MULTIVARIATE EWMA CONTROL CHART AND APPLICATION TO A SEMICONDUCTOR MANUFACTURING PROCESS

Multivariate EWMA Control Chart and Application to a Semiconductor Manufacturing Process

by

Ick Huh

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Master of Science

McMaster University

MASTER OF SCIENCE (2010)
(Statistics)

McMaster University
Hamilton, Ontario

TITLE:	Multivariate EWMA Control Chart and Application to a Semiconductor Manufacturing Process
AUTHOR:	Ick Huh (McMaster University, Canada)
SUPERVISOR:	Dr. Román Viveros-Aguilera
NUMBER OF PAGES:	xi, 92

Abstract

The multivariate cumulative sum (MCUSUM) and the multivariate exponentially weighted moving average (MEWMA) control charts are the two leading methods to monitor a multivariate process. This thesis focuses on the MEWMA control chart. Specifically, using the Markov chain method, we study in detail several aspects of the run length distribution both for the on- and off- target cases. Regarding the on-target run length analysis, we express the probability mass function of the run length distribution, the average run length (ARL), the variance of run length (VRL) and higher moments of the run length distribution in mathematically closed forms. In previous studies, with respect to the off-target performance for the MEWMA control chart, the process mean shift was usually assumed to take place at the beginning of the process. We extend the classical off-target case and introduce a generalization of the probability mass function of the run length distribution, the ARL and the VRL. What Prabhu and Runger (1996) proposed can be derived from our new model. By evaluating the off-target ARL values for the MEWMA control chart, we determine the optimal smoothing parameters by using the partition method that provides an easy algorithm to find the optimal smoothing parameters and study how they respond as the process mean shift time changes. We compare the ARL performance of the MEWMA control chart with that of the multivariate Shewhart control chart to see whether the MEWMA chart is still effective in detecting a small mean shift as the process mean shift time changes. In order to apply the model to semiconductor manufacturing processes, we use a bivariate normal distribution to generate sample data and compare the MEWMA control chart with the multivariate Shewhart control chart to evaluate how the MEWMA control chart behaves when a delayed mean shift happens. We also apply the variation transmission model introduced by Lawless et al. (1999) to the semiconductor manufacturing process and show an extension of the model to make our application to semiconductor manufacturing processes more realistic. All the programming and calculations were done in ***R***

Key words: *Multivariate Exponentially Weighted Moving Average Control Chart; Multivariate Shewhart Control Chart; Average Run Length; Markov Chain; Optimal Smoothing Parameter; Semiconductor Manufacturing.*

Acknowledgement

First and foremost, I would like to express the deepest gratitude to my supervisor, Professor Román Viveros-Aguilera, for his wisdom, immense knowledge and encouragement throughout my thesis work. Without his thoughtful guidance, this work would not have been successful. I would like to acknowledge valuable advice from other members of my committee, Professor Aaron Childs and Professor Sui Feng. Also special thanks to Professor N. Balakrishnan for giving me helpful suggestions which have been useful for my future planning.

I thank the Department of Mathematics and Statistics at McMaster University for providing financial support. I am grateful to all my friends for their support. Last, but not the least, I would like to thank my family members, especially my wife, Vivian for supporting and encouraging me to pursue this degree. Without her support, none of this would have been possible.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Control charts and the mutivariate exponentially weighted moving average (MEWMA)

A multivariate control chart is an important tool for monitoring and improvement of the quality of products. In recent years, the importance of multivariate control charts has increased because more quality features are measured in mass production than ever before. These quality measures often exhibit substantial cross-correlations. For example, in semiconductor manufacturing, manufacturers make semiconductor devices around the clock through hundreds of processes. In this case, it would be more efficient to maintain a multivariate control chart than several univariate control charts because it is possible that individual control charts might not detect an assignable cause when quality characteristics are dependent. Several multivariate quality control charts have been proposed to monitor the mean vector of quality characteristics. The three most common multivariate control charts are the multivariate cumulative sum (MCUSUM) control chart, the multivariate exponentially weighted moving average (MEWMA) control chart and the multivariate Shewhart control chart. The latter is also known as Hotelling's χ^2 control chart. As the number of process variables grows, traditional multivariate control charts such as the multivariate Shewhart chart lose efficiency with respect to shift detection (Montgomery, 2005, pp. 486). The Shewhart control chart is poor at detecting small and moderate shifts in the mean vector. However, the MEWMA and the MCUSUM control charts are known to respond to small mean shifts very quickly. In this thesis, several aspects of the MEWMA control chart will be in detail studied.

1.2 Average run length (*ARL*)

The average run length (*ARL*) is a good tool to evaluate the performance of a statistical process control chart. The *ARL* is the average number of points that must be plotted before a point indicates an out-of-control condition (Montgomery, 2005, pp. 160). When a process control chart is set up, it is desirable that it produces a large *ARL* when the process is in-control while smaller *ARL* values are preferred when the process is out-of-control (Pham, 2006, pp. 337). A large in-control *ARL* reduces the false alarms while a small out-of-control *ARL* indicates quick detection of a change. Since evaluating *ARL* values is not elementary, let us consider the univariate Shewhart control chart for the purpose of illustrating how the *ARL* is calculated. In this case, it is well-known that the run length follows a geometric distribution. Thus, its expected value is

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

where p is the probability that any point exceeds the control limits.

For instance, when the process is in-control with $p = 0.005$, then the in-control *ARL* (called ARL_0) equals $\frac{1}{0.005} = 200$, which means that the control chart signals a false (out-of-control) alarm on average every 200 plotted points even though the process is in-control.

When the process is out-of-control, it is expected that more chart points will go out of the control limits. Thus, the out-of-control ARL (called ARL_1) will be smaller than ARL_0.

1.3 Semiconductor manufacturing

Materials used in electronics are classified into three types in terms of conductivity: conductors, insulators and semiconductors. Conductors are materials that can carry electrons easily thanks to the availability of free electrons such as copper and aluminum. Most metals are considered to be conductors. Insulators are materials that refuse to carry an electric current due to lack of free electorns such as glass and wood. Semiconductors are substances that are neither conductors nor insulators but they can have electrical properties by applying a certain voltage and doping impurity content (Bakshi and Godse, 2008, pp. 8). The two widely used simiconductor materials are silicon and germanium. Semicondutor devices are manufactured electronic components or integrated circuits by using semiconductor materials. Nowadays, semiconductor devices are considered as the cornerstone of electronics because most of our modern conveniences such as computers, cell phones, digital cameras, medical diagnostic equipment and all kind of domestic electric appliances are made of semicoductors. The reason that semiconductors are important is that we can alter their conductivity.

In semiconductor device fabrication, all the processing steps fall into one of the following categories: Deposition, ion-implantation, diffusion, photolithography and etching. Deposition is used to put down either a metal layer or an oxide (non-metal) layer on a wafer. Ion-implantation and diffusion are the operations that introduce dopants inside the wafer and grow a silicon oxide layer. Photolithography is the process that a light-sensitive material, called *photoresist* is applied in the wafer which is then exposed to ultraviolet light through an optical mask. Then the area of the photoresist exposed to light becomes soluble and is stripped off with solvents. Etching operation is used to create a circuit pattern that has been defined during the photolithography process by removing a thin film.

They form a process cycle and are employed on the wafer numerous times to make a semiconductor device. As a result, multiple layers are created and stacked directly on the wafer. Figures 1.1 and 1.2 are the cross-section of a semiconductor device. Specifically, Figure 1.2 is the cross-section of SRAM (Static Random Access Memory) which is taken by SEM (Scanning Electron Microscope). In this study, our focus is on the layer on which polysilicon gates (transistors) are patterned. In semiconductor manufacturing, the critical dimension (CD) of the gate width is the most critical parameter since the gate CD decides the overall speed of the integrated circuit and it has continued to shrink since the integrated circuit was introduced (Orshansky et al., 1999). The following website shows how quickly design rules for gate patterns have changed. Refer to `http://en.wikipedia.org/wiki/Semiconductor_device_fabrication` at wikipedia.

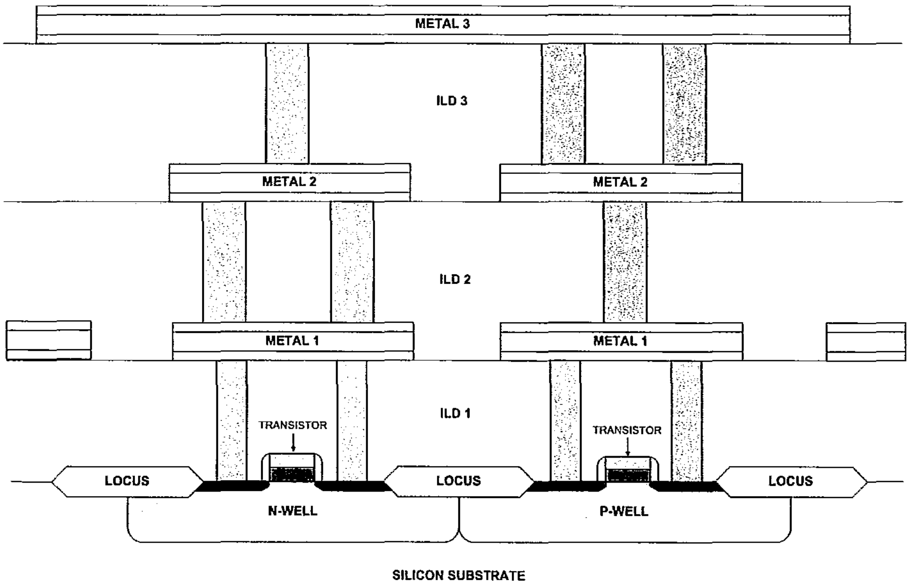

Figure 1.1: Cross-section of a semiconductor device.

The design rule of silicon chips was 10 μm in 1971 but chipmakers are now making 32 nm devices. That is, today's transistors are more than 300 times smaller than the ones made in 1971. Shrinkage in the gate width brings us difficulties to control it. Thus, a tighter control over the gate width has to be made to maximize process yield and throughput. Even a small amount of mean shift in the silicon gate width has to be detected. Therefore, the MEWMA control scheme which is very efficient at issuing a warning signal on a small amount of mean shift is the right choice and suitable for the semiconductor industry facing increasing quality demands. Figure 1.3 shows a finished product of semiconductor device.

As an application of the MEWMA control chart, we construct a simple bivariate normal distribution model, apply it to semiconductor manufacturing operations and provide simulation results. The main reason for the application to a semiconductor manufacturing process is that semiconductor manufacturing operations need a control chart that provides a high sensitivity in detecting a small mean shift.

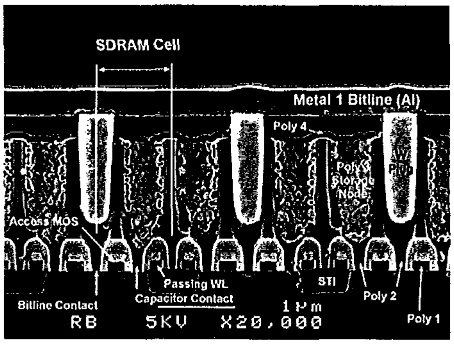

Figure 1.2: Cross-section of SRAM taken by SEM.

1.4 Propagation of variability in a process

One of the main goals of statistical process control is to effectively reduce the variability in a process. In order to do that, it is important to identify at which stages variation is

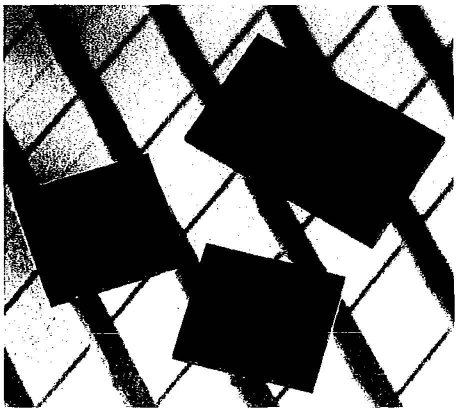

Figure 1.3: 4Gb DDR DRAM chip.

added substantially and how much variation is transmitted from previous stages. Lawless et al. (1999) introduced the variation transmission model and showed how variation in key product characteristics could be built across the production stages. However, the model is constructed based on that only one machine runs at final stage. In this thesis, we will extend the model and apply it to semiconductor manufacturing.

1.5 Thesis objectives and organization

Prabhu and Runger (1996) developed a Markov chain algorithm to evaluate the performance of the MEWMA control charts. The Markov chain algorithm provides an acceptable approximation for the average run length (ARL) of the MEWMA control chart. The on- and off-target ARL values can be computed by using the algorithm. However, when the performance of the off-target case is evaluated, the off-target case has to be extended in terms of the process transition time since the past studies only cover the off-target case where the process mean shifts instantaneously to a new value once manufacturing operations start; that is, the mean shift was assumed to take place at the very beginning of the process (the zero-state case). In real life, the process goes out-of-control after staying in-control for a while from the beginning and the change sustains until human intervention (the steady-state case). The process mean shift could take place at any time in process operation, not just at the beginning. For example, in the Markov chain model, when the process mean changes at transition time τ, we consider a transition matrix $\boldsymbol{P}_0$ until time $\tau - 1$ and a new transition matrix $\boldsymbol{P}_1$ afterwards.

The main objectives of this thesis are as follows:

1. To calculate analytically and numerically the run length distribution and the average run length for the off-target case.
2. To compare the MEWMA control with the multivariate Shewhart control chart for the off-target case.
3. To identify optimal smoothing parameter values for the off-target case.
4. To apply the MEWMA control chart to a bivariate semiconductor manufacturing process.
5. To develop a model for variation propagation with application to the semiconductor manufacturing process.

The study is arranged as follows. In Chapter 2, the general off-target case is discussed and we show derivations of the *ARL* and the variance for the general off-target case. Additionally, the optimal smoothing parameter and the comparison of the MEWMA control chart and the multivariate Shewhart control chart are discussed for different values of transition time. In Chapter 3, a bivariate normal distribution model for the MEWMA control chart is applied to a semiconductor fabrication process and the method is illustrated with simulated data. An extension of the variation transmission model is introduced in Chapter 4. Finally, Chapter 5 presents some conclusions.

Chapter 2

Analysis of MEWMA control chart

2.1 Overview of MEWMA control chart

The conventional Shewhart-type control charts such as the T^2 charts are pretty effective for detecting mean shifts. However, they are slow in reacting to small and moderate shifts in the process mean. In that regard, the MEWMA control chart was developed to provide more sensitivity to small mean shifts (Montgomery, 2005, pp. 504). Suppose that $\boldsymbol{X_t} = (X_1, X_2, \cdots, X_p)'$ is a $p-$dimensional random vector whose components are random variables at time t. Lowry et al. (1992) proposed a multivariate version of the univariate exponentially weighted moving average (EWMA) control chart. As for the MEWMA control chart, it is defined by

$$\boldsymbol{Z_t} = r\boldsymbol{X_t} + (1-r)\boldsymbol{Z_{t-1}} \tag{2.1}$$

where, r is a smoothing parameter $(0 < r \leq 1)$ and it is assumed that $\boldsymbol{Z_0} = \boldsymbol{0_p}$. The MEWMA control chart issues a warning signal when

$$Q_t = \boldsymbol{Z_t'}\Sigma_{\boldsymbol{Z_t}}^{-1}\boldsymbol{Z_t} > H \tag{2.2}$$

where H is a specified control limit and the covariance matrix, $\Sigma_{\boldsymbol{Z}_t}$ is given as $\left\{\frac{r[1-(1-r)^{2t}]}{(2-r)}\right\}\Sigma_{\boldsymbol{X}}$. However, in this thesis we use the following asymptotic covariance matrix:

$$\Sigma_{\boldsymbol{Z}_t} = \Big(\frac{r}{2-r}\Big)\Sigma_{\boldsymbol{X}}.$$

Note from Equation (2.1) that when we expand $\boldsymbol{Z}_t$ recursively, we get

$$\boldsymbol{Z}_t = r\boldsymbol{X}_t + r(1-r)\boldsymbol{X}_{t-1} + r(1-r)^2\boldsymbol{X}_{t-2} + \cdots + r(1-r)^{t-1}\boldsymbol{X}_1 + (1-r)^t\boldsymbol{Z}_0.$$

Thus, $\boldsymbol{Z}_t$ is a weighted average of the t quality measurements available with weights following a geometric form. However, in the literature the chart is known as *exponentially weighted.*

An important special case of the MEWMA control chart is the case that $r = 1$ leads to $\boldsymbol{Z}_t = \boldsymbol{X}_t$ and $Q_t = \boldsymbol{X}_t{}'\Sigma_{\boldsymbol{X}}^{-1}\boldsymbol{X}_t$. This is precisely the multivariate Shewhart control chart also known as the chi-squared control chart. Let us assume that

$$E(\boldsymbol{X}) = \boldsymbol{\mu} = \begin{cases} \boldsymbol{\mu}_0 & \text{when the process is on-target} \\ \boldsymbol{\mu}_1 & \text{when the process is off-target} \end{cases}$$

and $Var(\boldsymbol{X}) = \Sigma_{\boldsymbol{X}}$.

Consider the following transformation and let the transformed variable be $\Sigma_{\boldsymbol{X}}^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}_0)$. By the transformation, we obtain

$$E(\Sigma_{\boldsymbol{X}}^{-1/2}(\boldsymbol{X}-\boldsymbol{\mu}_0)) = \begin{cases} \Sigma_{\boldsymbol{X}}^{-1/2}E(\boldsymbol{X}-\boldsymbol{\mu}_0) = \Sigma_{\boldsymbol{X}}^{-1/2}(\boldsymbol{\mu}_0-\boldsymbol{\mu}_0) = 0 & \text{when the process is on-target} \\ \Sigma_{\boldsymbol{X}}^{-1/2}E(\boldsymbol{X}-\boldsymbol{\mu}_0) = \Sigma_{\boldsymbol{X}}^{-1/2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0) & \text{when the process is off-target} \end{cases}$$

and $Var(\Sigma_{\boldsymbol{X}}^{-1/2}(\boldsymbol{X}-\boldsymbol{\mu}_0)) = \Sigma_{\boldsymbol{X}}^{-1/2}\Sigma_{\boldsymbol{X}}(\Sigma_{\boldsymbol{X}}^{-1/2})' = \Sigma_{\boldsymbol{X}}^{-1/2}\Sigma_{\boldsymbol{X}}^{1/2}\Sigma_{\boldsymbol{X}}^{1/2}(\Sigma_{\boldsymbol{X}}^{-1/2})' = \boldsymbol{I}$. The non-centrality parameter c is defined as follows.

$$c = (\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\Sigma_{\boldsymbol{X}}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0). \tag{2.3}$$

Then, the noncentrality parameter of the transformed variable $\Sigma_{\boldsymbol{X}}^{-1/2}(\boldsymbol{X}-\boldsymbol{\mu}_0)$ is

$$\begin{aligned} c &= (\Sigma_{\boldsymbol{X}}^{-1/2}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)-0)'(\boldsymbol{I})^{-1}(\Sigma_{\boldsymbol{X}}^{-1/2}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)-0) \\ &= (\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\Sigma_{\boldsymbol{X}}^{-1/2}\Sigma_{\boldsymbol{X}}^{-1/2}(\boldsymbol{\mu}-\boldsymbol{\mu}_0) \\ &= (\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\Sigma_{\boldsymbol{X}}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0). \end{aligned}$$

The result is equivalent to the noncentrality parameter of $\boldsymbol{X}$ (Equation (2.3)). By this transformation, we can assume that $\boldsymbol{X}$ has mean zero and an identity covariance matrix since the performance of a MEWMA control chart is a function of $\boldsymbol{\mu}$ only through the noncentrality parameter (Lowry, 1992). Using that, Q_t in Equation (2.2) can be rewritten as $Q_t = (\frac{2-r}{r})||\boldsymbol{Z}_t||^2$. Thus, $Q_t > H$ is equivalent to $||\boldsymbol{Z}_t|| > \sqrt{\frac{r}{2-r}H}$. That is, $||\boldsymbol{Z}_t|| > H'$ where $H' = \sqrt{\frac{r}{2-r}H}$. In this thesis, we will use

$$q_t = ||\boldsymbol{Z}_t||$$

as the control chart statistic.

2.2 The Markov chain approximation algorithm

The main objective of statistical process control charts is to provide a way to detect process shifts as quickly as possible when the process is out-of-control. One way is through the average run length (ARL) of the control chart. Several attempts by using simulation have been made to determine on- and off-target average run length for multivariate control charts, such as MEWMA and multivariate cumulative sum (MCUSUM) control charts (Crosier, 1988; Hawkins, 1992; Pignatiello and Runger, 1990; Woodall and Ncube, 1985). However, the simulation method has the downside that we have to go through a long and tiresome process to obtain an upper control limit and a large number of simulated process runs are required to get an acceptable variance. Brooks and Evans (1972) used a Markov chain approximation for the ARL of a univariate CUSUM control chart and Lucas and Saccucci (1990) applied this method for the EWMA chart. Rigdon (1995a, 1995b) used integral equations to obtain ARL values for a MEWMA. Prabhu and Runger (1996) used the Markov chain model to determine the run length performance of a MEWMA control chart. There is a conceptual difference between the two approaches. To analyze a shift of the observed p dimensional mean vector (when the process becomes out-of-control), the Rigdon's method uses a change in the mean of two dependent random variables while the Markov chain approach uses a one-dimensional random variable and a $p-1$ dimensional random vector. However, the main drawback of the Rigdon's integration equation is that the equation can not be applied for the off-target setting. In this section, we review the Markov chain model that Prabhu and Runger (1996) proposed for the MEWMA control chart, which is the foundation for this study.

The control statistic q_t is non-negative and large values of it are indicative of out-of-control. Thus, an upper control limit (UCL) is used. In the Markov chain approach, the in-control range [0, UCL] is divided into subintervals which form the states of the chain. Let

$$g_1 = \text{width of the on-target states}$$

$$g_2 = \text{width of the off-target states}$$

$$\text{UCL} = \sqrt{H \times r/(2-r)}$$

$$p = \text{number of variables.}$$

2.2.1 On-target performance

Figure 2.1 illustrates the states when the process is in-control. Dividing the range $[0, \text{UCL}]$ into $m_1 + 1$ subintervals, m_1 of them have the same length g_1 and one of them has the length $\frac{g_1}{2}$. Thus,

$$\frac{g_1}{2} + m_1 g_1 = \text{UCL}$$
$$g_1 = \frac{2\text{UCL}}{2m_1 + 1}.$$

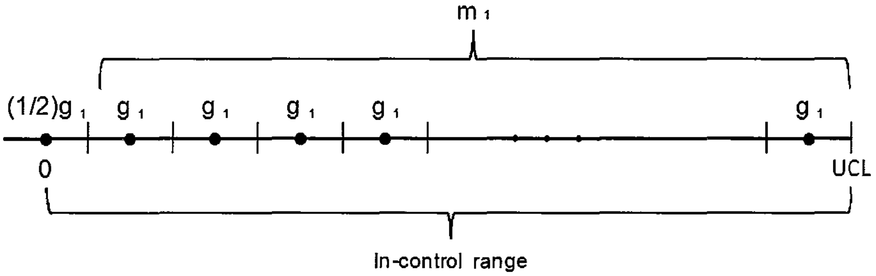

Figure 2.1: A illustration of the Partitioning the Control Region of a MEWMA (On-Target).

When the process is in-control ($\boldsymbol{\mu} = \mathbf{0}$) from the beginning, the on-target distribution of $q_t = ||\boldsymbol{Z_t}||$ can be approximated by using a Markov chain.

$$\begin{aligned} p(i,j) &= P(q_t \text{ in state } j | q_{t-1} \text{ in state } i) \\ &= P\{(j-0.5)g_1 < ||r\boldsymbol{X_t} + (1-r)\boldsymbol{Z_{t-1}}|| < (j+0.5)g_1 | q_{t-1} = g_1 i\} \end{aligned}$$

Let us denote $S(r)$ the p - dimensional sphere of radius $r > 0$. Since $\boldsymbol{Z_t}$ has a spherical distribution, the conditional distribution of $\boldsymbol{Z_t}$ given $||\boldsymbol{Z_t}||$ is the same as $||\boldsymbol{Z_t}||\boldsymbol{U}$, where $\boldsymbol{U}$ is the uniform random variable on the p - dimensional sphere with radius 1 (Eaton, 1983). Thus, given that $q_{t-1} = g_1 i$, the distribution of $\boldsymbol{Z_{t-1}}|(q_{t-1} = g_1 i)$ follows $ig_1\boldsymbol{U}$. We get

$$p(i,j) = P\{(j-0.5)g_1/r < ||\boldsymbol{X_t} + (1-r)ig_1\boldsymbol{U}/r|| < (j+0.5)g_1/r\} \tag{2.4}$$

where $\boldsymbol{X_t} \sim N_p(\mathbf{0}, \boldsymbol{I})$ and $\boldsymbol{U} \sim S(1)$. Assume that $\boldsymbol{X_t}$ is independent of $\boldsymbol{U}$. Then Equation

(2.4) can be rewritten as

$$p(i,j) = \int_{S(1)} \cdots \int f(u) P\{(j-0.5)g_1/r < ||\boldsymbol{X_t} + (1-r)ig_1\boldsymbol{u}/r|| < (j+0.5)g_1/r | \boldsymbol{U} = \boldsymbol{u}\} d\boldsymbol{u}.$$

Conditioning on $\boldsymbol{U} = \boldsymbol{u}$,

$$\boldsymbol{X_t} + (1-r)ig_1\boldsymbol{u}/r \sim N_p\big((1-r)ig_1\boldsymbol{u}/r, \boldsymbol{I}\big).$$

Hence, conditioning on $\boldsymbol{U} = \boldsymbol{u}$,

$$||\boldsymbol{X_t} + (1-r)ig_1\boldsymbol{u}/r||^2 \sim \chi^2(p,c) \quad \text{where c is the noncentrality parameter.}$$

The noncentrality parameter c can be calculated as follows.

$$c = \Big(\frac{1-r}{r} ig_1\boldsymbol{u}'\Big)\boldsymbol{I}\Big(\frac{1-r}{r}ig_1\boldsymbol{u}\Big) = \Big(\frac{1-r}{r}ig_1\Big)^2\boldsymbol{u}'\boldsymbol{u} = \Big(\frac{1-r}{r}ig_1\Big)^2 ||\boldsymbol{u}||^2 = \Big(\frac{1-r}{r}ig_1\Big)^2$$

where $\boldsymbol{u}$ represents any vector in the p - dimensional sphere of radius 1, that is $||\boldsymbol{u}|| = 1$.

Thus,

$$\begin{aligned} p(i,j) &= \int_{S(1)} \cdots \int f(u) P\left\{ \frac{(j-0.5)^2 g_1^2}{r^2} < ||\boldsymbol{X_t} + (1-r)ig_1\boldsymbol{u}/r||^2 < \frac{(j+0.5)^2 g_1^2}{r^2} \middle| \boldsymbol{U} = \boldsymbol{u} \right\} d\boldsymbol{u} \\ &= \int_{S(1)} \cdots \int f(u) P\{(j-0.5)^2 g_1^2/r^2 < \chi^2(p,c) < (j+0.5)^2 g_1^2/r^2\} d\boldsymbol{u} \\ &= P\{(j-0.5)^2 g_1^2/r^2 < \chi^2(p,c) < (j+0.5)^2 g_1^2/r^2\} \int_{S(1)} \cdots \int f(u) d\boldsymbol{u} \\ &= P\{(j-0.5)^2 g_1^2/r^2 < \chi^2(p,c) < (j+0.5)^2 g_1^2/r^2\}. \quad \Big(\text{since } \int_{S(1)} \cdots \int f(u) d\boldsymbol{u} = 1\Big) \end{aligned}$$

Therefore, for $i, j = 0, 1, 2, \ldots, m_1$, the probability of a transition from state i to state

j is denoted $p(i,j)$ and defined as follows.

$$p(i,j) = \begin{cases} P\{(j-0.5)^2 g_1^2/r^2 < \chi^2(p,c) < (j+0.5)^2 g_1^2/r^2\} & \text{if } j \neq 0 \\ P\{\chi^2(p,c) < (0.5)^2 g_1^2/r^2\} & \text{if } j = 0 \end{cases}$$

where $\chi^2(p,c)$ is a noncentral chi-squared random variable with p degrees of freeedom, noncentrality parameter $c = [(1-r)ig_1/r]^2$ and $g_1 = \frac{2\text{UCL}}{2m_1+1}$. Using the transitional probabilities, the $(m_1+1) \times (m_1+1)$ transition matrix $\boldsymbol{P}_0$ of the transient states of the chain can be constructed.

By using the above algorithm, the on-target average run length is given by

$$ARL_0 = \lim_{m_1 \to +\infty} \boldsymbol{s}'(\boldsymbol{I} - \boldsymbol{P}_0)^{-1}\mathbf{1} \quad \text{(Prabhu and Runger, 1996)} \tag{2.5}$$

where $\boldsymbol{s}$ is the starting probability vector and $\mathbf{1}$ denotes a vector of 1s of the dimension m_1+1. Derivation of Equation (2.5) is provided in section 2.3.

2.2.2 Off-target performance (zero-state case)

Figure 2.2 represents the two-dimensional range of $(Z_{t1}, ||\boldsymbol{Z}_{t2}||)$ with the axes Z_1 and $||\boldsymbol{Z}_2||$. For the Markov chain approximation of Z_{t1}, the number of states between -UCL and UCL is $2m_2+1$. States are 1, 2, $\ldots, 2m_2+1$. Thus, the width of each state, g_2 is $\frac{2\text{UCL}}{2m_2+1}$. For the Markov chain approximation of $||\boldsymbol{Z}_{t2}||$, the number of states are m_1+1, labelled as 0, 1, $\ldots m_1$. Thus, $g_1 = \frac{2\text{UCL}}{2m_1+1}$.

Suppose that the process is out-of-control ($\boldsymbol{\mu}_0$ changes to $\boldsymbol{\mu}_1$) from the beginning and let $\delta = ||\boldsymbol{\mu}_1||$. Then $\delta = ||\boldsymbol{\mu}_1|| = \sqrt{(\boldsymbol{\mu}_1 - \mathbf{0})'\boldsymbol{I}(\boldsymbol{\mu}_1 - \mathbf{0})} = \sqrt{\boldsymbol{\mu}_1{}'\boldsymbol{\mu}_1}$ which is the noncentrality parameter. Since the MEWMA is a function of the off-target mean ($=\boldsymbol{\mu}_1$) only through the noncentrality parameter, we can assume that $\boldsymbol{\mu}_1 = \delta\boldsymbol{e}$ where $\boldsymbol{e}$ is the p component unit vector $\boldsymbol{e}' = (1,0,0,\ldots,0)$. Thus, $\boldsymbol{Z}_t$ can be partitioned into a one-dimensional random variable Z_{t1} with non zero mean δ and $p-1$ dimensional random vector $\boldsymbol{Z}_{t2}$ with zero

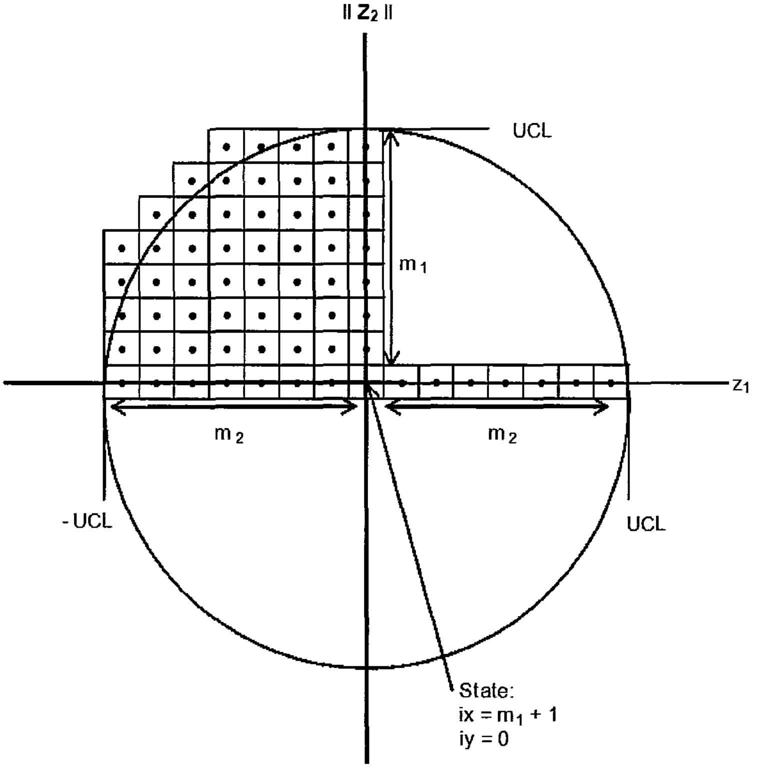

Figure 2.2: States in the Markov Chain Used for the Off-Target Case of a MEWMA (Prabhu and Runger, 1996).

mean, where $\delta = (\boldsymbol{\mu}'{\boldsymbol{\Sigma}_{\boldsymbol{X}}}^{-1}\boldsymbol{\mu})^{1/2}$. That is, $q_t = ||\boldsymbol{Z}_t|| = (Z_{t1}^2 + \boldsymbol{Z}_{t2}'\boldsymbol{Z}_{t2})^{1/2}$. The transitional probability of Z_{t1} from state i_x to state j_x, denoted by $h(i_x, j_x)$ is used to analyze the off-target control component. For $i_x, j_x = 1, 2, \ldots, 2m_2 + 1$,

$$\begin{aligned}
& h(i_x, j_x) = P(Z_{t1} \text{ in state } j_x | Z_{t-1} \text{ in state } i_x) \\
= \; & P[(-\text{UCL} + (j_x - 1)g_2 - (1-r)c_{i_x})/r - \delta) < X_{t1} - \delta < (-\text{UCL} + j_x g_2 - (1-r)c_{i_x})/r - \delta)] \\
= \; & \Phi\Big((-\text{UCL} + j_x g_2 - (1-r)c_{i_x})/r - \delta\Big) - \Phi\Big((-\text{UCL} + (j_x - 1)g_2 - (1-r)c_{i_x})/r - \delta\Big)
\end{aligned}$$

where Φ is the cumulative standard normal distribution function and $c_{i_x} = \text{-UCL} + (i_x - 0.5)g_2$.

Let $\boldsymbol{A}$ denote the $(2m_2+1) \times (2m_2+1)$ transition matrix of Z_{t1}. The transitional probability of $||\boldsymbol{Z}_{t2}||$ from state i_y to state j_y, denoted by $v(i_y, j_y)$ is used to analyze the on-target control components. For $i_y, j_y = 0, 1, 2, \ldots, m_1$,

$$v(i_y, j_y) = \begin{cases} P\{(j_y - 0.5)^2 g_1^2/r^2 < \chi^2(p-1, c) < (j_y + 0.5)^2 g_1^2/r^2\} & \text{if } j_y \neq 0 \\ P\{\chi^2(p-1, c) < (0.5)^2 g_1^2/r^2\} & \text{if } j_y = 0 \end{cases}$$

where $c = [(1-r)i_y g_1/r]^2$. Let $\boldsymbol{B}$ denote the $(m_1+1) \times (m_1+1)$ transition matrix of $||\boldsymbol{Z_{t2}}||$. Since Z_{t1} is independent of $\boldsymbol{Z_{t2}}$, the transitional probability of the bivariate chain $\{Z_{t1}, ||\boldsymbol{Z_{t2}}||\}$ from state $(i_x,\ i_y)$ to state $(j_x,\ j_y)$ is

$$p[(i_x, i_y), (j_x, j_y)] = h(i_x, j_x)v(i_y, j_y).$$

Let $\boldsymbol{P_1}$ be the transition matrix of the transient states of the bivariate chain. Using the condition $(i_x - (m_2+1))^2 g_2^2 + i_y^2 g_1^2 < \mathrm{UCL}^2$ and calculating the Kronecker product of $\boldsymbol{A}$ and $\boldsymbol{B}$ (Lee, 2009), $\boldsymbol{P_1}$ can be calculated. See Appendix A.1 for the definition of Kronecker product and Appendix B.1 for the $\boldsymbol{R}$ code for the Markov chain algorithm.

As a result, the off-target average run length is given by

$$ARL_1 = \lim_{m_1, m_2 \to +\infty} \boldsymbol{s}'(\boldsymbol{I} - \boldsymbol{P}_1)^{-1}\boldsymbol{1}. \tag{2.6}$$

2.3 On-target run length analysis

Assume that the process is operating on-target. Following Prabhu and Runger (1996), the Markov chain methods for the MEWMA control chart leads to

$$P(N > n) = \lim_{m_1 \to +\infty} \boldsymbol{s}'\boldsymbol{P_0}^n\boldsymbol{1},\ n = 0,\ 1,\ 2,\ \ldots \tag{2.7}$$

where N is the run length of the scheme, that is the number of runs until the false signals for the first time. Here $\boldsymbol{s}$ is the starting probability vector. $\boldsymbol{P_0}$ is the $(m_1+1) \times (m_1+1)$ transition matrix for the Markov chain, and $\boldsymbol{1}$ denotes a vector of 1s of the dimension

m_1+1. It turns out that the convergence is quite fast with values of 10 to 15 for m_1, giving satisfactory results. In this thesis, we used $m_1 = 25$. Then, the probability mass function for the run length N is

$$\begin{aligned} f(n) = P(N=n) &= P(N>n-1) - P(N>n) \\ &= \boldsymbol{s}'\boldsymbol{P}_0{}^{n-1}\boldsymbol{1} - \boldsymbol{s}'\boldsymbol{P}_0{}^{n}\boldsymbol{1} \\ &= \boldsymbol{s}'\boldsymbol{P}_0{}^{n-1}(\boldsymbol{I}-\boldsymbol{P}_0)\boldsymbol{1}, \quad n=1,2,\ldots \end{aligned}$$

where $\boldsymbol{P}_0{}^0 = \boldsymbol{I}$ and $\boldsymbol{I}$ is the $(m_1+1)\times(m_1+1)$ identity matrix. The on-target run length distributions are provided in Figure 2.3. It is observed that the on-target run length distribution is skewed to the right (positively skewed).

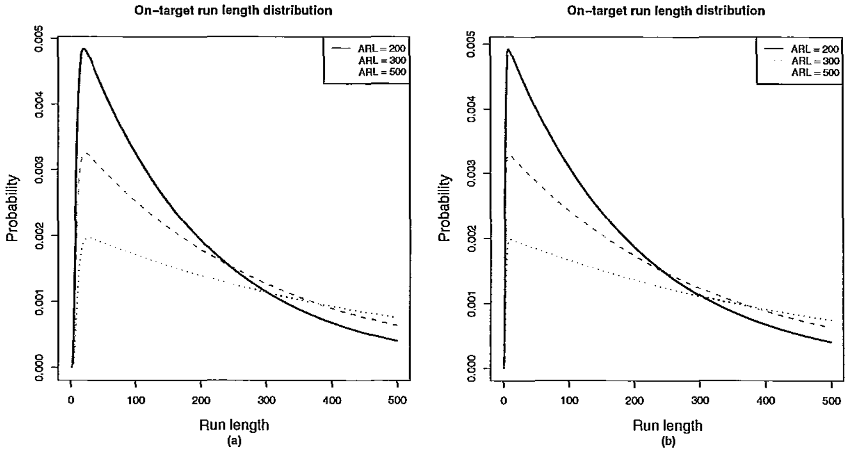

Figure 2.3: On-Target Run Length Distribution for MEWMA with $ARL_0 = 200, 300, 500$ and (a) $r = 0.1$ and (b) $r = 0.3$.

Let us consider the situation that the process is operating on-target and the process mean is $\boldsymbol{\mu} = \boldsymbol{\mu}_0$. Since Equation (2.7) is the survival function of N, we can use it to derive

$E(N)$. Thus, the on-target average run length (ARL_0) is

$$\begin{aligned}
ARL_0 = E(N) &= \sum_{n=1}^{\infty} nf(n) \\
&= 1f(1) + 2f(2) + 3f(3) + 4f(4) + \cdots \\
&= [f(1) + f(2) + f(3) + f(4) + \cdots] + [f(2) + f(3) + f(4) + \cdots] + [f(3) + f(4) + \cdots] + \cdots \\
&= \sum_{n=1}^{\infty} f(n) + \sum_{n=2}^{\infty} f(n) + \sum_{n=3}^{\infty} f(n) + \cdots = P(N \geq 1) + P(N \geq 2) + P(N \geq 3) + \cdots \\
&= \sum_{n=1}^{\infty} P(N \geq n) = \sum_{n=1}^{\infty} \boldsymbol{s}'\boldsymbol{P_0}^{n-1}\mathbf{1} = \boldsymbol{s}'\left(\sum_{n=1}^{\infty} \boldsymbol{P_0}^{n-1}\right)\mathbf{1} = \boldsymbol{s}'(\boldsymbol{I} + \boldsymbol{P_0} + \boldsymbol{P_0}^2 + \cdots)\mathbf{1} \\
&= \boldsymbol{s}'(\boldsymbol{I} - \boldsymbol{P_0})^{-1}\mathbf{1}.
\end{aligned}$$

An alternative way to derive $E(N)$ is presented in Appendix A.2.4.1.

The variance of run length ($=VRL_0$) when the process is operating on-target also can be derived as a closed form (See Appendix A.2.4.2).

$$\begin{aligned}
VRL_0 = Var(N) &= E(N^2) - [E(N)]^2 \\
&= 2\boldsymbol{s}'\boldsymbol{P_0}(\boldsymbol{I} - \boldsymbol{P_0})^{-2}\mathbf{1} + \boldsymbol{s}'(\boldsymbol{I} - \boldsymbol{P_0})^{-1}\mathbf{1} - [\boldsymbol{s}'(\boldsymbol{I} - \boldsymbol{P_0})^{-1}\mathbf{1}]^2 \\
&= 2\boldsymbol{s}'\boldsymbol{P_0}(\boldsymbol{I} - \boldsymbol{P_0})^{-2}\mathbf{1} + \boldsymbol{s}'(\boldsymbol{I} - \boldsymbol{P_0})^{-1}\mathbf{1}[1 - \boldsymbol{s}'(\boldsymbol{I} - \boldsymbol{P_0})^{-1}\mathbf{1}]. \qquad (2.8)
\end{aligned}$$

2.3.1 Moments of on-target run length

In the previous section, we derived the first and second moments of the distribution of run length. The higher moments also can be derived by the same approach. In particular, the third and fourth moments are used to measure skewness and kurtosis of the run length distribution respectively. The third moment of N is

$$E(N^3) = 6\boldsymbol{s}'\boldsymbol{P_0}^2(\boldsymbol{I} - \boldsymbol{P_0})^{-3}\mathbf{1} + 6\boldsymbol{s}'\boldsymbol{P_0}(\boldsymbol{I} - \boldsymbol{P_0})^{-2}\mathbf{1} + \boldsymbol{s}'(\boldsymbol{I} - \boldsymbol{P_0})^{-1}\mathbf{1}$$

(See Appendix A.2.4.3).

Skewness is the degree of asymmetry of a distribution and it is the standardized $3rd$ central moment of run length N. It is defined as

$$\gamma_1 = \frac{E[(N-\mu)^3]}{\sigma^3}, \text{ where } \sigma \text{ is the standard deviation.}$$

We compute

$$\begin{aligned} & E[(N-\mu)^3] = E(N^3) - 3\mu E(N^2) + 2\mu^3 \\ = \;& \mathbf{s}'[6\cdot \mathbf{P_0}^2(\mathbf{I}-\mathbf{P_0})^{-3} + 6\mathbf{P_0}(\mathbf{I}-\mathbf{P_0})^{-2} + (\mathbf{I}-\mathbf{P_0})^{-1}]\mathbf{1} \\ - \;& 3\big(\mathbf{s}'(\mathbf{I}-\mathbf{P_0})^{-1}\mathbf{1}\big)[2\mathbf{s}'\mathbf{P_0}(\mathbf{I}-\mathbf{P_0})^{-2}\mathbf{1} + \mathbf{s}'(\mathbf{I}-\mathbf{P_0})^{-1}\mathbf{1}] \\ + \;& 2\big(\mathbf{s}'(\mathbf{I}-\mathbf{P_0})^{-1}\mathbf{1}\big)^3. \end{aligned}$$

Thus, the skewness of the on-target run length distribution is

$$\begin{aligned} & \gamma_1 = \big\{\mathbf{s}'[6\mathbf{P_0}^2(\mathbf{I}-\mathbf{P_0})^{-3} + 6\mathbf{P_0}(\mathbf{I}-\mathbf{P_0})^{-2} + (\mathbf{I}-\mathbf{P_0})^{-1}]\mathbf{1} - [\mathbf{s}'(\mathbf{I}-\mathbf{P_0})^{-1}\mathbf{1}] \\ & [6\mathbf{s}'\mathbf{P_0}(\mathbf{I}-\mathbf{P_0})^{-2}\mathbf{1} + 3\mathbf{s}'(\mathbf{I}-\mathbf{P_0})^{-1}\mathbf{1} - 2(\mathbf{s}'(\mathbf{I}-\mathbf{P_0})^{-1}\mathbf{1})^2]\big\} \\ & /\big(2\mathbf{s}'\mathbf{P_0}(\mathbf{I}-\mathbf{P_0})^{-2}\mathbf{1} + \mathbf{s}'(\mathbf{I}-\mathbf{P_0})^{-1}\mathbf{1}[1 - \mathbf{s}'(\mathbf{I}-\mathbf{P_0})^{-1}]\big)^{3/2}. \end{aligned}$$

The fourth moment of N is

$$E(N^4) = 24\mathbf{s}'\mathbf{P_0}^3(\mathbf{I}-\mathbf{P_0})^{-4}\mathbf{1} + 36\mathbf{s}'\mathbf{P_0}^2(\mathbf{I}-\mathbf{P_0})^{-3}\mathbf{1} + 14\mathbf{s}'\mathbf{P_0}(\mathbf{I}-\mathbf{P_0})^{-2}\mathbf{1} + \mathbf{s}'(\mathbf{I}-\mathbf{P_0})^{-1}\mathbf{1}$$

(See Appendix A.2.4.4).

Kurtosis is a measure of the flatness of a distribution and it is the standardized $4th$ central moment of run length N. It is defined as

$$\kappa = \frac{E[(N-\mu)^4]}{\sigma^4}.$$

We compute

$$
\begin{aligned}
E[(N-\mu)^4] &= E(N^4) - 4\cdot\mu E(N^3) + 6\cdot\mu^2 E(N^2) - 3\cdot\mu^4 \\
&= \mu_4 - \mu(4\cdot\mu_3 - 6\cdot\mu\mu_2 + 3\cdot\mu^3)
\end{aligned}
$$

where, $\mu_4 = E(N^4), \mu_3 = E(N^3)$, and $\mu_2 = E(N^2)$.

Thus, the kurtosis of the on-target run length distribution is

$$\kappa = \frac{\mu_4 - \mu(4\cdot\mu_3 - 6\cdot\mu_2 + 3\cdot\mu^3)}{(\mu_2-\mu^2)^2}.$$

Additionally, excess kurtosis is $\kappa - 3 = \frac{\mu_4-\mu(4\cdot\mu_3-6\cdot\mu_2+3\cdot\mu^3)}{(\mu_2-\mu^2)^2} - 3$. All the moments derived above can be verified numerically by using the probability mass function. For example, given the condition that $H = 12.7378, ARL_0 = 200, r = 0.1$ and $p = 4$, $E(N^2) = 76,432.59, E(N^3) = 43,757,943$, and $E(N^4) = 33,401,107,743$ are obtained respectively. Table 2.1 shows the approximation of the moments.

m	$\sum_{n=1}^{m} nf(n)$	$\sum_{n=1}^{m} n^2 f(n)$	$\sum_{n=1}^{m} n^3 f(n)$	$\sum_{n=1}^{m} n^4 f(n)$
100	19.43	1,257.93	92,355.41	7,290,654.00
500	147.33	37,242.22	11,769,166.00	4202406879.00
1,000	193.39	68,358.97	33,581,291.00	20075724754.00
5,000	200.00	76,432.59	43,757,942.00	33,401,104,540.00
10,000	200.00	76,432.59	43,757,943.00	33,401,107,743.00
20,000	200.00	76,432.59	43,757,943.00	33,401,107,743.00

Table 2.1: Approximation of moments of on-target run length.

2.4 Off-target run length analysis

In section 2.3, we studied the on-target run length distribution. In this section, we will see the off-target run length distribution. As mentioned earlier, Prabhu and Runger (1997) evaluated the off-target performance by assuming that the process was out-of-control at the beginning of operation (the zero-state case). The analysis in this thesis extends the above method to the steady-state case that is where it is possible for a delayed shift to take place; that is, it is not necessary to happen at the beginning. Thus, we will generalize the notion of the off-target case.

Consider the situation where the process goes off-target from $\mu = \mu_0$ to $\mu = \mu_1$ at the time $t = \tau$ and the change sustains. Thus,

$$\mu = \begin{cases} \mu_0, & t = 1, 2, \dots, \tau - 1, \\ \mu_1, & t = \tau, \tau + 1, \dots . \end{cases}$$

Thus, the transition matrix $\boldsymbol{P}$ changes as well according to

$$\boldsymbol{P} = \begin{cases} \boldsymbol{P}_0, & t = 1, 2, \dots, \tau - 1, \\ \boldsymbol{P}_1, & t = \tau, \tau + 1, \dots . \end{cases}$$

Now, let us consider the situation where the process mean stays in-control until $t = \tau - 1$ and it shifts out-of-control from $t = \tau$ on.

As a result, the survivor function of run length N becomes

$$f_S(n) = P(N > n) = \begin{cases} \boldsymbol{s}'\boldsymbol{P}_0{}^n\boldsymbol{1}, & n = 1, 2, \dots, \tau - 1, \\ \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\boldsymbol{P}_1{}^{n-\tau+1}\boldsymbol{1}, & n = \tau, \tau + 1, \dots . \end{cases}$$

This leads to the following probability mass function for N.

For $n = 1, 2, \dots, \tau - 1$,

$$\begin{aligned} f(n) = P(N = n) = P(N > n - 1) - P(N > n) = \boldsymbol{s}'\boldsymbol{P}_0{}^{n-1}\boldsymbol{1} - \boldsymbol{s}'\boldsymbol{P}_0{}^{n}\boldsymbol{1} \\ = \boldsymbol{s}'\boldsymbol{P}_0{}^{n-1}(\boldsymbol{I} - \boldsymbol{P}_0)\boldsymbol{1}. \end{aligned}$$

For $n = \tau$,

$$\begin{aligned} f(n) = P(N = \tau) = P(N > \tau - 1) - P(N > \tau) = \boldsymbol{s}'\boldsymbol{P_0}^{\tau-1}\mathbf{1} - \boldsymbol{s}'\boldsymbol{P_0}^{\tau-1}\boldsymbol{P_1}\mathbf{1} \\ = \boldsymbol{s}'\boldsymbol{P_0}^{\tau-1}(\boldsymbol{I} - \boldsymbol{P_1})\mathbf{1}. \end{aligned}$$

For $n = \tau + 1, \tau + 2, \tau + 3, \ldots$

$$\begin{aligned} f(n) = P(N = n) = P(N > n - 1) - P(N > n) = \boldsymbol{s}'\boldsymbol{P_0}^{\tau-1}\boldsymbol{P_1}^{(n-1)-\tau+1}\mathbf{1} - \boldsymbol{s}'\boldsymbol{P_0}^{\tau-1}\boldsymbol{P_1}^{n-\tau+1}\mathbf{1} \\ = \boldsymbol{s}'\boldsymbol{P_0}^{\tau-1}\boldsymbol{P_1}^{n-\tau}(\boldsymbol{I} - \boldsymbol{P_1})\mathbf{1}. \end{aligned}$$

Note that the latter formula also applies to $n = \tau$. Thus the probability mass function of run length N is

$$f(n) = P(N = n) = \begin{cases} \boldsymbol{s}'\boldsymbol{P_0}^{n-1}(\boldsymbol{I} - \boldsymbol{P_0})\mathbf{1}, & \text{if } n = 1, 2, \ldots, \tau - 1, \\ \boldsymbol{s}'\boldsymbol{P_0}^{\tau-1}\boldsymbol{P_1}^{n-\tau}(\boldsymbol{I} - \boldsymbol{P_1})\mathbf{1}, & \text{if } n = \tau, \tau + 1, \ldots. \end{cases} \tag{2.9}$$

Note that when $\tau = \infty$ or $\tau = 1$, then the probability mass function reduces to the following forms.

$$f(n) = P(N = n) = \begin{cases} \boldsymbol{s}'\boldsymbol{P_0}^{n-1}(\boldsymbol{I} - \boldsymbol{P_0})\mathbf{1}, & \text{if } \tau = \infty \\ \boldsymbol{s}'\boldsymbol{P_1}^{n-1}(\boldsymbol{I} - \boldsymbol{P_1})\mathbf{1}, & \text{if } \tau = 1. \end{cases}$$

This result is consistent with what Prabhu and Runger (1996) proposed. Thus, Equation (2.9) is the general form of the probability mass function of run length N. Using Equation (2.9), the distribution of run length N can be plotted for different values of τ. Figures 2.4 and 2.5 show how the distribution of the off-target run length moves as τ value changes given the condition that $ARL_0 = 200$ and 500 respectively.

Figures 2.6 and 2.7 show the off-target run length distributions for the MEWMA when the process mean has shifted at $\tau = 50$ and $\tau = 100$ for varying values of smoothing parameter r given that $ARL_0 = 200$ and 500 respectively. Notice that the distribution shows a higher peak as a smaller smoothing parameter r is used.

Figures 2.8 and 2.9 show the off-target run length distributions for the MEWMA when the process mean has shifted at $\tau = 50$ and $\tau = 100$ for different values of mean shift δ given the condition that $ARL_0 = 200$ and 500 respectively. Notice that the distribution shows a higher peak as the amount of shift (δ) increases.

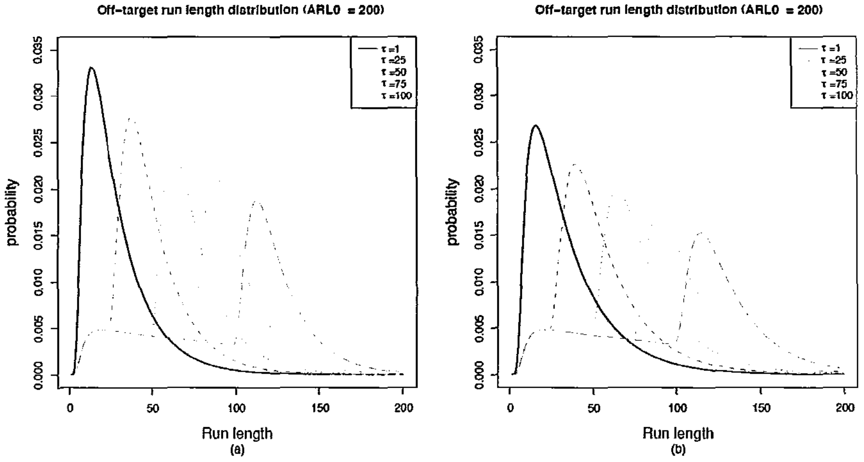

Figure 2.4: Off-Target Run Length Distribution for MEWMA when Process Mean has shifted by $\delta = 0.5$ with $ARL_0 = 200$, $r = 0.1$ and (a) $p = 2$ and (b) $p = 4$.

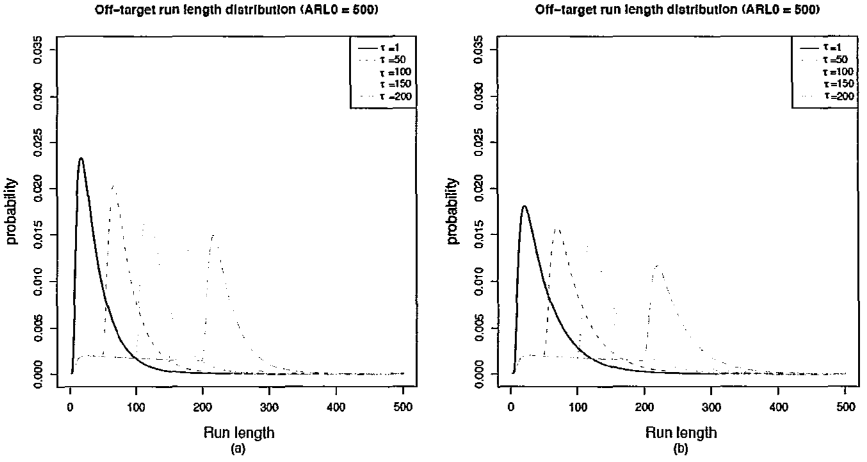

Figure 2.5: Off-Target Run Length Distribution for MEWMA when Process Mean has shifted by $\delta = 0.5$ with $ARL_0 = 500$, $r = 0.1$ and (a) $p = 2$ and (b) $p = 4$.

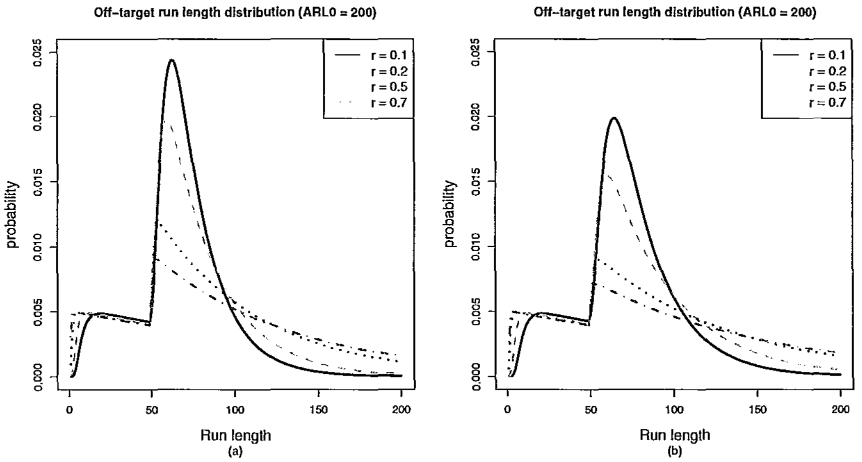

Figure 2.6: Off-Target Run Length Distribution for MEWMA when Process Mean has shifted by $\delta = 0.5$ at $\tau = 50$ with $ARL_0 = 200$ and (a) $p = 2$ and (b) $p = 4$.

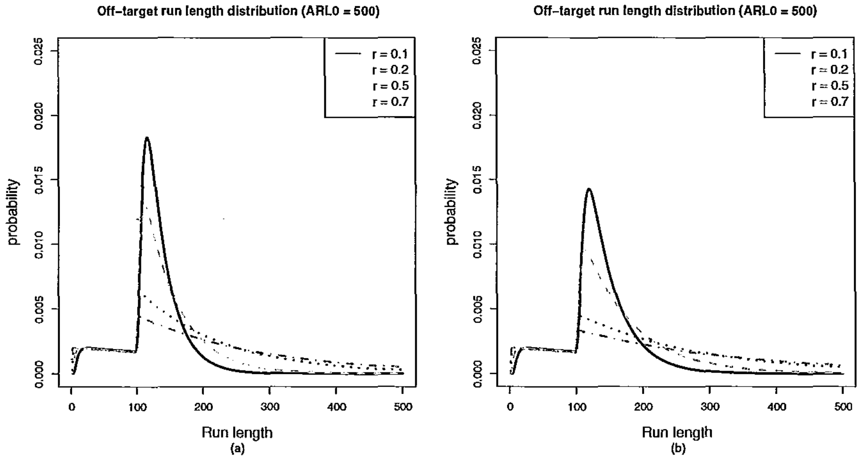

Figure 2.7: Off-Target Run Length Distribution for MEWMA when Process Mean has shifted by $\delta = 0.5$ at $\tau = 100$ with $ARL_0 = 500$ and (a) $p = 2$ and (b) $p = 4$.

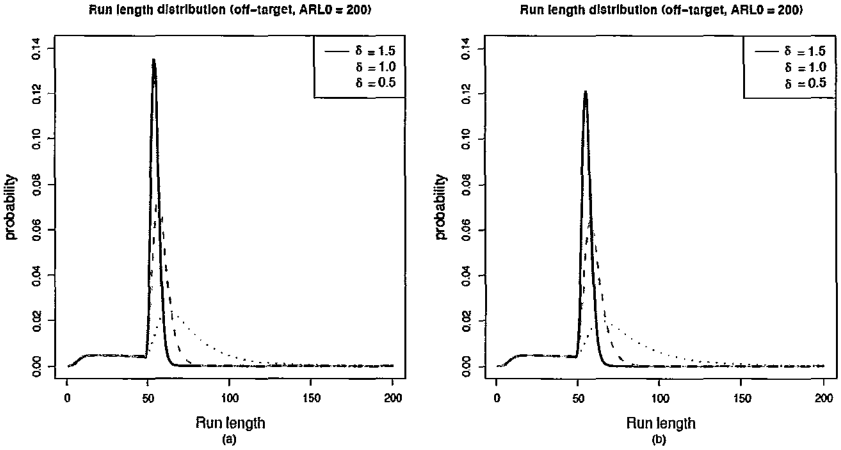

Figure 2.8: Off-Target Run Length Distribution for MEWMA when Process Mean has shifted at $\tau = 50$ with $ARL_0 = 200$, $r = 0.1$ and (a) $p = 2$ and (b) $p = 4$.

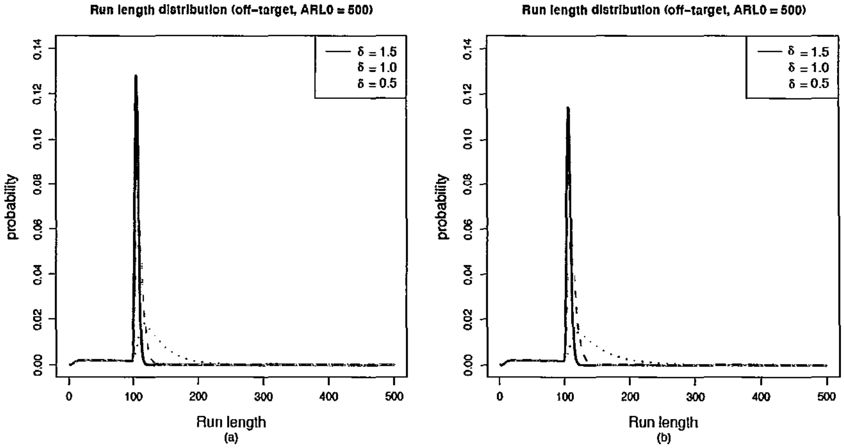

Figure 2.9: Off-Target Run Length Distribution for MEWMA when Process Mean has shifted at $\tau = 100$ with $ARL_0 = 500$, $r = 0.1$ and (a) $p = 2$ and (b) $p = 4$.

Table 2.2 shows the computed values of the off-target average run length (ARL_1) and the probability of false alarm for various τ values, where the probability of false alarm is defined as

$$\text{Probability of false alarm} = \sum_{n=1}^{\tau-1} f(n).$$

It also contains the effective average run length, $ARL_1 - \tau$ which is the average number of runs needed to detect a change in the mean vector after it has occurred. Table 2.2 is constructed given the condition that $ARL_0 = 500, p = 4, r = 0.1$ and $\delta = 0.5$.

Table 2.2: Probability of false alarm, ARL_1 and effective ARL_1 for each transition point (τ).

τ	Probability of false alarm	ARL_1	Effective ARL_1
1	0	51.7425	50.7425
10	0.0024533	59.1293	49.1293
15	0.0094964	63.5308	48.5308
20	0.0185961	67.9431	47.9431
25	0.0282821	72.3270	47.3270
30	0.0380723	76.6711	46.6711
35	0.0478228	80.9726	45.9726
40	0.0574919	85.2307	45.2307
45	0.0670679	89.4456	44.4456
50	0.0765480	93.6180	43.6180
55	0.0859324	97.7478	42.7478
60	0.0952215	101.8356	41.8356
65	0.1044162	105.8819	40.8819
70	0.1135175	109.8871	39.8871
75	0.1225263	113.8516	38.8516
80	0.1314436	117.7758	37.7758
85	0.1402702	121.6602	36.6602
90	0.1490071	125.5050	35.5050
95	0.1576553	129.3108	34.3108
100	0.1662156	133.0779	33.0779

Furthermore, the off-target average run length ARL_1 can be expressed as a closed form.

$$
\begin{aligned}
ARL_1 = E(N) &= \sum_{n=1}^{\infty} nf(n) = \sum_{n=1}^{\tau-1} nf(n) + \sum_{n=\tau}^{\infty} nf(n) \\
&= \sum_{n=1}^{\tau-1} n\boldsymbol{s}'\boldsymbol{P}_0{}^{n-1}(\boldsymbol{I}-\boldsymbol{P}_0)\mathbf{1} + \sum_{n=\tau}^{\infty} n\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\boldsymbol{P}_1{}^{n-\tau}(\boldsymbol{I}-\boldsymbol{P}_1)\mathbf{1} \\
&= \boldsymbol{s}'\left\{\left(\sum_{n=1}^{\tau-1} n\boldsymbol{P}_0{}^{n-1}\right)(\boldsymbol{I}-\boldsymbol{P}_0) + \boldsymbol{P}_0{}^{\tau-1}\left(\sum_{n=\tau}^{\infty} n\boldsymbol{P}_1{}^{n-\tau}\right)(\boldsymbol{I}-\boldsymbol{P}_1)\right\}\mathbf{1} \\
&= \left\{\boldsymbol{s}'[\boldsymbol{I}-\boldsymbol{P}_0{}^{\tau-1}](\boldsymbol{I}-\boldsymbol{P}_0)^{-1}\mathbf{1} + \boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P}_1)^{-1}\right\}\mathbf{1} \text{ (Appendix A.2.5.1).} \qquad (2.10)
\end{aligned}
$$

Alternatively, we can derive the off-target average run length ARL_1 by using the *law of total probability* . That is, $E(N)$ can be written as follows.

$$E(N) = E(N|N < \tau)P(N < \tau) + E(N|N \geq \tau)P(N \geq \tau).$$

We know that $P(N \geq \tau) = P(N > \tau - 1) = \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\mathbf{1}$ and $P(N < \tau) = 1 - \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\mathbf{1}$. The conditional distribution of N, given that $N < \tau$ is defined as

$$\begin{aligned} f(n|N < \tau) &= \frac{f(n)}{P(N < \tau)} \quad n = 1, 2, \cdots, \tau - 1 \quad \text{provided that } P(N < \tau) > 0 \\ &= \frac{1}{1 - \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\mathbf{1}}\boldsymbol{s}'[\boldsymbol{I} + (\tau - 1)\boldsymbol{P}_0{}^{\tau} - \tau\boldsymbol{P}_0{}^{\tau-1}](\boldsymbol{I} - \boldsymbol{P}_0)^{-1}\mathbf{1} \quad \text{(See Appendix A.2.6)}. \end{aligned}$$

Thus,

$$\begin{aligned} E(N|N < \tau)P(N < \tau) &= \frac{1}{1 - \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\mathbf{1}}\boldsymbol{s}'[\boldsymbol{I} + (\tau - 1)\boldsymbol{P}_0{}^{\tau} - \tau\boldsymbol{P}_0{}^{\tau-1}](\boldsymbol{I} - \boldsymbol{P}_0)^{-1}\mathbf{1}(1 - \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\mathbf{1}) \\ &= \boldsymbol{s}'[\boldsymbol{I} + (\tau - 1)\boldsymbol{P}_0{}^{\tau} - \tau\boldsymbol{P}_0{}^{\tau-1}](\boldsymbol{I} - \boldsymbol{P}_0)^{-1}\mathbf{1}. \end{aligned}$$

Now, the conditional distribution of N, given that $N \geq \tau$ is defined as

$$\begin{aligned} f(n|N \geq \tau) &= \frac{f(n)}{P(N \geq \tau)} \quad n = \tau, \tau + 1, \cdots \quad \text{provided that } P(N \geq \tau) > 0 \\ &= \frac{1}{\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\mathbf{1}}\boldsymbol{s}'[\boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I} - \boldsymbol{P}_1)^{-1} + (\tau - 1)\boldsymbol{P}_0{}^{\tau-1}] \quad \text{(See Appendix A.2.6)}. \end{aligned}$$

Thus,

$$\begin{aligned} E(N|N \geq \tau)P(N \geq \tau) &= \frac{1}{\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\mathbf{1}}\boldsymbol{s}'[\boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I} - \boldsymbol{P}_1)^{-1} + (\tau - 1)\boldsymbol{P}_0{}^{\tau-1}]\mathbf{1}(\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\mathbf{1}) \\ &= \boldsymbol{s}'[\boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I} - \boldsymbol{P}_1)^{-1} + (\tau - 1)\boldsymbol{P}_0{}^{\tau-1}]\mathbf{1}. \end{aligned}$$

Therefore,

$$\begin{aligned}
& ARL_1 = E(N) = E(N|N<\tau)P(N<\tau) + E(N|N \geq \tau)P(N \geq \tau) \\
= \; & \boldsymbol{s}'[\boldsymbol{I} + (\tau-1)\boldsymbol{P_0}^{\tau} - \tau\boldsymbol{P_0}^{\tau-1}](\boldsymbol{I}-\boldsymbol{P_0})^{-1}\mathbf{1} + \boldsymbol{s}'[\boldsymbol{P_0}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P_1})^{-1} + (\tau-1)\boldsymbol{P_0}^{\tau-1}]\mathbf{1} \\
= \; & \boldsymbol{s}'[\boldsymbol{I} + (\tau-1)\boldsymbol{P_0}^{\tau} - \tau\boldsymbol{P_0}^{\tau-1} + (\tau-1)\boldsymbol{P_0}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P_0})](\boldsymbol{I}-\boldsymbol{P_0})^{-1}\mathbf{1} + \boldsymbol{s}'\boldsymbol{P_0}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P_1})^{-1}\mathbf{1} \\
= \; & \boldsymbol{s}'[\boldsymbol{I} - \boldsymbol{P_0}^{\tau-1}](\boldsymbol{I}-\boldsymbol{P_0})^{-1}\mathbf{1} + \boldsymbol{s}'\boldsymbol{P_0}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P_1})^{-1}\mathbf{1}.
\end{aligned}$$

We obtain the same result as Equation (2.10).

As a special case, Equation (2.10) reduces to the following forms.

$$ARL_1 = \begin{cases} \boldsymbol{s}'(\boldsymbol{I}-\boldsymbol{P_1})^{-1}\mathbf{1}, & \text{if } \tau = 1 \\ \boldsymbol{s}'(\boldsymbol{I}-\boldsymbol{P_0})^{-1}\mathbf{1} = ARL_0, & \text{if } \tau = \infty\text{, since } \boldsymbol{P}^{\infty} = \mathbf{0}. \end{cases}$$

Note that above results agree with past studies by Runger and Prabhu (1996). Hence, Equation (2.10) is a generalization of of the on- and off-target ARL.

The off-target variance (VRL_1) of N is

$$\begin{aligned}
& VRL_1 = Var(N) = E(N^2) - [E(N)]^2 \\
= \; & \boldsymbol{s}'\{\boldsymbol{P_0}[2\boldsymbol{I} - (\tau-1)\tau\boldsymbol{P_0}^{\tau-2} + 2(\tau-2)\tau\boldsymbol{P_0}^{\tau-1} - (\tau-1)(\tau-2)\boldsymbol{P_0}^{\tau}](\boldsymbol{I}-\boldsymbol{P_0})^{-2} \\
+ \; & \boldsymbol{P_0}^{\tau-1}[2\boldsymbol{P_1}(\boldsymbol{I}-\boldsymbol{P_1})^{-2} + 2(\tau-1)\boldsymbol{P_1}(\boldsymbol{I}-\boldsymbol{P_1})^{-1} + \tau(\tau-1)\boldsymbol{I}] + [\boldsymbol{I}-\boldsymbol{P_0}^{\tau-1}](\boldsymbol{I}-\boldsymbol{P_0})^{-1} \\
+ \; & \boldsymbol{P_0}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P_1})^{-1}\}\mathbf{1} - \{\boldsymbol{s}'[\boldsymbol{I}-\boldsymbol{P_0}^{\tau-1}](\boldsymbol{I}-\boldsymbol{P_0})^{-1}\mathbf{1} + \boldsymbol{s}'\boldsymbol{P_0}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P_1})^{-1}\mathbf{1}\}^2
\end{aligned}$$

(See Appendix A.2.5.2).

As a special case, when $\tau = \infty$,

$$\begin{aligned} & VRL_1 = Var(N) \\ = \;& 2s'P_0(I - P_0)^{-2}1 + s'(I - P_0)^{-1}1[I - s'(I - P_0)^{-1}1] \\ = \;& 2s'P_0(I - P_0)^{-2}1 + s'(I - P_0)^{-1}1 - [s'(I - P_0)^{-1}1]^2 \\ = \;& VRL_0. \end{aligned}$$

Note that the result is equivalent to the on-target variance (See Equation (2.8)).

when $\tau = 1$,

$$\begin{aligned} & VRL_1 = Var(N) \\ = \;& s'\{P_0(2I - 2I)(I - P_0)^{-2} + 2P_1(I - P_1)^{-2} + I + (I - P_1)^{-1} - I\}1 \\ - \;& [s'\{(I - P_0)(I - P_0)^{-1} + (I - P_1)^{-1} - I\}1]^2 \\ = \;& 2s'P_1(I - P_1)^{-2}1 + s'(I - P_1)^{-1}1 - [s'(I - P_1)^{-1}1]^2. \end{aligned}$$

2.5 Comparison of MEWMA and Shewhart control chart

Suppose that X is a random variable and is the number of *Bernoulli* trials until the first success is observed, supported on the set $\{1, 2, 3, \dots\}$. Then the probability mass function of a *geometric random variable* X with success probability α is defined as

$P(X = x) = (1 - \alpha)^{x-1}\alpha,\ x = 1, 2, 3, \dots$. As we discussed earilier, the probability mass function of the on-target run length and the average run length for the MEWMA control chart are defined as

$$\begin{aligned} & f(n) = P(N = n) = s'(I - P_0)P_0^{\,n-1}1,\ \ n = 1, 2, \dots \\ & ARL_0 = E(N) = s'(I - P_0)^{-1}1 \quad \text{respectively.} \end{aligned}$$

The mean, variance, skewness and excess kurtosis of a geometric distribution with probability α are given in Table 2.3.

Table 2.3: Mean, variance, skewness and excess kurtosis of geometric distribution.

Geometric distribution with probability α			
Mean	Variance	Skewness	Excess kurtosis
$\frac{1}{\alpha}$	$\frac{1-\alpha}{\alpha^2}$	$\frac{2-\alpha}{\sqrt{1-\alpha}}$	$6+\frac{\alpha^2}{1-\alpha}$

Comparing the two distributions (i.e., the geometric distribution and the distribution of run length N) by matching the means leads to interesting results. For a given starting vector $\boldsymbol{s}$, a transition matrix $\boldsymbol{P}_0$ and a control limit H, then determine the α such that $E(X) = ARL_0$. As a result, we have $\alpha = \frac{1}{ARL_0} = \frac{1}{\boldsymbol{s}'(\boldsymbol{I}-\boldsymbol{P}_0)^{-1}\boldsymbol{1}}$.

Table 2.4: Comparison of MEWMA run length distribution and geometric distribution with $ARL_0 = 200$.

$p = 4, ARL_0 = 200, prob = 0.005$						
r	0.1	0.5	0.7	0.99	1	geometric dist
Variance	36432.61	39284.82	39593.28	39799.74	39800	39800
Skewness	1.998662	1.999953	1.999969	1.999979	2.000006	2.000006
Excess Kurtosis	5.994399	5.999810	5.999917	5.999963	6.000025	6.000025

Table 2.5: Comparison of MEWMA run length distribution and geometric distribution with $ARL_0 = 500$.

$p = 4, ARL_0 = 500, prob = 0.002$						
r	0.1	0.5	0.7	0.99	1	geometric dist
Variance	239636.3	248085.6	248916.5	249499.2	249500	249500
Skewness	1.99973	1.999992	1.999996	2.000001	2.000001	2.000001
Excess Kurtosis	5.998916	5.999968	5.999986	6.000004	6.000004	6.000004

Figures 2.10 and 2.11 illustrate that there is not much difference between the on-target run length distribution of a MEWMA and a geometric distribution by matching the mean

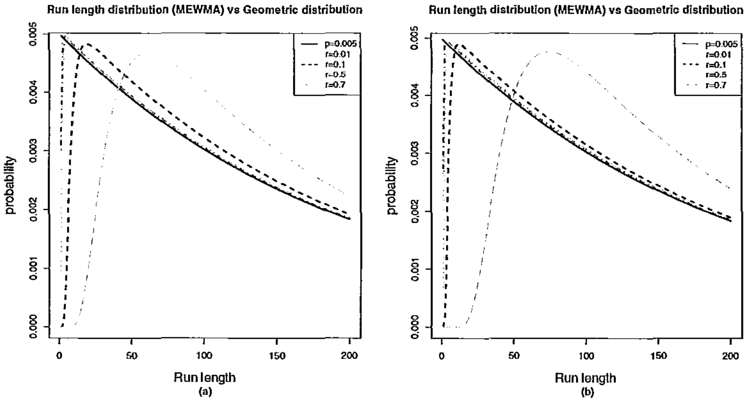

Figure 2.10: Comparison of on-target run length distribution with $ARL_0 = 200$ and geometric distribution with $prob = 0.005$ and (a) $p = 2$ and (b) $p = 4$.

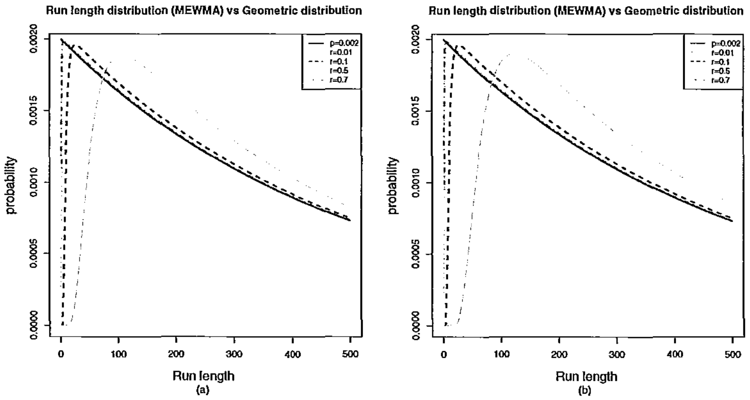

Figure 2.11: Comparison of on-target run length distribution with $ARL_0 = 500$ and geometric distribution with $prob = 0.002$ and (a) $p = 2$ and (b) $p = 4$.

unless the smoothing parameter r is very small. It is also observed that as the smoothing parameter r gets closer to 1, the on-target run length distribution for the MEWMA is becoming the geometric distribution. That is, as $r \rightarrow 1$, then $Z_t = r(X_t - \mu_0) + (1 - r)Z_{t-1} \rightarrow X_t - \mu_0$, which is the multivariate Shewhart control chart. On the other hand, as $r \rightarrow 0$, then $Z_t \rightarrow Z_{t-1}$, which means that all information used is past information. Thus, when r is 1, the on-target run length distribution for the MEWMA is equivalent to the geometric distribution. The variance, skewness and kurtosis of the two distributions are also computed in the Tables 2.4 and 2.5.

Now, let us consider the off-target distribution of the two control charts (i.e., the MEWMA control chart and the Shewhart control chart) for a given smoothing parameter r and ARL_0. The general form of the multivariate Shewhart statistic is defined as

$$T_t^2 = (\boldsymbol{X_t} - \boldsymbol{\mu_0})'\Sigma^{-1}(\boldsymbol{X_t} - \boldsymbol{\mu_0}), \text{ where } t = 1, 2, \ldots$$

and follows a chi-square distribution with p degrees of freedom when the process is on-target and $\boldsymbol{X}$ follows a multivariate normal distribution (Aparisi, 2004).

Since the on-target run length distribution for the Shewhart control chart follows a geometric distribution, the off-target run length distribution for the Shewhart chart can be constructed as follows.

$$g(n) = \begin{cases} p_0(1-p_0)^{n-1}, & n = 1, 2, \cdots, \tau - 1, \\ (1-p_0)^{\tau-1}p_1(1-p_1)^{n-\tau}, & n = \tau, \tau + 1, \ldots \end{cases}$$

where p_0 is the probability that any point exceeds the control limits when the process is in-control while p_1 is the the same probability when the process is out-of-control. Using the probability mass function, the out-of-control ARL_1 when the mean shift takes place at τ is

$$\begin{aligned}
ARL_1 &= \sum_{n=1}^{\infty} ng(n) = \sum_{n=1}^{\tau-1} n\big(p_0(1-p_0)^{n-1}\big) + \sum_{n=\tau}^{\infty} n(1-p_0)^{\tau-1}p_1(1-p_1)^{n-\tau} \\
&= \frac{\big(1 + (\tau-1)(1-p_0)^{\tau} - \tau(1-p_0)^{\tau-1}\big)}{p_0} + (1-p_0)^{\tau-1}\left(\tau + \frac{1-p_1}{p_1}\right) \\
&= \frac{\big(1 + (\tau-1)(1-p_0)^{\tau} - \tau(1-p_0)^{\tau-1} - (\tau-1)(1-p_0-1)(1-p_0)^{\tau-1}\big)}{p_0} + \frac{(1-p_0)^{\tau-1}}{p_1} \\
&= \frac{\big(1 - \tau(1-p_0)^{\tau-1} + (\tau-1)(1-p_0)^{\tau-1}\big)}{p_0} + \frac{(1-p_0)^{\tau-1}}{p_1} \\
&= \frac{\big(1 - (1-p_0)^{\tau-1}\big)}{p_0} + \frac{(1-p_0)^{\tau-1}}{p_1}, \quad \tau = 1, 2, \ldots.
\end{aligned}$$

As a special case, ARL_1 reduces to the following forms.

$$ARL_1 = \begin{cases} 1/p_0, & \text{if } \tau = \infty \\ 1/p_1, & \text{if } \tau = 1. \end{cases}$$

which is equivalent to the mean of the geometric distribution with probability p_0 and p_1 respectively. Additionally, the off-target VRL_1 when the mean shift occurs at time τ is

$$\begin{aligned} & VRL_1 = Var(N) = E[N(N-1)] + E(N) - E(N)^2 \\ = \;& p_0^{-2}(1-p_0)[2-(\tau-1)\tau(1-p_0)^{\tau-2} + 2(\tau-2)\tau(1-p_0)^{\tau-1} - (\tau-1)(\tau-2)(1-p_0)^{\tau}] \\ & +(1-p_0)^{\tau-1}[2(1-p_1)p_1^{-2} + 2(\tau-1)(1-p_1)p_1^{-1} + \tau(\tau-1)] \\ & +\left[(1-(1-p_0)^{\tau-1})p_0^{-1} + (1-p_0)^{\tau-1}p_1^{-1}\right]\left[1-\left((1-(1-p_0)^{\tau-1})p_0^{-1} + (1-p_0)^{\tau-1}p_1^{-1}\right)\right]. \end{aligned}$$

As a special case,

$$VRL_1 = \begin{cases} (1-p_0)/p_0^2, & \text{if } \tau = \infty \\ (1-p_1)/p_1^2, & \text{if } \tau = 1 \end{cases}$$

which is equivalent to the variance of the geometric distribution with probability p_0 and p_1 respectively.

Now, let us compare the performance of the MEWMA control chart with that of the multivariate Shewhart control chart when the mean shift happens. For example, we pick $r = 0.1$ (since 0.1 is the value most often used) and determine the MEWMA control limit that satisfies $ARL_0 = 200$. Then consider the geometric distribution matching with the same mean $ARL_0 = 200$. This is the geometric distribution with parameter $p_0 = \frac{1}{ARL_0} = 0.005$. Determine the control limit for the Shewhart control chart with $ARL_0 = 200$. Notice that the noncentrality parameter of the multivariate Shewhart control chart is defined as $c = (\boldsymbol{\mu} - \boldsymbol{\mu}_0)'\Sigma_{\boldsymbol{X}}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)$ while the noncentrality of the MEWMA is defined as $\delta = ((\boldsymbol{\mu} - \boldsymbol{\mu}_0)'\Sigma_{\boldsymbol{X}}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0))^{1/2}$. That is $c = \delta^2$.

Figures 2.12 – 2.14 show the comparison of the out-of-control ARL_1 values for the two distributions for different values for δ and different values of τ. It is well known that the MEWMA control chart is effective in detecting small shifts when the shift happens at τ =1 (Lowry et al., 1992). Just as in the case $\tau = 1$, the MEWMA control chart outperforms the Shewhart control chart as τ increases until the mean shift δ is 1.5. However, when the large mean shift takes place (i.e., δ is greater than 1.5), the Shewhart control chart is as good as the MEWMA at detecting large shifts in the mean or performs slightly better. Additionally, as τ increases (i.e., the mean shift is delayed more steps), the MEWMA control chart loses its sensitivity to a small mean shift.

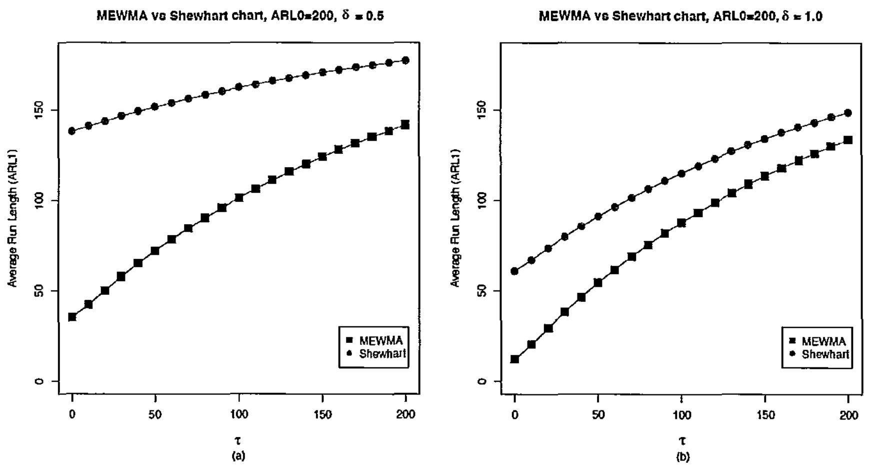

Figure 2.12: Off-Target ARL Comparision of MEWMA and Shewhart with $ARL_0 = 200$ (a) $\delta = 0.5$ and (b) $\delta = 1.0$.

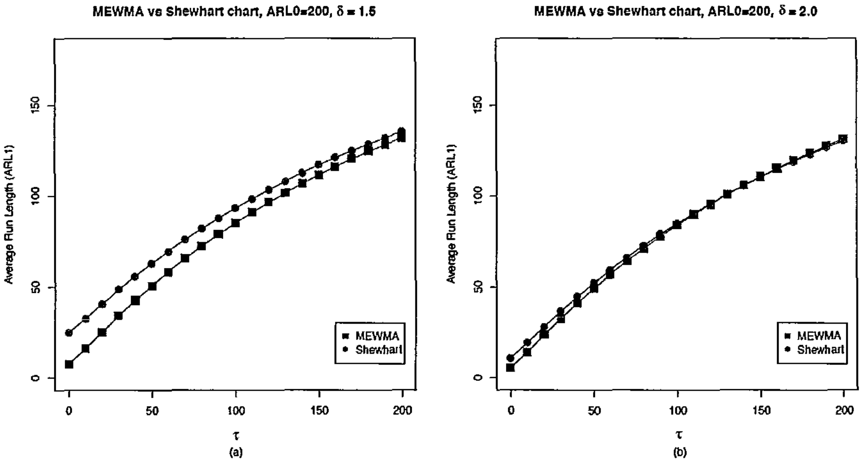

Figure 2.13: Off-Target *ARL* Comparision of MEWMA and Shewhart with $ARL_0 = 200$ (a) $\delta = 1.5$ and (b) $\delta = 2.0$.

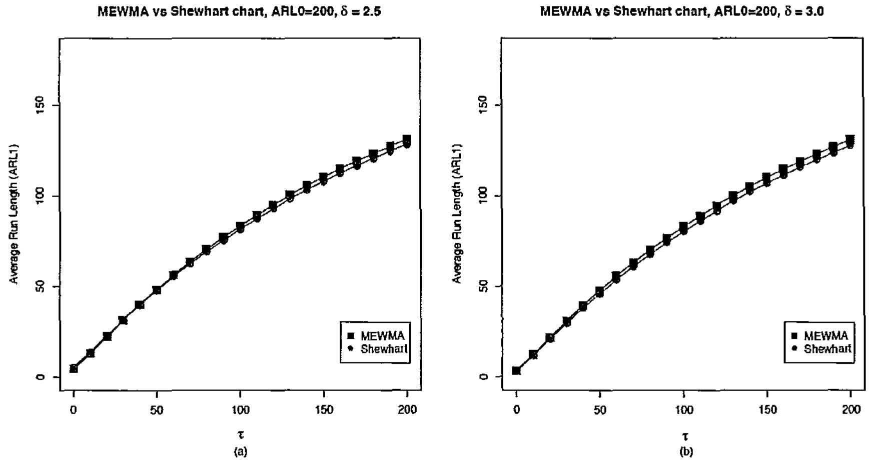

Figure 2.14: Off-Target *ARL* Comparision of MEWMA and Shewhart with $ARL_0 = 200$ (a) $\delta = 2.5$ and (b) $\delta = 3.0$.

2.6 Analysis of optimal smoothing parameter r

When a shift has taken place in the process mean, it is very important to detect the occurrence of the change as early as possible. In the MEWMA control charts, smaller values of r are more effective in detecting small shifts in the mean (Lowry et al., 1992). Thus, for a given ARL_0, we need to find the smoothing parameter that is associated with the smallest ARL_1. First, let us consider the case that the process goes off-target at the beginning of operation ($\tau = 1$). Tables C.5 - C.7 in Appendix C present optimum MEWMA control charts for various shifts (δ) and in-control values of ARL_0 (from 200 to 1,000). The smoothing parameter corresponding to a minimum ARL_1 for a given ARL_0 can be obtained by using the Markov chain algorithm and the partition method (The ***R*** code is provided in Appendix B.4).

The partition method generates a combination of a smoothing parameter r and a control limit H satisfying a given ARL_0 and find the optimal smoothing parameter. The basic idea of the method is as follows. For a fixed smoothing parameter r, the method inspects the middle point of a lower control limit H_{low} and a upper control limit H_{up} such that $ARL_{H_{low}} \leq ARL_0$ and $ARL_{H_{up}} \geq ARL_0$. Once H_{mid}, the middle point of two control limits is obtained, ARL can be calculated by using the Markov chain algorithm. If the difference of ARL_0 and the newly computed ARL is less than a very small number (i.e., $\epsilon < 10^{-3}$), the smoothing parameter r and the control limit H_{mid} is a pair that can satisfy the given ARL_0. Otherwise, keep doing the previous procedures until a sought pair is found. If this task is carried out until the method covers a whole range of smoothing parameter r ($0 < r \leq 1$), a number of combinations of r and H can be obtained. With the combinations obtained, ARL_1 values can be calculated for a given shift δ. Then, the smoothing parameter r for which ARL_1 is the smallest can be identified.

Now we are interested in how the optimal value of r behaves as transition point τ changes. Figures 2.15 and 2.16 show that the optimal values of r increases as τ increases in each case. It is observed that the optimal parameter r changes more slowly when the ARL_0 value gets bigger or p increases. That is, when we have smaller ARL_0 and p, the change of transition time gives a huge impact on deciding the optimal smoothing parameter r. For instance, when we have $ARL_0 = 500$ and $p = 10$, the smoothing parameter changes very little (from 0.10 to 0.13) as τ changes from 1 to 200 while it changes from 0.14 to 0.26 in case we have $ARL_0 = 200$ and $p = 2$. Notice that the usual practice is to pick the optimal value for $\tau = 1$. However, this is unrealistic since most process starts operating well for a while and later on out-of-control slippage occurs.

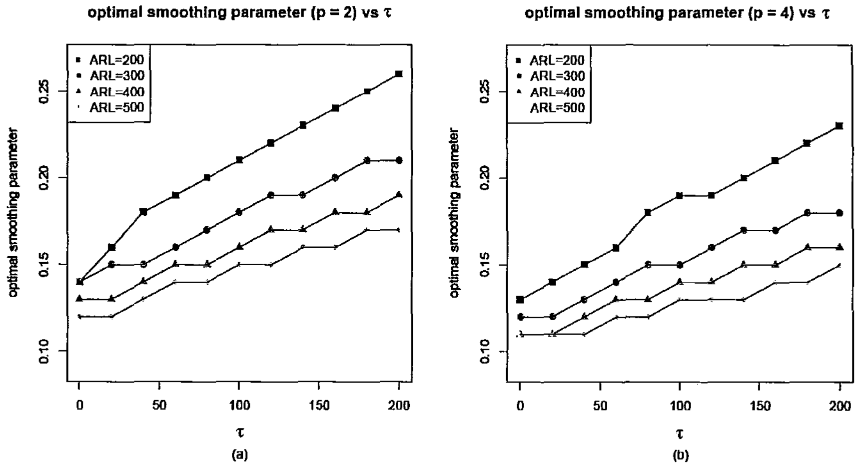

Figure 2.15: Comparison of optimal smoothing parameter as τ increases with (a) $p = 2$ and (b) $p = 4$.

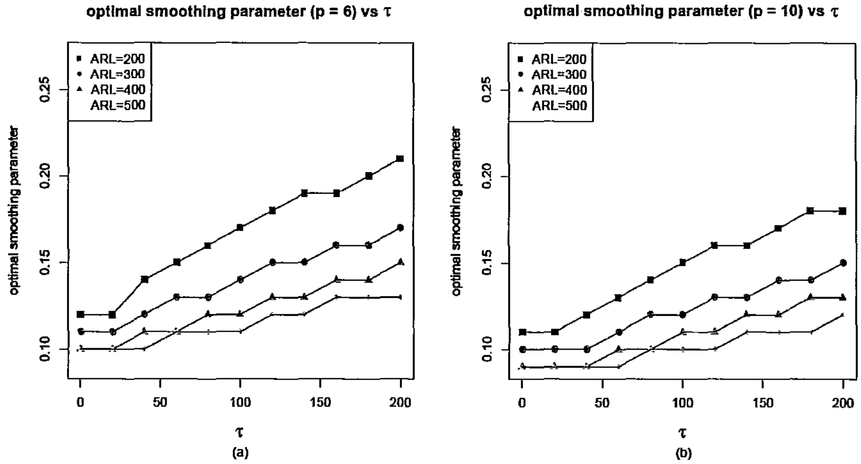

Figure 2.16: Comparison of optimal smoothing parameter as τ increases with (a) $p = 6$ and (b) $p = 10$.

Chapter 3

Application to semiconductor manufacturing

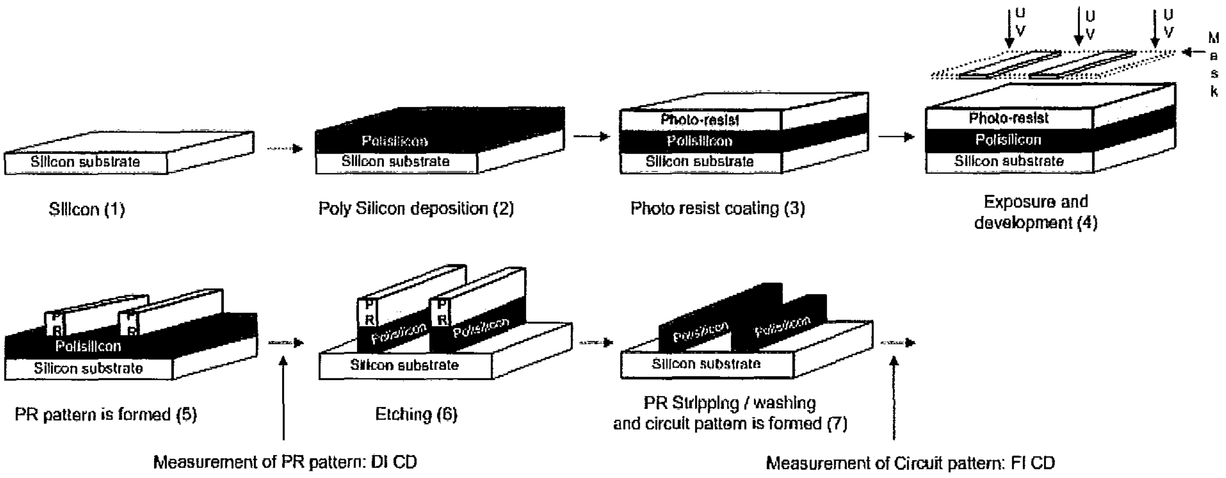

Figure 3.1: Semiconductor fabrication process.

Figure 3.1 illustrates the process steps involved in patterning of the transistor gate. A polysilicon layer is formed on a silicon wafer (2). The wafer is coated with a photoresist which is sensitive to ultraviolet light (3). A mask pattern is exposed and the photoresist is developed (4 and 5). By using gases in a plasma through the resist pattern and removing unwanted areas of film (etching), a circuit pattern is made (6 and 7) (Quirk and Serda, 2000). Among a number of semiconductor manufacturing processes, the patterning of polisilicon gates has been the most important and challenging process in semiconductor manufacturing since it defines the success of semiconductor manufacturing. The linewidth of a gate transistor before etching is called the developed inspection critical dimension (DI CD) and it is called the final inspection critical dimension (FI CD) after etching (Joung et al., 2004). As mentioned before, the tighter control for the DI CD and FI CD is required since the degree of integration on a chip increases. Notice that since optical lenses are used in photolithography, it is impossible to have the best focus over the entire wafer area because silicon wafers have rough surfaces and they also have bow and warpage. Therefore, in this study, we assume that we have the best focus in the central area of a wafer, which is normally happening in semiconductor manufacturing. Additionally, we do not expect any process particles to happen. Then we do a simple simulation to see how the MEWMA control chart performs.

Suppose $\boldsymbol{X} = (X_1, X_2)'$ is a 2×1 random vector representing the DI CD (X_1) and the FI CD (X_2). The quality characteristic X_1 (DI CD) is normally distributed with mean μ_{DI} and standard deviation σ_{DI}, where both μ_{DI} and σ_{DI} are known and correspond to in-control production (Greer et al., 2003). Moreover, the statistical model between the final inspection critical dimensioin (FI CD) and the developed inspection critical dimension (DI CD) is $X_2 = \alpha + \beta X_1 + \epsilon$, where ϵ is a random variable representing noise or environmental factor affecting FI CD and $\epsilon \sim N(0, \sigma_\epsilon^2)$. Thus, the random vector, $\mathbf{X}$ can be expressed as $\mathbf{X} = (X_1, X_2)' = (\text{DI CD, FI CD})' = (X_1, \alpha + \beta X_1 + \epsilon)'$. Suppose that $\sigma_\epsilon < \sigma_{DI}$ and ϵ is independent of X_1.

The mean and variance of X_2 and the covariance between X_1 and X_2 can be calculated as follows.

$$
\begin{aligned}
E(X_2) &= E(\alpha + \beta X_1 + \epsilon) = \alpha + \beta E(X_1) + E(\epsilon) = \alpha + \beta \mu_{DI} \\
Var(X_2) &= Var(\alpha + \beta X_1 + \epsilon) = \beta^2 Var(X_1) + Var(\epsilon) = \beta^2 \sigma_{DI}^2 + \sigma_\epsilon^2.
\end{aligned}
$$

The covariance between X_1 and X_2 is

$$
\begin{aligned}
&\sigma_{X_1,X_2}^2 = \sigma_{DI,FI}^2 = Cov(X_1, X_2) = Cov(X_1, \alpha + \beta X_1 + \epsilon) \\
&= Cov(X_1, \alpha) + Cov(X_1, \beta X_1) + Cov(X_1, \epsilon) \\
&= 0 + \beta Var(X_1) + 0 = \beta \sigma_{DI}^2. \quad \text{(since } X_1 \text{ is independent of } \epsilon)
\end{aligned}
$$

Therefore, the random vector, $\boldsymbol{X}$ is distributed as a bivariate normal distribution as follows.

$$
\boldsymbol{X} \sim N\Big(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_{\boldsymbol{X}}\Big), \text{ where } \boldsymbol{\Sigma}_{\boldsymbol{X}} = \begin{pmatrix} \sigma_{DI}^2 & \sigma_{DI,FI}^2 \\ \sigma_{DI,FI}^2 & \sigma_{FI}^2 \end{pmatrix} \text{ and } \boldsymbol{\mu}_0 = \begin{pmatrix} \mu_{DI} \\ \mu_{FI} \end{pmatrix}.
$$

Furthermore, the covariance matrix, $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ is

$$
\boldsymbol{\Sigma}_{\boldsymbol{X}} = \begin{pmatrix} \sigma_{DI}^2 & \beta \sigma_{DI}^2 \\ \beta \sigma_{DI}^2 & \beta^2 \sigma_{DI}^2 + \sigma_\epsilon^2 \end{pmatrix}.
$$

Let n be the sample size. In semiconductor manufacturing, normally several wafers are selected from a run to measure DI CD and FI CD at regular time intervals when the process is thought to be in-control. For the purpose of simulation, we assume that five wafers are selected from a run ($n = 5$) and the mean of five measurements is used. We also make use of experimental results from the U.S patent (7,541,286 B2) suggesting parameters values $\alpha = -0.03$ and $\beta = 0.98$. Suppose that X_1 is distributed as $N(130, 14.78)$ and ϵ is distributed as $N(0, 1)$, then the distribution of the sample mean $\bar{\boldsymbol{X}}$ is

$$\bar{\boldsymbol{X}} = \frac{\Sigma_{j=1}^{5}\boldsymbol{X}_j}{5} \sim N\left(\boldsymbol{\mu}_0 = \begin{pmatrix} 130.00 \\ 127.37 \end{pmatrix}, \Sigma_{\bar{\boldsymbol{X}}} = \begin{pmatrix} 2.956 & 2.897 \\ 2.897 & 3.100 \end{pmatrix}\right).$$

Figures 3.2 - 3.5 show the results of simulation to compare the performance of the MEWMA chart and the multivariate Shewhart control chart with the condition that the on-target $ARL = 200$ with $p = 2$ and $r = 0.1$. The control limits for a MEWMA and a multivariate Shewhart control chart are 8.66 and 10.6 respectively. That is, the multivariate Shewhart control chart issues an out-of-control signal when $T_t^2 = (\bar{\boldsymbol{X}}_t - \boldsymbol{\mu}_0)'\Sigma_{\bar{\boldsymbol{X}}}^{-1}(\bar{\boldsymbol{X}}_t - \boldsymbol{\mu}_0) > 10.6$, whereas the MEWMA chart procedure signals when $\boldsymbol{Z}_t{}'\Sigma_{\boldsymbol{Z}}^{-1}\boldsymbol{Z}_t > 8.66$, where $\Sigma_{\boldsymbol{Z}_t} = \frac{r}{2-r}\Sigma_{\bar{\boldsymbol{X}}}$. Suppose that the process is initially in-control and a shift in the mean happens at $\tau = 20$. When a small shift happens, it is observed that the MEWMA chart is superior to the multivariate Shewhart control chart. Otherwise, both control charts perform well. For example, when $\delta = 0.5$ (Figure 3.2), the MEWMA issued a signal at 53th run while no indication of an out-of-control condition was observed for the multivariate Shewhart control chart. When $\delta = 1.0$ (Figure 3.3), an out-of-control signal was generated at 30th run for the MEWMA while the multivariate Shewhart chart detected the shift at 43th run.

Now, when a relatively big shift happens ($\delta = 2.0$ and 3.0), both control chart issued a out-of-signal as quickly as possible (Figures 3.4 and 3.5). The above simulation results show good agreement with previously obtained results in Chapter 2 section 5.

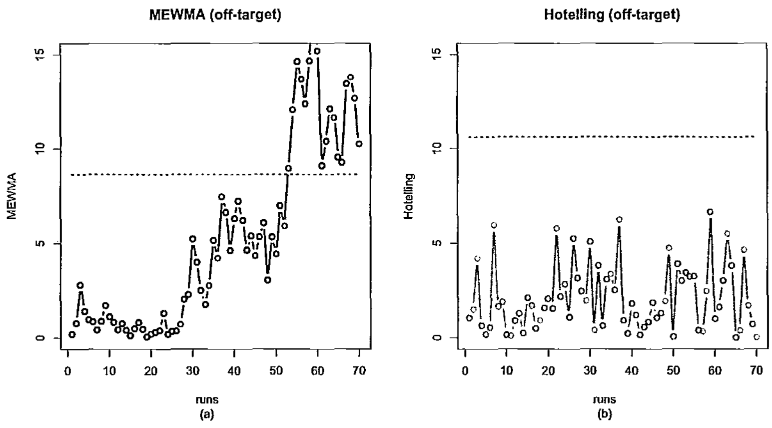

Figure 3.2: Comparision MEWMA and Hotelling control chart with $\delta = 0.5$ (a) MEWMA control chart and (b) Hotelling control chart.

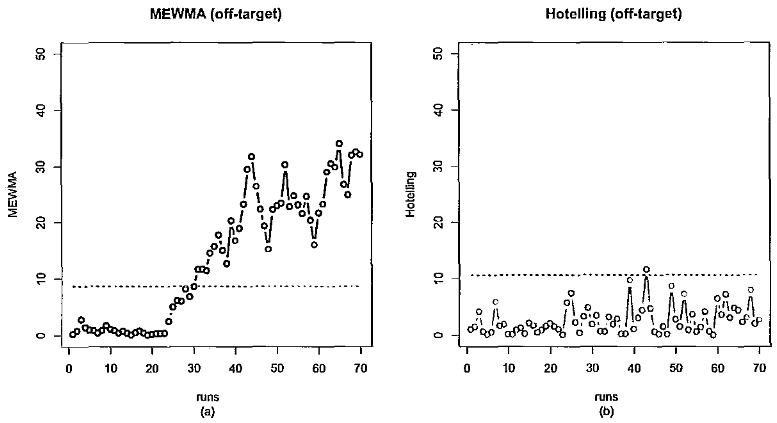

Figure 3.3: Comparision MEWMA and Hotelling control chart with $\delta = 1.0$ (a) MEWMA control chart and (b) Hotelling control chart.

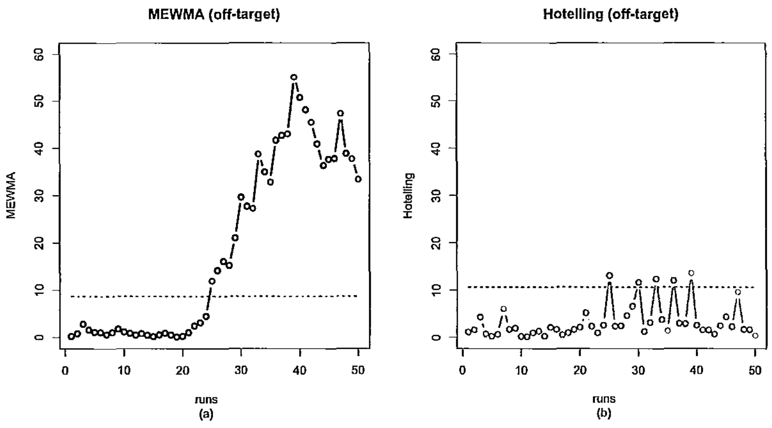

Figure 3.4: Comparision MEWMA and Hotelling control chart with $\delta = 2.0$ (a) MEWMA control chart and (b) Hotelling control chart.

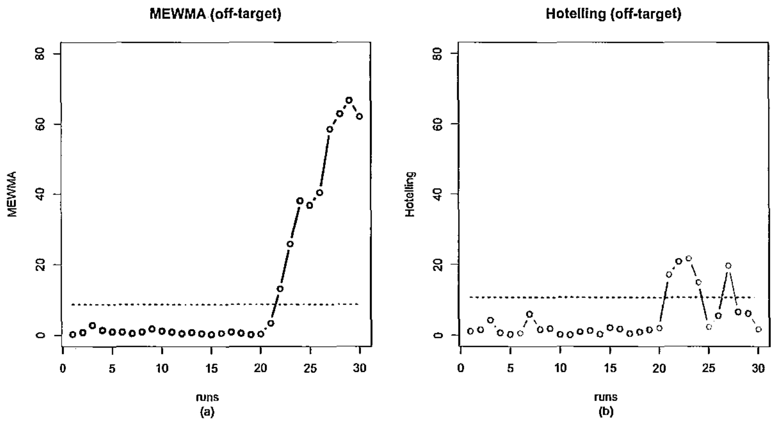

Figure 3.5: Comparision MEWMA and Hotelling control chart with $\delta = 3.0$ (a) MEWMA control chart and (b) Hotelling control chart.

Chapter 4

Propagation of variability

4.1 Introduction

So far, we have discussed a shift in the mean for the process monitoring. Additionally, throughout the manufacturing processes, it is also important to know which stage contributes most to variation. In terms of analysis of variation transmission in manufacturing processes, Lawless et al. (1999) discussed methodology for understanding how variation is added and transmitted across the manufacturing process. Let us assume we have discrete manufacturing stages as follows.

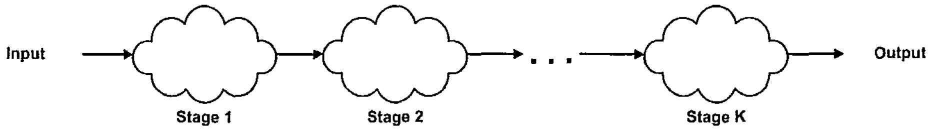

Figure 4.1: Manufacturing processes.

Let X be a quality characteristic of the output and X_k be the measurement at stage k. We have a target value for the quality characteristic but there is variation in the product. That is, there is variation in the quality characteristic, X. As the measurement of interest X passes through the above processes, each step makes a contribution to the variance of X. The objective of this chapter is to understand the amount of variation attributable to different stages of a manufacturing process and to introduce an extension of the variation transmission model suggested by Lawless et al. (1999) by using a simple linear regression model.

4.2 Variation transmission model

Figure 4.2 illustrates how the variation transmission model is applied to semiconductor manufacturing. For simplicity, we use two steppers and one etcher. Y_1 denotes the DI CD while Y_2 denotes the FI CD. The random variable Y_2 is a linear function of an independent variable Y_1 such that

$$Y_2 = \alpha + \beta Y_1 + e \tag{4.1}$$

where α and β are parameters and the random variable $e \sim N(0, \sigma_A^2)$. We assume that the

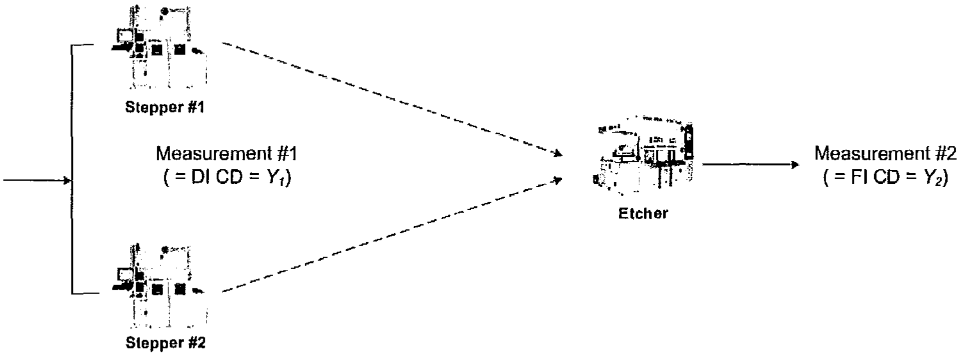

Figure 4.2: Photo and etch stages in a gate patterning process in model I.

DI CD is measured right after photo processing as Y_1 while the FI CD is measured after etching process as Y_2. By defining $\sigma_i^2 = Var(Y_i)$, we can obtain from Equation (4.1)

$$\sigma_2^2 = \beta^2 \sigma_1^2 + \sigma_A^2$$

where $\beta^2\sigma_1^2$ is the variance transmitted through the etcher and σ_A^2 is the variance added by the etching process. Let us define $Y_{1j} = Y_1|Z_1 = j$ where Z_1 is a random variable such that $Z_1 = j$ if a wafer is processed by a stepper j. μ_1 and σ_1^2 are the mean and variance of the first measurement Y_1 and it can be expressed as

$$\mu_1 = \frac{1}{2}\sum_{i=1}^{2}\mu_{1j} \tag{4.2}$$

$$\sigma_1^2 = \frac{1}{2}\sum_{j=1}^{2}\sigma_{1j}^2 + \frac{1}{2}\sum_{j=1}^{2}(\mu_{1j} - \mu_1)^2 \tag{4.3}$$

where $\mu_{1j} = E(Y_1|Z_1 = j)$ and $\sigma_{1j}^2 = Var(Y_1|Z_1 = j)$ (See section A.2.1 in Appendix).

In addition, using Equations (4.2) and (4.3), the mean and variance of Y_2 (=FI CD) are

$$\mu_2 = E(Y_2) = \alpha + \beta\Big(\frac{1}{2}\sum_{j=1}^{2}\mu_{1j}\Big)$$

$$Var(Y_2) = \underbrace{\beta^2\Big(\frac{1}{2}\sum_{j=1}^{2}\sigma_{1j}^2 + \frac{1}{2}\sum_{j=1}^{2}(\mu_{1j} - \mu_1)^2\Big)}_{\text{Variance transmitted}} + \underbrace{\sigma_A^2}_{\text{Variance added}} .$$

In this simple case, the variance added by the etch operation is determined completely by one etcher. The downside of this model is that possibly the etcher will be overloaded since it is the only machine running. Since many etchers and steppers are involved for mass production in semiconductor manufacturing, the model can be extended with the addition of etchers. Intuitively, it would be more complicated if more etchers were involved in the etching process. For simplicity, we have two steppers and two etchers (See Figure 4.3).

Let us define $Y_{2_{jk}} = Y_2|(Z_1 = j, Z_2 = k)$ and Z_2 is a random variable such that $Z_2 = k$ if a wafer is etched by etcher k. Thus, $Y_{2_{jk}}$ is the measurement of polysilicon gate line width (FI CD) processed by stepper j and etcher k ($j = 1, 2; k = 1, 2$). Since we have four possible combinations of steppers and etchers working in a pair, we can think of four linear equations as follows.

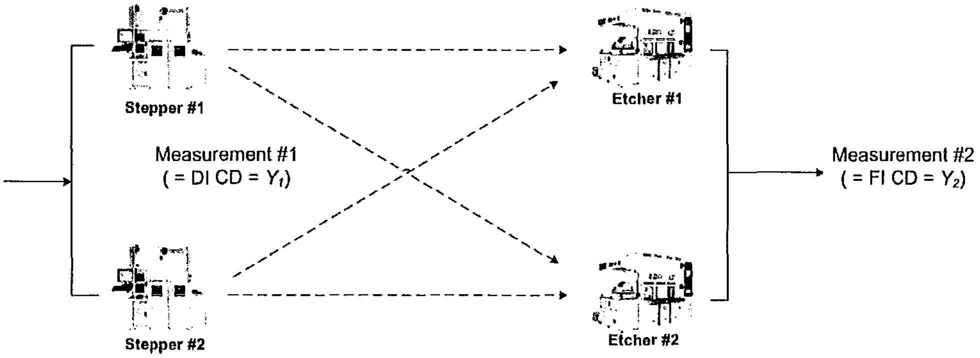

Figure 4.3: Photo and etch stages in a gate patterning process in model II.

Let us assume that

$$\begin{aligned}
Y_{2_{11}} &= Y_2|(Z_1 = 1, Z_2 = 1) = \alpha_1 + \beta_1 Y_{11} + e_1, \text{where} \quad e_1 \sim N(0, \sigma^2_{A_1}) \\
Y_{2_{12}} &= Y_2|(Z_1 = 1, Z_2 = 2) = \alpha_2 + \beta_2 Y_{11} + e_2, \text{where} \quad e_2 \sim N(0, \sigma^2_{A_2}) \\
Y_{2_{21}} &= Y_2|(Z_1 = 2, Z_2 = 1) = \alpha_3 + \beta_3 Y_{12} + e_1, \text{where} \quad e_3 \sim N(0, \sigma^2_{A_1}) \\
Y_{2_{22}} &= Y_2|(Z_1 = 2, Z_2 = 2) = \alpha_4 + \beta_4 Y_{12} + e_2, \text{where} \quad e_4 \sim N(0, \sigma^2_{A_2})
\end{aligned}$$

where α_i and β_i are parameters, $i = 1, 2, 3, 4$ and $\sigma^2_{A_k}$ is the variance added by etcher k, $k = 1, 2$. The expected value of $Y_{2_{jk}}$ from each combination is

$$\begin{aligned}
\mu_{2_{11}} &= E(Y_{2_{11}}) = E[Y_2|(Z_1 = 1, Z_2 = 1)] = \alpha_1 + \beta_1 \mu_{11} \\
\mu_{2_{12}} &= E(Y_{2_{12}}) = E[Y_2|(Z_1 = 1, Z_2 = 2)] = \alpha_2 + \beta_2 \mu_{11} \\
\mu_{2_{21}} &= E(Y_{2_{21}}) = E[Y_2|(Z_1 = 2, Z_2 = 1)] = \alpha_3 + \beta_3 \mu_{12} \\
\mu_{2_{22}} &= E(Y_{2_{22}}) = E[Y_2|(Z_1 = 2, Z_2 = 2)] = \alpha_4 + \beta_4 \mu_{12}.
\end{aligned}$$

The variance of $Y_{2_{jk}}$ from each combination is

$$\sigma^2_{2_{11}} = Var(Y_{2_{11}}) = Var[Y_2|(Z_1 = 1, Z_2 = 1)] = \beta_1^2\sigma_{11}^2 + \sigma^2_{A_1}$$

$$\sigma^2_{2_{12}} = Var(Y_{2_{12}}) = Var[Y_2|(Z_1 = 1, Z_2 = 2)] = \beta_2^2\sigma_{11}^2 + \sigma^2_{A_2}$$

$$\sigma^2_{2_{21}} = Var(Y_{2_{21}}) = Var[Y_2|(Z_1 = 2, Z_2 = 1)] = \beta_3^2\sigma_{12}^2 + \sigma^2_{A_1}$$

$$\sigma^2_{2_{22}} = Var(Y_{2_{22}}) = Var[Y_2|(Z_1 = 2, Z_2 = 2)] = \beta_4^2\sigma_{12}^2 + \sigma^2_{A_2}.$$

We assume that workload is evenly distributed through the combinations mentioned above and each stepper (etcher) is independent of the others, respectively. Thus, the expected value of Y_2 (FI CD) is

$$E(Y_2) = \frac{1}{4}\Big(\sum_{i=1}^{4}\alpha_i + \mu_{11}(\beta_1 + \beta_2) + \mu_{12}(\beta_3 + \beta_4)\Big) = \mu_2 \quad \text{(See section A.2.2 in Appendix).}$$

Thus, the variance of Y_2 is

$$\begin{aligned}
&Var(Y_2) = E(Y_2^2) - [E(Y_2)]^2 \\
&= \frac{1}{4}\Big\{\sigma_{11}^2(\beta_1^2 + \beta_2^2) + \sigma_{12}^2(\beta_3^2 + \beta_4^2) + 2\sum_{k=1}^{2}\sigma^2_{A_k} + (\alpha_1 + \beta_1\mu_{11})^2 + (\alpha_2 + \beta_2\mu_{11})^2 + (\alpha_3 + \beta_3\mu_{12})^2 \\
&+(\alpha_4 + \beta_4\mu_{12})^2\Big\} - \frac{1}{16}\big((\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) + \mu_{11}(\beta_1 + \beta_2) + \mu_{12}(\beta_3 + \beta_4)\big)^2
\end{aligned}$$

(See section A.2.2 in Appendix).

Furthermore, we can compute the overall variance added by the etching operation as well. The overall variance added by the etching operation can be obtained by subtracting the variance transmitted through the etching operation.

$$\begin{aligned}
&\sigma_A^2 = Var(Y_2) - \beta^2\sigma_1^2 \\
&= \frac{1}{4}\Big\{\sigma_{11}^2(\beta_1^2+\beta_2^2) + \sigma_{12}^2(\beta_3^2+\beta_4^2) + 2\sum_{k=1}^{2}\sigma_{A_k}^2 + (\alpha_1+\beta_1\mu_{11})^2 + (\alpha_2+\beta_2\mu_{11})^2 + (\alpha_3+\beta_3\mu_{12})^2 \\
&+(\alpha_4+\beta_4\mu_{12})^2\Big\} - \frac{1}{16}\big((\alpha_1+\alpha_2+\alpha_3+\alpha_4) + \mu_{11}(\beta_1+\beta_2) + \mu_{12}(\beta_3+\beta_4)\big)^2 - \beta^2\sigma_1^2 \\
&\text{where } \sigma_1^2 = \frac{1}{2}\Big(\sum_{j=1}^{2}\sigma_{1j}^2 + (\mu_{1j}-\mu_1)^2\Big).
\end{aligned}$$

The result shows that the variance of the random variable Y_2 becomes more complicated and cumbersome as the number of etchers in operation increases.

As a special case, if a stepper is paired with a specific etcher (for instance, stepper #1(2) only works with etcher #1(2)), more simplified model can be constructed (See Figure 4.4). Denote that Y_{2k} is the measurement of line width (FI CD) processed by etcher k ($k = 1, 2$). Suppose that

$$\begin{aligned}
Y_{21} &= Y_2|Z_2 = 1 = \alpha_1 + \beta_1 Y_{11} + e_1, \text{where} \quad e_1 \sim N(0, \sigma_{A_1}^2) \\
Y_{22} &= Y_2|Z_2 = 2 = \alpha_2 + \beta_2 Y_{12} + e_2, \text{where} \quad e_2 \sim N(0, \sigma_{A_2}^2)
\end{aligned}$$

where α_i and β_i are parameters, $i = 1, 2$ and $\sigma_{A_k}^2$ is the variance added by etcher k, $k = 1, 2$.

We see that

$$\begin{aligned}
\mu_{21} &= E(Y_2|Z_2 = 1) = \alpha_1 + \beta_1\mu_{11} \\
\mu_{22} &= E(Y_2|Z_2 = 2) = \alpha_2 + \beta_2\mu_{12} \\
\sigma_{21}^2 &= Var(Y_2|Z_2 = 1) = \beta_1^2\sigma_{11}^2 + \sigma_{A_1}^2 \\
\sigma_{22}^2 &= Var(Y_2|Z_2 = 2) = \beta_2^2\sigma_{12}^2 + \sigma_{A_2}^2.
\end{aligned}$$

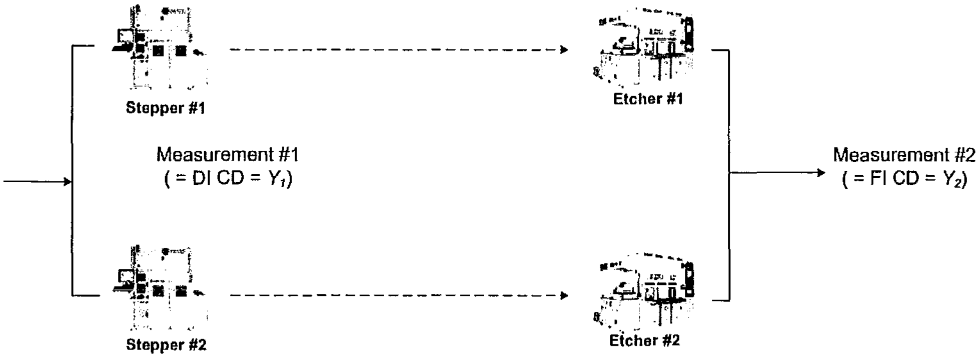

Figure 4.4: Photo and etch stages in a gate patterning process in model III.

If the workload is processed at equal proportion, $E(Y_2), Var(Y_2)$ and σ_A^2 are

$$
\begin{aligned}
E(Y_2) &= \frac{\alpha_1 + \alpha_2 + \beta_1\mu_{11} + \beta_2\mu_{12}}{2} \\
Var(Y_2) &= \frac{\sum_{k=1}^2 \beta_k^2\sigma_{1k}^2 + \sum_{k=1}^2 \sigma_{Ak}^2 + \sum_{k=1}^2(\alpha_k + \beta_k\mu_{1k})^2}{2} \\
&- \frac{\{\sum_{k=1}^2(\alpha_k + \beta_k\mu_{1k})\}^2}{4} \qquad (4.4) \\
\sigma_A^2 &= \frac{\sum_{k=1}^2 \beta_k^2\sigma_{1k}^2 + \sum_{k=1}^2 \sigma_{Ak}^2 + \sum_{k=1}^2(\alpha_k + \beta_k\mu_{1k})^2}{2} - \frac{\{\sum_{k=1}^2(\alpha_k + \beta_k\mu_{1k})\}^2}{4} - \beta^2\sigma_1^2 \\
&\text{where } \sigma_1^2 = \frac{1}{2}\Big(\sum_{j=1}^{2} \sigma_{1j}^2 + (\mu_{1j} - \mu_1)^2\Big).
\end{aligned}
$$

4.3 Numerical example

The following is a simple numerical example for the second case introduced in the previous section. Suppose that we have

$$Y_1|(Z_1 = 1) = Y_{11} \sim N(130.2, 14.5)$$

$$Y_1|(Z_1 = 2) = Y_{12} \sim N(129.8, 15.0)$$

and each stepper has capability of processing 1,000 wafers on a daily basis and move them to a specific etcher and one measurement is captured from each wafer. We use the following straight line equations to estimate parameters $\alpha_1, \alpha_2, \beta_1$ and β_2 in equation.

$$\widehat{y}_{21} = \widehat{\alpha}_1 + \widehat{\beta}_1 y_{11}$$
$$\widehat{y}_{22} = \widehat{\alpha}_2 + \widehat{\beta}_2 y_{12}.$$

For the purpose of simulation, we use the following equations but we pretend not to know them.

$$Y_{21} = -0.02 + 0.97Y_{11} + e_1, \text{where} \quad e_1 \sim N(0, 1.5)$$
$$Y_{22} = -0.03 + 0.98Y_{12} + e_2, \text{where} \quad e_2 \sim N(0, 1.2).$$

Let us denote that Y_{1jk} is the measurement of Y_1 from stepper j for wafer k and Y_{2jk} is the measurement of Y_2 from etcher j for wafer k $\quad (k = 1, 2, \cdots, 1000; j = 1, 2)$.

Here are the estimates for parameters.

$$
\begin{aligned}
\widehat{\sigma}_{11}^2 &= \sum_{k=1}^{1000}(y_{11k}-\bar{y}_{11})^2/1000 = 15.6328\\
\widehat{\sigma}_{12}^2 &= \sum_{k=1}^{1000}(y_{12k}-\bar{y}_{12})^2/1000 = 15.1519\\
\widehat{\sigma}_{1}^2 &= \sum_{j=1}^{2}\sum_{k=1}^{1000}(y_{1jk}-\bar{y}_{1})^2/2000 = 15.4634\\
\widehat{\sigma}_{2}^2 &= \sum_{j=1}^{2}\sum_{k=1}^{1000}(y_{2jk}-\bar{y}_{2})^2/2000 = 16.3839 \qquad (4.5)\\
\widehat{\beta}_1 &= \left\{1000\sum_{k=1}^{1000} y_{11k}y_{21k} - \sum_{k=1}^{1000} y_{11k}\sum_{k=1}^{1000} y_{21k}\right\}\Big/\left\{1000\sum_{k=1}^{1000} y_{11k}^2 - \Big(\sum_{k=1}^{1000} y_{11k}\Big)^2\right\} = 0.9692\\
\widehat{\alpha}_1 &= \bar{y}_{21} - \widehat{\beta}_1\bar{y}_{11} = 0.0422\\
\widehat{\beta}_2 &= \left\{1000\sum_{k=1}^{1000} y_{12k}y_{22k} - \sum_{k=1}^{1000} y_{12k}\sum_{k=1}^{1000} y_{22k}\right\}\Big/\left\{1000\sum_{k=1}^{1000} y_{12k}^2 - \Big(\sum_{k=1}^{1000} y_{12k}\Big)^2\right\} = 0.9960\\
\widehat{\alpha}_2 &= \bar{y}_{22} - \widehat{\beta}_2\bar{y}_{12} = -2.1300\\
\widehat{\beta} &= \left\{2000\sum_{j=1}^{2}\sum_{k=1}^{1000} y_{1jk}y_{2jk} - \sum_{j=1}^{2}\sum_{k=1}^{1000} y_{1jk}\sum_{j=1}^{2}\sum_{k=1}^{1000} y_{2jk}\right\}\Big/\left\{2000\sum_{j=1}^{2}\sum_{k=1}^{1000} y_{1jk}^2 - \Big(\sum_{j=1}^{2}\sum_{k=1}^{1000} y_{1jk}\Big)^2\right\}\\
&= 0.9711.
\end{aligned}
$$

Now, by using Equation (4.4) and replacing the parameters by the estimates calculated above, we obtain the estimate of $Var(Y_2) = 16.3622$, which provides good agreement with the result (4.5). The estimate of variation transmitted ($\widehat{\beta}^2\widehat{\sigma}_1^2$) is 14.5825 and the estimate of variation added ($\widehat{\sigma}_A^2$) by etching operation is 1.7798. Thus, in our numerical example, the variance added by etching operation accounts for about 11% of the total variance of Y_2. It is observed that most contribution comes from Y_1 (DI CD).

Chapter 5

Conclusions

The multivariate exponentially weighted moving average (MEWMA) control chart is an extension of the well-known univariate EWMA chart applicable where product quality is characterized by two or more variables. It contains the well-known Shewhart chi-squared chart as a particular case. Several aspects of the run length distribution not studied before are discussed in detail in this thesis. The methods are applied to the problem of monitoring a semiconductor manufacturing process where bivariate quality is measured. The thesis also discusses methods to model and quantify variability built in a manufacturing process.

The previous study for the MEWMA by Runger and Prabhu (1997) was concentrated on two areas: the on-target run length analysis and the off-target run length analysis. For the on-target analysis, we derive the probability mass function, the second, third and fourth moment of the run length distribution as closed forms respectively. When the off-target case was analyzed before, it was assumed that the process mean shift happened at the beginning of the operation (the zero-state case). We introduce a general off-target form such that the mean shift can happen at any time, including the beginning of the operation (the steady-state case). Here is a generalization of the probability mass function of the run length distribution for the MEWMA control chart.

$$f(n) = P(N = n) = \begin{cases} \boldsymbol{s}'\boldsymbol{P}_0{}^{n-1}(\boldsymbol{I} - \boldsymbol{P}_0)\mathbf{1} & \text{if } n = 1,\, 2,\, \ldots, \tau - 1, \\ \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\boldsymbol{P}_1{}^{n-\tau}(\boldsymbol{I} - \boldsymbol{P}_1)\mathbf{1} & \text{if } n = \tau,\, \tau + 1,\, \ldots. \end{cases}$$

where $\boldsymbol{P}_0$ is an in-control transition matrix and $\boldsymbol{P}_1$ is an out-of-control transition matrix. With the general probability mass function, more derivations are made and all the results are consistent with those of past studies.

The MEWMA scheme is well-known for detecting a small shift and a good way to improve the ability to detect a small shift is to find an optimum smoothing parameter. In the general off-target case ($\tau \geq 1$), the smoothing parameter shows the optimum parameter value increases as transition time (τ) increases. Moreover, as either the in-control average run length (ARL_0) or the number of variables (p) increases, the optimum parameter value increases slowly to the change of transition time τ.

When a small shift happens at the beginning ($\tau = 1$), the MEWMA control chart is very effective in detecting the change. Our interest is how the control chart behaves in case that a process change happens at a different transition time. Even though the transition time changes, just as in the previous case, the MEWMA control chart still outperforms the multivariate Shewhart control chart in performance when the shift is small. Otherwise both control charts perform well.

As an application, we suggest a bivariate normal distribution model for the MEWMA control chart and apply the model to the main semiconductor manufacturing processes. Since the critical dimension of polysilicon gate has been continued to shrink, the impact of environment errors can not be negligible any more and the tighter control over the DI CD and FI CD is required. The model is defined as follows.

$$\boldsymbol{X} = (X_1, \alpha + \beta X_1 + \epsilon)' \sim N\Big(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_{\boldsymbol{X}}\Big), \text{ where } \boldsymbol{\mu}_0 \text{ is the process mean in control.}$$

The variation transmission model suggested by Lawless et al. (1999) is based on that products processed by a multiple of machines, move from one operation to the next stage and are processed by a machine. Since semiconductor manufacturing is composed of hundreds of processes, it is more realistic to consider the case that products are processed by a multiple of machines from one stage to another. We suggest an extension for the original variation transmission model. By using the extended model, the total variance transmitted can be calculated.

Appendix A

Derivations

A.1 Kronecker product

Let $\boldsymbol{A}$ be an $n \times p$ matrix and $\boldsymbol{B}$ be an $m \times q$ matrix . The $mn \times pq$ matrix

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{pmatrix} a_{1,1}\boldsymbol{B} & \cdots & a_{1,p}\boldsymbol{B} \\ a_{2,1}\boldsymbol{B} & \cdots & a_{2,p}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{n,1}\boldsymbol{B} & \cdots & a_{n,p}\boldsymbol{B} \end{pmatrix}$$

is called the Kronecker product of $\boldsymbol{A}$ and $\boldsymbol{B}$.

For example, let $\boldsymbol{A} = \begin{pmatrix} 2 & 1 & 3 \\ 4 & 2 & 5 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$.

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{pmatrix} 2 & 4 & 1 & 2 & 3 & 6 \\ 6 & 8 & 3 & 4 & 9 & 12 \\ 4 & 8 & 2 & 4 & 5 & 10 \\ 12 & 16 & 6 & 8 & 15 & 20 \end{pmatrix}.$$

A.2 Derivations

A.2.1 Derivation of the mean and variance in the transmission model I

Let $\mu_{1j} = E(Y_1|Z_1 = j)$ and $\sigma_{1j}^2 = Var(Y_1|Z_1 = j)$ where $j = 1, 2$. Given that the stepper j is in operation, μ_{1j} and σ_{1j}^2 are the expected value and variance of the measurement Y_1 respectively. Since each stepper processes the parts at equal proportion, it is assumed that $P(Z_1 = 1) = P(Z_1 = 2) = \frac{1}{2}$. The expected value of Y_1 is

$$E(Y_1) = \sum_{j=1}^{2}\{E(Y_1|Z_1 = j)P(Z_1 = j)\} = \frac{1}{2}\sum_{j=1}^{2} E(Y_1|Z_1 = j)$$

since $P(Z_1 = j) = \frac{1}{2}$.

$$\text{Thus, } E(Y_1) = (1/2)\sum_{j=1}^{2}\mu_{1j} = \mu_1. \tag{A.1}$$

The conditional expectation of Y_1^2 is

$$E(Y_1^2|Z_1 = j) = Var(Y_1|Z_1 = j) + [E(Y_1|Z_1 = j)]^2 = \sigma_{1j}^2 + \mu_{1j}^2.$$

Thus, the second moment of Y_1 can be expressed as follows.

$$\begin{aligned} E(Y_1^2) &= \sum_{j=1}^{2}\{E(Y_1^2|Z_1 = j)P(Z_1 = j)\} \\ &= \sum_{j=1}^{2}\{(\sigma_{1j}^2 + \mu_{1j}^2)P(Z_1 = j)\} = \frac{1}{2}\sum_{j=1}^{2}(\sigma_{1j}^2 + \mu_{1j}^2) \\ &= \frac{1}{2}\sum_{j=1}^{2}\sigma_{1j}^2 + \frac{1}{2}\sum_{j=1}^{2}\mu_{1j}^2. \end{aligned}$$

Thus, for the variance of Y_1, $\sigma_1^2 = Var(Y_1) = E(Y_1^2) - [E(Y_1)]^2 = \frac{1}{2}\sum_{j=1}^{2}\sigma_{1j}^2 + \frac{1}{2}\sum_{j=1}^{2}\mu_{1j}^2 - \mu_1^2.$

Since it is known that

$$
\begin{aligned}
&\sum_{j=1}^{2}(\mu_{1j}-\mu_1)^2 = \sum_{j=1}^{2}(\mu_{1j}^2 - 2\mu_1\mu_{1j} + \mu_1^2) \\
= \;& \sum_{j=1}^{2}\mu_{1j}^2 - 2\mu_1\sum_{j=1}^{2}\mu_{1j} + 2\mu_1^2 = \sum_{j=1}^{2}\mu_{1j}^2 - 2\mu_1(2\mu_1) + 2\mu_1^2 \\
= \;& \sum_{j=1}^{2}\mu_{1j}^2 - 4\mu_1^2 + 2\mu_1^2 = \sum_{j=1}^{2}\mu_{1j}^2 - 2\mu_1^2
\end{aligned}
$$

we obtain

$$\frac{1}{2}\sum_{j=1}^{2}\mu_{1j}^2 - 2\mu_1^2 = \frac{1}{2}\sum_{j=1}^{2}\mu_{1j}^2 - \mu_1^2 = \frac{1}{2}\sum_{j=1}^{2}(\mu_{1j}-\mu_1)^2.$$

Hence, the variance of Y_1 is

$$
\begin{aligned}
\sigma_1^2 &= Var(Y_1) = E(Y_1^2) - [E(Y_1)]^2 \\
&= \frac{1}{2}\sum_{j=1}^{2}\sigma_{1j}^2 + \frac{1}{2}\sum_{j=1}^{2}(\mu_{1j}-\mu_1)^2.
\end{aligned}
$$

A.2.2 Derivation of the mean and variance in the transmission model II

$$E(Y_2) = \sum_{j=1}^{2}\sum_{k=1}^{2}\{E(Y_2|Z_1=j, Z_2=k)P(Z_1=j)P(Z_2=k|Z_1=j)\}.$$

For simplicity, we assume that $P(Z_1=j)P(Z_2=k|Z_1=j)=\frac{1}{4}$ where $j,k=1,2$. Thus, we obtain

$$
\begin{aligned}
E(Y_2) &= \frac{1}{4}\sum_{j=1}^{2}\sum_{k=1}^{2}E(Y_2|Z_1=j, Z_2=k) \\
= \;& \frac{1}{4}\left(\sum_{i=1}^{4}\alpha_i + \mu_{11}(\beta_1+\beta_2) + \mu_{12}(\beta_3+\beta_4)\right) = \mu_2.
\end{aligned}
$$

The conditional expectation of Y_2^2 is

$$E(Y_2^2|Z_1 = j, Z_2 = k) = Var(Y_2|Z_1 = j, Z_2 = k) + [E(Y_2|Z_1 = j, Z_2 = k)]^2 = \sigma^2_{2_{jk}} + \mu^2_{2_{jk}}.$$

Thus, the second moment of Y_2 is

$$\begin{aligned}
E(Y_2^2) &= \sum_{j=1}^{2}\sum_{k=1}^{2}\{E(Y_2^2|Z_1 = j, Z_2 = k)P(Z_1 = j)P(Z_2 = k|Z_1 = j)\} \\
&= \sum_{j=1}^{2}\sum_{k=1}^{2}\{(\sigma^2_{2_{jk}} + (\mu_{2_{jk}})^2 P(Z_1 = j)P(Z_2 = k|Z_1 = j)\} \\
&= \frac{1}{4}\sum_{j=1}^{2}\sum_{k=1}^{2}(\sigma^2_{2_{jk}} + \mu^2_{2_{jk}}) \\
&= \frac{1}{4}\{\sigma^2_{11}(\beta_1^2 + \beta_2^2) + \sigma^2_{12}(\beta_3^2 + \beta_4^2) + 2\sum_{k=1}^{2}\sigma^2_{A_k} + (\alpha_1 + \beta_1\mu_{11})^2 + (\alpha_2 + \beta_2\mu_{11})^2 + (\alpha_3 + \beta_3\mu_{12})^2 \\
&+ (\alpha_4 + \beta_4\mu_{12})^2\}.
\end{aligned}$$

A.2.3 Identities

Identity 1. *For every positive integer a,*

$$I + P + P^2 + \cdots P^{a-1} = (I - P^a)(I - P)^{-1} = (I - P)^{-1}(I - P^a). \qquad \text{(A.2)}$$

Proof.

$$\begin{aligned}
&(I + P + P^2 + \cdots + P^{a-2} + P^{a-1})(I - P) \\
= \;& (I + P + P^2 + \cdots + P^{a-2} + P^{a-1}) - (I + P + P^2 + \cdots + P^{a-2} + P^{a-1} + P^a) \\
= \;& I - P^a.
\end{aligned}$$

Thus, $(I + P + P^2 + \cdots + P^{a-2} + P^{a-1})(I - P) = I - P^a$.

Multiplying on the right on both sides by $(I - P)^{-1}$ yields

$$(\boldsymbol{I}+\boldsymbol{P}+\boldsymbol{P}^2+\cdots+\boldsymbol{P}^{a-2}+\boldsymbol{P}^{a-1})=(\boldsymbol{I}-\boldsymbol{P}^a)(\boldsymbol{I}-\boldsymbol{P})^{-1}.$$

Identity 2. *For every positive integer a,*

$$\boldsymbol{I}+2\boldsymbol{P}+3\boldsymbol{P}^2+\cdots+a\boldsymbol{P}^{a-1}=[\boldsymbol{I}+(a\boldsymbol{P}-(a+1)\boldsymbol{I})\boldsymbol{P}^a](\boldsymbol{I}-\boldsymbol{P})^{-2}=(\boldsymbol{I}-\boldsymbol{P})^{-2}[\boldsymbol{I}+(a\boldsymbol{P}-(a+1)\boldsymbol{I})\boldsymbol{P}^a]. \tag{A.3}$$

Proof.

$$\begin{aligned}
& [\boldsymbol{I}+2\boldsymbol{P}+3\boldsymbol{P}^2+4\boldsymbol{P}^3+\cdots+(a-2)\boldsymbol{P}^{a-3}+(a-1)\boldsymbol{P}^{a-2}+a\boldsymbol{P}^{a-1}](\boldsymbol{I}-\boldsymbol{P})^{-2}\\
= & [\boldsymbol{I}+2\boldsymbol{P}+3\boldsymbol{P}^2+4\boldsymbol{P}^3+\cdots+(a-2)\boldsymbol{P}^{a-3}+(a-1)\boldsymbol{P}^{a-2}+a\boldsymbol{P}^{a-1}](\boldsymbol{I}-2\boldsymbol{P}+\boldsymbol{P}^2)\\
= & \boldsymbol{I}+2\boldsymbol{P}+3\boldsymbol{P}^2+4\boldsymbol{P}^3+\cdots+a\boldsymbol{P}^{a-1}-(2\boldsymbol{P}+4\boldsymbol{P}^2+6\boldsymbol{P}^3+\cdots+2(a-1)\boldsymbol{P}^{a-1}+2a\boldsymbol{P}^a)\\
& +\boldsymbol{P}^2+2\boldsymbol{P}^3+\cdots+(a-2)\boldsymbol{P}^{a-1}+(a-1)\boldsymbol{P}^a+a\boldsymbol{P}^{a+1}\\
= & \boldsymbol{I}-2a\boldsymbol{P}^a+(a-1)\boldsymbol{P}^a+a\boldsymbol{P}^{a+1}\\
= & \boldsymbol{I}-(a+1)\boldsymbol{P}^a+a\boldsymbol{P}^{a+1}\\
= & \boldsymbol{I}+[a\boldsymbol{P}-(a+1)\boldsymbol{I}]\boldsymbol{P}^a.
\end{aligned}$$

Thus, $[\boldsymbol{I}+2\boldsymbol{P}+3\boldsymbol{P}^2+4\boldsymbol{P}^3+\cdots+a\boldsymbol{P}^{a-1}](\boldsymbol{I}-\boldsymbol{P})^2=\boldsymbol{I}+[a\boldsymbol{P}-(a+1)\boldsymbol{I}]\boldsymbol{P}^a$.

Multiplying on the right both sides by $(\boldsymbol{I}-\boldsymbol{P})^{-2}$ yields

$\boldsymbol{I}+2\boldsymbol{P}+3\boldsymbol{P}^2+4\boldsymbol{P}^3+\cdots+a\boldsymbol{P}^{a-1}=[\boldsymbol{I}+(a\boldsymbol{P}-(a+1)\boldsymbol{I})\boldsymbol{P}^a](\boldsymbol{I}-\boldsymbol{P})^{-2}$.

Similarly, we can show that

$\boldsymbol{I}+2\boldsymbol{P}+3\boldsymbol{P}^2+4\boldsymbol{P}^3+\cdots+a\boldsymbol{P}^{a-1}=(\boldsymbol{I}-\boldsymbol{P})^{-2}[\boldsymbol{I}+(a\boldsymbol{P}-(a+1)\boldsymbol{I})\boldsymbol{P}^a]$.

Identity 3. *For every positive integer a,*

$$\begin{aligned}
&(2\cdot 1)\boldsymbol{I}+(3\cdot 2)\boldsymbol{P}+(4\cdot 3)\boldsymbol{P}^2+\cdots+a(a-1)\boldsymbol{P}^{a-2}\\
= \;& [2\boldsymbol{I}-a(a+1)\boldsymbol{P}^{a-1}+2(a-1)(a+1)\boldsymbol{P}^{a}-a(a-1)\boldsymbol{P}^{a+1}](\boldsymbol{I}-\boldsymbol{P})^{-3}\\
= \;& (\boldsymbol{I}-\boldsymbol{P})^{-3}[2\boldsymbol{I}-a(a+1)\boldsymbol{P}^{a-1}+2(a-1)(a+1)\boldsymbol{P}^{a}-a(a-1)\boldsymbol{P}^{a+1}]. \qquad \text{(A.4)}
\end{aligned}$$

Proof.

$$\begin{aligned}
&[(2\cdot 1)\boldsymbol{I}+(3\cdot 2)\boldsymbol{P}+(4\cdot 3)\boldsymbol{P}^2+\cdots+a(a-1)\boldsymbol{P}^{a-2}](\boldsymbol{I}-\boldsymbol{P})^{3}\\
= \;& [(2\cdot 1)\boldsymbol{I}+(3\cdot 2)\boldsymbol{P}+(4\cdot 3)\boldsymbol{P}^2+\cdots+a(a-1)\boldsymbol{P}^{a-2}](\boldsymbol{I}-3\boldsymbol{P}+3\boldsymbol{P}^2-\boldsymbol{P}^3)\\
= \;& 2\boldsymbol{I}-a(a+1)\boldsymbol{P}^{a-1}+2(a-1)(a+1)\boldsymbol{P}^{a}-a(a-1)\boldsymbol{P}^{a+1}. \qquad \text{(A.5)}
\end{aligned}$$

Multiplying by $(\boldsymbol{I}-\boldsymbol{P})^{-3}$ on the right-hand side of both sides yields

$(2\cdot 1)\boldsymbol{I}+(3\cdot 2)\boldsymbol{P}+(4\cdot 3)\boldsymbol{P}^2+\cdots+a(a-1)\boldsymbol{P}^{a-2}$

$=[2\boldsymbol{I}-a(a+1)\boldsymbol{P}^{a-1}+2(a-1)(a+1)\boldsymbol{P}^{a}-a(a-1)\boldsymbol{P}^{a+1}](\boldsymbol{I}-\boldsymbol{P})^{-3}$.

Similarly, we can show that

$(2\cdot 1)\boldsymbol{I}+(3\cdot 2)\boldsymbol{P}+(4\cdot 3)\boldsymbol{P}^2+\cdots+a(a-1)\boldsymbol{P}^{a-2}$

$=(\boldsymbol{I}-\boldsymbol{P})^{-3}[2\boldsymbol{I}-a(a+1)\boldsymbol{P}^{a-1}+2(a-1)(a+1)\boldsymbol{P}^{a}-a(a-1)\boldsymbol{P}^{a+1}]$.

Identity 4. *For every positive integer a,*

$$\begin{aligned}
&(3\cdot 2\cdot 1)\boldsymbol{I}+(4\cdot 3\cdot 2)\boldsymbol{P}+(5\cdot 4\cdot 3)\boldsymbol{P}^2+\cdots+a(a-1)(a-2)\boldsymbol{P}^{a-3}\\
= \;& [(3\cdot 2\cdot 1)\boldsymbol{I}-(a-1)a(a+1)\boldsymbol{P}^{a-2}+3(a-2)a(a+1)\boldsymbol{P}^{a-1}\\
& -3(a-2)(a-1)(a+1)\boldsymbol{P}^{a}+(a-2)(a-1)a\boldsymbol{P}^{a+1}](\boldsymbol{I}-\boldsymbol{P})^{-4}\\
= \;& (\boldsymbol{I}-\boldsymbol{P})^{-4}[(3\cdot 2\cdot 1)\boldsymbol{I}-(a-1)a(a+1)\boldsymbol{P}^{a-2}+3(a-2)a(a+1)\boldsymbol{P}^{a-1}\\
& -3(a-2)(a-1)(a+1)\boldsymbol{P}^{a}+(a-2)(a-1)a\boldsymbol{P}^{a+1}]. \qquad \text{(A.6)}
\end{aligned}$$

Proof.

$$\begin{aligned}
& [(3\cdot 2\cdot 1)\boldsymbol{I} + (4\cdot 3\cdot 2)\boldsymbol{P} + (5\cdot 4\cdot 3)\boldsymbol{P}^2 + \cdots + a(a-1)(a-2)\boldsymbol{P}^{a-3}] \\
& \times(\boldsymbol{I}-\boldsymbol{P})^4 = [(3\cdot 2\cdot 1)\boldsymbol{I} + (4\cdot 3\cdot 2)\boldsymbol{P} + (5\cdot 4\cdot 3)\boldsymbol{P}^2 + \cdots + a(a-1)(a-2)\boldsymbol{P}^{a-3}] \\
& \times(\boldsymbol{I} - 4\boldsymbol{P} + 6\boldsymbol{P}^2 - 4\boldsymbol{P}^3 + \boldsymbol{P}^4) \\
= {} & (3\cdot 2\cdot 1)\boldsymbol{I} - (a-1)a(a+1)\boldsymbol{P}^{a-2} + 3(a-2)a(a+1)\boldsymbol{P}^{a-1} \\
& -3(a-2)(a-1)(a+1)\boldsymbol{P}^{a} + (a-2(a-1)a\boldsymbol{P}^{a+1}.
\end{aligned}$$

Multiplying by $(\boldsymbol{I}-\boldsymbol{P})^{-4}$ on the right both sides yields

$$\begin{aligned}
& (3\cdot 2\cdot 1)\boldsymbol{I} + (4\cdot 3\cdot 2)\boldsymbol{P} + (5\cdot 4\cdot 3)\boldsymbol{P}^2 + \cdots + a(a-1)(a-2)\boldsymbol{P}^{a-3} \\
= {} & [(3\cdot 2\cdot 1)\boldsymbol{I} - (a-1)a(a+1)\boldsymbol{P}^{a-2} + 3(a-2)a(a+1)\boldsymbol{P}^{a-1} \\
& -3(a-2)(a-1)(a+1)\boldsymbol{P}^{a} + (a-2(a-1)a\boldsymbol{P}^{a+1}](\boldsymbol{I}-\boldsymbol{P})^{-4}.
\end{aligned}$$

Similarly, we can show that

$$\begin{aligned}
& (3\cdot 2\cdot 1)\boldsymbol{I} + (4\cdot 3\cdot 2)\boldsymbol{P} + (5\cdot 4\cdot 3)\boldsymbol{P}^2 + \cdots + a(a-1)(a-2)\boldsymbol{P}^{a-3} \\
= {} & (\boldsymbol{I}-\boldsymbol{P})^{-4}[(3\cdot 2\cdot 1)\boldsymbol{I} - (a-1)a(a+1)\boldsymbol{P}^{a-2} + 3(a-2)a(a+1)\boldsymbol{P}^{a-1} \\
& -3(a-2)(a-1)(a+1)\boldsymbol{P}^{a} + (a-2(a-1)a\boldsymbol{P}^{a+1}].
\end{aligned}$$

Identity 5. *For every positive integer a,*

$$\begin{aligned}
& (4\cdot 3\cdot 2\cdot 1)\boldsymbol{I} + (5\cdot 4\cdot 3\cdot 2)\boldsymbol{P} + (6\cdot 5\cdot 4\cdot 3)\boldsymbol{P}^2 + \cdots + a(a-1)(a-2)(a-3)\boldsymbol{P}^{a-4} \\
= \; & [(4\cdot 3\cdot 2\cdot 1)\boldsymbol{I} - (a-2)(a-1)a(a+1)\boldsymbol{P}^{a-3} + 4(a-3)(a-1)a(a+1)\boldsymbol{P}^{a-2} \\
& -6(a-3)(a-2)a(a+1)\boldsymbol{P}^{a-1} + 4(a-3)(a-2)(a+1)\boldsymbol{P}^{a} \\
& -(a-3)(a-2)(a-1)a\boldsymbol{P}^{a+1}](\boldsymbol{I}-\boldsymbol{P})^{-5} \\
= \; & (\boldsymbol{I}-\boldsymbol{P})^{-5}[(4\cdot 3\cdot 2\cdot 1)\boldsymbol{I} - (a-2)(a-1)a(a+1)\boldsymbol{P}^{a-3} + 4(a-3)(a-1)a(a+1)\boldsymbol{P}^{a-2} \\
& -6(a-3)(a-2)a(a+1)\boldsymbol{P}^{a-1} + 4(a-3)(a-2)(a+1)\boldsymbol{P}^{a} \\
& -(a-3)(a-2)(a-1)a\boldsymbol{P}^{a+1}]. \qquad \text{(A.7)}
\end{aligned}$$

Proof.

$$\begin{aligned}
& [(4\cdot 3\cdot 2\cdot 1)\boldsymbol{I} + (5\cdot 4\cdot 3\cdot 2)\boldsymbol{P} + (6\cdot 5\cdot 4\cdot 3)\boldsymbol{P}^2 + \cdots + a(a-1)(a-2)(a-3)\boldsymbol{P}^{a-4}] \\
& \times(\boldsymbol{I}-\boldsymbol{P})^5 = [(4\cdot 3\cdot 2\cdot 1)\boldsymbol{I} + (5\cdot 4\cdot 3\cdot 2)\boldsymbol{P} + (6\cdot 5\cdot 4\cdot 3)\boldsymbol{P}^2 + \cdots + a(a-1)(a-2)(a-3)\boldsymbol{P}^{a-4}] \\
& \times(\boldsymbol{I} - 5\boldsymbol{P} + 10\boldsymbol{P}^2 - 10\boldsymbol{P}^3 + 5\boldsymbol{P}^4 - \boldsymbol{P}^5) \\
= \; & (4\cdot 3\cdot 2\cdot 1)\boldsymbol{I} - (a-2)(a-1)a(a+1)\boldsymbol{P}^{a-3} + 4(a-3)(a-1)a(a+1)\boldsymbol{P}^{a-2} \\
& -6(a-3)(a-2)a(a+1)\boldsymbol{P}^{a-1} + 4(a-3)(a-2)(a+1)\boldsymbol{P}^{a} \\
& -(a-3)(a-2)(a-1)a\boldsymbol{P}^{a+1}.
\end{aligned}$$

Multiplying by $(\boldsymbol{I}-\boldsymbol{P})^{-5}$ on the right both sides yields

$$\begin{aligned}
& (4\cdot 3\cdot 2\cdot 1)\boldsymbol{I} + (5\cdot 4\cdot 3\cdot 2)\boldsymbol{P} + (6\cdot 5\cdot 4\cdot 3)\boldsymbol{P}^2 + \cdots + a(a-1)(a-2)(a-3)\boldsymbol{P}^{a-4} \\
= \; & [(4\cdot 3\cdot 2\cdot 1)\boldsymbol{I} - (a-2)(a-1)a(a+1)\boldsymbol{P}^{a-3} + 4(a-3)(a-1)a(a+1)\boldsymbol{P}^{a-2} \\
& -6(a-3)(a-2)a(a+1)\boldsymbol{P}^{a-1} + 4(a-3)(a-2)(a+1)\boldsymbol{P}^{a} \\
& -(a-3)(a-2)(a-1)a\boldsymbol{P}^{a+1}](\boldsymbol{I}-\boldsymbol{P})^{-5}.
\end{aligned}$$

Similarly, we can show that

$$
\begin{aligned}
&(4\cdot 3\cdot 2\cdot 1)\boldsymbol{I}+(5\cdot 4\cdot 3\cdot 2)\boldsymbol{P}+(6\cdot 5\cdot 4\cdot 3)\boldsymbol{P}^2+\cdots+a(a-1)(a-2)(a-3)\boldsymbol{P}^{a-4}\\
= \;&(\boldsymbol{I}-\boldsymbol{P})^{-5}[(4\cdot 3\cdot 2\cdot 1)\boldsymbol{I}-(a-2)(a-1)a(a+1)\boldsymbol{P}^{a-3}+4(a-3)(a-1)a(a+1)\boldsymbol{P}^{a-2}\\
&-6(a-3)(a-2)a(a+1)\boldsymbol{P}^{a-1}+4(a-3)(a-2)(a+1)\boldsymbol{P}^{a}-(a-3)(a-2)(a-1)a\boldsymbol{P}^{a+1}].
\end{aligned}
$$

Since the matrix $\boldsymbol{P}$ is made up of transient states, $\boldsymbol{P}^a \to \boldsymbol{0}$ as $a \to \infty$ (Karen, S. and Taylor, H. M., 1975, pp. 77). Taking the limit in Identities (1-5) yields the following identities.

$$\boldsymbol{I}+\boldsymbol{P}+\boldsymbol{P}^2+\cdots=\sum_{n=1}^{\infty}\boldsymbol{P}^{n-1}=(\boldsymbol{I}-\boldsymbol{P})^{-1}. \tag{A.8}$$

$$\boldsymbol{I}+2\boldsymbol{P}+3\boldsymbol{P}^2+\cdots=\sum_{n=1}^{\infty}n\boldsymbol{P}^{n-1}=(\boldsymbol{I}-\boldsymbol{P})^{-2}. \tag{A.9}$$

$$(2\cdot 1)\boldsymbol{I}+(3\cdot 2)\boldsymbol{P}+(4\cdot 3)\boldsymbol{P}^2+\cdots=\sum_{n=2}^{\infty}n(n-1)\boldsymbol{P}^{n-2}=2(\boldsymbol{I}-\boldsymbol{P})^{-3}. \tag{A.10}$$

$$(3\cdot 2\cdot 1)\boldsymbol{I}+(4\cdot 3\cdot 2)\boldsymbol{P}+(5\cdot 4\cdot 3)\boldsymbol{P}^2+\cdots=\sum_{n=3}^{\infty}n(n-1)(n-2)\boldsymbol{P}^{n-3}=6(\boldsymbol{I}-\boldsymbol{P})^{-4}. \tag{A.11}$$

$$(4\cdot 3\cdot 2\cdot 1)\boldsymbol{I}+(5\cdot 4\cdot 3\cdot 2)\boldsymbol{P}+(6\cdot 5\cdot 4\cdot 3)\boldsymbol{P}^2+\cdots=\sum_{n=4}^{\infty}n(n-1)(n-2)(n-3)\boldsymbol{P}^{n-4}=24(\boldsymbol{I}-\boldsymbol{P})^{-5}. \tag{A.12}$$

A.2.4 moments (on-target case)

Let N be a random variable and let k be a positive integer. Then the kth *moment* of N is defined as $E(N^k)$, if $E(N^k)$ exists and is finite.

A.2.4.1 The first moment

The first moment of random variable N is defined as follows.

$$E(N) = \sum_{n=1}^{\infty} nf(n) = \sum_{n=1}^{\infty} \boldsymbol{s}'\boldsymbol{P}_0{}^{n-1}(\boldsymbol{I} - \boldsymbol{P}_0)\mathbf{1}$$
$$= \boldsymbol{s}' \left(\sum_{n=1}^{\infty} n\boldsymbol{P}_0{}^{n-1} \right) (\boldsymbol{I} - \boldsymbol{P}_0)\mathbf{1}.$$

By equation (A.9)

$$\sum_{n=1}^{\infty} n\boldsymbol{P}_0{}^{n-1} = \boldsymbol{I} + 2\boldsymbol{P}_0 + 3\boldsymbol{P}_0{}^2 + \cdots = (\boldsymbol{I} - \boldsymbol{P}_0)^{-2}$$
$$E(N) = \boldsymbol{s}'(\boldsymbol{I} - \boldsymbol{P}_0)^{-2}(\boldsymbol{I} - \boldsymbol{P}_0)\mathbf{1} = \boldsymbol{s}'(\boldsymbol{I} - \boldsymbol{P}_0)^{-1}\mathbf{1}.$$

A.2.4.2 The second moment

For the second moment, $E(N^2)$, compute $E[N(N-1)]$.

$$E[N(N-1)] = \sum_{n=1}^{\infty} n(n-1)f(n) = \sum_{n=1}^{\infty} n(n-1)\boldsymbol{s}'\boldsymbol{P}_0{}^{n-1}(\boldsymbol{I} - \boldsymbol{P}_0)\mathbf{1}$$
$$= \boldsymbol{s}' \left(\sum_{n=1}^{\infty} n(n-1)\boldsymbol{P}_0{}^{n-1} \right) (\boldsymbol{I} - \boldsymbol{P}_0)\mathbf{1}.$$

By equation (A.10)

$$\sum_{n=1}^{\infty} n(n-1)\boldsymbol{P}_0{}^{n-1} = \sum_{n=2}^{\infty} n(n-1)\boldsymbol{P}_0{}^{n-1} = \boldsymbol{P}_0 \sum_{n=2}^{\infty} n(n-1)\boldsymbol{P}_0{}^{n-2}$$
$$= \boldsymbol{P}_0 \cdot 2(\boldsymbol{I} - \boldsymbol{P}_0)^{-3} = 2\boldsymbol{P}_0(\boldsymbol{I} - \boldsymbol{P}_0)^{-3}.$$

Thus,

$$E[N(N-1)] = \boldsymbol{s}'(2\boldsymbol{P}_0(\boldsymbol{I} - \boldsymbol{P}_0)^{-3})(\boldsymbol{I} - \boldsymbol{P}_0)\mathbf{1} = 2\boldsymbol{s}'\boldsymbol{P}_0(\boldsymbol{I} - \boldsymbol{P}_0)^{-2}\mathbf{1}.$$

Therefore, the second moment, $E(N^2)$ is

$$E(N^2) = 2\boldsymbol{s}'\boldsymbol{P}_0(\boldsymbol{I} - \boldsymbol{P}_0)^{-2}\boldsymbol{1} + E(N) = 2\boldsymbol{s}'\boldsymbol{P}_0(\boldsymbol{I} - \boldsymbol{P}_0)^{-2}\boldsymbol{1} + \boldsymbol{s}'(\boldsymbol{I} - \boldsymbol{P}_0)^{-1}\boldsymbol{1}.$$

A.2.4.3 The third moment

For the third moment, $E(N^3)$, consider $E[N(N-1)(N-2)]$.

$$\begin{aligned} E[N(N-1)(N-2)] &= \sum_{n=2}^{\infty} n(n-1)(n-2)f(n) = \sum_{n=2}^{\infty} n(n-1)(n-2)\boldsymbol{s}'\boldsymbol{P}_0{}^{n-1}(\boldsymbol{I} - \boldsymbol{P}_0)\boldsymbol{1} \\ &= \boldsymbol{s}'\left(\sum_{n=2}^{\infty} n(n-1)(n-2)\boldsymbol{P}_0{}^{n-1}\right)(\boldsymbol{I} - \boldsymbol{P}_0)\boldsymbol{1}. \end{aligned}$$

By equation (A.11)

$$\begin{aligned} \sum_{n=2}^{\infty} n(n-1)(n-2)\boldsymbol{P}_0{}^{n-1} &= \sum_{n=3}^{\infty} n(n-1)(n-2)\boldsymbol{P}_0{}^{n-1} = \boldsymbol{P}_0{}^2 \sum_{n=3}^{\infty} n(n-1)(n-2)\boldsymbol{P}_0{}^{n-3} \\ &= \boldsymbol{P}_0{}^2 \cdot 6(\boldsymbol{I} - \boldsymbol{P}_0)^{-4} = 6\boldsymbol{P}_0{}^2(\boldsymbol{I} - \boldsymbol{P}_0)^{-4}. \end{aligned}$$

Thus,

$$E[N(N-1)(N-2)] = \boldsymbol{s}'\left(6\boldsymbol{P}_0{}^2(\boldsymbol{I} - \boldsymbol{P}_0)^{-4}\right)(\boldsymbol{I} - \boldsymbol{P}_0)\boldsymbol{1} = 6\boldsymbol{s}'\boldsymbol{P}_0{}^2(\boldsymbol{I} - \boldsymbol{P}_0)^{-3}\boldsymbol{1}.$$

Therefore, the third moment, $E(N^3)$ is

$$\begin{aligned} E(N^3) &= 6\boldsymbol{s}'\boldsymbol{P}_0{}^2(\boldsymbol{I} - \boldsymbol{P}_0)^{-3}\boldsymbol{1} + 3E(N^2) - 2E(N) \\ &= 6\boldsymbol{s}'\boldsymbol{P}_0{}^2(\boldsymbol{I} - \boldsymbol{P}_0)^{-3}\boldsymbol{1} + 3\left(2\boldsymbol{s}'\boldsymbol{P}_0(\boldsymbol{I} - \boldsymbol{P}_0)^{-2}\boldsymbol{1} + \boldsymbol{s}'(\boldsymbol{I} - \boldsymbol{P}_0)^{-1}\boldsymbol{1}\right) - 2\boldsymbol{s}'(\boldsymbol{I} - \boldsymbol{P}_0)^{-1}\boldsymbol{1} \\ &= 6\boldsymbol{s}'\boldsymbol{P}_0{}^2(\boldsymbol{I} - \boldsymbol{P}_0)^{-3}\boldsymbol{1} + 6\boldsymbol{s}'\boldsymbol{P}_0(\boldsymbol{I} - \boldsymbol{P}_0)^{-2}\boldsymbol{1} + \boldsymbol{s}'(\boldsymbol{I} - \boldsymbol{P}_0)^{-1}\boldsymbol{1}. \end{aligned}$$

A.2.4.4 The fourth moment

For the fourth moment, $E(N^4)$, consider $E[N(N-1)(N-2)(N-3)]$.

$$E[N(N-1)(N-2)(N-3)] = \sum_{n=3}^{\infty} n(n-1)(n-2)(n-3)f(n)$$

$$= \sum_{n=3}^{\infty} n(n-1)(n-2)(n-3)\boldsymbol{s}'\boldsymbol{P_0}^{n-1}(\boldsymbol{I}-\boldsymbol{P_0})\mathbf{1}$$

$$= \boldsymbol{s}'\left(\sum_{n=3}^{\infty} n(n-1)(n-2)(n-3)\boldsymbol{P_0}^{n-1}\right)(\boldsymbol{I}-\boldsymbol{P_0})\mathbf{1}.$$

By equation (A.12)

$$\sum_{n=3}^{\infty} n(n-1)(n-2)(n-3)\boldsymbol{P_0}^{n-1} = \sum_{n=4}^{\infty} n(n-1)(n-2)(n-3)\boldsymbol{P_0}^{n-1}$$

$$= \boldsymbol{P_0}^3\sum_{n=4}^{\infty} n(n-1)(n-2)(n-3)\boldsymbol{P_0}^{n-4} = \boldsymbol{P_0}^3\cdot 24(\boldsymbol{I}-\boldsymbol{P_0})^{-5} = 24\boldsymbol{P_0}^3(\boldsymbol{I}-\boldsymbol{P_0})^{-5}. \tag{A.13}$$

Thus,

$$E[N(N-1)(N-2)(N-3)] = \boldsymbol{s}'\left(24\boldsymbol{P_0}^3(\boldsymbol{I}-\boldsymbol{P_0})^{-5}\right)(\boldsymbol{I}-\boldsymbol{P_0})\mathbf{1} = 24\boldsymbol{s}'\boldsymbol{P_0}^3(\boldsymbol{I}-\boldsymbol{P_0})^{-4}\mathbf{1}.$$

Therefore, the fourth moment, $E(N^4)$ is

$$E(N^4) = 24\boldsymbol{s}'\boldsymbol{P_0}^3(\boldsymbol{I}-\boldsymbol{P_0})^{-4}\mathbf{1} + 6E(N^3) - 11E(N^2) + 6E(N)$$

$$= 24\boldsymbol{s}'\boldsymbol{P_0}^3(\boldsymbol{I}-\boldsymbol{P_0})^{-4}\mathbf{1} + 6\left(6\boldsymbol{s}'\boldsymbol{P_0}^2(\boldsymbol{I}-\boldsymbol{P_0})^{-3}\mathbf{1} + 6\boldsymbol{s}'\boldsymbol{P_0}(\boldsymbol{I}-\boldsymbol{P_0})^{-2}\mathbf{1} + \boldsymbol{s}'(\boldsymbol{I}-\boldsymbol{P_0})^{-1}\mathbf{1}\right)$$

$$-11\left(2\boldsymbol{s}'\boldsymbol{P_0}(\boldsymbol{I}-\boldsymbol{P_0})^{-2}\mathbf{1} + \boldsymbol{s}'(\boldsymbol{I}-\boldsymbol{P_0})^{-1}\mathbf{1}\right) + 6\boldsymbol{s}'(\boldsymbol{I}-\boldsymbol{P_0})^{-1}\mathbf{1}$$

$$= 24\boldsymbol{s}'\boldsymbol{P_0}^3(\boldsymbol{I}-\boldsymbol{P_0})^{-4}\mathbf{1} + 36\boldsymbol{s}'\boldsymbol{P_0}^2(\boldsymbol{I}-\boldsymbol{P_0})^{-3}\mathbf{1} + 14\boldsymbol{s}'\boldsymbol{P_0}(\boldsymbol{I}-\boldsymbol{P_0})^{-2}\mathbf{1} + \boldsymbol{s}'(\boldsymbol{I}-\boldsymbol{P_0})^{-1}\mathbf{1}.$$

A.2.5 Moments (off-target case)

A.2.5.1 The first moment

$$
\begin{aligned}
E(N) &= \sum_{n=1}^{\infty} n f(n) = \sum_{n=1}^{\tau-1} n f_1(n) + \sum_{n=\tau}^{\infty} n f_2(n) \\
&= \sum_{n=1}^{\tau-1} n \boldsymbol{s}' \boldsymbol{P_0}^{n-1}(\boldsymbol{I}-\boldsymbol{P_0})\boldsymbol{1} + \sum_{n=\tau}^{\infty} n \boldsymbol{s}' \boldsymbol{P_0}^{\tau-1}\boldsymbol{P_1}^{n-\tau}(\boldsymbol{I}-\boldsymbol{P_1})\boldsymbol{1} \\
&= \boldsymbol{s}'\left\{\left(\sum_{n=1}^{\tau-1} n\boldsymbol{P_0}^{n-1}\right)(\boldsymbol{I}-\boldsymbol{P_0}) + \boldsymbol{P_0}^{\tau-1}\left(\sum_{n=\tau}^{\infty} n\boldsymbol{P_1}^{n-\tau}\right)(\boldsymbol{I}-\boldsymbol{P_1})\right\}\boldsymbol{1}.
\end{aligned}
$$

Using Identity 2 (A.3), we get

$$
\begin{aligned}
& \sum_{n=1}^{\tau-1} n\boldsymbol{P_0}^{n-1} = \boldsymbol{I} + 2\boldsymbol{P_0} + 3\boldsymbol{P_0}^2 + \cdots + (\tau-1)\boldsymbol{P_0}^{\tau-2} \\
= & \left(\boldsymbol{I} + (\tau-1)\boldsymbol{P_0}^{\tau} - \tau\boldsymbol{P_0}^{\tau-1}\right)(\boldsymbol{I}-\boldsymbol{P_0})^{-2}.
\end{aligned}
$$

Using equations (A.8) – (A.9) we also get

$$
\begin{aligned}
& \sum_{n=\tau}^{\infty} n\boldsymbol{P_1}^{n-\tau} = \sum_{n=\tau}^{\infty}(n-\tau+1)\boldsymbol{P_1}^{n-\tau} + \sum_{n=\tau}^{\infty}(\tau-1)\boldsymbol{P_1}^{n-\tau} \\
= & (\boldsymbol{I} + 2\boldsymbol{P_1} + 3\boldsymbol{P_1}^2 + \cdots) + (\tau-1)(\boldsymbol{I} + \boldsymbol{P_1} + \boldsymbol{P_2} + \cdots) \\
= & (\boldsymbol{I}-\boldsymbol{P_1})^{-2} + (\tau-1)(\boldsymbol{I}-\boldsymbol{P_1})^{-1}.
\end{aligned}
$$

Thus,

$$\begin{aligned}
& E(N) = \boldsymbol{s}'\{[\boldsymbol{I} + (\tau-1)\boldsymbol{P}_0{}^{\tau} - \tau\boldsymbol{P}_0{}^{\tau-1}](\boldsymbol{I}-\boldsymbol{P}_0)^{-2}(\boldsymbol{I}-\boldsymbol{P}_0) \\
& +\boldsymbol{P}_0{}^{\tau-1}[(\boldsymbol{I}-\boldsymbol{P}_1)^{-2} + (\tau-1)(\boldsymbol{I}-\boldsymbol{P}_1)^{-1}](\boldsymbol{I}-\boldsymbol{P}_1)\}\mathbf{1} \\
= \; & \boldsymbol{s}'\{[\boldsymbol{I} + (\tau-1)\boldsymbol{P}_0{}^{\tau} - \tau\boldsymbol{P}_0{}^{\tau-1}](\boldsymbol{I}-\boldsymbol{P}_0)^{-1} + \boldsymbol{P}_0{}^{\tau-1}[(\boldsymbol{I}-\boldsymbol{P}_1)^{-1} + (\tau-1)]\}\mathbf{1} \\
= \; & \boldsymbol{s}'\{[\boldsymbol{I} + \tau\boldsymbol{P}_0{}^{\tau} - \boldsymbol{P}_0{}^{\tau} - \tau\boldsymbol{P}_0{}^{\tau-1}](\boldsymbol{I}-\boldsymbol{P}_0)^{-1} + \boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P}_1)^{-1} + (\tau-1)\boldsymbol{P}_0{}^{\tau-1}\}\mathbf{1} \\
= \; & \boldsymbol{s}'\{[\boldsymbol{I} - \boldsymbol{P}_0{}^{\tau} - \tau\boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P}_0)](\boldsymbol{I}-\boldsymbol{P}_0)^{-1} + \boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P}_1)^{-1} + \tau\boldsymbol{P}_0{}^{\tau-1} - \boldsymbol{P}_0{}^{\tau-1}\}\mathbf{1} \\
= \; & \boldsymbol{s}'\{(\boldsymbol{I} - \boldsymbol{P}_0{}^{\tau})(\boldsymbol{I}-\boldsymbol{P}_0)^{-1} - \tau\boldsymbol{P}_0{}^{\tau-1} + \boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P}_1)^{-1} + \tau\boldsymbol{P}_0{}^{\tau-1} - \boldsymbol{P}_0{}^{\tau-1}\}\mathbf{1} \\
= \; & \boldsymbol{s}'\{(\boldsymbol{I} - \boldsymbol{P}_0{}^{\tau})(\boldsymbol{I}-\boldsymbol{P}_0)^{-1} - \boldsymbol{P}_0{}^{\tau-1} + \boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P}_1)^{-1}\}\mathbf{1} \\
= \; & \boldsymbol{s}'\{(\boldsymbol{I} - \boldsymbol{P}_0{}^{\tau})(\boldsymbol{I}-\boldsymbol{P}_0)^{-1} - \boldsymbol{P}_0{}^{\tau-1} + \boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P}_1)^{-1}\}\mathbf{1} \\
= \; & \boldsymbol{s}'(\boldsymbol{I}-\boldsymbol{P}_0)^{-1}\mathbf{1} - \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau}(\boldsymbol{I}-\boldsymbol{P}_0)^{-1}\mathbf{1} - \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\mathbf{1} + \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P}_1)^{-1}\mathbf{1} \\
= \; & \boldsymbol{s}'(\boldsymbol{I}-\boldsymbol{P}_0)^{-1}\mathbf{1} - \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau}(\boldsymbol{I}-\boldsymbol{P}_0)^{-1}\mathbf{1} - \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P}_0)(\boldsymbol{I}-\boldsymbol{P}_0)^{-1}\mathbf{1} + \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P}_1)^{-1}\mathbf{1} \\
= \; & \boldsymbol{s}'[\boldsymbol{I} - \boldsymbol{P}_0{}^{\tau} - \boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P}_0)](\boldsymbol{I}-\boldsymbol{P}_0)^{-1}\mathbf{1} + \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P}_1)^{-1}\mathbf{1} \\
= \; & \boldsymbol{s}'[\boldsymbol{I} - \boldsymbol{P}_0{}^{\tau} - \boldsymbol{P}_0{}^{\tau-1} + \boldsymbol{P}_0{}^{\tau}](\boldsymbol{I}-\boldsymbol{P}_0)^{-1}\mathbf{1} + \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P}_1)^{-1}\mathbf{1} \\
= \; & \boldsymbol{s}'[\boldsymbol{I} - \boldsymbol{P}_0{}^{\tau-1}](\boldsymbol{I}-\boldsymbol{P}_0)^{-1}\mathbf{1} + \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P}_1)^{-1}\mathbf{1}.
\end{aligned}$$

$$\text{Therefore, } E(N) = \boldsymbol{s}'[\boldsymbol{I} - \boldsymbol{P}_0{}^{\tau-1}](\boldsymbol{I}-\boldsymbol{P}_0)^{-1}\mathbf{1} + \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P}_1)^{-1}\mathbf{1}. \qquad \text{(A.14)}$$

A.2.5.2 The second moment

$$\begin{aligned}
& E[N(N-1)] = \sum_{n=1}^{\infty} n(n-1)f(n) = \sum_{n=1}^{\tau-1} n(n-1)f(n) + \sum_{n=\tau}^{\infty} n(n-1)f(n) \\
= \; & \sum_{n=1}^{\tau-1} n(n-1)\boldsymbol{s}'\boldsymbol{P}_0{}^{n-1}(\boldsymbol{I}-\boldsymbol{P}_0)\mathbf{1} + \sum_{n=\tau}^{\infty} n(n-1)\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\boldsymbol{P}_1{}^{n-\tau}(\boldsymbol{I}-\boldsymbol{P}_1)\mathbf{1} \\
= \; & \boldsymbol{s}'\left\{\left(\sum_{n=1}^{\tau-1} n(n-1)\boldsymbol{P}_0{}^{n-1}\right)(\boldsymbol{I}-\boldsymbol{P}_0) + \boldsymbol{P}_0{}^{\tau-1}\left(\sum_{n=\tau}^{\infty} n(n-1)\boldsymbol{P}_1{}^{n-\tau}(\boldsymbol{I}-\boldsymbol{P}_1)\right)\right\}\mathbf{1}.
\end{aligned}$$

Using Identity 3 (A.4), we get

$$
\begin{aligned}
&\sum_{n=1}^{\tau-1} n(n-1)\boldsymbol{P}_0{}^{n-1} = (2\cdot 1)\boldsymbol{P}_0 + (3\cdot 2)\boldsymbol{P}_0{}^2 + (4\cdot 3)\boldsymbol{P}_0{}^3 + \cdots + (\tau-1)(\tau-2)\boldsymbol{P}_0{}^{\tau-2} \\
= \;& \boldsymbol{P}_0[(2\cdot 1)\boldsymbol{I} + (3\cdot 2)\boldsymbol{P}_0 + (4\cdot 3)\boldsymbol{P}_0{}^2 + \cdots + (\tau-1)(\tau-2)\boldsymbol{P}_0{}^{\tau-3}] \\
= \;& \boldsymbol{P}_0[(2\cdot 1)\boldsymbol{I} + (3\cdot 2)\boldsymbol{P}_0 + (4\cdot 3)\boldsymbol{P}_0{}^2 + \cdots + (\tau-1)(\tau-2)\boldsymbol{P}_0{}^{\tau-3}] \\
= \;& \boldsymbol{P}_0[2\boldsymbol{I} - (\tau-1)\tau\boldsymbol{P}_0{}^{\tau-2} + 2(\tau-2)\tau\boldsymbol{P}_0{}^{\tau-1} - (\tau-1)(\tau-2)\boldsymbol{P}_0{}^{\tau}](\boldsymbol{I}-\boldsymbol{P}_0)^{-3}.
\end{aligned}
$$

While using the identities in equations (A.8) - (A.10), we get

$$
\begin{aligned}
&\sum_{n=\tau}^{\infty} n(n-1)\boldsymbol{P}_1{}^{n-\tau} = \sum_{n=\tau}^{\infty}(\underbrace{n-\tau+1}+\tau-1)(\underbrace{n-1-\tau+1}+\tau-1)\boldsymbol{P}_1{}^{n-\tau} \\
= \;& \sum_{n=\tau}^{\infty}[(n-\tau+1)(n-\tau) + (n-\tau+1)(\tau-1) + (\tau-1)(n-\tau) + (\tau-1)^2]\boldsymbol{P}_1{}^{n-\tau} \\
= \;& \sum_{n=\tau}^{\infty}[(n-\tau+1)(n-\tau) + (n-\tau)(\tau-1) + \tau - 1 + (\tau-1)(n-\tau) + \tau^2 - 2\tau + 1]\boldsymbol{P}_1{}^{n-\tau} \\
= \;& \sum_{n=\tau}^{\infty}[(n-\tau+1)(n-\tau) + 2(n-\tau)(\tau-1) + \tau(\tau-1)]\boldsymbol{P}_1{}^{n-\tau} \\
= \;& \sum_{n=\tau}^{\infty}(n-\tau+1)(n-\tau)\boldsymbol{P}_1{}^{n-\tau} + 2(\tau-1)\sum_{n=\tau}^{\infty}(n-\tau)\boldsymbol{P}_1{}^{n-\tau} + \tau(\tau-1)\sum_{n=\tau}^{\infty}\boldsymbol{P}_1{}^{n-\tau} \\
= \;& (1\cdot 0)\boldsymbol{I} + (2\cdot 1)\boldsymbol{P}_1 + (3\cdot 2)\boldsymbol{P}_1{}^2 + (4\cdot 3)\boldsymbol{P}_1{}^3 + (5\cdot 4)\boldsymbol{P}_1{}^4 + \cdots \\
& +2(\tau-1)[0\cdot\boldsymbol{I} + \boldsymbol{P}_1 + 2\boldsymbol{P}_1{}^2 + 3\boldsymbol{P}_1{}^3 + \cdots] + \tau(\tau-1)[\boldsymbol{I} + \boldsymbol{P}_1 + \boldsymbol{P}_1{}^2 + \boldsymbol{P}_1{}^3 \cdots] \\
= \;& \boldsymbol{P}_1[(2\cdot 1) + (3\cdot 2)\boldsymbol{P}_1 + (4\cdot 3)\boldsymbol{P}_1{}^2 + \cdots] + 2(\tau-1)\boldsymbol{P}_1[\boldsymbol{I} + 2\boldsymbol{P}_1 + 3\boldsymbol{P}_1{}^2] \\
& +\tau(\tau-1)[\boldsymbol{I} + \boldsymbol{P}_1 + \boldsymbol{P}_1{}^2 + \cdots] \\
= \;& 2\boldsymbol{P}_1(\boldsymbol{I}-\boldsymbol{P}_1)^{-3} + 2(\tau-1)\boldsymbol{P}_1(\boldsymbol{I}-\boldsymbol{P}_1)^{-2} + \tau(\tau-1)(\boldsymbol{I}-\boldsymbol{P}_1)^{-1}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&E[N(N-1)] = \boldsymbol{s}'\{\boldsymbol{P}_0[2\boldsymbol{I} - (\tau-1)\tau\boldsymbol{P}_0^{\tau-2} + 2(\tau-2)\tau\boldsymbol{P}_0{}^{\tau-1} - (\tau-1)(\tau-2)\boldsymbol{P}_0{}^{\tau}](\boldsymbol{I}-\boldsymbol{P}_0)^{-2} \\
&+\boldsymbol{P}_0{}^{\tau-1}[2\boldsymbol{P}_1(\boldsymbol{I}-\boldsymbol{P}_1)^{-2} + 2(\tau-1)\boldsymbol{P}_1(\boldsymbol{I}-\boldsymbol{P}_1)^{-1} + \tau(\tau-1)\boldsymbol{I}]\}\boldsymbol{1}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&E(N^2) = \boldsymbol{s}'\{\boldsymbol{P}_0[2\boldsymbol{I} - (\tau-1)\tau\boldsymbol{P}_0^{\tau-2} + 2(\tau-2)\tau\boldsymbol{P}_0{}^{\tau-1} - (\tau-1)(\tau-2)\boldsymbol{P}_0{}^{\tau}](\boldsymbol{I}-\boldsymbol{P}_0)^{-2} \\
&+\boldsymbol{P}_0{}^{\tau-1}[2\boldsymbol{P}_1(\boldsymbol{I}-\boldsymbol{P}_1)^{-2} + 2(\tau-1)\boldsymbol{P}_1(\boldsymbol{I}-\boldsymbol{P}_1)^{-1} + \tau(\tau-1)\boldsymbol{I}] + [\boldsymbol{I}-\boldsymbol{P}_0{}^{\tau-1}](\boldsymbol{I}-\boldsymbol{P}_0)^{-1} \\
&+\boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P}_1)^{-1}\}\boldsymbol{1}.
\end{aligned}
$$

A.2.6 Conditional probability mass function and conditional expection

The conditional distribution of N, given that $N < \tau$ is defined as

$$
f(n|N<\tau) = \frac{f(n)}{P(N<\tau)} = \frac{\boldsymbol{s}'\boldsymbol{P}_0{}^{n-1}(\boldsymbol{I}-\boldsymbol{P}_0)\boldsymbol{1}}{1-\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\boldsymbol{1}} \qquad n = 1,2,\ldots,\tau-1.
$$

The conditional expectation of N, given that $N < \tau$ is defined as

$$
\begin{aligned}
E(N|N<\tau) &= \sum_{n=1}^{\tau-1} nf(n|N<\tau) = \sum_{n=1}^{\tau-1} n\frac{\boldsymbol{s}'\boldsymbol{P}_0{}^{n-1}(\boldsymbol{I}-\boldsymbol{P}_0)\boldsymbol{1}}{1-\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\boldsymbol{1}} \\
&= \frac{1}{1-\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\boldsymbol{1}}\sum_{n=1}^{\tau-1} n\boldsymbol{s}'\boldsymbol{P}_0{}^{n-1}(\boldsymbol{I}-\boldsymbol{P}_0)\boldsymbol{1} = \frac{1}{1-\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\boldsymbol{1}}\boldsymbol{s}'\left(\sum_{n=1}^{\tau-1} n\boldsymbol{P}_0{}^{n-1}\right)(\boldsymbol{I}-\boldsymbol{P}_0)\boldsymbol{1} \\
&= \frac{1}{1-\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\boldsymbol{1}}\boldsymbol{s}'\{\boldsymbol{I} + 2\boldsymbol{P}_0 + 3\boldsymbol{P}_0{}^2 + \cdots + (\tau-1)\boldsymbol{P}_0{}^{\tau-2}\}(\boldsymbol{I}-\boldsymbol{P}_0)\boldsymbol{1} \\
&= \frac{1}{1-\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\boldsymbol{1}}\boldsymbol{s}'\left\{\,\boldsymbol{I} + \big((\tau-1)\boldsymbol{P}_0 - \tau\boldsymbol{I}\big)\boldsymbol{P}_0{}^{\tau-1}\right\}(\boldsymbol{I}-\boldsymbol{P}_0)^{-1}\boldsymbol{1}. \qquad \text{(A.15)}
\end{aligned}
$$

Additionally, the conditional distribution of N, given that $N \geq \tau$ is defined as

$$
f(n|N\geq\tau) = \frac{f(n)}{P(N\geq\tau)} = \frac{\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\boldsymbol{P}_1{}^{n-\tau}(\boldsymbol{I}-\boldsymbol{P}_1)\boldsymbol{1}}{\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\boldsymbol{1}} \qquad n = \tau,\tau+1,\ldots
$$

The conditional expectation of N, given that $N \geq \tau$ is defined as

$$
\begin{aligned}
E(N|N \geq \tau) = \sum_{n=\tau}^{\infty} n f(n|N \geq \tau) &= \sum_{n=\tau}^{\infty} n \frac{\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\boldsymbol{P}_1{}^{n-\tau}(\boldsymbol{I}-\boldsymbol{P}_1)\mathbf{1}}{\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\mathbf{1}} \\
&= \frac{1}{\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\mathbf{1}} \sum_{n=\tau}^{\infty} n\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\boldsymbol{P}_1{}^{n-\tau}(\boldsymbol{I}-\boldsymbol{P}_1)\mathbf{1} = \frac{1}{\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\mathbf{1}} \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\left(\sum_{n=\tau}^{\infty} n\boldsymbol{P}_1{}^{n-\tau}\right)(\boldsymbol{I}-\boldsymbol{P}_1)\mathbf{1} \\
&= \frac{1}{\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\mathbf{1}} \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\{\tau\boldsymbol{I} + (\tau+1)\boldsymbol{P}_1 + (\tau+2)\boldsymbol{P}_1{}^2 + \cdots\}(\boldsymbol{I}-\boldsymbol{P}_1)\mathbf{1} \\
&= \frac{1}{\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\mathbf{1}} \boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\{(\boldsymbol{I}-\boldsymbol{P}_1)^{-2} + (\tau-1)(\boldsymbol{I}-\boldsymbol{P})^{-1}\}(\boldsymbol{I}-\boldsymbol{P}_1)\mathbf{1} \\
&= \frac{1}{\boldsymbol{s}'\boldsymbol{P}_0{}^{\tau-1}\mathbf{1}} \boldsymbol{s}'\{\boldsymbol{P}_0{}^{\tau-1}(\boldsymbol{I}-\boldsymbol{P}_1)^{-1} + (\tau-1)\boldsymbol{P}_0{}^{\tau-1}\}\mathbf{1}. \qquad \text{(A.16)}
\end{aligned}
$$

Appendix B

Programming (R- code)

B.1 Markov chain algorithm for the calculation of ARL

```
ARL_Calculator<-function(r,h,d,p){

# r is a smoothing parameter

# h is a control limit

# d is a mean shift

# p is the number of variables

# S is a starting vector


UCL<-sqrt(h*r/(2-r))

m1<-25

m2<-25

g1<-2*UCL/(2*m1+1)

g2<-2*UCL/(2*m2+1)

H<-matrix(data=NA,nrow=2*m1+1,ncol=2*m1+1)


#[Defining an identity matrix and a vector of 1s]
```

```
z<-(2*m1+1)*(m2+1)

n1<-c(z)

I <- matrix(0,nrow=n1,ncol=n1)

I[row(I)==col(I)]<-1

one<- matrix(1,nrow=z,ncol=1)


#[Transition Matrix of Wt1]

range1<-2*m1+1 # range1 is the number of states of Wt1

for (i in 1:range1){

c_i<- -UCL+(i-0.5)*g1

for (j in 1:range1){

up<-(-UCL+j*g1-(1-r)*c_i)/r-delta

down<-(-UCL+(j-1)*g1-(1-r)*c_i)/r-delta

H[i,j]<-pnorm(up,mean=0,sd=1)-pnorm(down,mean=0,sd=1)

}

}

#[Transition Matrix of Wt2]

range2<-m2+1 # range2 is the number of states of Wt2

V<-matrix(data=NA,nrow=range2,ncol=range2)

for (i in 0:m2){

c<-((1-r)*i*g2/r)^2

for (j in 0:m2){

if (j==0) {

V[i+1,1]<-pchisq((0.5*(g2)/r)^2, df=p-1,ncp=c)

}

else {

up<-((j+0.5)*g2/r)^2
```

```
down<-((j-0.5)*g2/r)^2

V[i+1,j+1]<-pchisq(up,df=p-1,ncp=c)-pchisq(down,df=p-1,ncp=c)

}

}

}

E<- kronecker(H,V)  # Operating kronecker product


#[Finding transient states]

counter<-1

for (alpha in 1:range1){

for (beta in 0:m2){

if ((alpha-(m1+1))^2*g1^2+(beta*g2)^2 >= UCL^2){

E[,counter]<-0

E[counter,]<-0

}

counter<-counter+1

}

}


temp<-solve(I-E)

temp%*%one

S<- matrix(0,nrow=z,ncol=1)

start<-m1*(m2+1)+1

S[start,1]<-1

ARL<-t(S)%*% temp%*%one

ARL

}
```

B.2 On-target run length distribution

```
#On-target run length distribution

OT_Dist<-function(x){

RL<-x

Sfunc <-NULL

Pfunc<-NULL

Tmatrix<-I

range<-RL+1

for (i in 1:range){

Sfunc[i]<-t(S)%*% Tmatrix %*%one

Tmatrix <- Tmatrix %*%E # E is an on-target transition matrix

}

for (i in 1:n){

Pfunc[i]<-Sfunc[i]- Sfunc[i+1]  #f(N) = P(N > n-1) - P(N > n)

}

plot(spline(seq(1,RL,by=1),Pfunc,n=200),type="l",col="blue",ylab="probability"

,xlab="run length")

}
```

B.3 Off-target run length distribution

```
#Off-target run length distribution

Off_Dist<-function(x,t){

RL<-x

tau<-t # tau is a mean shift time

range1<-tau-1

Sfunc<-NULL
```

```
Pfunc <-NULL
Tmatrix0<-I
Tmatrix1<-Q # Q is a off-target transition matrix
Tmatrix2<-I
range2<-RL+1
for (i in 1:range1){
Tmatrix2<- Tmatrix2%*%E
}
for (i in 1:range2){
if (i <=tau){
Sfunc[i]<-t(S)%*% Tmatrix0%*%one
Tmatrix0<- Tmatrix0%*%E
}
else
{ Sfunc[i]<-t(S)%*% Tmatrix2%*%Tmatrix1%*%one
Tmatrix1<- Tmatrix1%*%Q}
}
for (i in 1:RL){
Pfunc[i]<- Sfunc[i]- Sfunc[i+1] #f(N) = P(N > n-1) - P(N > n)
}
plot(spline(seq(1,RL,by=1),Pfunc,n=200),type="l",col="blue",ylab="probability",
xlab="run length")
}
```

B.4 The partition method

```
# The partition method

# The following code is constructed

# based on the condition that in-control ARL = 200, p = 4

delta<-1.5

p<-4

h_max<-30 # Initial upper control limit value

h_min<-0.01 # Initial lower control limit value

h_tem<-0

ARL_tem<-0

ARL_0<-200 # In-control ARL value

epsilon<-0.01 # Degree of precision

k<-1

R_opt<-0 # Initialization of smoothing parameter

ARL<-NULL

H<-NULL

ARL_opt<-ARL_0

ARL1_opt<-0

for (i in 1:100){

r<-i/100

h_tem<-(h_max + h_min)/2

ARL_tem<-ARL_Calculator(r,h_tem,0,p)


while(abs(ARL_0-ARL_tem)> epsilon)

{

if (ARL_tem > ARL_0)
```

```
{
h_max<-h_tem
h_tem<-(h_max + h_min)/2
ARL_tem<- ARL_Calculator(r,h_tem,0,p)
}
else
{
h_min<-h_tem
h_tem<-(h_max + h_min)/2
ARL_tem<- ARL_Calculator(r,h_tem,0,p)
}
}
ARL[k]<-ARL_Calculator(r,h_tem,delta,p)
if(ARL_opt > ARL[k])  #If-statement to find the optimal smoothing parameter
{
ARL_opt<-ARL[k]
R_opt<-r
}
H[k]<-h_tem
k=k+1
h_max<-30
h_min<-0.01
h_tem<-0
}
R_opt
ARL_opt
```

Appendix C

Tables

Table C.1: Optimal MEWMA Control Charts for $\tau\,(1 \leq \tau \leq 200)$, $p = 2$.

$p = 2, \delta = 1, ARL_0 = 200.00$											
$\tau =$	1	20	40	60	80	100	120	140	160	180	200
r	0.14	0.16	0.18	0.19	0.20	0.21	0.22	0.23	0.24	0.25	0.26
H	9.160	9.347	9.502	9.570	9.632	9.690	9.744	9.793	9.839	9.882	9.922
ARL_{min}	9.986	27.477	44.307	59.464	73.126	85.447	96.560	106.586	115.635	123.801	131.173

$p = 2, \delta = 1, ARL_0 = 300.00$											
$\tau =$	1	20	40	60	80	100	120	140	160	180	200
r	0.14	0.15	0.15	0.16	0.17	0.18	0.19	0.19	0.20	0.21	0.21
H	10.091	10.184	10.184	10.267	10.343	10.413	10.476	10.476	10.535	10.588	10.588
ARL_{min}	10.953	28.840	46.675	63.324	78.863	93.371	106.917	119.568	131.384	142.420	152.730

$p = 2, \delta = 1, ARL_0 = 400.00$											
$\tau =$	1	20	40	60	80	100	120	140	160	180	200
r	0.13	0.13	0.14	0.15	0.15	0.16	0.17	0.17	0.18	0.18	0.19
H	10.647	10.647	10.745	10.834	10.834	10.914	10.986	10.986	11.052	11.052	11.113
ARL_{min}	11.639	29.715	48.086	65.532	82.098	97.834	112.786	126.987	140.484	153.307	165.494

$p = 2, \delta = 1, ARL_0 = 500.00$											
$\tau =$	1	20	40	60	80	100	120	140	160	180	200
r	0.12	0.12	0.13	0.14	0.14	0.15	0.15	0.16	0.16	0.17	0.17
H	11.047	11.047	11.153	11.249	11.249	11.334	11.334	11.412	11.412	11.482	11.482
ARL_{min}	12.173	30.371	49.061	67.001	84.212	100.734	116.591	131.814	146.425	160.457	173.922

Table C.2: Optimal MEWMA Control Charts for $\tau\,(1 \leq \tau \leq 200)$, $p = 4$.

$p = 4, \delta = 1, ARL_0 = 200.00$											
$\tau =$	1	20	40	60	80	100	120	140	160	180	200
r	0.13	0.14	0.15	0.16	0.18	0.19	0.19	0.20	0.21	0.22	0.23
H	13.203	13.324	13.433	13.532	13.703	13.778	13.778	13.846	13.909	13.967	14.021
ARL_{min}	12.061	29.304	46.078	61.167	74.751	86.985	98.010	107.949	116.910	124.995	132.290

$p = 4, \delta = 1, ARL_0 = 300.00$											
$\tau =$	1	20	40	60	80	100	120	140	160	180	200
r	0.12	0.12	0.13	0.14	0.15	0.15	0.16	0.17	0.17	0.18	0.18
H	14.156	14.156	14.282	14.395	14.495	14.495	14.586	14.669	14.669	14.744	14.744
ARL_{min}	13.197	30.834	48.652	65.263	80.758	95.213	108.707	121.292	133.046	144.019	154.267

$p = 4, \delta = 1, ARL_0 = 400.00$											
$\tau =$	1	20	40	60	80	100	120	140	160	180	200
r	0.11	0.11	0.12	0.13	0.13	0.14	0.14	0.15	0.15	0.16	0.16
H	14.777	14.777	14.913	15.032	15.032	15.139	15.139	15.234	15.234	15.320	15.320
ARL_{min}	14.002	31.837	50.193	67.619	84.153	99.856	114.776	128.928	142.379	155.157	167.293

$p = 4, \delta = 1, ARL_0 = 500.00$											
$\tau =$	1	20	40	60	80	100	120	140	160	180	200
r	0.11	0.11	0.11	0.12	0.12	0.13	0.13	0.13	0.14	0.14	0.15
H	15.361	15.361	15.361	15.491	15.491	15.606	15.606	15.606	15.708	15.708	15.800
ARL_{min}	14.637	32.585	51.262	69.188	86.379	102.879	118.706	133.902	148.477	162.471	175.906

Table C.3: Optimal MEWMA Control Charts for $\tau\,(1 \leq \tau \leq 200)$, $p = 6$.

$p = 6, \delta = 1, ARL_0 = 200.00$											
$\tau =$	1	20	40	60	80	100	120	140	160	180	200
r	0.12	0.12	0.14	0.15	0.16	0.17	0.18	0.19	0.19	0.20	0.21
H	16.635	16.635	16.914	17.032	17.138	17.234	17.321	17.401	17.401	17.474	17.541
ARL_{min}	13.581	30.617	47.356	62.394	75.919	88.093	99.058	108.935	117.834	125.857	133.092

$p = 6, \delta = 1, ARL_0 = 300.00$											
$\tau =$	1	20	40	60	80	100	120	140	160	180	200
r	0.11	0.11	0.12	0.13	0.13	0.14	0.15	0.15	0.16	0.16	0.17
H	17.683	17.683	17.836	17.970	17.970	18.091	18.198	18.198	18.295	18.295	18.382
ARL_{min}	14.855	32.284	50.085	66.674	82.134	96.548	109.999	122.545	134.255	145.184	155.389

$p = 6, \delta = 1, ARL_0 = 400.00$											
$\tau =$	1	20	40	60	80	100	120	140	160	180	200
r	0.10	0.10	0.11	0.11	0.12	0.12	0.13	0.13	0.14	0.14	0.15
H	18.356	18.356	18.520	18.520	18.664	18.664	18.791	18.791	18.904	18.904	19.005
ARL_{min}	15.755	33.380	51.726	69.136	85.650	101.333	116.211	130.346	143.762	156.507	168.614

$p = 6, \delta = 1, ARL_0 = 500.00$											
$\tau =$	1	20	40	60	80	100	120	140	160	180	200
r	0.10	0.10	0.10	0.11	0.11	0.11	0.12	0.12	0.13	0.13	0.13
H	19.001	19.001	19.001	19.158	19.158	19.158	19.296	19.296	19.417	19.417	19.417
ARL_{min}	16.456	34.200	52.865	70.787	87.961	104.448	120.254	135.424	149.984	163.948	177.355

Table C.4: Optimal MEWMA Control Charts for $\tau \, (1 \leq \tau \leq 200)$, $p = 10$.

$p = 10, \delta = 1, ARL_0 = 200.00$										
$\tau =$	1	20	40	60	80	100	120	140	160	180	200
r	0.11	0.11	0.12	0.13	0.14	0.15	0.16	0.16	0.17	0.18	0.18
H	22.881	22.881	23.069	23.234	23.381	23.512	23.630	23.630	23.737	23.833	23.833
ARL_{min}	15.949	32.616	49.300	64.266	77.704	89.784	100.655	110.437	119.240	127.170	134.314

$p = 10, \delta = 1, ARL_0 = 300.00$										
$\tau =$	1	20	40	60	80	100	120	140	160	180	200
r	0.10	0.10	0.10	0.11	0.12	0.12	0.13	0.13	0.14	0.14	0.15
H	24.084	24.084	24.084	24.278	24.447	24.447	24.596	24.596	24.728	24.728	24.846
ARL_{min}	17.457	34.524	52.290	68.834	84.247	98.609	111.992	124.478	136.116	146.978	157.109

$p = 10, \delta = 1, ARL_0 = 400.00$										
$\tau =$	1	20	40	60	80	100	120	140	160	180	200
r	0.09	0.09	0.09	0.10	0.10	0.11	0.11	0.12	0.12	0.13	0.13
H	24.844	24.844	24.844	25.054	25.054	25.234	25.234	25.392	25.392	25.530	25.530
ARL_{min}	18.515	35.771	54.097	71.477	87.974	103.606	118.452	132.535	145.904	158.605	170.649

$p = 10, \delta = 1, ARL_0 = 500.00$										
$\tau =$	1	20	40	60	80	100	120	140	160	180	200
r	0.09	0.09	0.09	0.09	0.10	0.10	0.10	0.11	0.11	0.11	0.12
H	25.589	25.589	25.589	25.589	25.788	25.788	25.788	25.959	25.959	25.959	26.108
ARL_{min}	19.337	36.723	55.354	73.257	90.423	106.873	122.662	137.795	152.311	166.246	179.613

Table C.5: Optimal MEWMA Control Chart, $p = 2$, $\tau = 1$.

Condition: $p = 2, m_1 = m_2 = 25, \tau = 1$										
δ	ARL_0	200	300	400	500	600	700	800	900	1000
0.5	r	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04
	H	7.3813	8.3930	9.1058	9.2427	9.6975	10.0801	10.4102	10.7003	10.9589
	ARL_{min}	26.7482	30.3318	33.0044	35.0648	36.7685	38.2371	39.5312	40.6910	41.7433
1	r	0.14	0.13	0.12	0.12	0.12	0.11	0.11	0.11	0.11
	H	9.1602	9.9887	10.5367	11.0468	11.4612	11.6952	11.9983	12.2647	12.5023
	ARL_{min}	9.9857	10.9458	11.6374	12.1734	12.6198	12.9915	13.3169	13.6073	13.8697
1.5	r	0.25	0.24	0.22	0.22	0.21	0.20	0.20	0.20	0.19
	H	9.8824	10.7268	11.2672	11.7522	12.1029	12.3901	12.6793	12.9338	13.1133
	ARL_{min}	5.4381	5.8736	6.1842	6.4261	6.6237	6.7923	6.9370	7.0660	7.1814
2	r	0.38	0.35	0.34	0.32	0.32	0.31	0.30	0.30	0.30
	H	10.2466	11.0450	11.6313	12.0602	12.4436	12.7466	13.0056	13.2528	13.4736
	ARL_{min}	3.5265	3.7737	3.9496	4.0863	4.1981	4.2927	4.3751	4.4475	4.5128
2.5	r	0.53	0.49	0.47	0.45	0.44	0.43	0.42	0.41	0.41
	H	10.4333	11.2332	11.8071	12.2477	12.6132	12.9206	13.1854	13.4176	13.6341
	ARL_{min}	2.5152	2.6842	2.8021	2.8923	2.9656	3.0271	3.0802	3.1269	3.1688
3	r	0.68	0.65	0.63	0.61	0.59	0.58	0.57	0.56	0.5
	H	10.5147	11.3238	11.8984	12.3426	12.7042	13.0125	13.2789	13.5135	13.7230
	ARL_{min}	1.8812	2.0064	2.0962	2.1656	2.2221	2.2696	2.3104	2.3461	2.3779

Table C.6: Optimal MEWMA Control Chart, $p = 4$, $\tau = 1$.

Condition: $p = 4, m_1 = m_2 = 25, \tau = 1$										
δ	ARL_0	200	300	400	500	600	700	800	900	1000
0.5	r	0.05	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03
	H	11.2683	11.9742	12.8373	13.4946	14.0240	14.4667	14.2589	14.6039	14.9105
	ARL_{min}	32.4380	36.9214	40.0895	42.6254	44.7534	46.5966	48.1545	49.4744	50.6670
1	r	0.13	0.12	0.11	0.11	0.10	0.10	0.10	0.09	0.09
	H	13.2030	14.1561	14.7776	15.3613	15.6894	16.0888	16.4322	16.5786	16.8494
	ARL_{min}	12.0608	13.1970	14.0026	14.6367	15.1436	15.5777	15.9584	16.2920	16.5843
1.5	r	0.22	0.21	0.20	0.19	0.19	0.18	0.18	0.17	0.17
	H	13.9672	14.9324	15.5921	16.0856	16.5332	16.8515	17.1768	17.4021	17.6579
	ARL_{min}	6.5072	7.0104	7.3673	7.6436	7.8705	8.0599	8.2255	8.3715	8.5006
2	r	0.33	0.31	0.30	0.29	0.28	0.28	0.27	0.27	0.26
	H	14.3900	15.3181	15.9808	16.4855	16.8908	17.2546	17.5420	17.8186	18.0381
	ARL_{min}	4.1790	4.4617	4.6620	4.8172	4.9439	5.0512	5.1434	5.2252	5.2980
2.5	r	0.45	0.42	0.41	0.39	0.38	0.38	0.37	0.37	0.36
	H	14.6098	15.5253	16.1841	16.6766	17.0855	17.4413	17.7356	18.0061	18.2347
	ARL_{min}	2.9739	3.1563	3.2842	3.3827	3.4632	3.5311	3.5898	3.6419	3.6881
3	r	0.61	0.57	0.55	0.52	0.51	0.50	0.49	0.48	0.47
	H	14.7408	15.6532	16.3008	16.7933	17.2023	17.5464	17.8431	18.1036	18.3356
	ARL_{min}	2.2343	2.3749	2.4720	2.5453	2.6041	2.6532	2.6953	2.7319	2.7645

Table C.7: Optimal MEWMA Control Chart, $p = 10$, $\tau = 1$.

Condition: $p = 10, m_1 = m_2 = 25, \tau = 1$										
δ	ARL_0	200	300	400	500	600	700	800	900	1000
0.5	r	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.03
	H	20.0880	21.7696	22.9044	22.9779	23.6952	24.2893	24.7957	25.2364	25.6260
	ARL_{min}	42.6234	48.8044	53.4589	56.8531	59.6636	62.0885	64.2288	66.1493	67.8949
1	r	0.11	0.10	0.09	0.09	0.08	0.08	0.08	0.08	0.08
	H	22.8809	24.0840	24.8443	25.5895	25.9643	26.4705	26.9039	27.2824	27.6182
	ARL_{min}	15.9489	17.4573	18.5150	19.3374	20.0059	20.5562	21.0384	21.4680	21.8561
1.5	r	0.19	0.17	0.16	0.16	0.15	0.15	0.15	0.14	0.14
	H	23.9212	25.0468	25.8607	26.5520	27.0199	27.4886	27.8909	28.1530	28.4681
	ARL_{min}	8.5373	9.1846	9.6393	9.9870	10.2713	10.5084	10.7159	10.8970	11.0565
2	r	0.28	0.26	0.25	0.24	0.23	0.23	0.23	0.22	0.22
	H	24.4532	25.6059	26.4251	27.0403	27.5348	27.9834	28.3690	28.6664	28.9681
	ARL_{min}	5.4302	5.7870	6.0369	6.2291	6.3858	6.5164	6.6306	6.7293	6.8180
2.5	r	0.38	0.35	0.34	0.33	0.32	0.32	0.31	0.31	0.30
	H	24.7488	25.8805	26.6942	27.3136	27.8117	28.2476	28.6026	28.9324	29.2058
	ARL_{min}	3.8214	4.0470	4.2054	4.3275	4.4270	4.5103	4.5823	4.6455	4.7021
3	r	0.48	0.45	0.43	0.42	0.41	0.41	0.40	0.40	0.39
	H	24.9031	26.0392	26.8349	27.4522	27.9508	28.3788	28.7374	29.0510	29.3401
	ARL_{min}	2.8821	3.0363	3.1440	3.2268	3.2944	3.3513	3.4006	3.441	3.4829

Bibliography

[1] Agrawal, R., Lawless, J.F., and Mackay, R, J. (1999). Analysis of variation transmission in manufacturing processes - Part II. *Journal of Quality Technology* **31**(2), 143–154.

[2] Aparisi, F., Champ, C. W., and Garcia-Diaz, J. C. (2004). A performance analysis of Hotelling's χ^2 control chart with supplementary runs rules. *Quality Engineering* **16**(3), 359–368.

[3] Bakshi, U. A. and Godse, A. P. (2008). *Electronics Engineering,* India: Technical publications.

[4] Brook, D. and Evans, D. A. (1972). An approach to the probability distribution of CUSUM run lengths. *Biometrika* **59**, 539–549.

[5] Crosier, R. B. (1988). Multivariate generalizations of cumulative sum quality control schemes. *Technometrics* **30**, 291–303.

[6] Eaton, M. L. (1983). *Multivariate Statistics,* New York: Wiley.

[7] Greer, J., Korkin, A., and Labanowski, J. K. (2003). *Nano Giga Challenges in Microelectronics,* New York: Elsevier.

[8] Hawkins, D.M. (1992). Evaluation of average run length of cumulative sum charts for an arbitrary data distribution. *Communications in Statistics, Part B - Simulation and Computation* **21**, 1004–1020.

[9] Joung, Y., Mun, S., and Kang, S. (2004). Improvement of gate CD imbalance for a 0.35 μm logic technology. *Materials Science in Semiconductor Processing* **7**, 51–54.

[10] Karlin, S. and Taylor, H. M. (1975). *A First Course In Stochastic Process,* 2nd edition. New York, New York: Academic Press, Inc.

[11] Kurihara, M., Izawa, M., Tanaka, J., Kawai, K., and Fujiwara, N. (2007). Gate CD control considering variation of gate and STI structure. *IEEE Transactions on Semiconductor Manufacturing* **20**(3).

[12] Lawless, J.F., Mackay, R, J., and Robinson, J. A. (1999). Analysis of variation transmission in manufacturing processes - Part I. *Journal of Quality Technology* **31**(2), 131–142.

[13] Lee, C. (2009). Method for manufacturing semiconductor device using KRF light source. *US Patent* 7,541,286 B2.

[14] Lee, M. H. (2009). Multivariate EWMA charts with variable sampling intervals. *Economic Quality Control* **24**(2), 231–241.

[15] Lowry, C. A., Woodall, W. H., Champ, C. W., and Ridgon, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics* **34**(1), 46–53.

[16] Lucas, J. M. and Saccucci, M. S. (1990). Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics* **32**, 1–12.

[17] Montgomery, D. C. (2005). *Statistical Quality Control,* 5th edition. Hoboken, New Jersey: Wiley.

[18] Orshansky, M., Milor, L., Nguyen, L., Hill, G., Peng, Y., and Hu, C. (1999). Intra-field gate CD variability and its impact on circuit performance. *IEDM Tech Dig* 479.

[19] Pham, H. (2006). *Handbook of Engineering Statistics,* New York: Springer.

[20] Prabhu, S. S. and Runger, G. C. (1996). A Markov chain model for the multivariate exponentially weighted moving averages control chart. *Journal of American Statistical Association* **91,** 1701–1706.

[21] Prabhu, S. S. and Runger, G. C. (1997). Designing a multivariate EWMA control chart. *Journal of Quality Technology* **29**(1), 8–15.

[22] Pignatiello, J. J. and Runger, G. C. (1995). Comparisons of multivariate CUSUM charts. *Journal of Quality Technoligy* **22,** 173–186.

[23] Quirk, M. and Serda, J. (2000). *Semiconductor Manufacturing Technology,* 1st edition. Upper Saddle River, New Jersey: Prentice Hall.

[24] Rigdon, S. E. (1995a). A double integral equation for the average run length of a multivariate exponentially weighted moving average control chart. *Statistics and Probability Letters* **24,** 365–373.

[25] Rigdon, S. E. (1995b). An integral equation for the in-control average run length of a multivariate exponentially weighted moving average control chart. *Journal of Statistical Computation and Simulation* **52,** 351–365.

[26] Woodall, W. H. and Ncube, M. M. (1985). Multivariate CUSUM quality control procedures. *Technometrics* **27,** 285–292.