A STUDY ON NUMBER OF SAMPLES NEEDED FOR APPLICATION OF EWMAST AND ARMA CHARTS IN MONITORING AUTOCORRELATED PROCESSES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

119 221

BY

119221

BURCU KANTAR

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF INDUSTRIAL ENGINEERING

DECEMBER 2002

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Tayfur Öztürk
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Çağlar Güven
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Prof. Dr. Taylan Ula	Assoc. Prof. Dr. Gülser Köksal
Co-Supervisor	Supervisor

Examining Committee Members

Prof. Dr. Ömer Saatçioğlu	
Assoc. Prof. Dr. Gülser Köksal	
Prof. Dr. Taylan Ula	
Assoc. Prof. Dr. Yasemin Serin	
Asst. Prof. Dr. Esra Karasakal	

ABSTRACT

A STUDY ON NUMBER OF SAMPLES NEEDED FOR APPLICATION OF EWMAST AND ARMA CHARTS IN MONITORING AUTOCORRELATED PROCESSES

Kantar, Burcu

M.Sc., Department of Industrial Engineering

Supervisor : Assoc. Prof. Dr. Gülser Köksal

Co-Supervisor : Prof. Dr. Taylan Ula

December 2002, 109 pages

Traditional control charts assume independence of the observations obtained from the monitoring process. If the observations are autocorrelated, then these charts do not work well and give misleading results. Therefore, when the observations exhibit an autocorrelation over time, some other control charts should be employed or modifications should be made in the control limits of the traditional control charts. Moreover, the performance of the charts is significantly affected by the number of samples used in the estimation of control limits. In this thesis, EWMAST chart and ARMA chart, which are used for monitoring autocorrelated observations, are considered and the effect of number of samples (with sample size 1) on the performance of these charts is investigated.

Keywords: Statistical Process Control, Autocorrelated Process, Run Length, Control Chart.

ÖZ

EWMAST VE ARMA ŞEMALARININ ÖZİLİNTİLİ SÜREÇLERİ İZLEMEK İÇİN UYGULANMASINDA GEREKLİ ÖRNEKLEM SAYISI ÜZERİNE BİR ÇALIŞMA

Kantar, Burcu

Yüksek Lisans, Endüstri Mühendisliği Bölümü

Tez Yöneticisi : Doç. Dr. Gülser Köksal

Ortak Tez Yöneticisi : Prof. Dr. Taylan Ula

Aralık 2002, 109 sayfa

Geleneksel kontrol şemaları izlenen süreçten elde edilen gözlemlerin bağımsız olduğunu varsayar. Eğer gözlemler özilintiliyse, bu şemalar iyi çalışmaz ve yanıltıcı sonuçlar verir. Dolayısıyla, gözlemler zaman içerisinde özilinti gösterdiklerinde, başka kontrol şemaları kullanılmalı veya geleneksel kontrol şemalarının kontrol sınırlarında değişiklikler yapılmalıdır. Ayrıca, şemaların performansı, kontrol sınırlarının tahmininde kullanılan örneklem sayısından önemli şekilde etkilenir. Bu tezde, özilintili gözlemlerin izlenmesinde kullanılan EWMAST şeması ve ARMA şeması dikkate alınmıştır ve örneklem sayısının (örneklem büyüklüğü 1 alınarak) bu şemaların performansına olan etkisi incelenmiştir.

Anahtar Kelimeler: İstatistiksel Süreç Kontrolü, Özilintili Süreç, Koşu Uzunluğu, Kontrol Şemaları.

To my family and Feza,

ACKNOWLEDGEMENTS

I would like to express my thanks and appreciation to my supervisor Assoc. Prof. Dr. Gülser Köksal for her invaluable contributions and support throughout the study.

I am grateful to my thesis co-supervisor Prof. Dr. Taylan Ula, for his exceptional suggestions and criticisms overall.

I extend my sincere thanks to my family, for their moral support, motivation and for sharing my problems with patience and understanding.

I would like to thank Gülden Evirgen for her motivation and worthy comments and most of all for her very special friendship.

I express my gratitude to my friend, Özlem Türker for she always did listen to me and gave suggestions for solving the problems that I encountered through the study.

Finally, I thank Feza Aytaçoğlu for he did really believe in me and did not spare his endless affectionate support throughout my master study.

TABLE OF CONTENTS

ABSTRACT.....iii
ÖZ.....iv
DEDICATION.....v
ACKNOWLEDGEMENTS.....vi
TABLE OF CONTENTS.....vii
LIST OF TABLES.....ix
LIST OF FIGURES.....x
CHAPTER
1. INTRODUCTION.....1
1.1 Aim of the Study.....1
1.2 Literature Survey.....3
2. STATISTICAL PROCESS CONTROL AND TIME SERIES.....8
2.1 Statistical Process Control.....8
2.1.1 Basic Definitions.....8
2.1.2 Statistical Process Control and Control Charts.....10
2.2 Basic Definitions for Time Series.....15
2.2.1 Stationarity and Invertibility.....15
2.2.2 The First-Order Autoregressive Process.....17
2.2.3 The First-Order Autoregressive Moving Average Model.....19
3. CONTROL CHARTS FOR AUTOCORRELATED DATA: THE EWMAST CHART AND THE ARMA CHART.....21
3.1 The EWMAST Chart.....21

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

3.2 The ARMA Chart.....26
4. THE EFFECT OF NUMBER OF SAMPLES ON THE PERFORMANCE OF EWMAST CHART AND ARMA CHART.....31
4.1 Simulation.....32
4.2 The Effect of Number of Samples on the Performance of EWMAST Chart.....33
4.3 The Effect of Number of Samples on the Performance of ARMA Chart.....56
5. CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK.....83
5.1 Conclusions.....83
5.2 Suggestions for Future Work.....86
REFERENCES.....88
APPENDICES
A. DERIVATION OF THE VARIANCE OF EWMA STATISTIC WHEN THE PROCESS DATA IS INDEPENDENT.....93
B. DERIVATION OF THE AUTOCOVARIANCE FUNCTION AND THE VARIANCE OF EWMA STATISTIC.....95
C. DERIVATION OF THE AUTOCOVARIANCE FUNCTION AND THE STEADY-STATE VARIANCE OF ARMA STATISTIC.....100
D. ALGORITHM FOR THE PARAMETER DESIGN OF ARMA CHARTS.....104
E. PLOTS OF SIGNAL-TO-NOISE RATIOS.....105

LIST OF TABLES

TABLE

4.1 Statistics for the run length distribution of EWMAST chart (Known control limits).....38

4.2 Steady-state variances of EWMA statistics for different autocorrelation levels (Known control limits).....40

4.3 Performance of the EWMAST chart with control limits estimated from n samples and (λ=0.2).....41

4.4 Statistics for the run length distribution of the EWMAST chart when h=0.0.....46

4.5 Statistics for the run length distribution of the EWMAST chart when h=1.0.....50

4.6 Statistics for the run length distribution of the EWMAST chart when h=3.0.....54

4.7 Selected ARMA chart parameters.....62

4.8 Performance of the ARMA chart with control limits estimated from n samples when h=0.0 and 1.0.....64

4.9 Statistics for the run length distribution of the ARMA chart when h=0.0 and chart parameters are selected for h=1.0.....68

4.10 Statistics for the run length distribution of the ARMA chart when h=1.0.....72

4.11 Performance of the ARMA chart with control limits estimated from n samples when h=0.0 and 3.0.....74

4.12 Statistics for the run length distribution of the ARMA chart when h=0.0 and chart parameters are selected for h=3.0.....78

4.13 Statistics for the run length distribution of the ARMA chart when h=3.0.....81

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

LIST OF FIGURES

FIGURES

2.1 A typical control chart.....11

4.1 Percentiles, mean and median of run length distribution for EWMAST chart when control limits are known.....37

4.2 Histogram of the run lengths of EWMAST chart when h=1.0 and ϕ=0.75.....39

4.3 Percentiles, mean and median of run length distribution for EWMAST chart when h=0.0.....44

4.4 Percentiles, mean and median of run length distribution for EWMAST chart when h=1.0.....48

4.5 Percentiles, mean and median of run length distribution for EWMAST chart when h=3.0.....52

4.6 Percentiles, mean and median of run length distribution for ARMA chart when h=0.0 and chart parameters are selected for h=1.0.....66

4.7 Percentiles, mean and median of run length distribution for ARMA chart when h=1.0.....70

4.8 Percentiles, mean and median of run length distribution for ARMA chart when h=0.0 and chart parameters are selected for h=3.0.....76

4.9 Percentiles, mean and median of run length distribution for ARMA chart when h=3.0.....80

E.1 Plots of signal-to-noise ratios for ϕ=-0.95, h=1.0.....105

E.2 Plots of signal-to-noise ratios for ϕ=-0.75, h=1.0.....107

E.3 Plots of signal-to-noise ratios for ϕ=0.75, h=1.0.....108

E.4 Plots of signal-to-noise ratios for ϕ=0.95, h=1.0.....109

CHAPTER 1

INTRODUCTION

1.1 Aim of the Study

Traditional control charts assume independence in the process data. However, sometimes the quality characteristic concerned exhibits autocorrelation over time. Conventional control charts may not perform well in the presence of autocorrelation. Specifically, these control charts will give misleading results in the form of too many false alarms if the data are autocorrelated (Montgomery, 2001).

The assumption of uncorrelated observations may not be satisfied in some manufacturing processes. Montgomery (2001) gives some examples of autocorrelated processes. These include chemical processes where consecutive measurements on process or product characteristics are often highly autocorrelated, or automated test and inspection procedures, where every quality characteristic is measured on every unit in time order of production. Faltin et.al. (1997) state that the conditions that result in autocorrelated data may be a result of the measurement system, the dynamics of the process, or both. If the autocorrelation does not exist in the nature of the process, but it occurs, then it should be detected. For this purpose traditional control charts are used. However, if the autocorrelation is inherent in the process, then some other control charting techniques must be employed that account for this autocorrelation, since traditional SPC procedures are not effective for

monitoring, controlling and improving process quality in the case of inherent autocorrelation in the process.

Several methods have been proposed for monitoring and controlling the mean of the autocorrelated processes. The most recent two methods are the EWMAST chart (Zhang, 1998) and the ARMA chart (Jiang et.al., 2000). EWMAST chart is an EWMA (Exponentially Weighted Moving Average) chart that is applicable to stationary process data. That is why 'ST' is used at the end of the word EWMA. ARMA chart is based on monitoring an ARMA (Autoregressive Moving Average) statistic of the original observations.

In most evaluations and comparisons of control chart performance, it is assumed that the in-control values of the parameters are known. However, the parameters must be estimated in practice. In this thesis, the EWMAST chart (Zhang, 1998) and the ARMA chart (Jiang et.al., 2000) are discussed and their performances are studied when the control limits are estimated. Estimation of the unknown parameters and the control limits is performed by different numbers of samples (with sample size 1) and the effect of number of samples on the performance of the charts is investigated. Run length statistics (i.e. Average run length, median run length, ...) are used as performance measure where run length is defined as the number of observations (samples) plotted until a plotted observation exceeds the control limits.

In practical use, conclusions obtained from this study are valid if the monitoring process is a stationary AR(1) process with normally distributed residuals and the control limits are established by the observations from a stable process.

In the next section of Chapter 1, the review of literature is provided. In Chapter 2, basic concepts relevant to SPC and Time Series are given.

Descriptions of EWMAST chart and ARMA chart are provided in Chapter 3. Chapter 4 represents the simulation results comparing the performance of each control chart for different numbers of samples and also includes the design procedure of the ARMA chart. Finally, conclusions and recommendations for relevant future studies are given in Chapter 5.

1.2 Literature Survey

Statistical Process Control (SPC), which is a sub-area of Statistical Quality Control (SQC), consists of statistical methods for monitoring and improving process performance over time. The fundamentals of SPC and control chart concept were initially developed by Walter Shewhart in 1920's and 1930's. Until 1970's there was little research conducted in this area. However, since 1980's research activity has greatly increased due to the quality revolution, which was caused by an increasingly competitive international marketplace.

Woodall and Montgomery (1999) give an overview of current research on control charting methods for process monitoring and improvement, and provide recommendations and references for the purpose of encouraging research in the area of control charting and SPC. According to them, some potential research areas in SPC are multivariate methods, effect of estimation error, short-run methods, variable sampling methods, economic design methods, change-point estimation, engineering process control and SPC, nonparametric approaches and autocorrelated (serially correlated) data. In this thesis, two of the univariate control charts proposed for autocorrelated data are considered in terms of the effect of estimation error. These control charts are the EWMAST chart (Zhang, 1998) and the ARMA chart (Jiang et.al.) as mentioned previously.

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

Control charts traditionally used for monitoring mean of a process are individual chart (Shewhart individual chart), $\overline{X}$-chart, exponentially weighted moving average (EWMA) chart and cumulative sum (CUSUM) chart. Individual control chart is constructed by plotting single observations and setting control limits on these observations. For constructing $\overline{X}$-chart, subsamples are formed and subsample means are plotted on a control chart. CUSUM control chart incorporates all the information in the sequence of sample values by plotting the cumulative sums of the deviations of the sample values from a target value. Finally, EWMA control chart is constructed by plotting EWMA statistic which is defined as $z_i=\lambda x_i+(1-\lambda)z_{i-1}$, where λ is between 0 and 1 and x_i are the original observations. All these control charts assume that the observations obtained from the monitoring process are independent.

Harris and Ross (1991) investigate the impact of autocorrelation on the performance of CUSUM and EWMA charts, and point out that the average and the median run lengths of these charts are sensitive to the presence of autocorrelation. Yashchin (1993) discuss the run length characteristics of a CUSUM chart when the underlying data have serial correlation.

Various control charts have been proposed for monitoring the mean of the autocorrelated data in recent years. Typical approach for dealing with the autocorrelated data is to fit an appropriate time series model to the observations and apply traditional control charts to the residuals from this model. Alwan and Roberts (1988) suggest using two charts for implementation, Common-Cause Chart (CCC) and Special-Cause Chart (SCC). CCC is a run chart (a chart that displays observations over a specified period of time) of the observations, with the predictions from the time series model superimposed, and SCC is an individual control chart applied to the residuals. Following this study, Montgomery and Mastrangelo (1991) propose a single control chart, which is

equivalent to the two-chart approach. This method approximates the true underlying time series model with EWMA and determines the centerline and control limits for each period. The centerline and control limits are variable.

SCC or in other words residual chart has been studied by several authors. Wardell et.al.(1994) derived the run length distributions of the residual chart, given that the assignable cause to be detected is a shift in the process mean. Zhang (1997) discussed the detection capability of residual chart for stationary processes. In addition to SCC, traditional EWMA or CUSUM charts are also applicable to the residuals since residuals from the properly fitted time series model are independent.

Another approach to deal with the autocorrelated data is to widen the control limits such that they encompass the systematic non-random behaviour resulting from autocorrelation. Vasilopoulos and Stamboulis (1978) propose the $\overline{X}$ Chart with modified limits when the underlying process is described by an AR(2) (second-order autoregressive) model. However, they ignore the autocorrelation between subsample means. Alwan and Radson (1992) investigate the time series behaviour of the subsample means for AR(1) process and propose to monitor the autocorrelated processes by plotting the averages of small batches of data separated by skipping observations. Runger and Willemain (1995) propose the Unweighted Batch Means (UBM) Chart which aims to diminish the autocorrelation between subsample means or batch means with no skipping but by taking much larger batch sizes. Following this study, again Runger and Willemain (1996) show for AR(1) process that generally better performance can be achieved by taking much larger batch sizes without skipping.

The other attempt is to modify the classical Shewhart Chart such that it can also be applied to autocorrelated data. Schmid (1995) analyze the

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

behaviour of the run length of the modified Shewhart chart and compare this chart with the classical Shewhart chart applied to the residuals.

Zhang (1998) propose a control chart for stationary process data and call it EWMAST chart. In that article, comparisons are made among the EWMAST chart, residual chart (Shewhart chart applied to the residuals, SCC) and modified Shewhart chart (Shewhart individual chart or X chart). It is concluded that when the process autocorrelation is not very strong and the mean shifts are not large, EWMAST chart performs better than the other charts.

The most recent approach for monitoring the autocorrelated data is the ARMA chart, which is proposed by Jiang et.al. (2000). This new method is based on monitoring an ARMA statistic of the original observations. It is demonstrated that the residual chart and EWMAST chart are special cases of the ARMA chart, and the ARMA chart with appropriate parameter values performs better than the residual chart and EWMAST chart in general. Jiang and Tsui (2001) further study some general properties that relate the control limit to the average run length of the ARMA chart applied to both independent and autocorrelated processes. Moreover, Jiang (2001) developed a Markov chain model for evaluating the run length performance of the ARMA chart applied to an ARMA(p, q) process.

Some authors give comparisons of the performances of the proposed charts with each other and with the classical control schemes. Wardell et.al. (1992) compare the performance of the modified Shewhart chart and EWMA chart (EWMAST chart) with the performance of residual chart, and residual chart is shown to be preferred in most cases when a shift in the process mean exceeds 2 standard deviations. Reynolds and Lu (1997) compare EWMA control chart based on the original observations with EWMA chart applied to

the residuals when the observations from the process are modelled by AR(1). Zhang (2000) makes comparisons among the EWMAST chart, the CUSUM residual chart and EWMA residual chart as well as the Shewhart residual chart and modified Shewhart chart. It is concluded that for negative autocorrelation levels, EWMAST chart performs slightly better than all the residual charts.

In the comparisons of control chart performance, usually, it is assumed that the in-control values of the parameters are known. However, the parameters must be estimated in practice. The effect of this estimation on control chart performance has been studied only for relatively few types of control charts. Quesenberry (1993) shows that $\bar{X}$ chart with subgroups of size n require about 400/(n-1) samples and X charts or individual charts require about 300 values to estimate control limits that perform like known limits. However, these results are for uncorrelated processes. For autocorrelated processes, Kramer and Schmid (2000) discuss the behaviour of the Shewhart residual chart and the modified Shewhart chart if the parameters of the underlying process are unknown and thus have to be estimated. Kantar (2002) studies the effect of number of preliminary observations on the performance of the residual chart, modified Shewhart chart and UBM chart when the underlying process is best described by an AR(1) model. In that study, performances of these charts are compared and suggestions for practical use are given.

CHAPTER 2

STATISTICAL PROCESS CONTROL AND TIME SERIES

2.1 Statistical Process Control

In this section, some basic definitions relevant to Statistical Process Control (SPC) and control charts will be given, development and implementation as well as the objectives of the control charts will be briefly explained.

2.1.1 Basic definitions

Process: A process is any set of conditions, or set of causes, which work together to produce a given result. In a manufacturing plant we usually think first of a series of fabricating and assembling operations which result, for example, in the production of cable, electron tubes, switchboards and other apparatus or equipment (Thomas et.al., 1984).

Process Variation: All processes vary. No two things are ever alike. Even though the difference between the sizes of two machined parts may be very small, they are different. There are two kinds of variation with regard to process control (Griffith, 2000).

Random variation: This is a variation that has several common (chance) causes. Common causes include vibration, heat and

humidity that either individually or collectively cause variation. These common causes are very difficult to identify and correct.

Nonrandom variation: This is a variation that is not built into the process. Causes of this type of variation are called special (assignable) causes. Special causes of variation are those that can be directly identified and corrected (particularly when you know exactly when they are occurring).

In-Control Process: A process that is operating with only chance causes of variation present is said to be in statistical control (Montgomery, 2001).

Out-of-Control Process: A process that is operating in the presence of assignable causes is said to be out of control (Montgomery, 2001).

Quality Characteristics: Every product possesses a number of elements that jointly describe what the user or consumer thinks of as quality. These parameters are often called quality characteristics (Montgomery, 2001).

Run Length: The run length (RL) of a control chart procedure is defined as the number of observations that have to be plotted until the plotted quality characteristic exceeds the control limits for the first time (Ledolter and Burrill, 1999).

Run Length Distribution: Because of the random nature of the observations, the run length follows a probability distribution. The run length can take any positive integer value 1, 2, 3,... The performance of control charts is usually characterized by their run length distribution (Ledolter and Burrill, 1999).

T.C. YÜKSEKÖĞRETİM KURU
DOKÜMANTASYON

ARL: Average number of points that must be plotted before a point indicates an out-of-control condition (Montgomery, 2001). Mean of the run length distribution.

2.1.2 Statistical Process Control and Control Charts

Many of the techniques developed by mathematical statisticians for the analysis of data may be used in the control of product quality. The expression *statistical quality control* (SQC) may be used to cover all uses of statistical techniques for this purpose (Grant and Leavenworth, 1988).

Statistical process control (SPC), which is also known under the alternative term, SQC, is the application of statistics to the control of process variability. SPC techniques help us detect unusual behaviour in the process and determine whether process outcomes are affected by special causes (Ledolter and Burrill, 1999).

SPC is a powerful collection of problem-solving tools useful in achieving process stability and improving capability through the reduction of variability. A control chart is one of the major tools of SPC (Montgomery, 2001).

A control chart is a time-sequence plot of important quality characteristics that are obtained from the process at periodic intervals (Ledolter and Burrill, 1999).

A typical control chart is given in Figure 2.1. The chart contains a *center line* that represents the average value of the quality characteristic corresponding to the in-control state. Two horizontal lines shown in the figure are called the *upper control limit* (UCL) and the *lower control limit* (LCL).

These control limits are chosen so that if the process is in control, nearly all of the points will fall between them. As long as the points plot within the control limits, the process is assumed to be in-control, and no action is necessary. However, a point that plots outside of the control limits is interpreted as evidence that the process is out-of-control, and investigation and corrective action are required to find and eliminate the assignable causes (Montgomery, 2001). Therefore, a merit of the control chart is that it tells when to leave a process alone as well as when to take action to correct trouble. 3-sigma control limits for individual measurements, when the process parameters are known, are basically given as follows:

$$UCL = \mu + 3\sigma$$

$$LCL = \mu - 3\sigma$$

where μ and σ are process mean and process standard deviation, respectively.

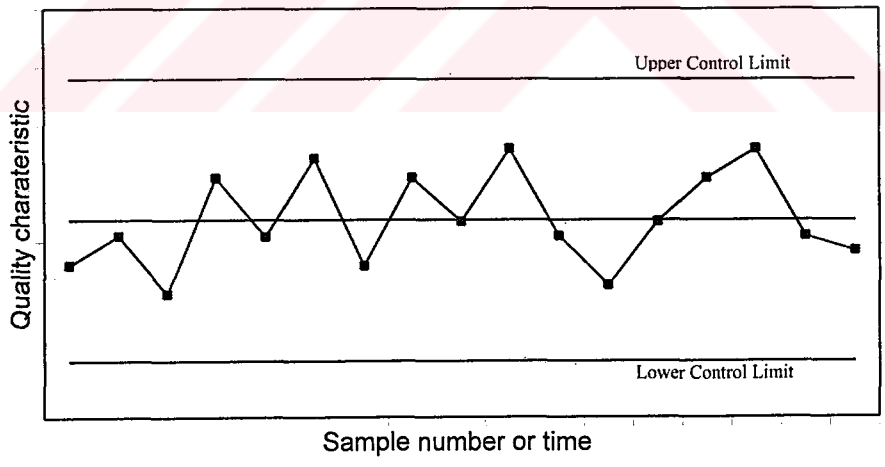


Figure 2.1 A typical control chart

In real life, true process parameters are never known and they have to be estimated. Throughout this thesis study, process mean, μ, and process

standard deviation, σ, will be estimated by sample mean and sample standard deviation, respectively. If $y_1, y_2,...,y_n$ represent the observations in a sample, then the sample mean and sample standard deviation are defined as follows:

Sample mean:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad (2.1)$$

Sample standard deviation:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}} \qquad (2.2)$$

Estimation of the control limits is one of the critical issues in the implementation. Unless sufficient number of observations is used in the estimation of control limits, they will not be reliable. Estimating the control limits farther from the center line with respect to the location of the real control limits decreases the risk of a point falling beyond the control limits, indicating an out of control condition when the process is in control. On the other hand, it increases the risk of a point falling between the control limits when the process is out of control. If the control limits are estimated closer to the center line, the opposite effect is obtained.

In practice, if the state of process changes from in-control to out-of control, then it should be detected as soon as possible by the help of the control chart. On the other hand, if it does not change, the probability that any point exceeds control limits should be small so that number of false alarms observed is small. This concept can be related to the run length, whose definition was given previously. For the former case, that is, if the state of process changes

from in-control to out-of control, then the run length should be small. However, if the process stays in-control, then the run length should be large.

Phases of Control Charts

There are two distinct phases of control chart practice. When preliminary observations are used to construct control charts, it is customary to treat the control limits as *trial control limits.* They allow us to determine whether the process was in control when the initial observations were selected. If all points plot inside the control limits and no systematic behaviour is evident, then it is concluded that the process was in control in the past, and the trial control limits are suitable for controlling current or future production. This analysis of past data is referred to as a *Phase I* analysis (Montgomery, 2001). If some points plot outside the control limits during the Phase I analysis, then assignable causes are looked for by examining each of the out-of-control points. If an assignable cause is found for the out-of-control point, the point is ignored and the trial control limits are recalculated. This process is continued until all points plot in-control. Thus, the objective of Phase I analysis is to obtain an in-control set of data to establish control limits for future monitoring purposes. These control limits are used in *Phase II* to test whether the process remains in-control, when future observations are drawn during the second phase. Therefore, Phase II consists of using the control chart to detect any departure of the underlying process from standard values (Lowry and Montgomery, 1995).

When and Where to Use Control Charts

It would be impractical to think of using a control chart at every work station in a plant. There is also no need to use control charts at a point in a manufacturing process at which it is highly unlikely that the process could ever

go out of control. Control charts should be used where trouble is likely to occur. When control charts are first implemented it is also important that they be used where the potential for cost reduction is substantial so that management can see the importance of the charts and support their continued use in a company (Ryan, 1989).

Benefits from the Use of Control Charts

Careful monitoring of products and reacting to unusual events by finding and eliminating their root causes leads to quality improvements in the processes that generate the products. There are several reasons why control charts have been used successfully for more than half a century (Ledolter and Burrill, 1999). These reasons can be listed as follows:

- They raise important issues and force their resolution. To apply control chart techniques, it is necessary to decide specifically what is to be controlled, the standards to be used, the measurements to be taken, how they are to be taken, and what will be done with them. The use of control charts forces communication and establishes responsibility on these important issues.

- They focus attention on the process rather than the product. There are two fundamental causes of poor products. A poor product can result from improper operation of a satisfactory production process. More often, however, a poor product is not due to poor workmanship, but to a poor production process. The distinction between poor workmanship and a poor process is highlighted by the use of control charts.

- They constitute a set of techniques that can be thought to people and that people can apply in a specified manner. As much as one might devalue a

'cookbook' approach to quality, experience shows that people want a prescribed procedure to assist them in achieving quality.

2.2 Basic Definitions for Time Series

In this study, it is assumed that the underlying process is best described by a stationary AR(1) model. Therefore, definition of stationarity and basic structure of AR(1) model are given below. Moreover, in order to understand the structure of the ARMA chart, which is defined in Chapter 3, ARMA(1,1) model is also described briefly and definition of invertibility is given.

2.2.1 Stationarity and Invertibility

The basic idea of stationarity is that the probability laws governing the process do not change with time, that is the process is in statistical equilibrium (Cryer, 1986).

If $\{Y_t\}$ is a *stationary* series, the mean, variance, and autocorrelation can usually be well approximated by sufficiently long time averages based on a single realization. A stochastic process having a finite mean and variance is covariance stationary if for all t, t-s, and j (Enders, 1995),

$$E(Y_t) = E(Y_{t-s}) = \mu$$

$$E\left[(Y_t - \mu)^2\right] = E\left[(Y_{t-s} - \mu)^2\right] = \sigma_y^2$$

$$\text{or } V(Y_t) = V(Y_{t-s}) = \sigma_y^2$$

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

$$E[(Y_t - \mu)(Y_{t-s} - \mu)] = E[(Y_{t-j} - \mu)(Y_{t-j-s} - \mu)] = \gamma_s$$

or $Cov(Y_t, Y_{t-s}) = Cov(Y_{t-j}, Y_{t-j-s}) = \gamma_s$

where μ, σ_y^2 and all γ_s are constants, representing mean, variance and lag s covariance of the process, respectively.

The series $\{Y_t\}$ is *invertible* if it can be represented by a finite-order or convergent autoregressive process (Enders, 1995).

Autocorrelation

The sth autocorrelation of a covariance-stationary process (denoted by ρ_s) is defined as its sth autocovariance divided by the variance (Hamilton, 1994):

$$\rho_s = \frac{\gamma_s}{\gamma_0}$$

The terminology arises from the fact that ρ_s is the correlation between Y_t and Y_{t-s}:

$$Corr(Y_t, Y_{t-s}) = \frac{Cov(Y_t, Y_{t-s})}{\sqrt{Var(Y_t)}\sqrt{Var(Y_{t-s})}} = \frac{\gamma_s}{\sqrt{\gamma_0}\sqrt{\gamma_0}} = \rho_s$$

Since ρ_s is a correlation, $|\rho_s| \leq 1$ for all s.

2.2.2 The First-Order Autoregressive Process

A first-order autoregression, denoted AR(1), satisfies the following difference equation (Hamilton, 1994):

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t$$

where the sequence $\{\varepsilon_t\}$ satisfies,

$$E(\varepsilon_t) = 0$$

$$E(\varepsilon_t^2) = \sigma_\varepsilon^2$$

$$E(\varepsilon_t \varepsilon_\tau) = 0 \quad \text{for } t \neq \tau$$

The mean of a stationary AR(1) process is $\mu = c/(1-\phi)$, where stationarity condition for AR(1) process is $|\phi| < 1$.

The variance is

$$\gamma_0 = E(Y_t - \mu)^2 = \frac{\sigma_\varepsilon^2}{1-\phi^2} \qquad (2.3)$$

while the sth autocovariance is

$$\gamma_s = E(Y_t - \mu)(Y_{t-s} - \mu) = \left[\frac{\phi^s}{1-\phi^2}\right]\sigma_\varepsilon^2, \qquad s \geq 0 \qquad (2.4)$$

It follows from (2.3) and (2.4) that the autocorrelation function is,

$$\rho_s = \frac{\gamma_s}{\gamma_0} = \phi^s \quad , \qquad s \geq 0$$

Estimation

Consider any sequence of data $y_1, y_2, \ldots, y_n$. Assuming stationarity, the autocorrelation function ρ_s for a variety of lags s=1,2,... can be estimated by the sample autocorrelation function r_s (Cryer, 1986):

$$r_s = \frac{\sum_{t=s+1}^{n}(y_t - \bar{y})(y_{t-s} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \qquad \text{for } s=1, 2,\ldots \qquad (2.5)$$

Following from (2.5), r_1, which is the estimator of ρ_1, can be written as;

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

Note that for AR(1) process, first lag autocorrelation is equal to ϕ. That is, $\rho_1 = \phi$. Therefore, for the estimation of ϕ, sample autocorrelation function can be used:

$$\hat{\phi} = r_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \qquad (2.6)$$

Throughout this thesis, first lag autocorrelation or ϕ is estimated by r_1.

2.2.3 The First-Order Autoregressive Moving Average Model

The defining equation of the First-Order Autoregressive Moving Average [ARMA(1, 1)] Model can be written as follows (Cryer, 1986):

$$Y_t = \phi Y_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}$$

where the sequence $\{\varepsilon_t\}$ satisfies,

$$E(\varepsilon_t) = 0$$

$$E(\varepsilon_t^2) = \sigma_\varepsilon^2$$

$$E(\varepsilon_t\varepsilon_\tau) = 0 \quad \text{for } t \neq \tau$$

The variance is

$$\gamma_0 = \frac{\left(1 - 2\theta\phi + \theta^2\right)}{1 - \phi^2}\sigma_\varepsilon^2 \qquad (2.7)$$

while the sth autocovariance is

$$\gamma_s = \left[\frac{(1 - \theta\phi)(\phi - \theta)}{1 - \phi^2}\right]\phi^{s-1}\sigma_\varepsilon^2 \,, \; s \geq 1 \qquad (2.8)$$

It follows from (2.7) and (2.8) that the autocorrelation function is,

$$\rho_s = \frac{\gamma_s}{\gamma_0} = \frac{(1 - \theta\phi)(\phi - \theta)}{\left(1 - 2\theta\phi + \theta^2\right)}\phi^{s-1}, \qquad s \geq 1$$

Stationarity and invertibility conditions for an ARMA(1, 1) process are $|\phi|<1$ and $|\theta|<1$, respectively.

CHAPTER 3

CONTROL CHARTS FOR AUTOCORRELATED DATA: THE EWMAST CHART AND THE ARMA CHART

In recent years, several methods have been proposed for monitoring autocorrelated processes and the most recent two methods are the EWMAST chart (Zhang, 1998) and the ARMA chart (Jiang et.al., 2000). In this study, the effect of number of samples on the estimation of control limits and thus on the performance of these charts have been investigated when the underlying process is best described by an AR(1) model. This chapter gives explanations of the EWMAST and ARMA charts with the corresponding control limits.

3.1 The EWMAST Chart

Nien Fan Zhang (1998) has proposed EWMAST chart for monitoring the autocorrelated process observations. This chart is an EWMA chart for stationary processes.

The EWMA chart has been studied for several years. Hunter (1986), Crowder (1989), and Lucas and Saccucci (1990) have discussed the performance and properties of EWMA chart.

Suppose $\{Y_t\}$ is the series of observations with mean=μ and variance=σ_y^2 and it is normally distributed. The EWMA of Y_t is defined as follows:

$$Z_t = (1-\lambda)Z_{t-1} + \lambda Y_t, \quad t=1, 2, \ldots$$

where $Z_0=\mu$ and $\lambda(0<\lambda\leq 1)$ is a constant. It follows that $E(Z_t)=\mu$ which is shown in the following equations.

$$Z_0=\mu$$

$$Z_1=(1-\lambda)\, Z_0+\lambda Y_1$$
$$E(Z_1)=(1-\lambda)\mu+\lambda\mu=\mu$$

$$Z_2=(1-\lambda)\, Z_1+\lambda Y_2$$
$$E(Z_2)=(1-\lambda)\mu+\lambda\mu=\mu$$

$$\vdots$$

$$Z_t=(1-\lambda)\, Z_{t-1}+\lambda Y_t$$
$$E(Z_t)=(1-\lambda)\mu+\lambda\mu=\mu \ , t=1, 2, 3,\ldots$$

EWMA of Y_t, which is equal to Z_t, can be rewritten as follows by replacing Z_t recursively:

$$Z_t = \lambda\sum_{j=0}^{t-1}(1-\lambda)^j Y_{t-j} + (1-\lambda)^t Z_0 \text{ where } Z_0 = \mu = 0$$

Notice that, since the series, $\{Y_t\}$, is normally distributed and Z_t is a linear combination of Y_t, t=1, 2,..., EWMA statistic, Z_t, is also distributed as normal.

If Y_t's are independent, then the variance of EWMA statistic is given as follows:

$$\sigma_z^2 = \text{var}(Z_t) = \left[\frac{\lambda}{2-\lambda}\right]\left[1-(1-\lambda)^{2t}\right]\sigma_y^2 \qquad (3.1)$$

When t is large, an approximate variance is:

$$\sigma_z^2 = \text{var}(Z_t) \approx \left[\frac{\lambda}{2-\lambda}\right]\sigma_y^2 \qquad (3.2)$$

Derivation of (3.1) can be found in Appendix A.

The EWMA chart can be established based on (3.1), and (3.2) when t is large. EWMA statistics are plotted and 3-sigma control limits of an EWMA chart are determined as follows:

$$UCL = \mu + 3\sigma_z$$
$$LCL = \mu - 3\sigma_z$$

where σ_z is the standard deviation of Z_t.

If Y_t's are not independent, then the EWMAST chart is applied. The control limits of the EWMAST chart are different from those of the ordinary EWMA chart when the data are autocorrelated.

T.C. YÜKSEK ... DOKÜMANTASYON MERKEZİ

Suppose Y_t is a stationary, autocorrelated process, with constant mean=μ and autocovariance function, γ_k. Since Y_t is a stationary process $E(Y_t)=\mu$ for all t and the autocovariance function, $\gamma_k=cov(Y_t,Y_{t+k})=E[(Y_t-\mu)(Y_{t+k}-\mu)]$, only depends on the lag k.

Under the assumption of stationarity, the autocovariance function of Z_t is given by Zhang (1998) as follows:

$$\begin{aligned} Cov(Z_t, Z_{t+\tau}) = \frac{\lambda}{2-\lambda}\sigma_y^2 \Bigg\{ & \sum_{k=0}^{t-1} \rho_{k+\tau}(1-\lambda)^k \left[1-(1-\lambda)^{2(t-k)}\right] \\ & + \sum_{k=1}^{\tau} \rho_{\tau-k}(1-\lambda)^k \left[1-(1-\lambda)^{2t}\right] \\ & + \sum_{k=\tau+1}^{t-1+\tau} \rho_{\tau-k}(1-\lambda)^k \left[1-(1-\lambda)^{2(t+\tau-k)}\right] \Bigg\} \end{aligned} \qquad (3.3)$$

where ρ_k is the autocorrelation of Y_t at lag k.

The variance of Z_t is obtained when $\tau = 0$ in equation (3.3), and it is equal to:

$$\begin{aligned} \sigma_z^2 &= var(Z_t) \\ &= \frac{\lambda}{2-\lambda}\sigma_y^2 \left\{ 1-(1-\lambda)^{2t} + 2\sum_{k=1}^{t-1} \rho_k (1-\lambda)^k \left[1-(1-\lambda)^{2(t-k)}\right] \right\} \end{aligned} \qquad (3.4)$$

Derivation of (3.3) and (3.4) can be found in Appendix B.

The steady-state variance or the variance of Z_t as $t \rightarrow \infty$ is written as follows:

$$\sigma_z^2 = \text{var}(Z_t) = \frac{\lambda}{2-\lambda}\sigma_y^2\left\{1+2\sum_{k=1}^{\infty}\rho_k(1-\lambda)^k\right\} \qquad (3.5)$$

Assuming that Y_t is normally distributed, Z_t is also normally distributed with mean = μ and variance given in (3.4). The EWMAST chart is constructed by charting Z_t. The centerline is at μ, and 3-sigma control limits are:

$$UCL = \mu + 3\sigma_z$$
$$LCL = \mu - 3\sigma_z$$

where σ_z is the standard deviation of Z_t which is the square root of (3.4), and the square root of (3.5) for large t.

Note that, if the process observations are independent, then $\rho_k=0$ for $k\geq1$ and therefore equation (3.4) becomes equal to equation (3.1) and equation (3.5) becomes equal to equation (3.2). Thus, the ordinary EWMA chart is a special case of the EWMAST chart when the sequence of observations, Y_t, are independent.

In practice, the real values of the parameters are not known and they have to be estimated. Estimation is performed by using preliminary observations. In this study, mean=μ and standard deviation=σ_y of the process data are estimated by the sample mean and sample standard deviation given in (2.1) and (2.2), respectively. Autocorrelation function of Y_t is estimated by sample autocorrelation function given in (2.5). Variance of Z_t is estimated by replacing σ_y and ρ_k with the estimated values of them

in equation (3.5). Control limits of the EWMAST chart are constructed based on these estimations.

3.2 The ARMA Chart

The ARMA Chart has been proposed by Jiang et.al.(2000). In this approach, the successive values to be plotted to the chart are the ARMA statistics which are obtained by the application of the generalized first-order autoregressive moving average [ARMA(1,1)] process to the observations.

Suppose $\{Y_t\}$ is the series of autocorrelated observations with mean=μ and variance=σ_y^2 and it is normally distributed. In the study by Jiang et.al. (2000), ARMA statistic applied to the underlying process, Y_t, is represented as follows:

$$Z_t = \theta_0 Y_t - \theta Y_{t-1} + \phi_1 Z_{t-1} \qquad (3.6)$$
$$= \theta_0 (Y_t - \beta Y_{t-1}) + \phi_1 Z_{t-1}$$

where $\beta = \theta / \theta_0$ and $\theta_0 = 1 + \theta - \phi_1$.

To guarantee that the monitoring process, Z_t, is invertible and stationary, the constraints $|\beta|<1$ for invertibility and $|\phi|<1$ for stationarity should hold.

By replacing Z_t in (3.6) recursively, equation (3.7) is obtained.

$$Z_t = \theta_0 Y_t + \alpha \sum_{k=1}^{t-1} \phi_1^{k-1} Y_{t-k} \qquad (3.7)$$

where $\alpha = \phi_1\theta_0 - \theta$ and $\theta_0 = 1 + \theta - \phi_1$

Expectation of Y_t is equal to μ, and if the initial value of Z_t is taken to be μ ($Z_1=\mu$), then the expected value of Z_t is also equal to μ. This is shown as follows:

$Z_1=\mu$

$Z_2 = \theta_0 Y_2 - \theta Y_1 + \phi_1 Z_1$

$E(Z_2)= (1 + \theta - \phi_1)E(Y_2) - \theta E(Y_1) + \phi_1\mu$

$(1 + \theta - \phi_1)\mu - \theta\mu + \phi_1\mu = \mu$

$Z_3 = \theta_0 Y_3 - \theta Y_2 + \phi_1 Z_2$

$E(Z_3)= (1 + \theta - \phi_1)E(Y_3) - \theta E(Y_2) + \phi_1\mu$

$(1 + \theta - \phi_1)\mu - \theta\mu + \phi_1\mu = \mu$

$\vdots$

$Z_t = \theta_0 Y_t - \theta Y_{t-1} + \phi_1 Z_{t-1}$

$E(Z_t)= (1 + \theta - \phi_1)E(Y_t) - \theta E(Y_{t-1}) + \phi_1\mu$

$(1 + \theta - \phi_1)\mu - \theta\mu + \phi_1\mu = \mu$

Assuming that Y_t is a stationary process with an autocovariance function of $\gamma_k=Cov(Y_t, Y_{t+k})$ and an autocorrelation function of $\rho_k=\gamma_k/\gamma_0$ ($\gamma_0=\sigma_y^2$), autocovariance function of Z_t is given by Jiang et.al. (2000) as follows:

$$\mathrm{Cov}(Z_t, Z_{t+\tau}) = \mathrm{Cov}\left[\theta_0 Y_t + \alpha\sum_{k=1}^{t-1}\phi_1^{k-1}Y_{t-k}, \theta_0 Y_{t+\tau} + \alpha\sum_{k=1}^{t+\tau-1}\phi_1^{k-1}Y_{t+\tau-k}\right]$$

$$= \theta_0^2\gamma_\tau + \theta_0\alpha\left[\sum_{k=1}^{t-1}\phi_1^{k-1}\gamma_{\tau+k} + \sum_{k=1}^{t+\tau-1}\phi_1^{k-1}\gamma_{\tau-k}\right] + \alpha^2\sum_{i=1}^{t-1}\sum_{j=1}^{t+\tau-1}\phi_1^{i+j-2}\gamma_{\tau-j+i}$$

$$= \sigma_y^2\left\{\theta_0^2\rho_\tau + \theta_0\alpha\left[\sum_{k=1}^{t-1}\phi_1^{k-1}\rho_{\tau+k} + \sum_{k=1}^{t+\tau-1}\phi_1^{k-1}\rho_{\tau-k}\right] + \alpha^2\sum_{i=1}^{t-1}\sum_{j=1}^{t+\tau-1}\phi_1^{i+j-2}\rho_{\tau-j+i}\right\} \qquad (3.8)$$

Variance of Z_t is obtained when $\tau = 0$ in equation (3.8).

$$\sigma_z^2 = \left\{\theta_0^2 + 2\theta_0\alpha\sum_{k=1}^{t-1}\phi_1^{k-1}\rho_k + \alpha^2\sum_{i=1}^{t-1}\sum_{j=1}^{t-1}\phi_1^{i+j-2}\rho_{j-i}\right\}\sigma_y^2 \qquad (3.9)$$

Similarly, the steady-state variance or the variance of Z_t as $t \to \infty$ is written as follows:

$$\sigma_z^2 = \left\{\theta_0^2 + 2\theta_0\alpha\sum_{k=1}^{\infty}\phi_1^{k-1}\rho_k + \alpha^2\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\phi_1^{i+j-2}\rho_{j-i}\right\}\sigma_y^2$$

$$= \left\{\theta_0^2 + \frac{\alpha^2}{1-\phi_1^2} + 2\left(\theta_0\alpha + \frac{\phi_1\alpha^2}{1-\phi_1^2}\right)\sum_{k=1}^{\infty}\phi_1^{k-1}\rho_k\right\}\sigma_y^2 \qquad (3.10)$$

Derivation of (3.8) and (3.10) can be found in Appendix C.

The ARMA chart can be constructed based on (3.9), and (3.10) when t is large. Then, 3-sigma control limits of the ARMA chart are as follows:

$$UCL = \mu + 3\sigma_z$$
$$LCL = \mu - 3\sigma_z$$

where σ_z is the standard deviation of Z_t which is the square root of (3.10) when t is large.

In real life, since the parameters of the process are not known they are estimated. As it is stated before, mean=μ and standard deviation=σ_y of the process data are estimated by the sample mean and sample standard deviation given in (2.1) and (2.2), respectively. Autocorrelation function of Y_t is estimated by sample autocorrelation function given in (2.5). The steady-state variance of Z_t given in (3.10) is estimated by replacing σ_y and ρ_k with the estimated values of them. Control limits of the ARMA chart are constructed based on these estimates.

EWMA and ARMA statistics as well as the control limits of the EWMAST and ARMA charts have been provided so far. Note that, EWMAST chart is a special class of ARMA charts. If ϕ_1=1-λ and θ=0, ARMA statistic will be same as the EWMA statistic.

As given previously, ARMA statistic is $Z_t = \theta_0 Y_t - \theta Y_{t-1} + \phi_1 Z_{t-1}$. Since $\theta_0 = 1 + \theta - \phi_1$, this can be rewritten as follows:

$$Z_t = Y_t + \theta Y_t - \phi_1 Y_t - \theta Y_{t-1} + \phi_1 Z_{t-1}$$

If ϕ_1=1-λ and θ=0, then

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

Finally, ARMA statistic is obtained as,

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

which is equal to the EWMA statistic given previously.

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

CHAPTER 4

THE EFFECT OF NUMBER OF SAMPLES ON THE PERFORMANCE OF EWMAST CHART AND ARMA CHART

In this chapter, the simulation procedure is explained, and the effect of number of samples (with sample size equals one) on the performance of EWMAST chart and ARMA chart is analyzed through simulation. Note that, the number of samples refers to the number of preliminary observations that is used in the estimation of unknown parameters and control limits. In practice, for the discussions in this chapter to be valid, control limits should be set by the observations from a stable process, and the process data should be stationary.

In relevant literature, it is seen that the performance of the charts is measured by Average Run Length (ARL). Therefore, ARL for each chart is estimated by simulation. However, the results have indicated that ARL by itself is insufficient for measuring the performance of the charts. For that reason, standard deviation of run lengths (SDRL), minimum run length (MINRL) and maximum run length (MAXRL) are also tabulated, and figures showing the 2.5 and 97.5 percentiles, mean and median of the run lengths are also given.

4.1 Simulation

A computer program code is developed in Visual Fortran version 6.1 to carry out simulations. The run lengths obtained by this code are analyzed and relevant statistics are calculated. The figures in this chapter are obtained by the statistical analysis software SPSS.

In this study, process measurements are assumed to come from an AR(1) model with standard normally distributed errors. n+50 normal errors are generated with variance, σ_ε^2, equal to 1, and the corresponding AR(1) observations are obtained from the model:

$$
\begin{aligned}
&Y_0 = \mu = 0 \\
&Y_t = \phi Y_{t-1} + \varepsilon_t \quad , \qquad t=1, 2,\ldots,n \qquad\qquad (4.1) \\
&\varepsilon_t \sim N(0,1)
\end{aligned}
$$

Without loss of generality, mean of the observations, μ, is taken to be zero. The initial value is also taken to be zero. After observations are generated, initial 50 values are omitted for the purpose of eliminating the effect of initial observation.

Simulations are performed by selecting different ϕ values so that, different autocorrelation levels are considered. According to Wheeler (1995), the adjustment for autocorrelation will not be appreciable unless the sample autocorrelation is greater than 0.7 or so. Therefore, high negative and positive autocorrelation levels (-0.95, -0.75, 0.75, 0.95) are taken into consideration in this study.

Generation of n observations from AR(1) model is replicated 10000 times and for each of them, the run length is determined. Statistics like ARL, SDRL, and so on are obtained by using these 10000 run lengths.

In this study, the shift in the process mean is assumed as a step shift. Originally, if there is no shift in the process mean, expectation of Y_t equals μ, and for this study, it is 0. However, when a step shift in the mean occurs in the process at time t+1, expectation of Y_{t+1} becomes $\mu^*= \mu+h\sigma_y$, here $\mu^*=h\sigma_y$, where $\sigma_y=\sqrt{1/(1-\phi^2)}$ is the standard deviation of the original observations generated from AR(1) model with standard normal errors.

To be able to obtain this step mean shift in the process at time t+1, Y_{t+1} and successive observations are generated as in the following equations:

$$Y_{t+1} = \mu^* + \phi Y_t + \varepsilon_{t+1}$$

$$Y_{t+k} = \mu^*(1-\phi) + \phi Y_{t+k-1} + \varepsilon_{t+k} \ , \quad k>1$$

Simulations are performed for h equals 0, 1 and 3. When h equals 0, it means that there is no shift in the process mean, and h is taken to be 1 and 3 for the purpose of representing small and large shifts, respectively.

4.2 The Effect of Number of Samples on the Performance of EWMAST Chart

In this section, the performance of the EWMAST chart for different number of samples and for different autocorrelation levels is investigated by simulation.

Remember that, 10000 observation set have been generated from AR(1) model. For each set EWMAST chart is constructed by plotting the corresponding EWMA statistics and estimating the control limits. The variance of the EWMA statistic for autocorrelated processes is given in (3.4). For an AR(1) model with standard normal errors, (3.4) can be rewritten as follows:

$$\sigma_z^2 = \mathrm{var}(Z_t) = \left[\frac{\lambda}{(2-\lambda)(1-\phi^2)}\right]\left\{1-(1-\lambda)^{2t}\right.$$

$$\left. +2\sum_{k=1}^{t-1}\phi^k(1-\lambda)^k\left[1-(1-\lambda)^{2(t-k)}\right]\right\} \qquad (4.2)$$

Similarly, the steady-state variance of the EWMA statistic given in (3.5) is obtained for an AR(1) process as follows:

$$\sigma_z^2 = \mathrm{var}(Z_t) = \frac{\lambda}{2-\lambda}\sigma_y^2\left\{1+2\sum_{k=1}^{\infty}\left[\phi(1-\lambda)\right]^k\right\}$$

$$= \frac{\lambda}{2-\lambda}\sigma_y^2\left\{1+\frac{2\phi(1-\lambda)}{1-\phi(1-\lambda)}\right\}$$

$$= \frac{\lambda}{2-\lambda}\sigma_y^2\left\{\frac{1+\phi(1-\lambda)}{1-\phi(1-\lambda)}\right\} \qquad (4.3)$$

After setting the control limits, new observations are generated with mean $\mu^*=h\sigma_y$ and to obtain a run length these new observations are counted until an observation plots out of control.

If there is no shift in the process mean, then $E(Y_t)=0$ and similarly expectation of EWMA statistic, $E(Z_t)$, is equal to 0.

Suppose a constant shift of $h\sigma_y$ has occurred in the original observations at time t+1, then $E(Y_{t+j})= \mu^*=h\sigma_y$ for j=1, 2, 3,.... EWMA statistics are affected in a different way from this shift in the mean. Expectation of the EWMA statistic is derived as follows:

$$E(Z_{t-j}) = 0 \text{ for } j=0, 1, 2,\ldots, t$$

$$Z_{t+1} = (1-\lambda)\, Z_t + \lambda\, Y_{t+1}$$
$$E(Z_{t+1}) = (1-\lambda)\, E(Z_t) + \lambda\, E(Y_{t+1}) = \lambda\mu^*$$

$$Z_{t+2} = (1-\lambda)\, Z_{t+1} + \lambda\, Y_{t+2}$$
$$E(Z_{t+2}) = (1-\lambda)\, E(Z_{t+1}) + \lambda\, E(Y_{t+2}) = (1-\lambda)\lambda\mu^* + \lambda\mu^*$$

$$Z_{t+3} = (1-\lambda)\, Z_{t+2} + \lambda\, Y_{t+3}$$
$$E(Z_{t+3}) = (1-\lambda)\, E(Z_{t+2}) + \lambda\, E(Y_{t+3}) = (1-\lambda)\,[(1-\lambda)\lambda\mu^* + \lambda\mu^*] + \lambda\mu^*$$
$$= (1-\lambda)^2\lambda\mu^* + (1-\lambda)\lambda\mu^* + \lambda\mu^*$$

$$\vdots$$

$$Z_{t+k} = (1-\lambda)\, Z_{t+k-1} + \lambda\, Y_{t+k}$$
$$E(Z_{t+k}) = (1-\lambda)\, E(Z_{t+k-1}) + \lambda\, E(Y_{t+k})$$
$$= (1-\lambda)^{k-1}\lambda\mu^* + (1-\lambda)^{k-2}\lambda\mu^* + \ldots + (1-\lambda)^{k-k}\lambda\mu^*$$
$$= \lambda\mu^*[1+(1-\lambda)+(1-\lambda)^2+\ldots+(1-\lambda)^{k-1}]$$
$$= \lambda\mu^* \sum_{j=0}^{k-1}(1-\lambda)^j = \lambda\mu^* \frac{1-(1-\lambda)^k}{1-(1-\lambda)} = \mu^*\left[1-(1-\lambda)^k\right]$$

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

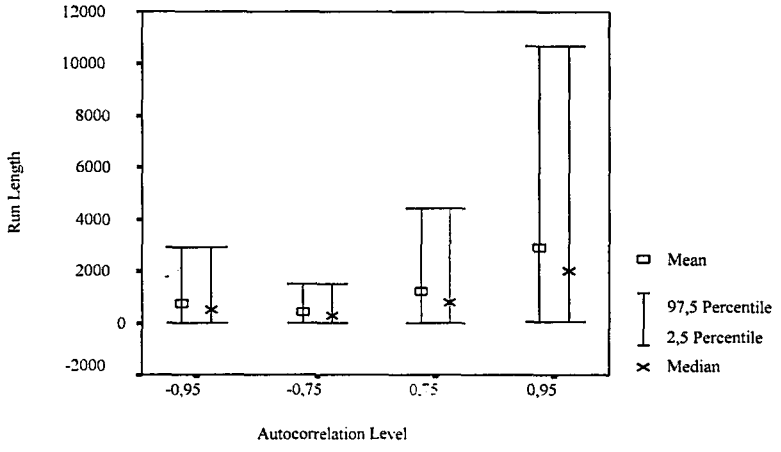
Finally,

$$E(Z_{t+k})= \mu^{*}\left[1-(1-\lambda)^{k}\right]= h\sigma_{y}\left[1-(1-\lambda)^{k}\right]= h\sqrt{\frac{1}{1-\phi^{2}}}\left[1-(1-\lambda)^{k}\right], k\geq 1$$

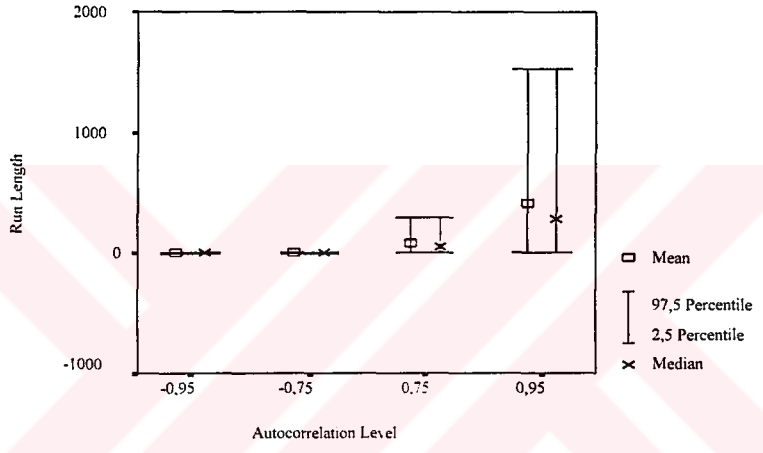
Expectation of EWMA statistic or expected shift in the EWMA statistics depends on the charting parameter, λ, and the autocorrelation level, ϕ. The variance of EWMA statistic, σ_z^2, given in (4.2) and the steady state variance given in (4.3) are also functions of λ and ϕ. Therefore, the performance of the EWMAST chart depends on the parameters λ and ϕ. In this study, λ is selected to be 0.2 as suggested by Zhang (1998).

Before working on the effect of number of samples on the performance of the EWMAST chart, the effect of the autocorrelation level, ϕ, is investigated. For this purpose, Figure 4.1 is given to illustrate the EWMAST chart performance for different values of ϕ. Note that, in Figure 4.1 it is assumed that all the parameters and thus the control limits are known and there is no estimation. In addition to Figure 4.1, Table 4.1 is given for more detailed information about the distribution of run length when the parameters are known.

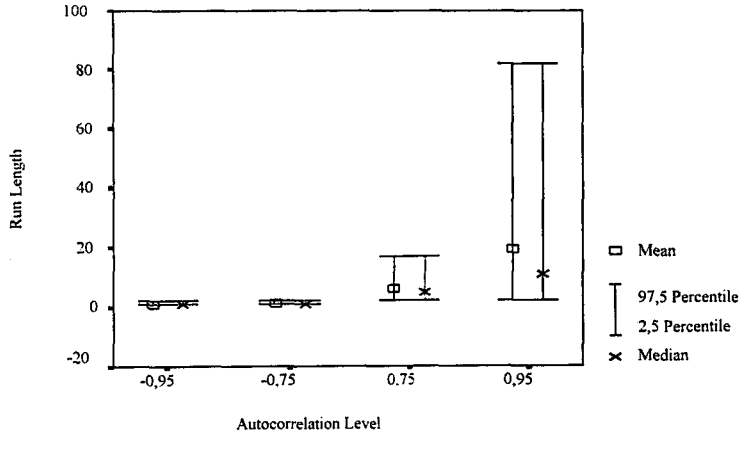
Note that, while examining the figures in this chapter one should be careful about the scale of the y-axis which denotes the run length. Scales may not be the same for every figure.



(a) h=0.0



(b) h=1.0



(c) h=3.0

Figure 4.1 Percentiles, Mean and Median of Run Length Distribution for EWMAST Chart When Control Limits are Known.

Table 4.1 Statistics for the Run Length Distribution of EWMAST Chart. (Known Control Limits)

h		ϕ			
		-0,95	**-0,75**	**0,75**	**0,95**
0,0	**Mean**	794	420	1221	2940
	Median	551	287	830	2017
	Std. Deviation	795	420	1223	2942
	Skewness	2,04	1,96	2,00	1,84
	Minimum	1,00	1,00	1,00	1,00
	Maximum	10310	3827	12590	24190
	2.5 Percentile	19,0	11,0	30,0	75,0
	97.5 Percentile	2914	1547	4459	10701
1,0	**Mean**	2,39	3,40	84,4	420
	Median	2,00	3,00	60,0	293
	Std. Deviation	0,67	1,20	78,9	421
	Skewness	-0,20	0,48	1,80	2,00
	Minimum	1,00	1,00	1,00	1,00
	Maximum	5,00	10,0	683	3978
	2.5 Percentile	1,00	1,00	5,00	10,0
	97.5 Percentile	3,00	6,00	295	1528
3,0	**Mean**	1,03	1,28	6,33	19,28
	Median	1,00	1,00	5,00	11,0
	Std. Deviation	0,17	0,45	3,88	22,0
	Skewness	5,49	1,01	1,96	2,63
	Minimum	1,00	1,00	1,00	1,00
	Maximum	2,00	2,00	41,0	239
	2.5 Percentile	1,00	1,00	2,00	2,00
	97.5 Percentile	2,00	2,00	17,0	82,0

As can be seen from Figure 4.1, run lengths are larger when the autocorrelation in the process is positive. Therefore, ARL, median of the run lengths, variation in the run length distribution and percentiles are larger for positive values of ϕ. When there is no shift in the process mean (h=0.0), it is desired to have large run lengths, hence small false alarm rate. However, if a shift occurs in the process mean, it should be detected as soon as possible. That is, run length should be small. From Figure 4.1, one can observe that, detection capability of the EWMAST chart is not good for positive autocorrelation levels. Furthermore, if positive autocorrelation is high (ϕ=0.95), then the performance gets worse. Numerical values of the run length statistics are given

in Table 4.1. When there is a small shift in the process mean (h=1.0), ARL for ϕ=0.75 and ϕ=0.95 are equal to 84 and 420, respectively, whereas ARL for ϕ= -0.75 and ϕ= -0.95 are equal to 3.40 and 2.39, respectively. There is an impressive difference between shift detection performance of the EWMAST chart for positive and negative autocorrelation levels. For example, when a large shift occurs in the process mean (h=3.0), it is detected in the 1st or 2nd observation all the time, if the autocorrelation in the process is negative. That is, the run length is equal to 1 or 2. However, for positive values of ϕ, it is possible to experience larger run lengths, such that, 97.5% of the run lengths lies between 2 and 82 for ϕ=0.95 and it lies between 2 and 17 for ϕ=0.75.

From Figure 4.1 and Table 4.1, one can also see that, distribution of the run length is positively skewed in general. Skewness changes with the value of ϕ and shift size. To give an example, histogram of the run lengths for ϕ=0.75 and h=1.0 is given in Figure 4.2.

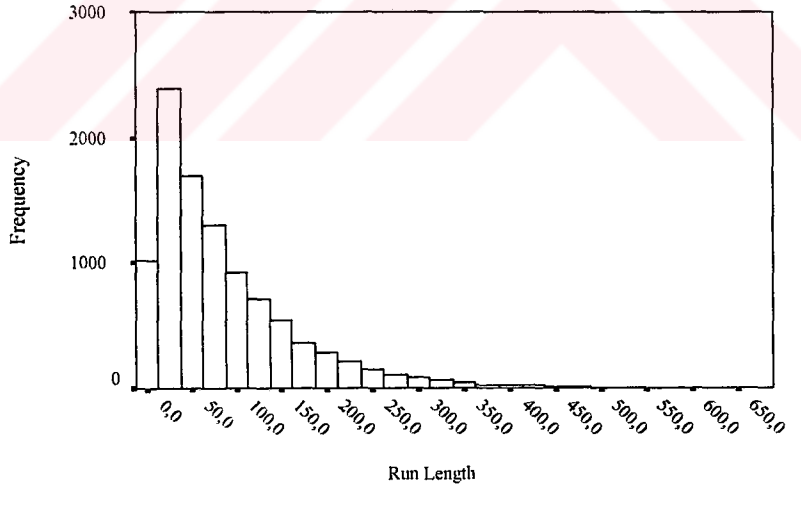


Figure 4.2 Histogram of the Run Lengths of EWMAST Chart When h=1.0 and ϕ=0.75.

Having larger run lengths when the autocorrelation in the process is positive can be explained by the variance of the EWMA statistic. The steady-state variances of EWMA statistic for different values of ϕ are calculated by using (4.3) and they are tabulated in Table 4.2.

Table 4.2 Steady-State Variances of EWMA Statistics for Different Autocorrelation Levels (Known Control Limits)

ϕ	σ_z^2
-0.95	0.1554
-0.75	0.0635
0.75	1.0159
0.95	8.3571

As can be seen in Table 4.2, variance of the EWMA statistic is larger for positive autocorrelation levels. Therefore, the control limits of the EWMAST chart is wider for positive values of ϕ than it is for negative ones. Since control limits are wider, larger run lengths are observed when the autocorrelation in the process is positive.

In order to demonstrate the performance of the EWMAST chart when the control limits are estimated from the samples in hand, Table 4.3 is obtained. For different number of samples and for different autocorrelation levels, run lengths are simulated and corresponding statistics like ARL, SDRL, MINRL and MAXRL are given in Table 4.3. Last lines for each ϕ value give the run length statistics when the estimated control limits and autocorrelation level are exactly the same as their theoretical values.

Table 4.3 Performance of the EWMAST Chart with Control Limits Estimated from n Samples (λ=0.2).

h		0,0				1,0				3,0			
ϕ	n	MIN.RL	ARL	MAX.RL	SDRL	MIN.RL	ARL	MAX.RL	SDRL	MIN.RL	ARL	MAX.RL	SDRL
0,95	200	1,00	17152	*	393112	1,00	1779	2282506	30026	1,00	19,9	13794	146
	600	1,00	4267	781059	20531	1,00	617	242525	3019	1,00	18,8	546	28,4
	1000	1,00	3557	353052	10096	1,00	509	120371	1630	1,00	18,9	529	25,7
	1800	1,00	3385	139092	6334	1,00	498	67467	1021	1,00	19,7	308	25,1
	2600	1,00	3217	84663	4838	1,00	464	37793	708	1,00	20,7	246	24,9
	3400	1,00	3144	71916	4272	1,00	447	6574	537	1,00	19,6	303	23,8
	5000	1,00	3064	46649	3744	1,00	445	7943	519	1,00	19,6	304	23,3
	7000	1,00	3004	80261	3575	1,00	440	5377	481	1,00	19,7	224	23,7
	∞	1,00	2940	24190	2942	1,00	420	3978	421	1,00	19,3	239	22,0
0,75	200	1,00	2581	4767404	49025	1,00	112	14475	306	1,00	6,04	69,0	4,52
	600	1,00	1371	56323	2403	1,00	90,0	1568	111	1,00	6,23	43,0	4,04
	1000	1,00	1283	38242	1885	1,00	87,4	1826	101	1,00	6,24	39,0	3,95
	1800	1,00	1270	34048	1618	1,00	88,0	1062	93,1	1,00	6,31	48,0	3,92
	2600	1,00	1252	14293	1439	1,00	86,9	922	89,5	1,00	6,35	34,0	3,95
	3400	1,00	1230	15283	1359	1,00	86,5	841	86,1	1,00	6,35	53,0	4,02
	5000	1,00	1239	17308	1334	1,00	85,4	884	84,2	1,00	6,39	52,0	4,06
	7000	1,00	1211	17518	1298	1,00	86,8	877	85,4	1,00	6,31	57,0	3,90
	∞	1,00	1221	12590	1223	1,00	84,4	683	78,9	1,00	6,33	41,0	3,88
-0,75	200	1,00	476	15081	798	1,00	3,39	11,0	1,24	1,00	1,26	3,00	0,44
	600	1,00	436	7147	522	1,00	3,39	10,0	1,22	1,00	1,27	3,00	0,45
	1000	1,00	421	4451	461	1,00	3,39	10,0	1,22	1,00	1,27	2,00	0,44
	1800	1,00	432	4424	454	1,00	3,41	10,0	1,22	1,00	1,28	3,00	0,45
	2600	1,00	420	4557	440	1,00	3,40	11,0	1,21	1,00	1,28	3,00	0,45
	3400	1,00	419	3906	427	1,00	3,40	10,0	1,22	1,00	1,28	3,00	0,45
	5000	1,00	430	4071	435	1,00	3,41	9,00	1,21	1,00	1,27	3,00	0,45
	7000	1,00	419	4317	429	1,00	3,40	9,00	1,20	1,00	1,27	3,00	0,45
	∞	1,00	420	3827	420	1,00	3,40	10,0	1,20	1,00	1,28	2,00	0,45
-0,95	200	1,00	16328	23415780	333625	1,00	2,37	7,00	0,79	1,00	1,05	2,00	0,23
	600	1,00	1383	207341	4509	1,00	2,38	6,00	0,71	1,00	1,04	2,00	0,19
	1000	1,00	1058	70935	2012	1,00	2,39	5,00	0,69	1,00	1,04	2,00	0,19
	1800	1,00	934	24835	1251	1,00	2,39	5,00	0,68	1,00	1,04	2,00	0,18
	2600	1,00	890	15724	1118	1,00	2,39	5,00	0,68	1,00	1,03	2,00	0,18
	3400	1,00	855	11855	989	1,00	2,39	5,00	0,68	1,00	1,03	2,00	0,18
	5000	1,00	837	13436	915	1,00	2,40	5,00	0,68	1,00	1,03	2,00	0,17
	7000	1,00	817	13403	886	1,00	2,39	5,00	0,68	1,00	1,03	2,00	0,18
	∞	1,00	794	10310	795	1,00	2,39	5,00	0,67	1,00	1,03	2,00	0,17

* : greater than 29000000

According to Table 4.3, shift detection capability of the EWMAST chart is better if the autocorrelation in the process is negative since run length statistics are smaller for negative values of ϕ when there is a shift in the process. SDRL and ARL increase in general as the number of samples used in the estimation of control limits decreases. Furthermore, when the number of samples is small it is more probable to observe very large run lengths which is not desired if there is a shift in the process mean.

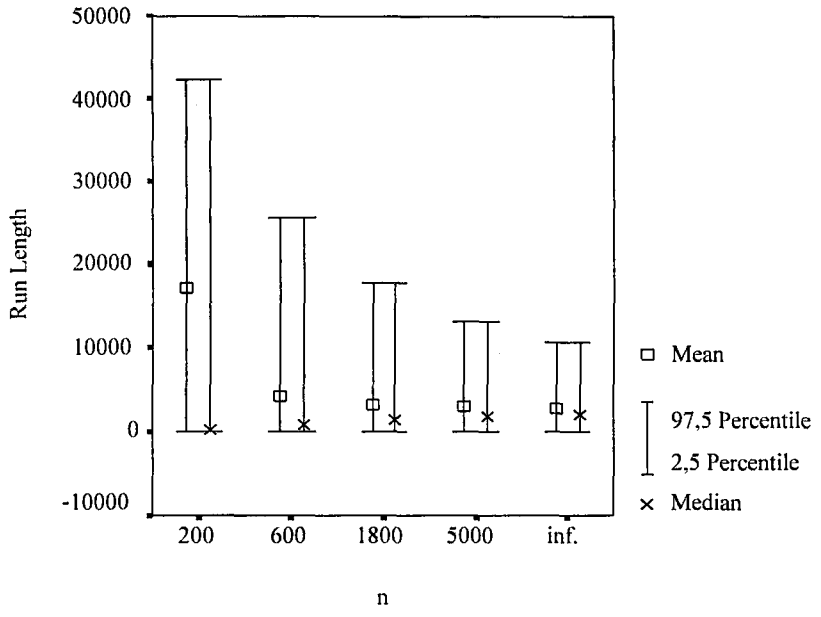
As can be seen from Table 4.3, if the autocorrelation in the process is negative, then the shift detection capability of the EWMAST chart does not seem to be affected so much with the number of samples since ARL, SDRL and MAXRL do not change much. Moreover, as the number of samples decreases in-control ARL reaches large values which is desired. Although having large variation in the run length distribution is not desired, which is the case when the number of samples is small and h=0.0, small number of samples, even n=200, may be sufficient to obtain the expected EWMAST chart performance when negative autocorrelation is present in the process. Because if number of samples is small, then run lengths are large in general when the process is in-control and at the same time mean and standard deviation of the out-of-control run lengths are similar to the out-of-control ARL and SDRL obtained with the known control limits.

From Table 4.3, it can also be seen that, if positive autocorrelation is present in the process and shift size is small (h=1.0), then ARL and SDRL increase as number of samples decreases, and this increase is more severe when positive autocorrelation is high (ϕ=0.95). However, when shift size to be detected is large (h=3.0), ARL and SDRL do not change much with the number of samples unless positive autocorrelation is high and number of samples is small (n=200). In addition, when there is no shift in the process mean ARL, SDRL and MAXRL reach large values for small number of samples.

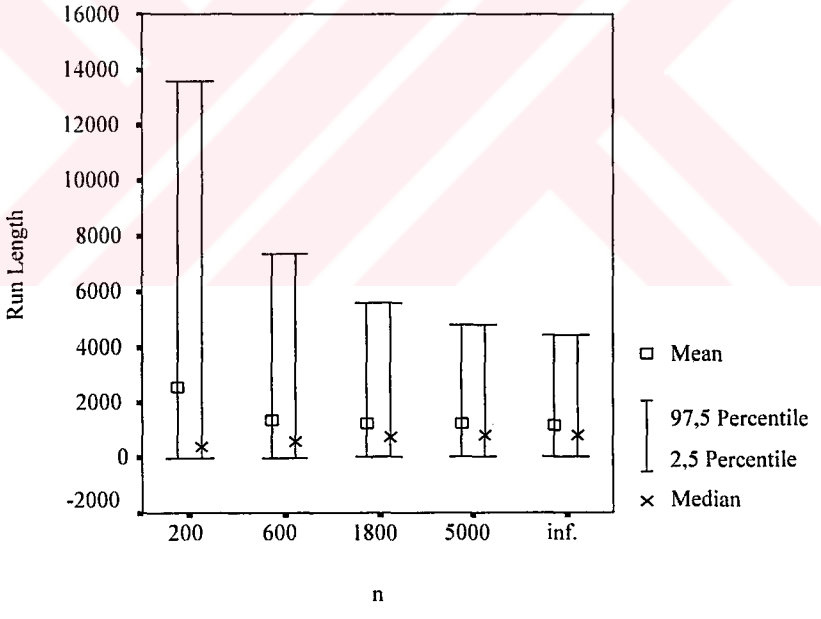
T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

In order to understand more explicitly the behaviour of the run length distribution when different number of samples are used in the estimation, figures showing the 2.5 and 97.5 percentiles, mean and median of the run lengths are obtained for different number of samples, autocorrelation levels and shift sizes. In addition to these figures, more detailed tables showing the numerical values of the run length statistics are also given.

Figure 4.3 shows the percentiles, mean and median of the run lengths for different number of samples and different values of ϕ when there is no shift in the process mean (h=0.0). Furthermore, numerical values of the relevant run length statistics are tabulated in Table 4.4.

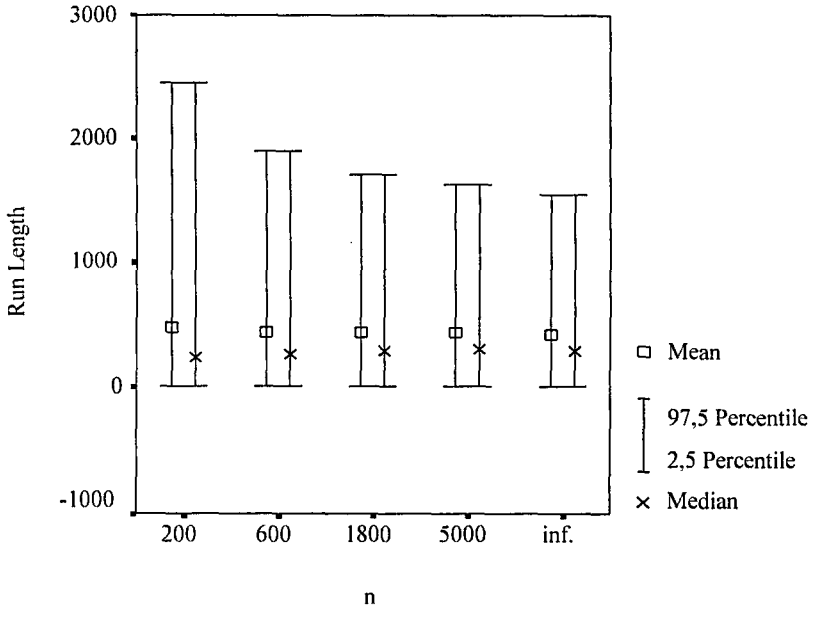


(a) ϕ=0.95

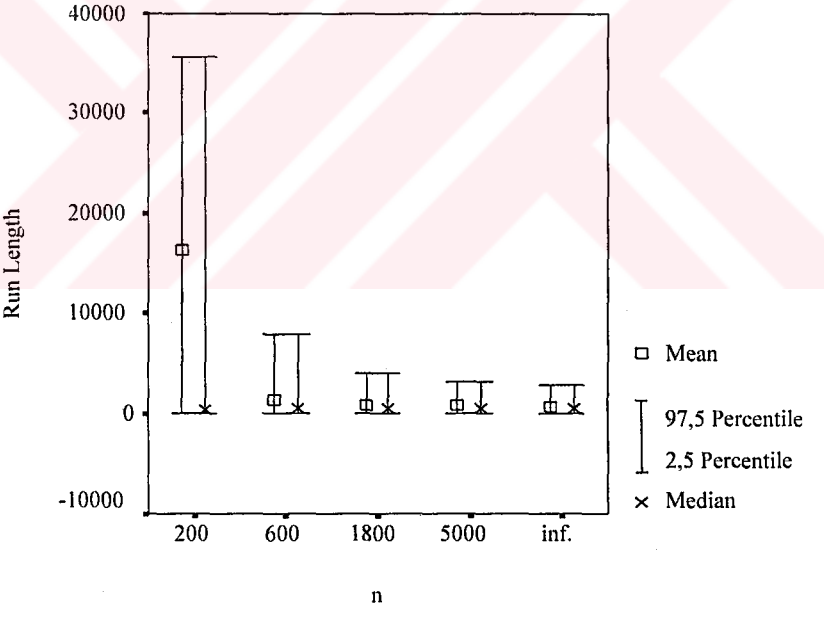


(b) ϕ=0.75

Figure 4.3 Percentiles, Mean and Median of Run Length Distribution for EWMAST Chart When h=0.0.



(c) ϕ= -0.75



(d) ϕ= -0.95

Figure 4.3 (Cont.)

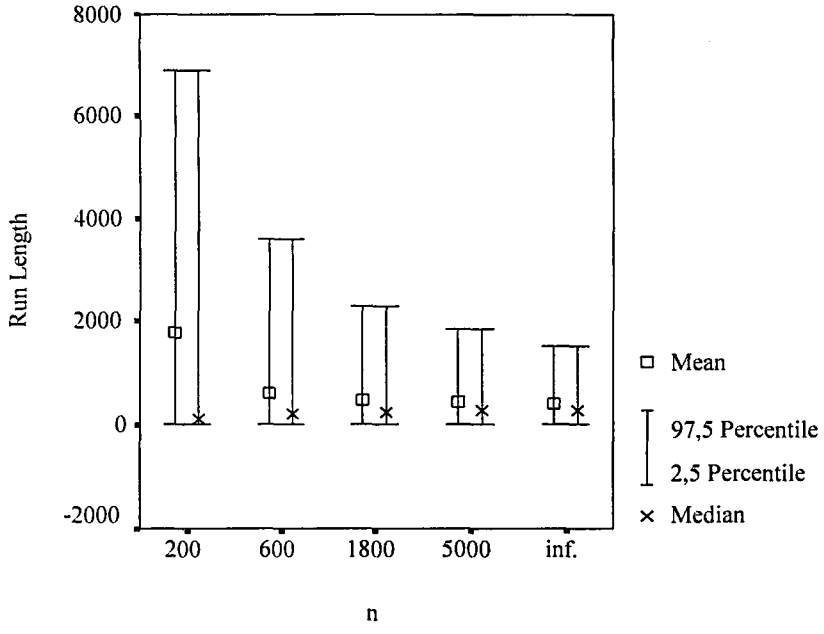
Table 4.4 Statistics for the Run Length Distribution of the EWMAST Chart When h=0.0.

ϕ		n				
		200	**600**	**1800**	**5000**	**∞**
0,95	**Mean**	17152	4267	3385	3064	2940
	Median	356	953	1531	1889	2017
	Std. Deviation	393113	20531	6334	3744	2942
	Skewness	55,3	19,5	7,07	3,38	1,84
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	*	781059	139092	46649	24190
	2.5 Percentile	10,0	24,0	45,0	61,0	75,0
	97.5 Percentile	42267	25728	17802	13115	10702
0,75	**Mean**	2581	1371	1270	1239	1221
	Median	386	625	771	818	830
	Std. Deviation	49026	2403	1618	1334	1223
	Skewness	92,2	6,33	4,09	2,59	2,00
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	4767404	56323	34048	17308	12590
	2.5 Percentile	10, 0	19, 0	25,0	28,0	30,0
	97.5 Percentile	13578	7382	5608	4827	4459
-0,75	**Mean**	476	436	432	430	420
	Median	236	264	283	296	287
	Std. Deviation	798	522	454	435	420
	Skewness	6,25	3,15	2,17	1,95	1,96
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	15081	7147	4424	4071	3827
	2.5 Percentile	9,00	11,0	11,0	11,0	11,0
	97.5 Percentile	2444	1904	1712	1631	1547
-0,95	**Mean**	16328	1383	934	837	794
	Median	423	496	549	547	551
	Std. Deviation	333627	4509	1251	915	795
	Skewness	48,5	20,0	4,77	2,57	2,04
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	23415780	207341	24835	13436	10310
	2.5 Percentile	10,0	15,0	17,0	17,0	19,0
	97.5 Percentile	35673	7938	4081	3327	2914

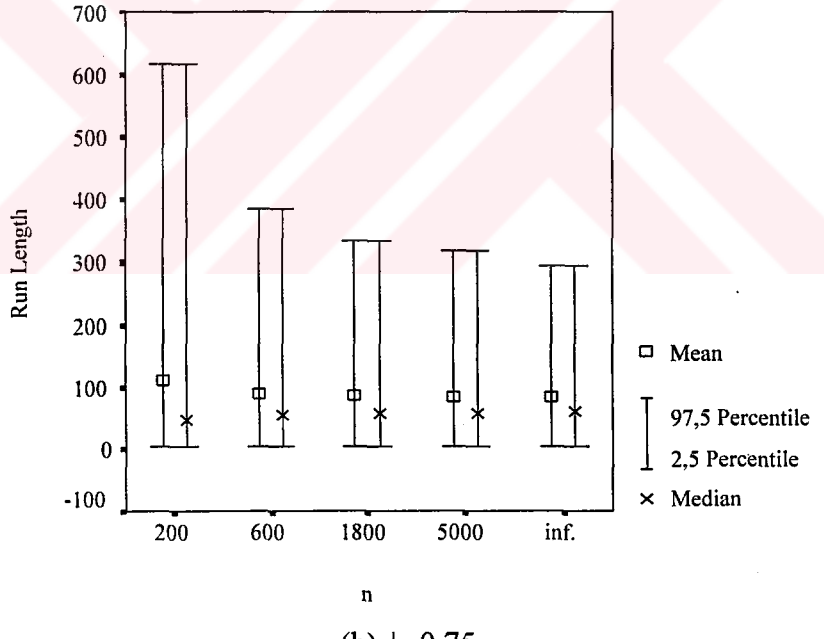
*: Greater than 29000000

Figure 4.3 and Table 4.4 indicate that as the number of samples to estimate the control limits gets smaller, the run length distribution becomes more skewed and variable. Especially when n=200, the behaviour of the run length distribution is quite different from the run length distribution obtained with the known limits and this difference is more severe for high positive or negative autocorrelation levels. According to Figure 4.3, when the number of samples used in the estimation is greater than or equal to 600, ARL and Median Run Length (MRL) does not seem to change much especially if the autocorrelation in the process is negative. However, interval between 2.5 and 97.5 percentiles gets larger as number of samples gets smaller. For example, in Table 4.4, when n=600 and ϕ= -0.95, 95% of the run length distribution lies between 15 and 7938 whereas it lies between 19 and 2914 when the control limits are known. That is, if the control limits are estimated with the samples at hand and the number of samples is not adequately large, then it is more probable to experience larger run lengths then expected. This is acceptable, if there is no shift in the process mean, however, if there occurs a shift in the process mean, then this is not desired.

The behaviour of the run length distribution in the presence of a small shift (h=1.0) in the process mean is illustrated in Figure 4.4 for different number of samples and different autocorrelation levels. In addition to Figure 4.4, in Table 4.5, detailed information is provided about the run length distribution by giving the numerical values of the run length statistics.

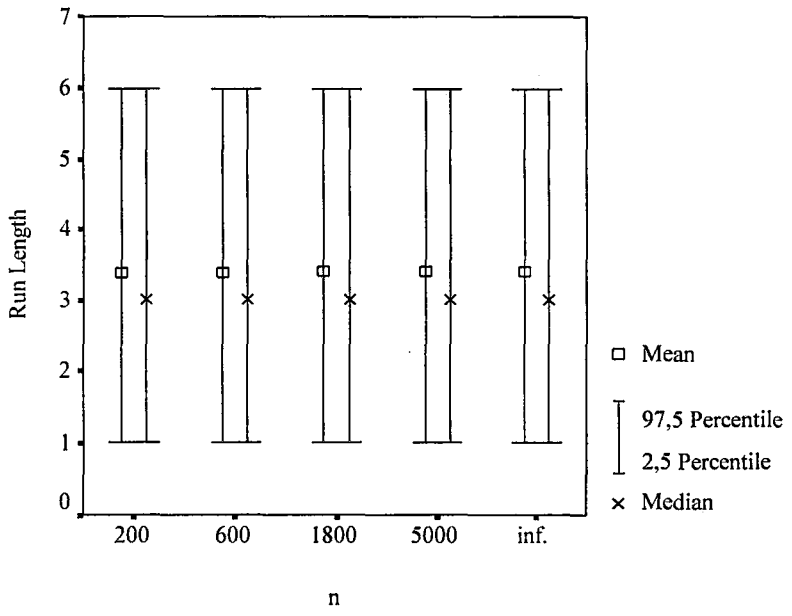


(a) ϕ=0.95

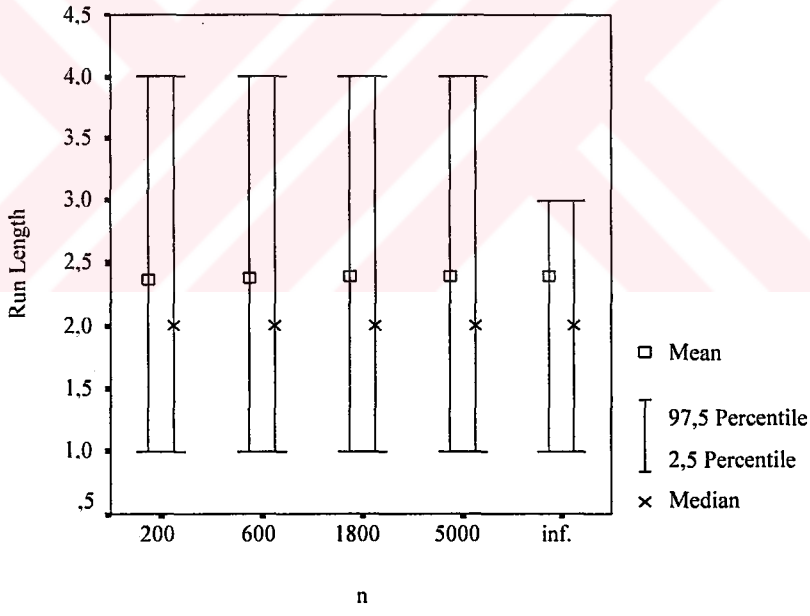


(b) ϕ=0.75

Figure 4.4 Percentiles, Mean and Median of Run Length Distribution for EWMAST Chart When h=1.0.



(c) ϕ= -0.75



(d) ϕ= -0.95

Figure 4.4 (Cont.)

Table 4.5 Statistics for the Run Length Distribution of the EWMAST Chart When h=1.0.

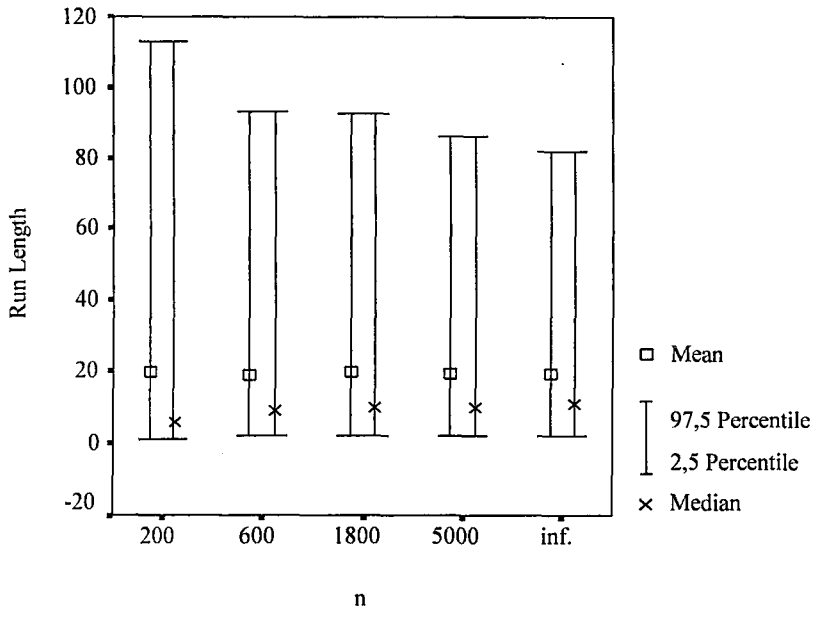
ϕ		n				
		200	**600**	**1800**	**5000**	**∞**
0,95	**Mean**	1779	617	498	445	420
	Median	114	204	262	284	293
	Std. Deviation	30027	3019	1021	519	421
	Skewness	56,1	54,8	31,0	3,20	1,99
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	2282506	242525	67467	7943	3978
	2.5 Percentile	4,00	6,00	8,00	9,00	10,0
	97.5 Percentile	6882	3596	2296	1850	1528
0,75	**Mean**	112	90,0	88,0	85,4	84,4
	Median	47,0	56,0	59,0	59,0	60,0
	Std. Deviation	306	111	93,1	84,2	78,9
	Skewness	21,1	3,85	2,61	2,18	1,80
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	14475	1568	1062	884	683
	2.5 Percentile	4,00	4,00	5,00	5,00	5,00
	97.5 Percentile	616	386	334	318	295
-0,75	**Mean**	3,39	3,39	3,41	3,41	3,40
	Median	3.00	3,00	3,00	3,00	3,00
	Std. Deviation	1,24	1,22	1,22	1,21	1,20
	Skewness	0,66	0,49	0,51	0,49	0,48
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	11,0	10,0	10,0	9,00	10,0
	2.5 Percentile	1,00	1,00	1,00	1,00	1,00
	97.5 Percentile	6,00	6,00	6,00	6,00	6,00
-0,95	**Mean**	2,37	2,38	2,39	2,40	2,39
	Median	2,00	2,00	2,00	2,00	2,00
	Std. Deviation	0,79	0,71	0,68	0,68	0,67
	Skewness	0,29	0,01	-0,13	-0,15	-0,20
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	7,00	6,00	5,00	5,00	5,00
	2.5 Percentile	1,00	1,00	1,00	1,00	1,00
	97.5 Percentile	4,00	4,00	4,00	4,00	3,00

According to Figure 4.4 and Table 4.5, when the autocorrelation in the process is positive, the behaviour of the run length distribution is affected from the number of samples used in the estimation of control limits. Such that, as the number of samples gets smaller, the run length distribution becomes more skewed and variable, ARL increases and MRL decreases. Particularly, the run length distribution obtained with the control limits which are estimated by 200 samples behaves quite different from that obtained with the known limits and this difference is more severe for high positive autocorrelation level. For example, when number of samples is 200 and ϕ=0.95, 95% of the run lengths are between 4 and 6882, however, if the control limits were known, then 95% of the run length distribution would lie between 10 and 1528. If the autocorrelation in the process is negative, run lengths are much smaller than the run lengths obtained when the autocorrelation is positive. Furthermore, run length distributions obtained with different number of samples behave similarly for negative autocorrelation levels. For example, when ϕ= -0.75, 97.5% of the run lengths are less than or equal to 6 whatever the number of samples is.

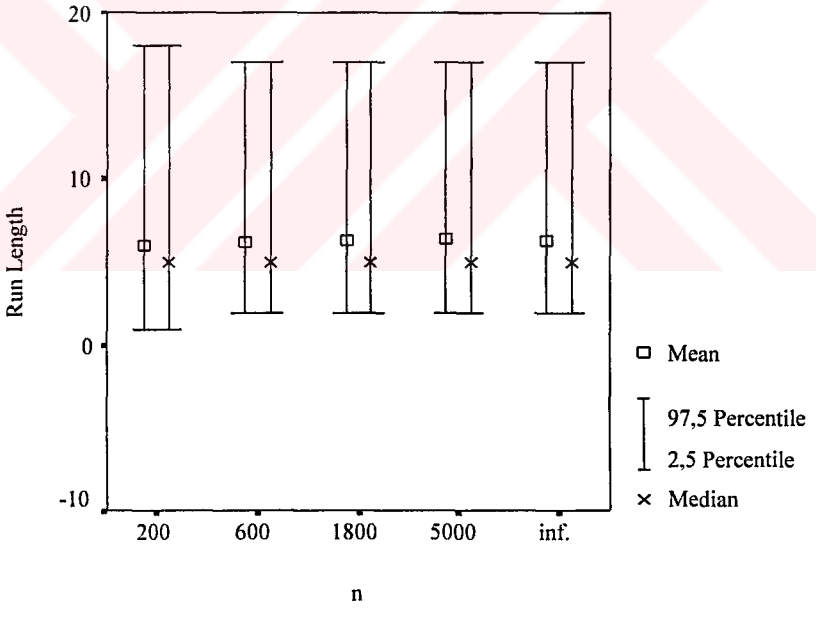
If there is a shift in the process mean, then it should be detected as soon as possible. That is, run length should be small. Therefore, considering all the explanations in the previous paragraph, one can conclude that, detection capability of the EWMAST chart get worse as number of samples gets smaller when the autocorrelation in the process is positive. However, if there is negative autocorrelation in the process, then the performance of the EWMAST chart is not affected by the number of samples used in the estimation of control limits as long as it is greater than or equal to 200.

Figure 4.5 illustrates the behaviour of the run length distribution for different number of samples and different autocorrelation levels when h=3.0. Moreover, in Table 4.6, values of the relevant run length statistics are given.

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

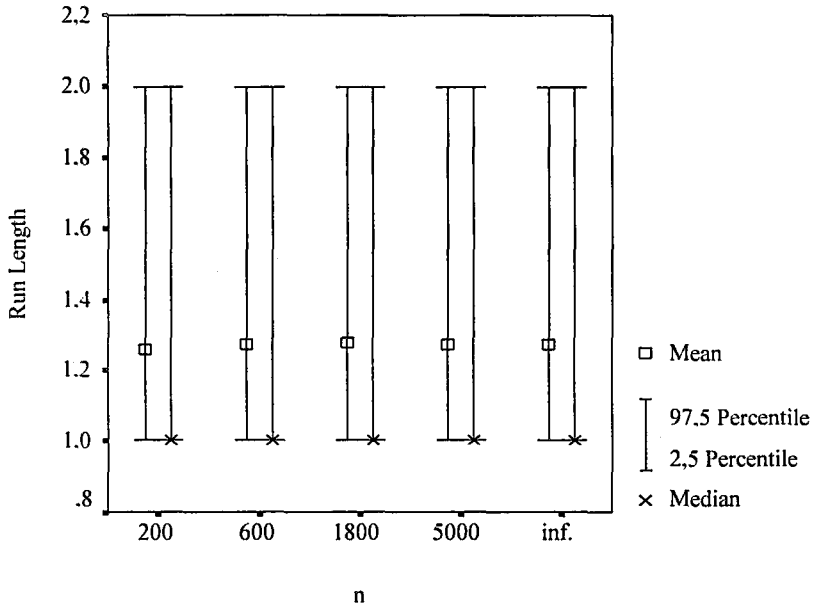


(a) ϕ=0.95

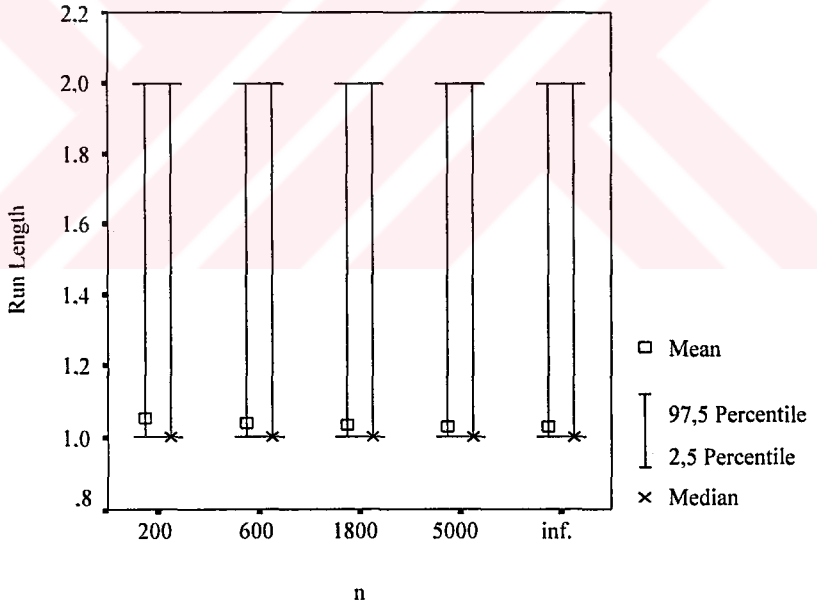


(b) ϕ=0.75

Figure 4.5 Percentiles, Mean and Median of Run Length Distribution for EWMAST Chart When h=3.0.



(c) ϕ= -0.75



(d) ϕ= -0.95

Figure 4.5 (Cont.)

Table 4.6 Statistics for the Run Length Distribution of the EWMAST Chart When h=3.0.

ϕ		n				
		200	**600**	**1800**	**5000**	∞
0,95	**Mean**	19,9	18,8	19,7	19,6	19,3
	Median	6,00	9,00	10,0	10,0	11,0
	Std. Deviation	146	28,4	25,1	23,3	22,0
	Skewness	84,1	5,04	3,23	2,90	2,63
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	13794	546	308	304	239
	2.5 Percentile	1,00	2,00	2,00	2,00	2,00
	97.5 Percentile	113	93,0	93,0	86,0	82,0
0,75	**Mean**	6,04	6,23	6,31	6,39	6,33
	Median	5,00	5,00	5,00	5,00	5,00
	Std. Deviation	4,52	4,04	3,92	4,06	3,88
	Skewness	3,02	2,12	2,07	2,13	1,96
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	69,0	43,0	48,0	52,0	41,0
	2.5 Percentile	1,00	2,00	2,00	2,00	2,00
	97.5 Percentile	18,0	17,0	17,0	17,0	17,0
-0,75	**Mean**	1,26	1,27	1,28	1,27	1,28
	Median	1,00	1,00	1,00	1,00	1,00
	Std. Deviation	0,44	0,45	0,45	0,45	0,45
	Skewness	1,11	1,03	1,01	1,02	1,01
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	3,00	3,00	3,00	3,00	2,00
	2.5 Percentile	1,00	1,00	1,00	1,00	1,00
	97.5 Percentile	2,00	2,00	2,00	2,00	2,00
-0,95	**Mean**	1,05	1,04	1,04	1,03	1,03
	Median	1,00	1,00	1,00	1,00	1,00
	Std. Deviation	0,23	0,19	0,18	0,17	0,17
	Skewness	3,97	4,74	5,05	5,38	5,49
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	2,00	2,00	2,00	2,00	2,00
	2.5 Percentile	1,00	1,00	1,00	1,00	1,00
	97.5 Percentile	2,00	2,00	2,00	2,00	2,00

As can be seen in Figure 4.5 and Table 4.6, when h=3.0, the behaviour of the run length distributions for different number of samples are similar unless there is high positive autocorrelation (ϕ=0.95) in the process. When ϕ equals 0.95, ARL and MRL do not change much as number of samples used in the estimation of the control limits gets smaller. However, as number of samples decreases run length distribution becomes more skewed and variable. If shift to be detected is large and the autocorrelation in the process is negative or not high, then the performance of the EWMAST chart with the estimated control limits is similar to that of with the known limits even when the number of samples used in the estimation is equal to 200.

As a conclusion, if the autocorrelation in the process is negative, then using 200 samples with sample size 1 to estimate the control limits will be sufficient to have similar shift detection performance of the EWMAST chart with the known control limits. However, when the autocorrelation in the process is positive and high (ϕ=0.95), and shift to be detected is small (h=1.0), it is suggested to use at least 1800 samples to estimate the control limits of the EWMAST chart. This number can be reduced to 600 if the positive autocorrelation is lower (ϕ=0.75). If shift to be detected is large (h=3.0) and there is no high positive autocorrelation in the process, then using 200 samples to estimate the control limits will be adequate in order to have a similar performance of the EWMAST chart with the known control limits. However, if the process is positively autocorrelated and this autocorrelation is high, then it is suggested to use at least 600 samples for the estimation of the control limits.

4.3 The Effect of Number of Samples on the Performance of ARMA Chart

In this section of the thesis, an ARMA chart is designed and by a simulation study the effect of number of samples on the performance of the ARMA chart is analyzed.

Design of the ARMA chart is performed by modifying the algorithm given by Jiang et.al. (2000). Afterwards, the ARMA chart is constructed by plotting the ARMA statistics calculated from the original observations and by estimating the control limits. The steady-state variance of the ARMA statistic is given in (3.10). For an AR(1) model with standard normal errors, (3.10) can be rewritten as follows:

$$\sigma_z^2 = \left\{\theta_0^2 + \frac{\alpha^2}{1-\phi_1^2} + 2\left(\theta_0\alpha + \frac{\phi_1\alpha^2}{1-\phi_1^2}\right)\sum_{k=1}^{\infty}\phi_1^{k-1}\rho_k\right\}\sigma_y^2$$

$$= \left\{\theta_0^2 + \frac{\alpha^2}{1-\phi_1^2} + 2\left(\theta_0\alpha + \frac{\phi_1\alpha^2}{1-\phi_1^2}\right)\rho_1\sum_{k=1}^{\infty}(\phi\phi_1)^{k-1}\right\}\sigma_y^2$$

$$= \left\{\theta_0^2 + \frac{\alpha^2}{1-\phi_1^2} + 2\left(\theta_0\alpha + \frac{\phi_1\alpha^2}{1-\phi_1^2}\right)\frac{\phi}{1-\phi\phi_1}\right\}\sigma_y^2 \qquad (4.4)$$

After setting the control limits, run lengths for ARMA chart are determined, similarly, as they are determined for EWMAST chart.

If a step shift of $h\sigma_y$ occurs in the process mean at time t+1, then expectation of the original observations will be $h\sigma_y$. That is, $E(Y_{t+j})=\mu^*=h\sigma_y$

for j=1, 2, 3,.... However, expectation of the ARMA statistics or shift in the ARMA statistics will be different from $h\sigma_y$ and it is derived as follows:

The ARMA statistic given in (3.6) is,

$$Z_t = \theta_0 Y_t - \theta Y_{t-1} + \phi_1 Z_{t-1}$$

and $E(Z_t)=\mu=0$ as shown in Chapter 3.

$$Z_{t+1} = \theta_0 Y_{t+1} - \theta Y_t + \phi_1 Z_t$$
$$E(Z_{t+1}) = \theta_0 E(Y_{t+1}) - \theta E(Y_t) + \phi_1 E(Z_t) = \theta_0 \mu^*$$

$$Z_{t+2} = \theta_0 Y_{t+2} - \theta Y_{t+1} + \phi_1 Z_{t+1}$$
$$E(Z_{t+2}) = \theta_0 E(Y_{t+2}) - \theta E(Y_{t+1}) + \phi_1 E(Z_{t+1}) = \theta_0 \mu^* - \theta \mu^* + \phi_1 \theta_0 \mu^*$$

Remember that $\theta_0 = 1 + \theta - \phi_1$. Therefore,

$$E(Z_{t+2}) = \mu^*[1+\theta-\phi_1-\theta+\phi_1(1+\theta-\phi_1)] = \mu^*[1+\phi_1(\theta-\phi_1)]$$

$$Z_{t+3} = \theta_0 Y_{t+3} - \theta Y_{t+2} + \phi_1 Z_{t+2}$$
$$E(Z_{t+3}) = \theta_0 E(Y_{t+3}) - \theta E(Y_{t+2}) + \phi_1 E(Z_{t+2})$$
$$= (1+\theta-\phi_1)\mu^* - \theta\mu^* + \phi_1\mu^*[1+\phi_1(\theta-\phi_1)]$$
$$= \mu^*\left[1+\phi_1^2(\theta-\phi_1)\right]$$

$$\vdots$$

$$Z_{t+k} = \theta_0 Y_{t+k} - \theta Y_{t+k-1} + \phi_1 Z_{t+k-1}$$

$$E(Z_{t+k}) = \theta_0 E(Y_{t+k}) - \theta E(Y_{t+k-1}) + \phi_1 E(Z_{t+k-1})$$

$$= (1+\theta-\phi_1)\mu^* - \theta\mu^* + \phi_1\mu^* \left[1+\phi_1^{k-2}(\theta-\phi_1)\right]$$

$$= \mu^* \left[1+\phi_1^{k-1}(\theta-\phi_1)\right]$$

Finally, the shift pattern on the charting process Z_t is,

$$E(Z_{t+k}) = \mu^* \left[1+\phi_1^{k-1}(\theta-\phi_1)\right] = h\sigma_y \left[1+\phi_1^{k-1}(\theta-\phi_1)\right], \; k \geq 1 \qquad (4.5)$$

Then for an AR(1) process with standard normal errors,

$$E(Z_{t+k}) = h\sqrt{\frac{1}{1-\phi^2}}\left[1+\phi_1^{k-1}(\theta-\phi_1)\right], \; k \geq 1 \qquad (4.6)$$

As can be seen in (4.6), expectation of ARMA statistic or expected shift in the ARMA statistics depends on the charting parameters θ and ϕ_1 as well as the process parameter ϕ. The variance of ARMA statistic given in (4.4) is also a function of ϕ, θ and ϕ_1. Hence, the performance of the ARMA chart depends on all three parameters.

In order to choose appropriate ARMA chart parameters, a heuristic approach is proposed by Jiang et.al. (2000). This approach will be explained in the following paragraphs.

As shown previously, the shift pattern on the charting process Z_t after the time of the shift t+1 follows equation (4.5). The transient shift is defined as $\mu_T = \theta_0\mu^*$ and the steady-state shift as $\mu_S = \mu^*$. Thus, the transient and steady-state signal-to-noise ratios are defined as follows:

Transient signal-to-noise ratio is defined as,

$$R_T=\mu_T/\sigma_z$$

and steady-state signal-to-noise ratio as,

$$R_S=\mu_S/\sigma_z$$

When μ_T, μ_S and σ_z are replaced with their values, these two ratios can be rewritten as follows:

$$R_T = \frac{\theta_0 h \sigma_y}{\sigma_y \sqrt{\theta_0^2 + \frac{\alpha^2}{1-\phi_1^2} + 2\left(\theta_0\alpha + \frac{\phi_1\alpha^2}{1-\phi_1^2}\right)\sum_{k=1}^{\infty}\phi_1^{k-1}\rho_k}}$$

$$= \frac{\theta_0 h}{\sqrt{\theta_0^2 + \frac{\alpha^2}{1-\phi_1^2} + 2\left(\theta_0\alpha + \frac{\phi_1\alpha^2}{1-\phi_1^2}\right)\sum_{k=1}^{\infty}\phi_1^{k-1}\rho_k}} \qquad (4.7)$$

$$R_S = \frac{h \sigma_y}{\sigma_y \sqrt{\theta_0^2 + \frac{\alpha^2}{1-\phi_1^2} + 2\left(\theta_0\alpha + \frac{\phi_1\alpha^2}{1-\phi_1^2}\right)\sum_{k=1}^{\infty}\phi_1^{k-1}\rho_k}}$$

$$= \frac{h}{\sqrt{\theta_0^2 + \frac{\alpha^2}{1-\phi_1^2} + 2\left(\theta_0\alpha + \frac{\phi_1\alpha^2}{1-\phi_1^2}\right)\sum_{k=1}^{\infty}\phi_1^{k-1}\rho_k}} \qquad (4.8)$$

If the underlying process is AR(1) process with standard normal errors, then the equations (4.7) and (4.8) are reduced to

$$R_T = \frac{\theta_0 h}{\sqrt{\theta_0^2 + \frac{\alpha^2}{1-\phi_1^2} + 2\left(\theta_0\alpha + \frac{\phi_1\alpha^2}{1-\phi_1^2}\right)\frac{\phi}{1-\phi\phi_1}}} \qquad (4.9)$$

$$R_S = \frac{h}{\sqrt{\theta_0^2 + \frac{\alpha^2}{1-\phi_1^2} + 2\left(\theta_0\alpha + \frac{\phi_1\alpha^2}{1-\phi_1^2}\right)\frac{\phi}{1-\phi\phi_1}}} \qquad (4.10)$$

According to Jiang et.al. (2000), these two ratios are critical to choose appropriate parameters of an ARMA chart for autocorrelated processes. The transient ratio measures the capability of the chart to detect the shift in the first few runs. If the chart fails to signal the shift in the early runs, the steady-state ratio becomes important for detecting the shift efficiently in the later runs. Jiang et.al. (2000) have developed a heuristic algorithm, which is based on transient and steady-state signal-to-noise ratios, in order to choose appropriate ARMA chart parameters. This algorithm can be found in Appendix D.

Unfortunately, this algorithm is difficult to implement in practice and it sometimes does not give a single solution of appropriate charting parameters (ϕ_1, θ). In the balancing part of the algorithm one should check the decreasing rate of R_T, however, it is not clear what rate is high or low. Therefore, in this study, a small modification has been made in the algorithm in order to obtain a table which gives the appropriate ARMA chart parameters for different process parameter values of AR(1) process. According to this modification, out-of-control ARL of possible parameter candidates obtained from the algorithm are simulated and appropriate charting parameters are selected so that the simulated out-of-control ARL is minimum. While estimating ARL, control

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

limits are set as $\hat{\mu} \pm L\hat{\sigma}_z$ where L is equal to 3. One should note that the Shewhart chart with 3-sigma control limits has an in-control ARL of 370. According to the personal communication with Jiang, W., to achieve the same in-control ARL, the L value for the ARMA chart is always less than 3. This is also mentioned in the study by Jiang and Tsui (2001). Using 3 as the L value for the ARMA chart would lead to less false alarms but also affects the sensitivity of the ARMA chart in detecting the shifts. Therefore, comparing the out-of-control ARL of ARMA chart using the same L value for different parameters may not be reasonable. However, the major concern of this study is the practical use of the chart. Therefore, it is assumed that many would tend to choose L as 3 unless there is a comparably easy way of choosing L. Taking L as 3 will always give an in-control ARL of ARMA chart greater than 370 whatever the parameters of the ARMA chart are. Thus, it may be assumed that false alarm rate of the chart is already small for different chart parameters when L is 3 and hence out-of-control run lengths can be compared for selecting the appropriate chart parameters. According to the personal communication with Jiang, W., in practice, 3-sigma is a reasonable alternative of control limits for an ARMA chart once its parameters are determined. However, better choices are also possible. Jiang and Tsui (2001) discuss the choice of L.

By implementing the algorithm as defined in the previous paragraph, the parameters of the ARMA chart are determined for some process parameter values of an AR(1) process, and tabulated in Table 4.7. One should also note that, R_T and R_S are functions of the shift size, h. Therefore, selected chart parameters in Table 4.7 may be different for different shift sizes.

Table 4.7 Selected ARMA Chart Parameters

Process Parameter	**Shift Size**	**Charting Parameters**		
ϕ	**h**	ϕ_1	θ	$\beta=\theta/\theta_0$
-0,95	1	-0,55	-0,77	-0,98
	2	0,00	-0,49	-0,95
	3	0,00	-0,49	-0,95
-0,85	1	-0,05	-0,52	-0,97
	2	-0,71	-0,84	-0,97
	3	0,00	-0,46	-0,85
-0,75	1	0,26	-0,33	-0,79
	2	-0,52	-0,72	-0,91
	3	0,00	-0,43	-0,75
-0,55	1	0,54	-0,20	-0,74
	2	-0,17	-0,49	-0,72
	3	-0,75	-0,82	-0,88
0,55	1	0,94	0,21	0,78
	2	0,79	0,29	0,58
	3	-0,54	-0,37	-0,32
0,75	1	0,98	0,27	0,93
	2	0,97	0,72	0,96
	3	0,00	3,00	0,75
0,85	1	0,99	0,24	0,96
	2	0,97	0,72	0,96
	3	0,00	5,67	0,85
0,95	1	0,99	-0,005	-0,98
	2	0,00	19,00	0,95
	3	0,00	19,00	0,95

Note that, to form Table 4.7, transient and steady-state ratios are calculated for different charting parameters (ϕ_1 and $\beta=\theta/\theta_0$), process parameters (ϕ) and shift sizes (h). Plots of transient and steady-state ratios (R_T, R_S) for different charting parameters (ϕ_1 and β) when h=1.0 and ϕ=-0.95, -0.75, 0.75 and 0.95 are obtained and they are given in Appendix E.

Selected ARMA chart parameters are different for different shift sizes. Therefore, two different ARMA charts are constructed for each process parameter, ϕ, in order to demonstrate the performance of the ARMA chart for detecting small (h=1.0) and large (h=3.0) shifts when the control limits are estimated from the samples at hand.

Table 4.8 gives ARL, SDRL, MINRL and MAXRL for different values of ϕ and for different number of samples when h=0.0 and 1.0. Charting parameters in Table 4.8 are taken from Table 4.7.

Table 4.8 Performance of the ARMA Chart with Control Limits Estimated from n Samples when h=0.0 and 1.0.

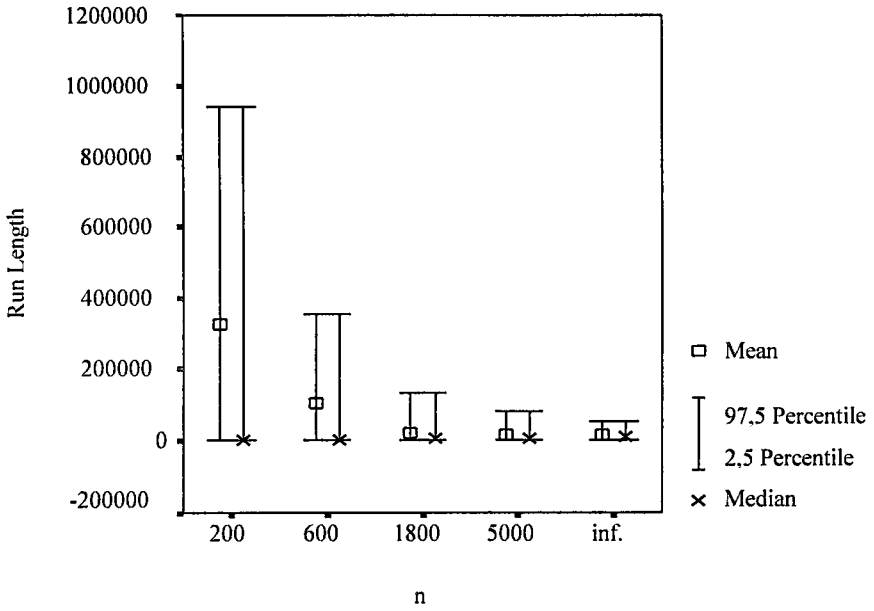
Process Parameter	Charting Parameters			Number of Samples	h							
					0,0				1,0			
ϕ	ϕ_1	θ	$\beta=\theta/\theta_0$	n	MIN.RL	ARL	MAX.RL	SDRL	MIN.RL	ARL	MAX.RL	SDRL
0,95	0,99	-0,005	-0,98	200	1,00	325022	*	2656662	1,00	16886	*	549792
				600	1,00	103907	*	1212006	1,00	921	1501108	18155
				1000	1,00	52848	*	661767	1,00	503	269564	3017
				1800	1,00	23091	6657885	128118	1,00	470	212637	2229
				2600	1,00	18054	2001620	50089	1,00	443	66693	858
				3400	1,00	16641	1868982	38068	1,00	399	11543	399
				5000	1,00	15724	662380	28587	1,00	408	5848	384
				7000	1,00	15394	412754	22960	1,00	397	5706	357
				∞	1,00	13963	113823	13978	1,00	376	2810	288
0,75	0,98	0,27	0,93	200	1,00	1645	4144321	44281	1,00	53,1	1119	53,5
				600	1,00	810	23426	1241	1,00	51,1	535	40,6
				1000	1,00	790	33454	1072	1,00	50,6	398	37,8
				1800	1,00	795	13407	944	1,00	50,5	319	37,0
				2600	1,00	805	13547	902	1,00	52,1	396	36,3
				3400	1,00	811	12025	901	1,00	50,3	296	35,1
				5000	1,00	817	9177	858	1,00	50,0	402	34,9
				7000	1,00	799	7831	835	1,00	50,6	333	35,8
				∞	1,00	807	7530	807	1,00	50,3	324	35,4
-0,75	0,26	-0,33	-0,79	200	1,00	444	10066	591	1,00	2,89	16,0	1,40
				600	1,00	399	4923	434	1,00	2,85	16,0	1,34
				1000	1,00	389	4479	399	1,00	2,82	12,0	1,29
				1800	1,00	382	3658	399	1,00	2,85	11,0	1,30
				2600	1,00	386	3702	394	1,00	2,84	17,0	1,30
				3400	1,00	371	3876	379	1,00	2,83	15,0	1,29
				5000	1,00	382	3423	383	1,00	2,83	13,0	1,28
				7000	1,00	381	3642	384	1,00	2,82	17,0	1,28
				∞	1,00	375	3482	374	1,00	2,83	16,0	1,32
-0,95	-0,55	-0,77	-0,98	200	1,00	825	149048	2329	1,00	1,76	13,0	0,73
				600	1,00	483	7597	546	1,00	1,70	6,00	0,63
				1000	1,00	444	4928	470	1,00	1,69	6,00	0,62
				1800	1,00	420	3554	429	1,00	1,68	5,00	0,61
				2600	1,00	412	3925	413	1,00	1,67	6,00	0,61
				3400	1,00	414	4515	412	1,00	1,67	7,00	0,60
				5000	1,00	404	3981	412	1,00	1,67	5,00	0,62
				7000	1,00	397	3857	399	1,00	1,67	6,00	0,61
				∞	1,00	402	3571	396	1,00	1,67	6,00	0,60

*: Greater than 29900000

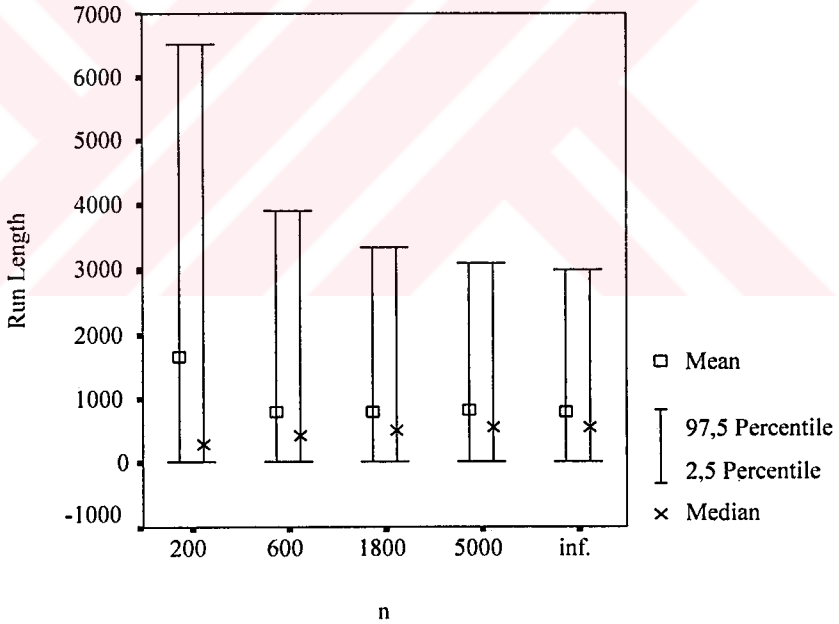
Table 4.8 indicates that the performance of the ARMA chart in detecting the shift of size $1\sigma_y$ is better when the autocorrelation in the process is negative. The run lengths when h=1.0 are smaller for negative values of ϕ. ARL and SDRL increase as the number of samples used in the estimation of the control limits decreases. However, this increase in the ARL and SDRL is very much smaller when the autocorrelation in the process is negative. In the presence of negative autocorrelation, detection capability of the ARMA chart with estimated control limits is similar to the detection capability of it with the known limits even when the number of samples used in the estimation is 200. However, in the presence of positive autocorrelation this is not the case. If there is positive autocorrelation in the process, the performance of the ARMA chart is affected by the number of samples. Furthermore, this effect of number of samples is more when the positive autocorrelation in the process is high (ϕ=0.95).

To illustrate the behaviour of the run length distributions of ARMA chart for different autocorrelation levels and number of samples when h=0.0 and 1.0, figures showing the 2.5 and 97.5 percentiles, mean and median of the run lengths are obtained. Moreover, detailed information about the run length distributions is provided in tables by giving the numerical values of run length statistics.

Figure 4.6 shows the 2.5 and 97.5 percentiles, mean and median of the run lengths for different number of samples and different values of ϕ when there is no shift in the process mean (h=0.0). In addition to Figure 4.6, Table 4.9 is given to see the values of the run length statistics.

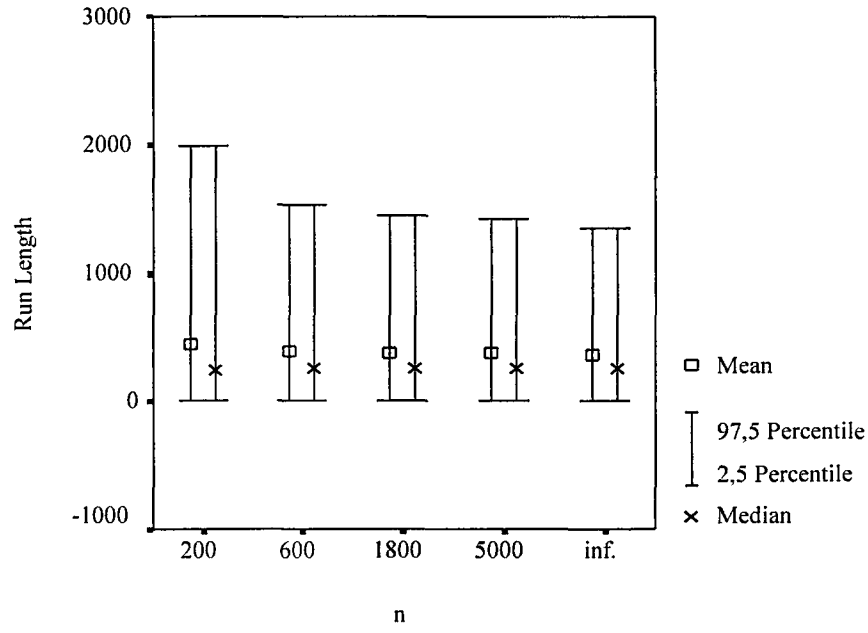


(a) ϕ=0.95

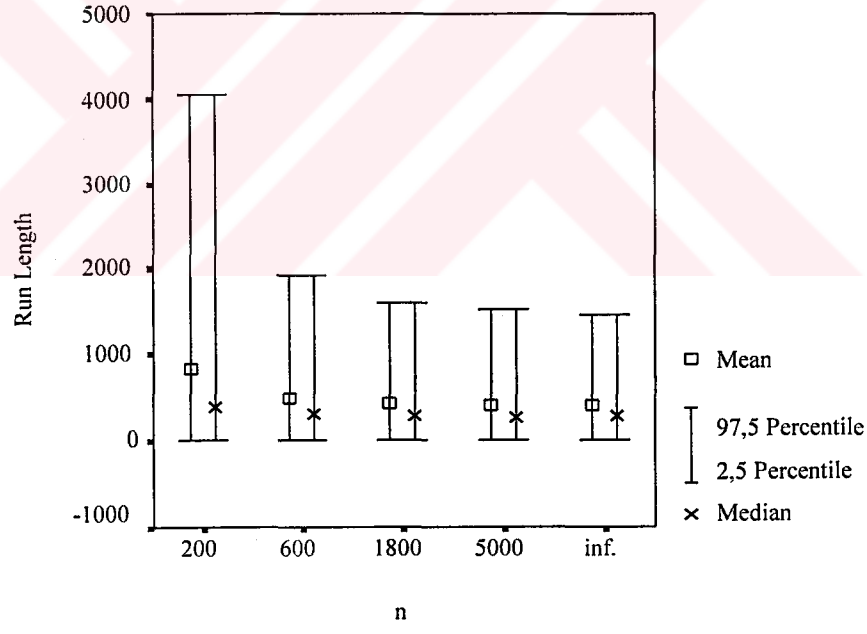


(b) ϕ=0.75

Figure 4.6 Percentiles, Mean and Median of Run Length Distribution for ARMA Chart When h=0.0 and Chart Parameters are Selected for h=1.0.



(c) ϕ=-0.75



(d) ϕ=-0.95

Figure 4.6 (Cont.)

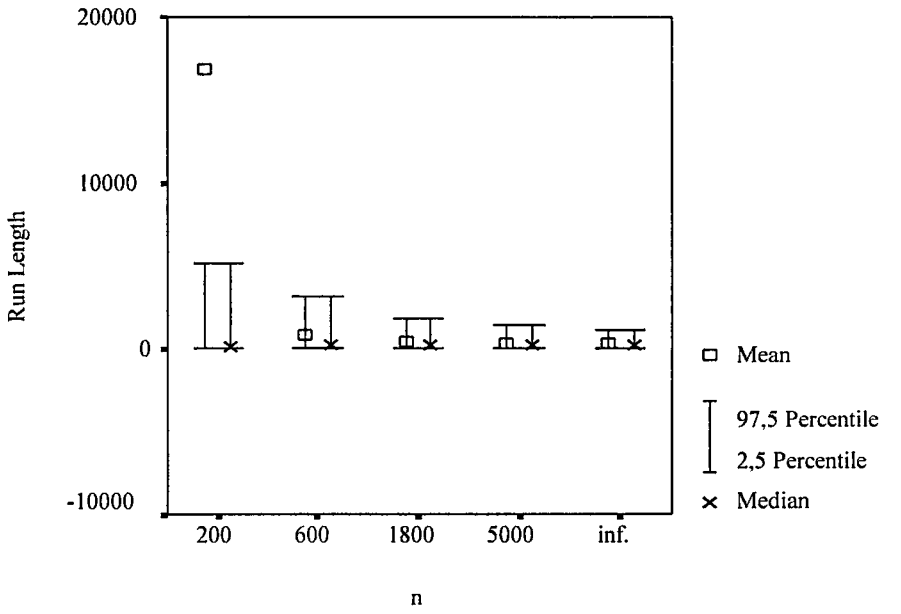
Table 4.9 Statistics for the Run Length Distribution of the ARMA Chart When h=0.0 and Chart Parameters are Selected for h=1.0.

ϕ		n				
		200	**600**	**1800**	**5000**	**∞**
0,95	**Mean**	325022	103907	23091	15724	13963
	Median	888	2495	5177	7532	9569
	Std. Deviation	2656667	1212014	128119	28587	13978
	Skewness	9,97	20,0	33,1	8,47	1,93
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	*	*	6657885	662380	113823
	2.5 Percentile	48,0	85,0	153	241	315
	97.5 Percentile	942640	355379	134825	79778	51550
0,75	**Mean**	1645	810	795	817	807
	Median	302	433	503	555	559
	Std. Deviation	44281	1241	944	858	807
	Skewness	85,7	5,55	3,42	2,29	1,96
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	4144321	23426	13407	9177	7530
	2.5 Percentile	11,0	15,0	16,0	16,0	19,0
	97.5 Percentile	6522	3913	3353	3094	2991
-0,75	**Mean**	444	399	382	382	375
	Median	254	259	259	267	258
	Std. Deviation	591	434	399	383	374
	Skewness	4,27	2,53	2,29	2,00	1,94
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	10066	4923	3658	3423	3482
	2.5 Percentile	9,00	10,0	10,0	11,0	10,0
	97.5 Percentile	1997	1540	1453	1434	1360
-0,95	**Mean**	825	483	420	404	402
	Median	381	309	290	278	282
	Std. Deviation	2329	546	429	412	396
	Skewness	34,1	2,99	2,12	2,12	1,91
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	149048	7597	3554	3981	3571
	2.5 Percentile	13,0	12,0	12,0	10,0	10,0
	97.5 Percentile	4044	1919	1604	1524	1464

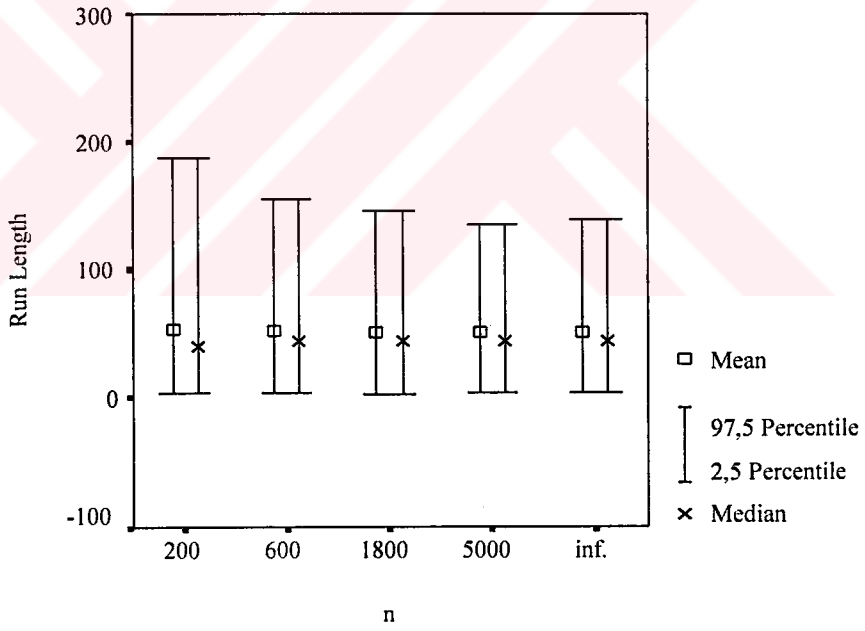
*: Greater than 29900000

As can be seen from Figure 4.6 and Table 4.9, if the number of samples to estimate the control limits gets smaller, then the run length distribution becomes more skewed and variable. Particularly when n=200, the behaviour of the run length distribution is quite different from the run length distribution obtained with the known limits, and this difference is more severe for high positive or negative autocorrelation levels. When the autocorrelation in the process is negative and number of samples used in the estimation is greater than or equal to 600, ARL and MRL does not seem to change much with the number of samples. However, the difference between 2,5 and 97.5 percentiles gets larger as number of samples gets smaller. If there is positive autocorrelation in the process, then SDRL and ARL reach very large values especially when n=200. For example, when ϕ=0.95 and the control limits of the ARMA chart are estimated by 200 samples, 95% of the run length distribution for h=0.0 lies between 48 and 942640, whereas, it lies between 315 and 51550 if the control limits are known and not estimated. From all these explanations it is clear that the behaviour of the run length distribution obtained with the estimated control limits will be different from that with the known limits. However, although having large variation in the distribution is not good, this difference may not be so bad when there is no shift in the process mean. Because, if there is no shift in the process mean, then it is desired to have large run lengths which is the case when number of samples to estimate the control limits is small. Therefore, it may seem to be fine to use small number of samples in the estimation, however using small number of samples may hurt the detection capability of the ARMA chart when a shift occurs in the process mean.

The behaviour of the run length distribution of ARMA chart in the presence of a small shift (h=1.0) in the process mean is illustrated in Figure 4.7 for different number of samples and different autocorrelation levels. Following Figure 4.7, values of the run length statistics are tabulated in Table 4.10.

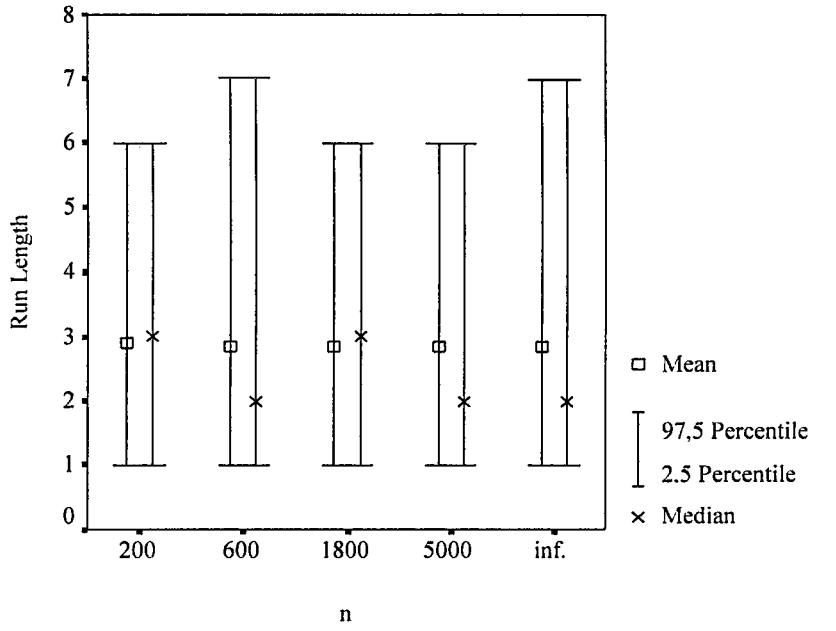


(a) ϕ=0.95

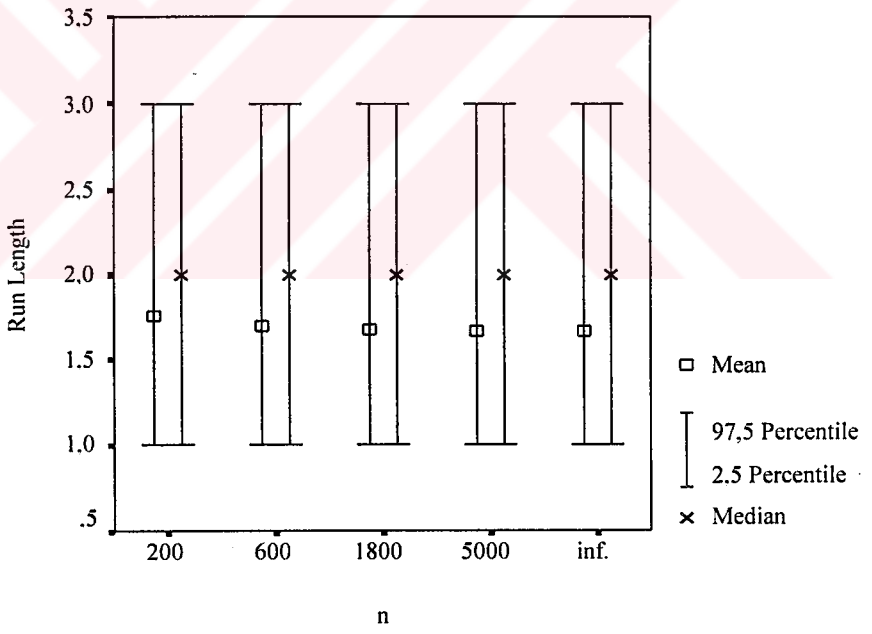


(b) ϕ=0.75

Figure 4.7 Percentiles, Mean and Median of Run Length Distribution for ARMA Chart When h=1.0.



(c) ϕ=-0.75



(d) ϕ=-0.95

Figure 4.7 (Cont.)

Table 4.10 Statistics for the Run Length Distribution of the ARMA Chart When h=1.0.

ϕ		n				
		200	**600**	**1800**	**5000**	**∞**
0,95	**Mean**	16886	921	470	408	376
	Median	160	233	280	296	297
	Std. Deviation	549793	18155	2229	384	288
	Skewness	50,3	68,8	86,6	3,04	1,78
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	*	1501108	212637	5848	2810
	2.5 Percentile	24,0	33,0	40,0	47,0	51,0
	97.5 Percentile	5200	3223	1870	1438	1135
0,75	**Mean**	53,1	51,1	50,5	50,0	50,3
	Median	39,0	43,0	43,0	43,0	44,0
	Std. Deviation	53,5	40,6	37,0	34,9	35,4
	Skewness	3,90	2,07	1,43	1,29	1,35
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	1119	535	319	402	324
	2.5 Percentile	3,00	3,00	2,00	3,00	3,00
	97.5 Percentile	187	155	145	135	139
-0,75	**Mean**	2,89	2,85	2,85	2,83	2,83
	Median	3,00	2,00	3,00	2,00	2,00
	Std. Deviation	1,40	1,34	1,30	1,28	1,32
	Skewness	1,88	1,84	1,64	1,55	1,84
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	16,0	16,0	11,0	13,0	16,0
	2.5 Percentile	1,00	1,00	1,00	1,00	1,00
	97.5 Percentile	6,00	7,00	6,00	6,00	6,98
-0,95	**Mean**	1,76	1,70	1,68	1,67	1,67
	Median	2,00	2,00	2,00	2,00	2,00
	Std. Deviation	0,73	0,63	0,61	0,62	0,60
	Skewness	1,81	0,78	0,58	0,72	0,62
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	13,0	6,00	5,00	5,00	6,00
	2.5 Percentile	1,00	1,00	1,00	1,00	1,00
	97.5 Percentile	3,00	3,00	3,00	3,00	3,00

*: Greater than 29900000

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

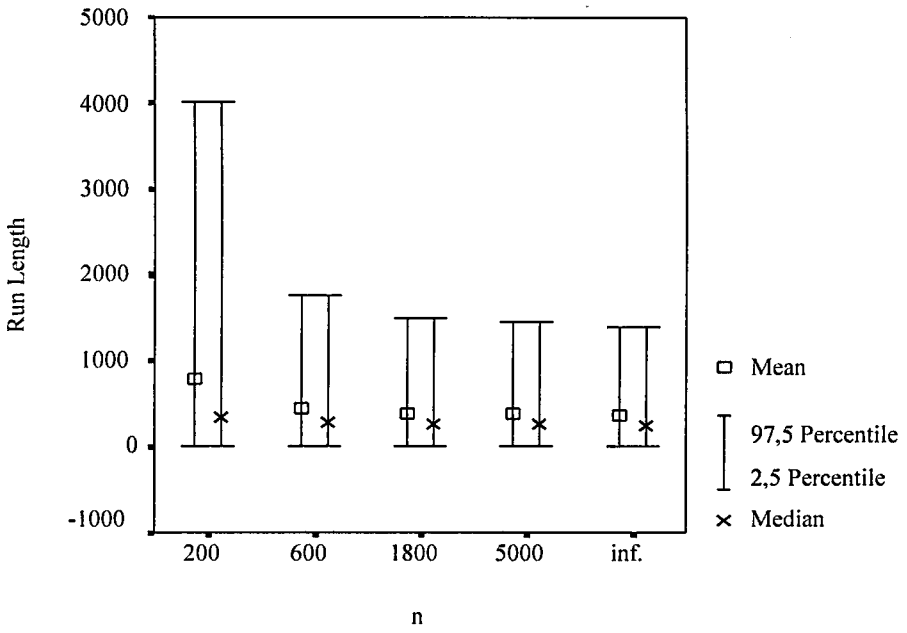
According to Figure 4.7 and Table 4.10, ARL and MRL does not seem to change much with the number of samples unless there is high positive autocorrelation in the process. When ϕ equals 0.95 and number of samples used in the estimation is 200, observed ARL as well as the standard deviation of the run lengths are very high and this is not good in the presence of a shift in the process mean. In fact, if there is high positive autocorrelation in the process, then even using 600 samples in the estimation of the control limits may not be sufficient to have similar performance of the ARMA chart when the control limits are known. However, for lower level of positive autocorrelation (ϕ=0.75), using 200 samples may be enough, but surely larger than 200 would give a better performance. When the autocorrelation in the process is negative, number of samples used in the estimation does not affect the performance of the ARMA chart. For example, if negative autocorrelation in the process is high (ϕ=-0.95), then 97.5% of the run lengths will be less than or equal to 3 whatever the number of samples is.

Large shifts in the process mean may be of concern instead of small shifts. If this is the case, ARMA chart parameters may be adjusted to detect the large shifts (h=3.0) efficiently. ARMA chart has been constructed after changing the chart parameters for detecting large shifts and the effect of number of samples on the performance of the chart has been studied. Table 4.11 gives ARL, SDRL, MINRL and MAXRL for different values of ϕ and for different number of samples when h=0.0 and 3.0. Charting parameters in Table 4.11 are taken from Table 4.7.

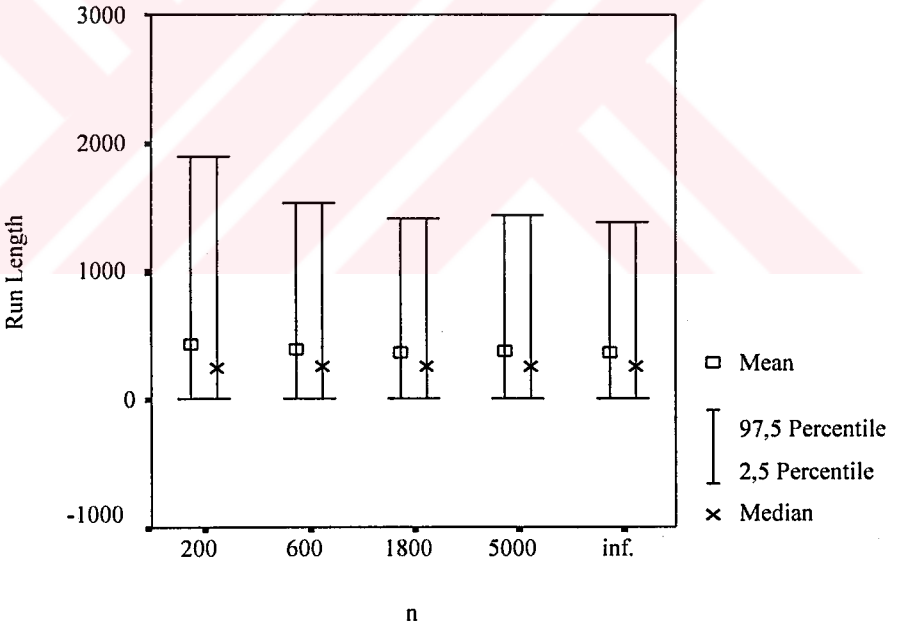
Table 4.11 Performance of the ARMA Chart with Control Limits Estimated from n Samples when h=0.0 and 3.0.

Process Parameter	Charting Parameters			Number of Samples	h							
					0,0				3,0			
ϕ	ϕ_1	0	$\beta=0/\theta_0$	n	MIN.RL	ARL	MAX.RL	SDRL	MIN.RL	ARL	MAX.RL	SDRL
0,95	0,00	19,0	0,95	200	1,00	783	80604	1862	1,00	1,00	1,00	0,00
				600	1,00	450	7965	486	1,00	1,00	1,00	0,00
				1000	1,00	414	4720	436	1,00	1,00	1,00	0,00
				1800	1,00	392	4585	410	1,00	1,00	1,00	0,00
				2600	1,00	394	4055	404	1,00	1,00	1,00	0,00
				3400	1,00	380	3511	390	1,00	1,00	1,00	0,00
				5000	1,00	385	3978	392	1,00	1,00	1,00	0,00
				7000	1,00	381	3523	389	1,00	1,00	1,00	0,00
				∞	1,00	372	3410	371	1,00	1,00	1,00	0,00
0,75	0,00	3,00	0,75	200	1,00	438	10590	564	1,00	3,49	239	13,9
				600	1,00	392	5711	421	1,00	3,17	303	12,1
				1000	1,00	380	4056	395	1,00	3,20	262	12,1
				1800	1,00	374	4180	389	1,00	3,12	278	12,4
				2600	1,00	383	3474	392	1,00	3,12	168	11,2
				3400	1,00	372	3511	382	1,00	2,96	268	10,8
				5000	1,00	379	3978	386	1,00	3,09	251	11,9
				7000	1,00	378	3523	388	1,00	2,98	182	10,8
				∞	1,00	372	3410	371	1,00	3,02	227	11,3
-0,75	0,00	-0,43	-0,75	200	1,00	438	10590	566	1,00	1,06	2,00	0,24
				600	1,00	392	5711	422	1,00	1,07	2,00	0,25
				1000	1,00	380	4619	398	1,00	1,06	2,00	0,24
				1800	1,00	375	4180	390	1,00	1,06	2,00	0,24
				2600	1,00	382	3440	391	1,00	1,06	2,00	0,24
				3400	1,00	372	3511	381	1,00	1,06	2,00	0,24
				5000	1,00	379	3978	386	1,00	1,06	2,00	0,24
				7000	1,00	378	3523	386	1,00	1,06	2,00	0,24
				∞	1,00	372	3410	371	1,00	1,06	2,00	0,24
-0,95	0,00	-0,49	-0,95	200	1,00	776	73495	1812	1,00	1,00	1,00	0,00
				600	1,00	454	9665	506	1,00	1,00	1,00	0,00
				1000	1,00	414	5311	437	1,00	1,00	1,00	0,00
				1800	1,00	392	4180	410	1,00	1,00	1,00	0,00
				2600	1,00	392	3729	401	1,00	1,00	1,00	0,00
				3400	1,00	381	3511	394	1,00	1,00	1,00	0,00
				5000	1,00	385	3978	391	1,00	1,00	1,00	0,00
				7000	1,00	382	3523	390	1,00	1,00	1,00	0,00
				∞	1,00	372	3410	371	1,00	1,00	1,00	0,00

As can be seen from Table 4.11 the performance of the ARMA chart for detecting a large shift (h=3.0) is not affected by the number of samples. When h equals 3.0, ARL and SDRL do not change with the number of samples used in the estimation of the control limits. When the autocorrelation in the process is high in negative or positive direction (ϕ=-0.95 or ϕ=0.95), the shift in the mean is detected in the first sample whatever the number of samples is. For every autocorrelation level, in-control ARL increases and it becomes more probable to observe large run lengths as number of samples decreases. When there is no shift in the process mean, it is preferred to have large run lengths since it is not desired to experience false alarms. However, variation in the run length distribution also increases as the number of samples decreases. Figure 4.8 illustrates the behaviour of the run length distribution for different number of samples and autocorrelation levels when there is no shift in the process mean. Moreover, corresponding run length statistics are tabulated in Table 4.12.

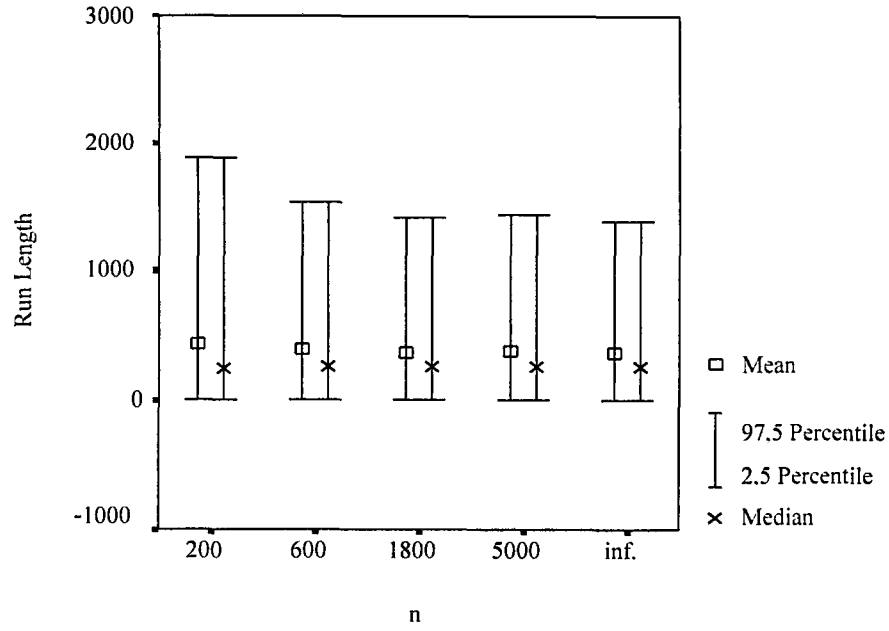


(a) ϕ=0.95

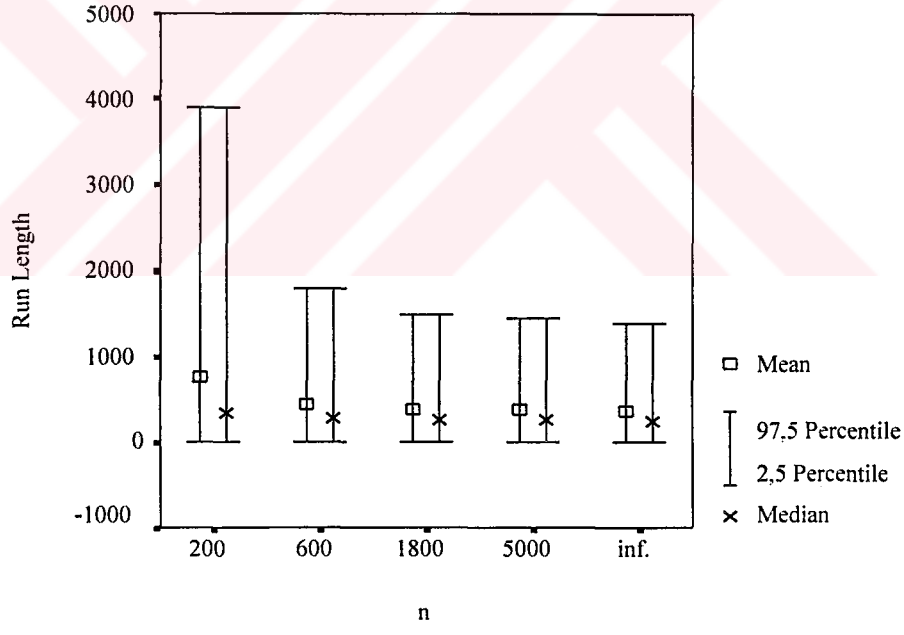


(b) ϕ=0.75

Figure 4.8 Percentiles, Mean and Median of Run Length Distribution for ARMA Chart When h=0.0 and Chart Parameters are Selected for h=3.0.



(c) ϕ=-0.75



(d) ϕ=-0.95

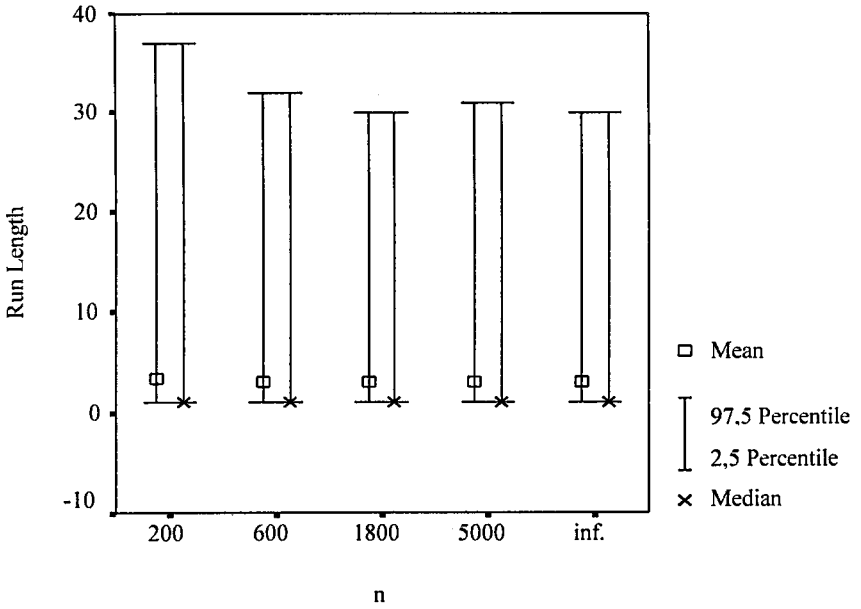
Figure 4.8 (Cont.)

Table 4.12 Statistics for the Run Length Distribution of the ARMA Chart When h=0.0 and Chart Parameters are Selected for h=3.0.

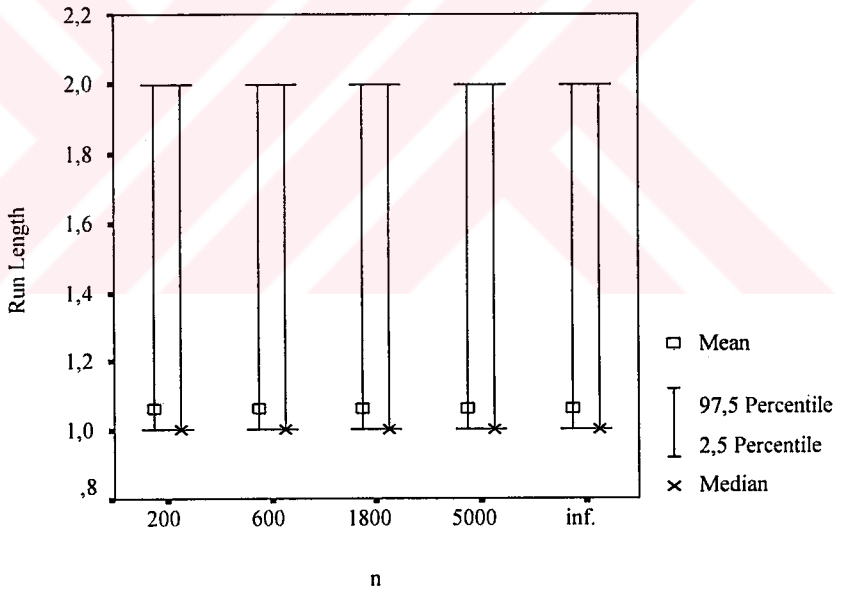
ϕ		n				
		200	**600**	**1800**	**5000**	**∞**
0,95	**Mean**	783	450	392	385	372
	Median	359	294	266	270	257
	Std. Deviation	1862	486	410	392	371
	Skewness	19,1	2,61	2,33	2,13	1,96
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	80604	7965	4585	3978	3410
	2.5 Percentile	12,0	11,0	10,0	10,0	11,0
	97.5 Percentile	4007	1764	1496	1461	1396
0,75	**Mean**	438	392	374	379	372
	Median	254	257	256	267	257
	Std. Deviation	564	421	389	386	371
	Skewness	3,98	2,49	2,28	2,16	1,96
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	10590	5711	4180	3978	3410
	2.5 Percentile	9,00	10,0	10,0	10,0	11,0
	97.5 Percentile	1899	1532	1410	1437	1396
-0,75	**Mean**	438	392	375	379	372
	Median	255	256	256	267	257
	Std. Deviation	566	422	390	386	372
	Skewness	4,08	2,49	2,28	2,16	1,96
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	10590	5711	4180	3978	3410
	2.5 Percentile	9,00	9,00	10,0	10,0	11,0
	97.5 Percentile	1891	1532	1416	1439	1396
-0,95	**Mean**	776	454	392	385	372
	Median	359	292	266	271	257
	Std. Deviation	1812	506	410	391	371
	Skewness	19,7	3,01	2,29	2,13	1,96
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	73495	9665	4180	3978	3410
	2.5 Percentile	12,0	11,0	10,0	10,0	11,0
	97.5 Percentile	3900	1803	1497	1458	1396

Figure 4.8 and Table 4.12 indicate that, run length distribution when ϕ=0.95 behaves similar to the run length distribution when ϕ=-0.95. In both cases the spread of the distribution is large and the distribution is much more skewed when number of samples used in the estimation is 200. For all values of ϕ, if control limits are estimated by 600 or more samples, then the performance of the ARMA chart with these estimated limits will be similar to that with the known control limits.

To observe the ARMA chart performance in the presence of a large shift in the process mean Figure 4.9 and Table 4.13 are given. Note from Table 4.11 that, when there is a large shift in the process mean and the autocorrelation in the process is high in negative or positive direction (ϕ=-0.95, ϕ=0.95), all of the run lengths are obtained as 1. That is, the shift is always detected in the first sample. This behaviour in the run length distribution does not change with the number of samples. Therefore, in Figure 4.9 and Table 4.13, percentiles, median and mean of the run lengths are obtained for different number of samples only for ϕ=0.75 and -0.75.



(a) ϕ=0.75



(b) ϕ=-0.75

Figure 4.9 Percentiles, Mean and Median of Run Length Distribution for ARMA Chart When h=3.0.

Table 4.13 Statistics for the Run Length Distribution of the ARMA Chart When h=3.0.

ϕ		n				
		200	**600**	**1800**	**5000**	**∞**
0,75	**Mean**	3,49	3,17	3,12	3,09	3,02
	Median	1,00	1,00	1,00	1,00	1,00
	Std. Deviation	13,9	12,1	12,4	11,9	11,3
	Skewness	7,88	8,92	9,96	8,60	8,45
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	239	303	278	251	227
	2.5 Percentile	1,00	1,00	1,00	1,00	1,00
	97.5 Percentile	37,0	32,0	30,0	31,0	30,0
-0,75	**Mean**	1,06	1,07	1,06	1,06	1,06
	Median	1,00	1,00	1,00	1,00	1,00
	Std. Deviation	0,24	0,25	0,24	0,24	0,24
	Skewness	3,61	3,52	3,59	3,62	3,62
	Minimum	1,00	1,00	1,00	1,00	1,00
	Maximum	2,00	2,00	2,00	2,00	2,00
	2.5 Percentile	1,00	1,00	1,00	1,00	1,00
	97.5 Percentile	2,00	2,00	2,00	2,00	2,00

According to Figure 4.9 and Table 4.13, detection capability of the ARMA chart is better when ϕ equals -0.75. Because, run lengths are smaller. When ϕ equals -0.75, a shift of size $3\sigma_y$ is detected in the first or at most second sample all the time and same behaviour is obtained even when number of samples used in the estimation is equal to 200. When ϕ equals 0.75, median of the run lengths does not change with the number of samples and very small increase in the mean of the run lengths is observed as number of samples decreases. 97.5% of the run lengths are less than about 30-32 unless number of samples is small (n=200). When n equals 200, 97.5 percentile of the run lengths is equal to 37. In general, it can be concluded that, if there is a large shift (h=3.0) in the process mean, then the detection capability of the ARMA chart is not affected much by the number of samples used in the estimation of the control limits.

So far, the performances of EWMAST chart and ARMA chart have been examined and the effect of number of samples on the performance of these charts has been discussed. Note that, EWMAST chart is a special case of ARMA chart and this is shown in Chapter 3. According to Jiang et.al (2000), it is possible to derive an ARMA chart with appropriate parameter values (ϕ_1 and θ) that outperforms the EWMAST chart. In this study, it is possible to observe this claim. Because, results of the simulation study indicate that the performance of the ARMA chart in detecting the shifts is better than the performance of the EWMAST chart, in general. This can be concluded simply by comparing Tables 4.8 and 4.11 with Table 4.3. However, designing an ARMA chart is not easy. ARMA statistic has two parameters to be selected appropriately and the algorithm given by Jiang et.al. (2000) for the selection of the appropriate charting parameters is difficult to implement. In this study, a table of the appropriate charting parameters has been provided. However, this table is limited to a certain set of process parameter values and it is valid for only AR(1) processes. Therefore, in practical use ARMA chart may not be preferred because of the difficulties in choosing the appropriate charting parameters.

CHAPTER 5

CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK

5.1 Conclusions

Traditional control charts assume that the underlying in-control process observations are uncorrelated. However, in some processes it is possible to observe autocorrelation inherently. Traditional control charts may not perform well in the presence of autocorrelation. Therefore, some other control charts, which account for autocorrelation, have been proposed in recent years. Two of these charts are EWMAST chart and ARMA chart.

In this study, the performances of the EWMAST and ARMA charts have been studied when the control limits are estimated and when the underlying process is best described by a stationary AR(1) model with normally distributed errors. The parameter of the EWMA statistic, λ, is taken to be 0.2 as suggested by Zhang (1998). ARMA chart parameters, ϕ_1 and θ, are selected on the basis of an algorithm given by Jiang et.al.(2000) and with some small modifications in that algorithm a table giving the appropriate charting parameters for several process parameter values of an AR(1) process is provided.

It is observed that the variability in the run length distributions of the control charts is high and the run length distribution is positively skewed. Therefore, using only average run length (ARL) as a performance measure is

not found to be appropriate. In addition to ARL, statistics like standard deviation of the run length distribution (SDRL), minimum and maximum run lengths, median run length, 2.5 and 97.5 percentiles of the run length distribution and skewness are also specified by simulation studies for the purpose of understanding the behaviour of the run length distribution. For each control chart, these statistics are found by simulation when different numbers of samples are used for estimating the control limits. By this way, the effect of number of samples on the behaviour of the run length distribution, thus, on the performance of the charts is understood more explicitly.

Results of the simulation studies indicate that both EWMAST and ARMA charts have better performance in detecting the shifts when the autocorrelation in the process is negative. Because, smaller run lengths are obtained in the presence of negative autocorrelation. Moreover, if the autocorrelation in the process is negative, then using 200 samples with sample size 1 to estimate the control limits will be sufficient for both control charts in order to have similar performance of the charts with the known control limits.

When the autocorrelation in the process is positive and high (ϕ=0.95), and shift to be detected is small (h=1.0), it is suggested to use at least 1800 samples to estimate the control limits of the EWMAST chart, so that the estimated control limits perform like the known limits. This number can be reduced to 600 if the positive autocorrelation is lower (ϕ=0.75). When ARMA chart is employed for process monitoring in the same situation, it also requires at least 1800 samples for the estimation of the control limits in order to have similar performance of the ARMA chart with the known control limits. However, for lower level of positive autocorrelation (ϕ=0.75), using 200 samples may be enough, but surely larger than 200 would give a better performance.

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

If shift to be detected is large (h=3.0) and the monitoring process is high positively autocorrelated, then it is suggested to use at least 600 samples in the estimation of the control limits of the EWMAST chart. However, if the process is not high positively autocorrelated, then using 200 samples to estimate the control limits will be adequate in order to have a similar performance of the EWMAST chart with the known control limits.

If ARMA chart is employed for monitoring an autocorrelated process and the autocorrelation in the process is high in negative or positive direction (ϕ=-0.95, ϕ=0.95), then large shifts are always detected in the first sample after shift occurs whatever the number of samples used in the estimation is.

In general, it can be concluded that both charts require larger number of samples to estimate the control limits if the shift to be detected is small (h=1.0). If sufficient number of samples is not used in the estimation of the control limits, then the shifts in the process mean may not be detected on time.

As mentioned previously, EWMAST chart is a special case of ARMA chart and it is possible to derive an ARMA chart with appropriate parameter values (ϕ_1 and θ) that outperforms the EWMAST chart. Results of the simulation study also indicate that the performance of the ARMA chart in detecting the shifts is better than the performance of the EWMAST chart, in general. However, designing an ARMA chart is not easy. ARMA statistic has two parameters to be selected appropriately and the algorithm given by Jiang et.al. (2000) for the selection of the appropriate charting parameters is difficult to implement. In this study, a table of the appropriate charting parameters has been provided. However, this table is limited to a certain set of process parameter values and it is valid for only AR(1) processes. Therefore, in practical use ARMA chart may not be preferred because of the difficulties in choosing the appropriate charting parameters.

5.2 Suggestions for Future Work

It is believed that some further research is needed in the area of autocorrelated processes.

In this study, only step shift in the mean is considered as an out-of-control condition. This can be extended to gradual shift in the mean and shift in the process variation.

The performances of the univariate control charts for monitoring an autocorrelated process can be studied for different estimators of variance, such as mean squared successive differences. Instead of using sample standard deviation for establishing the control limits, some other estimator of the standard deviation that perform better than the sample standard deviation can be searched for.

If the number of preliminary observations for estimating the control limits is too small, then a short run problem exists. In addition to the autocorrelation in the process if there exists a short run problem, then some different process monitoring techniques may be available and they can be studied.

Univariate control charts are used for monitoring only one quality characteristic at a time. However, sometimes several quality characteristics explain the overall quality of a product or a process together. In such cases, multivariate monitoring is required. Moreover, there may exist an autocorrelation over time for the quality characteristics. Therefore, multivariate monitoring of several autocorrelated quality characteristics can be studied.

This study includes some discussions about the design of an ARMA chart. It is understood that design procedure has a big space for further research. It is not practical or easy to implement. Hence, ARMA chart can further be improved for better use in practice.

REFERENCES

Alwan, L. C., and Radson, D. (1992), "Time-Series Investigation of Subsample Mean Charts", IIE Transactions, Vol.24, No.5, pp 66-80.

Alwan, L. C., and Roberts, H. V. (1988), "Time-Series Modelling for Statistical Process Control", Journal of Business & Economic Statistics, Vol.6, No.1, pp 87-95.

Crowder, S. V. (1989), "Design of Exponentially Weighted Moving Average Schemes", Journal of Quality Technology, Vol.21, No.3, pp 155-162.

Cryer, J. D. (1986), Time Series Analysis, PWS-KENT, Boston.

Enders, W. (1995), Applied Econometric Time Series, John Wiley & Sons, New York.

Faltin, F. W., Mastrangelo C. M., Runger, G. C., and Ryan, T. P. (1997), "Considerations in the Monitoring of Autocorrelated and Independent Data", Journal of Quality Technology, Vol.29, No.2, pp 131-133.

Grant, E. L., and Leavenworth, R. S. (1988), Statistical Quality Control, 6th Edition, McGraw-Hill, New York.

Griffith, G. K. (2000), The Quality Technician's Handbook, 4th Edition, Prentice-Hall, New Jersey.

Hamilton, J. D. (1994), Time Series Analysis, Princeton University Press, New Jersey.

Harris, T. J., and Ross, W. H. (1991), "Statistical Process Control Procedures for Correlated Observations", The Canadian Journal of Chemical Engineering, Vol.69, pp 48-57.

Hunter, J. S. (1986), "The Exponentially Weighted Moving Average", Journal of Quality Technology, Vol.18, No.4, pp 203-210.

Jiang, W. (2001), "Average Run Length Computation of ARMA Charts for Stationary Processes", Communications in Statistics: Simulation and Computation, Vol.30, No.3, pp 699-716.

Jiang, W., Tsui, K. L. (2001), "Some Properties of the ARMA Control Chart", Nonlinear Analysis, Vol.47, pp 2073-2088.

Jiang, W., Tsui, K. L., and Woodall W. H. (2000), "A New SPC Monitoring Method: The ARMA Chart", Technometrics, Vol.42, No.4, pp 399-410.

Kantar, B. (2002), "The Effect of Sample Size on the Performance of Univariate Control Charts for Autocorrelated Data", M.Sc. Thesis, Department of Statistics, METU.

Kramer, H., and Schmid, W. (2000), "The Influence of Parameter Estimation on the ARL of Shewhart Type Charts for Time Series", Statistical Papers, Vol.41, pp 173-196.

Ledolter, J., and Burrill, C. W. (1999), Statistical Quality Control Strategies and Tools for Continual Improvement, John Wiley & Sons, New York.

Lowry, C. A., and Montgomery, D. C. (1995), "A Review of Multivariate Control Charts", IIE Transactions, Vol.27, pp 800-810.

Lucas, J. M., and Saccucci M. S. (1990), "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements", Technometrics, Vol.32, No.1, pp 1-12.

Montgomery, D. C. (2001), Introduction to Statistical Quality Control, 4th Edition, John Wiley & Sons, New York.

Montgomery, D. C., and Mastrangelo, C. M. (1991), "Some Statistical Process Control Methods for Autocorrelated Data", Journal of Quality Technology, Vol.23, No.3, pp 179-193.

Quesenberry, C. P. (1993), "The Effect of Sample Size on Estimated Limits for $\overline{X}$ and X Control Charts", Journal of Quality Technology, Vol.25, No.4, pp 237-247.

Reynolds, M. R. Jr., and Lu, C. W. (1997), "Control Charts for Monitoring Processes with Autocorrelated Data", Nonlinear Analysis, Theory, Methods & Applications, Vol.30, No.7, pp 4059-4067.

Runger, G. C., and Willemain, T. R. (1995), "Model-Based and Model-Free Control of Autocorrelated Processes", Journal of Quality Technology, Vol.27, No.4, pp 283-292.

Runger, G. C., and Willemain, T. R. (1996), "Batch-Means Control Charts for Autocorrelated Data", IIE Transactions, Vol.28, pp 483-487.

Ryan, T. P. (1989), Statistical Methods for Quality Improvement, John Wiley & Sons, New York.

Schmid, W. (1995), "On the Run Length of a Shewhart Chart for Correlated Data", Statistical Papers, Vol.36, pp 111-130.

Thomas, D. W., Barrett, C. S., Blankenstein, E. E., Chapman, A. T., Cole, C. C., Drummond, F. H., LeSage, C. W., Ordecki, A. S., Schin, R., Small, B. B., and Stonehill, F. (1984), Statistical Quality Control Handbook, AT&T Technologies, Indiana.

Vasilopoulos, A. V., and Stamboulis, A. P. (1978), "Modification of Control Chart Limits in the Presence of Data Correlation", Journal of Quality Technology, Vol.10, No.1, pp 20-30.

Wardell, D. G., Moskowitz, H., and Plante, R. D. (1992), "Control Charts in the Presence of Data Correlation", Management Science, Vol.38, No.8, pp 1084-1105.

Wardell, D. G., Moskowitz, H., and Plante, R. D. (1994), "Run-Length Distributions of Special-Cause Control Charts for Correlated Processes", Technometrics, Vol.36, No.1, pp 3-17.

Wheeler, D. J. (1995), Advanced Topics in SPC, SPC Press.

Woodall, W. H., and Montgomery, D. C. (1999), "Research Issues and Ideas in Statistical Process Control", Journal of Quality Technology, Vol.31, No.4, pp 376-386.

Yashchin, E. (1993), "Performance of CUSUM Control Schemes for Serially Correlated Observations", Technometrics, Vol.35, No.1, pp 37-52.

Zhang, N. F. (1997), "Detection Capability of Residual Control Chart for Stationary Process Data", Technometrics, Vol.24, Iss.4, pp 475-492.

Zhang, N. F. (1998), "A Statistical Control Chart for Stationary Process Data", Technometrics, Vol.40, No.1, pp 24-38.

Zhang, N. F. (2000), "Statistical Control Charts for Monitoring the Mean of a Stationary Process", Journal of Statistical Computation and Simulation, Vol.66, pp 249-258.

APPENDIX A

DERIVATION OF THE VARIANCE OF EWMA STATISTIC WHEN THE PROCESS DATA IS INDEPENDENT

In this appendix, the variance of the EWMA statistic given in (3.1) is derived when the process observations are independent.

Suppose $\{Y_t\}$ is the series of observations with mean=μ and variance=σ_y^2, then the EWMA of Y_t is defined as follows:

$Z_t = (1-\lambda)Z_{t-1} + \lambda Y_t$, $t=1, 2, \ldots$

where $Z_0=\mu$ and $\lambda(0<\lambda\leq 1)$ is a constant.

The EWMA of Y_t can be rewritten as follows by replacing Z_t recursively.

$$Z_t = \lambda\sum_{j=0}^{t-1}(1-\lambda)^j Y_{t-j} + (1-\lambda)^t Z_0$$

Considering that Z_0 is not random,

$$\sigma_z^2 = \text{var}(Z_t) = \lambda^2\sum_{j=0}^{t-1}\text{var}\left[(1-\lambda)^j Y_{t-j}\right] = \lambda^2\sigma_y^2\sum_{j=0}^{t-1}(1-\lambda)^{2j}$$

$$= \lambda^2 \sigma_y^2 \frac{1-(1-\lambda)^{2t}}{1-(1-\lambda)^2} = \left[\frac{\lambda}{2-\lambda}\right]\left[1-(1-\lambda)^{2t}\right]\sigma_y^2$$

$$\sigma_z^2 = \text{var}(Z_t) = \left[\frac{\lambda}{2-\lambda}\right]\left[1-(1-\lambda)^{2t}\right]\sigma_y^2$$

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

APPENDIX B

DERIVATION OF THE AUTOCOVARIANCE FUNCTION AND THE VARIANCE OF EWMA STATISTIC

In this appendix, the autocovariance function and the variance of the EWMA statistic defined in (3.3) and (3.4), respectively, are derived when the underlying process, Y_t, is a stationary process with mean=μ, variance=σ_y^2, autocovariance function γ_k=cov(Y_t,Y_{t+k}) and autocorrelation function $\rho_k=\gamma_k/\gamma_0$ ($\gamma_0=\sigma_y^2$) at lag k.

1. Derivation of the autocovariance function of EWMA statistic

The EWMA of Y_t, which is obtained by recursively replacing Z_t, is given in section 3.1.

$$Z_t = \lambda\sum_{j=0}^{t-1}(1-\lambda)^j Y_{t-j} + (1-\lambda)^t Z_0$$

Considering that Z_0 is not random,

$$\mathrm{Cov}(Z_t, Z_{t+\tau}) = \mathrm{Cov}\left[\lambda\sum_{j=0}^{t-1}(1-\lambda)^j Y_{t-j}, \lambda\sum_{i=0}^{t+\tau-1}(1-\lambda)^i Y_{t+\tau-i}\right]$$

$$= \mathrm{Cov}\Big[\lambda\left(Y_t + (1-\lambda)Y_{t-1} + \ldots + (1-\lambda)^{t-1}Y_1\right),$$
$$\lambda\left(Y_{t+\tau} + (1-\lambda)Y_{t+\tau-1} + \ldots + (1-\lambda)^{t+\tau-1}Y_1\right)\Big]$$

$$\begin{aligned}
&= \lambda^2\gamma_\tau\left[1 + (1-\lambda)^2 + (1-\lambda)^4 + \ldots + (1-\lambda)^{2t-2}\right] \\
&+ \lambda^2\gamma_{\tau+1}\left[(1-\lambda) + (1-\lambda)^3 + \ldots + (1-\lambda)^{2t-3}\right] \\
&+ \lambda^2\gamma_{\tau+2}\left[(1-\lambda)^2 + (1-\lambda)^4 + \ldots + (1-\lambda)^{2t-4}\right] \\
&\vdots \\
&+ \lambda^2\gamma_{\tau+t-1}\left[(1-\lambda)^{t-1}\right] \\
&+ \lambda^2\gamma_{\tau-1}\left[(1-\lambda) + (1-\lambda)^3 + \ldots + (1-\lambda)^{2t-1}\right] \\
&+ \lambda^2\gamma_{\tau-2}\left[(1-\lambda)^2 + (1-\lambda)^4 + \ldots + (1-\lambda)^{2t}\right] \\
&\vdots \\
&+ \lambda^2\gamma_0\left[(1-\lambda)^\tau + (1-\lambda)^{\tau+2} + \ldots + (1-\lambda)^{2t+\tau-2}\right] \\
&+ \lambda^2\gamma_{-1}\left[(1-\lambda)^{\tau+1} + (1-\lambda)^{\tau+3} + \ldots + (1-\lambda)^{2t+\tau-3}\right] \\
&+ \lambda^2\gamma_{1-t}\left[(1-\lambda)^{t+\tau-1}\right]
\end{aligned}$$

$$\begin{aligned}
= \lambda^2\Bigg\{&\sum_{k=0}^{t-1}\gamma_{\tau+k}(1-\lambda)^k\sum_{i=0}^{t-k-1}(1-\lambda)^{2i} + \sum_{j=1}^{\tau}\gamma_{\tau-j}(1-\lambda)^j\left[\frac{1-(1-\lambda)^{2t}}{1-(1-\lambda)^2}\right] \\
&+ \sum_{j=\tau+1}^{t+\tau-1}\gamma_{\tau-j}(1-\lambda)^j\sum_{l=0}^{t+\tau-1-j}(1-\lambda)^{2l}\Bigg\}
\end{aligned}$$

$$=\lambda^2\left\{\sum_{k=0}^{t-1}\gamma_{\tau+k}(1-\lambda)^k\left[\frac{1-(1-\lambda)^{2(t-k)}}{1-(1-\lambda)^2}\right]\right.$$

$$+\frac{1-(1-\lambda)^{2t}}{1-(1-\lambda)^2}\sum_{j=1}^{\tau}\gamma_{\tau-j}(1-\lambda)^j$$

$$\left.+\sum_{j=\tau+1}^{t+\tau-1}\gamma_{\tau-j}(1-\lambda)^j\left[\frac{1-(1-\lambda)^{2(t+\tau-j)}}{1-(1-\lambda)^2}\right]\right\}$$

Remember that $\gamma_k=\gamma_0\rho_k=\sigma_y^2\rho_k$. Therefore,

$$\mathrm{Cov}(Z_t, Z_{t+\tau})=\lambda^2\left\{\sum_{k=0}^{t-1}\sigma_y^2\rho_{\tau+k}(1-\lambda)^k\left[\frac{1-(1-\lambda)^{2(t-k)}}{1-(1-\lambda)^2}\right]\right.$$

$$+\frac{1-(1-\lambda)^{2t}}{1-(1-\lambda)^2}\sum_{j=1}^{\tau}\sigma_y^2\rho_{\tau-j}(1-\lambda)^j$$

$$\left.+\sum_{j=\tau+1}^{t+\tau-1}\sigma_y^2\rho_{\tau-j}(1-\lambda)^j\left[\frac{1-(1-\lambda)^{2(t+\tau-j)}}{1-(1-\lambda)^2}\right]\right\}$$

$$=\frac{\lambda}{2-\lambda}\sigma_y^2\left\{\sum_{k=0}^{t-1}\rho_{\tau+k}(1-\lambda)^k\left[1-(1-\lambda)^{2(t-k)}\right]\right.$$

$$+\sum_{j=1}^{\tau}\rho_{\tau-j}(1-\lambda)^j\left[1-(1-\lambda)^{2t}\right]$$

$$\left.+\sum_{j=\tau+1}^{t+\tau-1}\rho_{\tau-j}(1-\lambda)^j\left[1-(1-\lambda)^{2(t+\tau-j)}\right]\right\}$$

$$\mathrm{Cov}(Z_t, Z_{t+\tau}) = \frac{\lambda}{2-\lambda}\sigma_y^2\left\{\sum_{k=0}^{t-1}\rho_{k+\tau}(1-\lambda)^k\left[1-(1-\lambda)^{2(t-k)}\right]\right.$$

$$+\sum_{k=1}^{\tau}\rho_{\tau-k}(1-\lambda)^k\left[1-(1-\lambda)^{2t}\right]$$

$$\left.+\sum_{k=\tau+1}^{t-1+\tau}\rho_{\tau-k}(1-\lambda)^k\left[1-(1-\lambda)^{2(t+\tau-k)}\right]\right\}$$

2. Derivation of the variance of EWMA statistic

Covariance function of EWMA statistic can be rewritten as follows:

$$\mathrm{Cov}(Z_t, Z_{t+\tau}) = \mathrm{Cov}\left[\lambda\sum_{j=0}^{t-1}(1-\lambda)^j Y_{t-j}, \lambda\sum_{i=0}^{t+\tau-1}(1-\lambda)^i Y_{t+\tau-i}\right]$$

$$= \lambda^2\sum_{j=0}^{t-1}\sum_{i=0}^{t+\tau-1}(1-\lambda)^{i+j}\gamma_{j-i+\tau}$$

When $\tau = 0$ in covariance function above, the variance of EWMA statistic is obtained. That is, $\tau = 0 \Rightarrow \mathrm{Cov}(Z_t, Z_{t+\tau}) = \mathrm{Cov}(Z_t, Z_t) = \mathrm{Var}(Z_t) = \sigma_z^2$

$$\mathrm{Var}(Z_t) = \sigma_z^2 = \lambda^2\sum_{j=0}^{t-1}\sum_{i=0}^{t-1}(1-\lambda)^{i+j}\gamma_{j-i} = \lambda^2\sum_{j=0}^{t-1}\sum_{i=0}^{t-1}(1-\lambda)^{i+j}\sigma_y^2\rho_{j-i}$$

$$= \lambda^2\sigma_y^2\sum_{j=0}^{t-1}\sum_{i=0}^{t-1}(1-\lambda)^{i+j}\rho_{j-i}$$

$$= \lambda^2\sigma_y^2\Big[\Big(1+(1-\lambda)^2+(1-\lambda)^4+\ldots+(1-\lambda)^{2t-2}\Big)$$
$$+2\rho_1\Big((1-\lambda)+(1-\lambda)^3+(1-\lambda)^5\ldots+(1-\lambda)^{2t-3}\Big)$$
$$+2\rho_2\Big((1-\lambda)^2+(1-\lambda)^4+\ldots+(1-\lambda)^{2t-4}\Big)+\cdots+2\rho_{t-1}(1-\lambda)^{t-1}\Big]$$

$$= \frac{\lambda\sigma_y^2}{2-\lambda}\Big[1-(1-\lambda)^{2t}\Big]+\lambda\sigma_y^2 2\sum_{k=1}^{t-1}\rho_k(1-\lambda)^k\underbrace{\sum_{i=0}^{t-1-k}(1-\lambda)^{2i}}_{\frac{1-(1-\lambda)^{2(t-k)}}{1-(1-\lambda)^2}}$$

$$= \frac{\lambda}{2-\lambda}\sigma_y^2\Big[1-(1-\lambda)^{2t}\Big]+\frac{\lambda}{2-\lambda}\sigma_y^2 2\sum_{k=1}^{t-1}\rho_k(1-\lambda)^k\Big[1-(1-\lambda)^{2(t-k)}\Big]$$

$$\mathrm{Var}(Z_t)=\sigma_z^2=\frac{\lambda}{2-\lambda}\sigma_y^2\left\{1-(1-\lambda)^{2t}+2\sum_{k=1}^{t-1}\rho_k(1-\lambda)^k\Big[1-(1-\lambda)^{2(t-k)}\Big]\right\}$$

APPENDIX C

DERIVATION OF THE AUTOCOVARIANCE FUNCTION AND THE STEADY-STATE VARIANCE OF ARMA STATISTIC

In this appendix, the autocovariance function and the steady-state variance of the ARMA statistic defined in (3.8) and (3.10), respectively, are derived when the underlying process, Y_t, is a stationary process with mean=μ, variance=σ_y^2, autocovariance function γ_k=cov(Y_t,Y_{t+k}) and autocorrelation function $\rho_k=\gamma_k/\gamma_0$ ($\gamma_0=\sigma_y^2$) at lag k.

Note that, in section 3.2, ARMA statistic is given as follows:

$$Z_t = \theta_0 Y_t + \alpha \sum_{k=1}^{t-1} \phi_1^{k-1} Y_{t-k}$$

where $\alpha = \phi_1\theta_0 - \theta$ and $\theta_0 = 1 + \theta - \phi_1$

1. Derivation of the autocovariance function of ARMA statistic

$$\mathrm{Cov}(Z_t, Z_{t+\tau}) = \mathrm{Cov}\left[\theta_0 Y_t + \alpha \sum_{i=1}^{t-1} \phi_1^{i-1} Y_{t-i}, \theta_0 Y_{t+\tau} + \alpha \sum_{j=1}^{t+\tau-1} \phi_1^{j-1} Y_{t+\tau-j}\right]$$

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

$$= \theta_0^2\gamma_\tau + \theta_0\alpha\left[\overbrace{Cov\left(Y_t, \sum_{j=1}^{t+\tau-1}\phi_1^{j-1}Y_{t+\tau-j}\right)}^{A} + \overbrace{Cov\left(Y_{t+\tau}, \sum_{i=1}^{t-1}\phi_1^{i-1}Y_{t-i}\right)}^{B}\right]$$

$$+\alpha^2\overbrace{Cov\left(\sum_{i=1}^{t-1}\phi_1^{i-1}Y_{t-i}, \sum_{j=1}^{t+\tau-1}\phi_1^{j-1}Y_{t+\tau-j}\right)}^{C}$$

$$A = Cov(Y_t, Y_{t+\tau-1} + \phi_1 Y_{t+\tau-2} + \phi_1^2 Y_{t+\tau-3} + \ldots + \phi_1^{t+\tau-2} Y_1)$$

$$= \gamma_{\tau-1} + \phi_1\gamma_{\tau-2} + \phi_1^2\gamma_{\tau-3} + \ldots + \phi_1^{t+\tau-2}\gamma_{1-t}$$

$$= \sum_{k=1}^{t+\tau-1}\phi_1^{k-1}\gamma_{\tau-k}$$

$$B = Cov(Y_{t+\tau}, Y_{t-1} + \phi_1 Y_{t-2} + \phi_1^2 Y_{t-3} + \ldots + \phi_1^{t-2} Y_1)$$

$$= \gamma_{\tau+1} + \phi_1\gamma_{\tau+2} + \phi_1^2\gamma_{\tau+3} + \ldots + \phi_1^{t-2}\gamma_{t+\tau-1}$$

$$= \sum_{k=1}^{t-1}\phi_1^{k-1}\gamma_{\tau+k}$$

$$C = Cov(Y_{t-1} + \phi_1 Y_{t-2} + \ldots + \phi_1^{t-2} Y_1, Y_{t+\tau-1} + \phi_1 Y_{t+\tau-2} + \ldots + \phi_1^{t+\tau-2} Y_1)$$

$$= \gamma_\tau + \phi_1\gamma_{\tau-1} + \phi_1^2\gamma_{\tau-2} + \ldots + \phi_1^{t+\tau-2}\gamma_{-t} + \phi_1\gamma_{\tau+1} + \phi_1^2\gamma_\tau + \ldots + \phi_1^{2t+\tau-4}\gamma_0$$

$$= \sum_{i=1}^{t-1}\sum_{j=1}^{t+\tau-1}\phi_1^{i+j-2}\gamma_{\tau-j+i}$$

$$\mathrm{Cov}(Z_t, Z_{t+\tau}) = \theta_0^2\gamma_\tau + \theta_0\alpha\left[\sum_{k=1}^{t+\tau-1}\phi_1^{k-1}\gamma_{\tau-k} + \sum_{k=1}^{t-1}\phi_1^{k-1}\gamma_{\tau+k}\right]$$

$$+\alpha^2\sum_{i=1}^{t-1}\sum_{j=1}^{t+\tau-1}\phi_1^{i+j-2}\gamma_{\tau-j+i}$$

Note that, $\gamma_k=\gamma_0\rho_k=\sigma_y^2\rho_k$. Therefore,

$$\mathrm{Cov}(Z_t, Z_{t+\tau}) = \sigma_y^2\left\{\theta_0^2\rho_\tau + \theta_0\alpha\left[\sum_{k=1}^{t+\tau-1}\phi_1^{k-1}\rho_{\tau-k} + \sum_{k=1}^{t-1}\phi_1^{k-1}\rho_{\tau+k}\right]\right.$$

$$\left.+\alpha^2\sum_{i=1}^{t-1}\sum_{j=1}^{t+\tau-1}\phi_1^{i+j-2}\rho_{\tau-j+i}\right\}$$

2. *Derivation of the steady-state variance of ARMA statistic*

$$\sigma_z^2 = \left\{\theta_0^2 + 2\theta_0\alpha\sum_{k=1}^{\infty}\phi_1^{k-1}\rho_k + \overbrace{\alpha^2\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\phi_1^{i+j-2}\rho_{j-i}}^{D}\right\}\sigma_y^2$$

$$D=\alpha^2\left(\rho_0 + \phi_1\rho_1 + \phi_1^2\rho_2 + \ldots + \phi_1\rho_1 + \phi_1^2\rho_0 + \phi_1^3\rho_1 + \ldots + \phi_1^2\rho_2 + \phi_1^3\rho_1 + \ldots\right)$$

$$=\alpha^2\left[\rho_0\left(1+\phi^2+\phi^4+\ldots\right)+2\rho_1\left(\phi+\phi^3+\phi^5+\ldots\right)+2\rho_2\left(\phi^2+\phi^4+\phi^6\ldots\right)+\ldots\right]$$

$$=\frac{\alpha^2}{1-\phi_1^2}+2\alpha^2\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\phi_1^{2l+k-2}\rho_k$$

$$= \frac{\alpha^2}{1-\phi_1^2} + 2\alpha^2\phi_1^{-2}\sum_{k=1}^{\infty}\phi_1^k\rho_k\sum_{l=1}^{\infty}\phi_1^{2l}$$

$$\sigma_z^2 = \left\{\theta_0^2 + 2\theta_0\alpha\sum_{k=1}^{\infty}\phi_1^{k-1}\rho_k + \frac{\alpha^2}{1-\phi_1^2} + 2\alpha^2\phi_1^{-2}\sum_{k=1}^{\infty}\phi_1^k\rho_k\sum_{l=1}^{\infty}\phi_1^{2l}\right\}\sigma_y^2$$

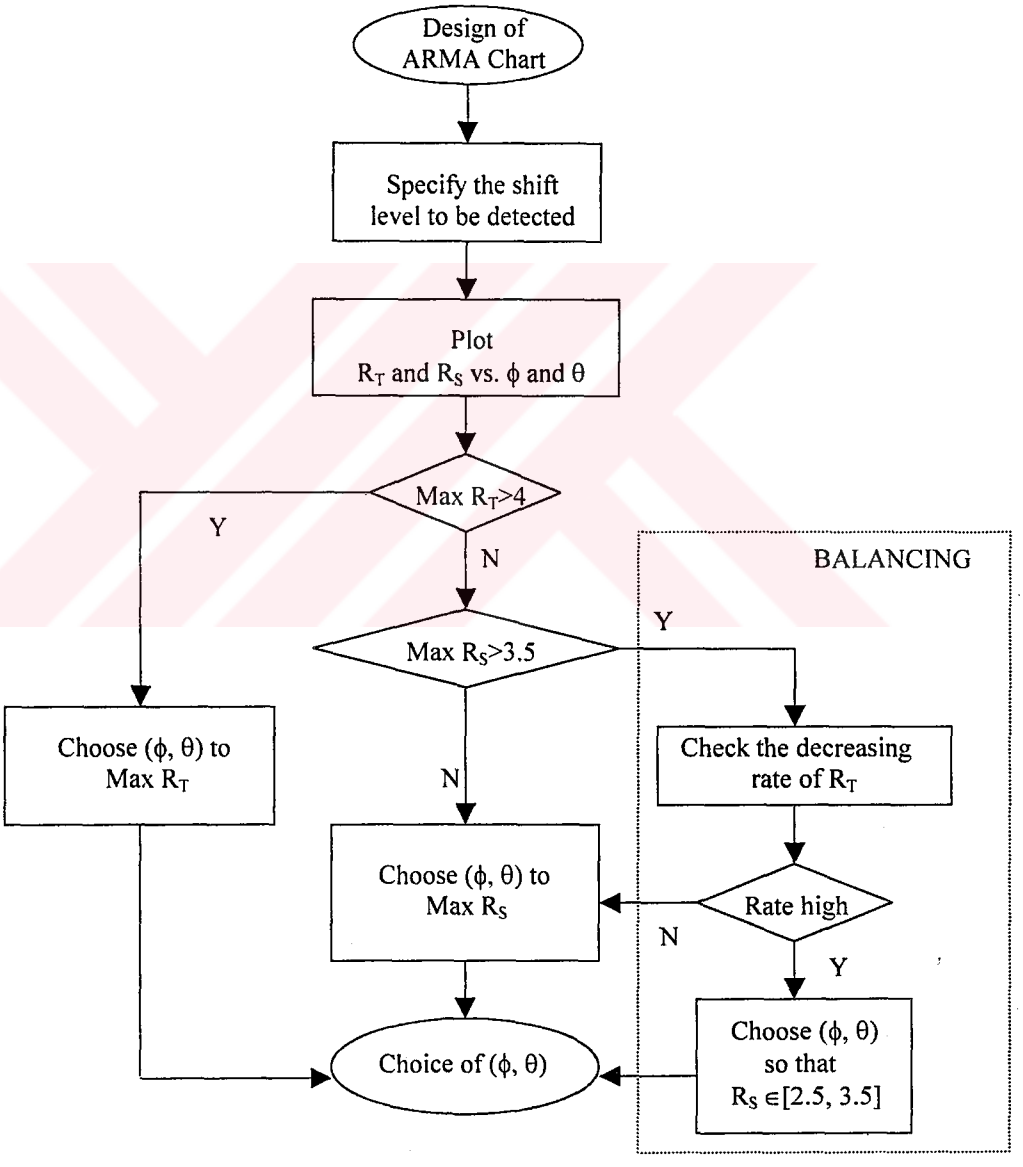
$$= \left\{\theta_0^2 + \frac{\alpha^2}{1-\phi_1^2} + 2\theta_0\alpha\sum_{k=1}^{\infty}\phi_1^{k-1}\rho_k + \frac{2\alpha^2\phi_1}{1-\phi_1^2}\sum_{k=1}^{\infty}\phi_1^{k-1}\rho_k\right\}\sigma_y^2$$

Then the steady-state variance of ARMA statistic becomes,

$$\sigma_y^2 = \left\{\theta_0^2 + \frac{\alpha^2}{1-\phi_1^2} + 2\left(\theta_0\alpha + \frac{\phi_1\alpha^2}{1-\phi_1^2}\right)\sum_{k=1}^{\infty}\phi_1^{k-1}\rho_k\right\}\sigma_y^2$$

APPENDIX D

ALGORITHM FOR THE PARAMETER DESIGN OF ARMA CHARTS

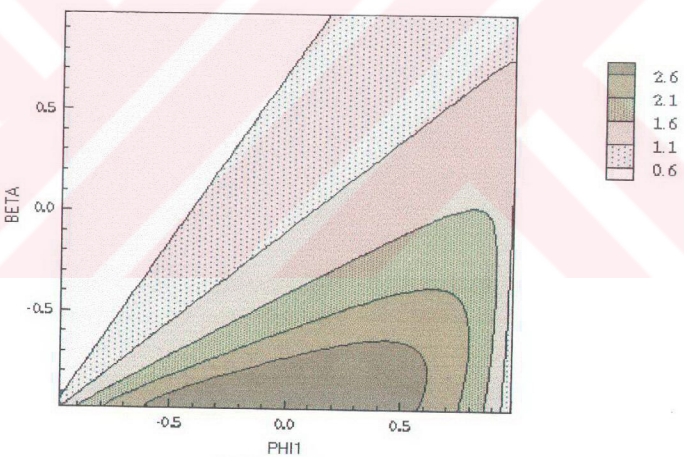


Source: Jiang, W., Tsui, K. L., and Woodall W. H. (2000)

APPENDIX E

PLOTS OF SIGNAL-TO-NOISE RATIOS

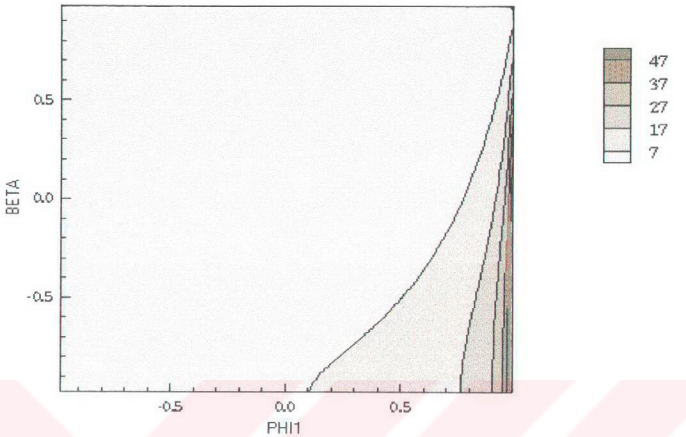
This appendix includes the plots of the transient and steady-state signal-to-noise ratios (R_T and R_S), which are used in the selection of ARMA chart parameters, when h=1.0 and ϕ=-0.95, -0.75, 0.75 and 0.95.



(a) Value of R_T

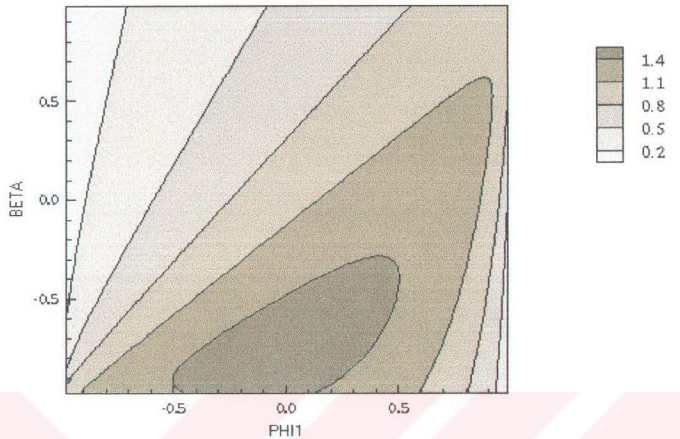
Figure E.1 Plots of Signal-to-Noise Ratios for ϕ=-0.95, h=1.0

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

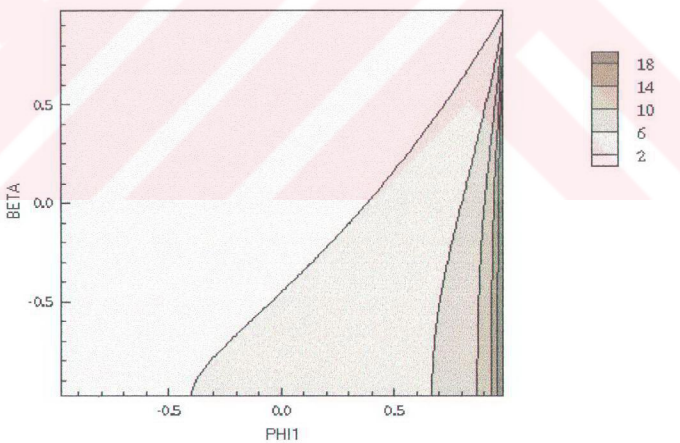


(b) Value of R_S

Figure E.1 (Cont.)



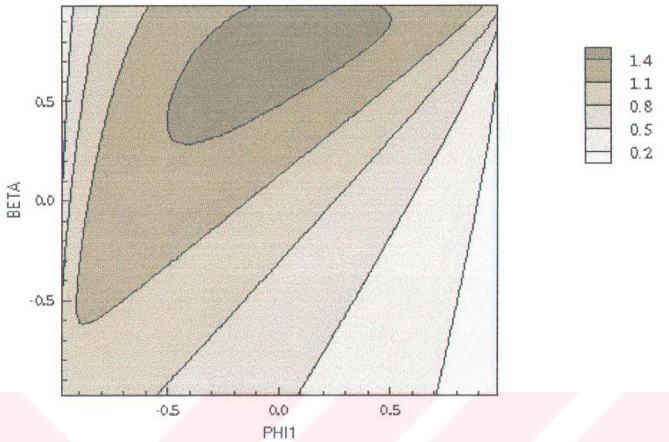
(a) Value of R_T



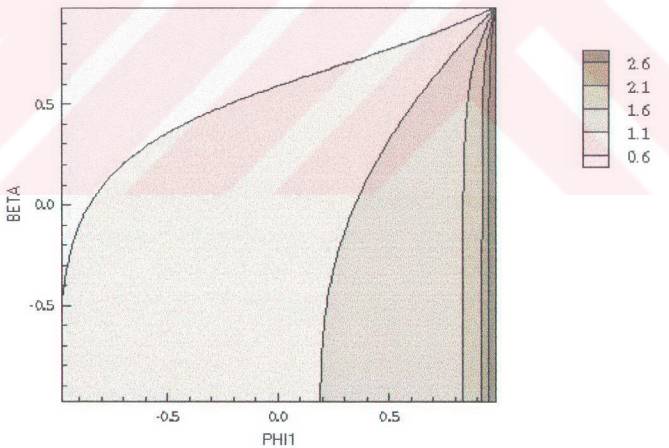
(b) Value of R_S

Figure E.2 Plots of Signal-to-Noise Ratios for ϕ=-0.75, h=1.0

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

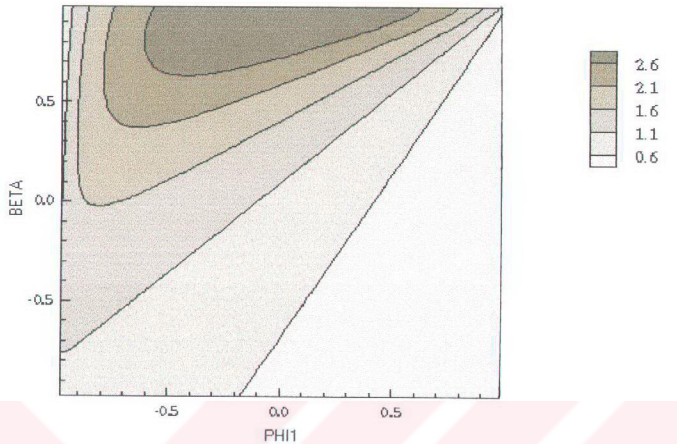


(a) Value of R_T

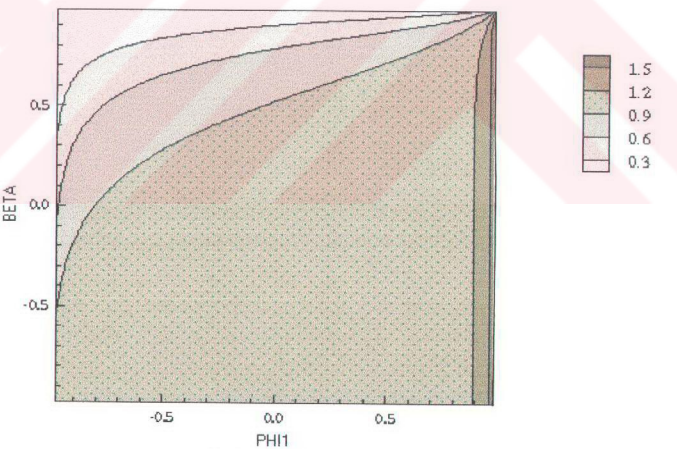


(b) Value of R_S

Figure E.3 Plots of Signal-to-Noise Ratios for ϕ=0.75, h=1.0



(a) Value of R_T



(b) Value of R_S

Figure E.4 Plots of Signal-to-Noise Ratios for ϕ=0.95, h=1.0

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANT