

Spring 2008

Monitoring the Process Mean of Autocorrelated Data

Jesse Dorian King

Follow this and additional works at: <https://digitalcommons.georgiasouthern.edu/etd>

Recommended Citation

King, Jesse Dorian, "Monitoring the Process Mean of Autocorrelated Data" (2008).
Electronic Theses and Dissertations. 648.
<https://digitalcommons.georgiasouthern.edu/etd/648>

This thesis (open access) is brought to you for free and open access by the Graduate Studies, Jack N. Averitt College of at Digital Commons@Georgia Southern. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons@Georgia Southern. For more information, please contact digitalcommons@georgiasouthern.edu.

MONITORING THE PROCESS MEAN OF AUTOCORRELATED DATA

by

JESSE D. KING

(Under the Direction of Dr. Charles W. Champ)

ABSTRACT

When modeling the stochastic behavior of a sequence $\{X_t\}$ of the quality measurement X on the output of a production process, it is usually assumed the measurements taken over time are independent and identically distributed. Multiple authors have pointed out that significant autocorrelation can affect the performance of traditional control charting procedures. One family of models for time series data are the autoregressive integrated moving average (ARIMA) models. These models are well suited to model production processes, in which the observations are autocorrelated. It is our interest to examine these models. Meaning is given to the process being in-control and out-of-control in terms of the parameters of the model. The performance of the Shewhart \bar{X} chart and CUSUM \bar{X} chart are compared. This includes determining the number of unobserved values between samples for the charts to perform as they would be expected if the samples were independent. Some recommendations are given.

INDEX WORDS: ARMA, ARIMA, Autocovariance, Autocorrelation, Shewhart, CUSUM, Control Charts, Gapping, Simulation

MONITORING THE PROCESS MEAN OF AUTOCORRELATED DATA

by

JESSE D. KING

B.S., Georgia Southern University, 2005

B.B.A., Georgia Southern University, 2005

A Thesis Submitted to the Graduate Faculty of Georgia Southern University in Partial
Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

STATESBORO, GEORGIA

2008

© 2008

Jesse D. King

All Rights Reserved

MONITORING THE PROCESS MEAN OF AUTOCORRELATED DATA

by

JESSE D. KING

Major Professor: Dr. Charles Champ

Committee: Dr. B. Oluyede
Dr. P. HumphreyElectronic Version Approved:
May 2008

DEDICATION

This thesis is dedicated to my parents whose encouragement motivated me to achieve my goals in education and in life. It is also dedicated to the rest of my family for their tolerance they have shown me for the last 2 years. Finally, thanks to my wife, who had to deal with me tuning her out during intensely productive sessions.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Charles W. Champ for his insightful guidance and constant encouragement. Without his ideas and help I could not have done this thesis project. I would also like to thank Dr. Broderick Oluyede and Dr. Patricia Humphrey for their suggestions and willingness to be on my committee. I applaud the faculty, staff and fellow graduates of the Department of Mathematics for their support.

TABLE OF CONTENTS

	Page
DEDICATION	5
ACKNOWLEDGMENTS.....	6
CHAPTER	
1 INTRODUCTION.....	9
2 MODEL, SAMPLING PROCEDURES, AND APPROXIMATION	
METHODS.....	10
INTRODUCTION.....	10
ARIMA(p,d,q) MODEL.....	10
AUTOCOVARIANCE AND AUTOCORRELATION FUNCTIONS.....	11
COMMON AND ASSIGNABLE CAUSES OF VARIABILITY	17
SIMULATION METHOD	19
SAMPLING PROCEDURES.....	21
ESTIMATING THE PROCESS PARAMETERS	22
CONCLUDING REMARKS.....	24
3 QUALITY CONTROL CHARTS	25
INTRODUCTION	25
SHEWHART \bar{X} CHART.....	26
CUSUM \bar{X} CHARTS.....	29
CONCLUDING REMARKS.....	31
4 DESIGNING A CONTROL CHART FOR AUTOCORRELATED DATA .	32
INTRODUCTION	32

	SIMULATION PROGRAM.....	33
	SIMULATION STUDY TO DETERMINE GAP SIZE	36
	PERFORMANCE STUDY	43
	CONCLUDING REMARKS.....	44
5	CONCLUSION.....	45
	GENERAL CONCLUSIONS.....	45
	AREAS FOR FURTHER RESEARCH	45
	REFERENCES.....	48
	APPENDIX	
	A MATLAB PROGRAM.....	54

CHAPTER 1

INTRODUCTION

Several authors have studied control charts for monitoring a process in which the quality measurements of the output are autocorrelated. These include, among others, Alwan (1992), Alwan and Roberts (1988), Box and Luceño (1997), Lu and Reynolds (1999), Montgomery and Mastrangelo (1991), Runger and Prabhu (1996), Runger and Willemain (1995), Runger, Willemain, and Prabhu (1995), Wardell, Moskowitz, and Plante (1994), and Zhang (1998).

It would be useful to develop a method that would allow the practitioner to use well known control charts and procedures on autocorrelated processes. We believe this may be easily accomplished when gapping between samples is practical that is often the case in industry. It is the purpose of this thesis to develop a method for designing a CUSUM \bar{X} chart for monitoring the mean of an autocorrelated process.

In Chapter 2, we discuss the autocorrelated data model both when the process is in- and out-of-control. A method is given for simulating a time series from an $ARIMA(p,0,q)$ data model. CUSUM \bar{X} charts are discussed in Chapter 3. In the case the practitioner chooses to use periodic sampling, a method is given in Chapter 4 for designing a CUSUM \bar{X} chart. A concluding chapter follows which gives several areas for further research.

CHAPTER 2

MODEL, SAMPLING PROCEDURES, AND APPROXIMATION METHODS

2.1. INTRODUCTION

When modeling the stochastic behavior of a sequence $\{X_t\}$ of the quality measurement X on the output of a production process, it is usually assumed the measurements taken over time are independent and identically distributed. Multiple authors have pointed out that significant autocorrelation can affect the performance of traditional control charting procedures. One family of models for time series data are the autoregressive integrated moving average (ARIMA) models. These models are well suited to model production processes, in which the observations are autocorrelated. It is our interest to examine these models. The general form of the model is described in the next section. In Section 3, we examine the autocovariance and autocorrelation functions and show how these functions can be obtained as functions of the parameters of the model. Models for in-control and out-of-control are developed in the fourth section. A method is given in Section 5 for simulating an $ARIMA(p,0,q)$ time series both when the process is in- and out-of-control. Sampling procedures are outlined in Section 6. While we are only interested in the case where the parameters are known, we provide in Section 7 some discussion of methods for estimating the parameters of an $ARIMA(p,0,q)$ model. Some concluding remarks are given in the last section.

2.2. $ARIMA(p,d,q)$ MODEL

The general form of the $ARIMA(p,d,q)$ model is given by

$$\phi_p(B)\nabla^d X_t = \phi_0 + \theta_q(B)a_t,$$

where $\nabla^d = (1 - B)^d$. The backward shift operator B , the autoregressive operator, $\phi_p(B)$ and the moving average operator $\theta_q(B)$ are defined, respectively, by

$$B^k X_t = X_{t-k}, \phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p, \text{ and } \theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B.$$

The values p , d , and q are nonnegative integers. The sequence $\{a_t\}$, referred to as “shocks,” is assumed to be a sequence of independent and identically distributed random variables each with mean 0 and common variance σ_a^2 . The ARIMA(p, d, q) models are discussed in Box and Jenkins (1994) and Wei (1990), among others.

Our study of the performance of various quality control charts will be limited to ARIMA($p, 0, q$) models. These models are a special sub-family of ARIMA models that are known as autoregressive moving average (ARMA) models. These models have the form

$$X_t = \phi_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}.$$

We assume the stochastic process $\{X_t\}$ is stationary. Box and Jenkins (1994) show that an ARIMA($p, 0, q$) process is stationary if the roots of $\phi(B) = 0$ are outside the unit circle. It follows that $E(X_t) = E(X_{t+k})$ and $\gamma_k = \text{cov}(X_t, X_{t-k})$ holds for all values of t .

Further, we assume that any collection of n random variables in the sequence $\{X_t\}$ follows an n -variate (multivariate) normal distribution.

2.3. AUTOCOVARANCE AND AUTOCORRELATION FUNCTIONS

To determine the mean of the distribution of X_t , we observe that

$$E(X_t) = \phi_0 + \phi_1 E(X_{t-1}) + \dots + \phi_p E(X_{t-p}) + E(a_t) - \theta_1 E(a_{t-1}) - \dots - \theta_q E(a_{t-q})$$

Under our assumptions, we have that

$$E(X_t) = E(X_{t-1}) = \dots = E(X_{t-p}) \text{ and } E(a_t) = E(a_{t-1}) = \dots = E(a_{t-q}) = 0.$$

It then follows that

$$\mu_{t,X} = E(X_t) = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p}.$$

As we see here, the mean $\mu_{t,X}$ is not a function of t . This is due to the stationary assumption and the assumption that $E(a_t) = 0$ for all t . Note that

$$\phi_0 = (1 - \phi_1 - \dots - \phi_p)\mu_{t,X}.$$

To obtain the autocovariances, we define

$$X_t^* = X_t - \mu_{t,X}$$

and observe that

$$X_t^* = \phi_1 X_{t-1}^* + \dots + \phi_p X_{t-p}^* + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}. \quad (2.3.1)$$

It is not difficult to show that

$$\gamma_k = \text{cov}(X_t, X_{t-k}) = \text{cov}(X_t^*, X_{t-k}^*) = E(X_t^* X_{t-k}^*).$$

Now, multiplying the Equation (2.3.1) by X_{t-k}^* and taking expectations, we have

$$\begin{aligned} E(X_t^* X_{t-k}^*) &= \phi_1 E(X_{t-1}^* X_{t-k}^*) + \dots + \phi_p E(X_{t-p}^* X_{t-k}^*) \\ &\quad + E(a_t X_{t-k}^*) - \theta_1 E(a_{t-1} X_{t-k}^*) - \dots - \theta_q E(a_{t-q} X_{t-k}^*) \end{aligned} \quad (2.3.2)$$

It will be convenient to let

$$\begin{aligned} \xi_{k+i} &= E(a_{t-i} X_{t-k}), \text{ if } k \leq i; \\ &= 0, \text{ if } k > i. \end{aligned}$$

Equation (2.3.2) gives us a system of equations of the form

$$\vec{\gamma}_p = (\mathbf{A}_{1,p} + \mathbf{A}_{2,p})\vec{\gamma}_p + \mathbf{B}_{p,q}\vec{\xi}_q.$$

The vectors $\vec{\gamma}_p = [\gamma_0, \gamma_1, \dots, \gamma_p]^T$ and $\vec{\xi}_q = [\xi_0, \xi_1, \dots, \xi_q]^T$ have dimensions $(p+1) \times 1$ and $(q+1) \times 1$, respectively. The matrix A_p has dimensions $(p+1) \times (p+1)$ with the $(i, j)^{th}$ component $a_{1,p}(i, j)$ given by

$$\begin{aligned} a_{1,p}(i, j) &= \phi_{i+j-2}, \text{ if } 0 < i \leq p+1 \text{ or } 0 < j \leq p-i+2; \\ &= 0, \text{ otherwise.} \end{aligned}$$

The matrix $A_{2,p}$ has dimensions $(p+1) \times (p+1)$ with the $(i, j)^{th}$ component $a_{2,p}(i, j)$ given by

$$\begin{aligned} a_{2,p}(i, j) &= \phi_{i-j}, \text{ if } 2 < i \leq p, 2 \leq j \leq i-1, \text{ and } p > 2; \\ &= 0, \text{ otherwise.} \end{aligned}$$

The matrix $B_{p,q}$ has dimensions $(p+1) \times (q+1)$ with the $(i, j)^{th}$ component $b_{p,q}(i, j)$ given by

$$\begin{aligned} b_{p,q}(i, j) &= 1, \text{ if } i = j = 1; \\ &= -\theta_{i+j-2}, \text{ if } 1 < i \leq p; 2 \leq j \leq \min\{i-1, q\}; \\ &= 0, \text{ otherwise.} \end{aligned}$$

To determine ξ_k , we consider the sequence of equations

$$\begin{aligned} E(a_{t-k} X_t^*) &= \phi_1 E(a_{t-k} X_{t-1}^*) + \dots + \phi_p E(a_{t-k} X_{t-p}^*) \\ &\quad + E(a_{t-k} a_t) - \theta_1 E(a_{t-k} a_{t-1}) - \dots - \theta_q E(a_{t-k} a_{t-q}). \end{aligned}$$

It follows that

$$\begin{aligned}
\xi_0 &= \sigma_a^2 \\
\xi_1 &= \phi_1 \xi_0 - \theta_1 \sigma_a^2 \\
&\vdots \\
\xi_q &= \phi_1 \xi_{q-1} + \dots + \phi_q \xi_0 - \theta_q \sigma_a^2 \\
\xi_{q+1} &= \phi_1 \xi_q + \dots + \phi_{q+1} \xi_0 \\
&\vdots \\
\xi_p &= \phi_1 \xi_{p-1} + \dots + \phi_p \xi_0 \\
\xi_k &= \phi_1 \xi_{k-1} + \dots + \phi_p \xi_{k-p}, \text{ for } k > p.
\end{aligned} \tag{2.3.3}$$

The values of ξ_k can be obtained iteratively from the equations in (2.3.3). Further, ξ_k can be expressed as a function ς_k of the autoregressive parameters ϕ_1, \dots, ϕ_p and the moving average parameters $\theta_1, \dots, \theta_q$ times the variance σ_a^2 of a_t . That is,

$$\varsigma_k = \varsigma_k(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q) = \xi_k / \sigma_a^2 \tag{2.3.4}$$

To solve our system of equations, we express our model in the form

$$X_t = \phi_0 + \phi_1 X_{t-1} + \dots + \phi_r X_{t-r} + a_t - \theta_1 a_{t-1} - \dots - \theta_r a_{t-r},$$

where $r = \max\{p, q\}$. For the case in which $r = p > q$, the parameters $\theta_{q+1}, \dots, \theta_r$ are zero; and for the case in which $r = q > p$, the parameters $\phi_{p+1}, \dots, \phi_r$ are zero. The system of equations can be expressed in matrix form as

$$\begin{aligned}
\vec{\gamma}_r &= (\mathbf{A}_r + \mathbf{B}_r) \vec{\gamma}_r + \mathbf{C}_{r,r} \vec{\xi}_r = (\mathbf{A}_r + \mathbf{B}_r) \vec{\gamma}_r + \mathbf{C}_{r,r} \vec{\varsigma}_r \sigma_a^2 \text{ and} \\
\gamma_k &= \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} \text{ for } k > r
\end{aligned} \tag{2.3.5}$$

ξ_k is determined by the equations in (2.3.3), and $\vec{\varsigma}_r = [\varsigma_0, \varsigma_1, \dots, \varsigma_r]^T$.

The results in (2.3.5) are a special case of the results derived by McLeod (1975) which also includes the seasonal model. Ansley (1980) extended the results of McLeod (1975) to the multivariate ARIMA($p, 0, q$) (MARIMA($p, 0, q$)) model. A FORTRAN program is

given by Pate and Davies (1988) implementing, with some modifications, the method of Ansley (1980).

It will be useful to define the transformed stochastic process $\{W_t\}$ by

$$W_t = (X_t - \mu_{t,X}) / \sigma_a. \quad (2.3.6)$$

For the ARIMA($p,0,q$) model, we have that

$$W_t = \phi_1 W_{t-1} + \dots + \phi_p W_{t-p} + b_t - \theta_1 b_{t-1} - \dots - \theta_q b_{t-q},$$

where $b_t = a_t / \sigma_a$. It is easy to see that $b_t \sim N(0,1)$.

The following two theorems hold for the stochastic process $\{W_t\}$.

Theorem 2.3.1: $E(W_t) = 0$ and $\tau_k = \text{cov}(W_t, W_{t-k}) = \text{cov}(X_t, X_{t-k}) / \sigma_a^2 = \gamma_k / \sigma_a^2$ for $k = 0, 1, 2, \dots$

Proof of Theorem 2.3.1: It follows that

$$E(W_t) = E[(X_t - \mu_{t,X}) / \sigma_a] = (E(X_t) - \mu_{t,X}) / \sigma_a = 0$$

since $E(X_t) = \mu_{t,X}$. Using Equation (2.3.6), we have

$$\begin{aligned} \tau_k &= \text{cov}(W_t, W_{t-k}) = E(W_t W_{t-k}) = E\left(\frac{X_t - \mu_{t,X}}{\sigma_a} \frac{X_{t-k} - \mu_{t-k,X}}{\sigma_a}\right) \\ &= E((X_t - \mu_{t,X})(X_{t-k} - \mu_{t-k,X})) / \sigma_a^2 = \text{cov}(X_t, X_{t-k}) / \sigma_a^2 = \gamma_k / \sigma_a^2 \end{aligned}$$

The autocorrelation function ρ_k is defined by $\rho_k = \gamma_k / \gamma_0$. Using the results in Theorem 2.3.1, yields the following results.

Theorem 2.3.2: $\rho_k = \tau_k / \tau_0$ for $k = 0, 1, 2, \dots$

By dividing Equation (2.3.5) by σ_a , we have the system of equations

$$\vec{\tau}_r = (\mathbf{A}_{1,r} + \mathbf{A}_{1,r})\vec{\tau}_r + \mathbf{B}_{r,r}\vec{\zeta}_r,$$

where $\vec{\tau}_r = [\tau_0, \tau_1, \dots, \tau_r]^T$.

For most of our analyzes, we will be considering ARMA($p, 0, q$) models in which both p and q are less than or equal to 2. For these cases, our system of equations can be written as

$$\begin{aligned} \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix} &= \left(\begin{bmatrix} 0 & \phi_1 & \phi_2 \\ \phi_1 & \phi_2 & 0 \\ \phi_2 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \phi_1 & 0 \end{bmatrix} \right) \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix} \\ &+ \begin{bmatrix} 1 & -\theta_1 & -\theta_2 \\ -\theta_1 & -\theta_2 & 0 \\ -\theta_2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ \phi_1 - \theta_1 \\ \phi_1(\phi_1 - \theta_1) + \phi_2 - \theta_2 \end{bmatrix} \sigma_a^2 \end{aligned} \quad (2.3.7)$$

with $\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}$ for $k > 2$. The solution $\vec{\gamma}_2 = [\gamma_0, \gamma_1, \gamma_2]^T$ can be expressed as

$$\begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} 1 & -\phi_1 & -\phi_2 \\ -\phi_1 & 1 - \phi_2 & 0 \\ -\phi_2 & -\phi_1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 1 & -\theta_1 & -\theta_2 \\ -\theta_1 & -\theta_2 & 0 \\ -\theta_2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ \phi_1 - \theta_1 \\ \phi_1(\phi_1 - \theta_1) + \phi_2 - \theta_2 \end{bmatrix} \sigma_a^2$$

Completing these computations in symbolic form gives expressions that are quite messy. In general, the results are determined numerically for given model parameters. However, some special cases lead to simple closed form expressions for the autocovariance function.

For example, the case in which $p = 1$ and $q = 1$, equation (2.3.7) reduces to the system

$$\begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix} = \left(\begin{bmatrix} 0 & \phi_1 & 0 \\ \phi_1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \phi_1 & 0 \end{bmatrix} \right) \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix} + \begin{bmatrix} 1 & -\theta_1 & 0 \\ -\theta_1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ \phi_1 - \theta_1 \\ \phi_1(\phi_1 - \theta_1) \end{bmatrix} \sigma_a^2. \quad (2.3.8)$$

Equation (2.3.8) follows directly from Equation (2.3.7) by setting $\phi_2 = \theta_2 = 0$. The solution is given by

$$\gamma_0 = \frac{1 - 2\phi_1\theta_1 + \theta_1^2}{1 - \phi_1^2} \sigma_a^2 \text{ and } \gamma_k = \frac{\phi_1^{k-1}(\phi_1 - \theta_1)(1 - \phi_1\theta_1)}{1 - \phi_1^2} \sigma_a^2 \quad (2.3.9)$$

for $k \geq 1$. Further, setting $\theta_1 = 0$ in equations (2.3.7, 2.3.8, and 2.3.9) yields the results for the ARIMA(1,0,0) and setting $\phi_1 = 0$ in equations (2.3.7, 2.3.8, and 2.3.9) yields the results for the ARIMA(0,0,1) model.

2.4. COMMON AND ASSIGNABLE CAUSES OF VARIABILITY

Shewhart (1931) introduced the ideas of common and assignable causes of variability for industrial processes. Basically, one can view a common cause of variability as one that can be removed by redesigning the process whereas an assignable cause is one that can be removed when discovered and the quality of the process improved. In order to provide the practitioner with an analysis of a control charting procedure, both common and assignable causes must be interpreted in terms of the parameters of the data model.

In general, it is assumed that if the process is in-control it follows a given ARIMA model; and when the process changes to an out-of-control state it follows another ARIMA model. Runger (2002) discusses in-control and out-of-control in terms of the parameters of a simple dynamic model. In this section, we will examine what is meant by the process being in- and out-of-control in terms of the parameters of ARIMA($p,0,q$)

models. In particular, we look at out-of-control scenarios that are changes in the distribution of the shock, a_t .

Under the independent normal data model, it is typically assumed the process is in-control if $\mu_X = \mu_{0,X}$ and $\sigma_X = \sigma_{0,X}$, for fixed values $\mu_{0,X}$ and $\sigma_{0,X}$. Thus when the process is in-control, we can express our model as

$$X_t = \mu_{0,X} + a_t,$$

where $a_t \sim N(0, \sigma_{0,a}^2)$. In this case, the variability of the distribution of the random variable a_t is due a natural (common) cause(s). Note that under this model that

$$\sigma_{0,a}^2 = \sigma_{0,X}^2.$$

It is typical to assume that when an assignable cause(s) affects the process it is reflected in a change in the distribution of a_t from $a_t \sim N(0, \sigma_{0,a}^2)$ to $a_t \sim N(\delta\sigma_{0,a}, \lambda^2\sigma_{0,a}^2)$, where $\delta \neq 0$ or $0 < \lambda \neq 1$. It then follows that

$$X_t \sim N(\mu_{0,X} + \delta\sigma_{0,X}, \lambda^2\sigma_{0,X}^2)$$

with the process being in-control if $\delta = 0$ and $\lambda = 1$. For autocorrelated data following an ARIMA($p,0,q$) model, we can make the same assumptions about the distribution of the random variable a_t (often referred to as a shock) as we did in the independent data model. That is, a process change occurs when the distribution of the random shocks changes from $a_t \sim N(0, \sigma_{0,a}^2)$ to $a_t \sim N(\delta\sigma_{0,a}, \lambda^2\sigma_{0,a}^2)$ for $\delta \neq 0$ or $0 < \lambda \neq 1$. We will refer to this as the out-of-control shock (OCS) model.

2.5. SIMULATION METHOD

We are interested in simulating time series from a univariate ARIMA($p,0,q$) process with given parameters $\phi_0, \phi_1, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$ and σ_a^2 . Anderson (1979), Beall (1982), and Burn (1986) discuss methods for “warming-up” a time series. Basically, in warming-up a time series, the initial values X_{1-p}, \dots, X_0 are set to given values, say for example the mean of the process and then N values are generated. For large enough N , it is assumed that the time series from the $N+1$ value of X have approximately the stochastic properties of the desired time series.

McLeod (1975) alludes to a method for simulating an ARIMA($p,0,q$) time series. He states that his method for obtaining the autocorrelation function “is useful in simulating initial values of a time series in simulation studies” He provides no further guidance in simulating these initial values. In this section, we discuss a method for simulating these initial values, X_{1-p}, \dots, X_0 given in Kreiger (1992). This method treats the initial values X_{1-p}, \dots, X_0 as a $p \times 1$ vector $[X_{1-p}, \dots, X_0]^T$. Under the normality assumption, we have that

$$\mathbf{X} = \begin{bmatrix} X_0 \\ X_{-1} \\ \vdots \\ X_{1-p} \end{bmatrix} \sim N_p \left(\mu_X = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \Sigma = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{bmatrix} \right).$$

The covariance matrix Σ is positive definite and hence has positive real eigenvalues. We define \mathbf{C} to be a diagonal matrix of the eigenvalues of Σ and \mathbf{V} the matrix whose columns are the associated normalized eigenvectors of Σ . The transformation of \mathbf{X} to

$$\mathbf{Z} = (\mathbf{VC}^{1/2})^{-1}(\mathbf{X} - \mu_X)$$

gives a $p \times 1$ vector $\mathbf{Z} = [Z_{1-p}, \dots, Z_0]^T$ with

$$\begin{bmatrix} Z_0 \\ Z_{-1} \\ \vdots \\ Z_{1-p} \end{bmatrix} \sim N_p \left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \right).$$

To initialize our time series generator, we generate the observed values of the p independent standard normal random variables Z_{1-p}, \dots, Z_0 and transform these values by

$$\mathbf{X} = \boldsymbol{\mu}_X + (\mathbf{VC}^{1/2})\mathbf{Z}$$

yielding the simulated values of the p initial values X_{1-p}, \dots, X_0 for the simulated time series. Next the observed values of the independent normal random variables a_0, \dots, a_{1-q} , are generated from a $N(0, \sigma_a^2)$ distribution while a_t is generated from a $N(\delta\sigma_{0,a}, \lambda^2\sigma_{0,a}^2)$. The first value of the simulated time series is then calculated by

$$X_1 = \phi_0 + \phi_1 X_0 + \dots + \phi_p X_{1-p} + a_1 - \theta_1 a_0 - \dots - \theta_q a_{1-q}$$

The t th value of the simulated time series is then generated iteratively by simulating independently the observed value of the random variable a_t from a $N(\delta\sigma_{0,a}, \lambda^2\sigma_{0,a}^2)$ distribution and calculating the simulated value of X_t as

$$X_t = \phi_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}.$$

A similar method could be used to simulate values from a multivariate ARIMA($p,0,q$) (MARIMA($p,0,q$)). The method of Ansley (1980) implemented by Pate and Davies (1988) in FORTRAN would be used to calculate the autocorrelation matrices. These results would then be used in the procedure to simulate the initial values of the multivariate time series. Other authors have given methods for generating multivariate

time series. These include Krogstad (1989) and Mastrangelo and Forrest (2002) who give a FORTRAN program to generate the series. These methods are only useful for generating in-control data whereas our method generates out-of-control data assuming that $a_t \sim N(\delta\sigma_{0,a}, \lambda^2\sigma_{0,a}^2)$. In the APPENDIX we provide a MATLAB program that implements the method outlined in this section for generating both in- and out-of-control ARIMA($p,0,q$) time series.

2.6. SAMPLING PROCEDURES

It is typical for a practitioner to periodically take the quality measurements on a sample of items from the output of the process to monitor for a change in quality. Under the correlated data model, it will be important to maintain the order of the output relative to these measurements. A simple sampling procedure is one in which the practitioner selects, in the order of output, a sample of size n and waits to take the next sample after m items have been produced. It will be convenient to represent the quality measurements on the items in the i th sample as the vector

$$\mathbf{X}_i = [X_{(i-1)(n+m)+1}, X_{(i-1)(n+m)+2}, \dots, X_{in+(i-1)m}]^T$$

and the unobserved measurements of the next m items in the order of output by the vector

$$\mathbf{G}_i = [X_{(i-1)(n+m)+(n+1)}, X_{(i-1)(n+m)+(n+2)}, \dots, X_{i(n+m)}]^T.$$

These items and/or their measurements will be referred as the i th “gap.” When every item from the output of the process is measured, this is referred to as 100% sampling. For this sampling procedure, 100% sampling occurs when $n = 1$ and $m = 0$.

Another sampling method presented by Apley and Tsung (2002) takes as the i th sample as the vector

$$X_i = [X_{1+i-1}, X_{2+i-1}, \dots, X_{n+i-1}]^T.$$

Clearly, these samples overlap. Using this sampling method, every item of the output of the process is measured; consequently, this is a 100% sampling procedure.

2.7. ESTIMATING THE PROCESS PARAMETERS

It is of importance to have estimates of the in-control parameters of the process when they are not available. We assume that we will have available N quality measurements X_1, X_2, \dots, X_N on items from the output of the process in the order of production from an in-control process. Various methods have been proposed for estimating the autocorrelation function. Jenkins and Watts (1968) concluded the most satisfactory estimate for ρ_k is

$$\hat{\rho}_k = r_k = \frac{c_k}{c_0}$$

where

$$c_k = \frac{1}{N} \sum_{t=1}^{N-k} (X_t - \bar{X}_0)(X_{t+k} - \bar{X}_0)$$

is the estimate for γ_k , $k = 1, 2, \dots, K$, and

$$\bar{X}_0 = \frac{1}{N} \sum_{t=1}^N X_t.$$

The values c_k and r_k are called the sample autocovariance and autocorrelation functions, respectively.

The mean and variance of \bar{X}_0 are given by

$$\mu_{0,\bar{X}_0} = E_0(\bar{X}_0) = \frac{1}{N} \sum_{t=1}^N E_0(X_t) = \mu_{0,X} \text{ and}$$

$$\sigma_{0,\bar{X}_0}^2 = V_0(\bar{X}_0) = \frac{\gamma_0}{N} + 2 \sum_{j=1}^{N-1} \frac{N-j}{N} \frac{\gamma_j}{N} = \frac{\gamma_0}{N} (1 + 2 \sum_{j=1}^{N-1} \frac{N-j}{N} \rho_j).$$

For c_k , we have that

$$E(c_k) = \frac{1}{N} \sum_{t=1}^{N-k} E[(X_t - \bar{X}_0)(X_{t+k} - \bar{X}_0)].$$

As we will see, we will be interested in the time series based on the transformed stochastic process, $\{W_t\}$. The estimate of the autocorrelation function of the transformed stochastic process $\{W_t\}$ is the same as for $\{X_t\}$. The estimate of the autocovariance function τ_k of the stochastic process $\{W_t\}$ is given by

$$\hat{\tau}_k = \frac{1}{N} \sum_{t=1}^{N-k} (W_t - \bar{W}_0)(W_{t+k} - \bar{W}_0)$$

for $k = 1, 2, \dots, K$, with $\bar{W}_0 = \frac{1}{N} \sum_{t=1}^N W_t$. We note here that \bar{X}_0 provides an estimate for $\mu_{0,X}$ while \bar{W}_0 provides an estimate for 0.

The mean \bar{X}_i and the variance S_i^2 are two statistics that are often used to summarize the information found in the sample measurements. For the i th sample, we have

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{(i-1)(n+m)+j} \text{ and } S_i^2 = \frac{1}{n-1} \sum_{j=1}^n (X_{(i-1)(n+m)+j} - \bar{X}_i)^2.$$

Theorem 2.7.1. If ARIMA($p,0,q$) process is in-control, then

$$\mu_{0,\bar{X}} = E(\bar{X}_i) = \mu_{0,X} \text{ and } \sigma_{0,\bar{X}}^2 = V(\bar{X}_i) = \frac{\gamma_0}{n} (1 + 2 \sum_{j=1}^{n-1} \frac{n-j}{n} \rho_j).$$

Proof of Theorem 2.7.1: In the case of the mean, we have

$$\mu_{0,\bar{X}} = E(\bar{X}_i) = \frac{1}{n} \sum_{j=1}^n E(X_{(i-1)(n+m)+j}) = \mu_{0,X}.$$

For the variance, it can be seen that

$$\begin{aligned} \sigma_{0,\bar{X}}^2 &= V(\bar{X}_i) = \frac{1}{n^2} \sum_{u=1}^n \sum_{v=1}^n \text{cov}(X_{(i-1)(n+m)+u}, X_{(i-1)(n+m)+v}) = \frac{1}{n^2} \sum_{u=1}^n \sum_{v=1}^n \gamma_{|u-v|} \\ &= \frac{1}{n^2} (n\gamma_0 + 2 \sum_{j=1}^{n-1} (n-j)\gamma_j) = \frac{\gamma_0}{n} + 2 \frac{\gamma_0}{n} \sum_{j=1}^{n-1} \frac{n-j}{n} \frac{\gamma_j}{\gamma_0} = \frac{\gamma_0}{n} (1 + 2 \sum_{j=1}^{n-1} \frac{n-j}{n} \rho_j). \end{aligned}$$

It is not our interest to study in this thesis the performance of control charts for autocorrelated data when the process parameters are estimated. This will be a topic for later research.

2.8. CONCLUDING REMARKS

The ARIMA($p,0,q$) model was studied in its use as a model for process data. A method for simulating both in- and out-of-control ARIMA($p,0,q$) time series does not exist in the literature. We have outlined a method for simulating a time series from an ARIMA($p,0,q$) model under the assumptions the process is stationary and any finite collection of the random variables follow a multivariate normal distribution. Further, we provide a MATLAB program implementing this method. Some sampling methods found in the literature were discussed along with methods for estimating the parameters of an ARIMA($p,0,q$) model.

CHAPTER 3

QUALITY CONTROL CHARTS

3.1 INTRODUCTION

The quality control chart introduced by Walter A. Shewhart in the 1920's (see Shewhart (1931)) is a statistical method that has been found useful by practitioners in the production of quality goods and services. Duncan (1986) pointed out that there are three uses of the quality control charts. First, it aids the practitioner in bringing a process into a state of statistical in-control. Secondly, it may be used as an aid in establishing the meaning of the process being in a state of statistical in-control. Thirdly, the control chart is used to aid the practitioner in monitoring the process for a change in the process from a state of statistical in-control to a state of statistical out-of-control. Charts used in the first two cases are referred to as retrospective or Phase I control charts. Using a set of data initially collected on the output of the process, they help answer the question "were these items produced by an in-control process." In the latter case, these charts are used to monitor for a change in the process and are referred to as prospective or Phase II charts.

For a continuous quality measurement, the most commonly recommended charts for monitoring the process mean are the Shewhart \bar{X} chart (see Shewhart (1931) and Montgomery (2001)), Shewhart \bar{X} chart with runs rules (see Champ and Woodall (1987)), the cumulative sum (CUSUM) \bar{X} chart of Page (1954), the CUSUM \bar{X} chart of Crosier (1986), and the exponential weighted moving average (EWMA) \bar{X} of Roberts (1959) (see Lucas and Saccucci (1990)). A family of cumulative sum type charts was introduced by Champ, Woodall, and Mohsen (1991) that includes as members the

Shewhart, CUSUM, and EWMA \bar{X} charts. Discussions of the parameters estimated versions of these charts can be found in Bagshaw and Johnson (1975), Ghosh, Reynolds, and Hui (1981), Ng and Case (1992), Burrough (1993), Burroughs, Rigdon, and Champ (1993, 1995), Quesenberry (1993), Chen (1997), Jones, Champ, and Rigdon (1999),

The Hotelling's T^2 chart (see Hotelling (1947)), the multivariate CUSUM \bar{X} charts of Crosier (1988), Healy (1987), and Pignatiello and Runger (1990), and the multivariate EWMA chart of Lowry, et al (1992) are commonly discussed charts for monitoring the mean vector of the distribution of a vector of quality measurements. Champ and Jones-Farmer (2007) give the estimated parameters version of these charts and equivalent forms of these charts that are useful in analyzing the run length properties of these charts.

In this chapter, we discuss the univariate charts for monitoring the mean of the quality measurement under an $ARMA(p,0,q)$ data model. Also, we examine how one could use various multivariate charts to monitor the mean of the process. In the next section, we discuss the Shewhart \bar{X} chart, followed by the CUSUM \bar{X} chart in section three. Section four includes some concluding remarks.

3.2. SHEWHART \bar{X} CHART

One of the most commonly recommended charts for monitoring the mean of a quality measurement is the \bar{X} chart introduced by Dr. Walter A. Shewhart in the early 1920's. The chart is a plot of the means of the quality measurements taken on the items in a sample from the output of the production process verses the sample number. Letting \bar{X}_i be the mean of the i th sample, the Shewhart \bar{X} chart is a plot of the points (i, \bar{X}_i)

for $i = 1, 2, 3, \dots$. We first consider the case in which our data are measurements on the samples of n items in order of output of the form

$$\mathbf{X}_i = [X_{(i-1)(n+m)+1}, X_{(i-1)(n+m)+2}, \dots, X_{(i-1)(n+m)+n}]^T$$

followed by a gap of m items that are not measured. The mean of the i th sample is then given by

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{(i-1)(n+m)+j}.$$

The control limits are functions of the in-control values of the mean and standard deviation of the sample mean. Letting $\mu_{0,\bar{X}}$ and $\sigma_{0,\bar{X}}$ represent the mean and standard deviation respectively of the distribution of the sample when the process is in-control, the lower (LCL) and upper (UCL) control limits of the \bar{X} chart are given by

$$LCL = \mu_{0,\bar{X}} - k\sigma_{0,\bar{X}} \text{ and } UCL = \mu_{0,\bar{X}} + k\sigma_{0,\bar{X}}.$$

The chart signals at the first sample i in which $\bar{X}_i < LCL$ or $\bar{X}_i > UCL$. The value k is a chart parameter that is selected by the practitioner.

The in-control mean and variance for the sample mean are given by

$$\mu_{0,\bar{X}} = E_0(\bar{X}_i) = \frac{1}{n} \sum_{j=1}^n E_0(X_{(i-1)(n+m)+j}) = \mu_{0,X} = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p}; \text{ and}$$

$$\sigma_{0,\bar{X}}^2 = V_0(\bar{X}_i) = \frac{\gamma_0}{n} + 2 \sum_{j=1}^{n-1} \frac{n-j}{n} \frac{\gamma_j}{n} = \frac{\gamma_0}{n} (1 + 2 \sum_{j=1}^{n-1} \frac{n-j}{n} \rho_j)$$

If the measurements are stochastically independent, then

$$\mu_{0,\bar{X}} = \mu_{0,X}; \text{ and } \sigma_{0,\bar{X}}^2 = V_0(\bar{X}_i) = \frac{\gamma_0}{n}.$$

It now follows that the control limits for the Shewhart \bar{X} chart can be expressed as

$$LCL = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p} - k_D \sqrt{\frac{\gamma_0}{n} \left(1 + 2 \sum_{j=1}^{n-1} \frac{n-j}{n} \rho_j\right)} \text{ and}$$

$$UCL = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p} + k_D \sqrt{\frac{\gamma_0}{n} \left(1 + 2 \sum_{j=1}^{n-1} \frac{n-j}{n} \rho_j\right)}$$

under the autocorrelated data model. Under the independent data model, the control limits are

$$LCL = \phi_0 - k_I \sqrt{\frac{\gamma_0}{n}} \text{ and } UCL = \phi_0 + k_I \sqrt{\frac{\gamma_0}{n}}$$

since $\mu_{0,X} = \phi_0$. If the values of k_I and k_D are related by

$$k_I = \sqrt{1 + 2 \sum_{j=1}^{n-1} \frac{n-j}{n} \rho_j} k_D, \quad (3.2.1)$$

then the two charts are using the same control limits.

We observe that the sample mean can be expressed as

$$\bar{X}_i = \mu_{0,X} + \sigma_a \bar{W}_i,$$

with

$$\bar{W}_i = \frac{1}{n} \sum_{j=1}^n W_{(i-1)(n+m)+j},$$

where the stochastic process $\{W_t\}$ is defined by $W_t = (X_t - \mu_X) / \sigma_a$ (see Section 2.3). It

then follows that the event described by the inequality

$$LCL < \bar{X}_i < UCL$$

can be described by the inequality

$$-k_D < \frac{\bar{W}_i}{\sqrt{\frac{\tau_0}{n} \left(1 + 2 \sum_{j=1}^{n-1} \frac{n-j}{n} \rho_j\right)}} < k_D. \quad (3.2.2)$$

when the process is in-control. For the independence data model, inequality (3.2.2) can be written as

$$-k_D < \frac{\overline{W}_i}{\sqrt{\tau_0/n}} < k_D \quad (3.2.3)$$

Note the two charts are equivalent if k_I and k_D are related by equation (3.2.1).

Inequalities (3.2.2) and (3.2.3) will be useful in simulating estimates for various run length parameters for both the in- and out-of-control cases.

3.3 CUSUM \overline{X} CHARTS

Page (1954) introduced the cumulative sum (CUSUM) \overline{X} control chart. A tabular form of this chart was given by Ewan and Kemp (1960). The plotted statistic C_i^+ at time i for the upper one-sided CUSUM \overline{X} chart is defined by

$$C_0^+ = 0 \text{ and } C_i^+ = \max \left\{ 0, C_{i-1}^+ + \frac{\overline{X}_i - u_{0,\overline{X}}}{\sigma_{0,\overline{X}}} - k_D^+ \right\}.$$

The upper one-sided CUSUM \overline{X} chart signals at sampling stage i if $C_i^+ > h_D^+$. The values $k_D^+ > 0$ and $h_D^+ \geq 0$ are referred to as the (upper) reference value and the (upper) control limit (UCL), respectively. For the lower one-sided CUSUM \overline{X} chart, the plotted statistic C_i^- is defined by

$$C_0^- = 0 \text{ and } C_i^- = \min \left\{ 0, C_{i-1}^- + \frac{\overline{X}_i - u_{0,\overline{X}}}{\sigma_{0,\overline{X}}} + k_D^- \right\}.$$

A signal is given at time i for the lower one-sided CUSUM \overline{X} chart if $C_i^- < h_D^-$. The values of $k_D^- > 0$ and $h_D^- \leq 0$ are the (lower) reference value and (lower) control limit

(LCL), respectively. The values k_D^- , h_D^- , k_D^+ , and h_D^+ are known as chart parameters which are selected by the practitioner. The two-sided CUSUM \bar{X} chart of Page (1954) plots on the same time plot the points (i, C_i^-) and (i, C_i^+) for $i = 1, 2, 3, \dots$. The chart signals at the first sampling stage t if $C_i^- < h^-$ or $C_i^+ > h^+$. Note that the Shewhart \bar{X} chart is a special case of the two-sided CUSUM \bar{X} chart of Page (1954). This is seen by setting $k_D^- = k_D^+ = k$ and $h_D^- = h_D^+ = 0$.

We note here that the statistic $(\bar{X}_i - u_{0,\bar{X}}) / \sigma_{0,\bar{X}}$ can be expressed as

$$\frac{\bar{X}_i - u_{0,\bar{X}}}{\sigma_{0,\bar{X}}} = \frac{\bar{W}_i}{\sqrt{\frac{\tau_0}{n} (1 + 2 \sum_{j=1}^{n-1} \frac{n-j}{n} \rho_j)}}. \quad (3.3.1)$$

when the process is in-control. For the case in which the data are independent and the process is in-control, we have

$$\frac{\bar{X}_i - u_{0,\bar{X}}}{\sigma_{0,\bar{X}}} = \frac{\bar{W}_i}{\sqrt{\tau_0/n}}. \quad (3.3.2)$$

These expressions will be useful obtaining the run length properties of the chart via simulation.

Lucas and Crosier (1982) proposed a head-start for the CUSUM \bar{X} charts. For the upper one-sided chart, the initial value of the CUSUM statistic is set to a value between 0 and h^+ , that is, $0 < C_0^+ \leq h^+$. Similarly, for the lower one-sided chart to have a head-start, the practitioner sets $h^- \leq C_0^+ < 0$.

Crosier (1986) developed a two-sided CUSUM \bar{X} scheme that is based on a single cumulative sum. The plotted statistic C_i is defined by

$$C_i = \left\{ \begin{array}{ll} 0, & -k \leq C_{i-1} + \frac{\bar{X}_i - u_{0,\bar{X}}}{\sigma_{0,\bar{X}}} \leq k \\ C_{i-1} + \frac{\bar{X}_i - u_{0,\bar{X}}}{\sigma_{0,\bar{X}}} + k, & C_{i-1} + \frac{\bar{X}_i - u_{0,\bar{X}}}{\sigma_{0,\bar{X}}} + k < 0 \\ C_{i-1} + \frac{\bar{X}_i - u_{0,\bar{X}}}{\sigma_{0,\bar{X}}} - k, & C_{i-1} + \frac{\bar{X}_i - u_{0,\bar{X}}}{\sigma_{0,\bar{X}}} - k > 0 \end{array} \right.$$

with $C_0 = 0$. The chart signals at time i if $C_i < -h$ or $C_i > h$. The values $k > 0$ and $h \geq 0$ will be referred to as the reference value and the control limit, respectively. A value of the CUSUM statistic greater than h gives evidence the mean may have shifted to the high side and less than $-h$ that the mean has shifted to the low side. We will not consider further the CUSUM \bar{X} of Crosier (1986).

3.4 CONCLUDING REMARKS

It is our intent to study the family of two-sided CUSUM \bar{X} charts of Page (1954) for monitoring for a change in the mean of autocorrelated data. As is well known, the family of Shewhart \bar{X} charts is subfamily of this family and will be discussed as a special case. The members are indexed by the chart parameters.

CHAPTER 4

DESIGNING A CONTROL CHART FOR AUTOCORRELATED DATA

4.1. INTRODUCTION

Methods for designing a control chart are typically based on the run length distribution of the chart. The run length is the number of samples taken until the chart first signals. The most commonly used criterion for evaluating the performance of a chart is based on the average run length (ARL). This method selects the chart parameters such that the chart has a specified in-control ARL and for some out-of-control scenario the ARL is minimized.

We are interested in using the ARL performance of the chart for a different purpose. It is our interest to examine the ARL performance of the chart when the practitioner has decided to take the quality measurements on n consecutive items from the output of the process, not measure the next m consecutive items, measure the next n consecutive items, etc. It is these data that will be used to make a decision about the quality of the process. The question of interest is “What is the minimum value of m for a given value of n for which a chart will have the same in-control ARL performance as expected if there is stochastic independence between samples?” As will be seen, the gap size m is a function of the process parameters. Hence, it will be desirable to determine the smallest gap size m needed over a wide range of process parameters that may occur in applications. Once a chart has been defined for a given sample size and gap size with a given in-control ARL, the chart is examined to see how well it detects various out-of-control scenario.

In this chapter, the design of the Shewhart and CUSUM \bar{X} charts based on samples of size n with gaps between samples of size $m \geq 0$ will be examined. As previously stated, it is of interest to determine the minimum gap size needed such that the ARL of the chart will be the same in-control ARL if the samples were in fact independent. Simulation will be used to study this problem. A MATLAB program, described in the next section, was written to carry out these simulations. In Section 4.3, the results of the simulation study to determine the gap size and chart parameters are reported. A performance study is reported in Section 4.4 for various out-of-control scenarios. Some concluding remarks are made in Section 4.5.

4.2. SIMULATION PROGRAM

A MATLAB program, which is included in the Appendix, was developed to determine each of the following for a given $ARIMA(p,0,q)$ model, chart parameters, and sampling method.

1. Determines the covariance and correlation functions for the given $ARIMA(p,0,q)$ model in terms of the process parameters.
2. Calculates the largest covariance between samples after gapping.
3. Uses simulation to estimate the ARL of the chart(s). The program reports the results.
4. Estimates, using simulation, the chart performance for shifts in the mean and/or variance of the shock model.

In this section, we will give a discussion of the construction of the program followed by some examples of its use.

The program is constructed using the results in Chapters 2 and 3. The method of simulating the time series data is discussed in Section 2.5, the sampling method is discussed in Section 2.6, and the meaning of out-of-control can be found in Section 2.4. The CUSUM \bar{X} chart is described in Section 3.3 which has the Shewhart \bar{X} chart as a special case.

The program prompts the user for the following input. Note that there are some limitations to the program.

- p – the number of AR parameters (0, 1, 2, ... , 10)
- ϕ 's – the AR parameters
- q – the number of MA parameters (0, 1, 2, ... , 10)
- θ 's – the MA parameters
- n and m of the sampling procedure
- δ - the number of in-control standard deviations the mean of the distribution of shocks has shifted with $\delta = 0$ if process is in-control
- $\lambda > 0$ - the ratio of the out-of-control standard deviation of the shocks to the in-control standard deviation with $\lambda = 1$ if the process is in-control
- The chart parameters h and k with $h \geq 0$ and $k > 0$
- The number of simulations

The output of the program consist of

- the covariance matrix of size $n \times n$
- γ_{m+1} - the covariance of $X_{(i-1)(n+m)+n}$ and $X_{(i+1-1)(n+m)+1}$
- simulated estimate of the ARL of the chart

For example, suppose the time series is modeled as an ARIMA(1,0,2) with parameters $\phi_1 = 0.2$, $\theta_1 = 0.1$, $\theta_2 = 0.3$, and $\sigma_{0,a} = 1$. The sampling procedure sets the sample size to $n = 5$ and the gap size to $m = 10$. The CUSUM \bar{X} chart has $k = 1.500$ and $h = 2$. The number of in-control standard deviations the mean of the distribution of shock has shifted is $\delta = 0.2$. The process is considered to be in-control with respect to the variance of the distribution of the shocks, that is, $\lambda = 1$. The number of simulations is selected to be 10,000. Figure 4.2.1 shows the inputs to the program and the program's outputs.

```

EDU>> ARMACUSUM
Enter p (0-10): 1
Enter next Phi value: .2
Enter q (0-10): 2
Enter next Theta value: .1
Enter next Theta value: .3
Enter Delta value: .2
Enter Lambda value: 1
Enter n: 5
Enter m: 10
Enter h: 2
Enter k: 1.500
Enter number of simulations: 10000
Variance-Covariance Matrix n by n:
  1.0417  0.2083  0.0417  0.0083  0.0017
  0.2083  1.0417  0.2083  0.0417  0.0083
  0.0417  0.2083  1.0417  0.2083  0.0417
  0.0083  0.0417  0.2083  1.0417  0.2083
  0.0017  0.0083  0.0417  0.2083  1.0417
Gamma(m+1):
  2.1333e-008
Chart ARL:
  71.8564

```

Figure 4.2.1

4.3 SIMULATION STUDY TO DETERMINE GAP SIZE

It is well known that sample size n affects the performance of a chart under the independent data model whereas the gap size has no affect. In this section, we will demonstrate that gap size also affects the performance of a chart when the data are autocorrelated. This will be illustrated through a series of examples. The simulation program described in Section 4.2 is used to estimate the ARL. In our first example, we give the design of a Shewhart \bar{X} chart under the independent normal data model and then examine how this chart performs under an autocorrelated normal data model with the same process variance as the independent data model.

The Shewhart \bar{X} chart is discussed in Section 3.2. For the case in which $\mu_{0,\bar{X}} = 0$ and $\sigma_{0,\bar{X}} = 1.3159/\sqrt{n}$. The control limits for the chart are

$$LCL = -k_I \frac{1.3159}{\sqrt{n}} \text{ and } UCL = k_I \frac{1.3159}{\sqrt{n}}. \quad (4.3.1)$$

If $k_I = z_{0.005} = 2.576$, then the ARL of the chart under the independent normal model is 100 regardless of the gap size m . While the out-of-control ARL is a function of the sample size n , it also does not depend on the gap size m . Further, it is not difficult to show that as the sample size n increases the ARL for a given shift decreases. That is, the ability of the chart to detect a given shift in the mean requires, on average, fewer samples as n increases.

The performance of the Shewhart \bar{X} chart defined in (4.3.1) is then examined if it were used to monitor an ARIMA(1,0,0) process with parameters $\phi_0 = 0$, $\phi_1 = 0.65$, and

$\sigma_{0,a} = 1$. Under this model, the control limits for the Shewhart \bar{X} chart that are the same as the one defined in (4.3.1) designed under the independent data model are

$$\begin{aligned} LCL &= -k_D \sqrt{1 + 2 \sum_{j=1}^n \frac{n-j}{n} 0.65^j} \frac{1.3159}{\sqrt{n}} = -k_I \frac{1.3159}{\sqrt{n}} \text{ and} \\ UCL &= k_D \sqrt{1 + 2 \sum_{j=1}^n \frac{n-j}{n} 0.65^j} \frac{1.3159}{\sqrt{n}} = k_I \frac{1.3159}{\sqrt{n}} \end{aligned} \quad (4.3.2)$$

with

$$k_I = k_D \sqrt{1 + 2 \sum_{j=1}^n \frac{n-j}{n} 0.65^j} . \quad (4.3.3)$$

Note that

$$g(n) = \sqrt{1 + 2 \sum_{j=1}^n \frac{n-j}{n} 0.65^j} = \sqrt{\frac{1}{1-0.65} - \frac{1}{n} \frac{0.65(1-0.65^n)}{(1-0.65)^2}}$$

is an increasing function of n with limit 10 as $n \rightarrow \infty$. Thus, for the charts whose control limits are defined in (4.3.1) and (4.3.2) to have the same control limits with k_I fixed, then the chart parameter k_D must be decreased as n increases. It would then be expected that if the data are autocorrelated and the chart defined in (4.3.2) is used to monitor for a change in the mean, then the ARL would decrease as n increases. This is illustrated in Table 4.3.1. with this example for a gap size of $m = 0$. The results are based on 10,000 simulated time series using the program described in Section 4.2.

Table 4.3.1: $\phi_1 = 0.65$, $\sigma_{0,a} = 1$, $m = 0$, $k = 2.575$, $N_{sim} = 10000$

n:	1	2	3	4	5	6	7	8	9	10
ARL:	127	28	15	11	9	8	7	7	7	6

Similar results hold for other values of $-1 < \phi_1 < 1$.

In the case that $-1 < \phi_1 < 1$ for the ARIMA(1,0,0) model, it can be shown that the correlation between the i^{th} and the $(i+r)^{\text{th}}$ sample means is given by

$$\text{corr}(\bar{X}_i, \bar{X}_{i+r}) = \frac{1}{n^2} \phi_1^{r(n+m)-1} \left(\frac{1 - \phi_1^n}{1 - \phi_1^2} \right)^2. \quad (4.3.3)$$

The correlation between the sample means is one measure of the dependence between samples. Another measure is the correlation between the n^{th} value of the i^{th} sample and the first value of the $(i+r)^{\text{th}}$ sample given by

$$\text{corr}(X_{(i-1)(n+m)+n}, X_{(i+r-1)(n+m)+1}) = \phi_1^{r(n+m)-n+1} \quad (4.3.4)$$

for the ARIMA(1,0,0). We see that as n (for $r > 1$), m , and/or r increases the correlation between the samples as measured by both the correlations in (4.3.3) and (4.3.4) decreases towards zero. It then follows that for a fixed value of n the larger the gap size m the weaker the correlation between the samples. In fact, it is not difficult to show that for ε (small) there exist values of m such that $|\text{corr}(\bar{X}_i, \bar{X}_{i+r})| < \varepsilon$ for fixed values of n and r . A similar statement can be made for the correlation in (4.3.4).

For example, for $n = 2$, $r = 1$, and $\varepsilon = 10^{-5}$, the values of m in which $|\text{corr}(\bar{X}_i, \bar{X}_{i+r})| < \varepsilon$ are at least 26. While the sample means are not uncorrelated, under the normal model their joint distribution is approximately the product of their marginal when the correlation is relatively close to zero. Thus, with respect to the value of ε , the samples are approximately independent. Similarly, the values of m that are solution to the inequality $|\text{corr}(X_{(i-1)(n+m)+n}, X_{(i+r-1)(n+m)+1})| < \varepsilon$ are at least 28.

Hence, for “relatively large” values of m , we would expect the ARL to be approximately the same. This is illustrated for this example with a sample of size $n = 2$ by the results in Table 4.3.2 and Figure 4.3.1. As we see in this example, the in-control ARL

Table 4.3.2: $\phi_1 = 0.65$, $n = 2$, $k_D = 2.575$, $N_{sim} = 10000$

m :	0	1	2	3	4	5	6	7	8
ARL:	26.5	23.9	23.1	22.5	22.5	22.2	22.4	22.1	22.4

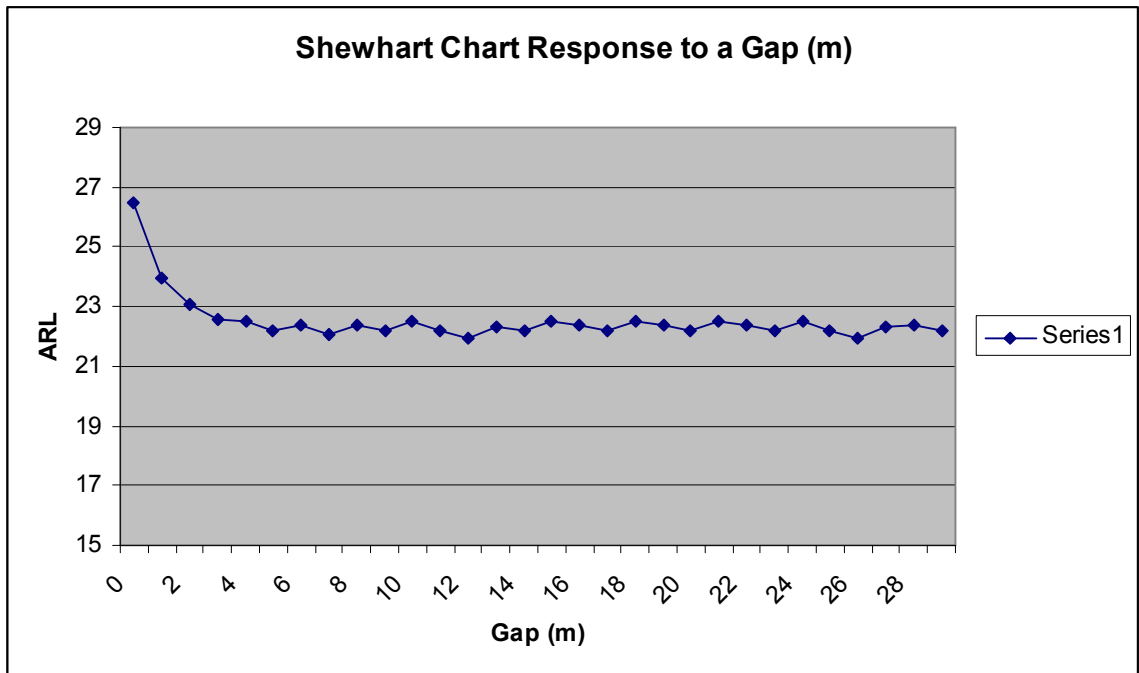


Figure 4.3.1: $\phi_1 = 0.65$, $n = 2$, $k = 2.575$, $N_{sim} = 10000$

As illustrated in both Table 4.3.2 and Figure 4.3.1, the ARL begins to level off beginning with a gap of $m = 6$ (this is an estimated gap size since the ARL was determined using simulation). Thus, examining either $\text{corr}(\bar{X}_i, \bar{X}_{i+r})$ or $\text{corr}(X_{(i-1)(n+m)+n}, X_{(i+r-1)(n+m)+1})$ for determining the gap size relative to the ARL may suggest much larger gap sizes than needed for the ARL to level off. Thus, the minimum gap size will be determined using the ARL. The aforementioned correlations will be used when necessary to place an upper bound on the minimum gap size.

To determine the gap size for a chart that has an in-control ARL of 100, we use an iterative method. A control chart is selected and the minimum gap size is estimated for which the in-control ARL levels off. This procedure of selecting a chart and finding the minimum gap size for which the ARL levels off is repeated until a chart is found for which the estimated in-control ARL has leveled off at 100. Table 4.3.3 for the given process gives the illustrated how the ARL changes as the k_D changes. It is interesting to note that in each case the value of minimum value of m is 6.

Table 4.3.3: $\phi_1 = 0.65$, $n = 2$, $m = 6$; $N_{sim} = 10000$

k_D :	3.285	3.295	3.3087	3.310	3.320
ARL:	95	98	100	102	104

Although the reported results in the previous example are for a sample size of $n = 2$, results of simulation results not reported here suggest that the gap size m does not

depend on the sample size n . On the other hand, the gap size is affected by the value of ϕ_1 . Figure 4.3.2 gives a plot of gap size m versus ϕ_1 with $|\phi_1| \leq 0.75$.

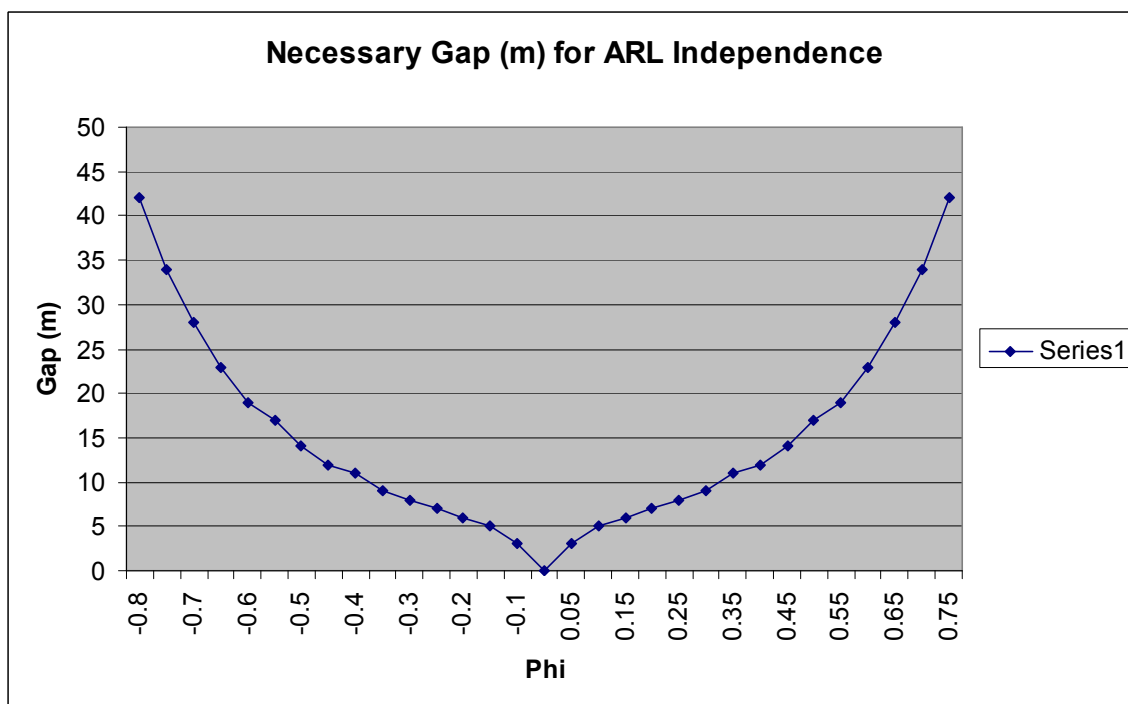


Figure 4.3.2: $|\phi_1| \leq 0.75$, $n = 1, 2, 3, \dots$

We include the results for $0.75 < |\phi_1| < 1$. These results are reported in Table 4.3.5.

Table 4.3.4: $0.75 < |\phi_1| < 1$; $n = 1, 2, 3, \dots$

ϕ_1 :	-0.99	-0.95	-0.90	-0.85	-0.80	0.80	0.85	0.90	0.95	0.99
m:	1535	269	125	78	56	56	78	125	269	1535

It is apparent from Figure 4.3.2 and Table 4.3.5 the gap size m is a function of $|\phi_1|$.

In our next example, we illustrate how the model affects gap size. Consider an ARIMA(1,0,1) model with parameters, $\phi_0 = 0$, $\phi_1 = 0.80$, $\theta_1 = -0.45$, and a sample size of $n = 4$. We have that the correlation $\text{corr}(X_{(i-1)(n+m)+n}, X_{i(n+m)+1})$ between the last value in sample $i-1$ and sample i correlation ρ_{m+1} of the ARIMA(1,0,1) model. Using the covariances given in Equation (2.3.9) for the ARIMA(1,0,1) model, the correlations function ρ_{m+1} is given by

$$\rho_{m+1} = \frac{\phi_1^m (\phi_1 - \theta_1)(1 - \phi_1 \theta_1)}{1 - 2\phi_1 \theta_1 + \theta_1^2}$$

for $m \geq 0$. Solving the inequality

$$|\text{corr}(X_{(i-1)(n+m)+n}, X_{(i+r-1)(n+m)+1})| < \varepsilon$$

for m , we have

$$m \geq \frac{\ln(\varepsilon) - \ln(|\rho_1|)}{\ln(|\phi_1|)}.$$

These results suggest that the minimum value of m for this example is 52.

On the other hand, the in-control ARL leveled off at an estimated gap size of $m = 8$.

Using the method in the previous example, we found that for the Shewhart \bar{X} chart with estimated values of $k = 4.682$ and $m = 8$ the in-control ARL of leveled off at 100. This is significantly different from the necessary k used in the previous example which supports that the chart and gap size are process dependent.

4.4 PERFORMANCE STUDY

Having selected a charting procedure (sampling method and chart parameters) under a given $ARIMA(p,0,q)$ model, it is of interest to study its out-of-control performance. We limit our study to only changes in the mean of the shocks beginning at time 1. This out-of-control scenario is discussed in Section 2.6. That is, we are assuming that the shock a_t follows a $N(\delta\sigma_{0,a}, \sigma_{0,a}^2)$. The $ARIMA(1,0,0)$ example in the previous section is used to illustrate the affect of a shift in the process. Table 4.3.6 gives the out-of-control ARL for the described Shewhart \bar{X} chart. For example, we see from Table 4.3.6 that the in-control ARL is 100 and the out-of-control ARL is 32 for a shift of $0.3\sigma_{0,a}$.

Table 4.4.1: $\phi_1 = 0.65$, $n = 2$, $m = 6$, $k = 3.3087$, $N_{sim} = 10000$

δ :	0	0.1	0.2	0.3	0.4	0.5
ARL:	100	83	53	32	21	14

A CUSUM \bar{X} chart was also designed to have an in-control ARL with a gap size of 6. The performance of this chart is illustrated in Table 4.3.7. We see that for a shift of $0.3\sigma_{0,a}$ the out-of-control ARL of the chart is 15.

Table 4.4.2: $\phi_1 = 0.65$, $n = 2$, $m = 6$, $h = 4$, $k = 0.745$, $N_{sim} = 10000$

δ :	0	0.1	0.2	0.3	0.4	0.5
ARL:	100	59	26	15	10	7

It is clear in this example, that the CUSUM \bar{X} chart out performs the Shewhart \bar{X} chart for small to moderate shifts in the mean. In the independent data case, this is a well known result. However, it is not known if this holds in general in the dependent data case. Further study would be needed to answer this question.

4.5 CONCLUDING REMARKS

A program that was written to estimate the ARL of CUSUM \bar{X} chart for monitoring the mean of an ARIMA($p,0,q$) process was discussed. A method for selecting a chart and the corresponding sampling method that was based on the ARL was outlined and illustrated using examples.

CHAPTER 5

CONCLUSION

5.1. GENERAL CONCLUSIONS

There is a considerable amount of literature on the performance of various control charts when the data are independent. It has been shown that the data generated by some industrial processes do not follow the independent data model. We have examined a method for designing a chart, in particular, the CUSUM \bar{X} chart when the data are autocorrelated following an ARIMA($p,0,q$) model assuming in-control parameters are known. This method is useful in designing a chart if the practitioner takes a sample from an autocorrelated process in the order of output and waits a period of time to take the next sample.

5.2. AREAS FOR FURTHER RESEARCH

A method was presented for simulating a time series of quality measurements on the output of a production process when the data are assumed to follow an ARIMA($p,0,q$) data model. The initial values need to start the series are simulated. The method allows for the simulation of both in- and out-of-control process data. Other authors have discussed methods for “warming-up” the generator before simulating the time series. A comparison of these methods would be of interest. This would require programs that implement the competing methods.

Little work has been done in the development of Phase I control charts for processes with autocorrelated quality measurements. Maragah and Woodall (1992) have

studied Phase I charts designed for independent data when in fact the data are autocorrelated. We would like to study Phase I charts that would be useful for autocorrelated data.

The point in time at which the process changes from an in-control state to an out-of-control state is of interest to the practitioner for determining the reason for the change. Various methods can be found in the literature for determining the change point for the independent data model. We are interested in examining the usefulness of these methods under the dependent data models.

When the types of shifts we have studied occur, this affects the distribution of the quality measurement. The process mean and the process autocovariances become functions of time. It is our intent to give closed formed expressions for these functions as functions of time.

It has been recommended that along with the ARL of the chart one should examine the standard deviation of the run length (SDRL) as well as percentage points of the run length distribution. Simulation can be used to obtain estimates of these parameters. As with the ARL, we plan also to examine how gapping affects these run length parameters.

Each sample can be treated as an n -variate quality measurement. The various quality control charts for multivariate data can then be designed to monitor for a shift in the mean of the quality measurement. Similar questions can be answered about gap size and how well these charts perform when the process is out-of-control.

Methods similar to those given in this paper could be developed for a multivariate quality measurement. A family of models useful in modeling multivariate time series data

is the multivariate ARIMA (MARIMA) models. These models are discussed by Box and Jenkins (1994), Brockwell and Davis (1987), and Wei (1990), among others.

While simulation has been used to estimate the run length distribution, analytical methods based on integral equations have been developed for particular time series models. It is our interest to extend these methods to the more general ARIMA($p,0,q$) models.

As demonstrated in Section 4.3, the chart including the gap size is process dependent. It is our interest to study this relationship between a process and the charting procedure to be used for the purpose of selecting the chart given the process.

REFERENCES

- Alwan, L.C. (1992), "Effects of Autocorrelation on Control Chart Performance," *Communication in Statistics Theory and Methods* **21**, 1025-1049.
- Alwan, L.C. and Roberts, H.V. (1988), "Time Series Modeling for Statistical Process Control," *Journal of Business and Economic Statistics* **6**, 87-95.
- Anderson, O.D. (1979), "On Warming-Up Time Series Simulations Generated by Box-Jenkins Models", *Journal of the Operational Research Society* **30**, 587-589.
- Ansley, C.F. (1980), "Computation of the Theoretical Autocovariance Function for a Vector ARMA Process," *Journal of Statistical Computation and Simulation* **12**, 15-24.
- Apley, Daniel W. and Tsung, Fugee (2002), "The Autoregressive T^2 Chart for monitoring Univariate Autocorrelated Processes," *Journal of Quality Technology* **34**, 80-96.
- Bagshaw, M. and Johnson, R.A. (1977), "Sequential Procedures for Detecting Parameter Changes in a Time-Series Model," *Journal of the American Statistical Association* **72**, 593-597.
- Beall, C.W. (1982), "A Regression Technique for Determining Steady State Conditions in Time Series Simulations," *1982 Winter Simulation Conference of the Institute of Electrical and Electronic Engineers Inc, New York*, 439-448.
- Box, G.E.P., and Jenkins, G.M. (1994), *Time Series Analysis, Forecasting and Control, Third Edition*, Prentice Hall: Englewood Cliffs, NJ.
- Box, G. and Luceño, A (1997), *Statistical Control: By Monitoring and Feedback Adjustment*. Wiley-Interscience, 1st edition.

- Brockwell, P.J. and Davis, R.A. (1987), *Time Series: Theory and Methods*, Springer-Verlag: New York.
- Burn, D.A. (1986), "A Johnson Curve Approach to Warming Up Time Series Simulation," *Computer Science and Statistics: Proceedings of the 18th Symposium on the Interface of the American Statistical Association*, T.J. Boardman, Editor, 302-306.
- Burroughs, T.E. (1993), "Evaluating Quality Control Charts with No Standards Given," Masters Thesis, Department of Mathematics and Statistics, Southern Illinois University at Edwardsville, Edwardsville, IL.
- Burroughs, T.E., Rigdon, S.E., and Champ, C.W. (1993), "An Analysis of Shewhart Charts with Runs Rules When No Standards Are Given," *Proceedings of the Quality and Productivity Section of the American Statistical Association*, August 8-12, San Francisco, CA, 16-19.
- Burroughs, T.E., Rigdon, S.E., and Champ, C.W. (1995), "Analysis of the Shewhart S-Chart with Runs Rules When No Standards Are Given," *Proceedings of the Twenty-Sixth Annual Meeting of the Midwest Decision Sciences Institute*, May 4-6, St. Louis, Missouri, 268-270.
- Champ, C.W. and Jones-Farmer, L.A. (2007), "Properties of Multivariate Control Chart with Estimated Parameters," *Sequential Analysis* **26**, 153-169.
- Champ, C.W. and Woodall, W.H. (1987), Exact Results for Shewhart Control Charts with Supplementary Runs Rules, *Technometrics*, **29**.
- Champ, C.W., Woodall, W.H., and Mohsen, H.A. (1991), "A Generalized Quality Control Procedure," *Statistics and Probability Letters*, **11**, 211-218.

- Chen, G. (1997), "The Mean and Standard Deviation of the Run Length Distribution of \bar{X} Charts when Control Limits are Estimated," *Statistica Sinica* **7**, 789-798.
- Crosier, R.B. (1986), "A New Two-Sided Cumulative Sum Quality Control Scheme," *Technometrics* **28**, 187-194.
- Crosier, R. B. (1988). Multivariate Generalizations of Cumulative Sum Quality-Control Schemes, *Technometrics* **30**: 291-303.
- Duncan, A.J. (1986), *Quality Control and Industrial Statistics*, Fifth Edition, Irwin: Homewood, IL.
- Ewan, W. D. and Kemp, K. W., (1960), Sampling inspection of continuous processes with no autocorrelation between successive results. *Biometrika*, **47**, 363-380.
- Ghosh, B. K., Reynolds, M. R., Jr., and Hui Y. V.(1981), "Shewhart \bar{X} -Charts with Estimated Process Variance," *Communications in Statistics-Theory and Methods*, A10, 1797-1822.
- Healy, J.D. (1987). A Note on Multivariate CUSUM Procedures, *Technometrics* 29: 409-412.
- Hotelling, H. (1947). Multivariate Quality Control Illustrated by Air Testing of Sample Bombsights, in Selected Techniques of Statistical Analysis, Eisenhart,C., Hastay, H., Wallis, W.A., (eds), McGraw-Hill: New York: 111-184
- Jenkins, G.M. and Watts, D.G. (1968), *Spectral Analysis and Its Applications*, Holden-Day: San Francisco.
- Jones, L.A., Champ, C.W., and Rigdon, S.E. (2001), "The Performance of Exponentially Weighted Moving Average Charts With Estimated Parameters," *Technometrics* **43**, pp. 156-167.

- Jones, L.A., Champ, C.W., and Rigdon, S.E. (2004), "The Run Length Distribution of the CUSUM Chart with Estimated Parameters," *Journal of Quality Technology* **36**, 95-108
- Krieger, C.A. (1992). "A CUSUM Quality Control Procedure for Monitoring an ARMA(p,q) Process," Masters Theses, Department of Mathematics and Statistics, Southern Illinois University, Edwardsville, Illinois.
- Krogstad, H.E. (1989), Simulation of Multivariate Gaussian Time Series," *Communications Statistics* **18**, 929-941.
- Lowry, C.A., Woodall, W. H., Champ, C. W., and Rigdon, S.E. (1992), A Multivariate Exponentially Weighted Moving Average Control Chart, *Technometrics* 34: 46-53.
- Lu, C.W. and Reynolds, M.R., Jr. (1999), "EWMA Control Charts for Monitoring the Mean of Autocorrelated Processes," *Journal of Quality Technology* 31, 166-188.
- Lucas, J. M. and Crosier, R. B. (1982), "Fast Initial Response for Cusum Quality Control Schemes: Give You Cusum a Head Start," *Technometrics* **24**, 199-205.
- Lucas, J.M., and Saccucci, M.S. (1990), "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements (with discussions)," *Technometrics*, **32**, 1-29.
- Maragah, H.D. and Woodall, W.H. (1992), "The Effect of Autocorrelation on the Retrospective \bar{X} -Chart," *Journal of Statistical Computation and Simulation* **40**, 29-42.

- Mastrangelo, C.M. and Forrest, D.R. (2002), "Multivariate Autocorrelated Processes: Data and Shift Generation," *Journal of Quality Technology* **34**, 216-220.
- McLeod, I. (1975), "Derivation of the Theoretical Autocovariance Function of Autoregressive Moving Average Time Series," *Applied Statistics* **24**, 255-256 (Corr: 77V26, pp 194).
- Montgomery, D.C. (1991), *Statistical Quality Control*, Wiley: New York.
- Montgomery, D.C. and Mastrangelo, C.M. (1991), "Some Statistical Process Control Methods for Autocorrelated Data," *Journal of Quality Technology* **23**, 170-193.
- Ng, C. H. and Case, K. E. (1992), "Control Limits and the ARL: Some Surprises." *First Industrial Engineering Research Conference Proceedings*, Norcross, GA: Institute of Industrial Engineers, 127-129.
- Page, E.S. (1954), Continuous Inspection Schemes, *Biometrika*, **41**, 100-115.
- Pate, M.B. and Davies, N. (1988), "Algorithm AS 232: Computation of Population and Sample Correlation and Partial Correlation Matrices in MARMA(p,q) Time Series **37**, 127-138.
- Pignatiello, J. J. and Runger, G. C. (1990), "Comparisons of Multivariate CUSUM Charts," *Journal of Quality Technology* **22**: 173-186.
- Quesenberry, C. P. (1993), "The Effect of Sample Size on Estimated Limits for \bar{X} and X Control Charts," *Journal of Quality Technology* **25**, 237-247.
- Roberts, S.W. (1959), "Control Chart Tests Based on Geometric Moving Averages," *Technometrics* **1**, 239-250.

- Runger, G.C. (2002), Assignable causes and autocorrelation: Control charts for observations or residuals. *Technometrics* **34**, 165–170
- Runger, G.C. and Prabhu, S.S. (1996). A Markov Chain Model for the Multivariate Exponentially Weighted Moving Average Control Chart, *Journal of the American Statistical Association* **91**: 1701-1706.
- Runger, G. C. and Willemain, T. R. (1995), Model-based and model-free control of autocorrelated processes. *Journal of Quality Technology* **27**, 283-292.
- Runger, Willemain, and Prabhu (1995), Average run lengths for CUSUM control charts applied to residuals. *Communication in Statistics-Theory and Methods* **24**, 273-282.
- Shewhart, W.A. (1931), *Economic Control of Quality of Manufactured Product*, Van Nostrand-Reinhold, NY.
- Wardell, D. G., Moskowitz, H. and Plante, R. D. (1994), “Run Length Distributions of Residual Control Charts for Autocorrelated Processes,” *Journal of Quality Technology* **26**, 308-317.
- Wei, W.S. (1990), *Time Series Analysis: Univariate and Multivariate Methods*, Addison-Wesley: Reading, MA.
- Zhang, N. F. (1998). "A Statistical Control Chart for Stationary Process Data," *Technometrics*, **40**, 24-38.

APPENDIX A

MATLAB PROGRAM

```

%ARMACUSUM.m
%This program prompts the user for process parameters, sampling
procedure,
%and CUSUM chart paramters. It outputs the Variance-Covariance matrix,
%Gamma(m+1) and the estimated ARL.
p=input('Enter p (0-10): ');
clear phi
for i=1:p
    phi(p)=input('Enter next Phi value: ');
end
phi(1,p+1:10)=0;
q=input('Enter q (0-10): ');
clear theta
for i=1:q
    phi(q)=input('Enter next Theta value: ');
end
theta(1,p+1:10)=0;

d=input('Enter Delta value: ');
lam=input('Enter Lambda value: ');
n=input('Enter n: ');
m=input('Enter m: ');
h=input('Enter h: ');
ks=input('Enter k: ');
sim=input('Enter number of simulations: ');

%Solve equation 1-phi1*B-phi2*B^2..... to check for stationarity
phi_check=['1-(' num2str(phi(1)) ') *x+(' num2str(phi(2)) ...
    ') *x^2+(' num2str(phi(3)) ') *x^3+(' num2str(phi(4)) ...
    ') *x^4+(' num2str(phi(5)) ') *x^5+(' num2str(phi(6)) ...
    ') *x^6+(' num2str(phi(7)) ') *x^7+(' num2str(phi(8)) ...
    ') *x^8+(' num2str(phi(9)) ') *x^9+(' num2str(phi(10)) ...
    ') *x^10=0'];
sss=(solve(phi_check));
sss=double(sss);
ssr=real(sss);
ssi=imag(sss);
%Check that solutions are outside of complex unit circle
ggg=1;
tf=1;
[xx zz]=size(sss);
while and(tf==1,ggg<=xx)
    cir=((ssr(ggg))^2+(ssi(ggg))^2);
    if cir>1
        tf=1;
        ggg=ggg+1;
    else
        tf=0;
    end
end

```

```

end

ss=1;
%CREATE MATRIX Aa
for j=1:p+1
    for i=1:p+1
        if and ((j==1),(i==1))
            aa(i,j)=0;
        else
            if i+j<=p+2
                aa(i,j)=phi(i+j-2);
            else
                aa(i,j)=0;
            end
        end
    end
end
end

for j=2:p+1
    for i=3:p+1
        if i>j
            bb(i,j)=phi(i-j);
        else
            bb(i,j)=0;
        end
    end
end

if p>=3
A=aa+bb;
else
    A=aa;
end
B=0;
%CREATE MATRIX B
for j=1:p+1
    for i=1:p+1
        if and ((j==1),(i==1))
            B(i,j)=1;
        else
            if i+j<=p+2
                B(i,j)=-theta(i+j-2);
            else
                B(i,j)=0;
            end
        end
    end
end
end

E(1)=ss;

if p+1>1

```



```

E(2)=(phi(1)-theta(1));
end

if p+1>2
E(3)=(phi(2)-theta(2))+phi(1)*E(2);
end

if p+1>3
E(4)=(phi(3)-theta(3))+phi(1)*E(3)+phi(2)*E(2);
end

if p+1>4
E(5)=(phi(4)-theta(4))+phi(1)*E(4)+phi(2)*E(3)+phi(3)*E(2);
end

if p+1>5
E(6)=(phi(5)-theta(5))+phi(1)*E(5)+phi(2)*E(4)+phi(3)*E(3) ...
    +phi(4)*E(2);
end

if p+1>6
E(7)=(phi(6)-theta(6))+phi(1)*E(6)+phi(2)*E(5)+phi(3)*E(4) ...
    +phi(4)*E(3)+phi(5)*E(2);
end

if p+1>7
E(8)=(phi(7)-theta(7))+phi(1)*E(7)+phi(2)*E(6)+phi(3)*E(5) ...
    +phi(4)*E(4)+phi(5)*E(3)+phi(6)*E(2);
end

if p+1>8
E(9)=(phi(8)-theta(8))+phi(1)*E(8)+phi(2)*E(7)+phi(3)*E(6) ...
    +phi(4)*E(5)+phi(5)*E(4)+phi(6)*E(3)+phi(7)*E(2);
end

if p+1>9
E(10)=(phi(9)-theta(9))+phi(1)*E(9)+phi(2)*E(8)+phi(3)*E(7) ...
    +phi(4)*E(6)+phi(5)*E(5)+phi(6)*E(4)+phi(7)*E(3)+phi(8)*E(2);
end

if p+1>10
E(11)=(phi(10)-theta(10))+phi(1)*E(10)+phi(2)*E(9)+phi(3)*E(8) ...
    +phi(4)*E(7)+phi(5)*E(6)+phi(6)*E(5)+phi(7)*E(4)+phi(8)*E(3) ...
    +phi(9)*E(2);
end

g=(eye(p+1,p+1)-A)^-1*B*E';
G=0;
p_st=0;
%CREATE MATRIX G
if and(p==0, q==0)
    check=0;
else
for j=1:p
    for i=1:p

```

```

        if (j==i)
            G(i,j)=g(1);
        else
            if i>j
                G(i,j)=g(i-j+1);
            else
                G(i,j)=0;
            end
        end
    end
end
end
end
G=G+G'-g(1)*eye(p,p);
[V D]=eig(G);
p_st=D^(1/2)*V;
end

%Generate Gammas to compute control chart variance
bg(1:(n+m+1),1)=0;
for i=1:(n+m+1)
    if i<=p+1
        bg(i)=g(i);
    else
        for i4=1:p
            bg(i)=bg(i)+phi(i4)*bg(i-i4);
        end
    end
end
end
%Put Gammas in a MATRIX BG
clear BG
for j=1:n
    for i=1:n
        if (j==i)
            BG(i,j)=bg(1);
        else
            if i>j
                BG(i,j)=bg(i-j+1);
            else
                BG(i,j)=0;
            end
        end
    end
end
end
end

%VARIANCE-COVARIANCE MATRIX
BG=BG+BG'-bg(1)*eye(n,n);

%SIMULATION LOOP
%Standard deviation for stopping criteria: summed BG (variance-
covariance
%matrix)
ss=0;
for i5=1:n
    for j5=1:n
        ss=ss+BG(i5,j5);
    end
end
end

```

```

    end
end
ss=((ss)^.5)/(n);

%USE THIS FOR INDEPENDENT CHART ASSUMPTION. If dependent assumption
used,
%put a % sign in front of the next line to comment the line out.
ss=((BG(1,1)/n)^.5);

for t=1:sim
%Generate random z's for x1 and x2 jointly distributed
z_gen=randn(p,1);
%Actual generation of x1, x2, ...,xp
x_gen=p_st*z_gen;

for i2=1:p
    x(i2)=x_gen(i2,1);
end
a=0;
%Generate random a_t's to use for generating series
for i3=1:20%%
    a(i3)=randn(1,1);
end

k=1;
Cp=0;
Cn=0;
%Begin generating time series values
while (k<25000)
    for i=(k-1)*m+(k-1)*n+1:(k-1)*m+(k-1)*n+n+m
        if (i<p+1)
            a(i+10)=a(i+10);
            x(i)=x(i);
        else
            a(i+10)=randn(1,1)*lam+d;
        x(i)=0;
        for i4=1:p
            x(i)=x(i)+phi(i4)*x(i-i4);
        end
        for i4=1:q
            x(i)=x(i)-theta(i4)*a(i-i4+10);
        end

        x(i)=x(i)+a(i+10);

    end
end

MMM(k)=mean(x((k-1)*m+(k-1)*n+1:(k-1)*m+(k-1)*n+n));
%
    if or(Cp > h,Cn <-h)
        kk=k;
        k=1000000;
    else
        Cp=max(0,Cp+(MMM(k))/(ss)-ks);
    end
end

```

```
        Cn=min(0,Cn+(MMM(k))/(ss)+ks);
        k=k+1;
        kk=k;
end
%_____
end
arl(t)=kk;
end

%Calculate mean ARL
ARL=mean(arl(1:t));
Gamma_gap=bg(m+2);

%Output Variance-Covariance Matrix n by n
display(' ')
display('Variance-Covariance Matrix n by n:')
BG
%Output largest covariance in samples after gap
display('Gamma(m+1):')
Gamma_gap
%Output ARL
display('Chart ARL:')
ARL
```
