

SCHOOL OF INFORMATION SCIENCES

& TECHNOLOGY

DEPARTMENT OF STATISTICS POSTGRADUATE PROGRAM

An Algorithmic Approach of the Effect of Parameter Estimation of the EWMA Control Chart

Ву

Georgios Samsonidis

A THESIS

Submitted to the Department of Statistics of the Athens University of Economics and Business in partial fulfilment of the requirements for the degree of Master of Science in Statistics Athens, Greece

2021







ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ

Αλγοριθμική προσέγγιση της Επίδρασης της Εκτίμησης Παραμέτρων στο EWMA Διάγραμμα Ελέγχου

Γεώργιος Σαμσονίδης

Διατριβή Που υποβλήθηκε στο Τμήμα Στατιτικής του Οικονομικού Πανεπιστημίου Αθηνών ως μέρος των απαιτήσεων για την απόκτηση Διπλώματος Μεταπτυχιακών Σπουδών στη Στατιστική Αθήνα 2021





Dedication

This thesis is dedicated to my parents, to my sister and to my friends who have always been a source of support and encouragement during my university life.



Acknowledgements

I would like to thank my supervisor Professor Psarakis Stelios for guidance, help, understanding and immediate response to every question throughout this work. I would also like to thank my family and my friends for patience and support.

George



Vita

I graduated in 2019 from Athens University of Economics and Business. I studied Informatics, specialized in Computer Systems and Networks and Operational Research and Economics of Information Technology. In 2019-2020 I attended the postgraduate program in Statistics at Athens University of Economics and Business the requirements of which are fulfilled with this thesis.



Abstract

Georgios Samsonidis

An Algorithmic Approach of the Effect of Parameter Estimation of the EWMA Control Chart

Athens 2021

Statistical Quality Control is a necessary tool for companies that compete in today's economy. With it, the companies can ensure the quality and reliability of their products. The control is implemented in the production process and enables taking corrective measures when the produced product is considered out of the control limits. The control processes are done through the control diagrams (i.e. Shewhart, CUSUM, EWMA). The first control diagram was presented during the second Industrial Revolution and specifically in 1924 by Walter A. Shewhart. Today, writing this study, we understand that we are closer to the fourth Industrial Revolution the result of which will be the increasing drop of human interaction in all parts of the production process. One of the problems we are going to focus on in this study is the unknown parameters that frequently must be estimated to set the limits of control diagrams, and their estimation is being achieved through a data sample. There are many control diagrams used in statistical quality control, but for our study we are going to focus on the Exponentially Weighted Moving Average diagram (EWMA). Estimating the parameters for the EWMA chard is going to be achieved through an algorithmic process, and, by using older studies as a base of reference and the results of this study, we are going to try to produce reliable results for parameter estimation.



Περίληψη

Γιώργος Σαμσονίδης

Αλγοριθμική προσέγγιση της Επίδρασης της Εκτίμησης Παραμέτρων στο EWMA Διάγραμμα Ελέγχου

Ο Στατιστικός Έλεγχος Ποιότητας είναι ένα αναγκαίο εργαλείο στην σημερινή εποχή για τις επιχειρήσεις. Με τον Στατιστικό έλεγχο ποιότητας οι επιχειρήσεις μπορούν να διασφαλίσουν την ποιότητα και αξιοπιστία των προϊόντων τους. Ο έλεγχος αυτός εφαρμόζεται στην παραγωγική διαδικασία και με τον τρόπο αυτο υπάρχει η δυνατότητα παρέμβασης όταν το παραγώμενο προϊόν θεωρηθεί εκτός ορίων. Η διαδικασίες ελέγχου γίνονται μέσα απο τα διαγράμματα ελέγχου (πχ. Shewhart, CUSUM, EWMA). Το πρωτό διάγραμμα ελέγχου παρουσιάστηκε κατά την 2η Βιομηγανική επανάσταση και συγκεκριμένα το 1924 απο τον Walter A. Shewhart. Σήμερα που γράφουμε αυτή την μελέτη γνωρίζουμε ότι είμαστε κόντα στην 4η Βιομηγανική επανάσταση της οποίας το αποτέλεσμα θα είναι η μείωση της αλληλεπίδρασης του ανθρωπού σε κάθε κομμάτι μιας παραγωγικής διαδικασίας. Ένα απο τα προβλήματα που θα μελετήσουμε όμως είναι τα όρια των διαγραμμάτων ελέγχου τα οποία αρκετές φορές παρουσιάζουν άγνωστες παραμέτρους οι οποίες πρέπει να εκτιμηθούν και η εκτίμηση τους γίνεται με την βοήθεια κάποιου δείγματος των δεδομένων. Υπάρχουν αρκετά διαγράμματα ελέγχου για τον στατιστικό έλεγχο ποιότητας, όμως στην μελέτη μας θα επικεντρωθούμε στο διάγραμμα Εκθετικής Εξομάλυνσης (EWMA). Θα εστιάσουμε στην εκτίμηση των παραμέτρων για το διάγραμμα EWMA μέσα απο μια αλγοριθμική προσέγγιση και θα προσπαθήσουμε με βάση παλαιότερες μελέτες αλλά και το πόρισμα των αποτελεσμάτων της εν λόγω μέλετης να παρέχουμε αξιόπιστα αποτελέσματα εκτίμησης παραμέτρων.





Contents

Introduction	1
Chapter 1	2
1.1 Statistical process control	2
1.2 Phase I and Phase II control diagrams	3
1.3 Classification of control charts	4
1.4 Run Length in Statistical Process Control	4
1.5 The Problem of Statistical Process Control	6
Chapter 2	7
2.1 Shewhart chart	7
2.2 CUSUM chart	9
2.3 EWMA chart	12
2.4 Multivariate Control Charts	15
2.5 EWMA-Poisson chart	16
2.6 Parameter Estimation for EWMA control chart	18
Chapter 3	20
3.1Effect of parameter estimation on control diagrams for the mean value	20
3.2 The distribution of run length when estimating the parameters of the control diag	rams 20
3.3 Studies in parameter estimation	21
3.4 Effect of parameter estimation on control diagrams for dispersion	23
Chapter 4	25
- 4.1 CSEWMA control chart	25
4.2 Combination of EWMA and CUSUM diagrams for simultaneous media and dispers monitoring	ion 26
4.3 Combined EWMA – CUSUM (CMEC) chart	26
4.4 Conclusion	27
Chapter 5	29
5.1 Introduction	29
5.2 I-Control chart and parameter estimators in phase I	31
5.3 Phase I : outlier detection and parameters estimation with Algorithmic usage	34
5.4 Phase II: AARLO and SDARLO computation with Monte Carlo simulation	36
5.5 Conclusion	46
Conclusion and Suggestions	47
Appendix	

OF ECO

MO AOLUN NSS

ography56



List of Figures

Figure 1: Shewhart control chart for marbles	8
Figure 2:Example of Ewma control chart	14
Figure 3: I-Chart with L = 3 and without outlier filtering	35
Figure 4:I-Chart with L = 3 and after outlier filtering	36



List of Tables

Table 2: Results from simulation of algorithm for sample 5037Table 3: Results from simulation of algorithm for sample 10038Table 4: Results from simulation of algorithm for sample 20038Table 5: Results from simulation of algorithm for sample 50039Table 6:Results from simulation of algorithm for sample 100040Table 7: Results from simulation of algorithm for sample 100040Table 8: Results from simulation of algorithm for sample 100040Table 8: Results from simulation of algorithm for sample 100042Table 9:Results from simulation of algorithm for sample 20043Table 10: Results from simulation of algorithm for sample 20043Table 11: Results from simulation of algorithm for sample 50043Table 12:Results from simulation of algorithm for sample 100044Table 13:Results from simulation of algorithm for sample 100044	Table 1: Standard Deviation Estimators	32
Table 3: Results from simulation of algorithm for sample 10038Table 4: Results from simulation of algorithm for sample 20038Table 5: Results from simulation of algorithm for sample 50039Table 6:Results from simulation of algorithm for sample 100040Table 7: Results from simulation of algorithm for sample 100040Table 8: Results from simulation of algorithm for sample 5042Table 9:Results from simulation of algorithm for sample 10042Table 10: Results from simulation of algorithm for sample 20043Table 11: Results from simulation of algorithm for sample 50043Table 12:Results from simulation of algorithm for sample 100044Table 13:Results from simulation of algorithm for sample 100044	Table 2: Results from simulation of algorithm for sample 50	37
Table 4: Results from simulation of algorithm for sample 20038Table 5: Results from simulation of algorithm for sample 50039Table 6:Results from simulation of algorithm for sample 100040Table 7: Results from simulation of algorithm for sample 1000040Table 8: Results from simulation of algorithm for sample 5042Table 9:Results from simulation of algorithm for sample 10042Table 10: Results from simulation of algorithm for sample 10043Table 11: Results from simulation of algorithm for sample 50043Table 12:Results from simulation of algorithm for sample 100044Table 13:Results from simulation of algorithm for sample 100044	Table 3: Results from simulation of algorithm for sample 100	38
Table 5: Results from simulation of algorithm for sample 50039Table 6:Results from simulation of algorithm for sample 100040Table 7: Results from simulation of algorithm for sample 1000040Table 8: Results from simulation of algorithm for sample 5042Table 9:Results from simulation of algorithm for sample 10042Table 10: Results from simulation of algorithm for sample 20043Table 11: Results from simulation of algorithm for sample 50043Table 12: Results from simulation of algorithm for sample 100044Table 13: Results from simulation of algorithm for sample 100044	Table 4: Results from simulation of algorithm for sample 200	38
Table 6:Results from simulation of algorithm for sample 100040Table 7: Results from simulation of algorithm for sample 1000040Table 8: Results from simulation of algorithm for sample 5042Table 9:Results from simulation of algorithm for sample 10042Table 10: Results from simulation of algorithm for sample 20043Table 11: Results from simulation of algorithm for sample 50043Table 12:Results from simulation of algorithm for sample 100044Table 13:Results from simulation of algorithm for sample 100045	Table 5: Results from simulation of algorithm for sample 500	39
Table 7: Results from simulation of algorithm for sample 1000040Table 8: Results from simulation of algorithm for sample 5042Table 9:Results from simulation of algorithm for sample 10042Table 10: Results from simulation of algorithm for sample 20043Table 11: Results from simulation of algorithm for sample 50043Table 12:Results from simulation of algorithm for sample 100044Table 13:Results from simulation of algorithm for sample 100045	Table 6:Results from simulation of algorithm for sample 1000	40
Table 8: Results from simulation of algorithm for sample 5042Table 9:Results from simulation of algorithm for sample 10042Table 10: Results from simulation of algorithm for sample 20043Table 11: Results from simulation of algorithm for sample 50043Table 12:Results from simulation of algorithm for sample 100044Table 13:Results from simulation of algorithm for sample 100045	Table 7: Results from simulation of algorithm for sample 10000	40
Table 9:Results from simulation of algorithm for sample 10042Table 10: Results from simulation of algorithm for sample 20043Table 11: Results from simulation of algorithm for sample 50043Table 12:Results from simulation of algorithm for sample 100044Table 13:Results from simulation of algorithm for sample 100045	Table 8: Results from simulation of algorithm for sample 50	42
Table 10: Results from simulation of algorithm for sample 20043Table 11: Results from simulation of algorithm for sample 50043Table 12:Results from simulation of algorithm for sample 100044Table 13:Results from simulation of algorithm for sample 100045	Table 9:Results from simulation of algorithm for sample 100	42
Table 11: Results from simulation of algorithm for sample 50043Table 12:Results from simulation of algorithm for sample 100044Table 13:Results from simulation of algorithm for sample 1000045	Table 10: Results from simulation of algorithm for sample 200	43
Table 12:Results from simulation of algorithm for sample 100044Table 13:Results from simulation of algorithm for sample 1000045	Table 11: Results from simulation of algorithm for sample 500	43
Table 13:Results from simulation of algorithm for sample 10000 45	Table 12:Results from simulation of algorithm for sample 1000	44
	Table 13:Results from simulation of algorithm for sample 10000	45

Table A.1: Values of AARL for SSD estimator	48
Table A.2: Values of AARL for RMSSD estimator	48
Table A.3: Values of AARL for MdMR estimator	49
Table A.4: Values of AARL for MnMR estimator	50
Table A.5: Values of AARL for MAD estimator	50
Table A.6: Values of AARL for IQR estimator	51
Table A.7: Values of AARL for MnMR estimator	52
Table A.8: Values of AARL for MdMR estimator	52
Table A.9: Values of AARL for RMSSD estimator	53
Table A. 10: Values of AARL for MAD estimator	54
Table A.11: Values of AARL for IQR estimator	54
Table A. 12: Values of AARL for SSD estimator	55



Introduction

In this study we are going to focus in Statistical Quality Control, and with one of its most useful tools, the control diagrams. The first control diagram was created in 1924 by Walter A. Shewhart. Other than Shewhart there are other control diagrams like the Exponentially Weighted Moving Average (EWMA) in which the study is being focused on, or Cumulative Sum (CUSUM). The dissertation is structured as follows. In the 1st chapter we are noting the use of control diagrams in phase I and phase II, the steps necessary to create a control diagram, the problems that are inherent in statistical control as well as the Average Run Length. In the 2nd chapter there is an in depth look in the Shewhart, CUSUM and specifically EWMA charts. In the 3rd chapter we are analyzing the effects of parameter estimation in control charts and in Average Run Length. In the 4th chapter we are looking into the combination of Shewhart – EWMA and EWMA – CUSUM control diagrams. In the 5th and final chapter, we are explaining the algorithm simulated for the six estimators and analyzing the results of each instance.



Chapter 1

1.1 Statistical process control

The two most deciding factors for a consumer to choose to buy a product that satisfy their needs are the perceived quality of the product and the price attached to it. Companies have sought business methods to produce the highest quality of products with the lowest possible cost, creating items that have a competitive advantage due to their "value-for-money" ratio. This search for more efficient and quality assured production methods resulted in new frameworks for operating and administrating the various enterprises, geared toward continual quality improvement.

In these procedures, statistical quality control is a critical factor, as it is the most wellknown and older way of regulating manufacturing processes to ensure product quality. Statistical quality control is a collection of statistical data analysis organized in three subsets, each of which are composed of statistical approaches geared toward different stages of the manufacturing process. The three aforementioned subsets are:

- 1. Experiment Design and Analysis
- 2. Statistical Process Control
- 3. Acceptance Sampling

The focus of this dissertation is Statistical Process control, and particularly on its most critical instrument, control charts, which are used to detect variations in the manufacturing process. Physical variability is always present in any productive process, regardless of how well designed, maintained or monitored it is. To be clear, no matter of how effective a process is, no two items created will be identical and there is always a measurable variance between them. That is, in a term, physical variability. Other types of variability may arise in a process that are not attributable to random factors. Variability of this type is mainly caused by the following factors:

a) Improperly tuned machinery, b) Machine operator mistake, c) Low quality or faulty raw material

This type of variability is called specific variability. A process includes only physical variability is considered to be under statistical control (in control process) or alternately it works in a stable state, while one that includes specific variability is considered to be

out of statistical control (out of control process) or alternately that it works in an unstable state. Control charts are an important tool in recognizing the exact type of variability present in the process and can allow people in charge of the production to take corrective actions.

Statistical quality control and its procedures are highly valuable in the modern world, as businesses try to deliver quality products and services in a highly competitive market. When statistical control is used in the production process it allows for fast intervention when the processes are judged to be in an unstable state, controlling the quality of the products and services associated with it. Control charts are one of the most critical tools available to quality assurance professionals, as their dependability is essential for the work. The control points of control diagrams are typically based on estimates for the parameters, which are frequently imprecise. Their parameters are estimated using a reference sample, and their precision has a major influence on the performance of the control diagram. As a result, academics have made great efforts in recent years on investigating the influence of using parameter estimates in constructing the control limits for various control charts and improving their performance. The purpose of this dissertation is to give a critical assessment of the approaches that have been established in this subject.

1.2 Phase I and Phase II control diagrams

There are two phases to control a production process using control diagrams, phase I and phase II.

Phase I: This phase examines samples collected in the past to check whether the process was in or out of statistical control. Phase I control charts help the process manager bring the process into statistical control. When this is achieved the control limits and the center line can be used for future process monitoring. This phase of the charts is also called retrospective. During phase I the administrator should study the process well before deciding when it was in statistical control and when not.

Phase II: In this phase the control charts are used to continuously check if the process remains within statistical control. Since phase I, we have identified the control limits



and the center line so by taking samples at any time the process manager can easily test whether the process remains in statistical control or not.

1.3 Classification of control charts

Control charts are classified into different categories according to their characteristics. In more detail:

1. Depending on the type of variable (continuous or categorical) that describes the quality characteristic we are interested in, we have control charts for variables and control charts for attributes.

2. If samples larger than the unit are taken from the production process, we refer to control charts for rational subgroups, while if the samples are one size, then we refer to control charts for individual observations.

3. If the measurements taken at time t are dependent on the measurements taken at time t - 1 then we refer to control charts for auto-correlated processes. Otherwise, we refer to control charts for uncorrelated processes.

4. If the measurements taken concern only one quality characteristic then we refer to univariate control charts, while if the measurements refer to more than one characteristic then we refer to multivariate control charts.

5. If the measurements come from a known distribution then we refer to parametric control charts, while if the distribution of measurements is unknown we refer to non-parametric control charts.

1.4 Run Length in Statistical Process Control

One of the most famous performance indicators that is used in control charts is the Average Run Length (ARL). Usually, the distribution of Run Length of the estimated charts is right skewed and not geometric as it happens in the case that the parameters are already known. It is important to know the value of ARL in a process to be able to estimate the possibility the process to end up out of control. The general function for control chart like Shewhart is:



$$ARL = \frac{1}{p}$$

when p is the probability that a point will exceed the limits. There are two types of ARLs for investigating the performance of a control chart:

In-control ARL (ARL_0) : It is the average number of control statistics plotted until a false signal (out of limits point) is seen when the process is in-control.

Out-of-control ARL (ARL_1) : It is the average number of control statistics plotted to detect that the process is out of control when the process is out of control.

We can have to types of error in ARL calculations. Type I and Type II error. A Type I error is when the control chart signals as out-of-control when the process is actually incontrol. On the other hand, Type-II error is the situation where the control graph evaluates the process as under control and does not give a signal when the process is actually out of control.

When a Shewhart control chart has 3 sigma control limits (L = 3), the probability α for normally distributed observations is 0.0027. So when L parameter is taken value 3, *ARL*₀ is 370.4.

$$ARL_0 = \frac{1}{a}$$

If ARL_0 is equal to 370 then the process it signals out-of-control every 370 observations. On the other hand, If the probability of Type II error is equal to 1-p then general ARL_1 formula is:

$$ARL_1 = \frac{1}{1-p}$$

While designing the control charts, it is tried to determine the parameters with the highest ARL0 values to give less false signals when the process is under control, and the lowest ARL1 values to give a fast signal when the process is out of control.



1.5 The Problem of Statistical Process Control

In any productive process, no matter how well designed or well maintained, there will always be some form of physical variability in it. This physical variability is the component of many small causes which are referred to as random causes of variability. Physical variability is usually small in size and cannot be reduced or eliminated. A process that works only in the presence of physical variability is called a process within control. In a productive process, however, it is possible for other forms of variability to occur beyond physics, which will not be due to accidental causes and which lead to the systematic change of one or more factors that determine the quality of the product. This kind of non-physical variability is called specific and the causes that lead to it are called specific or systemic causes variability, such are usually the following:(a) incorrectly tuned machines, (b) machine operator errors, and (c) poor quality or defective raw material. When a process operates in the presence of specific variability, we say that it is out of (statistical) control. During the design phase of a product, it is quite important to set control limits for the quality characteristics of that product. These limits are the upper and lower control limits, which are denoted as UCL and LCL respectively, and between them must be the values of the quality characteristic being investigated in order for the final product to be acceptable. Also, in the design phase is defined a desired value for the quality characteristic called target value and is usually in the middle of the space defined by the control limits (LCL, UCL). Under conditions of physical variability, the vast majority of quality attribute values are within control limits, but this is not the case under conditions of specific variability. In case of specific variability there may be a change in the parameters of the distribution that follows the values of the quality characteristic we study. In the case of normal distribution e.g. there may be a shift either in the middle of the distribution, or in its variation, or in both. In any case, the effect of specific variability results in an increase in the products produced that have quality characteristic values outside the control limits. The following figure clearly shows the effect of the specific causes of variability on the mean value $\mu 0$ in the control and on the standard deviation $\sigma 0$ at different times, in detail we have: (a) at time t1 the mean is shifted to the position $\mu 1 > \mu_0$, (b) at time t2 the mean is shifted to position $\mu 0$ while the standard deviation is shifted to position $\sigma 1 > \sigma 0$, and (c) at time t3 the mean is shifted to position $\mu 2 < \mu_o$.



Chapter 2

2.1 Shewhart chart

In 1924 Bell Telephone Laboratories became familiar with the legendary control chart, discovered by Walter A. Shewhart, an American scientist or most of the times known as the father of statistical control chart. The necessity of a method that would be able to control the quality of a product led Shewhart to create a chart that until now has a major role in the fields of manufacturing, engineering, healthcare and procedural research.

The Shewhart control chart is the most famous chart and a major discovery which allows during a process to identify significant and sometimes abrupt changes. On the other hand, Shewhart chart is not so effective when it comes to identify small changes when monitoring the behavior of a product throughout the process. Nevertheless, it is beneficial because it determines if a process remains in statistical control and monitors attributes such as the mean and standard deviation, which are related to the run length distribution.

When it comes to the process of a production the focus remains in observing a crucial quantity of a random variable X, which is a measurable characteristic, regarding the way it behaves during the production of a product. During the production there is a random selection of products, at different times having random samples X1X2, ... of the values of X whom measurements is the base of the process of monitoring the crucial quantity. Those random samples can be used to estimate the value $W_t = g(x_t)$, a function of a random variable that estimates critical quantities such as the mean and the variance of X.

Therefore, when observing several samples, which are acquired by the W function, it's achievable to observe how the critical quantity behaves. For instance, if we want to produce marbles that have an X diameter and we want to observe how the mean value of this diameter behaves, we will select samples of random sizes of marbles where n (n>1) in different time points. Then using the function $W_t = g(X_t) = (X_{t_1} + X_{t_2} + ... + X_{t_m})$ which is an objective evaluator of X's mean, we can observe how the average value behaves.

This is a graphic depiction of how a control chart or Shewhart chart looks like:





Figure 1: Shewhart control chart for marbles

In the diagram above we can observe points that are connected with line and this points are our observed values. The center line or average level of the process usually represents the mean value of W. The top and bottom lines shown in the diagram are called the upper and lower control limit (UCL and LCL, respectively). As long as the values of W are within the control limits we can assume that the process remains under control and we do not need to take any corrective action. However, if a point is found outside the control limits, we say that there is an indication that the process is out of control and we must conduct research to find out the specific causes of variability that are responsible for this behavior and if it is necessary to take corrective action. However, it should be noted that even if all points on the chart are within control but behave in a systematic or non-random way then this is also an indication that the process is out of control.

Below we can observe UCL and LCL limits type:

$$UCL = \mu_0 + L\sigma_0$$
$$CL = \mu_0$$



$$LCL = \mu_0 - L\sigma_0$$

The quantities μ_0 state the mean value and σ_0 state the standard deviation of the function W shown in the control diagram. The quantity L indicates the distance of the control limits from the center line in standard deviation units. Usually L=3, so we are talking about Shewhart control charts with 3σ control limits. In addition to the sigma limit model for constructing Shewhart control charts, there is also the probability limits for a normal or approximate normal distribution of W

$$UCL = \mu_0 + z_{a/2}\sigma_0$$
$$CL = \mu_0$$
$$LCL = \mu_0 - z_{a/2}\sigma_0$$

In Shewhart control charts we distinguish two major categories depending on whether the X attribute is a continuous or discrete random variable. If the random variable X is continuous with mean μ and fluctuation σ 2, then there are Shewhart control charts to monitor the mean and dispersion of X. In case the random variable X is distinct, there are Shewhart control charts to monitor the percentage of defective products produced by the production process, as well as the number (and average number) of defects in a control unit Antzoulakos (2003), Damianou (1996), Kaffes (1996). The simplest and most common Shewhart control chart is the control chart for monitoring the mean value of a continuous X attribute, which we will briefly develop in the next paragraph through an example.

2.2 CUSUM chart

While Shewhart charts are ideal to identify large shifts during a statistical process they are deficient in detecting small shifts in the parameters of such a process. The suggestion of a chart that can identify small shifts in the process was presented by E.S. Page in 1954. The CUmulative SUM or CUSUM chart is a memory-type chart with such a sensitivity to be able to detect those small and non-extreme shifts. CUSUM control charts are improved monitoring tools and belong to the category of charts that

were created to monitor binomial data. They define if a process is in control or out of control depended on collected sums, while monitoring the process.

In comparison with the Shewhart chart, CUSUM chart can provide more data regarding the process, because each one of the points that are plotted in the graph is extracting information from many samples, something that Shewhart chart is not suitable to do and that is one of the most important reasons that CUSUM chart is better when it comes to the detection of small shifts. Those charts are the most suitable ones in case someone needs to analyze separate (individual) observations.

In CUSUM chart, all sample values are taken into account by plotting the cumulative sums of deviations of sample values from the target value on the control charts:

$$C_i = \sum_{j=1}^i (\bar{x}_j - \mu_0)$$

Depending on whether the process is under control or not, the *Ci* values, which show the deviation from the target value, show different trends. When the process is in control (the process mean is equal to the target value μ_0), the cumulative sum is a zero mean random walk. A positive trend in cumulative total Ci is seen if the process mean increases (if the process mean is greater than the target value, $\mu_1 > \mu_0$). If the process mean decreases ($\mu_1 < \mu_0$ if the process mean is less than the target value), a negative trend is seen in the cumulative total C_i . According to the trends detected in the process, deviations in the process average can be determined.

The construction of the diagram is based on the following algorithm (Algorithmic Construction Method):

$$C_i^+ = max[0, x_i - (\mu_0 - K) + C_{i-1}^+]$$
$$C_i^- = max[0, (\mu_0 - K) - x_i + C_{i-1}^-]$$



The initial values of the C_i values are taken as zero ($C_i^+ = C_i^- = 0$). Where, K is called the reference value and usually takes a value between the mean μ_1 that we want to determine and the target value μ_0 .

$$K = \frac{\delta}{2}\sigma = \frac{|\mu_1 - \mu_0|}{2}$$
$$\mu_1 = \mu_0 + \delta\sigma$$
$$\delta = \frac{|\mu_1 - \mu_0|}{\sigma}$$

If in the diagram we construct we use the C_i^+ and C_i^- , then we use the two-sided CUSUM control diagram, while if we use only one of the two, then we use the one-sided CUSUM control diagram. One-sided CUSUM diagrams are suitable for the detection of increases (the C_i^+) or reductions (the C_i^-) while the duplex can detect either increases or decreases in the middle of the process. In a one-sided CUSUM control diagram, we show only one of them on the figure C_i^+ or C_i^- and consider whether it exceeds the value H. If one of the C_i^+ and C_i^- values exceed the limits then the process is considered out of control.

If C_i^+ ηC_i^- is equal to 0, then the process is in control and we discard 1. But if either is different from 0, then consider if $C_i^+ > \text{ or } C_i^- < -$ where $H = h\sigma$. The value *h* is known as the decision interval and plays the role of control limit. It is worth noting that even in the case of CUSUM control diagrams, the distribution of the number of points until we have for the first time an indication out of process control is not Geometric and therefore the *ARL* is not the average value of a Geometric distribution. For example, if we want to find a change in the mean for = 1, then the pair we will use will be k = 0.5and h = 4.77, as long as these are the prices they will give *ARL*0 = 370.

To calculate the ARL of a two-sided CUSUM control chart, we usually calculate the ARLs for the upper one-sided and the lower one-sided CUSUM control chart, i.e. the ARL^+ and ARL^- respectively.

Another option we have in CUSUM control charts is the application of the Fast Initial Response (FIR) method. In case we study a process, which even after the necessary corrective actions is out of control, we can calculate the C_i^+ and C_i^- starting from $C_0^+ \neq 0$

and $C_0^- \neq 0$. We usually choose $C_0^+ = \frac{H}{2}$ and $C_0^- = -\frac{H}{2}$. Applying this method, in case the process will be out of control again, the increased values for C_0^+ and C_0^- will help us to detect the change in the middle of the process faster. If the process is in control, the above method may cause an initial disturbance in the control diagram but will not affect us.

Finally, it is worth noting that the CUSUM scale has generally limited use since although it is sensitive to detecting increases in process variability, it is not particularly effective in detecting reductions (small and / or medium). In addition, the above methodology may allow us to detect the change in process variation, but it is not always easy to distinguish this change from the change in the medium (ie the indication that the process is out of control may be due to a change in the medium). and not in a change of its variation).

2.3 EWMA chart

Another memory-type control chart like CUSUM is the EWMA control chart. In 1959 Roberts introduced the exponentially weighted moving average control chart which is also an alternative option when someone wants to detect small shifts when observing the process average or mean. It uses past data that are summed up to the current data. While both CUSUM and EWMA are very similar charts there is an important difference between them when it comes to individual observation. The EWMA chart is insensitive when assuming the regularity of observations and that makes it more suitable in that case than CUSUM. Furthermore, in EWMA control charts, before a signal out of control is being found in the process, there is no geometric distribution.

On the other hand, EWMA control chart is more difficult to interpret in comparison to Shewhart chart. Also, there is a delay in finding the shift in EWMA chart when the mean is moved on one side and the values are on the other side. Despite those peculiarities and despite EWMA being overlooked in the past, it is nowadays one of the most useful tools in quality control. Later work as well as the execution of computer science within the research facility space have invigorated intrigued in this quality control strategy.



In contrast to the Levey-Jennings diagram, the EWMA diagram itself does not use the control value. Instead, the control values (xi) are transformed into z_i values according to the following equation:

$$z_i = \lambda x_i + (1 - \lambda) z_{i-1},$$

i, represents the observation value over time, \mathbf{z}_i represents the new weighted average and λ the weighting parameter (λ takes values between $0 < \lambda \leq 1$). The weighting factor, indicated by the Greek letter, is the coefficient that defines the reliance of each z_i value on its predecessor. The EWMA chart's capacity to detect tiny shifts or variations of the control values around the target value, i.e., systematic mistakes, is responsible for the elimination process. The characterization moving instrument is derived from the value zi, while the characterization barycentric is derived from the coefficient of gravity. The exponential characterisation is derived from the exponential functions used to calculate the Method's control limits. The EWMA diagram's creation necessitates the continuous solution of complicated equations. Before beginning any calculations, the average value $(z_0 = \bar{x})$ and standard deviation $(\sigma_{z_i}^2 = \sigma_0 \left(\frac{\lambda}{2-\lambda}\right) \left[1 - (1-\lambda)^{2i}\right])$ of the laboratory's control samples must be determined. Although it is standard practice to compute and utilize the analyzer manufacturer's control limits, it is better to calculate and in the laboratory itself. As previously stated, the control material values are represened in the EWMA diagram as z_i values. The goal value of the control limits is indicated as CL and is the first value of equation 1, i.e. the first value z_0 (Central Limit). The mean value of the control limits is equal to Z. The x_i control values are more satisfying the closer the z_i values are near CL. The gravity coefficient is a number that ranges from 0 to 1. The upper control (Upper Control Limit or UCL) and the lower control (Lower Control Limit or LCL) are given by the equations:

$$UCL = \mu_0 + L\sigma \sqrt{\frac{\lambda}{(2-\lambda)} [1 - (1-\lambda)^{2i}]}$$
$$CL = \mu_0$$

$$LCL = \mu_0 - L\sigma \sqrt{\frac{\lambda}{(2-\lambda)} [1 - (1-\lambda)^{2i}]}$$



The above limits are the variable limits, ie for $\lambda \neq 1$. Nevertheless, the percentage (1- λ) ^ 2i tends to 0 as i increases. Therefore, the limits are stabilized and are given by the following relations:

$$UCL = \mu_0 + L\sigma \sqrt{\frac{\lambda}{(2-\lambda)}}$$
$$CL = \mu_0$$

$$LCL = \mu_0 - L\sigma \sqrt{\frac{\lambda}{(2-\lambda)}}$$



Figure 2: Example of Ewma control chart



The parameters *L* and λ are used in the design of EWMA diagrams. Depending on the values given in the parameters, they will have different results in ARL. In certain studies, the ARL performance of EWMA control diagrams has been calculated with different values designed for the parameter λ . λ usually takes the value 0.05, 0.10, 0.20. In practice, from the studies that have been done, it has been determined that λ gives better results in the range $0.05 \le \lambda \le 0.25$, with λ usually taking the value 0.05, 0.10, 0.20. If we have the case where $\lambda = 1$ and L = 3, then the ARL tables show the same results as the Shewhart tables.

According to Montgomery (2013) the values of λ and L are calculated so that when we have an in-control procedure the ARL is equal to 370. In the Ewma control chart the performance of the ARL is better in small deviations but does not respond as fast as the Shewhart diagrams in large deviations. Finally, when $\lambda = 0$ the values zi and the limits UCL and LCL are equal to the mean value μ .

2.4 Multivariate Control Charts

Multivariate control charts are used to track activities with a variety of quality characteristics. These control charts are made in the same way as analogous univariate ones, and they are separated into two sections. Phase I entails gathering an in-control reference sample of m subgroups of size n and estimating the values of the process parameters in the in-control state using the data. It is possible to utilize a sample size of one, individual observations, or many observations. With the help of control limits and the parameter values in phase II, the process may be monitored. Following we can see the vector with quality features:

$$x_{ik} = (x_{1ik}, x_{2ik}, x_{3ik}, \dots, x_{pik})$$

with k = 1, 2, ..., m and i = 1, 2, ..., n

We must know that $x_{ik}s$ is following a $N(\mu, \Sigma)$ distribution, m is the number of subgroups in phase I (m must be greater or equal to one and m(n-1) > p) and p is the number of quality characteristics. Below we can see the chart statistic of the chi-square control chart if the parameters are known:

$$x_j^2 = n(\bar{x} - \mu)' \Sigma^{-1}(\bar{x} - \mu)$$

where

$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_p \end{bmatrix}$$

With the Upper control limit of the chart:

$$UCL = x_{a,p}^2$$

Maximum likelihood estimator is used for every quality characteristic when parameters are unknown. The equation of the sample mean is:

$$\bar{\bar{x}} = \frac{1}{m} \sum_{k=1}^{m} \bar{x}_{jk}$$

but in this case:

$$\bar{x}_{jk} = \frac{1}{n} \sum_{k=1}^{n} x_{i \ jk}$$

with j^{th} be the characteristic and k^{th} is the subgroup.

The vector \overline{x} give us the multivariate sample mean:

$$\bar{\bar{x}} = \left[\bar{\bar{x}}_1, \bar{\bar{x}}_2, \dots, \bar{\bar{x}}_p\right]$$

The equation for the sample covariance of p x p matrix is:

$$\bar{S} = \frac{1}{m-1} \sum_{k=1}^{m} (x_{ik} - \bar{x}) (x_{ik} - \bar{x})'$$

In phase I limits of the control chart are:

$$UCL = \frac{p(m-1)(n-1)}{mn - m - p + 1} F_{a,p,mn-m-p+1}$$
$$LCL = 0$$

2.5 EWMA-Poisson chart

In addition to the normal distribution the data in a productive process of an industry can also come from the Poisson distribution. Below we will describe how the EWMA diagram works when the data comes from a Poisson distribution.

The Shewhart c-chart was initially used for these cases, but the Poisson-Ewma chart showed that it detects small shifts of the mean faster than the Shewhart c-chart. So we assume that we have data that follows the Poisson (μ) distribution. Below we can observe the statistical representation of Ewma control chart:

$$Z_t = (1 - \lambda)Z_{t-1} + \lambda X_t$$

with t = 1, 2, 3

where $Z_0 = \mu_0$. The starting value of Z_t is Z_0 and when the process is in-control $Z_0 = \mu_0$. Furthermore λ is the smoothing parameter. Below is the expected value of EWMA and variance:

$$E(Z_t) = \mu_0$$
$$Var(Z_t) = \frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2t}] \mu_0$$

But if we have large values of t then the variance will be:

$$Var(Z_{\infty}) = \frac{\lambda}{2-\lambda}\mu_0$$

In the occasion that we want to create the control limits of the chart, then we will use the second case of the variance. The control limits of Poisson EWMA are the following:

$$h_U = \mu_0 + A_U \sqrt{\frac{\lambda}{2 - \lambda} \mu_0}$$
$$h_L = \mu_0 - A_L \sqrt{\frac{\lambda}{2 - \lambda} \mu_0}$$

If $Z_t > h_U$ or $Z_t < h_L$ then the Ewma chart signals us about a value out of the limits.

The appropriate values for A_L , A_U and λ are chosen, so to secure the desired ARL_0 . In addition, the values of λ depend on how quickly it is necessary to detect an average shift. When λ receives low values then we have faster detection of small and medium shifts. If the parameters are unknown, then the maximum likelihood will be the estimator for in-control mean. $\hat{\mu}_0 = \frac{1}{n} \sum_{i=1}^n x_i$



n represents the sample size in phase I.

$$h_U = \hat{\mu}_0 + A_U \sqrt{\frac{\lambda}{2 - \lambda} \hat{\mu}_0}$$
$$h_L = \hat{\mu}_0 - A_L \sqrt{\frac{\lambda}{2 - \lambda} \hat{\mu}_0}$$

2.6 Parameter Estimation for EWMA control chart

As we mentioned earlier in case we don't know the mean and standard deviation we can then estimate those values. Suppose we have x_{ij} for i = 1, 2, 3, ..., m and j = 1, 2, ..., n, which are data that have been extracted from a normal distribution. We also know that the parameter λ has a range from 0 to 1.

$$Y_i = \lambda W_i + (1 - \lambda) Y_{i-1}$$

The first difference with the general function that is used in EWMA is x_i is being replaced from W_i which function is:

$$W_i = \frac{\bar{X}_i - \hat{\mu}_0}{\frac{\hat{\sigma}_0}{\sqrt{n}}}$$

From the function above $\hat{\mu}_0$ and $\hat{\sigma}_0$ are the parameters of phase I. According to Mandla Diko, Chakraborti, Ronald Does (2018):

$$\hat{\mu}_0 = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n x_{ij}$$
$$\hat{\sigma}_0 = \sqrt{\frac{1}{m} \sum_{i=1}^m S_i^2} = S_p$$

The S_i^2 is the variance from each sample of phase I. The aggregate standard deviation estimator, which is one of the estimators for σ_0 that are used most of the times, gives for the mean squared error the lowest values (Mahmoud et al18).



Diko et al, 19 on the other hand highlights that based on Montgomery,20 Schoonhoven et al21, 22 $\hat{\sigma}_0 = \frac{s_p}{c_4(m(n-1)+1)}$ is of equal value and the reason is that m (n - 1) is usually large in those functions and therefore the constant c4 (m (n - 1) + 1) is identical to 1.

The function of W_i normally is:

$$W_i = \frac{1}{Q} \left(\gamma T_i + \delta - \frac{Z}{\sqrt{m}} \right)$$

Where

$$T_i = \frac{\overline{X}_i - \mu}{\frac{\sigma}{\sqrt{n}}}, \ Q = \frac{S_p}{\sigma_0}, \ Z = \frac{\widehat{\mu}_0 - \mu_0}{\frac{\sigma_0}{\sqrt{mn}}}, \ \gamma = \frac{\sigma}{\sigma_0}, \ \delta = \frac{\mu - \mu_0}{\frac{\sigma_0}{\sqrt{n}}}$$



Chapter 3

3.1Effect of parameter estimation on control diagrams for the mean value

The majority of studies that have based their research in the supposition that the parameters are known and the process is in-control, have used the developed control charts in phase II. For instance, we are interested in the process parameter when the mean μ and the standard deviation σ , are in-control and a quality characteristic follows the normal distribution. The hypothesis that the in-control values of the parameters are known make simpler the development and therefore the calculations of several quantities that are related in the control diagrams. In practice, however, the parameters are rarely known and the control limits are usually based on the estimated parameters. However, when it comes to the practical part because the parameters are most of the time unknown the control limits are mostly calculated using the estimated parameters. When the known parameters are replaced from the estimations the performance of the charts is affected by the changeability of the estimators and that causes a difference between the performance of the charts which use known parameters. Also, many researchers have highlighted that there is a difficulty in defining the control limits, while using a short amount of data or using data from samples that are not in any way representative.

Shewhart (1939) pointed out that: "In most practical cases, the most difficult part of all is choosing the sample to be used as the basis for determining the control limits."

We will be completely sure for the accuracy of the control limits, when a representable sample has been gathered.

An empirical rule used mainly for control charts is to take m = 20 to m = 30 preliminary samples of size n = 4 to 6 n = 6 in phase I (Montogomery, 2005). However, these rules are mainly based on empirical data and are usually insufficient. The inadequacy of these rules for the selection of the number m and the size n of the samples led to the conclusion that the effect of the parameter estimation should be seriously considered when designing a control diagram.

3.2 The distribution of run length when estimating the parameters of control chart

If the run length distribution is geometric, then the mean ARL flow length fully characterizes the flow length distribution and is the acceptable efficiency measure. For simplification of comparisons but also for a better picture of the run length distribution, concise values of the run length distribution are usually used in order to evaluate in an

UNIV

effective way the performance of the control charts using estimated parameters. For example, the average run length (ARL), the standard deviation of the run length (SD or SDRL) and the percentages of run length distribution are often used. Therefore, if the value T that represents the run length, if only the data follow a normal distribution and we have 3s as control limits in phase II, follows the geometric distribution with parameter p = 0.0027

which represents the probability of a point in the diagram to be found outside the control limits, then the mean value of T and its standard deviation will be respectively:

$$E(T) = \frac{1}{p} = 370.4$$

 $SD(T) = \frac{\sqrt{1-p}}{p} = 369.9$

Additionally, its second moment of T is the following

$$E(T^2) = \frac{2p}{n^2}$$

When the control limits of the diagram are unknown and must be estimated then the run length distribution is not geometric but is unknown. This means that when the parameters of the distribution in phase I, for example, the mean and standard deviation must be estimated, we cannot calculate the mean ARL run length and its standard SDRL deviation. That is the reason several authors have conducted studies to carefully calculate these values when estimating the parameters in phase I. The accuracy of the approaches is studied mainly through simulation methods.

3.3 Studies in parameter estimation

Prochan & Savage (1960) examined the impact that the size n and the number m from the samples they had collected, have on the performance of X control chart, regarding. the possibility that the sample of mean X is out of control limits, when in reality we are in-control when the unknown parameter of the standard deviation σ is estimated through the range $\overline{R_m}$ and Sp. Having given values of size n of the samples they provided values for the number m of the samples that needed, so to have a specific value of the former possibility incase the process is in-control.

However, Prochan and Savage (1960) in their study did not consider the correlation between the probability that the sample mean Xi of sample i exceeds the upper UCL control limit and the probability that the sample mean of another sample j exceeds UCL control upper limit. For this reason, the results of this study are of limited use. Hillier (1969) addressed the above problem of estimated parameters in the case of X and R control diagrams. More specifically, he presented a method that was able to calculate the possibility of type I error, in the case of the \overline{X} diagram, using the R range to estimate the standard deviation of the process.

However, the weakness of the method was the same as in Prochan and Savage (1960). Ghosh et al (1981) in their study gave formulas for calculating the flow length distribution of X control charts with unknown variation. The purpose of their study was to examine the already known criteria that have been proposed and that calculate the performance of a control chart in terms of their suitability with special reference to Shewhart X charts for controlling the mean value when the process dispersion is unknown. Also, another important element in their article refers to the creation of an economic model that assesses the impact on net income and costs associated with the construction of a control chart. The idea for economic models was first started by Duncan (1956) and later by other authors (Chiu & Cheung (1977). Another approach to the effect of parameter estimation was proposed by Chakraborti (2007). Chakraborti considered that traditional measures for the performance of a control chart, such as the mean ARL flow length and the standard SDRL deviation, do not provide a complete and satisfactory picture of the chart performance as the flow length distribution is asymmetric to the right. Also, there is a more general concern that it is possible to lose the important information that can be extracted from the performance of a control chart if one focuses too much on the average ARL run length. For this reason, the author suggested that the percentiles of the run length distribution provide a much better indication of the performance of the control chart.


3.4 Effect of parameter estimation on control diagrams for dispersion

The R, S and S2 control charts are widely used in practice to monitor the dispersion of a process. The usual technique in process control is to estimate the control limits resulting from preliminary samples, which is phase I, and to use these estimates to create the control limits in phase II. However, few studies have been performed on the flow length distribution in dispersion control diagrams when the control limits are based on estimated parameters.

Chen (1998) considered the case where the standard deviation σ is unknown. As an aftermath the control limits that are used for monitoring the variation of the process in phase I are the:

UCL=Unσ CL=σ LCL=Lnσ

where σ is the estimation of standard deviation σ , each one of them based on m preliminary samples of size n, the Xij, $1 \le j \le n$, $1 \le i \le m$, and Un, Ln are suitable constants that control the probability α of the type I error.

For a control chart that has the estimated control limits mentioned above, we consider the possibility

Fi: the estimation of the standard deviation σi $(1 \le i \le m)$ is out of control, either above or below the UCL and LCL respectively.

The probability of {Fi} occurrence is calculated from the relation

 $P(Fi) = P(\sigma i < LCL \text{ or } \sigma i > UCL).$

We also define the random variable W which indicates the number of samples until the first Fi probability.

The distribution of the random variable W is not geometric, because the contingencies {Fi} are not independent of each other.

Hillier (1969) studied the problem of defining m, n to produce a specific type I error in the R diagram. In particular, he observed that when someone uses the estimated control limits, the constants Un and Ln arising from the case of known parameters, (i.e. the known standard deviation σ), cease to create the desired probability of type I error. However, assuming the normality of the data, he managed to modify the constants Un and Ln for different values m and n to guarantee the maintenance of the desired probability a of the type I error. Regardless of the alterations of the constants Un and Ln, the distribution of the random variable W is not geometric.



Chapter 4

4.1 CSEWMA control chart

In chapter 2 we mentioned that the difference between the Shewhart and EWMA diagrams is that the Shewhart charts are detecting faster large deviation in the mean value while EWMA charts are better at detecting small deviations of the mean value. Lucas and Saccucci are the first who mentioned the possible combinations of those two diagrams, presenting a combined diagram named CSEWMA. Let's examine the structure and equations of the CSEWMA. Consider that $x_{t1}, x_{t2}, ..., x_{tn}$ is our sample that consists of observations following normal distribution. If \bar{x}_t is the mean of the sample then the corresponding standardized mean will be:

$$D_t = \frac{\bar{x}_t - \mu_0}{\sigma_0} / \sqrt{n}$$

with $D_t \sim N(\delta, 1)$ and $\delta = \frac{\sqrt{n}(\mu - \mu_0)}{\sigma}$

where δ is representing the shift of mean.

In two conditions we have a signal from the CSEWMA:

The equation for Z_t is:

$$Z_t = (1 - \lambda)Z_{t-1} + \lambda D_t$$

Furthermore, we know for the initial value of Z being $Z_0 = 0$ and the parameter $\lambda \in (0,1]$. The run length until we have an alarm from the CSEWMA chart is given by the equation:

$$RL_{CS} = min(RL_{SH}, RL_{EW})$$

with:

- $RL_{SH} = \inf\{t: |D_t| > k\}$
- $RL_{EW} = \inf \{t : |Z_t| > h \sqrt{\frac{\lambda}{(1-\lambda)}}$



Here we have to specify two distinct cases. The first is when $h = \infty$ where the CSEWMA chart turns into a Shewhart chart, and the second case is when the $k = \infty$ where the CSEWMA chart turns into a EWMA chart.

4.2 Combination of EWMA and CUSUM diagrams for simultaneous media and dispersion monitoring

One of the most common used techniques to monitor the occurrence of causes of variability in a process is to use more than one diagram of the same type. The mean chart can be affected from a shift in variance in a combined and shared monitoring. This can happen because both can be shifted at the same time, due to the participation of the variance in the control limits of the mean diagram. These diagrams are required to monitor the medium and the dispersion of the process simultaneously.

Memory control charts such as EWMA and CUSUM use a combination of previous and current information to study the subject of interest. Abbas et.al proposed an advanced control chart, which is the combination of EWMA and CUSUM control charts. This mixed control chart showed that can achieve the best results when it comes to detect smalls shifts in the process mean.

The diagrams of Zaman, et al. are based on the proposal of Abbas, et al., from segregated to combined process supervision. There are three proposed diagrams: the Combined Mixed EWMA-CUSUM (Combined Mixed EWMA-CUSUM, CMEC), the Combined Mixed Double EWMA-CUSUM (Mixed Double EWMA-CUSUM, CMDEC) and the Combined CUSUM (Combined CUSUM).

The following transformation is being used, so that *xij* represents a specific qualitative feature of the process we want to monitor. *xij* follows normal $N(\mu+\delta\mu,\sigma\delta\mu)$, with mean μ and standard deviation σ . Assume the process is under control: $\delta\mu = 0$ and $\delta\gamma = 1$. Suppose that i = 1, 2, ..., n its indicators *xij*, divided into crowd groups *i* and size *n*.

4.3 Combined EWMA – CUSUM (CMEC) chart

To be able to construct the combination of EWMA and CUSUM control charts, we will have to combine the equations of those charts. The combination provides us with the following equations:

$$CMECL_{i}^{+} = \max [0, (Y_{i} - \mu) - K_{i} + CMECL_{i-1}^{+}]$$

 $CMECL_{i}^{-} = \min [0, (Y_{i} - \mu) + K_{i} + CMECL_{i-1}^{-}]$



$$CMECV_{i}^{+} = \max \left[0, (Z_{i} - \mu) - K_{i} + CMECV_{i-1}^{+}\right]$$
$$CMECV_{i}^{-} = \min \left[0, (Y_{i} - \mu) + K_{i} + CMECL_{i-1}^{-}\right]$$
where, $K_{i} = k * \sqrt{Variance(Y_{i})} = k * \sigma_{Y}^{2} \left[\frac{\lambda}{2-\lambda} \left(1 - (1-\lambda)^{i}\right)\right]$

From the equation of K_i , k is a constant. If $CMECL_i^+$, $CMECV_i^+$ exceed the H_i or, $CMECL_i^-$, $CMECV_i^-$ are lower than $-H_i$ then the process is out-of-control. The equation for H_i is the following:

$$H_i = h * \sigma_Y^2 \left[\frac{\lambda}{2 - \lambda} \left(1 - (1 - \lambda)^i \right) \right]$$

4.4 Conclusion

In conclusion, to achieve a sufficient performance compared to that of the known parameters it is necessary to obtain more data in phase I and in general the more parameters are estimated the larger sample sizes are required. For example, multivariate diagrams require more data in phase I than uni-variate control diagrams. Thus Chen makes it necessary to take at least 75 samples in phase I with a sample size of at least 5 in the dispersion control diagrams (R, S and S₂). Regarding the X control diagrams for the mean value, he concludes that the effect of parameter estimation is more severe for small shifts in the process than for larger ones and in addition that the effect is greater for SDRL values compared to those for ARL. In contrast, Maravelakis et al. (2002) suggested larger sample sizes in the dispersion control diagrams, ie 100 samples of at least 20 size as for smaller samples they considered the effect to be quite severe resulting in false alarms in the process, while in the individual observation control diagrams they considered that at least 300 comments. These propositions are of course in contradiction with the empirical rule for the selection of the number of samples which are m = 20 or m = 30 with quantities n = 4 or n = 5. In addition, for the detection of small displacements of the process it is suggested to use diagrams more sensitive than Shewhart control charts such as CUSUM or EWMA charts. Research in the area of the effect of parameter estimation on control charts is not exhaustive in the context of this postgraduate thesis. The study of the flow length distribution margin is of particular interest as in the literature several authors, while dealing with this issue, did not make suggestions for the selection of the samples, because they did not consider the distribution margin. Another research area that needs more study by researchers

concerns robust estimators and their use in aberrant measurements in phase I (Rocke (1989, 1992), Tatum (1997), Vargas (2003), Davis and Adams (2005). In addition, the sensitization rules of a Shewhart control chart with estimated parameters used by researchers are the reason why the chart does not perform as well as in the case of known parameters. Moreover, the use of multiple rules at the same time makes it difficult to analyze the properties of control diagrams with estimated parameters, which is why further study is needed on this subject (Champ and Woodall (1987), Burroughs et al. (1993)). Also, more emphasis has been placed on media tracking diagrams than on diagrams to monitor dispersion, but it is also worth noting that multivariate control diagrams for monitoring dispersion have not been adequately studied. Finally, most authors agree that it is necessary to take as many samples as possible in phase I which guarantee a satisfactory performance in phase II. In case the data in phase I is not sufficient, either corrections in the control limits or their renewals are suggested when more samples become available. (Neduraman and Pignatiello (2001), Tsai et al. (2004,2005), Albers and Kallenberg (2004a, 2004c).



Chapter 5

5.1 Introduction

With the coming of the fourth industrial revolution, which is a global phenomenon, and the integration of robots and artificial intelligence in the working society, the workplace as well as the society itself will face many challenges in order to respond to the high demands of this continuous growth and transformation of professions and working positions.

The effects of the fourth industrial revolution will have an impact in manufacturing developments and as a result the industry from businesses to ethnic systems will be forced to change. Innovation is becoming a necessity for the future of companies and national systems engaging them to new approaches for forward thinking and be pioneering. The phenomenon of globalization has played a key role in the trend of innovation in countries that have developed their markets around the world.

All industrial revolutions were a stepping-stone that led to more modern and improved methods of production in multiple fields such as industry, healthcare, transportation, agriculture resulting in production development and economic growth. The Second industrial revolution was a mainspring for important discoveries and between them was Shewhart's Control Chart. Shewhart introduced his famous chart in a period that history experts refer to as the aftermath of economic growth and technological revolution.

The needs of specialist knowledge in the field of science were a pressure to universities to give more fundings for programs that would lead to the evolution of new ways on handling this rapidly increasing environment. Bell System company used statistics in its effort to create a telephone switching office in a more efficient way, while at that time the engineers that worked for this project were not yet specializing in statistics. Shewhart had collected this kind of previous data that assist in his studies culminate to the discovery of his control chart.

Although Shewhart's chart was a significant discovery, it wasn't immediately acceptable from the employees of Bell Laboratories. The workers were worried that this new development would decrease their wages, because the adoption of using the control chart was delaying their work, which was directly linked to their wages.



With the reported evolution in data collection processes, we note the production systems used having gradually more qualitative features that are being automatically observed while human interaction decreases.

As we have mentioned earlier before the creation of a control diagram is split into two phases, phase I in which we are checking for outliers and we estimate the unknown parameters, and phase II where we draw a control diagram with the parameters estimated to be able to monitor this process online. In phase 2 the statistical testing of the drawn control diagram is being drawn in time and is compared to the respective control limits the breaching of which is an indication of an out-of-control process.

We are noting the connection between the past (meaning data collected beforehand) and the estimated operation of the future. With the technological progress there is a effort to minimize human interaction in please II and have it automated as much as possible with appropriate algorithms who can support a goal like this. Comparatively, it is very hard for such a process to be appropriate in phase II where good knowledge of statistical operations is necessary to estimate the parameters.

Following that we will focus on Exponentially Weighted Moving Average control chart (EWMA) specifically when the parameters are not known and have to be estimated. A lot of estimators were made on estimating parameters for the EWMA control chart. Our test will estimate an algorithm who will analyze the variability of the EWMA control chart by estimating unknown parameters from samples of data where we will contaminate the sample with outliers to be more representative. Our test runs will run either with clearing the data from outliers (not with simple clearing but with a specific method discussed later) or with letting the data as is (with the outliers in our sample). The process that we are going to use in phase 1 is algorithmic and does not require human interaction while running, as the algorithm is going to receive the sample and execute all the steps that we have programmed. To enhance the results from our sample, in every test run we will try different sample size and different contamination ratio, and we will evaluate 6 different estimators of standard deviation. In the following chapters we will analyze the process of executing outlier detection and parameter estimation for phase I and then we will create an EWMA control chart for phase II of statistical control.



5.2 I-Control chart and parameter estimators in phase I

The I-chart is specialized in finding large variations of the mean. The limits of the chart follow the following equations:

$$UCL = \mu_0 + L\sigma_0$$
$$CL = \mu_0$$
$$LCL = \mu_0 - L\sigma_0$$

With μ_0 being the mean value, σ_0 being the standard deviation and L being the distance of the control units from the center line.

We have two parameters that are necessary to estimate, the mean and the standard deviation. For the mean the type of parameter that is used by the algorithm is the following:

$$\bar{x} = \sum_{i=1}^{m} x_i / m$$

where x_i is an observation from our sample with i = 1, 2, ..., m

For the value of μ_0 the estimation is considerably simpler compared to estimating the σ_0 value because for the latter a lot of studies have been made but no one concludes with conviction on what is the best estimator. In the following table we are going to present the 6 estimators this dissertation is going to use.



Table 1: Standard Deviation Estimators

Estimator	Formula
Sample Standard Deviation (SSD)	s/c_4
Mean Moving Range (MnMR)	$\frac{MR}{d_2}$
Median Moving Range (MdMR)	$1.047 * median(MR_i)$
Root Mean Square Successive Differences (RMSSD)	$\frac{\sqrt{\frac{1}{2(m-1)}\sum_{i=1}^{m-1}(x_{i+1}-x_i)^2}}{1-\frac{3}{8m}}$
Median Absolute Deviation (MAD)	$\frac{median(x_i - x)}{0.6745}$
Interquartile Range (IQR)	$\frac{\hat{x}_{(0.75)} - \hat{x}_{(0.25)}}{1.349}$

One of the most common and well-known estimators for the I-chart is Sample Standard Deviation (SSD) estimator, being very effective when our data are resulting from normal distribution and do not contain outliers. As we can see in table 1 the equation of the estimator is:

$$\hat{\sigma} = \frac{S}{c_4}$$
, where $s = \sqrt{\frac{\sum_{i=1}^{m} (x_i - \bar{x})^2}{m-1}}$ and $c_4 = \sqrt{\frac{2}{m-1}} * \frac{\Gamma(m/2)}{\Gamma(m-1/2)}$

with c_4 being a unbaised constant and $\Gamma(m/2)$ being the gamma function for i =1,2, ..., m being the value of each consecutive observation of our sample.

The next estimator is Mean Moving Range (MnMR) who, after considerable studies, has been found not very effective for data who are following the normal distribution and are statistically independent from each other. Cryer and Ryan et al. (1990) have also mentioned the ineffectiveness of MnMR compared to SSD. In effect, the best and safer results will be to calculate both estimators in practice and if we receive INIVERS inconsistent results take it as a sign that our data need closer analysis. The MnMR equation is:

$$\hat{\sigma}_{MR} = rac{MR}{d_2}$$
 , with $MR_i = |x_{i+1} - x_i|$ and $d_2 = 1.128$

 d_2 is a constant applied when our data follow the normal distribution and MR is the Mean Moving Range.

The Mean Moving Range estimator (MdMR) shows better ANTAPOKRISI and, as a result, better results from the last two estimator (SSD, MnMR) when there is a middle percentage of outliers in our data, following the studies of Bruce et al. (1997). When we have a high percentage of outliers the best results seem to be by using the MnMR based on Boyles et al. (1997). The value 1.047 in the estimator's equation is a unbaised constant used with the condition that our observations are normally distributed. The Root Mean Square Successive Differences (RMSSD) estimator based on the studies of Atalay, Testik, Duran et al. (2020) performs slightly better than MnMR when the process is in-control, but also when we have a small percentage of outliers in our data and the process is out of control. The equation of RMSSD estimator is:

$$\frac{\sqrt{\frac{1}{2(m-1)}\sum_{i=1}^{m-1}(x_{i+1}-x_i)^2}}{1-\frac{3}{8m}}$$
, with m being the total number of observations.

The next estimator is Median Absolute Deviation (MAD) who is found by the equation:

$$\frac{median(|x_i - x|)}{0.6745}$$

And is inadequate, based on Braun and Park et al. (2008) for data with outliers. Additionally, based on Boyles et al. (1997) the MAD estimator is unreliable compared to MnMR when our data follow a pattern. The last estimator we are going to study is Interquantile Range (IQR) with equation:

$$\frac{\hat{x}_{(0.75)} - \hat{x}_{(0.25)}}{1.349}$$



Based on the studies of Karagöz D. et al. (2016) the IQR estimator is performing better for a large sample size of data regardless of the deviation of such data, especially when we have a high percentage of outliers in the sample.

5.3 Phase I : outlier detection and parameters estimation with Algorithmic usage

The arrangement of the algorithm for phase I, in particular the process to identify and remove the outliers, is with the following method. First of all, for the needs of our tests we create a sample of data following a normal distribution with median 0 and standard deviation 1. Afterwards we randomly select some observations and replace them with outliers, to have a more realistic results from our sample compared to real data. The outliers are following normal distribution with median 4 and standard deviation 1. The size of our sample and number of outliers are different in consecutive testing to see how the estimators are responding to each sample size. For example, our sample might contain 50, 100, 200, 500, 1000, 10.000 observations while the Contamination Ratios¹ (CR) is respectively getting values 0%, 2%, 4% and 10%. The case where CR is equal to 0% is for a sample with no outliers.

Using the I-Chart, we calculate the initial limits (UCL, LCL). It is necessary for the calculation of the limits to give an initial L value, and the initial L values are between 2 and 4. According to Murat Atalay, Murat Caner Testik, Serhan Duran (2020) and Yao, Y., & Chakraborti, S. (2021) best value for L is 3. Afterwards we examine the observations one by one to see if any of them are out of the limits. When we have no observation that is out of limits, we move to the next step, but if there is at least one we remove it and calculate the limits with the rest of the observation. This process is executed until no observation is out of limits.

In following, the estimation of the parameters is done by the algorithm, possibly having outliers in the data but after the process of their elimination. As we said before we used the estimators SSD, MnMR, MdMR, RMSSD, MAD και IQR.

In the following 2 diagrams we can see how the process of phase I is executed through diagram depiction. In the first diagram there is no elimination of the outliers neither



¹ Contamination Ratios (CR) is the percentage of outliers present in our sample.

recalculating the limits that is why we have values outside the limits. The sample that was used for the creation of the diagrams is a 200-value sample with mean 0 and standard deviation 1. From this sample 20 values were randomly selected and replaced with values with mean 4 and standard deviation 1. In the second diagram we do not have values outside the limits and we note that the values of the limits and the size of the sample are different. That is because all values outside the limits were eliminated from our sample. The algorithm to eliminate outlier values is executing this process.



Figure 3: I-Chart with L = 3 and without outlier filtering





Figure 4:I-Chart with L = 3 and after outlier filtering

5.4 Phase II: AARLO and SDARLO computation with Monte Carlo simulation The values of the parameters λ and L_{ewma} were chosen initially so the ARLO equals 370 and using this as a base we are going to calculate the results of the estimators. The estimation process is being studied after the outlier filtering in our sample but without yet implemented. The difference with other studies in the current study is that the choices we have made in phase I is affecting the efficiency of phase 2 of EWMA due to its design. The results where this is considered is AARLO and SDARLO.

The algorithm in each loop executed it generates a new sample and new contaminated data, and every time calculating the estimators before and after the outlier filtering and afterwards calculating the ARL. Using the Knoth library (spc, 2004) and with the command xewma.arl we are calculating the arl. The values of our sample are 50, 100, 200, 500, 1000, 10000 but, to be clear, working with sample sizes over 500 is rare in practice and over 1000 extremely unlikely. Nevertheless, the samples will be analyzed so we can study the effect on the estimators. The values of L for estimating the limits in phase I as we have said before are from 2 to 4 with step 0.1. The following values of L we are going to try are 2.7022, 2.8595, 2.959 and 3. Furthermore the λ , which is

UNIV

necessary with L_{ewma} to calculate the ARL, gets values 0.05, 0.1, 0.2, 0.25, 0.4 and 1. The sum of those combinations together with the Contamination Ratio (CR) are giving us 12096 combinations. Every combination with the method of Monte Carlo was simulated 250 times so we can have the best and most realistic results. After every loop the algorithm saves the ARL value and then we calculate the AARL0 which is the mean of ARL values. In the same fashion we calculate the SDARL0.

In the following tables we showcase the 3 best estimators for the corresponding combination. The way we selected the best estimators is by calculating the median of ARL0 we can check which estimator's values most closely approximate 370. With this process the results were compiled in the following tables. The data has been subjected to outlier filtering in phase I.

	CR	0%		2%		4%		10%	
Μ	λ	Estimators	L	Estimators	L	Estimators	L	Estimator	L
50	0.1	RMSSD	2	IQR	2	SSD	2.3	MAD	2.1
		MnMR	2.8	SSD	2.2	MAD	2.1	IQR	2.4
		SSD	2.7	MnMR	2.1	IQR	2.2	SSD	2.1
	0.2	MnMR	2.9	MdMR	2.1	IQR	3.4	IQR	3.1
		SSD	4	IQR	2.5	MAD	2.3	MAD	2.1
		MAD	3.5	RMSSD	2.5	SSD	2.3	RMSSD	2
	0.4	SSD	2.8	IQR	2.2	IQR	2	SSD	2
		IQR	2.9	SSD	3	SSD	2.9	IQR	3.9
		MdMR	3.9	RMSSD	3	RMSSD	2.3	MAD	2.3
	1	SSD	4	MdMR	2	MnMR	2	RMSSD	2
		RMSSD	2.3	IQR	2.1	IQR	2.7	SSD	2.1
		IQR	2.3	MnMR	3	RMSSD	2.7	MnMR	2

Table 2: Results from simulation of algorithm for sample 50



	CR	0%		2%		4%		10%	
m	λ	Estimators	L	Estimators	L	Estimators	L	Estimators	L
100	0.1	SSD	2.3	MAD	2.8	IQR	3.1	SSD	2.3
		MdMR	2.7	RMSSD	2.2	SSD	2.5	RMSSD	2.2
		IQR	3.6	IQR	3.8	RMSSD	2.2	IQR	2.4
	0.2	RMSSD	3	MnMR	3.6	IQR	2.8	SSD	2.3
		MdMR	3.7	MAD	3.8	MnMR	2.2	MnMR	2.4
		SSD	3.5	SSD	2.4	SSD	2.4	IQR	3.6
	0.4	MAD	2.8	SSD	3.2	MnMR	3	MnMR	2.3
		MnMR	3.9	MdMR	3.2	RMSSD	3	SSD	2.5
		RMSSD	2.5	IQR	2.6	SD	2.8	MdMR	2.2
	1	IQR	2.5	IQR	2.2	MAD	2.2	RMSSD	2.2
		MnMR	2.3	MnMR	3.8	IQR	3.2	IQR	2
		SSD	2.8	MAD	2.4	SSD	3.2	SSD	2.5

Table 3: Results from simulation of algorithm for sample 100

Table 4: Results from simulation of algorithm for sample 200

	CR	0%		2%		4%		10%	
m	λ	Estimators	L	Estimators	L	Estimators	L	Estimators	L
200	0.1	MnMR	2.3	MAD	3	RMSSD	2.3	MnMR	3.9
		RMSSD	2.5	IQR	3.5	MAD	3.4	SSD	2.7
		MAD	2.8	MnMR	3.2	IQR	2.3	RMSSD	2.4
	0.2	MAD	3	MdMR	2.2	RMSSD	3.2	MAD	2.1
		RMSSD	2.4	RMSSD	3	IQR	2.6	IQR	2.4
		MnMR	3.2	MnMR	3	MdMR	3.6	MnMR	2.4
	0.4	MAD	2.5	MAD	4	MAD	3.3	IQR	2.6
		IQR	2.6	MdMR	2.7	MdMR	3.7	MAD	2.4
		RMSSD	2.8	RMSSD	2.8	SSD	2.7	RMSSD	2.6
					38				UNIVERS

1	MdMR	3.7	MdMR	3	MdMR	2.4	RMSSD	2.5
	RMSSD	3.1	IQR	3.9	IQR	3.7	MAD	2.7
	MnMR	3.2	MAD	3.2	MAD	3.7	IQR	4

Table 5: Results from simulation of algorithm for sample 500

	CR	0%		2%		4%		10%	
m	λ	Estimators	L	Estimators	L	Estimators	L	Estimators	L
500	0.1	MdMR	3.4	MnMR	2.5	IQR	2.9	RMSSD	2.3
		MAD	3.7	IQR	2.1	MAD	3.1	IQR	2.1
		RMSSD	2.4	MAD	2.9	MdMR	2.8	MdMR	2.7
	0.2	IQR	2.9	IQR	2.9	MnMR	2.9	MAD	2.3
		MnMR	3.7	MAD	2.9	RMSSD	3.4	IQR	3.1
		MdMR	2.9	MdMR	2.9	MAD	2.7	RMSSD	2.9
	0.4	MdMR	2.9	MAD	2.7	IQR	2.6	MnMR	2.6
		MnMR	3.5	MdMR	3.8	MdMR	2.6	MAD	3.3
		RMSSD	2.9	MnMR	2.9	RMSSD	2.7	MdMR	2.6
	1	MnMR	3	MnMR	2.9	IQR	2.7	IQR	2.5
		IQR	3.9	IQR	3.2	RMSSD	2.8	MAD	2.5
		MAD	3.4	MdMR	2.9	MAD	2.8	MdMR	2.5



	CR	0%		2%		4%		10%	
m	λ	Estimators	L	Estimators	L	Estimators	L	Estimators	L
1000	0.1	RMSSD	3.7	IQR	3.2	MAD	3.1	MAD	2.6
		MAD	3.9	RMSSD	3.2	MnMR	2.3	MdMR	2.6
		MnMR	3.7	MAD	3.2	IQR	3.2	IQR	2.7
	0.2	MnMR	4	MAD	2.8	MdMR	2.8	MnMR	2.7
		RMSSD	3.7	MnMR	2.5	MnMR	2.5	IQR	2.6
		MAD	3.5	IQR	2.8	MAD	2.7	MAD	2.6
	0.4	MnMR	3.4	RMSSD	2.8	MnMR	2.7	IQR	3.4
		RMSSD	3	MdMR	2.8	MdMR	2.7	MdMR	2.5
		IQR	3.2	IQR	2.9	RMSSD	3.6	MAD	2.4
	1	MnMR	3.5	MnMR	2.9	IQR	2.7	MAD	3
		MdMR	3	IQR	3.4	MnMR	3.3	IQR	3
		RMSSD	3.4	MAD	2.8	MAD	3	RMSSD	2.7

Table 6:Results from simulation of algorithm for sample 1000

Table 7: Results from simulation of algorithm for sample 10000

	CR	0%		2%		4%		10%	
m	λ	Estimators	L	Estimators	L	Estimators	L	Estimators	L
10000	0.1	IQR	3.9	IQR	3	MdMR	2.4	IQR	2.7
		MAD	3.6	MAD	3	MAD	2.9	MAD	2.7
		MnMR	3.9	RMSSD	3.1	IQR	2.9	MdMR	2.7
	0.2	RMSSD	3.8	IQR	2.3	MnMR	2.6	MdMR	2.7
		MAD	3.7	MAD	2.3	IQR	2.8	IQR	3.1
		IQR	3.7	RMSSD	4	MAD	2.8	MAD	3.1
	0.4	IQR	3.8	IQR	2.6	MnMR	2.9	RMSSD	2.9
		MnMR	3.8	MAD	2.6	RMSSD	2.8	MAD	2.5

40

UNIVE

C

	Mad	3.7	MnMR	2.8	MAD	2.8	IQR	2.5
1	RMSSD	4	MdMR	3	MAD	3.3	MdMR	3
	MnMR	3.8	MAD	4	IQR	3.3	RMSSD	2.8
	MAD	3.3	IQR	4	MnMR	2.9	MAD	2.8

The above tables are divided by sample size. The sample sizes are 50, 100, 200, 500, 1000, 10000. Furthermore, the estimator SSD was used for sample size equal to or less than 200 because in large sample sizes the 2nd part of its equation is approximating the infinite $(\Gamma(m/2)/\Gamma((m-1)/2))$. Studies have shown that SSD can be utilized for larger sample sizes by approximating it, but we decided for our studies to see its capability in small - realistic sample sizes.

For those sample sizes we note very good results for the SSD estimator, who is very effective for data with no outliers (as in our case by utilizing outlier filtering) and data that come from normal distribution. For sample sizes above 200 we are noting the estimators IQR and MAD are more and more present as the sample size increases. The IQR is more effective as sample size increases as we have said before. However, there is no clear superiority of an estimator as the rest of the estimators are present as well. IQR and MAD's presence is a little more frequent with the rest of the estimators being reasonably close.

In addition, we are noting that the value of L is fluctuating between 2 to 3 and in some rare cases gets values over 3. An important note is that as the contamination ratio increases the value of L is getting values from 2.6 to 3.

Considering the previous points, we conclude that for data with no outliers the SSD estimator is reliable for small sample sizes. As our sample size increases, appropriate results can be gained by using the IQR and MAD estimators, a result that is in odds with the theories and studies by researchers.



The following tables are without outlier filtering:

	CR	0%		2%		4%		10%	
m	λ	Estimators	L	Estimators	L	Estimators	L	Estimators	L
50	0.1	MnMR	2	MAD	2.8	MAD	3.8	MAD	2.5
		IQR	2.8	IQR	3.5	IQR	3.4	IQR	2
		RMSSD	3.7	-	-	-	-	-	-
	0.2	SSD	2.4	MAD	3.4	IQR	3.7	IQR	3.3
		MnMR	2.4	IQR	3.1	MAD	38 0	MAD	3.8
		MAD	3.5	-	-	-	-	-	-
	0.4	IQR	2.9	MAD	2	IQR	2.7	IQR	3.5
		MnMR	3.9	IQR	2.9	MAD	2.5	MAD	2.1
		MdMR	2.9	-	-	-	-	-	-
	1	SSD	3.4	IQR	3.5	IQR	3.4	-	-
		IQR	2.8	MAD	3.5	MAD	3.9	-	-
		MAD	2.4	-	-	-	-	-	-

Table 8: Results from simulation of algorithm for sample 50

Table 9: Results from simulation of algorithm for sample 100

	CR	0%		2%		4%		10%	
m	λ	Estimators	L	Estimators	L	Estimators	L	Estimators	L
100	0.1	MnMR	3.2	IQR	3.8	IQR	2.3	MdMR	2.5
		RMSSD	2.8	MAD	2.7	MAD	3.2	-	-
		IQR	2.6	-	-	-	-	-	-
	0.2	SSD	3.5	MAD	4	MAD	2.4	IQR	3.7
		MAD	2.5	IQR	3	IQR	3.3	MAD	2.8
		MdMR	3.7	-	-	-	-	-	-
	0.4	MnMR	2.6	MdMR	3.1	MAD	2	MAD	2

UNIV

SSD	2.1	IQR	3.6	IQR	2	IQR	2.9
MAD	3.1	MAD	3.6	-	-	-	-
IQR	3.1	MAD	3.2	IQR	3	IQR	3.8
MnMR	3.3	MdMR	2.1	MAD	2.2	MAD	3.8
MAD	3.9	IQR	3.2	-	-	-	-
	SSD MAD IQR MnMR MAD	SSD 2.1 MAD 3.1 IQR 3.1 MnMR 3.3 MAD 3.9	SSD2.1IQRMAD3.1MADIQR3.1MADMnMR3.3MdMRMAD3.9IQR	SSD 2.1 IQR 3.6 MAD 3.1 MAD 3.6 IQR 3.1 MAD 3.2 MnMR 3.3 MdMR 2.1 MAD 3.9 IQR 3.2	SSD 2.1 IQR 3.6 IQR MAD 3.1 MAD 3.6 - IQR 3.1 MAD 3.2 IQR MNMR 3.3 MdMR 2.1 MAD MAD 3.9 IQR 3.2 -	SSD 2.1 IQR 3.6 IQR 2 MAD 3.1 MAD 3.6 - - IQR 3.1 MAD 3.2 IQR 3 MNMR 3.3 MdMR 2.1 MAD 2.2 MAD 3.9 IQR 3.2 - -	SSD 2.1 IQR 3.6 IQR 2 IQR MAD 3.1 MAD 3.6 - - - IQR 3.1 MAD 3.2 IQR 3 IQR MNMR 3.3 MdMR 2.1 MAD 2.2 MAD MAD 3.9 IQR 3.2 - - -

Table 10: Results from simulation of algorithm for sample 200

	CR	0%		2%		4%		10%	
m	λ	Estimators	L	Estimators	L	Estimators	L	Estimators	L
200	0.1	RMSSD	2.2	IQR	3	IQR	3.5	MnMR	3.9
		IQR	2.8	MAD	2.8	MAD	2.3	MdMR	2.8
		MdMR	3.1	-	-	-	-	-	-
	0.2	MnMR	3	MdMR	2.8	MdMR	3.3	MdMR	2.8
		MAD	2.8	IQR	2.2	IQR	2.1	-	-
		RMSSD	3	MAD	2.2	MAD	4	-	-
	0.4	IQR	3.3	IQR	3.4	IQR	2.9	MAD	2.3
		MdMR	2.7	MAD	2.9	MAD	2.4	IQR	4
		MAD	3.3	MnMR	2.9	MdMR	2.3	-	-
	1	MdMR	3.8	MAD	4	MdMR	3.9	MAD	3.7
		MnMR	2.3	MnMR	2.4	IQR	2.7	IQR	4
		RMSSD	2.3	IQR	2.2	-	-	-	-

Table 11: Results from simulation of algorithm for sample 500

CR	0%	2%	4%	10%	I OF ECO
		43	3		A DATION A BUSINESS &

m	λ	Estimators	L	Estimators	L	Estimators	L	Estimators	L
500	0.1	MnMR	2.1	MdMR	3.4	IQR	3.3	MnMR	2.4
		IQR	2	MAD	2.9	MAD	3.3	-	-
		MAD	3.7	IQR	3.2	MdMR	2.9	-	-
	0.2	MnMR	2.1	IQR	3.9	IQR	3.1	MdMR	3.7
		RMSSD	3.7	MAD	2.1	MAD	3.5	-	-
		IQR	2.8	MdMR	2	-	-	-	-
	0.4	RMSSD	3.4	MnMR	3.2	IQR	2.3	IQR	3.7
		MnMR	3	MAD	2.2	MAD	2.3	MAD	2.9
		IQR	3.7	IQR	3.4	-	-	MdMR	2.5
	1	MdMR	2.8	MdMR	3.6	IQR	2.4	IQR	2.4
		IQR	3.9	MnMR	2.6	MAD	2.1	MAD	2.4
		MnMR	3.6	MAD	2.4	MdMR	3.4	-	-

Table 12:Results from simulation of algorithm for sample 1000

	CR	0%		2%		4%		10%	
m	λ	Estimators	L	Estimators	L	Estimators	L	Estimators	L
1000	0.1	IQR	2.4	MnMR	2.6	MAD	3	-	-
		-	-	MAD	4	IQR	3	-	-
		-	-	IQR	4	-	-	-	-
	0.2	RMSSD	2.3	MAD	2.1	MAD	2.3	MdMR	2.2
		IQR	2	IQR	2.1	IQR	2.3	-	-
		MnMR	3.4	-	-	-	-	-	-
	0.4	MnMR	2.2	MnMR	3.6	IQR	3.4	IQR	2.3
		RMSSD	3	MdMR	2.5	MAD	3.4	MAD	2.3

	IQR	3.8	MAD	3.7	-	-	-	-
1	MnMR	2.1	MdMR	3	IQR	3.7	-	-
	IQR	3	MnMR	3.8	MAD	3.9	-	-
	MAD	3	-	-	-	-	-	-

Table 13: Results from simulation of algorithm for sample 10000

	CR	0%		2%		4%		10%	
Μ	λ	Estimators	L	Estimators	L	Estimators	L	Estimators	L
10000	0.1	MnMR	3.6	MdMR	3.1	MAD	2	-	-
		RMSSD	3.9	MAD	3.8	IQR	2	-	-
		MdMR	2	IQR	3.8	-	-	-	-
	0.2	MnMR	2.6	MAD	2.6	MAD	3.6	-	-
		MdMR	2.4	IQR	3.9	IQR	3	-	-
		RMSSD	3.1	-	-	-	-	-	-
	0.4	MnMR	3	MdMR	3.9	-	-	-	-
		MdMR	3.5	MnMR	3.3	-	-	-	-
		RMSSD	3.3	-	-	-	-	-	-
	1	RMSSD	3.7	MdMR	3.9	IQR	3.9	-	-
		MdMR	3.4	-	-	MAD	3.9	-	-
		MnMR	2.1	-	-	-	-	-	-

As before, the tables are divided by sample size so we can show the results. In contrast to our earlier process where our data has been subjected to outlier filtering, here we generate the sample from a normal distribution with median 0 and standard deviation 1 and we want to see the performance of the estimators with no outlier filtering. We have also in this case corresponding contamination ratios (0%, 2%, 4%, 10%). Contrasting the earlier findings, in this case we have a clearer idea about which estimators are performing better and most reliably. The estimators MAD and IQR are those who

UNIVER

outclass the others in this case. We are noting that for the contamination ratio of 0% almost all estimators are giving reliable results with the median ARL being very frequently equal to the target value 370. As the contamination ratio increases then the estimators MAD and IQR are giving the most reliable results with their median ARL being very close to 370. In addition, the L value doesn't seem to follow some pattern (having specified reliable values) and that is because our sample hasn't been subjected to outlier filtering.

5.5 Conclusion

The fifth and final chapter of this dissertation was dedicated to the process and the algorithm we developed and I personally tried to verify results and earlier studies of statistics, and also test some processes such as in phase I where we used the estimator RMSSD to find the limits (UCL, LCL). The tests were inspired by Murat Caner Testik, Ozge Kara, Sven Knoth (2020). From the results we calculated we can reliably come to some conclusions and to build on them. If we have a sample not subjected to outlier filtering then the most reliable results can be gained by using the MAD and IQR estimators. These two estimators have a lot of common systems in them and as such in some cases they were the only reliable estimators. When we subject our sample to outlier filtering and our sample is equal to or smaller than 200 observations, reliable results can be gained by using the SSD estimator, and if our sample is larger than 200, as in some samples in our study with sample sizes of 500, 1000 and 10000 observations, we are noting that there is no clear winner between the estimators as they are all relatively reliable with a slighter frequency of MAD and IQR.



Conclusion and Suggestions

In this study, that was created for the purposes of this dissertation, the results of analyzing phase I for the effectiveness of the Exponentially Weighted Moving Average (EWMA) that was used in phase II, were evaluated contrasted with some alternative scenarios. The Control Diagram I-Chart is the one chosen for analyzing phase I and several combination of values for the Contamination Ratio were chosen, for L as well as for the parameter λ of EWMA.

Following this process, the results were analyzed, drawn and went through some suggestions and notes. The focus was in choosing the estimator for the standard deviation, given that the I-Chart is affecting the limits of phase II. The effectiveness of Phase II, where we tested several scenarios calculating AARL and SDARL whose results were given in tables. As first effect of this study was the discovery that the elimination of outliers in the way mentioned above during phase 1 results in a reliable parameter estimation. Given that those estimations affect the online monitoring of the process during phase II, we need to look closer into the process for phase I contrasting literature around the subject. The estimators of standard deviation were different depending on the scenario, where the size of filtered observations are lower than 200 values we recommend using the SSD estimator and choose lower values for L. Strong showing for data sample size over 200 observations by all estimators with CR taking values 0%, 2%, 4%, 10%. Most of all, the estimators MAD and IQR are a little better performing than the rest, especially for CR 10% and/or increased sample size. Furthermore, wit was noted that for a small sample size the results aren't very reliable. It may be suggested for future researchers to seek estimators that are more effective for smaller sample size processes.



Appendix

	2.7022	2.8595	2.959	3
2	64,4414	82,69109	94,80434	100,5087
2,1	77,8762	113,542	160,4204	145,275
2,2	107,5139	149,8493	212,0034	213,2947
2,3	133,7106	206,1682	259,5514	234,5969
2,4	143,5617	223,0261	296,2152	371,6208
2,5	168,9281	255,2299	354,2886	362,8662
2,6	196,1345	319,1291	394,0878	391,5676
2,7	208,9984	297,8669	457,3353	483,2783
2,8	219,6068	340,9977	479,9774	538,6473
2,9	236,1027	369,1831	446,6991	554,6396
3	268,5752	350,114	588,3317	637,3544
3,1	271,488	466,1268	515,5752	699,3616
3,2	266,9821	437,1156	600,9969	692,3458
3,3	273,7118	526,8844	756,5237	764,7354
3,4	350,2084	466,8577	698,2866	901,6194
3,5	332,655	570,9091	739,9037	806,9856
3,6	351,1526	558,4322	893,2291	1110,351
3,7	361,534	668,9412	894,816	1076,25
3,8	409,0095	681,6797	1029,475	1156,373
3,9	467,6342	692,0224	1229,672	1617,923
4	442,9343	801,741	1343,928	1281,801

Table A.1: Values of AARL for SSD estimator

with m = 100, CR = 0.02 and λ = 0.2 after outlier filtering

Table A.2

Table A.2: Values of AARL for RMSSD estimator

	-			
	2.7022	2.8595	2.959	3
2	55,63545	67,09377	92,62638	110,3925
2,1	79,7585	107,6421	149,0456	145,0311
2,2	120,3945	140,2147	189,3909	221,6035
2,3	128,1836	195,5056	264,364	263,7521
2,4	149,07	240,4398	305,6734	342,778
2,5	164,7339	258,4784	361,4471	379,7866
2,6	193,4699	300,0601	377,787	431,5412
2,7	218,2075	304,8358	425,1919	470,4791
2,8	219,3297	349,2558	455,4121	494,1727



2,9	236,8006	375,0603	536,3558	587,5396
3	257,5362	373,3104	532 <i>,</i> 8678	636,5759
3,1	253,5841	406,4257	524,6986	697,4279
3,2	298,6886	426,6664	563,3193	644,0837
3,3	309,5657	480,3632	670,1184	739,3127
3,4	294,411	494,3414	698,2193	777,3717
3,5	305,5984	481,2029	694,0679	844,4572
3,6	382,476	563,5336	769,7364	975,0265
3,7	392,4586	653,0534	897,7594	1054,22
3,8	434,7229	616,791	1097,764	1234,806
3,9	427,8941	700,2575	1041,332	1257,51
4	450,0799	742,7255	1053,183	1371,682

Table A.3: Values of AARL for MdMR estimator

	2.7022	2.8595	2.959	3
2	77,47429	101,4939	146,0278	185,363
2,1	108,2786	182,7237	247,4714	251,5839
2,2	180,5767	191,0642	300,1332	370,2643
2,3	176,3963	321,4765	447,7372	441,0348
2,4	201,0884	351,0209	453,0259	520,4756
2,5	228,5486	406,3518	543,1087	553,6429
2,6	252,6393	402,2067	503,2559	612,3968
2,7	283,1194	423,3074	558,3605	622,442
2,8	271,3361	467,003	655,388	608,8334
2,9	289,262	475,7245	785,2493	890,3547
3	338,8971	431,8876	718,9326	818,3
3,1	311,5524	483,204	644,1413	804,3403
3,2	369,6586	488,2373	614,0596	781,4583
3,3	343,5024	527,925	730,5997	731,4972
3,4	324,3296	534,467	726,0553	902,0971
3,5	343,5654	569,0107	733,5383	988,5313
3,6	368,107	518,7889	799,6638	995 <i>,</i> 673
3,7	352,4825	638,842	949,4525	1003,944
3,8	370,9662	528,5721	868,3432	1025,437
3,9	338,9365	525,3007	778,5596	1023,442
4	327,268	537,108	785,695	1055,105



Table A.4: Values of AARL for MnMR estimator

	2.7022	2.8595	2.959	3
2	61,14807	75,33221	105,4094	127,8757
2,1	87,39931	124,8409	174,6676	166,587
2,2	135,047	154,6435	213,0772	257,4664
2,3	140,0106	217,8257	303,2224	295,0222
2,4	157,6933	262,0169	337,7965	380,1487
2,5	177,3131	282,9735	391,3871	413,0466
2,6	205,7116	322,5945	404,8313	453,379
2,7	228,9193	325,9845	451,8336	502,2961
2,8	228,6669	367,4988	488,7629	514,1365
2,9	242,8217	387,5167	565,9031	624,7133
3	263,0795	373,6194	544,1359	638,7696
3,1	255,738	403,6939	542,3952	690,6341
3,2	299,6271	423,6356	553,2843	628,0661
3,3	305,3226	469,6093	634,9981	694,8931
3,4	285,7388	470,2821	676,5834	750,1394
3,5	299,8191	473,1873	641,912	807,4091
3,6	351,0991	503,2305	701,8091	880,5663
3,7	352,7022	577,1078	800,9403	921,8311
3,8	376,2546	540,2974	928,047	1025,039
3,9	355,9628	591,3252	850,0213	1033,851
4	369,194	597,3487	827,9754	1090,01

with m = 200, CR = 0.02 and λ = 0.2 after outlier filtering

Table A.5: Values of AARL for MAD estimator

	2.7022	2.8595	2.959	3
2	101,4102	138,9107	165,6301	236,8351
2,1	135,3065	178,9099	294,8492	313,8931
2,2	208,0223	229,554	322,75	414,6353
2,3	183,1974	296,8947	455,9946	470,9928
2,4	193,1934	341,2763	444,3377	565,8438
2,5	221,3895	348,4558	498,4325	529,4629
2,6	235,9662	348,1018	508,4163	662,1624
2,7	253,6313	403,6003	515,4928	595,873
2,8	256,8641	391,6879	577,4229	631,2082
2,9	262,2372	414,9128	565,3909	613,2586



3	262,4783	406,6499	562,5159	647,2063
3,1	259,8917	410,4903	603,3392	732,7225
3,2	263,3863	411,9237	552,6614	645,0312
3,3	269,324	433,2265	636,433	604,7797
3,4	265,854	483,2427	585,966	620,7615
3,5	267,8977	439,5631	572,117	679,2715
3,6	309,0117	447,473	537,882	715,4995
3,7	292,8955	452,2593	605,2504	763,8661
3,8	285,9015	435,7561	586,2229	743,6202
3,9	268,366	472,0634	642,0773	690,996
4	268,9788	482,8333	546,3301	724,2017

Table A.6: Values of AARL for IQR estimator

	2.7022	2.8595	2.959	3
2	96,79772	137,7004	169,1506	231,6901
2,1	132,4041	175,7735	287,6423	306,3879
2,2	197,947	222,2506	316,0154	397,4176
2,3	177,8525	284,0597	411,3889	455,0045
2,4	194,1893	325,2988	418,1416	560,4967
2,5	214,3072	351,6993	492,2204	515,3494
2,6	233,4601	346,2991	487,9319	623,1133
2,7	251,5967	398,2545	503,5077	565,36
2,8	255,7086	387,9511	569,466	601,6993
2,9	256,9415	389,0406	567,6472	584,5296
3	264,2619	401,1544	554,2471	636,981
3,1	251,2651	400,9109	585,1821	699,7736
3,2	253,7854	404,6752	530,5123	611,7573
3,3	264,5258	439,9114	597,4378	596,3712
3,4	262,2481	496,3392	557,6219	615,6942
3,5	269,3542	417,6019	552,8538	664,8631
3,6	302,318	422,1748	520,038	693,996
3,7	279,8087	437,0774	576,6372	720,6239
3,8	279,299	418,6323	576,8025	730,9605
3,9	257,2797	475,5363	611,1878	661,5731
4	284,7876	454,9979	542,3828	695,3276



Table A.7: Values of AARL for MnMR estimator

	2.7022	2.8595	2.959	3
2	462,9074	768,4247	1155,823	1374,837
2,1	457,6095	794,9669	1188,182	1242,379
2,2	507,0485	713,6941	1174,071	1424,967
2,3	464,2426	831,5748	1232,552	1267,255
2,4	458,395	859,2018	1142,117	1353,111
2,5	449,851	793,0606	1130,236	1272,852
2,6	480,1322	759,8825	1130,372	1334,562
2,7	500,0605	781,2367	1149,233	1374,791
2,8	476,2855	817,7257	1134,735	1320,821
2,9	446,5381	785,8396	1244,604	1502,369
3	469,183	759,8885	1150,004	1372,018
3,1	470,998	785,2871	1118,525	1398,887
3,2	508,314	802,3702	1115,386	1223,15
3,3	497,4463	814,3873	1143,257	1213,66
3,4	455,5257	758,4165	1108,049	1302,869
3,5	468,8638	790,1383	1069,642	1358,343
3,6	495,2728	763,1767	1083,274	1425,514
3,7	474,9455	837,884	1225,368	1464,933
3,8	493,1033	807,353	1308,469	1448,875
3,9	458,6102	808,2031	1183,357	1396,153
4	452,9462	754,3327	1073,201	1409,408

with m = 200, CR = 0.02 and λ = 0.2 without outlier filtering

Table A.8: Values of AARL for MdMR estimator

	2.7022	2.8595	2.959	3
2	350,1403	579,9065	912,183	1218,82
2,1	348,5209	568,8651	853,8459	1022,027
2,2	434,3245	487,8046	891,678	1147,376
2,3	347,6505	696,2851	994,282	998,742
2,4	359,8159	639,1904	902,5965	1052,868
2,5	347,0617	781,2475	950,1066	982,734
2,6	381,7465	584,3745	834,2889	1085,533
2,7	386,5748	653,1884	842,0457	921,4107
2,8	370,963	651,6406	944,0195	882,3357
2,9	359,5424	637,2667	1108,66	1213,478



3	420,3334	553,5777	933,8194	1088,449
3,1	382,6523	622,7461	812,3654	993,0971
3,2	448,3859	585,9118	815,926	1010,739
3,3	408,5126	618,652	836,5497	857,5182
3,4	362,2069	601,8749	817,4811	1109,721
3,5	386,8516	624,2037	849,3308	1104,448
3,6	409,9998	570,6701	881,9873	1150,728
3,7	372,4324	686,2251	1020,31	1105,736
3,8	401,7704	564,9371	930,9833	1115,091
3,9	358,7114	563,8554	833,6468	1089,948
4	341,8426	561,9088	814,711	1095,35

Table A.9: Values of AARL for RMSSD estimator

	2.7022	2.8595	2.959	3
2	682,9085	1202,92	1830,131	2230,659
2,1	681,6549	1281,812	1881,845	1916,874
2,2	713,0003	1155,409	1928,938	2238,883
2,3	698,9653	1295,839	1919,376	1974,089
2,4	647,7848	1367,287	1746,889	2162,4
2,5	671,0211	1202,957	1771,199	2122,904
2,6	736,3825	1193,85	1791,782	2192,407
2,7	731,0147	1200,315	1887,727	2594,953
2,8	694,3658	1260,654	1744,386	2267,364
2,9	648,381	1188,378	1922,501	2393,272
3	673,5307	1213,176	1865,7	2266,222
3,1	712,2378	1218,794	1768,059	2294,773
3,2	728,6472	1294,716	1809,111	2009,784
3,3	710,0547	1247,799	1838,701	2011,565
3,4	669,9261	1139,913	1702,258	2074,826
3,5	665,6277	1282,269	1734,082	2221,75
3,6	692,8392	1195,176	1678,262	2326,631
3,7	694,3665	1309,63	1984,401	2506,664
3,8	710,4448	1304,536	2125,677	2328,934
3,9	683,1394	1249,067	1951,55	2241,186
4	660,7424	1175,389	1737,163	2355,782



Table A. 10: Values of AARL for MAD estimator

	2.7022	2.8595	2.959	3
2	263,3499	453,6943	526,2659	706,7897
2,1	269,1914	380,5234	680,203	762,6484
2,2	307,6862	374,0527	638,1479	768,1378
2,3	262,0026	462,9894	706,9723	754,072
2,4	263,31	437,0569	618,8128	789,1231
2,5	258,0061	414,0074	614,3673	685,9953
2,6	273,7008	412,4052	606,4466	757,167
2,7	285,6733	441,6887	577,5257	683,2393
2,8	278,1037	425,8711	669,1197	726,6338
2,9	281,7341	448,8126	620,3072	679,7978
3	283,1887	435,2411	616,464	672,7621
3,1	265,6389	436,2866	603,3692	752,3436
3,2	277,4291	440,2015	582,7199	646,9058
3,3	280,498	438,822	631,1579	602,8408
3,4	279,1347	487,5586	589,0972	687,9037
3,5	273,0984	436,4078	567,5969	689,4223
3,6	310,6536	440,4684	550,5284	724,6499
3,7	293,2875	459,4443	622,3219	789,9442
3,8	286,7588	436,2851	598,7264	780,7454
3,9	265,2279	459,7591	636,3231	707,0271
4	269,7831	477,1412	544,9138	712,7648

with m = 200, CR = 0.02 and λ = 0.2 without outlier filtering

Table A.11:

Table A.11: Values of AARL for IQR estimator

	2.7022	2.8595	2.959	3
2	256,1001	419,0673	524,2203	677,539
2,1	259,7941	386,1282	660,6804	725,9259
2,2	300,3262	370,7903	614,9291	763,2443
2,3	252,556	443,7871	654,4541	720,8398
2,4	261,5922	424,8705	612,8835	771,4252
2,5	252,1914	410,7197	606,7861	665,5249
2,6	265,8453	401,7466	585,1441	749,4729
2,7	280,5536	435,5888	562,3952	663,2144
2,8	274,8569	433,5093	643,178	685,0049
2,9	272,8603	430,3942	622,5912	648,4708



3	281,2313	439,0797	600,4702	667,1229
3,1	257,0122	426,7629	590,6863	758,5202
3,2	270,8283	430,3906	561,691	599,0682
3,3	273,8827	430,7617	613,6583	589,8942
3,4	266,9264	507,6348	551,7388	663,3478
3,5	270,4531	418,3603	549,4086	673,165
3,6	300,8035	418,3941	548,125	697,1193
3,7	276,7578	438,9137	610,7347	750,6765
3,8	282,0427	414,8911	571,6746	764,2863
3,9	252,6918	473,6422	603,9607	682,619
4	280,1364	445,2562	543,5402	685,1567

Table A. 12: Values of AARL for SSD estimator

	2.7022	2.8595	2.959	3
2	681,0816	1280,162	1789,241	2534,617
2,1	614,1191	1269,378	1781,634	1969,04
2,2	686,09	1281,262	2038,113	2215,189
2,3	789,5704	1340,262	2033,215	2271,016
2,4	670,8004	1334,913	1983,026	2578,593
2,5	778,5133	1262,656	1954,791	2290,509
2,6	657,1008	1349,949	1915,852	2372,227
2,7	739,574	1179,234	2019,519	2331,203
2,8	663,8199	1177,768	1813,132	2420,364
2,9	642,9511	1161,85	1765,776	2111,142
3	717,8113	1339,242	1804,327	2430,445
3,1	690,5184	1493,653	1732,196	2240,293
3,2	656,3558	1217,275	2163,102	2164,436
3,3	663,5395	1280,112	2011,155	2261,525
3,4	655,9843	1210,746	1617,55	2605,518
3,5	695,3678	1202,53	1691,793	2345,723
3,6	715,6431	1244,655	1897,973	2382,786
3,7	606,7026	1124,298	1763,187	2231,349
3,8	740,0087	1312,279	2104,017	2296,482
3,9	684,4374	1162,486	1925,722	2595,885
4	662,1312	1369,693	1876,035	2409,933

with m = 100, CR = 0.02 and λ = 0.2 without outlier filtering



Bibliography

- [1] Abbas, N., Riaz, M. and Does, R., 2012. CS-EWMA Chart for Monitoring Process Dispersion. Quality and Reliability Engineering International, 29(5), pp.653-663.
- [2] Abbas, N., Riaz, M. and Does, RJMM. (2013a). CS-EWMA Chart forMonitoringProcess Dispersion. Quality and Reliability Engineering international, Vol. 29(5), pp. 653-663.
- [3] Albers, W. and Kallenberg, W.C.M. (2004a). Are Estimated Control Charts in Control? Statistics 38, pp.67-79.
- [4] Ali, S., Pievatolo, A. and Göb, R., 2016. An Overview of Control Charts for High-quality Processes. Quality and Reliability Engineering International, 32(7), pp.2171-2189.
- [5] Bradford, P. and Miranti, P., 2019. Information in an Industrial Culture: Walter A. Shewhart and the Evolution of the Control Chart, 1917–1954. Information & Culture, 54(2), pp.179-219.
- [6] Braun, W. J., & Park, D. (2008). Estimation of σ for individuals charts. Journal of Quality Technology, 40, 332–344.
- [7] Brondoni, S. and Zaninotto, E., 2018. Ouverture de 'The 4th Industrial Revolution. Business Model Innovation & Global Competition'. Symphonya. Emerging Issues in Management, (2), p.1.
- [8] Bryce, R. G., Gaudard, M. A., & Joiner, B. L. (1997). Estimating the standard deviation for individuals control charts. Quality Engineering, 10, 331–341.
- [9] Burroughs, T. E., Rigdon, S.E., Champ, C.W. (1993). An analysis of Shewhart Charts with Runs Rules When No Standards Are Given, Proceedings of the Quality and Productivity Section of the American Statistical Association, August 8-12, San Francisco, CA, pp. 16-19.
- [10] C.H. Weiß and M.C. Testik, CUSUM monitoring of first-order integer-valued autoregressive processes of Poisson counts, J. Qual. Technol. 41(4) (2009), pp. 389–400.
- [11] C.H. Weiß, Controlling correlated processes of Poisson counts, Qual. Reliab. Eng. Int. 23(6) (2007), pp. 741–754.
- [12] C.H. Weiß, EWMA monitoring of correlated processes of Poisson counts, Qual. Technol. Quant. Manage. 6(2) (2009),pp. 137–153.
- [13] C.H. Weiß, Thinning operations for modelling time series of counts a survey, Adv. Stat. Anal. 92(3) (2008), pp. 319–341.
- [14] C.M. Borror, C.W. Champ, and S.E. Ridgon, Poisson EWMA control charts, J. Qual. Technol. 30(4) (1998),pp. 352–361.
- [15] Capizzi, G. and Masarotto, G., 2010. Combined Shewhart–EWMA control charts with estimated parameters. Journal of Statistical Computation and Simulation, 80(7), pp.793-807.
- [16] Cembrowski S, Westgard J, Eggert A, Toren C. Trend detection in control data: optimization and interpretation of Trigg's technique for trend analysis. Clin Chem 1975, 21:1396-1405.
- [17] Chakraborti, S. (2000). Run Length, Average Run Length and False Alarm Rate of Shewhart X Chart: Exact Derivations by Conditioning, Communications in Statistics-Simulation and Computation 29, pp. 61-81.
- [18] Chakraborti, S. (2007). Run length distribution and percentiles: The Shewhart X chart with unknown parameters, Quality Engineering 19, pp 119-127.

- [19] Chakraborti, S., 2007. Run Length Distribution and Percentiles: The Shewhart Chart with Unknown Parameters. Quality Engineering, 19(2), pp.119-127.
- [20] Champ, C.W. and Woodall, W.H. (1987). Exact Results for Shewhart Control Charts with Supplementary Runs Rules, Technometrics 29, pp. 393-399.
- [21] Chen G., Cheng W.S. and Xie H. (2001). Monitoring process mean and reliability with one EWMA chart. Journal of Quality Technology. Vol. 33(2), pp.223-233.
- [22] Chen, G. (1997). The mean and Standard Deviation of the Run Length Distribution of X Charts When Control Limits Are Estimated, Statistica Sinica 7, pp.789-798.
- [23] Chen, G., Cheng, S.W. and Xie, H. (2004). A new EWMA Control Chart for Monitoring Both Location and Dispersion. Quality Technology & Quantitative Management, Vol. 2, pp.217-231.
- [24] Cheng S.W. and Thaga K. (2010). The max CUSUM chart. Frontiers in Statistical Quality Control. Vol. 2010, pp. 85-98.
- [25] Chiu, W.K and Cheung, K.C. (1977). An economic study of X charts with warning limits, Journal of Quality Technology 9, pp. 166-171.
- [26] Crowder S. and Hamilton M. (1992). An EWMA for monitoring standard deviation. Journal of Quality Technology. Vol. 24, pp. 12-21.
- [27] Cryer, J. D., & Ryan, T. P. (1990). The Estimation of Sigma for an X Chart. Journal of
- [28] D.C. Montgomery and C.M. Mastrangelo, Some statistical process control methods for auto correlated data, J. Qual. Technol. 23(3) (1991), pp. 179–193.
- [29] Diko, M., Chakraborti, S. and Does, R., 2019. Guaranteed in-control performance of the EWMA chart for monitoring the mean. Quality and Reliability Engineering International, 35(4), pp.1144-1160.
- [30] Duncan, A.J. (1956). The economic design of X charts used to maintain current control of a process, Journal of American Statistical Association 51, pp. 228-242.
- [31] Karagöz, D. (2016). Robust X control chart for monitoring the skewed and contaminated process. Hacettepe Journal of Mathematics and Statistics 47(1), pp. 223-242
- [32] Lucas, J. and Saccucci, M., 1990. Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements, [online] 32(1), pp.1-12
- [33] M. Atalay, M. Caner Testik, S. Duran, CH. Weiß (2020). Guidelines for automating Phase I of control charts by considering effects on Phase-II performance of individuals control chart. Quality Engineering.
- [34] Maravelakis, P.E., Castagliola P. (2009). An EWMA Chart for Monitoring the Process Standard Deviation when Parameters are Estimated, Computational Statistics & Data Analysis 53(7), pp. 2653-2664.
- [35] Montogomery, D.C. (2005). Introduction to Statistical Quality Control, 5th Ed. John Wiley.
- [36] Neubauer A, The EWMA control chart: properties and comparison with other quality-control procedures by computer simulation. Clin Chem 1997, 43(4) 594-601.
- [37] Pitsis, T., Beckman, S., Steinert, M., Oviedo, L. and Maisch, B., 2020. Designing the Future: Strategy, Design, and the 4th Industrial Revolution—An

Introduction to the Special Issue. California Management Review, 62(2), pp.5-11.

- [38] Prochan, F. and Savage, I.R. (1960). Starting a control chart. Industrial quality control 17, pp 12-13. Quality Technology, 22, 187–192
- [39] Quesenberry, C.P. (1993). The effect of sample size on estimated limits for X and X control charts, Journal of Quality Technology 25, pp. 237-247.
- [40] R.L. Mason and J.C. Young, It depends the effect of dependent observations on process control, Qual. Prog. (2008),pp. 70–72.
- [41] Ramzy A. and Peiris T. (2014). Comparison of rectangular and elliptical control region EWMA schemes for joint quality Monitoring. Open Journal of Statistics. Vol. 4(11), pp. 970-976.
- [42] Roberts S. Control chart tests based on geometric moving averages. Technometrics 1959:1:239-250.
- [43] Rocke, D.M. (1989). Robust Control Charts. Technometrics 31, pp. 173-184.
- [44] S. and G.E.A. Papaleonida, SPC procedures for monitoring autocorrelated processes, Qual. Technol. Quant. Manage. 4(4) (2007), pp. 501–540.
- [45] Saleh, N., Mahmoud, M., Keefe, M. and Woodall, W., 2015. The Difficulty in Designing ShewhartXandXControl Charts with Estimated Parameters. Journal of Quality Technology, 47(2), pp.127-138
- [46] Sheu S.H., Huang C.J. and Hsu T.S. (2012). Extended maximum generally weighted moving average control chart for monitoring process mean and variability. Computer &Industrial Engineering. Vol. 62(1), pp. 216-225.
- [47] Tatum, L.G. (1997). Robust Estimation of the Process Standard Deviation for Control Charts, Technometrics 39, pp. 127-141.
- [48] Tsai, T.R., Wu, S.J. and Lin, H.C. (2004). An Alternative Control Chart Approach Based on Small Number of Subgroups, International Journal of Advanced Manufacturing Technology 26, pp. 1312-1316.
- [49] Vargas, J.A. (2003). Robust Estimation in Multivariate Control Charts for Individual Observations, Journal of Quality Technology 35, pp. 367-376.
- [50] Yao, Y., & Chakraborti, S. (2021). Phase I monitoring of individual normal data: Design and implementation. Quality Engineering, 33(3)
- [51] Zaman, B., Riaz, M. and Lee, M.H. (2017). On the Performance of Control Charts for Simultaneous Monitoring of Location and Dispersion Parameters. Quality Technology & Quantitative Management, Vol. 33, pp. 37-56.
- [52] Zhang, M., Megahed, F. and Woodall, W., 2013. Exponential CUSUM Charts with Estimated Control Limits. Quality and Reliability Engineering International, 30(2), pp.275-286.
- [53] Zhang, X. and Woodall, W., 2015. Dynamic probability control limits for riskadjusted Bernoulli CUSUM charts. Statistics in Medicine, 34(25), pp.3336-3348.
- [54] Αντζουλάκος, Δ. (2003). Στατιστικός Έλεγχος Ποιότητας, Σημειώσεις παραδόσεων, Πανεπιστήμιο Πειραιά, Πειραιάς.
- [55] Καφφές, Δ. Γ. (1996). Στατιστικός Ποιοτικός Έλεγχος, Σημειώσεις παραδόσεων, Πανεπιστήμιο Πειραιά, Πειραιάς.
- [56] Δαμιανού, Χ. (1996). Στατιστικός Έλεγχος Ποιότητας και Αξιοπιστία, Σημειώσεις παραδόσεων, Πανεπιστήμιο Αθηνών, Αθήνα.


