median	374.278	41.831	11.267	6.073	4.205	2.665	2.065
	Midrange	370.110	37.528	10.265	5.712	3.974	2.534	1.986
	HL	367.095	29.990	8.790	4.999	3.521	2.306	1.851
	TM	368.020	32.519	9.363	5.247	3.699	2.392	1.908
$n = 10$	**Mean**	367.292	14.845	5.156	3.177	2.364	1.708	1.148
	Median	367.663	19.538	6.390	3.837	2.774	1.946	1.468
	Midrange	368.529	26.582	7.919	4.568	3.275	2.175	1.756
	HL	373.781	15.607	5.393	3.277	2.438	1.753	1.194
	TM	370.949	16.720	5.640	3.421	2.528	1.816	1.261

Table 5.3: $SDRL$ values for the CUSUM chart based on different estimators under uncontaminated normal distribution with $n = 10$ at $ARL_0 \cong 370$

Chart parameters	Estimator	δ						
		0	0.25	0.5	0.75	1	1.5	2
$k_{\hat{\theta}} = 0.25$ $h_{\hat{\theta}} = 8.03$	**Mean**	360.523	6.590	1.902	0.994	0.641	0.437	0.164
	Median	356.984	9.085	2.477	1.264	0.824	0.533	0.241
	Midrange	360.781	12.457	3.195	1.611	1.021	0.564	0.459
	HL	350.709	6.980	1.991	1.032	0.669	0.469	0.142
	TM	356.536	7.674	2.143	1.099	0.706	0.501	0.130
$k_{\hat{\theta}} = 0.5$ $h_{\hat{\theta}} = 4.774$	**Mean**	367.992	9.498	1.963	0.939	0.577	0.465	0.355
	Median	363.205	13.603	2.766	1.245	0.764	0.403	0.499
	Midrange	361.144	20.159	3.764	1.623	0.975	0.479	0.446
	HL	367.477	10.194	2.141	0.988	0.626	0.450	0.395
	TM	365.022	11.056	2.278	1.043	0.647	0.421	0.439

Table 5.4: Percentile run length values for the CUSUM chart based on different estimators under uncontaminated normal distribution with $n = 10$, $k_{\hat{\theta}} = 0.5$ and $h_{\hat{\theta}} = 4.774$ at $ARL_0 \cong 370$

Estimator	Percentile	δ						
		0	0.25	0.5	0.75	1	1.5	2
Mean	P_{10}	44	6	3	2	2	1	1
	P_{25}	111	8	4	3	2	1	1
	P_{50}	255	12	5	3	2	2	1
	P_{75}	504	19	6	4	3	2	1
	P_{90}	833.1	27	8	4	3	2	2
Median	P_{10}	45	7	3	2	2	1	1
	P_{25}	112	10	4	3	2	2	1
	P_{50}	256	16	6	4	3	2	1
	P_{75}	501	25	8	5	3	2	2
	P_{90}	843	37	10	5	4	2	2
Midrange	P_{10}	46	8	4	3	2	2	1
	P_{25}	113	12	5	3	3	2	1
	P_{50}	255	21	7	4	3	2	2
	P_{75}	504	35	10	5	4	2	2
	P_{90}	828	53	13	7	4	3	2
HL	P_{10}	44	6	3	2	2	1	1
	P_{25}	111	9	4	3	2	1	1
	P_{50}	260	13	5	3	2	2	1
	P_{75}	518	20	7	4	3	2	1
	P_{90}	847.1	29	8	5	3	2	2
TM	P_{10}	44	6	3	2	2	1	1
	P_{25}	109	9	4	3	2	2	1
	P_{50}	258	14	5	3	2	2	1
	P_{75}	517	21	7	4	3	2	2
	P_{90}	846.1	31	9	5	3	2	2

Considering the results of Tables 5.3 and 5.4, we observe that the run length distribution of all the proposed CUSUM charts is positively skewed as long as there is some variation in the run lengths. By decreasing the value of $k_{\hat{\theta}}$ from 0.5 to 0.25, the standard deviation of all the proposed charts also decreases for small values of δ. The percentiles can also be used to compare the Median Run Length (MRL) for CUSUM chart with different estimators.

Lucas and Crosier (1982) proposed the use of a fast initial response (FIR) feature with the CUSUM charts in which they recommended not to set the initial values of CUSUM statistics equal to zero. They found that the choice of a head start $S_0 = H_{\hat{\theta}}/2$ is optimal in the sense that its effect on the ARL_0 is very minor but it significantly decreases the ARL_1 values. The intention of FIR was to enhance the CUSUM chart sensitivity in detecting the shifts that occur immediately after the start of the process. As an example, we provide the ARL values of FIR CUSUM based on different estimators for $n = 10$, $k_{\hat{\theta}} = 0.5$ and $S_0 = H_{\hat{\theta}}/2$ at $ARL_0 \cong 370$. The results are given in Table 5.5 where the uncontaminated normal distribution is taken as the parent environment.

Table 5.5: ARL values for the FIR CUSUM chart based on different estimators under uncontaminated normal distribution with $n = 10$, $k_{\hat{\theta}} = 0.5$ and $S_0 = H_{\hat{\theta}}/2$ at $ARL_0 \cong 370$

Estimator	$h_{\hat{\theta}}$	δ						
		0	0.25	0.5	0.75	1	1.5	2
Mean	4.86	371.772	9.920	3.088	1.919	1.436	1.031	1.000
Median	4.87	371.999	14.114	3.886	2.297	1.694	1.134	1.007
Midrange	4.9	370.930	19.588	4.853	2.699	1.969	1.295	1.043
HL	4.86	369.097	10.216	3.182	1.979	1.479	1.045	1.001
TM	4.86	372.281	11.417	3.364	2.069	1.523	1.068	1.002

It is gratifying to note that the FIR feature indeed enhances the performance of the proposed CUSUM charts. In the sequel we will not give the additional tables for the $SDRL$, percentile points of the run length and the FIR features, although they can be easily obtained according the same lines.

Variance contaminated normal environment: A $(\varphi)100\%$ variance contaminated normal distribution is one that contain $(1 - \varphi)100\%$ observations from $N(\mu_0, \sigma_0^2)$ and $(\varphi)100\%$ observations from $N(\mu_0, \tau\sigma_0^2)$, where $0 < \tau < \infty$. The ARL values of the CUSUM charts using different estimators for the variance contaminated normal environment are given in Tables 5.6 with $\varphi = 0.05$ and $\tau = 9$ and in Table 5.7 with $\varphi = 0.1$ and $\tau = 9$.

Table 5.6: ARL values for the CUSUM chart based on different estimators under 5% variance contaminated normal distribution with $k_{\hat{\theta}} = 0.5$ and $h_{\hat{\theta}} = 4.774$

Subgroup size	Estimator	δ						
		0	0.25	0.5	0.75	1	1.5	2
$n = 5$	**Mean**	303.375	40.319	10.911	5.900	4.129	2.630	2.037
	Median	353.227	44.057	11.895	6.454	4.424	2.789	2.125
	Midrange	198.394	77.199	20.560	9.772	6.300	3.746	2.733
	HL	342.091	35.703	10.112	5.604	3.918	2.511	1.975
	TM	354.750	36.583	10.288	5.700	3.955	2.527	1.988
$n = 10$	**Mean**	330.826	20.222	6.475	3.844	2.803	1.959	1.474
	Median	366.873	21.549	6.797	4.009	2.891	1.995	1.545
	Midrange	197.230	83.124	22.377	10.551	6.622	3.920	2.866
	HL	359.458	17.524	5.845	3.552	2.601	1.864	1.332
	TM	361.461	18.050	6.040	3.606	2.652	1.886	1.350

In Tables 5.6 and 5.7, the ARL_0 for the TM and HL CUSUM charts are affected least by the contamination, while the Midrange estimator is affected the most by this variance contamination. Similarly, in terms of ARL_1 values, TM and HL are outperforming all of the other estimators under discussion.

Table 5.7: ARL values for the CUSUM chart based on different estimators under 10% variance contaminated normal distribution with $k_{\hat{\theta}} = 0.5$ and $h_{\hat{\theta}} = 4.774$

Subgroup size	Estimator	δ						
		0	0.25	0.5	0.75	1	1.5	2
$n = 5$	**Mean**	296.883	51.619	13.442	7.054	4.816	2.984	2.226
	Median	340.203	49.462	12.875	6.927	4.704	2.924	2.222
	Midrange	222.020	99.674	29.238	13.460	8.273	4.683	3.340
	HL	306.853	42.497	11.465	6.248	4.288	2.723	2.077
	TM	335.199	41.117	11.131	6.152	4.195	2.688	2.067
$n = 10$	**Mean**	329.818	25.775	7.771	4.482	3.196	2.141	1.727
	Median	356.416	22.899	7.161	4.213	3.026	2.057	1.624
	Midrange	242.112	113.064	34.908	15.300	9.298	5.152	3.642
	HL	342.307	20.108	6.486	3.853	2.785	1.961	1.475
	TM	361.443	20.064	6.473	3.807	2.774	1.950	1.464

Location contaminated normal environment: A $(\varphi)100\%$ location contaminated normal distribution is the one that contains $(1-\varphi)100\%$ observations from $N(\mu_0, \sigma_0^2)$ and $(\varphi)100\%$ observations from $N(\mu_0 + \omega\sigma_0, \sigma_0^2)$ where $-\infty < \omega < \infty$. The ARL values of the CUSUM chart using different estimators for location contaminated normal environment are given in Tables 5.8 with $\varphi = 0.05$ and $\omega = 4$.

Table 5.8 shows none of the estimators is able to adequately detect the location contamination in the process when $n = 5$, as the ARL_0 for all the estimators is substantially lower than that for the uncontaminated environment. Increasing the subgroup size may be a better option as the midrange and median CUSUM charts have a reasonable ARL_0 for $n = 10$ but the ARL_1 performance of the midrange CUSUM chart is way too poor as compared to the median CUSUM.

Table 5.8: ARL values for the CUSUM chart based on different estimators under 5% location contaminated normal distribution with $k_{\hat{\theta}} = 0.5$ and $h_{\hat{\theta}} = 4.774$

Subgroup size	Estimator	δ						
		0	0.25	0.5	0.75	1	1.5	2
$n = 5$	**Mean**	317.549	48.560	13.543	7.138	4.801	2.973	2.222
	Median	301.845	49.447	13.245	6.928	4.682	2.933	2.210
	Midrange	315.737	78.997	25.153	12.361	7.818	4.429	3.124
	HL	285.025	45.634	12.369	6.565	4.485	2.806	2.124
	TM	297.391	43.218	11.775	6.330	4.372	2.742	2.085
$n = 10$	**Mean**	344.932	24.938	7.639	4.450	3.155	2.106	1.710
	Median	357.296	22.676	7.076	4.160	2.979	2.038	1.601
	Midrange	391.235	78.931	24.038	11.693	7.523	4.367	3.097
	HL	303.981	21.554	6.811	3.978	2.879	1.986	1.543
	TM	329.528	20.369	6.464	3.869	2.793	1.957	1.489

Table 5.9: ARL values for the CUSUM chart based on different estimators under special cause normal distribution with $\varphi = 0.01$, $k_{\hat{\theta}} = 0.5$ and $h_{\hat{\theta}} = 4.774$

Subgroup size	Estimator	δ						
		0	0.25	0.5	0.75	1	1.5	2
$n = 5$	**Mean**	221.158	52.954	12.241	6.338	4.295	2.699	2.065
	Median	365.619	41.318	11.252	6.094	4.230	2.671	2.052
	Midrange	168.113	132.188	48.782	15.514	8.731	4.740	3.301
	HL	356.893	31.754	9.094	5.156	3.607	2.359	1.888
	TM	364.754	33.872	9.704	5.368	3.762	2.418	1.921
$n = 10$	**Mean**	211.164	23.559	6.853	3.989	2.861	1.987	1.552
	Median	360.264	19.982	6.552	3.834	2.798	1.956	1.477
	Midrange	124.994	104.625	75.651	27.395	12.790	6.089	4.115
	HL	367.158	15.936	5.460	3.345	2.465	1.784	1.215
	TM	375.058	16.858	5.690	3.464	2.546	1.834	1.281

Table 5.10: ARL values for the CUSUM chart based on different estimators under special cause normal distribution with $\varphi = 0.05$, $k_{\hat{\theta}} = 0.5$ and $h_{\hat{\theta}} = 4.774$

Subgroup size	Estimator	δ						
		0	0.25	0.5	0.75	1	1.5	2
$n = 5$	**Mean**	132.301	80.360	32.039	13.488	7.879	4.404	3.095
	Median	354.591	45.811	12.137	6.518	4.436	2.803	2.130
	Midrange	104.606	90.670	76.205	58.787	38.613	14.055	7.940
	HL	292.820	45.662	11.647	6.180	4.230	2.680	2.056
	TM	322.227	38.766	10.660	5.794	4.076	2.573	2.008
$n = 10$	**Mean**	148.795	52.119	14.593	7.137	4.698	2.896	2.118
	Median	368.855	21.345	6.730	3.980	2.866	1.994	1.534
	Midrange	105.884	91.084	75.093	62.220	48.336	24.478	12.724
	HL	345.449	18.332	6.031	3.625	2.645	1.878	1.357
	TM	355.257	18.601	6.020	3.639	2.656	1.889	1.366

Special cause environment: Asymmetric variance disturbances are created in which each observation is drawn from $N(0,1)$ and has a φ probability of having a multiple of a $\chi^2_{(1)}$ variable added to it, with multiplier equal to 4. $ARLs$ of the CUSUM chart under this environment with $\varphi = 0.01$ and $\varphi = 0.05$ are given in Tables 5.9 and 5.10, respectively.

In presence of this special cause, the median CUSUM seems more robust while the midrange CUSUM is affected the most. TM and HL CUSUM has good detection ability with a reasonable ARL_0. TM and HL estimators are affected positively by the increase in subgroup size, i.e. their ARL_0 increase as we increase the value of n and vice versa.

Non-normal environments: To investigate the effect of using non-normal distributions we consider two cases: one by changing the kurtosis and the other by changing the symmetry of the distribution. For the case of disturbing the kurtosis we use Student's t distribution with 4 degrees of freedom (T_4) and the logistic distribution $(\text{Logis}(0,1))$, and for the disturbance in symmetry we use the chi-square distribution with 5 degrees of freedom $(\chi^2_{(5)})$. Tables 5.11 – 5.13 contains the ARL values for the proposed CUSUM charts under T_4, $\text{Logis}(0,1)$ and $\chi^2_{(5)}$, respectively, where the ARL_0 is kept fixed at 370.

Table 5.11: ARL values for the CUSUM chart based on different estimators under T_4 distribution with $n = 10$ and $k_{\hat{\theta}} = 0.5$ at $ARL_0 \cong 370$

Estimator	$h_{\hat{\theta}}$	δ						
		0	0.25	0.5	0.75	1	1.5	2
Mean	4.99	371.243	30.619	8.690	4.933	3.511	2.309	1.877
Median	4.83	371.753	24.568	7.443	4.366	3.123	2.110	1.682
Midrange	5.80	369.540	205.442	53.968	20.771	12.092	6.505	4.501
HL	4.846	371.437	22.654	7.019	4.138	2.986	2.043	1.607
TM	4.84	370.038	22.075	6.877	4.055	2.932	2.024	1.581

For the case of Student's t distribution TM CUSUM is performing the best among all of the other estimators followed by the HL CUSUM. Median CUSUM has also reasonable performance as compared to the others, but the midrange CUSUM seems to have worst performance for the said case (cf. Table 5.11). For the logistic distribution the HL and TM CUSUM outperform the median and midrange CUSUM charts, while the mean CUSUM reasonably maintains its performance (cf. Table 5.12). Similarly, TM and mean CUSUM charts show very good performance in case of chi-square distribution. HL is also performing well whereas midrange CUSUM has the worst performance (cf. Table 5.13).

Table 5.12: ARL values for the CUSUM chart based on different estimators under Standard Logistic distribution with $n = 10$ and $k_{\hat{\theta}} = 0.5$ at $ARL_0 \cong 370$

Estimator	$h_{\hat{\theta}}$	δ						
		0	0.25	0.5	0.75	1	1.5	2
Mean	4.790	371.335	47.114	12.424	6.671	4.541	2.860	2.166
Median	4.817	369.302	52.499	13.795	7.246	4.944	3.075	2.304
Midrange	4.970	370.869	126.070	35.936	15.877	9.760	5.415	3.808
HL	4.817	368.761	45.081	12.161	6.498	4.479	2.819	2.139
TM	4.813	371.226	45.926	12.258	6.630	4.493	2.817	2.150

Table 5.13: ARL values for the CUSUM chart based on different estimators under Chi-square distribution with $n = 10$ and $k_{\hat{\theta}} = 0.5$ at $ARL_0 \cong 370$

Estimator	$h_{\hat{\theta}}$	δ						
		0	0.25	0.5	0.75	1	1.5	2
Mean	4.85	370.487	115.012	36.877	17.094	10.376	5.630	3.937
Median	4.90	369.970	127.209	44.319	20.670	12.017	6.527	4.454
Midrange	5.19	369.765	196.527	96.916	49.213	28.913	13.261	8.395
HL	4.88	371.978	117.913	37.105	17.439	10.544	5.796	4.019
TM	4.87	370.087	115.104	37.648	17.085	10.494	5.760	4.000

At the end of this section, we provide the ARL curves of the CUSUM charts with different estimators under different environments discussed above. Figures 5.1 – 5.5 contain the ARL curves of the CUSUM charts based on different estimators with $n = 10$, $k_{\hat{\theta}} = 0.5$ and $h_{\hat{\theta}} = 4.774$. ARL curves of the CUSUM charts based on different estimators in form of figures are not given here and can be seen in Nazir, Riaz, Does and Abbas (2013).

From Figures 5.1 – 5.5, we see that the $ARLs$ of the midrange CUSUM are affected the most in case of some non-normal, contaminated and special cause environments. The mean CUSUM is also influenced badly under special cause environments. The ARL_0 values of the median, TM and HL CUSUMs seem less affected by the change of parent normal environment.

For a graphical comparison of the proposed charts with non-normal environments, the ARL curves of all the charts under T_4, Logis$(0,1)$ and $\chi^2_{(5)}$ distributions are given in Figures 5.6, 5.7 and 5.8 respectively, with $k_{\hat{\theta}} = 0.5$ and ARL_0 fixed at 370. Figures 5.6-5.8 are not given here and can be seen in Nazir, Riaz, Does and Abbas (2013). These figures clearly indicate that in general TM and HL CUSUMs are performing better than the other estimators under Student's t and Logistic distributions. All the estimators (except midrange) are performing equally well in case of chi-square parent environment.

Table 5.14: ARL values for the NPCUSUM chart under different environments with $n = 10$, $k = 0.5$ and $h = 10.65$

Environment	δ						
	0	0.25	0.5	0.75	1	1.5	2
Normal	371.051	20.492	8.266	5.465	4.274	3.33	3.03
5% Variance Contaminated	370.868	21.764	8.626	5.618	4.422	3.444	3.082
10% Variance Contaminated	370.739	22.915	9.037	5.866	4.571	3.548	3.156
5% Location Contaminated	172.225	15.595	7.558	5.205	4.162	3.289	3.025
Special Cause ($\varphi = 0.01$)	367.505	19.645	8.157	5.407	4.253	3.331	3.029
Special Cause ($\varphi = 0.05$)	243.051	16.922	7.716	5.244	4.183	3.307	3.029
T_4	370.492	22.744	9.075	5.94	4.685	3.682	3.263
Logis(0,1)	371.779	43.808	15.06	8.956	6.624	4.604	3.814

Finally, we provide the comparison of our proposed CUSUM charts with the non-parametric CUSUM mean chart (NPCUSUM) by Yang and Cheng (2011) under different environments discussed above in this section. Yang and Cheng (2011) calculated the $ARLs$ of NPCUSUM using the shift parameter p_1. For a valid comparison of NPCUSUM with our proposed charts, we have evaluated the ARL values of NPCUSUM chart using the shift parameter as δ which is the difference between μ_0 and μ_1. The $ARLs$ are calculated through

Monte Carlo simulations by doing $10{,}000$ replications of run lengths and are given in Table 5.14.

A general comparison reveals that the NPCUSUM is performing better than all the proposed charts for Logis$(0,1)$ while for all other scenarios, the proposed median, HL and TM CUSUMs have better performance than the NPCUSUM chart, across all scenarios and for most values of δ.

5.5 SUMMARY AND CONCLUSIONS

Natural variations are an inherent part of any process and by timely monitoring, evaluating and identifying sources of un-natural variations, the quality of the output of the process can be improved and waste (of time and cost) can be reduced significantly. Control charts are widely used to monitor a process. For monitoring the location and dispersion parameters of the process, two main types of charts, named as Shewhart-type control charts and memory control charts (CUSUM and EWMA), are used. In practice un-natural variations, special cause environments and outliers may be occasionally present. The charts having a robust design structure are used to cope with such environments. This chapter presents different robust design structure CUSUM-type charts and evaluates their performances in different environments. The findings of the chapter are that the mean CUSUM control chart performs efficiently in uncontaminated environments and the TM and HL CUSUM charts are good as compared to the mean CUSUM chart in this situation. The median CUSUM chart outperforms in relation to the other charts in the presence of special cause environment and outliers. The TM CUSUM chart is an alternative of median chart and is highly efficient in the presence of outliers. It is also a good option in non-normal parent environments. Finally, it may be concluded that the TM CUSUM chart is the best choice of controlling the location parameter of a process in normal, non-normal, special cause and outliers environments.

Chapter 6

Robust Memory-Type Control Charts for Location

Cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) control charts are commonly used to detect small changes in the parameters of production processes. Recently, a new control structure was introduced, named as mixed EWMA-CUSUM control chart, which combined both charts. The current study provides a detailed comparison of these three types of control charts under normal and contaminated normal environments. Performance measures like the (average) run length and the extra quadratic loss are used for comparison purposes. We investigate six different location estimators with the structures of the three memory charts and study their robustness properties. The results of this chapter have been submitted by Nazir, Abbas et al. (2014).

6.1. INTRODUCTION

CUSUM and EWMA charts are designed such that they use the past information along with the current information, which makes them very sensitive to shifts of small and moderate magnitudes in the process parameters. A number of modifications of the CUSUM and EWMA charts have been developed to further enhance the performance of these charts. Some of these enhancements may be seen in Lucas (1982), Lucas and Saccucci (1990), Steiner (1999), Capizzi and Masarotto (2003), Zhao et al. (2005), Machado and Costa (2008), Riaz et al. (2011), Abbas et al. (2011) and the references therein. Following these authors, Abbas et al. (2013) proposed a mixed EWMA-CUSUM control chart and concluded that mixing the two charts makes the proposed scheme even more sensitive to small shifts in the process mean as compared to the other schemes designed for similar purposes.

In practice, process parameters are unknown and they need to be estimated from samples, which are assumed to be in state of statistical control. Woodall and Montgomery (1999) name this stage as Phase I. The resultant estimates from Phase I establish the control limits that are used to monitor the process parameter of interest in the next stage: Phase II.

Jensen et al. (2006) studied estimation effects on control chart properties in Phase I and their Phase II impact. Schoonhoven et al. (2011) used different robust estimators for the location control chart in a Shewhart set up by considering limited information in Phase I and observed the performance of these estimators in Phase II. Recently, Nazir et al. (2013) proposed the use of some robust location estimators with the control structures of the CUSUM chart, in order to increase the robustness of the CUSUM chart against contamination or non-normality. However, they only considered the situation where a large number of samples are available in Phase I and they did not take into account the estimation effects of parameters.

The concern of this chapter is to assess the estimation effects of process parameters in Phase I and to check the impact and influence of robust estimators of the location parameter on the Phase II performance with the design structures of CUSUM, EWMA and mixed EWMA-CUSUM control charts under different environments. Generally, the performance and efficiency of control charts is assessed by the determinant, called average run length (ARL). The ARL is the mean of a random variable called run length (RL), where RL is the number of samples required before an alarm signal occurs. The in-control ARL is likely to be high, but can be fixed to a specific number for a given false alarm rate and is denoted by ARL_0. The out-of-control ARL is expected to be as small as possible and is denoted by ARL_1.

Wu et al. (2009), Ou et al. (2012) and Ahmad et al. (2013) highlighted some of the drawbacks of the ARL as it gives only the performance of a control chart for a specific shift size. Hence, they recommended measures which evaluate the performance of a control chart over a range of shifts. This measure is named as Extra Quadratic Loss (EQL) and is defined as:

$$EQL = \frac{1}{\delta_{\max} - \delta_{\min}} \int_{\delta_{\min}}^{\delta_{\max}} \delta^2 ARL(\delta) d\delta \tag{6.1}$$

where δ indicates the size of the shift (i.e. when the location parameter is shifted from the target value θ_0 to $\theta_0 + \delta\sigma$). When $\delta = 0$, the location parameter θ of the process is in control, otherwise the location parameter has changed and needs to be detected.

Besides using the ARL and EQL as efficiency indicators we will also take into account the RL distribution in order to have an even better comparison.

In the next section, we give the details regarding the control structure of three memory-type control charts and the Phase I and Phase II estimators. Design structures of these charts are provided in section 3. In section 4 we provide a comprehensive comparison of the three memory-type control charts (based on six different location estimates) in terms of ARLs, EQLs and RL distribution to study their robustness. Finally, the chapter is concluded in section 5.

6.2. CUSUM, EWMA AND MIXED EWMA-CUSUM CHARTS

Shewhart-type control charts are less efficient to detect small and moderate shifts in the process parameter(s). For that reason, some memory-type control charts are proposed. The most important ones include the CUSUM, EWMA and mixed EWMA-CUSUM charts, and the current section contains the details about these three structures.

Cumulative sum charts: Page (1954) presented the idea of accumulating the positive and negative deviations from the process location in two different statistics $\boldsymbol{C_i^+}$ and $\boldsymbol{C_i^-}$, respectively. These two statistics are defined as:

$$C_i^+ = \max\left[0, \left(\hat{\theta}_i - \theta_0\right) - K_{\hat{\theta}} + C_{i-1}^+\right], \quad C_i^- = \max\left[0, -\left(\hat{\theta}_i - \theta_0\right) - K_{\hat{\theta}} + C_{i-1}^-\right] \tag{6.2}$$

where i is the sample number and $\hat{\theta}$ is the location estimator used to monitor the process location parameter. The initial values for both of the statistics given in (6.2) are usually taken equal to the target value θ_0, i.e. $C_0^+ = C_0^- = \theta_0$. The statistics C_i^+ and C_i^- are plotted against the control limit $H_{\hat{\theta}}$ and an out-of-control signal is generated if either one of these statistics crosses the control limit. The standardized versions of the chart parameters ($K_{\hat{\theta}}$ and $H_{\hat{\theta}}$) are given as:

$$K_{\hat{\theta}} = k \times \sigma_{\hat{\theta}}, \quad H_{\hat{\theta}} = h \times \sigma_{\hat{\theta}} \tag{6.3}$$

Here k and h are the constants which are chosen to satisfy a pre-specified ARL_0.

Exponentially weighted moving average charts: Roberts (1959) proposed a control charting scheme, in which the plotting statistic is split into two components (i.e. present information and past information), and named it as the exponentially weighted moving average (EWMA) chart. The weights are assigned to the observations such that these weights decrease exponentially for the more dated observations. The control structure of the EWMA chart, consisting of a plotting statistic and the control limits, is given as:

$$Z_i = \lambda\hat{\theta}_i + (1-\lambda)Z_{i-1} \tag{6.4}$$

$$\left.\begin{aligned} LCL_i &= \theta_0 - L_{\hat{\theta}}\sqrt{\mathrm{Var}(\hat{\theta}) \times \tfrac{\lambda}{2-\lambda}\left(1-(1-\lambda)^{2i}\right)} \\ CL &= \theta_0 \\ UCL_i &= \theta_0 + L_{\hat{\theta}}\sqrt{\mathrm{Var}(\hat{\theta}) \times \tfrac{\lambda}{2-\lambda}\left(1-(1-\lambda)^{2i}\right)} \end{aligned}\right\} \tag{6.5}$$

where $\lambda \in (0,1]$ is the smoothing parameter of the chart. The initial value for the above plotting statistic in (6.4) is usually taken equal to the target value, i.e. $Z_0 = \theta_0$. $L_{\hat{\theta}}$ is the control limit coefficient and can be chosen to satisfy the pre-specified ARL_0. Note that for $\lambda = 1$, we obtain the Shewhart control chart and hence the Shewhart control chart is a special case of the EWMA control chart.

Mixed EWMA-CUSUM charts: Abbas et al. (2013) proposed a mixture of the CUSUM and EWMA charts and named it as the mixed EWMA-CUSUM chart. The two plotting statistics ($\boldsymbol{M_i^+}$ and $\boldsymbol{M_i^-}$) for this chart are given as:

$$M_i^+ = \max\left[0, (Z_i - \theta_0) - A_{\hat{\theta},i} + M_{i-1}^+\right], \quad M_i^- = \max\left[0, -(Z_i - \theta_0) - A_{\hat{\theta},i} + M_{i-1}^-\right] \tag{6.6}$$

where Z_i is defined as in (6.4). The statistics (given in (6.6)) are plotted against the control limit $B_{\hat{\theta},i}$ and an out-of-control signal is detected if either one of these statistics crosses the control limit. The standardized versions $A_{\hat{\theta},i}$ and $B_{\hat{\theta},i}$ are given as:

$$A_{\hat{\theta},i} = a \times \sqrt{\mathrm{Var}(\hat{\theta}) \times \frac{\lambda}{2-\lambda}\left(1-(1-\lambda)^{2i}\right)}, \ B_{\hat{\theta},i} = b \times \sqrt{\mathrm{Var}(\hat{\theta}) \times \frac{\lambda}{2-\lambda}\left(1-(1-\lambda)^{2i}\right)} \tag{6.7}$$

where a and b are constants like k and h in (3), respectively.

Previous equations (6.2)–(6.7) include θ_0 and $\mathrm{Var}(\hat{\theta})$. Consider that X_{ij}, $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, m$ denote the Phase I data, when the process is in an in-control state and let Y_{ij}, $i = 1,2, \dots, n$ and $j = 1,2, \dots$, denote the Phase II data.

We assume that the X_{ij} are normally distributed with mean θ_0 and variance σ^2, i.e. $N(\theta_0, \sigma^2)$. The unknown location parameter θ_0 is estimated from the mean of the sample means, i.e.

$$\hat{\theta}_0 = \bar{\bar{X}} = \frac{1}{m}\sum_{j=1}^{m} \bar{X}_j = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n} X_{ij}\right) \tag{6.8}$$

and the unknown dispersion parameter σ is based on the pooled sample standard deviation,

$$S_p = \left(\frac{1}{m}\sum_{j=1}^{m} S_j^2\right)^{1/2}$$

where S_j^2 is the jth sample variance defined by

$$S_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2$$

An unbiased estimator of σ is given by

$\hat{\sigma} = S_p / c_4(m(n-1)+1)$ (6.9)

where $c_4(q)$ is defined by

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

Note that $\text{Var}(\hat{\theta})$ is a function of dispersion parameter σ which is unknown and has to be estimated. The estimates $\hat{\theta}_0 = \bar{\bar{X}}$ and $\hat{\sigma} = S_p / c_4(m(n-1)+1)$ will be used in (6.2)-(6.7), respectively, instead of θ_0 and σ for the construction of the control limits in Phase I.

In Phase II we assume that Y_{ij} are independent and equally distributed as the X_{ij}, with the only difference that the location parameter may be shifted.

The sample mean $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ is one of the estimators of the population location that can replace $\hat{\theta}$ in (6.2) – (6.7) in Phase II. However, there are many other estimators that can also be used instead of $\hat{\theta}$ with the above mentioned three memory charts structures. The first estimator we consider out of those is the sample median. The sample median is defined as the middle order statistic $\tilde{Y} = Y_{\left(\frac{n+1}{2}\right)}$ for odd sample sizes and the average of the two middle order statistics $\tilde{Y} = \frac{1}{2}\left(Y_{\left(\frac{n}{2}\right)} + Y_{\left(\frac{n+2}{2}\right)}\right)$ in case of even sample sizes. The sample median is a robust estimator, because it is least affected by outliers (cf. Dixon and Massey (1969)). The next estimator is the sample midrange and is defined as $MR = \frac{Y_{(1)}+Y_{(n)}}{2}$, where $Y_{(1)}$ and $Y_{(n)}$ are the lowest and highest order statistics in a random sample of size n. It is highly sensitive to outliers as its design structure is based on only extreme values of data (cf. Ferrell (1953) for more details). We also include the estimator based on the median of the pairwise Walsh averages, which is defined as: $HL = \text{median}\big((Y_j + Y_k)/2\,, 1 \le j \le k \le n\big)$. The main advantage of the HL estimator is that it is robust against outliers in a sample. For more properties of HL see Hettmansperger and McKean (1998). The estimator HL is also known as the Hodges-Lehmann estimator. The next estimator included in this study is the trimean of a sample, which is the weighted average of the sample median and two quartiles and is defined as: $TM = \frac{Q_1+2Q_2+Q_3}{4}$, where Q_p $(p = 1,2,3)$ denote one of the three quartiles in a sample. For detailed properties of trimean (TM) see Wang et al. (2007). The last estimator used in this study is sample trimmed mean and is defined as $T_R M = \frac{1}{n-2T}\sum_{i=T+1}^{n-T} Y_{(i)}$, where $2T$ is the number of trimmed values and $Y_{(i)}$ is the ith order statistic in a sample of size n. We take $T = 1$ for $n = 5$ and $T = 2$ for $n = 10$, respectively.

Under normality, the means and (asymptotic) variances (cf. Song et al. (1982), Caperaa and Rivest (1995), Khattree and Rao (2003) and Wang et al. (2007)) of these estimators are given in Table 6.1.

Table 6.1: Expected values and the asymptotic variances of the estimators under in-control situation

Estimator	Expected value of the estimator	(asymptotic) variance of the estimator
Mean	θ_0	$\frac{\sigma^2}{n}$
Median	θ_0	$\frac{\pi\sigma^2}{2n}$
Midrange	θ_0	$\frac{\pi^2\sigma^2}{24\ln(n)}$
Hodges-Lehmann (HL)	θ_0	$\frac{\pi\sigma^2}{3n}$
Trimean	θ_0	$\frac{\pi\sigma^2}{2.61n}$
Trimmed mean	θ_0	$\frac{n(\sigma^2-2.9565)}{(n-2)^2}$ for $n = 5$ $\frac{n(\sigma^2-5.9209)}{(n-4)^2}$ for $n = 10$

6.3. DESIGN AND DERIVATION OF PHASE II LIMITS OF THE CHARTS

The design of the Phase II control charts involves a derivation of different factors: the CUSUM structure requires values of k and h (cf. (6.3)), the EWMA scheme needs λ and L (cf. (6.5)), and the mixed EWMA-CUSUM demands values of a and b (cf. (6.7)) for the construction of the control limits of these charts. Along with these factors, the Phase I process location parameter θ and dispersion parameter σ also have to be estimated.

We derive these factors in such a way that we obtain the intended value of $ARL_0 = 370$. The Phase I estimators are $\hat{\theta}_0 = \bar{\bar{X}}$ and $\hat{\sigma} = S_p/c_4(m(n-1)+1)$. We employ different estimators for the Phase II plotting statistic(s) and adopt the following settings, $k = 0.5$, $\lambda = 0.13$ and $a = 0.5$ as optimal constants to detect a shift size of $\delta = 1$ (i.e. a shift of 1σ), for respectively, CUSUM, EWMA and mixed EWMA-CUSUM charts, taking inspiration from Lucas (1982), Crowder (1989), and Abbas et al. (2013). We simulate the factors h for the CUSUM, L for the EWMA and b for the mixed EWMA-CUSUM control chart by considering $m = 50$ subgroups of sizes $n = 5$ and $n = 10$ from an uncontaminated normal environment with desired $ARL_0 = 370$. The values of these factors are given in Table 6.2.

Table 6.2: Factors of CUSUM, EWMA and mixed EWMA-CUSUM charts under uncontaminated normal environment with $m = 50$ at $ARL_0 = 370$

n	Chart	Phase II Estimators					
		Mean	Median	Midrange	HL	Trimean	Trimmed
5	CUSUM	$h = 5.000$	$h = 4.532$	$h = 5.058$	$h = 5.110$	$h = 4.480$	$h = 4.520$
	EWMA	$L = 2.895$	$L = 2.740$	$L = 2.912$	$L = 2.926$	$L = 2.790$	$L = 2.760$
	Mixed	$b = 36.30$	$b = 31.76$	$b = 35.90$	$b = 36.82$	$b = 32.20$	$b = 33.98$
10	CUSUM	$h = 5.078$	$h = 4.440$	$h = 5.168$	$h = 5.116$	$h = 4.680$	$h = 4.827$
	EWMA	$L = 2.916$	$L = 2.714$	$L = 2.935$	$L = 2.929$	$L = 2.790$	$L = 2.847$
	Mixed	$b = 36.70$	$b = 31.08$	$b = 35.26$	$b = 36.78$	$b = 33.50$	$b = 35.14$

6.4. PERFORMANCE EVALUATION OF MEMORY CONTROL CHARTS

This section gives the details regarding the performance evaluation of the three memory control charts for normal and contaminated normal environments.

The performance of the design structures is measured in terms of their ARLs and $EQLs$. As a baseline we use the most conventional in-control ARL, i.e. $ARL_0 = 370$, under normality for the performance evaluation and comparisons. The evaluation of ARL values is done using Monte Carlo simulations.

Now for the Phase II analysis, the estimated Phase I process location parameter θ and dispersion parameter σ are used for constructing the control limits of all the charts (where the control limits of the CUSUM are given in (6.3), the control limits of the EWMA are given in (6.5) and the control limits of the mixed EWMA-CUSUM control chart are given in (6.7)). Then, by applying the out-of-control condition (i.e. when an out-of-control signal occurs), we have noted the sample number where the plotting statistic crosses the control limits. This noted number is called run length which is replicated 10^5 times to get the run length distribution. The mean of that distribution, when the location has not been changed, is known as ARL_0. After that, we introduce different amounts of shifts in the process while keeping the control limits the same as we have used for the in-control case. It results in an evaluation of the ARL_1 performance of all the charts.

Note that the performance measure EQL is based upon the ARL_1 values and therefore EQL can be used for measuring the efficiency of a chart (cf. (6.1)). For robustness comparison, ARL_0 is used.

6.4.1. Normal and contaminated normal environments

The description of the environments, for which the performance of the CUSUM, EWMA and mixed EWMA-CUSUM control charts is evaluated, is given as follows:

Normal environment: Here, we provide the performance of the memory charts under a perfectly normal environment with mean θ and variance σ^2, i.e. $N(\theta, \sigma^2)$. Without loss of generality, we use $\theta = 0$ and $\sigma = 1$ throughout this article.

Diffuse symmetric variance contaminated normal environment: Here, $(1-\alpha)100\%$ observations in a sample come from the standard normal distribution, i.e. $N(0,1)$, and

$(\alpha)100\%$ observations of that sample are from $N(0,4)$, i.e. a normal distribution with inflated variance.

Localized variance contaminated normal environment: Here, a sample of size n with probability $(1-\alpha)100\%$ come from the standard normal distribution, i.e. $N(0,1)$, and otherwise a sample with probability $(\alpha)100\%$ is from $N(0,4)$, i.e. a normal distribution with inflated variance.

Diffuse asymmetric variance contaminated normal environment: Here, asymmetric variance disturbances are introduced in the process, i.e. $(1-\alpha)100\%$ observations in a sample come from the standard normal distribution, i.e. $N(0,1)$, and $(\alpha)100\%$ observations are from $N(0,1)$ having a multiple of a $\chi^2_{(1)}$ variable added to it with multiplier equal to 4.

Diffuse mean contaminated normal environment: Here, $(1-\alpha)100\%$ observations in a sample come from standard normal distribution, i.e. $N(0,1)$, and $(\alpha)100\%$ observations are from $N(4,1)$, i.e. the normal distribution with an inflated mean.

Localized mean contaminated normal environment: In this scenario, a sample of size n with probability $(1-\alpha)100\%$ come from the standard normal distribution, i.e. $N(0,1)$, and otherwise a sample with probability $(\alpha)100\%$ is from $N(4,1)$, i.e. a normal distribution with inflated mean.

Note that these environments are commonly used in these robust studies (cf. Schoonhoven et al. (2011) and Nazir et al. (2013)). We take in the comparisons $\alpha = 0.05$.

6.4.2. Performance comparison of the classical and robust control structures

In Section 2 we have described the 6 different location estimators which will be used in this study. In this subsection we will evaluate these estimators with the design structures of the CUSUM, EWMA and mixed EWMA-CUSUM control charts. The ARL and EQL based comparisons of the charts under the different environments discussed in 6.4.1 are given in Tables 6.3-6.9.

Keeping in mind that ARL_0 is a measure for robustness and EQL is a measure of efficiency, the following points cover the findings of Tables 6.3-6.9:

1. **Normal environment**: As we have explained earlier we have taken $ARL_0 = 370$. It may be concluded from Table 6.3 that, if there is no shift, indeed the ARL_0 is around 370. Clearly, the EQL performance of the EWMA control chart is best among the three control charts. The reason behind this superiority is that the formula of EQL (see (6.1)) gives a weight δ^2 to the ARL values and therefore the $ARLs$ for larger shifts have greater weights. This makes the EWMA control chart dominant over the CUSUM and mixed EMWA-CUSUM charts (in terms of $EQLs$) as the $ARLs$ for EWMA are smaller for larger values of δ. As far as the ARL performance is concerned, the EWMA chart is better than the other two charts (cf. Table 6.3). Under the uncontaminated normal environment, as it was expected, the sample mean performs best with the design structures of the CUSUM, EWMA and mixed EWMA-CUSUM charts as compared to all other estimators used (cf. Table 6.3). For small shifts in the process, i.e. $\delta = 0.25$, the EWMA chart with the trimmed mean estimator performs well, followed by the HL estimator with the EWMA structure. For small sample sizes, the mixed EWMA-CUSUM chart is slightly better than the CUSUM chart (cf. Table 6.3).
2. **Diffuse symmetric variance contaminated normal environment:** We see from Table 6.4, that the ARL_0 of the mixed EWMA-CUSUM chart with the HL estimator is not much affected by the presence of a diffuse symmetric variance contamination in the distribution. On the other hand, the in-control $ARLs$ for the other charts are disturbed and the gap between the pre-fixed ARL_0 (370) and the attained ARL_0 is

mostly significant. In most cases the number of false alarms has increased when there is a diffuse symmetric variance contamination present in the data but $\delta = 0$. The effect of this kind of variance contamination on the design structure of these charts is obvious. For example, the in-control ARL_0 of the CUSUM, EWMA and mixed EWMA-CUSUM charts with the sample mean as location estimator decreases respectively by 44.58%, 40.53% and 20.93% for $n = 5$ and approximately the same decrease can be seen for $n = 10$. This shows that the mixed EWMA-CUSUM is more robust to diffuse symmetric variance contaminations as compared to the CUSUM and EWMA charts. In the other words, the mixed EWMA-CUSUM reacts when there is a shift in the location parameter and does not react unnecessarily in the presence of variance contamination when we assume that this variance contamination is a part of the in-control process.

3. **Localized variance contaminated normal environment:** Table 6.5 reads that the in-control ARLs of the mixed EWMA-CUSUM chart is less affected as compared to the in-control ARLs of the CUSUM and EWMA charts. The influence of the estimator is limited for all charts when there are localized variance contaminations (the trimmed mean has the worst performance in the in-control situation). The CUSUM chart is producing many false alarms when the process is in-control. According to the EQL determinant, the EWMA chart is very effective in finding out-of-control situations.

4. **Diffuse asymmetric variance contaminated normal environment:** From Table 6.6, we see that the in-control ARLs of the CUSUM, EWMA and the mixed EWMA-CUSUM charts are highly affected by the presence of such type of contaminations. The best in-control behavior is obtained with the median as estimator. Again the mixed EWMA-CUSUM chart has the best in-control *ARLs* and the EWMA chart is

more efficient in detecting shifts as the corresponding EQLs are the smallest (cf. Table 6.6).

Table 6.3: ARL and EQL values of CUSUM, EWMA and mixed EWMA-CUSUM charts with $m = 50$ at $ARL_0 = 370$ under uncontaminated normal environment

n	Chart	Estimator	δ									EQL
			0	0.25	0.5	0.75	1	1.5	2	3	5	
5	CUSUM	Mean	370.960	39.833	9.073	5.043	3.551	2.322	1.861	1.124	1.000	10.798
		Median	370.186	56.332	12.110	6.264	4.264	2.682	2.055	1.374	1.000	11.877
		Midrange	369.489	50.029	11.147	5.973	4.118	2.633	2.039	1.358	1.000	11.766
		HL	371.060	42.524	9.614	5.314	3.714	2.413	1.920	1.188	1.000	11.071
		Trimean	371.919	43.767	9.487	5.125	3.571	2.317	1.841	1.116	1.000	10.776
		Trimmed	369.742	42.234	9.212	5.042	3.508	2.287	1.783	1.136	1.000	10.778
	EWMA	Mean	371.264	34.015	7.516	3.691	2.353	1.362	1.061	1.000	1.000	9.548
		Median	369.967	46.430	10.147	4.894	3.058	1.685	1.206	1.003	1.000	9.872
		Midrange	369.669	42.766	9.386	4.561	2.858	1.595	1.157	1.002	1.000	9.773
		HL	370.258	36.090	7.956	3.928	2.475	1.420	1.084	1.000	1.000	9.603
		Trimean	370.297	39.073	8.247	4.027	2.532	1.442	1.089	1.000	1.000	9.627
		Trimmed	369.247	35.086	7.694	3.828	2.433	1.412	1.087	1.000	1.000	9.592
	mixed EWMA-CUSUM	Mean	369.937	37.119	17.361	12.508	10.101	7.589	6.219	4.773	3.179	37.478
		Median	369.702	43.617	19.256	13.650	10.950	8.169	6.675	5.057	3.630	41.223
		Midrange	369.226	42.226	19.098	13.625	10.960	8.195	6.710	5.084	3.692	41.651
		HL	369.097	38.443	17.893	12.861	10.386	7.788	6.381	4.889	3.342	38.857
		Trimean	370.070	37.442	17.238	12.363	9.978	7.468	6.121	4.668	3.103	36.674
		Trimmed	370.697	37.969	17.469	12.524	10.106	7.571	6.206	4.707	3.287	37.854
10	CUSUM	Mean	369.951	17.700	5.521	3.357	2.477	1.802	1.232	1.000	1.000	9.707
		Median	370.690	24.379	6.683	3.886	2.802	1.944	1.457	1.003	1.000	9.959
		Midrange	370.496	31.178	8.398	4.782	3.400	2.252	1.825	1.082	1.000	10.595
		HL	369.035	18.640	5.741	3.471	2.552	1.845	1.293	1.000	1.000	9.769
		Trimean	370.352	19.469	5.766	3.458	2.532	1.814	1.264	1.000	1.000	9.744
		Trimmed	370.627	18.988	5.735	3.450	2.537	1.801	1.283	1.001	1.000	9.753
	EWMA	Mean	371.352	15.161	4.110	2.170	1.471	1.035	1.000	1.000	1.000	9.242
		Median	370.566	20.177	5.304	2.745	1.794	1.132	1.007	1.000	1.000	9.335
		Midrange	370.449	26.687	6.751	3.412	2.191	1.291	1.041	1.000	1.000	9.476
		HL	369.956	15.943	4.316	2.270	1.524	1.046	1.001	1.000	1.000	9.256
		Trimean	368.507	16.440	4.421	2.313	1.554	1.056	1.001	1.000	1.000	9.264
		Trimmed	371.715	16.275	4.399	2.303	1.550	1.059	1.001	1.000	1.000	9.264
	mixed EWMA-CUSUM	Mean	369.621	24.261	13.175	9.752	7.958	6.026	4.978	3.926	2.946	32.367
		Median	370.223	26.953	14.083	10.327	8.394	6.315	5.167	3.990	2.994	33.054
		Midrange	370.631	31.757	16.121	11.733	9.507	7.151	5.881	4.418	3.009	35.165
		HL	371.022	24.904	13.434	9.937	8.105	6.130	5.039	3.964	2.979	32.734
		Trimean	368.809	24.802	13.275	9.793	7.977	6.033	4.978	3.918	2.939	32.313
		Trimmed	370.225	25.018	13.421	9.898	8.070	6.097	5.010	3.874	2.864	31.839

Table 6.4: ARL and EQL values of CUSUM, EWMA and mixed EWMA-CUSUM charts with $m = 50$ at $ARL_0 = 370$ under diffuse symmetric variance contaminated normal environment

n	Chart	Estimator	δ									EQL
			0	0.25	0.5	0.75	1	1.5	2	3	5	
5	CUSUM	Mean	205.586	34.104	9.385	5.261	3.714	2.430	1.925	1.216	1.005	11.166
		Median	280.197	51.814	12.359	6.435	4.385	2.754	2.102	1.416	1.002	12.064
		Midrange	123.295	35.917	11.411	6.352	4.422	2.832	2.189	1.482	1.093	12.764
		HL	250.772	41.438	10.176	5.581	3.896	2.509	1.974	1.256	1.007	11.384
		Trimean	241.763	42.477	10.084	5.427	3.763	2.427	1.896	1.194	1.002	11.102
		Trimmed	258.276	40.950	9.651	5.241	3.655	2.363	1.833	1.179	1.001	10.975
	EWMA	Mean	220.788	29.779	7.384	3.712	2.370	1.380	1.073	1.001	1.000	9.555
		Median	290.172	42.363	10.032	4.889	3.048	1.694	1.211	1.005	1.000	9.871
		Midrange	140.017	32.355	9.020	4.536	2.882	1.625	1.181	1.008	1.000	9.792
		HL	260.660	33.133	7.900	3.916	2.495	1.428	1.092	1.001	1.000	9.606
		Trimean	300.958	35.151	8.143	4.015	2.552	1.454	1.099	1.001	1.000	9.630
		Trimmed	275.763	32.019	7.607	3.828	2.439	1.418	1.097	1.001	1.000	9.595
	mixed EWMA-CUSUM	Mean	292.322	36.840	17.435	12.534	10.119	7.592	6.227	4.767	3.189	37.525
		Median	475.142	46.897	19.939	14.062	11.254	8.392	6.860	5.181	3.728	42.313
		Midrange	280.581	39.496	18.468	13.209	10.638	7.966	6.535	4.962	3.530	40.236
		HL	415.371	39.541	18.153	13.036	10.514	7.888	6.469	4.923	3.457	39.661
		Trimean	414.258	38.303	17.427	12.506	10.080	7.549	6.184	4.710	3.212	37.445
		Trimmed	851.272	47.058	19.921	14.111	11.326	8.465	6.920	5.240	3.747	42.641
10	CUSUM	Mean	212.117	16.842	5.714	3.499	2.588	1.852	1.330	1.007	1.000	9.814
		Median	291.308	23.585	6.827	3.973	2.869	1.977	1.495	1.009	1.000	10.015
		Midrange	78.245	23.243	8.796	5.241	3.767	2.511	2.003	1.305	1.085	11.885
		HL	260.353	19.338	6.027	3.622	2.655	1.886	1.366	1.004	1.000	9.855
		Trimean	273.492	20.145	6.030	3.592	2.625	1.857	1.325	1.003	1.000	9.816
		Trimmed	277.029	19.642	5.975	3.578	2.617	1.840	1.340	1.003	1.000	9.819
	EWMA	Mean	224.377	14.445	4.113	2.194	1.489	1.047	1.001	1.000	1.000	9.246
		Median	301.759	19.693	5.317	2.738	1.804	1.138	1.009	1.000	1.000	9.338
		Midrange	90.298	20.550	6.582	3.454	2.254	1.337	1.071	1.004	1.000	9.509
		HL	274.427	15.485	4.302	2.276	1.532	1.055	1.001	1.000	1.000	9.258
		Trimean	283.828	15.900	4.399	2.325	1.561	1.061	1.001	1.000	1.000	9.265
		Trimmed	288.092	15.799	4.397	2.321	1.558	1.064	1.002	1.000	1.000	9.265
	mixed EWMA-CUSUM	Mean	294.426	24.424	13.203	9.765	7.960	6.023	4.976	3.918	2.938	32.298
		Median	503.884	28.285	14.623	10.678	8.659	6.519	5.333	4.030	2.987	33.327
		Midrange	216.461	28.891	15.075	11.009	8.949	6.732	5.533	4.168	2.982	33.908
		HL	428.580	25.382	13.677	10.105	8.235	6.223	5.104	3.966	2.957	32.701
		Trimean	455.428	25.495	13.578	10.007	8.145	6.152	5.050	3.938	2.926	32.400
		Trimmed	898.986	29.327	15.184	11.128	9.045	6.815	5.604	4.186	3.000	34.133

Table 6.5: ARL and EQL values of CUSUM, EWMA and mixed EWMA-CUSUM charts with $m = 50$ at $ARL_0 = 370$ under localized variance contaminated normal environment

n	Chart	Estimator	δ									EQL
			0	0.25	0.5	0.75	1	1.5	2	3	5	
5	CUSUM	Mean	186.939	34.273	9.005	5.036	3.558	2.325	1.862	1.132	1.000	10.813
		Median	183.256	45.501	11.868	6.240	4.257	2.687	2.065	1.377	1.001	11.876
		Midrange	187.388	41.651	10.939	5.962	4.142	2.640	2.046	1.358	1.001	11.761
		HL	189.957	35.935	9.582	5.313	3.726	2.418	1.921	1.195	1.000	11.084
		Trimean	183.341	36.725	9.414	5.132	3.581	2.324	1.842	1.125	1.000	10.793
		Trimmed	166.674	34.709	9.161	5.043	3.519	2.294	1.786	1.142	1.000	10.786
	EWMA	Mean	206.444	30.071	7.439	3.711	2.371	1.378	1.071	1.002	1.000	9.557
		Median	205.883	39.461	9.932	4.881	3.059	1.698	1.218	1.008	1.000	9.881
		Midrange	207.653	36.833	9.228	4.555	2.870	1.607	1.169	1.005	1.000	9.783
		HL	207.087	31.781	7.884	3.939	2.495	1.433	1.091	1.002	1.000	9.609
		Trimean	235.093	34.154	8.168	4.030	2.546	1.450	1.097	1.002	1.000	9.632
		Trimmed	187.948	30.133	7.663	3.829	2.448	1.422	1.098	1.003	1.000	9.600
	mixed EWMA-CUSUM	Mean	290.984	37.045	17.397	12.540	10.114	7.588	6.225	4.767	3.187	37.508
		Median	291.370	43.480	19.349	13.695	10.957	8.169	6.678	5.057	3.625	41.199
		Midrange	293.450	41.961	19.189	13.645	10.978	8.202	6.716	5.084	3.688	41.636
		HL	291.633	38.454	17.954	12.887	10.397	7.795	6.384	4.885	3.345	38.871
		Trimean	288.958	37.402	17.301	12.398	9.984	7.467	6.125	4.665	3.110	36.711
		Trimmed	289.030	38.031	17.549	12.551	10.119	7.574	6.206	4.708	3.290	37.871
10	CUSUM	Mean	191.302	17.000	5.517	3.374	2.485	1.800	1.240	1.001	1.000	9.715
		Median	184.090	22.743	6.673	3.899	2.814	1.946	1.459	1.007	1.000	9.969
		Midrange	191.757	28.383	8.326	4.799	3.416	2.258	1.824	1.090	1.000	10.616
		HL	190.179	17.692	5.744	3.482	2.561	1.846	1.299	1.002	1.000	9.779
		Trimean	185.838	18.452	5.763	3.470	2.541	1.815	1.271	1.002	1.000	9.752
		Trimmed	179.613	18.033	5.718	3.461	2.543	1.800	1.289	1.002	1.000	9.760
	EWMA	Mean	209.400	14.597	4.121	2.190	1.478	1.044	1.003	1.000	1.000	9.246
		Median	209.208	19.153	5.304	2.747	1.812	1.141	1.012	1.000	1.000	9.341
		Midrange	209.875	24.620	6.724	3.420	2.211	1.304	1.047	1.001	1.000	9.484
		HL	207.123	15.368	4.333	2.285	1.536	1.057	1.003	1.000	1.000	9.261
		Trimean	206.977	15.600	4.407	2.338	1.570	1.063	1.004	1.000	1.000	9.268
		Trimmed	200.900	15.547	4.398	2.327	1.567	1.069	1.005	1.000	1.000	9.269
	mixed EWMA-CUSUM	Mean	293.249	24.346	13.207	9.768	7.963	6.024	4.978	3.920	2.940	32.318
		Median	291.893	27.122	14.117	10.338	8.401	6.320	5.173	3.990	2.991	33.042
		Midrange	294.094	31.787	16.180	11.753	9.517	7.153	5.881	4.422	3.013	35.198
		HL	290.902	24.995	13.473	9.949	8.114	6.133	5.041	3.959	2.976	32.705
		Trimean	291.016	24.873	13.303	9.807	7.983	6.035	4.976	3.910	2.930	32.239
		Trimmed	289.129	25.112	13.446	9.913	8.071	6.098	5.012	3.872	2.860	31.818

5. **Diffuse mean contaminated normal environment:** Table 6.7 shows that the ARL_0s for the CUSUM, the EWMA and the mixed EWMA-CUSUM charts are highly

affected and are substantially lower than the pre-fixed $ARL_0 = 370$. This means that all three charts are unable to maintain their in-control properties under diffuse mean disturbances. Hence none of all are robust in such environment. The best performance is obtained with the mixed EWMA-CUSUM chart and the median as location estimator (cf. Table 6.7).

6. **Localized mean contaminated normal environment**: Under this environment none of the charts and none of the estimators behave properly. Table 6.8 shows for instance that when $\delta = 0$, the in-control ARLs are 18.88, 19.446, and 46.458, respectively, for the CUSUM, EWMA and mixed EWMA-CUSUM charts with the sample mean as estimator.

Table 6.6: ARL and EQL values of CUSUM, EWMA and mixed EWMA-CUSUM charts with $m = 50$ at $ARL_0 = 370$ under diffuse asymmetric variance contaminated normal environment

n	Chart	Estimator	δ									EQL
			0	0.25	0.5	0.75	1	1.5	2	3	5	
5	CUSUM	Mean	24.471	11.733	6.191	4.139	3.115	2.165	1.769	1.107	1.000	10.508
		Median	221.194	33.230	10.131	5.735	4.023	2.602	2.019	1.348	1.000	11.661
		Midrange	15.477	10.459	6.496	4.498	3.441	2.386	1.918	1.307	1.000	11.263
		HL	110.292	20.474	7.749	4.733	3.444	2.321	1.865	1.165	1.000	10.851
		Trimean	37.723	14.067	6.713	4.291	3.177	2.179	1.757	1.102	1.000	10.510
		Trimmed	130.813	22.258	7.672	4.577	3.298	2.212	1.741	1.124	1.000	10.613
	EWMA	Mean	24.765	10.197	4.986	3.014	2.092	1.307	1.053	1.000	1.000	9.417
		Median	229.483	27.581	8.391	4.370	2.840	1.627	1.188	1.003	1.000	9.759
		Midrange	15.167	9.325	5.382	3.457	2.433	1.494	1.133	1.001	1.000	9.559
		HL	116.648	17.650	6.174	3.407	2.274	1.367	1.072	1.000	1.000	9.497
		Trimean	42.583	12.694	5.629	3.292	2.246	1.373	1.076	1.000	1.000	9.480
		Trimmed	146.130	19.240	6.270	3.403	2.268	1.369	1.076	1.000	1.000	9.504
	Mixed EWMA-CUSUM	Mean	49.725	20.309	13.646	10.719	9.006	7.025	5.872	4.564	3.119	36.003
		Median	267.080	33.471	17.685	12.970	10.563	7.973	6.562	5.005	3.580	40.609
		Midrange	29.304	17.770	13.015	10.578	9.038	7.182	6.070	4.747	3.486	38.601
		HL	170.946	27.386	15.885	11.972	9.858	7.526	6.222	4.807	3.286	38.053
		Trimean	75.458	22.265	14.178	10.942	9.126	7.043	5.854	4.505	3.070	35.596
		Trimmed	214.485	28.489	15.887	11.840	9.691	7.368	6.081	4.645	3.257	37.302
10	CUSUM	Mean	17.351	7.027	3.978	2.789	2.196	1.653	1.178	1.000	1.000	9.560
		Median	223.303	16.272	5.874	3.617	2.676	1.897	1.412	1.002	1.000	9.876
		Midrange	7.919	5.992	4.329	3.293	2.654	1.968	1.652	1.061	1.000	10.163
		HL	134.052	11.551	4.903	3.182	2.412	1.778	1.247	1.000	1.000	9.683
		Trimean	143.537	12.270	4.970	3.179	2.401	1.750	1.223	1.000	1.000	9.662
		Trimmed	180.328	12.645	5.024	3.210	2.416	1.745	1.244	1.000	1.000	9.678
	EWMA	Mean	16.346	5.747	2.879	1.831	1.346	1.026	1.000	1.000	1.000	9.188
		Median	227.155	13.632	4.553	2.503	1.696	1.114	1.006	1.000	1.000	9.295
		Midrange	7.409	5.220	3.509	2.432	1.812	1.214	1.030	1.000	1.000	9.313
		HL	136.495	9.693	3.541	2.035	1.433	1.038	1.001	1.000	1.000	9.219
		Trimean	147.900	10.218	3.677	2.094	1.462	1.044	1.001	1.000	1.000	9.227
		Trimmed	186.448	10.652	3.731	2.103	1.470	1.049	1.001	1.000	1.000	9.231
	mixed EWMA-CUSUM	Mean	33.412	14.895	10.454	8.352	7.084	5.561	4.687	3.710	2.795	30.504
		Median	247.442	22.434	13.132	9.880	8.123	6.186	5.097	3.974	2.989	32.852
		Midrange	15.376	11.530	9.346	7.974	7.025	5.764	4.961	3.926	2.872	31.707
		HL	168.006	19.863	12.254	9.365	7.761	5.962	4.961	3.923	2.959	32.351
		Trimean	183.772	19.981	12.175	9.270	7.660	5.871	4.895	3.860	2.900	31.787
		Trimmed	213.733	20.632	12.428	9.431	7.783	5.950	4.930	3.835	2.838	31.448

Table 6.7: ARL and EQL values of CUSUM, EWMA and mixed EWMA-CUSUM charts with $m = 50$ at $ARL_0 = 370$ under diffuse mean contaminated normal environment

n	Chart	Estimator	δ									EQL
			0	0.25	0.5	0.75	1	1.5	2	3	5	
5	CUSUM	Mean	29.869	10.863	5.884	4.015	3.084	2.178	1.781	1.109	1.000	10.515
		Median	129.790	25.883	9.280	5.474	3.906	2.560	2.003	1.347	1.000	11.598
		Midrange	14.220	8.652	5.734	4.208	3.328	2.387	1.940	1.327	1.000	11.304
		HL	56.897	14.976	6.883	4.458	3.329	2.283	1.853	1.168	1.000	10.804
		Trimean	50.116	13.557	6.445	4.214	3.153	2.185	1.765	1.105	1.000	10.516
		Trimmed	74.154	16.684	6.946	4.350	3.201	2.178	1.727	1.121	1.000	10.557
	EWMA	Mean	28.515	8.955	4.473	2.804	1.999	1.289	1.050	1.000	1.000	9.388
		Median	139.620	21.544	7.613	4.162	2.756	1.605	1.182	1.004	1.000	9.718
		Midrange	12.579	6.918	4.359	3.040	2.262	1.458	1.126	1.001	1.000	9.503
		HL	57.550	12.565	5.392	3.154	2.180	1.350	1.068	1.000	1.000	9.457
		Trimean	54.612	11.949	5.266	3.154	2.186	1.357	1.071	1.000	1.000	9.459
		Trimmed	80.483	14.377	5.644	3.206	2.181	1.347	1.074	1.000	1.000	9.469
	mixed EWMA-CUSUM	Mean	50.128	19.385	13.337	10.568	8.918	6.982	5.846	4.530	3.105	35.785
		Median	188.471	29.807	16.913	12.623	10.348	7.871	6.495	4.976	3.555	40.281
		Midrange	28.989	17.001	12.695	10.426	8.951	7.155	6.056	4.747	3.436	38.302
		HL	94.549	23.317	14.818	11.443	9.526	7.357	6.119	4.741	3.251	37.513
		Trimean	75.189	21.222	13.828	10.783	9.018	6.989	5.824	4.478	3.063	35.430
		Trimmed	132.362	24.799	15.025	11.411	9.442	7.245	5.999	4.604	3.240	36.965
10	CUSUM	Mean	20.054	6.660	3.858	2.764	2.210	1.656	1.169	1.000	1.000	9.554
		Median	141.225	13.722	5.519	3.503	2.617	1.872	1.395	1.002	1.000	9.842
		Midrange	6.930	5.063	3.856	3.084	2.584	1.994	1.686	1.067	1.000	10.186
		HL	54.335	9.201	4.485	3.024	2.342	1.744	1.234	1.000	1.000	9.647
		Trimean	69.378	9.910	4.594	3.047	2.342	1.721	1.208	1.000	1.000	9.628
		Trimmed	89.792	10.409	4.675	3.087	2.357	1.717	1.229	1.000	1.000	9.645
	EWMA	Mean	17.750	5.138	2.648	1.729	1.309	1.023	1.000	1.000	1.000	9.177
		Median	143.580	11.508	4.225	2.402	1.663	1.106	1.005	1.000	1.000	9.279
		Midrange	5.373	3.691	2.724	2.084	1.662	1.191	1.026	1.000	1.000	9.270
		HL	52.685	7.524	3.187	1.922	1.389	1.034	1.000	1.000	1.000	9.203
		Trimean	67.893	8.091	3.332	1.991	1.423	1.040	1.001	1.000	1.000	9.212
		Trimmed	91.109	8.685	3.415	2.005	1.432	1.043	1.001	1.000	1.000	9.216
	mixed EWMA-CUSUM	Mean	31.933	14.480	10.317	8.282	7.032	5.534	4.669	3.685	2.788	30.369
		Median	162.964	20.720	12.671	9.653	7.987	6.117	5.064	3.962	2.984	32.730
		Midrange	15.963	11.686	9.470	8.088	7.138	5.851	5.024	4.001	2.989	32.607
		HL	74.366	17.564	11.540	9.000	7.528	5.840	4.889	3.879	2.932	31.969
		Trimean	90.709	17.877	11.558	8.958	7.459	5.768	4.827	3.809	2.865	31.356
		Trimmed	116.420	18.725	11.887	9.159	7.618	5.866	4.877	3.803	2.820	31.183

Table 6.8: ARL and EQL values of CUSUM, EWMA and mixed EWMA-CUSUM charts with $m = 50$ at $ARL_0 = 370$ under localized mean contaminated normal environment

n	Chart	Estimator	δ									EQL
			0	0.25	0.5	0.75	1	1.5	2	3	5	
5	CUSUM	Mean	18.880	13.540	7.098	4.472	3.300	2.237	1.815	1.119	1.000	10.627
		Median	19.162	14.951	8.537	5.345	3.880	2.557	1.999	1.356	1.000	11.587
		Midrange	19.045	14.596	8.109	5.156	3.777	2.519	1.982	1.340	1.000	11.503
		HL	18.948	13.827	7.353	4.676	3.445	2.321	1.872	1.177	1.000	10.870
		Trimean	18.847	13.803	7.273	4.535	3.318	2.233	1.797	1.111	1.000	10.599
		Trimmed	18.114	13.131	7.022	4.450	3.261	2.204	1.744	1.131	1.000	10.603
	EWMA	Mean	19.446	12.543	5.959	3.365	2.240	1.343	1.058	1.000	1.000	9.468
		Median	21.726	14.021	7.354	4.272	2.833	1.649	1.194	1.004	1.000	9.732
		Midrange	20.977	13.695	6.979	4.010	2.670	1.553	1.149	1.002	1.000	9.647
		HL	19.895	12.843	6.216	3.539	2.351	1.394	1.077	1.000	1.000	9.510
		Trimean	20.155	13.200	6.374	3.613	2.397	1.413	1.082	1.000	1.000	9.526
		Trimmed	19.055	12.190	5.994	3.445	2.300	1.384	1.083	1.000	1.000	9.500
	mixed EWMA-CUSUM	Mean	46.458	21.096	13.964	10.880	9.099	7.068	5.892	4.591	3.130	36.177
		Median	52.183	23.312	15.200	11.774	9.814	7.583	6.305	4.872	3.537	39.602
		Midrange	51.375	23.061	15.166	11.773	9.823	7.619	6.350	4.899	3.592	40.002
		HL	47.765	21.690	14.301	11.181	9.337	7.242	6.043	4.704	3.282	37.464
		Trimean	46.201	20.970	13.790	10.720	8.975	6.954	5.797	4.490	3.054	35.387
		Trimmed	46.662	21.302	13.978	10.898	9.081	7.051	5.876	4.534	3.216	36.450
10	CUSUM	Mean	18.920	10.382	4.838	3.142	2.380	1.760	1.221	1.000	1.000	9.655
		Median	19.034	11.838	5.632	3.574	2.663	1.895	1.433	1.003	1.000	9.877
		Midrange	19.024	13.164	6.696	4.296	3.178	2.175	1.786	1.079	1.000	10.460
		HL	18.964	10.608	4.993	3.241	2.447	1.804	1.280	1.000	1.000	9.714
		Trimean	18.964	10.752	5.015	3.224	2.427	1.777	1.254	1.000	1.000	9.689
		Trimmed	18.840	10.550	4.966	3.216	2.424	1.761	1.270	1.001	1.000	9.695
	EWMA	Mean	18.691	9.220	3.686	2.076	1.444	1.034	1.000	1.000	1.000	9.221
		Median	15.925	7.664	3.294	1.950	1.402	1.042	1.001	1.000	1.000	9.208
		Midrange	19.285	12.041	5.547	3.133	2.105	1.278	1.038	1.000	1.000	9.416
		HL	18.731	9.449	3.839	2.166	1.492	1.046	1.001	1.000	1.000	9.233
		Trimean	18.808	9.641	3.926	2.210	1.517	1.052	1.001	1.000	1.000	9.240
		Trimmed	18.444	9.432	3.883	2.202	1.520	1.057	1.001	1.000	1.000	9.240
	mixed EWMA-CUSUM	Mean	34.515	16.282	10.921	8.597	7.218	5.638	4.722	3.763	2.854	31.027
		Median	37.334	17.336	11.608	9.067	7.602	5.901	4.911	3.848	2.899	31.745
		Midrange	42.352	19.607	13.100	10.217	8.575	6.669	5.572	4.255	2.958	33.947
		HL	35.338	16.526	11.158	8.754	7.356	5.734	4.785	3.805	2.886	31.398
		Trimean	34.873	16.394	10.988	8.622	7.241	5.640	4.720	3.757	2.847	30.978
		Trimmed	35.105	16.578	11.110	8.716	7.314	5.696	4.760	3.729	2.783	30.610

To obtain a more global view of the run length distribution, along with the ARL, different indicators like the standard deviation of the run length (SDRL) and percentiles (denoted by $P_i, i = 5, 25, 50, 75, 95$) of the run lengths of the in-control process are reported in Table 6.9. These measures help studying the short and long run behavior of the run length distribution. For instance, the 5% percentiles of the run length distribution of the CUSUM, EWMA and mixed EWMA-CUSUM charts are on average about 20, 14, and 40 observations for all estimators used. (cf. Table 6.9).

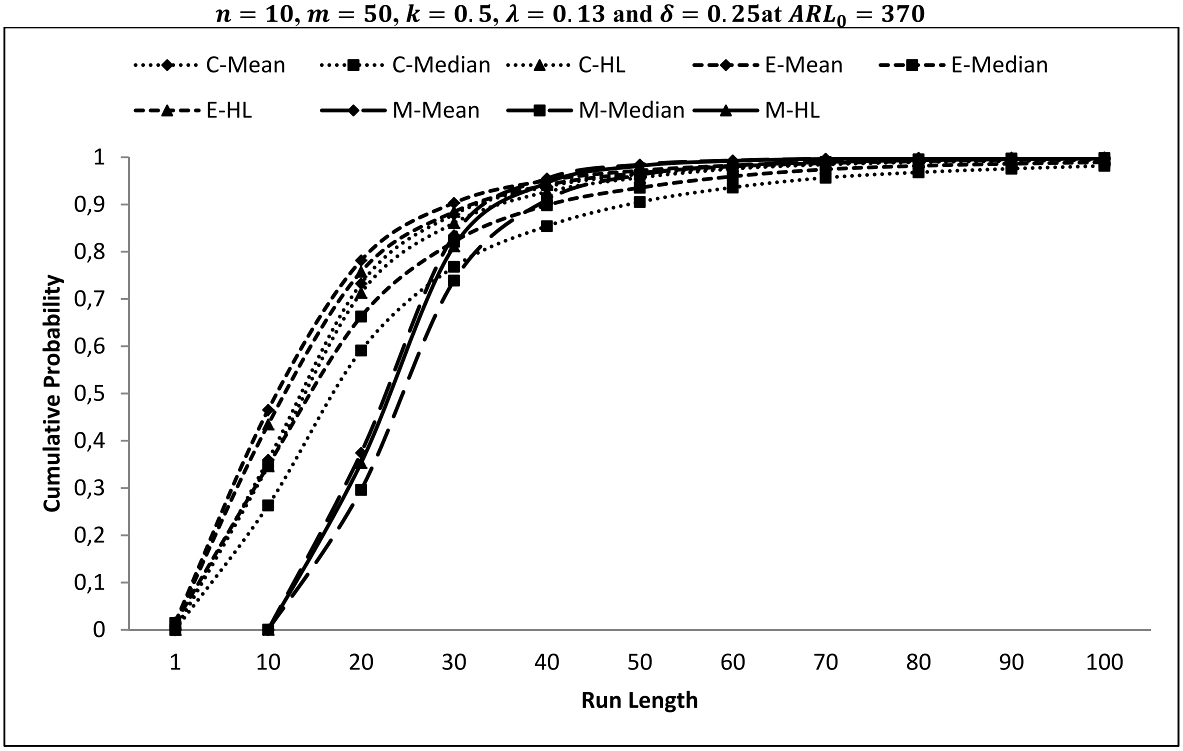
Table 6.9: Characteristics of In-Control Run Length Distribution under Uncontaminated Normal Environment for $n = 10$, $m = 50$, $k = 0.5$ and $\lambda = 0.13$ at $ARL_0 = 370$

Chart	Estimator	SDRL	Min	P_5	P_{25}	P_{50}	P_{75}	P_{95}	Max
CUSUM	Mean	441.06	3	21	87	220	482.25	1223.05	6516
	Median	425.46	2	21	95	234	499	1179	4488
	Midrange	414.8	2	24	100	238	497	1171	5086
	HL	443.5	3	20	89	220	484	1259.05	5047
	Trimean	424.06	2	21	90	222	487	1173	6226
	Trimmed	628.88	2	16	70	176	431	1322.2	16482
EWMA	Mean	463.69	1	14	82	209	478	1265.1	7765
	Median	451.07	1	14	87	222	492	1245	5868
	Midrange	427.63	1	15	91	231	497	1221	6477
	HL	445.82	1	15	83.75	221	489.25	1250.05	4362
	Trimean	454.56	1	14	81	209	478	1235.1	7772
	Trimmed	592.98	1	11	65	175	435	1342	12857
mixed EWMA-CUSUM	Mean	467.16	15	40	93	199.5	442	1270	5056
	Median	443.42	14	40	99	216	466	1217.05	6810
	Midrange	398.36	16	45	110	230	478	1168	5197
	HL	453.89	16	41	96	204	468	1249.2	5500
	Trimean	442	13	40	95	207.5	455.25	1214	4772
	Trimmed	497.97	14	39	89	194	445	1262.15	6974

To get more insight into the out-of-control run length distribution, Figure 6.1 presents the run length distribution curves of all the charts considering $n = 10$, $m = 50$, $k = 0.5$ and $\lambda = 0.13$ with $\delta = 0.25$ for normal environment. We only use three estimators: sample mean, sample median and HL estimator. In Figure 6.1, C, E and M represent, respectively, CUSUM, EWMA and mixed EWMA-CUSUM charts. The curves give the cumulative probability of detecting an out-of-control situation. A higher curve shows the superiority of a chart in terms of its quick detection of shifts in the process parameter.

It can be observed from Figure 6.1 that EWMA charts based on all estimators have higher probabilities for small run lengths to detect the shift than those of other memory charts under normality. For detecting a shift of magnitude $\delta = 0.25$ at a run length equal to 50, the mixed EMWA-CUSUM has larger probabilities as compared to the EWMA and CUSUM charts.

Figure 6.1: Run Length Curves for Memory charts under Uncontaminated Normal Environment when $n = 10, m = 50, k = 0.5, \lambda = 0.13$ and $\delta = 0.25$at $ARL_0 = 370$



6.5. SUMMARY AND CONCLUSION

Control charts are widely used in monitoring and controlling variations present in the process location and dispersion. Commonly applied control charts are the memory-less

(Shewhart-type) charts for targeting the large shifts and memory (EWMA and CUSUM) charts for aiming the smaller shifts. A combination of the EWMA and CUSUM control charts is applied to enhance the performance of the charts even further. The current study presents a comparison of the CUSUM, EWMA and mixed EWMA-CUSUM control charts based on different estimators. Different parent environments (normal and contaminated normal) are used to evaluate the performance of these charts in terms of their $ARLs$ and $EQLs$. The comparisons showed that there is no single control chart or estimator which behaves well in all environments. Under normality the EWMA control chart based on the sample mean is the best, although the differences with the other charts and estimators are insignificant (especially for small shifts). When there are localized or diffuse symmetric variances contaminations the mixed EWMA-CUSUM control chart is quite robust against these variance contaminations. Overall the best performance is obtained by the EWMA control chart based on the median estimator.

References

Abbas, N., Riaz, M., and Does, R.J.M.M. (2011). Enhancing the Performance of EWMA Charts. *Quality and Reliability Engineering International*, 27(6), 821 – 833.

Abbas, N., Riaz, M., and Does, R.J.M.M. (2013). Mixed Exponentially Weighted Moving Average – Cumulative Sum charts for Process Monitoring. *Quality and Reliability Engineering International*, 29(3), 355 – 365.

Abbasi, S. and Miller, A. (2012). On Proper Choice of Variability Control Chart for Normal and Non-Normal Processes. *Quality Reliability and Engineering International*, 28, 279-296.

Abbasi, S. A. and Miller, A. (2013). MDEWMA chart: An efficient and robust alternative to monitor process dispersion. *Journal of Statistical Computation and Simulation*, 83, 247–268.

Abbasi, S.A., Riaz, M., Miller, A., Ahmad, S., and Nazir, H.Z. (2014). EWMA Dispersion Control Charts for Normal and Non-normal processes. Accepted for publication in *Quality and Reliability Engineering International*.

Acosta-Mejia, C.A. (1998). Monitoring Reduction in Variability with the Range. *IIE Transactions,* 30, 515-523.

Acosta-Mejia, C.A. and Pignatiello, J.J.Jr. (2000). Monitoring Process Dispersion without Subgrouping. *Journal of Quality Technology,* 32, 89-102.

Acosta-Mejia, C.A., Pignatiello, J.J.Jr., and Rao, B.V. (1999). A Comparison of Control Charting Procedures for Monitoring Process Dispersion. *IIE Transactions,* 31, 569-579.

Ahmad, S., Riaz, M., Abbasi, S.A., and Lin, Z. (2013). On Monitoring Process Variability Under Double Sampling Scheme. *International Journal of Production Economics*, 142(2), 388 – 400.

Alloway, J. A., Jr. and Raghavachari, M. (1991). Control Chart Based on the Hodges–Lehmann Estimator. *Journal of Quality Technology 23*, 336–347.

Amin, R.W. and Searcy, A.J. (1991). A Nonparametric Exponentially Weighted Moving Average Control Scheme. *Communications in Statistics - Simulation and Computation*, 20(4), 1049-1072.

Antzoulakos, D.L., and Rakitzis, A.C.(2008). The Modified r out of m Control Chart. *Communication in Statistics -Simulations and Computations*, 37, 396-408.

Bickel, P.J. and Lehmann, E.L. (1976). Descriptive Statistics for Non-Parametric Models III: Dispersion. *Annals of Statistics*, 4, 1139–1158

Bickel, P.J. and Lehmann, E.L. (1979). Descriptive Statistics for Nonparametric Models III: Dispersion, in *Contributions to Statistics, Hájek Memorial Volume*, edited by J. Jurekova, Academia, Prague, 33 - 40.

Bissell, D. (1994). *Statistical methods for SPC and TQM*. 1st edition, Chapman & Hall, New York.

Caperaa, P. and Rivest, L.P. (1995). On the Variance of Trimmed Mean. *Statistics and Probability Letters*, 22(1), 79-85.

Capizzi, G. and Masarotto, G. (2003). An Adaptive Exponentially Weighted Moving Average Control Chart. *Technometrics*, 45(3), 199-207.

Chang, T.C. and Gan, F.F. (1995). A Cumulative Sum Control Chart for Monitoring Process Variance. *Journal of Quality Technology,* 27, 109-119.

Croux, C. and Rousseeuw, P. J. (1992). A Class of High-Breakdown Scale Estimators Based on Subranges. *Communications in Statistics—Theory and Methods*, 21, 1935-1951.

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

David, H.A. (1968). Gini's Mean Difference Rediscovered. *Biometrika*, 55, 573–575.

David, H.A. (1998). Early Sample Measures of Variability. *Statistical Science*, 13, 368–377.

Davis, C.M. and Adams, B.M. (2005). Robust Monitoring of Contaminate Data. *Journal of Quality Technology* 37, 163-174.

Dixon, W.J. and Massey, F.J. (1969). *Introduction to Statistical Analysis*, 3rd edition, McGraw-Hill, New York.

Ewan, W.D. and Kemp, K.W. (1960). Sampling Inspection of Continuous Processes with no Autocorrelation Between Successive Results. *Biometrika*, 47, 363-380.

Ferrell, E.B. (1953). Control Charts using Midranges and Medians. *Industrial Quality Control*, 9, 30-34.

Gini, C. (1912). Variabilità e Mutabilità, Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche. *Studi Economico-Giuridici della R. Università di Cagliari*, 3, 3–159.

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

Grove, D.M. (2013). Letter to the Editor. *Journal of Quality Technology*, 45(4), 395-396.

Hampel, F.R. (1974). The Influence Curve and its Role in Robust Estimation. *Journal of the American Statistical Association*, 69, 383-393.

Hawkins, D.M. (1993). Robustification of Cumulative Sum Charts by Winsorization. *Journal of Quality Technology*, 25(4), 248-261.

Hawkins, D.M. and Olwell, D.H. (1998). *Cumulative Sum Charts and Charting for Quality Improvement*. Springer Verlag, New York.

Hettmansperger, T.P. and McKean, J.W. (1998). *Robust Nonparametric Statistical Methods*. Wiley, New York.

Hodges, J.L. and Lehmann, E.L. (1963). Estimates of Location Based on Rank Tests. *The Annals of Mathematical Statistics 34*, 598–611.

Hunter, J.S. (1986). The Exponentially Weighted Moving Average. *Journal of Quality Technology,* 18, 203-210.

Janacek, G.J. and Miekle, S.C. (1997). Control Charts Based on Medians. *Statistician*, 46, 19-31.

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

Jensen, W.A., Jones-Farmer, L.A., Champ, C.W., and Woodall, W.H. (2006). Effects of Parameter Estimation on Control Charts Properties: A Literature Review. *Journal of Quality Technology*, 38(4), 349-364.

Jones-Farmer, L.A., Jordan, V. and Champ, C.W. (2009). Distribution-Free Phase I Charts for Subgroup Location. *Journal of Quality Technology*, 41(3), 304 – 316.

Khattree, R. and Rao, C.R. (2003). *Statistics in Industry- (Handbook of Statistics; v. 22)*. 1st ed., Elsevier, The Netherlands.

Khoo, M.B.C. (2004). Some Control Charts for the Process Mean and Variance Based on Downton's Estimator. *Engineering Management Conference*, 2004. Proceedings 2004 IEEE International, 3, 1071–1074.

Langenberg, P. and Iglewicz, B. (1986). Trimmed Mean X-bar and R Charts. *Journal of Quality Technology,* 18, 152-161.

Lee, M.H. (2011). Economic Design of Cumulative Sum Control Chart for Non-Normally Correlated Data. *Matematika*, 27(1), 79–96.

Lehmann, E.L. (1983), *Theory of Point Estimation.* Wiley, New York.

Levinson, W. A. and Polny, A. (1999). SPC for Tool Particle Counts. *Semiconductor International,* 22, 117–121.

Li, S.Y., Tang, L.C. and Ng, S.H. (2010). Non-Parametric CUSUM and EWMA Control Charts for Detecting Mean Shifts. *Journal of Quality Technology*, 42(2), 209 – 226.

Lucas, J.M. (1982). Combined Shewhart-CUSUM Quality Control Schemes. *Journal of Quality Technology*, 14(2), 51-59.

Lucas, J.M. and Crosier, R.B. (1982). Fast Initial Response for CUSUM Quality-Control Scheme. *Technometrics*, 24, 199-205.

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

MacEachern, S.N., Rao, Y., and Wu, C. (1997). A Robust-Likelihood Cumulative Sum Chart. *Journal of the American Statistical Association*, 102(480), 1440-1447.

Machado, M.A.G. and Costa, A.F.B. (2008). The Double Sampling and the EWMA Charts based On the Sample Variances. *International Journal of Production Economics*, 114(1), 134 – 148.

Mahmoud, M.A., Henderson, G.R., Epprecht, E.K., and Woodall, W.H. (2010). Estimating the Standard Deviation in Quality-Control Applications. *Journal of Quality Technology,* 42, 348–357.

Maravelakis, P.E., Panaretos, J. and Psarakis, S. (2002). Effect of Estimation of the Process Parameters on the Control Limits of the Univariate Control Charts for Process Dispersion. *Communication in Statistics Simulation and Computation,* 31, 443-461.

Maravelakis, P., Panaretos, J., and Psarakis, S. (2005). An Examination of the Robustness to Non-normality of the EWMA Control Charts for the Dispersion. *Communications in Statistics: Simulation and Computation,* 34, 1069–1079.

Mast, J. de and Roes, K. (2004). Robust Individuals Control Chart for Exploratory Analysis. *Quality Engineering,* 16, 407–421.

McDonald, D. (1990). A CUSUM Procedure Based on Sequential Ranks. *Naval Research Logistics*, 37, 627-646.

Midi, H. and Shabbak, A. (2011). Robust Multivariate Control Charts to Detect Small Shifts in Mean. *Mathematical Problems in Engineering*, DOI:10.1155/2011/923463.

Montanari, A. and Monari, P. (2005). Gini's Contribution to Multivariate Statistical Analysis. International Conference in Memory of Two Eminent Scientists: C. Gini and M. O. Lorenz, 23-26 May 2005, Certosa in Pontignano.

Montgomery, D.C. (2009). *Introduction to Statistical Quality Control*. 6th ed. Wiley, New York.

Moustafa, A. (2009). A Control Chart based on Robust Estimators for Monitoring the Process Mean of a Quality Characteristic. *International Journal of Quality and Reliability Management*, 26(5), 480-496.

Moustafa, A. and Mokhtar, B.A. (1999). New Robust Statistical Process Control Chart for Location. *Quality Engineering*, 12(2), 149-159.

Moustakides, G.V. (1986). Optimal Stopping Times for Detecting Changes in Distributions. *Annals of Statistics,* 14, 1379-1387.

Muhammad, F. and Riaz, M. (2006). Probability Weighted Moments Approach to Quality Control Charts. *Economic Quality Control*, 21, 251-260.

Nazir, H.Z., Riaz, M., Does, R.J.M.M. and Abbas, N. (2013). Robust CUSUM Control Charting. *Quality Engineering*, 25, 211-224.

Nazir, H.Z., Riaz, M. and Does, R.J.M.M. (2013). Robust CUSUM Control Charting for Process Dispersion. *Quality and Reliability Engineering International,* DOI: 10.1002/qre. 1596.

Nazir, H.Z., Schoonhoven, M., Riaz, M. and Does, R.J.M.M. (2014a). How to Set Up a Robust Shewhart Control Chart for Dispersion?. *Quality Engineering,* 26(1), 130-136.

Nazir, H.Z., Schoonhoven, M., Riaz, M. and Does, R.J.M.M. (2014b). A Stepwise Approach for Setting Up a Robust Shewhart Location Control Chart. *Quality Engineering,* 26(2), 246-252.

Nazir, H. Z., Abbas, N., Riaz, M. and Does, R.J.M.M. (2014). A Comparative Study of Memory-Type Control Charts under Normal, Contaminated and Non-Normal Environments. Submitted to *Quality Technology and Quality Management*.

Ng, C.H. and Case, K.E. (1989). Development and Evaluation of Control Charts using Exponentially Weighted Moving Averages. *Journal of Quality Technology,* 21:242-250.

Omar M.A. (2008). A Simple Robust Control Chart Based on MAD. *Journal of Mathematics and Statistics,* 4, 102-107.

Ou, Y.J., Wu, Z. and Tsung, F. (2012). A Comparison Study of Effectiveness and Robustness of Control Charts for Monitoring Process Mean. *International Journal of Production Economics*, 135(1), 479 – 490.

Page, E.S. (1954). Continuous Inspection Schemes. *Biometrika*, 41, 100-115.

Page, E.S. (1963). Controlling the Standard Deviation by CUSUM and Warning Lines. *Technometrics,* 5, 307-315.

Psarakis, S., Vyniou, A.K. and Castagliola, P. (2013). Some Recent Developments on the Effects of Parameter Estimation of Control Charts. *Quality and Reliability Engineering International,* 29(7), 951-970.

Quesenberry, C.P. (1993). The Effect of Sample Size on Estimated Limits for $\bar{X}$ and X Control Charts. *Journal of Quality Technology*, 25, 237–247.

Reynolds Jr, M.R. and Stoumbosb, Z.G. (2010). Robust CUSUM Charts for Monitoring the Process Mean and Variance. *Quality and Reliability Engineering International*, 26, 453 – 473.

Riaz, M. (2008). A Dispersion Control Chart. *Communications in Statistics—Simulation and Computation*, 37, 1239–1261.

Riaz, M., Abbas, N., and Does, R.J.M.M. (2011). Improving the Performance of CUSUM Charts. *Quality and Reliability Engineering International*, 27(4), 415 – 424.

Roberts, S.W. (1959), Control Chart Tests Based on Geometric Moving Averages. *Technometrics*,1(3), 239-250.

Rocke, D.M. (1989). Robust Control Charts. *Technometrics,* 31, 173–184.

Rocke, D.M. (1992). $\bar{X}_Q$ and R_Q Charts: Robust control charts. *The Statistician*, 41, 97-104.

Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*, 80, 1273-1283.

Rousseeuw, P.J. (1991). Tutorial to Robust Statistics. *Journal of Chemometrics*, 5, 1-20.

Schoonhoven, M. (2011). Estimation Methods for Statistical Process Control. Ph.D. Thesis, University of Amsterdam.

Schoonhoven, M. and Does, R.J.M.M. (2012). A Robust Standard Deviation Control Chart. *Technometrics*, 54, 73-82.

Schoonhoven, M. and Does, R.J.M.M. (2013a). A Robust $\bar{X}$ Control Chart. *Quality and Reliability Engineering International,* 29, 951-970.

Schoonhoven, M. and Does, R.J.M.M. (2013b). Response to the Letter to the Editor. *Journal of Quality Technology*, 45(4), 397-398.

Schoonhoven, M., Nazir, H.Z., Riaz, M. and Does, R.J.M.M. (2011). Robust Location Estimators for the $\bar{X}$ control chart. *Journal of Quality Technology*, 43, 363-379.

Schoonhoven, M., Riaz, M. and Does, R.J.M.M. (2011). Design and Analysis of Control Charts for Standard Deviation with Estimated Parameters. *Journal of Quality Technology*, 43, 307-333.

Schoonhoven, M., Riaz, M. and Does, R.J.M.M. (2009). Design Schemes for the $\bar{X}$ Control Chart. *Quality and Reliability Engineering International,* 25, 581-594.

Shamos, M.I. (1976). Geometry and Statistics: Problems at the Interface, in *Directions and Recent Results in Algorithms and Complexity*, edited by J. F. Traub. Academic Press, New York, 251-280.

Shewhart, W.A. (1931). Economic Control of Quality of Manufactured product. Prinveton, NJ: Van Nostrand.

Shu, L. and Jiang, W. (2008). A New EWMA Chart for Monitoring Process Dispersion. *Journal of Quality Technology* 40, 319–331.

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

Song, M.S., Chung, H.Y., and Bae, W.S. (1982). Subset Selection Procedures Based on Some Robust Estimators. *Journal of the Korean Statistical Society*, 11, 109-117.

Steiner, S.H. (1999). EWMA Control Charts with Time Varying Control Limits and Fast Initial Response. *Journal of Quality Technology*, 31(1), 75 – 86.

Tatum, L.G. (1997). Robust Estimation of the Process Standard Deviation for Control Charts. *Technometrics,* 39, 127–141.

Tukey, J.W. (1977). *Exploratory Data Analysis*. Addison Wesley Reading, MA.

Tuprah, K. and Ncube, M. (1987). A Comparison of Dispersion Quality Control Charts. *Sequential Analysis*, 6, 155-163.

Vargas, J.A. (2003). Robust Estimation in Multivariate Control Charts for Individual Observations. *Journal of Quality Technology,* 35, pp. 367-376.

Vining, G. (2009). Technical Advice: Phase I and Phase II Control Charts. *Quality Engineering*, 21, 478–479.

Wadsworth, H.M., Stephens, K.S. and Godfrey, A.B. (2001). *Modern Methods for Quality Control and Improvement*, 2nd ed. New York, Wiley.

Wang, T., Li, Y. and Cui, H. (2007). On Weighted Randomly Trimmed Means. *Journal of Systems Science and Complexity* 20, 47-65.

Weisberg, H.F. (1992). *Central Tendency and Variability*. Sage University Paper Series on Quantitative Application in Social Sciences, series no. 07-038, 7–83. Newbury Park, CA.

Woodall, W.H. and Montgomery, D.C. (1999). Research Issues and Ideas in Statistical Process Control. *Journal of Quality Technology*, 31, 376–386.

Wu, C., Zhao, Y. and Wang, Z. (2002) The Median Absolute Deviation and Their Applications to Shewhart Xbar Control Charts. *Communications in Statistics-Simulation and Computation*, 31(3), 425–442.

Wu, Z., Jiao, J., Yang, M., Liu, Y., and Wang, Z. (2009). An Enhanced Adaptive CUSUM Control Chart. *IIE Transactions*, 41(7), 642 – 653.

Yang, L., Pai, S., and Wang, Y.R. (2010). A Novel CUSUM Median Control Chart. *Proceedings of International Multiconference of Engineers and Computer Scientists*, Hong Kong.

Yang, S.F. and Cheng, S.W. (2011). A New Non-parametric CUSUM Mean Chart. *Quality and Reliability Engineering International*, 27, 867 – 875.

Yang, S.F., Lin, J.S. and Cheng, S.W. (2011). A New Nonparametric EWMA Sign Control Chart. *Expert Systems with Applications*, 38, 6239–6243.

Zhao, Y., Tsung, F., and Wang, Z. (2005). Dual CUSUM Control Schemes for Detecting a Range of Mean Shifts. *IIE Transactions*, 37(11), 1047 – 1057.

Zwetsloot, I.M., Schoonhoven, M. and Does, R.J.M.M. (2014a). A Robust Estimator for Location in Phase I based on an EWMA Chart. Accepted for publication in *Journal of Quality Technology*.

Zwetsloot, I.M., Schoonhoven, M. and Does, R.J.M.M. (2014b). A Robust Phase I Exponentially Weighted Moving Average Chart for Dispersion. Accepted for publication in *Quality and Reliability Engineering International*.

SAMENVATTING

Statistische procesbeheersing is een methode waarbij statistische technieken worden gebruikt om een proces te beheersen en om de kwaliteit van de producten te verbeteren. Van deze technieken is de Shewhart regelkaart de belangrijkste en het meest gebruikt. Het is een grafiek van metingen van een kwaliteitskarakteristiek van het proces op de verticale as uitgezet tegen de tijd op de horizontale as. De grafiek wordt aangevuld met regelgrenzen die de procesinherente variatie markeren. Zodra een meting buiten de regelgrenzen valt dan noemen we het proces niet beheerst. In de loop der jaren is hier veel onderzoek naar gedaan. Zo zijn er regelkaarten voorgesteld die gebaseerd zijn op de eerdere waarnemingen die hebben plaatsgevonden. De signalering hangt dus niet alleen af van de laatste meting maar het geheugen speelt ook een rol. Van deze regelkaarten zijn de CUSUM (CUmulative SUM) kaart en de EWMA (Exponentially Weighted Moving Average) kaart de meest populaire.

Dit proefschrift introduceert een aantal robuuste Shewhart, CUSUM en mixed CUSUM-EWMA regelkaarten voor de locatie- en spreidingsparameters van de kwaliteitskarakteristiek die gemonitord wordt. Robuust in de zin dat de regelkaarten zich goed moeten gedragen bij afwijkingen van normaal verdeelde kwaliteitskenmerken van een proces. Het proefschrift richt zich dus op het achterhalen welke regelkaarten en met welke schattingsmethoden men kan komen tot robuuste procesbeheersing.

In de hoofdstukken 2 en 3 worden Shewhart-type regelkaarten voor respectievelijk de spreiding en de locatie parameter van de kwaliteitskarakteristiek voorgesteld die robuust zijn voor lokale en verspreide verstoringen in de variantie of gemiddelde en die zich goed gedragen onder normaliteit. Voor de gebruiker wordt een stapsgewijze procedure gegeven om de robuuste regelkaarten voor deze situaties te ontwerpen.

In hoofdstukken 4 en 5 worden CUSUM-regelkaarten voor dezelfde situaties bestudeert. Ook hierbij wordt nagegaan in hoeverre de voorgestelde regelkaarten robuust zijn voor lokale en verspreide verstoringen in de variantie en het gemiddelde en hoe ze zich gedragen onder normaliteit.

Tot slot wordt in het laatste hoofdstuk drie regelkaarten, die ook gebruikmaken van een geheugen, vergeleken. Deze zijn de CUSUM, EWMA en een mengvorm van beide regelkaarten. De drie kaarten worden bestudeerd onder normaliteit en verstoringen van de normaliteit.

Curriculum Vitae

Hafiz Zafar Nazir was born in Okara, Punjab, Pakistan on 3rd July, 1983. He obtained his B.Sc. with Statistics and Mathematics as major subjects from the Government College University, Lahore, Pakistan in 2004; M.Sc. in Statistics from the Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan in 2006 and M.Phil in Statistics from the Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan in 2008. He is serving as a Lecturer in the Department of Statistics of the University of Sargodha, Sargodha, Pakistan from September 2009 to date. His current research interests include Statistical Process Control and Non-Parametric techniques.