CHANGE POINT DETECTION IN UNIVARIATE AND

MULTIVARIATE PROCESSES

by

JINHO KIM

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Industrial and Systems Engineering

Written under the direction of

Professor Elsayed A. Elsayed and

Professor Myong K. Jeong

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

Octorber, 2014

ABSTRACT OF THE DISSERTATION

Change Point Detection in Univariate and Multivariate Processes

By JINHO KIM

Dissertation Directors:

Dr. Elsayed A. Elsayed and Dr. Myong K. Jeong

In this dissertation, we present several methodologies for detecting the mean changes in univariate and multivariate processes, identifying fault variables in multivariate processes, and detecting the mean changes in multistage processes. We first propose an adaptive runs rule, which is motivated by the concept of supplementary runs rule, in order to make control charts more sensitive to small mean shifts. The adaptive runs rule assigns scores to consecutive runs based on the estimated shift size of the mean. We supplement the adaptive CUSUM (ACUSUM) chart with the adaptive runs rule to enhance its sensitivity in detecting small mean shifts.

We propose two new SPC procedures, MASC and AMASC, for detecting mean shift vectors based on the approximate sequential χ^2 test. Similar to the univariate CUSUM chart, a multivariate CUSUM chart can be designed to detect a specific size of the mean shift optimally based on a sequential likelihood ratio test for noncentrality.

However, in multivariate case, the probability ratio of a sequential test is intractable mathematically and the test statistic based on the ratio does not have a closed form expression which makes it impractical for real application. We drive an approximate log-likelihood ratio and propose a multivariate SPC chart based on the sequential χ^2 test.

We propose an adaptive step-down procedure using conditional T^2 statistics for the identification of fault variables. In a process with massive process, identifying which variable or a subset of variables causes an out-of-control signal is a challenging issue for quality engineers. The proposed adaptive step-down procedure selects a variable having no significant evidence of a change at each step based on the variables that are selected in previous steps.

Finally, we represent an autocorrelated multistage process as VAR(1) model and derive the propagation models of mean shifts to subsequent stages under the state space model. Further, we propose a new conditional CMEWMA chart to detect the shift of mean in a multistage process by incorporating unchanged stage information. The simulation results show that the proposed CMEWMA chart is efficient in detecting a wide range of small mean shifts compared with other MEWMA charts.

Acknowledgement

First and foremost, I would like to extend heartfelt appreciation to my advisors, Professor Elsayed A. Elsayed and Professor Myong K. Jeong, without whose support and guidance I would not have been where I am now. Professors Elsayed and Jeong have the determination and insight of true scholars and they are the most patient and supportive mentors who have kindly concerned and considered my academic improvements. Especially, I am truly grateful to Professor Myong K. Jeong, an exceptional researcher and teacher. Since the beginning of my studies at Rutgers University, he has encouraged me and helped me with his precision and unfailing support. I should also thank my dissertation committee members, Professor Hoang Pham and Professor Ming-gi Xie for their expertise, support, intellectual insights and efforts.

I would not have been able to finish this dissertation and my Ph. D degree without the support from my wife, Jina Hong, who has been there for me for the last nine years. I also thank all the members of my family who have been there for me and have given all of the love, support, encouragement, and dedication. I acknowledge all my friends in the Rutgers University for being great colleagues and friends. I would especially like to thank the members from the Praise Presbyterian church – Junghee, Changsub, Jeongsub, Suyeon - who have prayed and supported for the completion of my study and they have been everything I needed to go through tough times during the program.

Last but not the least, my family and the one above all of us, I thank my God for answering my prayers and giving me the strength to endure despite on the moment of wanting to give up the study. Thank you so much, my Dear Lord.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

1.1 Overview

The recent advances in instrumentation, data communication and sensors have resulted in significant improvements in product quality. The proper quality improvement strategy has emerged as a critical factor in the successful design, construction and operation and controlling the quality in a wide range of industries. A well-engineered and correctly specified modern control system will minimize the process variability, improve the process efficiency, avoid unexpected failure rates, keep operating and maintenance costs as low as possible.

Quality improvement can be defined as the reduction of the proportion to variability (Montgomery 2005). Process variability normally consists of both common causes and special (assignable) causes. The common causes are inherent variability naturally embedded in the process so that they cannot be removed from the process. On the other hand, the special causes occur at random times due to some assignable causes during the process. The assignable causes usually change process characteristics from the target and can be removed from the process when detected.

The action of monitoring the process with the assignable causes is performed by process charts such as Shewhart, CUSUM, and EWMA to detect whether the process level is changed from the target. These actions of process monitoring are called the

Statistical Process Control (SPC). SPC is often part of an organization's strategic thrust to improve quality and it is the activity to monitor processes with emphasis on methods and procedures. The main objective of SPC is to detect process changes quickly and remove causes of process disturbances. Control charts are widely used to monitor process characteristics to detect process changes. Control charts plot values of test statistics against time or the sample number of process outputs collected constantly and randomly.

Shewhart (1931) introduces the concept of a SPC chart to monitor the mean of a process with single parameter (univariate process). Since the development of the Shewhart chart, numerous control charts for univariate processes have been proposed (Page 1954, Roberts 1959, Lucas and Saccucci 1990, Sparks 2000, Shu and Jiang 2006, Jiang et al. 2008, Wu et al. 2009). Further research has been conducted to investigate multivariate SPC (MSPC) for multiple quality characteristics such as Hotelling's T^2, MCUSUM and MEWMA (Healy 1987, Crosier 1988, Pignatiello and Runger 1990, Lowry et al. 1992, Sullivan and Jones 2002, Hawkins and Maboudou-Tchao 2008, Jiang and Tsui 2008). In this dissertation, we propose efficient and effective approaches to detect changes in the process parameters as quickly as they occur for both univariate and multivariate processes.

In high-dimensional processes, which have massive process variables, identifying which variable or a subset of variables causes an out-of-control signal is a challenging issue for quality engineers. The proposed MSPC chart focuses on detecting mean shifts in multiple process parameters based on T^2 statistics. However, it has difficulty in identifying variables which cause the out-of-control signal when the mean shift is

detected. We propose an adaptive step-down procedure that can identify the source of changes efficiently.

Most of the research on univariate/multivariate quality characteristics has been limited to single-stage production processes. Several investigators have proposed SPC procedures for monitoring and controlling processes with autocorrelated data for a single stage process (Schmid 1995, Lu and Reynolds 2001, Kramer and Schmid 1997, Rosotowski and Schmid 2006). In summary, SPC processes that deal with autocorrelated observations are limited to a single-stage process. However, as manufacturing industries become more sophisticated, it is common to find a production process involving multiple stages such as those found in pharmaceutical manufacturing, chemical industry and semiconductor manufacturing. In order to develop advanced SPC methodologies for an autocorrelated multistage process, there are a number of challenges that must be addressed. First, due to the complexity of multistage processes and autocorrelations of observations, the relationship between the output variables and input variables is complicated. Multistage processes have a unique cascading property, i.e., outputs from operations at upstream stages may affect the quality of downstream stages, and product variation may propagate throughout the production stages (Hawkins 1993, Li and Tsung 2011). In case of an autocorrelated multistage process, this cascading property may produce more complex consequences because the variation propagation could be more time-dependent than usual discrete multistage processes. Consequently, identifying the faulty stages and the change time will be difficult. Our preliminary study in chapter 5 shows that conventional multistage control charts do not work well when the quality

characteristic exhibits even low levels of correlation over time and gives misleading results in the form of excessive false alarms when the data are positively correlated.

1.2 Dissertation outline

This dissertation is organized as follows. Chapter 2 presents an adaptive runs rule, which is motivated by the concept of supplementary runs rule, in order to develop control charts more sensitive to small mean shifts than existing approaches. Chapter 3 proposes a multivariate SPC chart based on a sequential test having an optimal property for testing shift vectors with a specific noncentrality parameter. Chapter 4 presents an adaptive step-down procedure using conditional T^2 statistics for fault variable identification. Chapter 5 proposes an SPC procedure for monitoring autocorrelated multistage processes. Finally, chapter 6 summarizes the research results and future work.

CHAPTER 2

UNIVARIATE SPC WITH ADAPTIVE RUNS RULE

2.1 Introduction

In this chapter, we assume that a process has a single quality characteristic X whose measurements follow a normal distribution $X \sim N(\mu, \sigma^2)$, and a sequence of measurements $\{X_t\}$ is independently and identically distributed (i.i.d.). When the process is in control, its mean and standard deviation are $\mu = \mu_0$ and $\sigma = \sigma_0$ respectively, otherwise the mean of the process is shifted to μ_1, where $\mu_1 \neq \mu_0$. Let $\delta = |\mu_1 - \mu_0| / \sigma_0$ represent the magnitude of the unknown shift. For simplicity, we assume that $\mu_0 = 0$ and $\sigma_0 = 1$. The cumulative sum (CUSUM) chart proposed by Page (1954) is widely used to detect small mean shifts efficiently. The conventional upper and lower-sided mean CUSUM statistics (Page 1954) can be written as

$$C_t^+ = \max[0, X_t - k + C_{t-1}^+] \text{ and } C_t^- = \min[0, X_t + k + C_{t-1}^-], \qquad (2.1)$$

where $k > 0$ and $C_0^+ = C_0^- = 0$. An out-of-control (OC) signal is triggered as soon as $C_t^+ > h$ or $C_t^- < -h$ where h is a predetermined control limit. Comprehensive investigations of the properties and the average run length (ARL) of the CUSUM charts are given in (Brook and Evans 1972, Woodall and Adams 1993, Nishina and Nishiyuki 2003). A main advantage of the conventional CUSUM chart is that it provides good detection performance when a particular shift level is known beforehand. In practice, the shift level

is usually unknown. Therefore the CUSUM chart may not perform well when the actual mean shift is different from the particular shift level.

An adaptive CUSUM (ACUSUM) chart (Sparks 2000, Shu and Jiang 2006) and its variants (Jiang *et al.* 2008, Wu *et al.* 2009) have been proposed for detecting a range of mean shifts efficiently. The ACUSUM chart adjusts the reference value of the conventional CUSUM chart dynamically based on the shift size estimation. The ARL performance of ACUSUM is analyzed by a two-dimensional Markov chain (MC) model (Shu and Jiang 2006). Simulation and MC analysis reveal that the ACUSUM chart is more robust and efficient in detecting a range of mean shifts than the conventional CUSUM chart. On the other hand, a drawback of the ACUSUM chart occurs when the range $[\delta_{min}, \delta_{max}]$, where $\delta_{min} < \delta_{max}$, of mean shifts is wide. The ACUSUM chart is often insensitive to mean shifts close to δ_{min}, since the ACUSUM chart is designed to provide a very good perfomance when the shift size is close to the initial value of shift size estimation (Shu and Jiang 2006). In general, the initial value is set to the midpoint of the region to balance the efficiency in detecting overall mean shifts within the region.

In some environments, the process is very sensitive to even small variations of the air ambient temperature and such small changes may lead to a direct impact on the quality of processed gas (Bakker 2006). Thus, it is critical to detect small changes in the process parameters or product characteristics since small changes may significantly result in deterioration of product quality (Park *et al*. 2012, Jeong *et al*. 2006). In this chapter, we incorporate a runs rule with the ACUSUM chart to improve its sensitivity in detecting smaller mean shifts within the range.

Incorporating supplementary runs rule with Shewhart and CUSUM control charts have been explored to improve the ability of the charts in detecting small mean shifts (Champ and Woodall 1987, Koutras *et al.* 2007, Riaz *et al.* 2010). Champ and Woodall (1987) introduce a general form of supplementary runs rules, $T(n,m,a,b)$, indicating that an OC signal is triggered if n of the last m statistics fall within the interval (a,b) as

$$\text{Rule no. } 1: R_1 = \{T(1,1,-\infty,-3), T(1,1,3,\infty)\},$$
$$\text{Rule no. } 2: R_2 = \{T(2,3,-3,-2), T(2,3,2,3)\},$$
$$\text{Rule no. } 3: R_3 = \{T(4,5,-3,-1), T(4,5,1,3)\},$$
$$\text{Rule no. } 4: R_4 = \{T(8,8,-3,0), T(8,8,0,3)\}.$$

Rule no. 1 represents the Shewhart control charts with 3-sigma control limits and Rule no. 4 is the consecutive runs rule. These rules can be combined to form several composite rules. For example, Rules no. 1, no. 2 and no. 3 can be combined to form $R_{123} = R_1 \cup R_2 \cup R_3$. Champ and Woodall (1987) conclude that, when the power of the control chart increases, the false alarm rate also increases. Khoo (2003) and Zhang and Wu (2005) propose an optimal design of the supplementary runs rules to setup the control limits to maintain a desired in-control (IC) ARL (ARL_0). Among several supplementary runs rules, runs rule 4, "eight consecutive points on the same side of the center line," is effective in detecting small mean shifts (Champ and Woodall 1987). Acosta-Mejia (2007) and Lim and Cho (2009) suggest a general form of runs rule 4, m-of-m runs rule which signals if m consecutive points fall beyond a predefined threshold level, called a warning limit, and combine the m-of-m runs rule with the Shewhart chart. Riaz *et al.* (2010) propose 2-of-2 and 2-of-3 runs rules to improve the performance of CUSUM charts. However, the performance of the CUSUM charts supplemented with two runs rules illustrates just a small improvement compared to the conventional CUSUM charts.

In existing control charts incorporating *m*-of-*m* runs rule, the control chart generates an OC signal either when an observation falls outside the control limits or when *m* consecutive observations fall on the same side of the center line. However, when the mean shift is known, the optimal number of consecutive runs on the same side of the center line may be different depending on both the level of mean shift and a specified ARL_0 as shown in Lim and Cho (2009). When the actual mean shift size is unknown, one may consider incorporating an ACUSUM chart with *m*-of-*m* runs rule by changing the number of consecutive runs adaptively through the estimation of mean shift level. However, changing the number of consecutive runs adaptively depending on the shift level of the mean can cause difficulty in setting up the control limits of the integrated control chart to maintain a desired ARL_0 under IC process. Also, the MC analysis of the integrated control chart to study the ARL_1 performance under different shift scenarios is difficult to perform.

In order to overcome these challenges, we propose a new incremental scoring procedure which adjusts a score adaptively depending on the estimated shift level of the mean rather than changing the number of consecutive runs, by adopting the concept in a zone control chart (Jaehn 1987). In this way, we change the number of consecutive runs indirectly based on the estimated shift level. In the proposed scoring procedure, we assign a score to a consecutive run using the estimated shift size and accumulate the scores until the total score reaches a threshold score limit. This procedure signals an OC condition when the total score exceeds the threshold value or when the point falls outside the control limits. Further, we provide some guidelines for the design the proposed adaptive runs rule.

The chapter is organized as follows. In section 2.2, the ACUSUM chart is reviewed and an initial value of shift size estimation is discussed. In section 2.3, the ACUSUM chart with the adaptive runs rule (ACUSUM-ACR) is presented with the discussion of a proposed adaptive runs rule, followed by an example to illustrate the implementation. In section 2.4, the performance of the proposed chart is compared with other control charts including CUSUM, Spark's ACUSUM, AEWMA, and ACUSUM-C charts. Finally, concluding remarks are given as well as suggestions for future research.

2.2 ACUSUM chart

2.2.1 ACUSUM chart based on EWMA estimator

The ACUSUM chart proposed by Sparks (2000) adjusts the reference value k of the conventional CUSUM chart using an estimate $\hat{\delta}_t$ of the current mean size. We illustrate this by considering the upper-sided CUSUM chart in equation (2.1). Let $k_t^+ = \hat{\delta}_t^+ / 2$, where $\hat{\delta}_t^+$ is the current estimate of a positive mean shift available at time t. The upper Sparks' ACUSUM chart can be defined as

$$Z_t^+ = \max\{0, Z_{t-1}^+ + (X_t - k_t^+)/h(k_t^+)\}, \tag{2.2}$$

where $h(k)$ is an operating function that defines the control limit of the upper CUSUM statistic. Shu and Jiang (2006) obtain an approximation of $h(k)$ as follows;

$$h(k) = \begin{cases} \ln(1 + 2k^2 ARL_0^+ + 2.332k)/2k - 1.166, & \text{if } k \neq 0 \\ (ARL_0^+)^{1/2} - 1.166, & \text{if } k = 0 \end{cases}, \tag{2.3}$$

where ARL_0^+ is a pre-specified ARL of the upper ACUSUM chart. Similarly, the lower-sided ACUSUM statistic is $Z_t^- = \min\{0, Z_{t-1}^- - (X_t - k_t^-)/h(k_t^-)\}$. An OC signal is triggered when $Z_t^+ > h$ or $Z_t^- < -h$, where h is a predetermined control limit.

Although different schemes can be used to estimate shift size, the EWMA statistic is widely used to estimate the process mean due to its simplicity and efficiency as

$$\hat{\delta}_t = \lambda X_t + (1 - \lambda)\hat{\delta}_{t-1}, \tag{2.4}$$

where $0 \le \lambda \le 1$ is a smoothing parameter. A small value of λ is efficient in detecting small shifts but less efficient in detecting relatively larger shifts. However, the traditional EWMA statistic suffers from the "inertia problem" when the difference between the value of EWMA statistic and the target value is large before a change occurs (Yashchin 1987, Capizzi and Masarotto 2003, Woodall and Mahmoud 2005). If the value of the EWMA statistic is far below the target value and the mean of a process is shifted in the opposite direction, then it may take much longer for the EWMA estimator to react to the mean shifts. To overcome this, Sparks (2000) recommends using the upper EWMA statistic with a minimum positive value $\hat{\delta}_{\min}^+ > 0$ for improving the sensitivity to shifts $\delta \ge \delta_{\min}^+$ as

$$\hat{\delta}_t^+ = \max\{\delta_{\min}^+, \lambda X_t + (1 - \lambda)\hat{\delta}_{t-1}^+\},$$

where $\hat{\delta}_0^+$ is the initial value for positive shifts, and $\hat{\delta}_{\min}^+ = \delta_{\min}$ when the shift range $[\delta_{\min}, \delta_{\max}]$ is of interest. Although Sparks (2000) and Shu and Jiang (2006) do not define $\delta_{\max}^+$, it is reasonable to choose $\delta_{\max}^+ = \delta_{\max}$ for the purpose of improving the sensitivity to shifts $\delta \le \delta_{\max}^+$. Thus, the upper EWMA estimator can be derived as

$$\hat{\delta}_t^+ = \min\{\max\{\delta_{\min}^+, \lambda X_t + (1-\lambda)\hat{\delta}_{t-1}^+\}, \delta_{\max}^+\}, \tag{2.5}$$

where $\delta_{\min}^+ = \delta_{\min}$ and $\delta_{\max}^+ = \delta_{\max}$ are pre-specified minimum and maximum positive shifts for improving the sensitivity in detecting shifts larger than $\delta_{\min}^+$ and smaller than $\delta_{\max}^+$. Similarly, for negative mean shifts, the lower EWMA estimator is defined as

$$\hat{\delta}_t^- = \max\{\min\{\delta_{\min}^-, \lambda X_t + (1-\lambda)\hat{\delta}_{t-1}^-\}, \delta_{\max}^-\}, \tag{2.6}$$

where $\hat{\delta}_0^- = -\hat{\delta}_0^+$ is the initial value for negative shifts, and $\delta_{\min}^- = -\delta_{\min}$ and $\delta_{\max}^- = -\delta_{\max}$. In this section, we focus our discussion on the ACUSUM chart based on the estimators defined in equations (2.5) and (2.6). In addition, we reset $\hat{\delta}_t^+ = \hat{\delta}_0^+$ when $Z_0^+ \leq 0$ and $\hat{\delta}_t^- = \hat{\delta}_0^-$ when $Z_0^- \geq 0$, as suggested by Sparks (2000).

2.2 Initial value setting for the EWMA estimator

The initial estimate of a shift size has a considerable effect on the ARL performance of the ACUSUM charts. In general, ACUSUM charts are sensitive in detecting mean shifts close to $\hat{\delta}_0$. For the detection efficiency over a range of mean shifts, the midpoint of the range $[\delta_{\min}, \delta_{\max}]$, i.e., $\hat{\delta}_0 = 0.5(\delta_{\min} + \delta_{\max})$, is commonly used as an initial value (Shu and Jiang 2006). For positive and negative shift estimators, we set $\hat{\delta}_0^+ = \hat{\delta}_0$ and $\hat{\delta}_0^- = -\hat{\delta}_0$, respectively. Since we can consider the term 0.5 as a certain weight assigned to $\delta_{\min}$ and $\delta_{\max}$, the initial value can be generalized as

$$\hat{\delta}_0 = (1-w)\delta_{\min} + w\delta_{\max}, \tag{2.7}$$

where $0 \leq w \leq 1$. When $w = 0.5$, the intial value is equal to the midpoint suggested by Shu and Jiang (2006). Table 2.1 shows the zero-state and steady-state ARL performance of ACUSUM charts with different values of $w = 0, 0.25, 0.5, 0.75, 1.0$ for various mean shifts when $\lambda = 0.2$ and $[\delta_{min}, \delta_{max}] = [0.5, 4.0]$. The zero-state ARL (ZARL) is the ARL obtained with an assumption that a process change occurs at the initial stage, while the steady-state ARL (SARL) is the ARL computed assuming a process change take places after the process has been in control for some time (Lucas and Saccucci 1990). The zero-state and steady-state ARLs are obtained from extensive simulation experiments with 100,000 replications.

It is observed that the minimum ARL value for each δ occurs when the initial value of the shift is around δ. When $w = 0$, the initial value is set to δ_{min}, while it is set to δ_{max} when $w = 1.0$. In general, a small value of w enhances the sensitivity of ACUSUM charts to small shifts, and a large value of w improves the sensitivity to large shifts. However, choosing $w = 0$ or 1.0 considerably deteriorates the performance of detecting either large or small shifts. Therefore, Shu and Jiang (2006) recommend choosing $\hat{\delta}_0$ as the midpoint of $[\delta_{min}, \delta_{max}]$, i.e. $w = 0.5$ for balancing the efficiency in detecting both small and large shifts.

2.3 ACUSUM chart with adaptive runs rule

The ACUSUM with $\hat{\delta}_0 = 0.5(\delta_{min} + \delta_{max})$ is insensitive to shifts close to δ_{min} when the range $[\delta_{min}, \delta_{max}]$ of mean shifts is wide. Specifically, when $\delta = 0.5$, the ARL = 135 for the range [0.5, 6.0] and 63 for the range [0.5, 4.0] as shown in Table 2.4. In order to

improve the efficiency in detecting a wide range of shifts, we propose an adaptive runs rule in this section.

2.3.1 Adaptive runs rule

An adaptive runs rule is based on the shift size estimation used in ACUSUM charts. We assign scores to consecutive runs using the estimated shift size and add these scores until the total score reaches a limit similar to the score function of a zone control chart (Jaehn 1987). We propose a new scoring function using the value of the current observation and the estimated shift size of the mean. For the upper adaptive runs rule for positive mean shift detection, we define the cumulative score as

$$N_t^+ = \begin{cases} \max\{0, N_{t-1}^+ + \phi(\hat{\delta}_t^+, X_t)\}, & \text{if } Z_t^+ > 0 \\ 0, & \text{if } Z_t^+ = 0 \end{cases}, \tag{2.8}$$

where $N_0^+ = 0$ and $\phi(\bullet)$ is a score function of the value of the current observation and the value of the current mean shift estimate. The score is added when the value of current statistic Z_t^+ is positive, otherwise the previous total score Z_{t-1}^+ is reset to zero.

When the mean changes and the magnitude of the change is known, the log-likelihood ratio (motivated by the log-likelihood ratio test) can be used as a good scoring function. Assume that X_t has a probability density function $f_0 \sim N(0,1)$ from an IC process, while X_t has a probability density function $f_1 \sim N(\hat{\delta}_t^+, 1)$ from an OC process. The log-likelihood ratio of X_t can be defined as

$$\phi(\hat{\delta}_t^+, X_t) = \log \frac{f_1(X_t)}{f_0(X_t)} = \log \frac{e^{-(X_t - \hat{\delta}_t^+)^2/2}}{e^{-X_t^2/2}} = \hat{\delta}_t^+ (X_t - \hat{\delta}_t^+/2). \tag{2.9}$$

Table 2.1 Zero- and steady-state ARLs of ACUSUM charts with different values of w

when $\lambda = 0.2$ for detecting shifts within the range [0.5, 4.0]

	$w = 0.0$, $h = 0.927$		$w = 0.25$, $h = 1.000$		$w = 0.5$, $h = 1.025$		$w = 0.75$, $h = 1.058$		$w = 1.0$, $h = 1.040$	
δ	ZARL	SARL	ZARL	SARL	ZARL	SARL	ZARL	SARL	ZARL	SARL
0.0	500.39	493.19	501.45	496.79	500.43	499.90	501.68	500.46	499.83	499.35
0.5	34.08	31.81	39.51	37.88	63.53	62.73	126.88	126.58	179.68	179.64
1.0	10.91	9.88	11.00	10.40	13.33	13.00	21.58	21.42	37.69	37.63
1.5	6.16	5.55	5.58	5.30	5.82	5.67	7.32	7.24	10.64	10.60
2.0	4.30	3.88	3.64	3.48	3.46	3.37	3.81	3.76	4.62	4.61
2.5	3.37	3.04	2.73	2.63	2.43	2.38	2.45	2.43	2.69	2.68
3.0	2.81	2.55	2.24	2.17	1.89	1.86	1.81	1.79	1.86	1.85
3.5	2.42	2.22	1.93	1.87	1.55	1.53	1.44	1.43	1.43	1.42
4.0	2.17	2.02	1.72	1.67	1.32	1.31	1.22	1.22	1.21	1.20

If $X_t > \hat{\delta}_t^+ / 2$, then the cumulative score is increased by $\phi(\bullet)$. Otherwise, N_t^+ is not changed or decreased by $\phi(\bullet)$. The adaptive runs rule triggers an OC signal whenever $N_t^+ > l$, where l is the limit of the accumulated score. Analogously, the cumulative score of the lower adaptive runs rule for negative mean shifts is defined as

$$N_t^- = \begin{cases} \max\{0, N_{t-1}^- + \phi(\hat{\delta}_t^-, X_t)\}, & \text{if } Z_t^- < 0 \\ 0, & \text{if } Z_t^- = 0 \end{cases}, \tag{2.10}$$

where $N_0^- = 0$.

It is interesting to study the effect of l on the ARL performance of ACUSUM. Table 2.2 compares the zero-state ARL values of ACUSUM-ACR with different values of l when $w = 0.5$, $\lambda = 0.2$, and $[\delta_{\min}, \delta_{\max}] = [0.1, 4.0]$. When $l = \infty$, the adaptive runs

rule could not generate OC signals so that ACUSUM-ACR provides the same ARL performance with Spark's ACUSUM, while only adaptive runs rule is responsible for OC signals when $h = \infty$. Note that both charts use the same EWMA estimator to estimate the process mean.

Table 2.2 also reports the proportions of OC signals generated by the adaptive runs rule. For instance, when $l = 6.04$ and $\delta = 0.5$, about 70% of detections are from the runs rule; this indicates that the adaptive runs rule plays an important role in detecting the small mean shift. In general, the proposed adaptive runs rule with smaller l is more sensitive in detecting small shifts. However, when the value of l is too small, it is difficult to maintain ARL_0 to a specific value regardless of the value of h. In Table 2.2, for $ARL_0 = 500$, the minimum value of l is 6.03. It is thus reasonable to choose l and h with the proportion of false alarms equal to 50% in order to improve sensitivity to small shifts. Table 2.3 shows the parameters l and h for various shift regions when $w = 0.5$ and $\lambda = 0.2$. The corresponding ARL_0 is equal to 500 and the proportion of false alarm by the adaptive runs rule is 50%.

Table 2.4 compares the zero- and steady-state ARL values of ACUSUM-ACR and ACUSUM for different values of δ_{max} when $w = 0.5$, $\lambda = 0.2$, and $\delta_{min} = 0.5$. The ACUSUM is inefficient in detecting shifts around δ_{min} when δ_{max} is large, and the adaptive runs rule makes the ACUSUM more sensitive to small shifts. Specifically, when the range is [0.5, 6.0], ACUSUM-ACR is more sensitive to shifts ≤ 2.0 than ACUSUM. On the other hand, when the range is small like [0.5, 2.0], the adaptive runs rule improves the performance of ACUSUM in a small region.

Table 2.2 Zero-state ARL comparisons of ACUSUM using adaptive runs rule with

different l and h ($w = 0.5$)

	ACUSUM with adaptive runs rule $\lambda = 0.2$, $[\delta_{min}, \delta_{max}] = [0.1, 4.0]$								Spark's ACUSUM
	$l = 6.03$, $h = \infty$		$l = 6.04$, $h = 1.12$		$l = 6.45$, $h = 1.06$		$l = 8.2$, $h = 1.03$		$l = \infty$, $h = 1.014$
δ	ARL	Prop (%)	ARL	Prop (%)	ARL	Prop (%)	ARL	Prop (%)	ARL
0.0	501.14	100	500.90	50.12	499.51	25.39	500.30	9.72	501.18
0.1	380.69	100	380.13	55.04	383.62	32.11	391.73	16.25	418.75
0.5	45.51	100	45.51	69.16	46.49	52.57	49.24	31.40	52.98
1.0	11.75	100	11.75	49.43	12.01	27.08	12.46	5.49	12.44
1.5	5.71	100	5.71	24.03	5.73	7.13	5.71	0.35	5.63
2.0	3.61	100	3.59	10.84	3.55	1.38	3.49	0.01	3.45
3.0	2.03	100	2.01	6.45	1.99	0.06	1.95	0.00	1.93
4.0	1.42	100	1.40	6.98	1.41	0.00	1.38	0.00	1.36

* Prop is the proportion of signals detected by the adaptive runs rule.

Table 2.3 Parameters of ACUSUM-ACR with various shift ranges when Prop = 50 %

δ_{min}	δ_{max}	ACUSUM-ACR		ACUSUM
		h	L	h
	2.0	1.09	6.92	0.996
0.50	4.0	1.12	5.78	1.025
	6.0	1.20	5.25	1.060
	2.0	1.10	8.34	1.130
0.25	4.0	1.13	5.94	1.017
	6.0	1.19	5.32	1.060

Table 2.4 Zero-state and steady-state ARL comparisons of ACUSUM-ACR and

ACUSUM with different $\delta_{\max}$ when $w = 0.5$ and $\lambda = 0.2$

δ	ACUSUM-ACR				ACUSUM	
	ZARL	Prop (%)	SARL	Prop (%)	ZARL	SARL
$[\delta_{\min}, \delta_{\max}] = [0.5, 6.0]$						
0	500.23	50.46	499.79	50.61	500.49	499.46
0.5	**118.91**	**57.08**	**118.55**	**57.12**	135.87	135.56
1.0	**20.79**	**58.55**	**20.63**	**58.69**	23.18	23.04
1.5	**6.99**	**49.56**	**6.91**	**49.83**	7.65	7.56
2.0	**3.69**	**34.49**	**3.64**	**34.98**	3.88	3.84
3.0	1.80	13.98	1.79	14.39	1.79	1.78
4.0	1.23	5.91	1.22	6.04	1.22	1.21
5.0	1.04	1.54	1.04	1.58	1.04	1.04
6.0	1.00	0.20	1.00	0.21	1.00	1.00
$[\delta_{\min}, \delta_{\max}] = [0.5, 4.0]$						
0.0	502.84	50.13	499.12	50.38	501.54	500.86
0.5	**54.88**	**68.63**	**53.81**	**68.96**	63.89	62.87
1.0	**12.21**	**55.65**	**11.84**	**56.47**	13.30	12.98
1.5	**5.67**	**33.85**	**5.47**	**35.20**	5.78	5.62
2.0	3.50	18.43	3.39	20.03	3.45	3.37
3.0	1.94	9.27	1.89	10.85	1.88	1.85
4.0	1.34	7.42	1.32	8.55	1.32	1.31
$[\delta_{\min}, \delta_{\max}] = [0.5, 2.0]$						
0.0	501.92	50.16	495.30	50.985	501.27	496.82
0.5	**33.89**	**63.15**	**31.72**	**64.975**	38.23	36.51
1.0	11.28	25.48	10.44	30.225	10.94	10.31
1.5	6.04	6.67	5.62	12.87	5.63	5.33
2.0	3.98	2.54	3.72	8.96	3.71	3.53

2.3.2 Parameter selections

The design of an ACUSUM-ACR chart involves the choice of parameters $\delta_{\min}$, $\delta_{\max}$, λ, l and h. In this section, we provide guidelines for the choice of the parameters of ACUSUM-ACR charts. Suppose $[\delta_1, \delta_2]$ is the range of potential mean shifts to be detected. Based on Sparks (2000) and Shu and Jiang (2006), we recommend the following guidelines for designing ACUSUM-ACR charts.

- Set $[\delta_{\min}, \delta_{\max}] = [\delta_1, \delta_2]$, and $\delta_{\min} = \delta_{\min}^+ = -\delta_{\min}^-$ and $\delta_{\max} = \delta_{\max}^+ = -\delta_{\max}^-$.

- Choose $\hat{\delta}_0 = (\delta_{\min} + \delta_{\max})/2$ to balance the efficiency in detecting shifts over the range. Set $\hat{\delta}_0^+ = \hat{\delta}_0$ and $\hat{\delta}_0^- = -\hat{\delta}_0$ for positive and negative shift estimators, respectively.

- Choose a EWMA parameter λ in the range $[0.05, 0.25]$. In general, $\lambda = 0.1$ and $\lambda = 0.2$ result in good performance of the chart

- Reset $\hat{\delta}_{t+1}^+ = \hat{\delta}_0^+$, $N_t^+ = 0$ when $Z_{t+1}^+ \leq 0$, and $\hat{\delta}_{t+1}^- = \hat{\delta}_0^-$, $N_t^- = 0$ when $Z_{t+1}^- \geq 0$. When $Z_{t+1}^+ \leq 0$ or $Z_{t+1}^- \geq 0$, the previous process is very unlikely to have positive or negative shifts so that it is reasonable to reset the mean estimation and the cumulative score to its initial value for the following observations.

- Select h and l to achieve the desired $ARL0$ with the proportion of false alarm generated by an adaptive runs rule equivalent to 50%.

Note that the above parameter selections are not intended to optimize detection for any particular range of shifts. In section 4, we show that these guidelines improve detection performance over the specified range $[\delta_{\min}, \delta_{\max}]$.

2.3.3 An illustrative example

To illustrate an adaptive runs rule scheme with ACUSUM, we use a data set of simulated observations taken from Lucas and Saccucci (1990). The data set is shown in Table 2.5 and contains 19 observations from normal distribution. The target value of the first 10 observations is 0 and the target value of the last 9 observations is shifted to one standard deviation. The parameters of the ACUSUM-ACR are chosen to be $w = 0.5$, $\lambda = 0.2$, $h = 0.862$ and $l = 5.0$ with $ARL_0 = 200$ for detecting mean shifts within $[0.5, 4.0]$.

Also in Table 2.5, the third and fourth columns show the values of ACUSUM statistics, Z_t^+ and Z_t^-, and the fifth and sixth columns contain the values of positive and negative estimates of mean shifts, respectively. The final two columns contain the values of the cumulative scores. For example, the score of the 11th observation for positive shifts is calculated as $\phi(\hat{\delta}_{11}^+, X_{11}) = \hat{\delta}_{11}^+(X_{11} - \hat{\delta}_{11}^+/2) = 2.04(1.2 - 2.04/2) = 0.3672$. As $N_{10}^+ = 0$, we obtain $N_{11}^+ = 0.3672$. As shown in Figure 2.1, which plots the values of Z_t^+ and N_t^+ for detecting positive shifts, the proposed adaptive runs rule gives an OC signal at the 16th observation, whereas the ACUSUM signals at the 17th observation.

Table 2.5 Example of an ACUSUM and adaptive runs rule using data (Lucas and

Saccucci 1990)

T	Observed Value	Z_t^+	Z_t^-	$\hat{\delta}_t^+$	$\hat{\delta}_t^-$	N_t^+	N_t^-
0		0	0			0	0
1	1	0	0	2.25	-2.25	0	0
2	-0.5	0	0	2.25	-2.25	0	0
3	0	0	0	2.25	-2.25	0	0
4	-0.8	0	0	2.25	-2.25	0	0
5	-0.8	0	0	2.25	-2.25	0	0
6	-1.2	0	-0.0324	2.25	-2.04	0	0
7	1.5	0.1618	0	2.10	-2.25	0.945	0
8	-0.6	0	0	2.25	-2.25	0	0
9	1	0	0	2.25	-2.25	0	0
10	-0.9	0	0	2.25	-2.25	0	0
11	1.2	0.0324	0	2.04	-2.25	0.3672	0
12	0.5	0	0	2.25	-2.25	0	0
13	2.6	0.6365	0	2.32	-2.25	3.3408	0
14	0.7	0.4310	0	2.00	-2.25	2.7460	0
15	1.1	0.4696	0	1.82	-2.25	3.0941	0
16	2	0.8444	0	1.85	-2.25	**5.0834**	0
17	1.4	**1.0102**	0	1.76	-2.25	5.9976	0
18	1.9	1.3493	0	1.79	-2.25	7.7965	0
19	0.8	1.3171	0	1.59	-2.25	7.8028	0
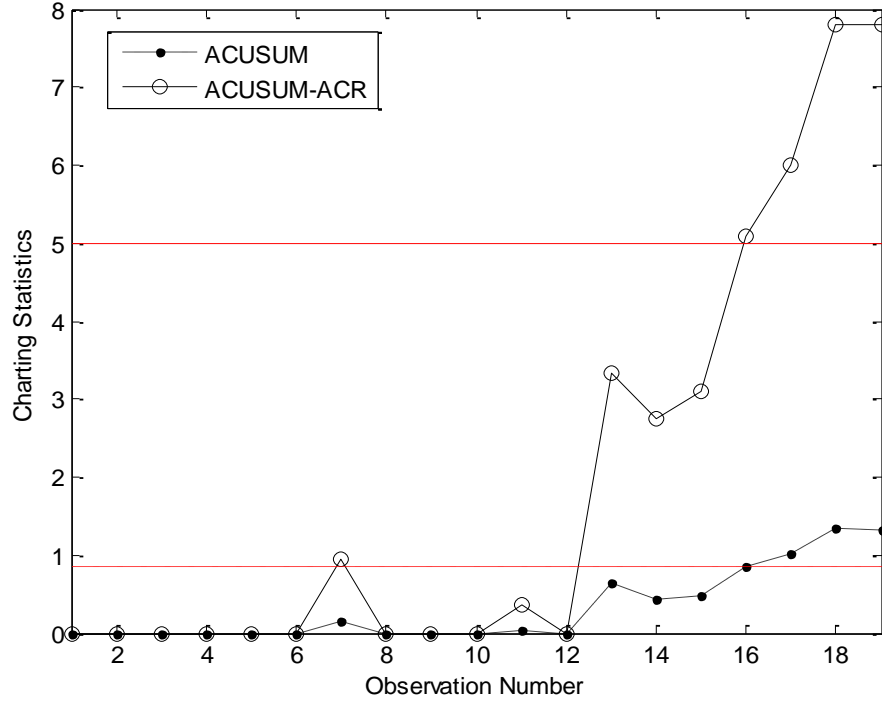
Figure 2.1 ACUSUM and ACUSUM-ACR charts. The solid and dashed horizontal

lines indicate control limits of ACUSUM-ACR and ACUSUM, respectively.

2.4 Performance comparisons

In the previous section, we compare ARL performance of the proposed

ACUSUM-ACR charts with that of ACUSUM charts for different ranges of mean shifts.

In this section, we compare the effectiveness of the ACUSUM-ACR in detecting both

small and large mean shifts with the conventional CUSUM and recent variants of

adaptive charts including AEWMA (Capizzi and Masarotto 2003) and ACUSUM-C

(Jiang et $al.$ 2008).

MC approximations of the ACUSUM and the ACUSUM-ACR are described in

Appendix A. The ACUSUM chart can be represented by the two-dimensional Markov

random vector $(\hat{\delta}_t^+, Z_t^+)$ to evaluate the ARL. The two-dimensional MC approaches (Shu

and Jiang 2006, Jiang *et al.* 2008) are extensions of existing univariate MC approaches (Brook and Evans 1972, Lucas and Saccucci 1990, Capizzi and Masarotto 2003). Similarly, the ACUSUM-ACR chart can be represented by the three-dimensional Markov random vector $(\hat{\delta}_t^+, Z_t^+, N_t^+)$ by extending the two-dimensional MC approach of the improved ACUSUM. The MC analysis of the ARL provides approximate results of Monte Carlo simulations.

2.4.1 Comparisons with CUSUM and ACUSUM charts

Figure 2.2 contrasts the (a) zero-state and (b) steady-state ARL values of ACUSUM-ACR chart and CUSUM charts with $k = \delta_{min}^+ / 2$ and $k = \delta_{max}^+ / 2$ for detecting shifts within the range $[\delta_{min}^+, \delta_{max}^+] = [0.5, 4.0]$. The ACUSUM-ACR chart is designed for providing good detection performance over a range of mean shifts [0.5, 4]. The zero-state ARL_0 values for both charts are maintained at 500. Both zero-state and steady-state ARL curves of the ACUSUM-ACR chart are almost always at the bottom. Similar observations are made for ACUSUM-ACR charts designed for moderate and small ranges of shifts.

2.4.2 Comparisons with recent variants of ACUSUM charts

In Figure 2.3, the zero-state and steady-state ARLs of ACUSUM-ACR are compared with AEWMA (Capizzi and Masarotto 2003) and ACUSUM-C (Jiang *et al.* 2008) charts. In order to detect shifts between 0.5 and 4.0 efficiently, we set $\delta_{min} = 0.5$ and $\delta_{max} = 4.0$, and select $\lambda = 0.2$. The parameters $h = 1.12$ and $l = 5.78$ of the ACUSUM-ACR chart are selected to satisfy an $ARL_0 = 500$. The zero-state and steady-

state ARLs are obtained from extensive simulation experiments with 100,000 replications.
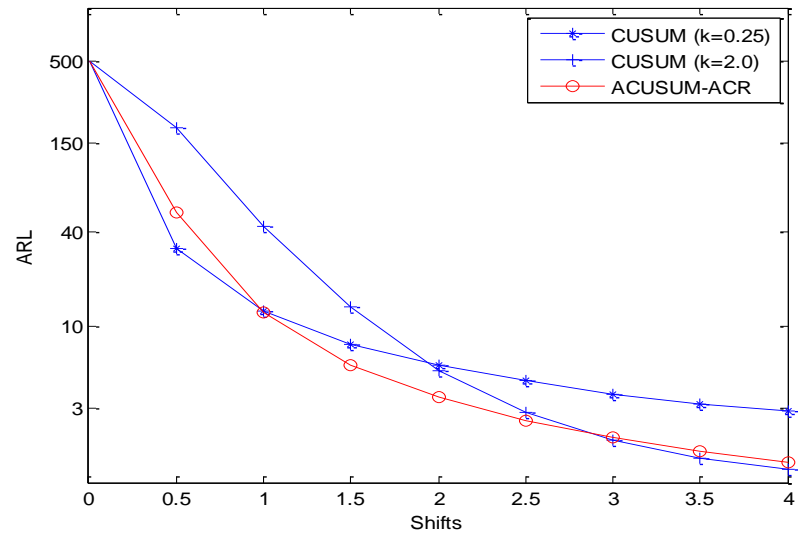
(a)
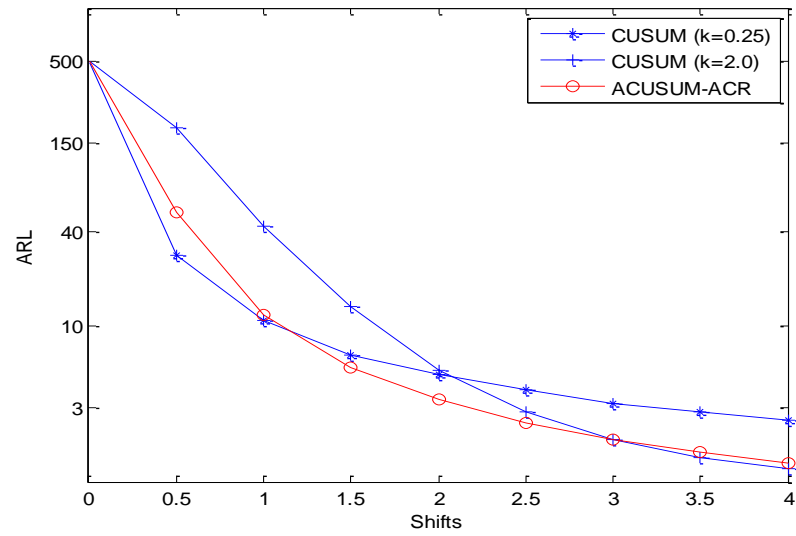
(b)

Figure 2.2 (a) Zero-state and (b) steady-state ARL comparisons between ACUSUM-ACR

and CUSUM with $k = 0.25$ and 2.0
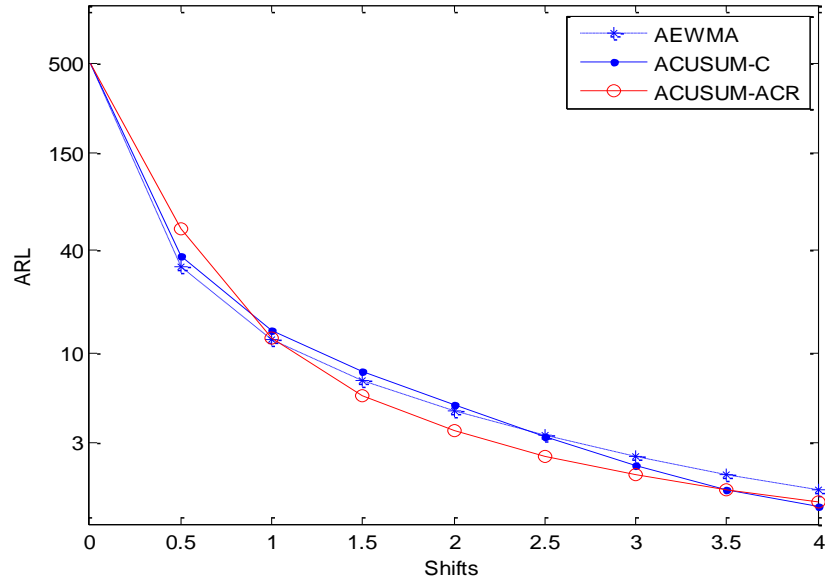
(a)
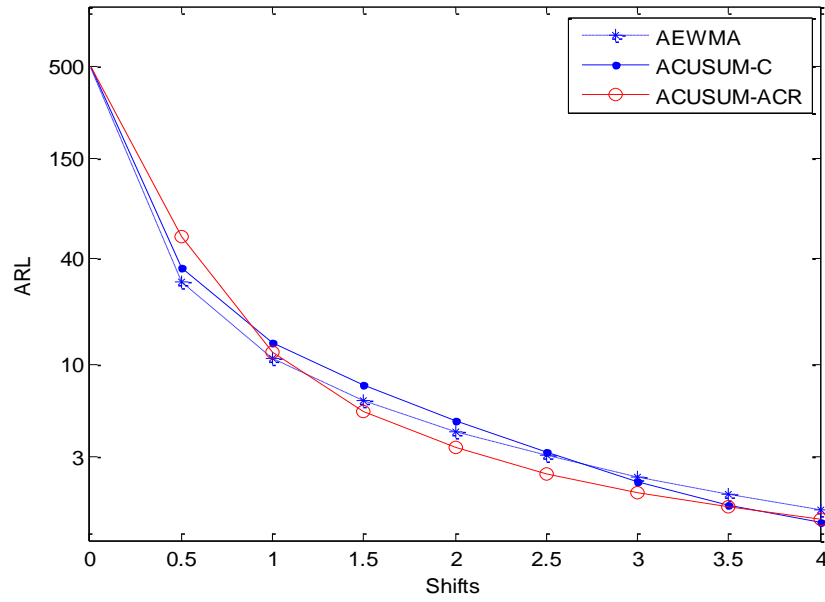
(b)

Figure 2.3 (a) Zero-state and (b) steady-state ARL comparisons between AEWMA,

ACUSUM-C, and ACUSUM-ACR for the range [0.5, 4.0]

Yashchin (1995) suggests the generalization of the EWMA (EWMA-C) statistic based on Huber's score function. The AEWMA chart proposed by Capizzi and Masarotto (2003) is an adaptive version of EWMA chart based on the EWMA-C statistic. Capizzi and Masarotto (2003) show that the AEWMA chart is more effective than the EWMA chart for detecting a wide range of mean shifts. The AEWMA chart with $\lambda = 0.0398$, $\gamma = 2.899$, $h = 0.4309$ and $\delta_{min}^{+} = 0.5$ in Figure 2.3 is optimized for detecting mean shifts over [0.5, 4]. Compared with the AEWMA chart, the detection performance of the ACUSUM-ACR chart is more effective for mean shifts when $1.0 \leq \delta \leq 4.0$.

Another adaptive chart based on the EWMA-C statistic for detecting a wide range of mean shifts is the ACUSUM-C chart, which was proposed by Jiang *et al.* (2008). The ACUSUM-C chart is an extension of Sparks' ACUSUM chart based on the EWMA-C statistic. According to Jiang *et al.* (2008), the ACUSUM-C chart with $\lambda = 0.2$, $\gamma = 2.5$, $\delta_{min}^{+} = 0.5$, and $h = 6.13$ provides an overall good performance in detecting mean shifts over the range [0.5, 4.0]. Interestingly, the ACUSUM-ACR is more sensitive than the ACUSUM-C chart in detecting mean shifts when $1.0 \leq \delta \leq 3.5$.

2.5 Conclusions

Traditionally, SPC charts are used to monitor processes in order to reduce process variability and improve product quality. Although the ACUSUM chart is more efficient in detecting a wider range of mean shifts than the conventional CUSUM chart, the ACUSUM is often insensitive to mean shifts close to the minimum level of the range when the range of the shift is relatively large. Some runs rules have been widely used with control charts to improve the performance in detecting small mean shifts. However, there is no efficient runs rule for supplementing the ACUSUM chart.

In this chapter, we propose a new adaptive runs rule and supplement the ACUSUM chart to improve its sensitivity for small mean shifts. The proposed adaptive runs rule is based on a score function from the estimated shift in consecutive runs. The performance of the ACUSUM-ACR is investigated and compared with CUSUM and variants of adaptive charts such as ACUSUM, AEWMA and ACUSUM-C charts. The comparisons reveal the effectiveness of the proposed adaptive runs rule in shift detection for a wide range of shift magnitudes.

The proposed adaptive runs rule can be designed to detect linear shifts of the process mean efficiently in future research. In addition, the rule can be extended to a multivariate control chart.

CHAPTER 3

MULTIVARIATE STATISTICAL PROCESS CONTROL CHARTS BASED ON THE APPROXIMATE SEQUENTIAL χ^2 TEST

3.1 Introduction

Multivariate statistical process control (MSPC) charts have been widely used to detect process changes by monitoring multiple quality characteristics and/or process parameters. In this chapter, we assume that a process has p quality characteristics and measurements, $\mathbf{X}_1, \mathbf{X}_2, \ldots$, are independently and identically distributed random vectors following a multivariate normal distribution with a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$, that is, $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. It is assumed that the first $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_\tau$ are from an in-control (IC) process with the process mean $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, while $\mathbf{X}_{\tau+1}, \mathbf{X}_{\tau+2}, \ldots$ are from an out-of-control (OC) process with $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. The covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ is assumed to be known and fixed over time, and the change time τ is unknown. For successive observations, multivariate control chart approaches for monitoring the mean of a multivariate normal process can be interpreted as repeated tests of significance of the form

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \text{ versus } H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0, \tag{3.1}$$

where $\boldsymbol{\mu}$ represents a multivariate normal process mean. The Hotelling's T^2 statistic is the classical test statistic for the hypotheses. It is defined as $T^2 = \left(\mathbf{X} - \boldsymbol{\mu}_0 \right)' \boldsymbol{\Sigma}_0^{-1} \left(\mathbf{X} - \boldsymbol{\mu}_0 \right)$, which follows a χ^2 distribution with p degrees of freedom (df). The average runs length (ARL) performance of control charts based on the T^2 statistic is dependent only on the distance of $\boldsymbol{\mu}$ from $\boldsymbol{\mu}_0$ (Lowry and Montgomery 1995), where distance is defined as the square root of the noncentrality parameter $\lambda^2 (\boldsymbol{\mu})$ of a χ^2 distribution given by

$$\lambda^2 (\boldsymbol{\mu}) = \left(\boldsymbol{\mu} - \boldsymbol{\mu}_0 \right)' \boldsymbol{\Sigma}_0^{-1} \left(\boldsymbol{\mu} - \boldsymbol{\mu}_0 \right).$$

In this chapter, distance $\lambda(\boldsymbol{\mu}_1)$ is used to represent the shift size of a mean vector $\boldsymbol{\mu}_1$ from $\boldsymbol{\mu}_0$.

The Hotelling's T^2 statistic is an optimal test statistic in detecting a mean change based on a single multivariate observation (Lowry and Montgomery 1995). However, with sequential observations, the T^2 statistic is not efficient in detecting small and moderate shifts of the mean vector, as it uses only the most recent observation. As alternatives to the Hotelling's control chart, control charts using information given by the entire sequence of observations like multivariate EWMA (MEWMA) (Lowry *et al.* 1992) and multivariate CUSUM (MCUSUM) (Alwan 1986, Healy 1987, Crosier 1988, Pignatiello and Runger 1990) have been proposed to detect small mean shifts efficiently. In MEWMA charts, test statistics are T^2 values of weighted sum of observations similar to a univariate EWMA scheme.

Woodall and Ncube (1985) suggest using p univariate CUSUM charts simultaneously to monitor p variables. This scheme generates an OC signal whenever any chart signals. Crosier (1988) proposes two multivariate CUSUM charts: cumulative

sum of T values (COT) and MCUSUM. The COT is based on the univariate CUSUM statistics of T values, and the MCUSUM is obtained by replacing the scalar values of a univariate CUSUM by vectors. Pignatiello and Runger (1990) also propose two multivariate CUSUM charts, MC1 and MC2. The MC2 chart accumulates the squared distance from an observation vector to the target mean vector similar to the univariate CUSUM chart. The MC1 chart accumulates the deviation vectors, $\mathbf{X} - \boldsymbol{\mu_0}$, rather than accumulate the squared distance. The MCUSUM chart provides similar ARL performance but is more complicated than the MC1 chart. Further, it is known that the MC1 chart shows the best zero-state ARL performance among multivariate CUSUM charts when processes are initially in OC state (Golosnoy *et al.* 2009). However, the MC1 chart shows poor steady-state ARL performance when p is large. The zero-state ARL is the ARL obtained with an assumption that a process change occurs at the initial stage, while the steady-state ARL is the computed ARL assuming a process change takes place after the process has been in control for some time (Lucas and Saccucci 1990).

Sequential tests, based on likelihood ratios, have been used in statistical process control (SPC) successfully (Page 1954, Healy 1987, Jiang and Tsui 2008, Mahmoud *et al*. 2008, Reynolds and Lou 2010, Ou *et al.* 2012, Tsui *et al.* 2012). The univariate CUSUM (Page 1954) is one of the most popular charts, which consists of a set of sequential probability ratio tests (SPRT) (Wald 1947). It is known that the CUSUM chart has certain optimality properties in terms of ARL (Lorden 1971, Pollack 1985, Moustakides 1986). Healy (1987) proposes a multivariate CUSUM chart using sequential log-likelihood ratios (log-LR), and the statistic accumulates a linear combination of the p normal random variables. It is well known that the test statistic

based on the SPRT is optimal for detecting a shift mean μ_1 in terms of ARL when μ_1 is known (Moustakides 1986).

A drawback of Healy's chart, which solves a simple hypothesis test as $H_0 : \mu = \mu_0$ versus $H_1 : \mu = \mu_1$, where $\mu_1 \neq \mu_0$, is that it is designed to detect a specific mean vector μ_1. When p is large, it becomes difficult to specify a meaningful single alternative since there are infinitely many points in p-space. When only a few components of the real shifted mean are different from those of μ_1, Healy's chart (1987) may perform far from optimal (Lowry and Montgomery 1995). Based on the drawback of Healy's scheme, it is natural to consider other procedures that operate within the surfaces of p-dimensional ellipsoids. For instance, the statement $\mu = \mu_0$ is equivalent to $\lambda(\mu) = 0$. Similarly, the alternative hypothesis would be of the same form but equal to a scalar value as $\lambda(\mu_1) = \lambda_1$ when $\mu_1 \neq \mu_0$. The hypotheses become

$$H_0 : \lambda = 0 \text{ versus } H_1 : \lambda = \lambda_1, \tag{3.2}$$

where $\lambda_1 \neq 0$. Alwan (1986) proposes a multivariate CUSUM chart using SPRT scheme based on the above hypotheses as a simple extension of univariate CUSUM. However, a certain optimality property of the SPRT is not satisfied under the hypotheses of equation (3.2) as the SPRT is designed for testing parameters, not for testing functions of parameters such as $\lambda(\mu)$, which is a function of the unknown parameter μ (Lai 1981).

Although the SPRT for testing the hypotheses in equation (3.1) is used in SPC charts effectively, there is no MSPC chart based on a sequential χ^2 test (Jackson and Bradley 1961) for testing the hypothesis in equation (3.2). A sequential χ^2 test is based on a probability ratio for the detection of a change in noncentrality parameter of a χ^2

distribution. The sequential χ^2 test has an asymptotic optimality property (Lai 1981). However, as Crosier (1988) points out, the probability ratio of a sequential test for the noncentrality parameter is intractable mathematically, and the test statistic based on the ratio is not clear as it has no closed form expression. Thus, even though an existing sequential χ^2 test can be used for developing an MSPC chart, it may not be appropriate for the on-line process monitoring if a closed form expression of the test statistic is not obtained.

In this chapter, we develop a closed form expression of approximate sequential χ^2 test based on the approximation to a noncentral χ^2 distribution. The approximation provides a closed form of the test statistic for testing the hypotheses in equation (3.2). By adapting the approximate test statistic, we propose a novel multivariate approximate sequential chi-square (MASC) chart for monitoring a specific shift size λ_1. The test statistic of the MASC chart is based on the cumulative moving average and its T^2 value with a reference used in CUSUM-type control charts. Then, we propose an adaptive MASC (AMASC) chart for detecting a wide range of shift by adjusting the reference value of the proposed chart dynamically motivated by the concepts of the adaptive CUSUM (ACUSUM) (Sparks 2000, Shu and Jiang 2006). By using an EWMA estimator of mean shift rather than the pre-specified reference value, we can adjust the reference value dynamically.

3.2 Approximate sequential χ^2 test

In this section, we construct the approximate sequential χ^2 test statistic which has a closed-form expression. A sequential statistical test is used for solving hypotheses testing problems when the sample number is not fixed a priori but depends on the data that have been already observed. The sequential χ^2 test is a multivariate sequential test based on the two simple hypotheses H_0 and H_1 in equation (3.2). After n data are observed, we obtain $\left\|\bar{\mathbf{X}}_n - \boldsymbol{\mu}_0\right\| = \left(\bar{\mathbf{X}}_n - \boldsymbol{\mu}_0\right)' \boldsymbol{\Sigma}_0^{-1} \left(\bar{\mathbf{X}}_n - \boldsymbol{\mu}_0\right)$, where $\bar{\mathbf{X}}_n = (\mathbf{X}_1 + \ldots + \mathbf{X}_n)/n$ is a sample mean vector based on n observations. Under H_1, $n\left\|\bar{\mathbf{X}}_n - \boldsymbol{\mu}_0\right\|$ follows a noncentral χ^2 distribution $\chi^2(p, n\lambda_1^2)$; while under under H_0, it follows a central χ^2 distribution. The test statistic based on the log-likelihood ratio is defined as

$$\Lambda_n = \log \frac{f(n\left\|\bar{\mathbf{X}}_n - \boldsymbol{\mu}_0\right\| \mid \lambda = \lambda_1)}{f(n\left\|\bar{\mathbf{X}}_n - \boldsymbol{\mu}_0\right\| \mid \lambda = 0)}, \tag{3.3}$$

where $f(\bullet)$ is a pdf of χ^2 distribution. The test statistic Λ_n consists of a generalized hypergeometric function defined in the form of infinite hypergeometric series (Jackson and Bradley 1961).

A closed form of Λ_n may be useful in practice for repeated tests. To obtain a closed form of Λ_n, we may approximate the generalized hypergeometric function using a specified number of hypergeometric series of the function. However, the error of the approximation can be large when the selected number of terms is small, while the computational overhead may be high and on-line monitoring may not be possible when the selected number of terms is large. To overcome this problem, several researchers have proposed various approximations to the noncentral χ^2 distribution (Patnaik 1949,

Cox and Reid 1987). Different schemes can be used to approximate a noncentral χ^2 distribution. In this section, we use Patnaik's approximation due to its computational efficiency (Cox and Reid 1987). It is known that the error of Patnaik's approximation to noncentral chi-squared cdf is $O(\lambda^2)$ as $\lambda \to 0$, $O(\lambda^{-1/2})$ as $\lambda \to \infty$ for a fixed p (Johnson and Kotz 1970).

Using Patnaik's approximation and rescaling, Λ_n can be approximated by (see Appendix B for details)

$$\tilde{\Lambda}_n = w_n \log\left(n\|\bar{\mathbf{X}}_n - \boldsymbol{\mu}_0\|\right) - k_n ,\tag{3.4}$$

where

$$k_n = \left(w_n \log 2 + 2\log\Gamma\left(\frac{p+w_n}{2}\right)\right) - 2\left(\log\frac{p+2\tilde{\lambda}_n^2}{p+\tilde{\lambda}_n^2} + \log\Gamma\left(\frac{p}{2}\right)\right)\tag{3.5}$$

and $w_n = \dfrac{\tilde{\lambda}_n^4}{p+2\tilde{\lambda}_n^2}$, $\tilde{\lambda}_n = \sqrt{n}\lambda_1$, and $\Gamma(\bullet)$ is a gamma function. When $n = 1$, $\tilde{\lambda}_1$ is equal to λ_1.

Note that the proposed statistic may be asymptotically optimal in testing a noncentrality parameter $\lambda = \lambda_1 + \varepsilon$ due to the error of the proposed approximation $\tilde{\Lambda}_n$ to Λ_n. Figure 3.1 compares the expected values of $\tilde{\Lambda}_n$ and Λ_n as functions of n with $\lambda_1 = 1.0$, when $p = 2$ and 20. For simplicity, we set $\boldsymbol{\mu}_0 = \mathbf{0}$. The solid line represents the expected values of Λ_n for different n values, and the dashed, dotted, and dash-dot lines are the expected values of $\tilde{\Lambda}_n$ for $\varepsilon = 0$, 0.5, and 1.0, respectively. Since the noncentrality parameter of $n\|\bar{\mathbf{X}}_n - \boldsymbol{\mu}_0\|$ is a linearly increasing function of n, the

difference between $\tilde{\Lambda}_n$ and Λ_n increases as n increases when $\varepsilon = 0$. When $\varepsilon = 1.0$, the difference becomes very large. The varying and large differences can cause difficulty in setting up decision regions of sequential tests and result in significant test errors. Interestingly, when $\varepsilon = 0.5$, the expected values of $\tilde{\Lambda}_n$ provides similar patterns with those of Λ_n for both small and large p, and the differences between $\tilde{\Lambda}_n$ and Λ_n appear to be unvarying. In this section, we select $\varepsilon = 0.5$ such that the test statistics for testing a noncentrality parameter λ_1 is based on

$$\tilde{\lambda}_n = \sqrt{n}(\lambda_1 + 0.5)$$.

The proposed approximate sequential χ^2 test has three decision rules: (i) Accept H_0 when $\tilde{\Lambda}_n \leq 0$; (ii) Accept H_1 when $\tilde{\Lambda}_n > h$; and (iii) continue by observing $\mathbf{X}_{n+1}$ when $0 < \tilde{\Lambda}_n \leq h$, where $(0, h)$ is called a continuation region.

(a)

(b)
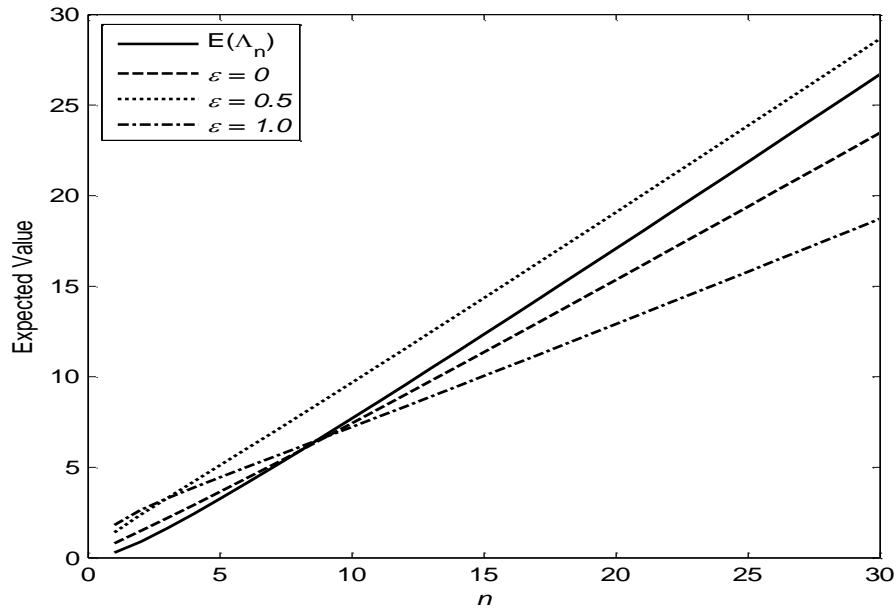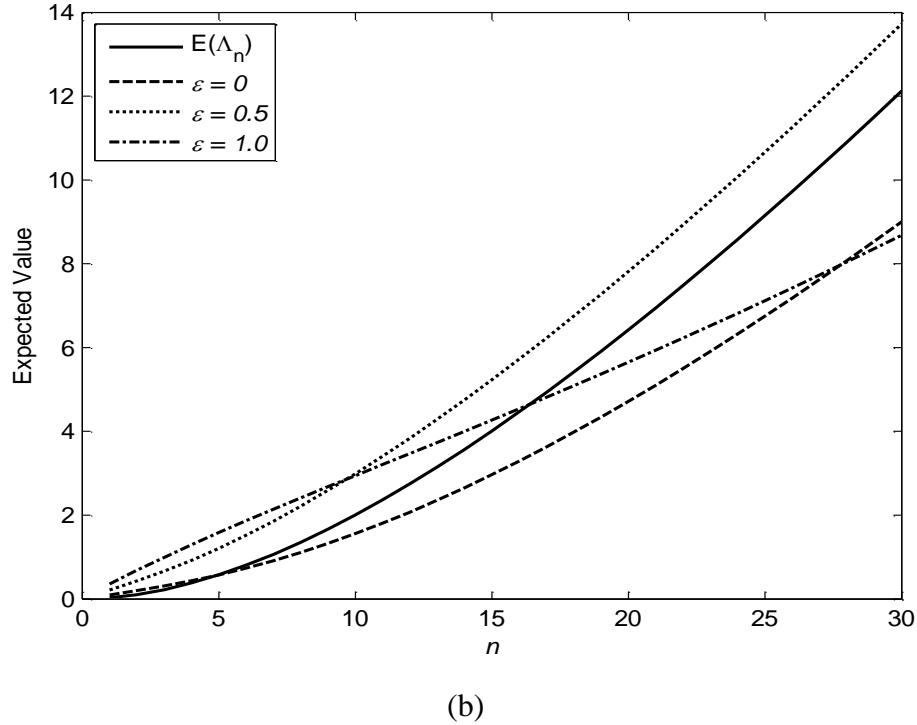
Figure 3.1 Expected values of Λ_n and $\tilde{\Lambda}_n$ with various ε when (a) $p = 2$ and (b) $p = 20$

3.3 MSPC using the approximate sequential χ^2 test

3.3.1 Multivariate approximate sequential chi-square (MASC) chart

As described in the previous section, we are interested in testing $H_0 : \lambda = 0$ versus

$H_1 : \lambda = \lambda_1$ based on the approximate sequential χ^2 test in the MSPC testing problem. In

sequential tests, we accept H_0 or H_1 based on the test statistic value. However, in a SPC

testing problem, we accept H_1 or restart a sequential test when the previously taken

decision is to accept H_0. When we accept H_1 for the first time, we stop observation and

do not restart a new cycle of the sequential test. For instance, the test statistic proposed

by Healy (1987) is reset to zero and the SPRT is restarted when the value of the test

statistic is less than zero. When the value of the test statistic is greater than a threshold

value (control limit), it alarms an OC signal by accepting H_1. Using this concept, we can integrate a sequential test into an MSPC chart.

To introduce the proposed MASC chart, we consider the cumulative moving average as

$$\mathbf{M}_{m_t,t} = \frac{1}{m_t}(\mathbf{X}_{t-m_t+1} + \ldots + \mathbf{X}_t), \tag{3.6}$$

where m_t is the number of observations starting from the recent sequential test. As shown in the approximate sequential χ^2 test based on the logarithm of the likelihood ratio, the test statistic of the proposed chart is defined as

$$MA_t = w_{m_t} \log\left(m_t \left\|\mathbf{M}_{m_t,t} - \boldsymbol{\mu}_0\right\|\right) - k_{m_t}, \tag{3.7}$$

where $k_{m_t} > 0$ is a reference value defined in equation (3.5). Since $\mathbf{M}_{m_t,t} - \boldsymbol{\mu}_0$ represents the difference between the accumulated sample average and the target mean, the norm $\left\|\mathbf{M}_{m_t,t} - \boldsymbol{\mu}_0\right\| = \left(\mathbf{M}_{m_t,t} - \boldsymbol{\mu}_0\right)' \boldsymbol{\Sigma}_0^{-1} \left(\mathbf{M}_{m_t,t} - \boldsymbol{\mu}_0\right)$ is considered a measure of the difference of our estimate of the mean of the process from the target mean. Based on the value of MA_t, the procedure of the proposed MASC chart is as follows:

- Stop and alarm an OC signal when $MA_t > h$;

- Update $m_{t+1} = m_t + 1$ when $0 < MA_t \leq h$;

- Update $m_{t+1} = 1$ when $MA_t \leq 0$,

where $h > 0$ is a control limit for an OC signal, respectively. As long as $MA_t < h$, we update the cumulative moving average as

$$\mathbf{M}_{m_{t+1},t+1} = \frac{m_{t+1}-1}{m_{t+1}}\mathbf{M}_{m_t,t} + \frac{1}{m_{t+1}}\mathbf{X}_{t+1}. \tag{3.8}$$

When $m_{t+1} = 1$, the previous moving average $\mathbf{M}_{m_t,t}$ is cancelled out so that $\mathbf{M}_{m_{t+1},t+1} = \mathbf{X}_{t+1}$. For simplicity, the value of MA_t defined in equation (3.7) can be set to zero whenever $MA_t \leq 0$. Hence the proposed MASC can be expressed as follows

$$MA_t = \max\left\{ w_{m_t} \log\left(m_t \left\| \mathbf{M}_{m_t,t} - \boldsymbol{\mu}_0 \right\| \right) - k_{m_t}, 0 \right\} \tag{3.9}$$

and

$$m_t = \begin{cases} m_{t-1} + 1 & \text{if } MA_{t-1} > 0 \\ 1 & \text{if } MA_{t-1} \leq 0 \end{cases}, \tag{3.10}$$

where $m_0 = 1$ and $MA_0 = 0$. The proposed MASC chart signals when $MA_t > h$.

Remarks: If we set $m_{t+1} = 1$ when $MA_t \leq h$ for all $t > 0$, then the moving average $\mathbf{M}_{m_t,t}$ is equal to $\mathbf{X}_t$. Since k_{m_t} and w_{m_t} are constants, by rescaling and transforming $\log(m_t \| \mathbf{M}_{m_t,t} - \boldsymbol{\mu}_0 \|)$, the test statistic can be equivalent to the test statistic of the Hotelling's χ^2 control chart. Appendix B describes the relationship between the popular MC1 chart (Pignatiello and Runger 1990) and the proposed chart in detail.

3.3.2 Design of parameters

The design of an MASC chart involves the choice of parameters h and λ_1. The basic design strategy is to select the shift size λ_1 to be detected quickly and choose h for satisfying the desired ARL_0. This approach produces MASC charts with good ARL performance when the size of the mean shift is close to λ_1. Using Tables 3.1 and 3.2, we can obtain the MASC parameters that result in the minimum OC ARL for mean shifts of the specified size λ_1. Table 3.1 provides the zero-state ARLs of the proposed MASC charts with different values of λ_1 when $p = 2, 5, 10$. The numerical results show that the

MASC charts with $\lambda_1 = 0.5$, 1.0, and 1.5 have the minimum ARLs in detecting shifts of

sizes $\lambda = 0.5$, 1.0, and 1.5, respectively. These results are similar to those obtained in

univariate CUSUM charts so that smaller and larger values of λ_1 are more sensitive in

detecting smaller and larger shifts, respectively. Table 3.2 provides h and the minimum

OC ARL (ARL_{min}) for detecting mean shifts of size $\lambda = 1.0$ with various $ARL_0 = 200$,

500, 700, and 1000. We choose $\lambda_1 = 1.0$ because this value provides good ARL

performance in detecting small shifts in the mean vector. All ARL values are obtained

using 20,000 simulations.

Table 3.1 Zero-state ARLs of MASC charts with different values of λ_1 when $p = 2, 5, 10$

	$\lambda_1 = 0.5$			$\lambda_1 = 1.0$			$\lambda_1 = 1.5$		
	$p = 2$	$p = 5$	$p = 10$	$p = 2$	$p = 5$	$p = 10$	$p = 2$	$p = 5$	$p = 10$
λ	$h = 6.3$	$h = 5.52$	$h = 4.81$	$h = 7.39$	$h = 6.67$	$h = 6.03$	$h = 8.05$	$h = 7.39$	$h = 6.81$
0.0	199.52	199.70	200.67	200.67	201.01	200.4	200.30	199.61	201.02
0.5	**28.27**	**36.45**	**44.44**	34.46	45.84	55.66	44.47	59.29	70.92
1.0	10.49	13.37	16.41	**9.76**	**12.46**	**15.23**	10.82	13.87	17.08
1.5	7.16	9.17	11.40	5.64	7.13	8.71	**5.24**	**6.55**	**8.01**
2.0	5.90	7.58	9.43	4.28	5.39	6.64	3.63	4.49	5.51

Table 3.2 Optimal parameters of MASC charts in detecting mean shifts of Size $\lambda=1.0$

with different ARL_0 when p = 2, 5, 10, 20

P		In-control ARL			
		200	500	700	1000
2	h	7.39	9.15	9.82	10.53
	ARL_{min}	9.76	11.75	12.50	13.29
5	h	6.67	8.42	9.07	9.82
	ARL_{min}	12.46	14.84	15.77	16.74
10	h	6.03	7.84	8.51	9.19
	ARL_{min}	15.23	18.34	19.43	20.49
20	h	5.37	7.15	7.80	8.53
	ARL_{min}	18.90	22.90	24.22	25.62

3.3.3 Properties of the proposed test statistic

The test statistic defined in equation (3.8) is based on the log transformation of

χ^2 distribution, since $m_t \parallel \mathbf{M}_{m_t,t} - \boldsymbol{\mu}_0 \parallel$ follows a χ^2 distribution with p degrees of

freedom when a process is in control. The $\log \chi^2$ distribution is considerably close to the

normal distribution for the following reasons: (i) it provides closer approximation to

normality based on the Kullback-Leibler information number, which is a measure of the

difference between two probability distributions (Hawkins and Wixley 1986); (ii) it can

be considered as truly normal, because the transformed variable of a positive-valued

variable from χ^2 distribution is defined over the whole range from $-\infty$ to ∞ (Keene

1995); and (iii) the log transformation is widely used for converting right-skewed

distributions with heavy right tails to be symmetric.

Suppose that Y is a random variable from $\log \chi^2$ distribution with p degrees of freedom. Then the pdf f_Y is given as

$$f_Y(y) = \frac{1}{2^{p/2}\Gamma(p/2)} e^{y(p/2-2)} e^{-\frac{e^y}{2}}.$$

The mean and variance of the $\log \chi^2$ distribution can be obtained using Taylor series approximation (Casella and Berger 2002), where the mean and variance of Y are $\log p$ and $2p^{-1}$, respectively. Interestingly, the variance decreases as p is increases. Figure 3.2 compares $\log \chi^2$ distributions when p=5, 10, 20. It is clear that a $\log \chi^2$ distribution with larger p is closer to normality and has smaller variance.
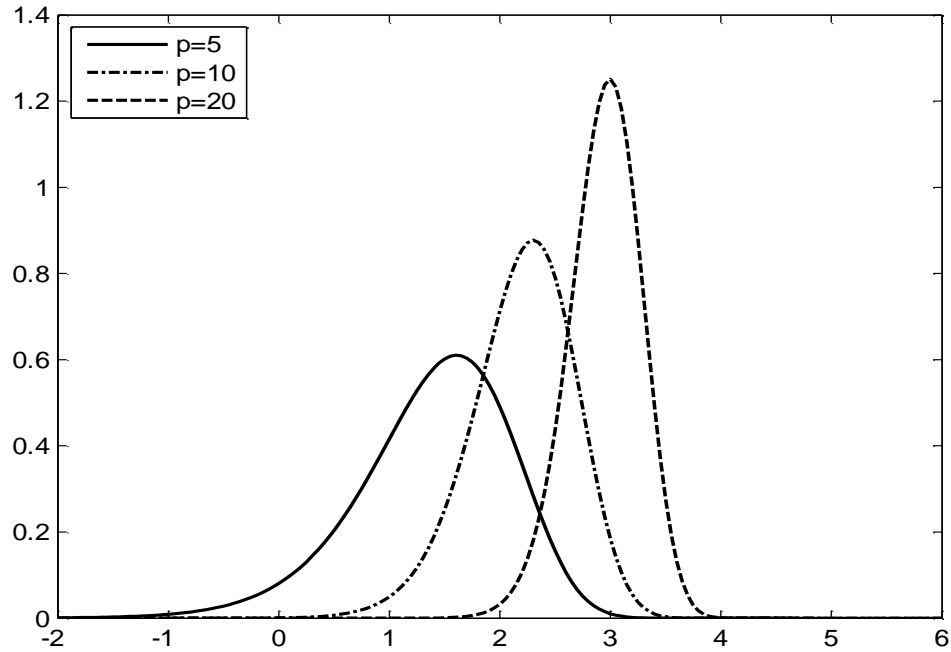
Figure 3.2 Probability density functions of $\log \chi^2$ distributions with p=5, 10, 20

Figure 3.3 compares the expected values of $w_{m_t} \log(m_t \| \mathbf{M}_{m_t,t} - \boldsymbol{\mu}_0 \|)$ and k_{m_t} when $p = 10$ and $\lambda_1 = 1.0$. The expected values of $w_{m_t} \log(m_t \| \mathbf{M}_{m_t,t} - \boldsymbol{\mu}_0 \|)$ under IC

and OC processes approximate to $w_{m_t} \log p$ and $w_{m_t} \log(p + m_t \lambda_1^2)$, respectively. It is

clear that the test statistic MA_t is often reset to zero since the expected value of MA_t,

which approximates to $w_{m_t} \log(p) - k_{m_t}$, becomes negative as m_t increases under an IC

process ($\lambda = 0$). However, when the process is out of control with $\lambda = 1.0$, a signaling

probability becomes larger since the expected value $w_{m_t} \log(p + m_t \lambda_1^2) - k_{m_t}$ increases as

m_t increases.
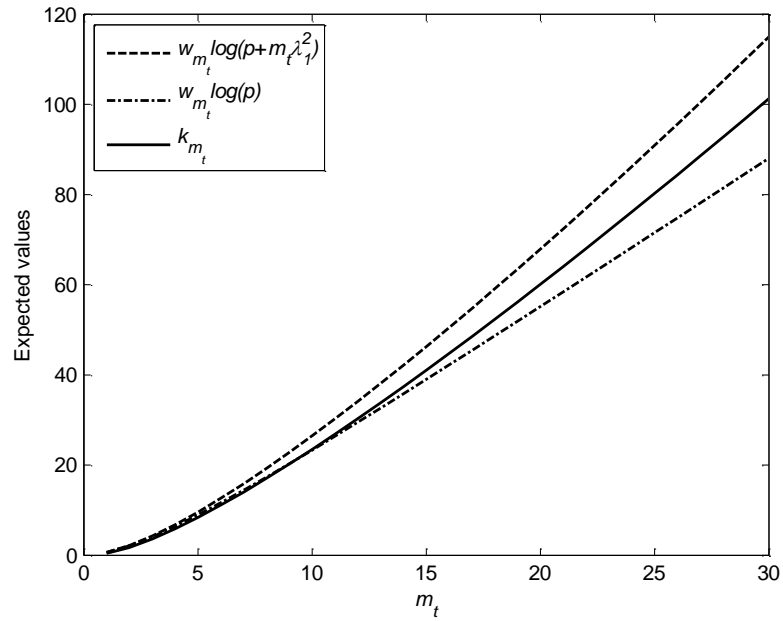
Figure 3.3 Expected values of $w_{m_t} \log\left(m_t || \mathrm{M}_{m_t,t} - \mu_0||\right)$ and k_{m_t} when $p = 10$ and

$$\lambda_1 = 1.0$$

3.3.4 An illustrative example

To illustrate the use of the MASC chart, we utilize a set of simulated observations

taken from Crosier (1988). The data set in Table 3.3 contains 10 observations from

bivariate normal distribution with a correlation coefficient of 0.5 and unit variances. The

process mean vector of the first 5 observations is [0, 0], while the mean vector of the last

5 observations is shifted to [1, 2]. The parameters of the MASC chart are chosen to be $\lambda_1 = 1.0$ and $h = 7.39$ with $ARL_0 = 200$.

Table 3.3 An example of an MASC chart using data from Crosier (1988)

	Observations			$\mathbf{M}_{m_t} = [M_1, M_2]$					
t	X_1	X_2	m_t	M_1	M_2	Y_t	w_{m_t}	k_{m_t}	MA_t
1	-1.19	0.59	1	-1.19	0.59	1.19	0.78	-0.55	1.47
2	0.12	0.90	2	-0.54	0.75	1.20	1.84	0.16	2.04
3	-1.69	0.40	3	-0.92	0.63	1.99	2.94	1.42	4.42
4	0.30	0.46	4	-0.62	0.59	1.76	4.05	3.04	4.06
5	0.89	-0.75	5	-0.31	0.32	0.70	5.17	4.94	0.00
6	0.82	0.98	1	0.82	0.98	0.10	0.78	-0.55	0.63
7	-0.30	2.28	2	0.26	1.63	1.81	1.84	0.16	3.18
8	0.63	1.75	3	0.38	1.67	2.22	2.94	1.42	5.10
9	1.56	1.58	4	0.68	1.65	2.40	4.05	3.04	6.66
10	1.46	3.05	5	0.83	1.93	2.93	5.17	4.94	10.19
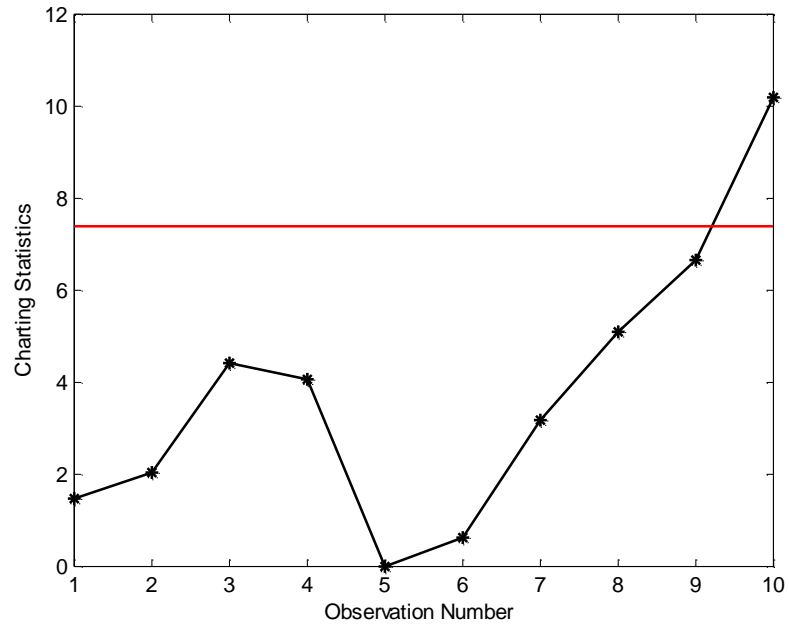
Figure 3.4 Plot of a MASC chart using data from Table 3.7

In Table 3.3, the fourth column contains the number of observations used for computing cumulative moving average $\mathbf{M}_{m_t,t}$, and the fifth and sixth columns show the values of $\mathbf{M}_{m_t,t} = [M_1, M_2]$. The seventh, eighth and ninth columns contain the values of $Y_t = \log(m_t \| \mathbf{M}_{m_t,t} - \boldsymbol{\mu}_0 \|)$, w_{m_t}, and k_{m_t} for computing MA_t, respectively. The final column shows the values of the test statistic MA_t. As shown in Figure 3.4, which plots the values of the MASC statistic, the proposed MASC chart gives an OC signal at the 10th observation. The Crosier's MCUSUM chart (1988) with $k = 0.5$ also signals at the 10th observation when $ARL_0 = 200$. Note that the MASC vector elements, M_1 and M_2, provide some clues to diagnose the direction of the shift when the MASC signals.

3.4 Adaptive MASC chart based on the estimated shift size

The MASC chart can be optimized when we have accurate information on the shift size λ. In practice, the magnitude of the mean shift is unknown. When a specified λ_1 is different from the magnitude λ of a real changed mean, the control charts based on the likelihood ratio methods can perform poorly. In the univariate case, adaptive schemes (Sparks 2000, Capizzi and Masarotto 2003, Shu and Jiang 2006, Jiang *et al*. 2008, Shu *et al*. 2008, Wu *et al*. 2009) have been widely used to overcome this problem. In this section, we introduce the adaptive version of the MASC chart.

The univariate ACUSUM chart proposed by Sparks (2000) adjusts the reference value of the conventional CUSUM chart using an estimate of the current mean shift size. When λ can be efficiently estimated at time *t*, say $\hat{\lambda}_t$, then we can obtain the reference value $k_{m_t}(\hat{\lambda}_t)$ at each time *t*. In this case, the AMASC statistic can be written as

$$AMS_t = \max\left\{ w_{m_t} \log\left(m_t \left\| \mathbf{M}_{m_t,t} - \boldsymbol{\mu}_0 \right\| \right) - k_{m_t}(\hat{\lambda}_t), 0 \right\}, \tag{3.11}$$

where $\hat{\lambda}_t^2$ is the estimator of the noncentrality parameter. It signals whenever $AMS_t > h$, where h is the control limit of the AMASC chart. For estimating a noncentrality parameter λ^2, we adopt the maximum likelihood estimator (MLE) in terms of the mean squared error (MSE) proposed by Saxena and Alam (1982) as

$$\psi_t = \max\left(0, \left\| \mathbf{x}_t - \boldsymbol{\mu}_0 \right\| - p \right). \tag{3.12}$$

where $\| \mathbf{x}_t - \boldsymbol{\mu}_0 \| = (\mathbf{x}_t - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1} (\mathbf{x}_t - \boldsymbol{\mu}_0)$. Analogous to the ACUSUM charts (Sparks 2000; Shu and Jiang 2006), the EWMA scheme, which is one of the most popular schemes in practice due to its simplicity, can be a good estimate of λ as $\hat{\lambda}_t = r\sqrt{\psi_t} + (1-r)\hat{\lambda}_{t-1}$, where r is an EWMA parameter with $0 \le r \le 1$. When $[\lambda_{\min}, \lambda_{\max}]$ is the range of shift sizes of interests, we define the EWMA statistic as

$$\hat{\lambda}_t = \min\left(\max\left(\lambda_{\min}, r\psi_t + (1-r)\hat{\lambda}_{t-1} \right), \lambda_{\max} \right), \tag{3.13}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are minimum and a maximum values for improving the sensitivity to shift sizes $\lambda_{\min} \le \lambda \le \lambda_{\max}$.

3.4.1 Parameter selections

The design of an AMASC chart involves the determination of the parameters $\lambda_{\min}$, $\lambda_{\max}$, r and h. In this section, we provide guidelines for the parameters of AMASC charts. Suppose $[a, b]$ is the range of potential mean shift sizes to be detected. Based on

Sparks (2000) and Shu and Jiang (2006), we recommend the following guidelines for designing AMASC charts.

- Set $[\lambda_{\min}, \lambda_{\max}] = [a, b]$.

- Choose r in the range [0.05, 0.2] (Shu and Jiang 2006). In general, $r = 0.1$ and $r = 0.2$ result in good performance of the chart.

- Reset $\hat{\lambda}_{t+1} = \hat{\lambda}_0$ when $AMS_t \leq 0$. When $AMS_t \leq 0$, the current process is very likely to be in control so that it is reasonable to reset the process variables to their initial values for the following observations.

- Choose $\hat{\lambda}_0 = (\lambda_{\min} + \lambda_{\max})/2$ to balance the efficiency in detecting shifts over the range.

- Select h to achieve the desired ARL_0.

Note that the above parameter selections are not intended for optimizing detection performance for any particular range of shifts. Next, we show that these guidelines can provide a reasonably good detection performance over the specified range $[\lambda_{\min}, \lambda_{\max}]$. The ARL performance of the proposed AMASC chart is slightly less than that of the MASC chart when the size of a mean shift is around λ_1, otherwise, the AMASC chart considerably outperforms the MASC.

Table 3.4 Zero-state and steady-state ARLs of an AMASC chart with different regions

when $r = 0.2$ and $\lambda_1 = 1.0$

λ	p = 2		p = 5		p = 10		p = 20	
	ZARL	SARL	ZARL	SARL	ZARL	SARL	ZARL	SARL
	$[\lambda_{min}, \lambda_{max}] = [\,0.5,\,3.0]$							
0.0	199.01	196.01	200.90	198.13	200.25	195.26	201.79	197.34
0.5	34.32	34.31	49.09	48.83	64.22	64.02	82.06	81.59
1.0	9.86	10.10	12.65	13.24	16.08	16.91	21.11	21.99
1.5	5.17	5.54	6.47	7.07	7.88	8.65	9.76	10.71
2.0	3.40	3.77	4.19	4.74	5.08	5.76	6.33	7.15
2.5	2.61	2.96	3.12	3.61	3.75	4.32	4.65	5.32
3.0	2.15	2.46	2.53	2.99	2.99	3.50	3.70	4.29
4.0	1.69	1.93	2.00	2.31	2.23	2.64	2.65	3.13
5.0	1.29	1.57	1.72	1.96	1.98	2.26	2.15	2.54
	$[\lambda_{min}, \lambda_{max}] = [\,0.5,\,4.0]$							
0.0	200.92	198.85	200.87	198.06	201.95	198.36	201.89	197.64
0.5	37.38	37.25	52.20	52.06	67.33	66.87	84.86	84.30
1.0	10.14	10.40	13.19	13.64	16.59	17.19	21.76	22.57
1.5	5.18	5.46	6.50	7.03	7.90	8.58	9.81	10.72
2.0	3.36	3.68	4.11	4.62	4.96	5.62	6.21	7.01
2.5	2.52	2.82	3.02	3.49	3.60	4.17	4.48	5.17
3.0	2.07	2.37	2.44	2.87	2.85	3.36	3.51	4.13
4.0	1.52	1.78	1.89	2.20	2.14	2.54	2.50	2.99
5.0	1.17	1.44	1.51	1.80	1.87	2.16	2.08	2.48
	$[\lambda_{min}, \lambda_{max}] = [\,0.5,\,5.0]$							
0.0	200.61	198.58	200.36	198.19	200.65	197.95	199.92	196.15
0.5	40.41	40.07	56.36	55.99	71.94	71.81	87.57	86.87
1.0	10.61	10.74	13.83	14.14	17.44	17.92	22.56	23.15
1.5	5.27	5.47	6.60	7.03	8.04	8.64	9.92	10.73
2.0	3.34	3.57	4.09	4.52	4.92	5.53	6.13	6.92
2.5	2.46	2.70	2.95	3.36	3.51	4.04	4.34	5.02
3.0	1.97	2.22	2.35	2.74	2.75	3.24	3.36	3.99
4.0	1.40	1.64	1.76	2.06	2.06	2.44	2.39	2.88
5.0	1.10	1.35	1.35	1.66	1.72	2.03	2.02	2.40

ZARL is zero-state ARL; SARL is steady-state ARL.

Table 3.4 compares the zero-state and steady-state ARLs of an AMASC chart for different shift range $[\lambda_{\min}, \ \lambda_{\max}]$ when $r = 0.2$ and $ARL_0 = 200$. When the range is [0.5, 3.0], control limits for $p = 2, 5, 10, 20$ are $h = 8.25, 7.41, 6.76, 6.13$, respectively. With the range [0.5, 4.0] and [0.5, 5.0], $h = 8.79, 7.93, 7.19, 6.47$ and $h = 9.2, 8.4, 7.67, 6.86$ for $p = 2, 5, 10, 20$, respectively. These comparisons show that the AMASC chart with larger $\lambda_{\max}$ is more effective in detecting shifts with large λ, but insensitive to small λ.

3.5 Performance comparisons

All control charts compared in this section are directional invariant. The ARL performance of a directionally invariant control chart depends only on the distance, not on the particular direction of the mean vector (Pignatiello and Runger 1990). Shifts in the process mean are based on the distance λ from the IC mean $\mathbf{\mu}_0 = \mathbf{0}$ which are of the form $\mathbf{\mu} = (\delta, 0, \ldots, 0)$. For simplicity, we assume that the covariance matrix is the identity matrix. In each independent simulation, the run length is recorded as the number of simulations observed until the control chart issues an OC signal. The estimated ARL is the average of the independent run lengths. We use $ARL_0 = 200$ and repeat 20,000 simulations to obtain ARLs. For simulating steady-state ARLs, the true shift location τ is fixed at 51, and any simulation in which shift signals occur before τ is discarded.

3.5.1 Comparisons of MASC charts with MCUSUM and MC1 charts

We now compare both zero-state and steady-state ARL performance of the proposed MASC chart with those of two multivariate CUSUM charts: Crosier's MCUSUM (1988) and MC1 (Pignatiello and Runger 1990) charts. All charts are optimally designed to detect shifts of size $\lambda = 1.0$, and these charts also provide good

ARL performance in detecting small shifts in the mean vector. Table 3.5 shows ARL comparisons for $p = 2$, 10, 20, and 50, respectively. Control limits of the MCUSUM charts are $h = 5.5$, 14.9, 24.7, 49.95 and those of the MC1 are $h = 4.77$, 9.55, 14.58, 28.72, for $p = 2$, 10, 20, 50, respectively.

It is clear from Table 3.5 that the MC1 chart provides the smallest zero-state ARL value for all p, but has poor steady-state ARL performance when p is large due to an inertia problem, which can refer to the resistance of a chart in signaling a process change (Yashchin 1987). The MC1 chart can build up a considerably large amount of inertia when the chart has run some time before a mean shift occurs (Woodall and Mahmoud 2005), which results in a lengthy delay in detecting the mean shift. For example, if a number of positive deviations from the target value in one variable are accumulated before a negative mean shift occurs, there is a delay in the signal. The inertia from the positive deviations needs to be canceled by sufficient negative deviations to either restart the chart or generate a negative sum that eventually grows beyond the control limit.

It is observed that the Crosier's MCUSUM chart provides relatively larger zero-state ARLs than the other charts, but smaller steady-state ARLs than the MC1 chart. One drawback of Crosier's MCUSUM chart is that the steady-state ARL_0 is considerably smaller than the zero-state when p is large. Thus, the MCUSUM chart in a steady-state process can cause more false alarms than that in a zero-state process. To maintain the steady-state ARL_0 to a specific level, the MCUSUM control limit must be increased and the increased control limit results in larger ARLs.

Table 3.5 ARL comparisons between MASC, MCUSUM and MC1 charts ($\lambda_1 = 1.0$)

	MASC		MCUSUM		MC1	
λ	ZARL	SARL	ZARL	SARL	ZARL	SARL
			$p = 2$			
0.0	200.67	197.49	203.52	197.22	197.74	193.99
0.5	34.46	33.97	29.97	28.62	31.39	31.34
1.0	9.76	9.81	9.87	9.41	9.31	9.78
1.5	5.64	5.68	5.79	5.54	5.26	5.70
2.0	4.28	4.24	4.14	3.98	3.70	4.08
2.5	3.59	3.50	3.25	3.14	2.89	3.24
3.0	3.21	3.05	2.70	2.62	2.41	2.72
			$p = 10$			
0.0	199.90	193.88	199.33	176.05	199.59	202.77
0.5	55.66	54.26	42.71	36.09	43.52	51.72
1.0	15.23	**15.28**	18.59	15.59	12.53	17.90
1.5	8.71	**8.77**	11.92	9.92	7.68	10.98
2.0	6.64	**6.47**	8.80	7.34	5.67	7.97
2.5	5.64	**5.29**	7.02	5.86	4.56	6.31
3.0	5.08	**4.58**	5.87	4.92	3.84	5.28
			$p = 20$			
0.0	202.03	194.17	198.62	160.56	200.89	212.36
0.5	66.96	64.58	56.06	43.61	47.71	74.68
1.0	18.90	**18.99**	27.17	20.79	15.26	29.26
1.5	11.03	**11.00**	17.97	13.60	10.05	17.72
2.0	8.48	**8.11**	13.49	10.23	7.64	12.76
2.5	7.24	**6.67**	10.79	8.20	6.21	10.04
3.0	6.47	**5.73**	9.04	6.83	5.26	8.31
			$p = 50$			
0.0	199.37	186.59	199.05	149.17	199.64	225.76
0.5	81.01	76.75	86.11	63.80	48.04	120.38
1.0	25.66	**25.17**	47.56	34.88	23.28	29.37
1.5	**15.36**	**14.93**	32.72	23.89	16.59	17.81
2.0	**12.00**	**11.23**	24.95	18.19	13.00	13.28
2.5	**10.23**	**9.16**	20.21	14.73	10.75	10.69
3.0	**9.15**	**7.86**	16.99	12.38	9.18	8.93

* ZARL is zero-state ARL; SARL is steady-state ARL.

Contrary to the MCUSUM and the MC1 charts, which have considerably different values between zero-state and steady-state ARLs when p is large, the proposed MASC chart provides similar zero-state and steady-state ARLs in detecting small shifts in the mean vector for all p. In addition, the MASC chart has the smallest steady-state ARLs when p is larger than 10 and $\lambda \geq 1.0$. Therefore the proposed MASC chart is most effective in detecting process mean shifts both when the process is initially out of control and when it is initially in control and a shift occurs later.

3.5.2 Comparisons of AMASC charts with MASC charts

In Figure 3.5, the zero-state ARLs of AMASC charts are compared with MASC charts when $p = 2$ and 10. In order to detect shifts between 0.5 and 3.0 efficiently, we set $\lambda_{min} = 0.5$ and $\lambda_{max} = 3.0$ and select $\lambda = 0.2$. The parameters $h = 8.25$ and 6.76 of the AMASC are selected to satisfy the zero-state $ARL_0 = 200$ when $p = 2$ and 10, respectively. The target shift sizes for MASC charts are chosen as $\lambda_1 = \lambda_{min}$ and $\lambda_1 = \lambda_{max}$, respectively. The two MASC charts have nearly the minimum ARL at shift sizes λ_{min} and λ_{max}. Figure 3.5 shows overall ARL values of the three charts. It is observed that the zero-state ARL curves of the AMASC charts are close to the bottom in wide range when $p = 2$ and 10.
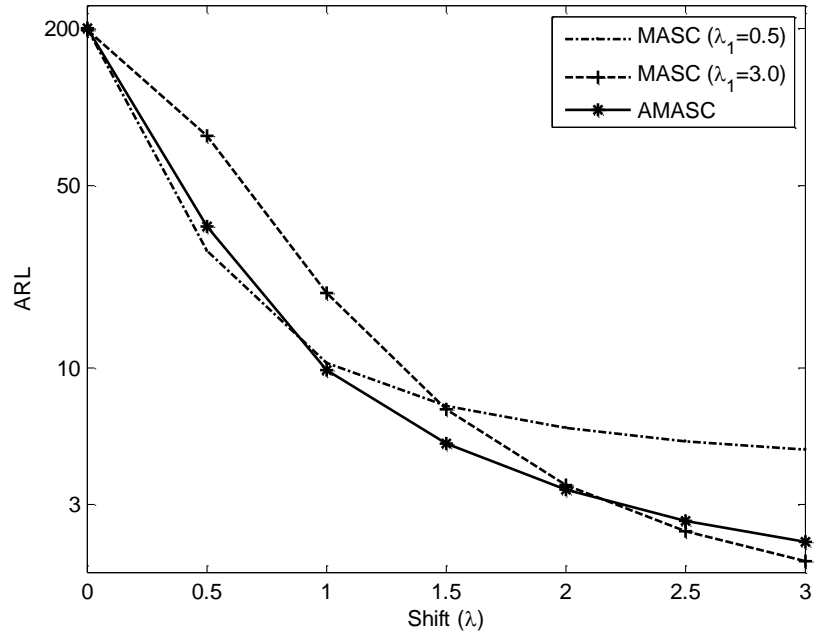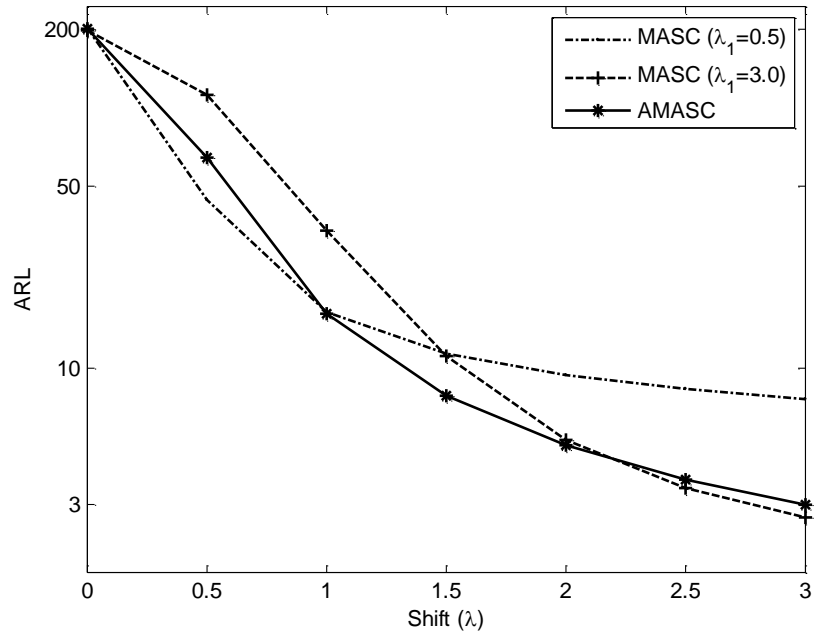
Figure 3.5 Zero-state ARL comparisons between MASC and AMASC charts for

detecting shifts within the range [0.5 3.0] when (a) $p = 2$ and (b) $p = 10$

3.5.3 Comparisons of AMASC charts with AMCUSUM charts

For detecting a wide range of mean shifts efficiently, Dai *et al*. (2011) propose an adaptive MCUSUM (AMCUSUM) chart based on Crosier's MCUSUM chart. Figures 3.6 and 3.7 compare the zero-state and steady-state ARL values of the AMASC charts and the AMCUSUM charts for detecting shifts of sizes within the range [0.5, 6.0]. Both charts use the same EWMA parameter $r = 0.2$. The zero-state ARL_0 for both charts is maintained at 200. The control limits of AMCUSUM are $h = 1.083$ and $h = 1.117$ when $p= 5$ and 10, respectively. The control limits of AMASC charts are $h = 8.8$ and 8.1 when $p = 5$ and 10, respectively. As shown in Figures 3.6 and 3.7, AMASC is more sensitive to small shifts of sizes $\lambda \geq 1.0$ but less sensitive to shift sizes $\lambda < 1.0$.

The steady-state comparisons shown in Figure 3.6 (b) and 3.7 (b) indicate that the steady-state ARL_0 of AMCUSUM charts are significantly smaller than 200, similar to Crosier's MCUSUM chart. Specifically, when $p=10$, the steady-state ARL_0 of the AMCUSUM chart is 160, while that of the AMASC chart is 199. Thus, to make the steady-state ARL_0 close to 200, the control limits of AMCUSM charts must be increased.

(a) (b)
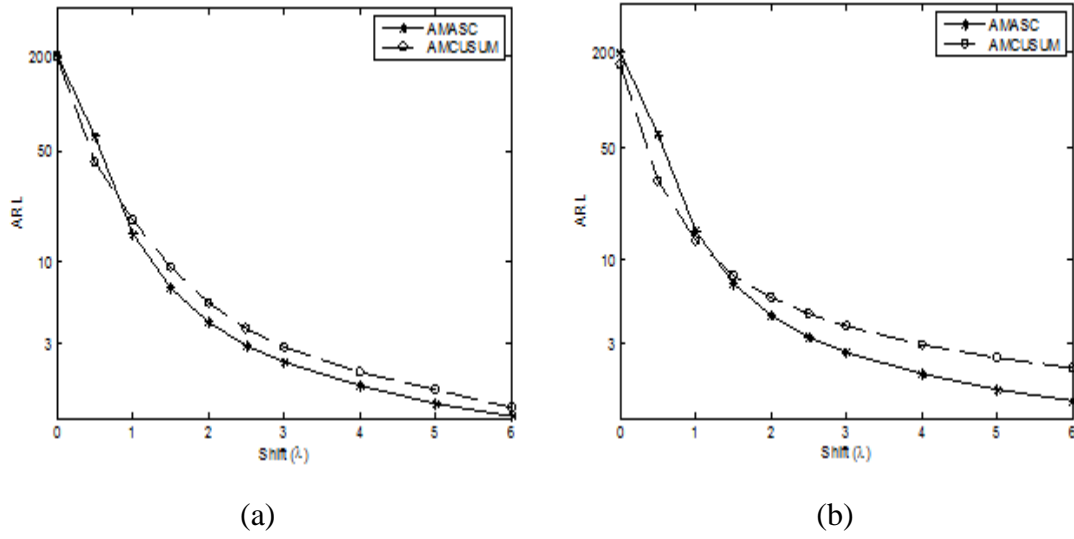
Figure 3.6 (a) Zero-state and (b) Steady-state ARL comparisons of AMCUSUM and

AMASC charts when $p = 5$

(a) (b)
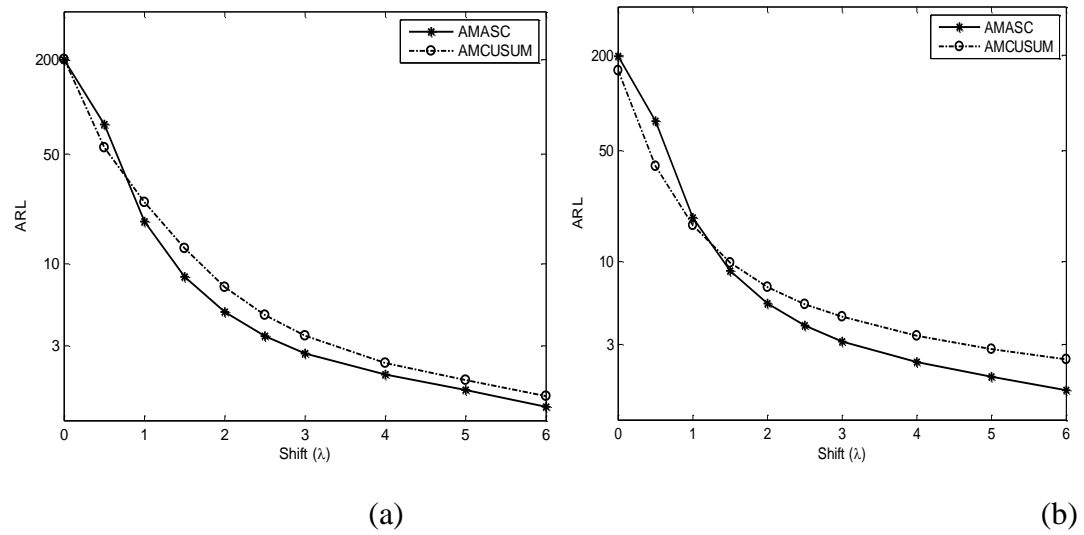
Figure 3.7 (a) Zero-state and (b) Steady-state ARL comparisons of AMCUSUM and

AMASC charts when $p = 10$

3.6 Conclusion

In this chapter, we present two new MSPC charts, MASC and AMASC, for detecting general mean shift vectors based on the approximate sequential χ^2 test, which uses an approximate likelihood ratio of a central and a noncentral χ^2 distribution. Because of the properties of the sequential χ^2 test, the proposed MASC chart provides good zero-state and steady-state ARL performance for detecting shifts in the mean vector that have a specific shift size λ_1 regardless of the dimension of measurements p, unlike the MCUSUM (Crosier 1988) and the MC1 (Pignatiello and Runger 1990), which have considerably different values between zero-state and steady-state ARL when p is large. Due to this robustness property of the MASC chart, it can quickly detect process mean shifts both when the process is initially out of control and when it is initially in control but shift occurs later.

We also propose an AMASC chart, which is an adaptive version of an MASC chart for detecting general mean shifts. The concept of the AMASC chart is to adaptively adjust the reference value of the proposed MASC chart by estimating the noncentrality parameter of the current process mean vector. The experimental results show that the proposed AMASC chart is efficient in detecting a wide range of mean shifts compared with the MASC chart and the AMCUSUM (Dai, *et al.* 2011), which is an adaptive version of Crosier's MCUSUM chart.

When a shift in the mean vector is detected, we may need to identify the variables that cause an out-of-control signal. In future work, we intend to explore diagnostic procedures to identify the out-of-control variables in both MASC and AMASC charts.

CHAPTER 4

FAULT VARIABLE IDENTIFICATION USING ADAPTIVE STEP-DOWN PROCEDURE

4.1 Introduction

MSPC has received considerable attention for monitoring multiple quality characteristics and/or process parameters. The primary objectives of MSPC are to detect a change in the process mean vector and to identify which variables are responsible for the change. MSPC charts alarm an OC signal when a shift is detected, but have difficulty identifying the variables which cause the OC signal. A variable whose mean is shifted is defined as a faulty variable.

Identifying the cause of an OC signal is a challenging problem for quality engineers in high-dimensional processes when an MSPC control chart detects a shifted process mean. Since control charts based on T^2 statistics have difficulty identifying variables that cause an OC signal, a variety of diagnostic procedures have been developed (Alt 1985, Doganaksoy *et al.* 1991, Hawkins 1991, 1993, Mason *et al.* 1995, Sullivan *et al.* 2007, Li *et al.* 2008). One popular approach for fault diagnosis is based on testing each individual variable. It identifies faulty variables that are significant (Alt 1985, Doganaksoy *et al.* 1991). A main drawback of this approach is that it ignores correlations among variables. In a highly-correlated structure, values of test statistics can be large,

even though there is no change in the mean vector. Another approach is based on testing every possible subset of variables (Murphy 1987, Chua and Montgomery 1992, Sullivan *et al.* 2007). Although this approach incorporates the correlation information among variables, it may be impractical due to intensive computations in a high-dimensional process.

Hawkins (1991, 1993) proposes a monitoring and diagnostic procedure under the assumption that only single variable is shifted in the mean vector. The procedure is based on regression-adjusted variables using the correlations among variables. Hawkins' regression-adjusted approach is effective in detecting and identifying a shift of single variable in the mean vector. When the maximum of the absolute values of regression-adjusted variables is significant, it signals and identifies the variable associated with the maximum as a changed variable. However, when the number of changed variables is unknown, one may select all significant regression-adjusted variables, which may result in poor identification performance when the means of several variables are simultaneously shifted or even when the mean of single variable that is highly correlated with other variables is shifted (Das and Prakash 2008).

Mason *et al*. (1995, 1997) propose a decomposition procedure based on all possible partitioning of T^2 statistic into independent unconditional and conditional T^2 terms defined in equation (4.2). This approach is referred to MYT decomposition (Mason *et al.* 1997). While the MYT decomposition is theoretically sound, it may not be practical when the number of variables, p, is large, since it needs to examine $p!$ decompositions. To reduce the number of computations in the MYT decomposition, Mason *et al.* (1997) propose a shortened sequential procedure. In a worst-case scenario,

however, the procedure still requires the same computations with the original MYT approach. Furthermore, the MYT approach has a concern about diagnostic capability (Li *et al*. 2008). Assuming the mean shifts of variables and no changes in the variable relationships, it is not clear which variable or a subset of variables is responsible for a significant conditional term. Since it examines all possible decompositions, an extremely large number of terms of all decompositions can be significant and all related variables can be identified as being changed. In this case, these terms and variables need to be evaluated and interpreted by process engineers. Mason *et al.* (1995, 1997) suggest using unconditional T^2 statistics for the identification of mean shifts of individual variables. However, their suggested method ignores correlation structures among variables, so it is equivalent with the procedures based on testing individual variables (Alt 1985, Doganaksoy *et al.* 1991).

A main issue based on MYT decomposition is to find a meaningful decomposition containing information on identifying which variable or a subset of variables are responsible for the process mean shift among different p! decompositions. Recently, Li *et al*. (2008) propose a causation-based decomposition by integrating the traditional MYT decomposition with a Bayesian causal network that defines the causal relationship between variables. Tan and Shi (2012) propose a Bayesian approach based on Bayesian hierarchical model to determine the shifted means and the directions of the shifts. Based on some prior knowledge and experience to specific processes that have known causal or Bayesian hierarchical property, one can investigate a smaller number of decompositions than the MYT approach. While these approaches are effective in some

process control situations, they are often not proper in certain processes that do not have known causal or Bayesian hierarchical property.

In high-dimensional processes, it is reasonable to assume that shifts in the mean vector occur in only a few variables, which is called the sparsity property (Zou and Qiu 2009). Wang and Jiang (2009) and Zou and Qiu (2009) propose process monitoring and diagnosis schemes based on variable selection methods and the sparsity assumption. Although both schemes provide diagnosis capability, they basically focus on the monitoring task.

In this chapter, we propose an adaptive step-down procedure for identifying variables whose means are shifted, under the assumption that a shift in the mean vector occurs in only a few variables and a multivariate SPC chart based on chi-square statistic like Hotelling's T^2 chart signals after detecting the shift. The proposed procedure selects a variable that has the strongest evidence of no mean change at each step. The variable selection is based on the variables that are selected in previous steps, where the previously selected variables have strong evidence of no change. The proposed procedure first searches for the group of variables that are not changed with strong evidence, and then identifies the variables that are responsible for the OC signal based on conditional T^2 statistics given the selected (or unchanged) variables. Our approach adopts a projection scheme (Runger 1996) and constructs conditional T^2 statistics. The proposed procedure yields a less computational complexity in a high-dimensional process, since it is based on the polynomial time algorithm. Thus it can be an effective diagnostic tool for the real time faulty variable identification in a high-dimensional process.

4.2 Conditional T^2 statistics with known group of unchanged variables

Assume that a process has p quality characteristics and the measurement, $\mathbf{X} = (X_1, \ldots, X_P)$ follows a multivariate normal distribution, $N_P(\mathbf{\mu}, \mathbf{\Sigma})$. When the process is in-control (IC), the mean vector is $\mathbf{\mu} = \mathbf{\mu}_0$, where $\mathbf{\mu}_0 = (\mu_1, \ldots, \mu_p)$, and the covariance matrix $\mathbf{\Sigma} = \mathbf{\Sigma}_0$, where $\mathbf{\Sigma}_0 = [\sigma_{ij}]_{1 \le i, j \le p}$, is known and fixed over time. When the process is out-of-control, the process mean vector is changed to $\mathbf{\mu}_1 \ne \mathbf{\mu}_0$. The Hotelling's T^2 statistic is decomposed to identify the fault variables when the control charts generate an out-of-control signal, and it is defined as

$$T^2 = (\mathbf{X} - \mathbf{\mu}_0)' \mathbf{\Sigma}_0^{-1} (\mathbf{X} - \mathbf{\mu}_0), \tag{4.1}$$

where T^2 statistic follows a χ^2 distribution with p degrees of freedom (df) when the process is in control. With a given false alarm rate α_0, it signals if $T^2 > \chi^2_{p,\alpha_0}$. Although T^2 is the optimal test statistic for a general multivariate shift of mean vector, it is not optimal when some variables are known to be unchanged in high-dimensional processes (Lowry and Montgomery 1995).

Suppose that $\mathbf{X} = (\mathbf{X}_\Omega, \mathbf{X}_\Gamma)$, where Ω and Γ are index sets of two partitions of $\mathbf{X}$. Then the mean vector and covariance matrix can be partitioned as $\mathbf{\mu_0} = (\mathbf{\mu_{0\Omega}}, \mathbf{\mu_{0\Gamma}})$ and

$$\mathbf{\Sigma_0} = \begin{bmatrix} \mathbf{\Sigma_{\Omega\Omega}} & \mathbf{\Sigma_{\Omega\Gamma}} \\ \mathbf{\Sigma_{\Gamma\Omega}} & \mathbf{\Sigma_{\Gamma\Gamma}} \end{bmatrix},$$

respectively. Without loss of generality, a shift of the mean vector occurs only in a subset of variables Ω and there are no changes in all variables of Γ, Runger (1996) propose a projection chart based on the conditional T^2 given Γ defined as

$$T^2_{\Omega|\Gamma} = T^2 - T^2_\Gamma, \tag{4.2}$$

where $T_\Gamma^2 = (\mathbf{X}_\Gamma - \boldsymbol{\mu}_{0\Gamma})'\boldsymbol{\Sigma}_{\Gamma\Gamma}^{-1}(\mathbf{X}_\Gamma - \boldsymbol{\mu}_{0\Gamma})$ is the T^2 of all variables in Γ. Under the null hypothesis, $T_{\Omega|\Gamma}^2$ follows the chi-square distribution with $|\Omega|$ df. It is shown that using $T_{\Omega|\Gamma}^2$ statistic is more powerful than using overall T^2 or T_Ω^2, which is the T^2 of all variables in Ω (Runger 1996). However, the $T_{\Omega|\Gamma}^2$ statistic also do not provide which variable or a group of variables in Ω are changed.

After a mean shift is detected by control charts based on Runger's $T_{\Omega|\Gamma}^2$ statistics, we propose a new conditional T^2-based diagnosis to specify the shifted variables by choosing regression-adjusted variables, which are regressed on the variables in Γ, with significantly different from the target value. For a variable i in Ω, the test statistic is based on the square of a regression adjusted variable $T_{i\cdot\Gamma}$, which is adjusted by the mean and standard deviation of the conditional distribution of X_i given $\mathbf{X}_\Gamma$ as

$$T_{i\cdot\Gamma} = \frac{(X_i - \mu_i) - \boldsymbol{\beta}_{i\cdot\Gamma}'(\mathbf{X}_\Gamma - \boldsymbol{\mu}_{0\Gamma})}{\left(\sigma_{ii} - \boldsymbol{\beta}_{i\cdot\Gamma}'\boldsymbol{\Sigma}_{\Omega\Gamma}^{(i)}\right)^{1/2}}, \tag{4.3}$$

where $\boldsymbol{\beta}_{i\cdot\Gamma}$ is a column vector of the regression coefficients of X_i on X_j for all $j \in \Gamma$, which are obtained as $\boldsymbol{\beta}_{i\cdot\Gamma} = \boldsymbol{\Sigma}_{\Gamma\Gamma}^{-1}\boldsymbol{\Sigma}_{\Omega\Gamma}^{(i)}$, where $\boldsymbol{\Sigma}_{\Omega\Gamma}^{(i)}$ is the ith column vector of $\boldsymbol{\Sigma}_{\Omega\Gamma}$ (Anderson (1984)). As shown in Mason et al. (1995), $T_{\Gamma\cup\{i\}}^2 = T_\Gamma^2 + T_{i\cdot\Gamma}^2$, so that we can simply obtain the square of $T_{i\cdot\Gamma}$ as

$$T_{i\cdot\Gamma}^2 = T_{\Gamma\cup\{i\}}^2 - T_\Gamma^2, \tag{4.4}$$

where $T_{\Gamma\cup\{i\}}^2$ is the T^2 of all variables in $\Gamma \cup \{i\}$. The distribution of $T_{i\cdot\Gamma}^2$ follows a chi-square distribution with one degree of freedom. With a given significant level α_1, we can choose a control limit χ_{1,α_1}^2 for $T_{i\cdot\Gamma}^2$ and determine shifted variables in Ω if $T_{i\cdot\Gamma}^2 > \chi_{1,\alpha_1}^2$ for all $i \in \Omega$.

Proposition 1.   Suppose that Γ is a subset of unchanged variables and Γ is known in advance. Then for $i \in \Omega$,

$$\Pr\left(T^2_{i\cdot\Gamma} > \chi^2_{1,\alpha_1}\right) \geq Pr\left(T^2_i > \chi^2_{1,\alpha_1}\right). \tag{4.5}$$

When $E(X_i) = \mu_0$, both $T^2_{i\cdot\Gamma}$ and T^2_i follow a central chi-square distribution with one degree of freedom. In this case, $\Pr\left(T^2_{i\cdot\Gamma} > \chi^2_{1,\alpha_1}\right) = Pr\left(T^2_i > \chi^2_{1,\alpha_1}\right)$. When $E(X_i) = \mu_0 + \delta_i$, where $\delta_i \neq 0$, the distributions of $T^2_{i\cdot\Gamma}$ and T^2_i depend on the noncentrality parameters, λ_1 and λ_2, of chi-square distributions, respectively. Since $\lambda_1 \geq \lambda_2$, $\Pr\left(T^2_{i\cdot\Gamma} > \chi^2_{1,\alpha_1}\right) \geq Pr\left(T^2_i > \chi^2_{1,\alpha_1}\right)$
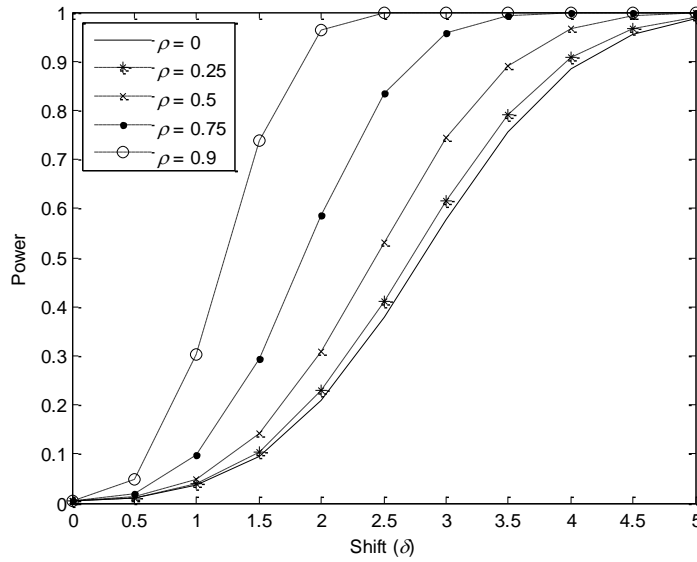
Figure 4.1 Power functions of $T^2_{1\cdot2}$ with various ρ when $X_1 \sim N(\delta_1, 1)$ and $X_2 \sim N(0,1)$

For instance, when $p = 2$, the correlation coefficient of two variables is $= \boldsymbol{\beta}'_{i\cdot\Gamma}\boldsymbol{\Sigma}^{(i)}_{S\Gamma}$. For simplicity, we also assume that $\boldsymbol{\mu_0} = \mathbf{0}$, and $\sigma_{ii} = 1$ for $i = 1, 2$. Without loss of generality, we assume that $\Omega = \{1\}$ and $\Gamma = \{2\}$, then

$$T_{1\cdot2} = \frac{X_1 - \rho X_2}{\sqrt{1 - \rho^2}}.$$

Since $\sqrt{1 - \rho^2} < 1$, if X_2 is close to target value, then $\Pr(\, T_{1\cdot 2}^2 \geq T_1^2 | X_2 = x_2) \rightarrow$ 1 as $x_2 \rightarrow 0$. Thus $T_{1\cdot 2}^2$ can be significant, although T_1^2 is not significant. The power functions of a shift δ can be defined as

$$\beta(\delta) = P(T_{1\cdot 2}^2 > \chi_{1,\alpha_1}^2),$$

where α_1 is a significant level. The function $\beta(\delta)$ is continuous in δ, and $\beta(\delta) = \alpha_1$, when $\delta = 0$. When $\rho = 0$, the distribution of $T_{1\cdot 2}^2$ is equivalent with that of T_1^2. The slope of power functions with larger ρ is steeper than those with smaller ρ.

As a special case, when $S = \{j\}$ and $\Gamma = \{1, \ldots, j-1, j+1, \ldots, p\}$, the conditional $T_{j\cdot\Gamma}^2$ statistic is closely related to Hawkins' regression-adjusted variable. Hawkins (1991, 1993) shows that the test static $T_{j\cdot\Gamma}^2$ is based on the optimal test statistic when a shift occurs only in the jth component of **X**. However, the conditional statistic is often far from optimal when $|S| > 1$.

4.3 Adaptive step-down procedure

In practice, the subset of variables that are not affected by assignable causes is often unknown. In this case, the step-down procedure (Sullivan *et al.* (2007)) and the MYT decomposition (Mason *et al.* (1995, 1997)) have been shown to be effective in the interpretation of the OC signal. However, interpretation efforts based on these approaches may require numerous computations. From the point of view of computational complexity, they depend on exponential-time algorithms, and this fact might discourage quality practitioners.

In this section, we propose an adaptive step-down procedure based on conditional T^2 statistics on a group of selected variables, $\hat{\Gamma}$. In each step, the proposed adaptive

procedure selects a variable with strong evidence of no change given previously selected variables. The variable at ith step is selected as

$$\gamma_i = \text{argmin}_{j \notin \hat{\Gamma}} \, T^2_{j \cdot \hat{\Gamma}}, \tag{4.6}$$

where $\hat{\Gamma} = \{\gamma_1, \gamma_2, \dots, \gamma_{i-1}\}$ is a group of variables selected by the previous step. When $i = 1$, the set $\hat{\Gamma}$ is empty. In this case, we set $T^2_{j \cdot \hat{\Gamma}} = T^2_j$, where T^2_j is the unconditional T^2 of an individual variable.

The procedure has two decision rules. The first rule is to detect fault variables. If the conditional $T^2_{\gamma_i \cdot \hat{\Gamma}}$ is larger than a threshold value χ^2_{1,α_1}, where α_1 are a significance level for fault variables, then the conditional $T^2_{j \cdot \hat{\Gamma}}$ for all $j \notin \hat{\Gamma}$ exceeds χ^2_{1,α_1}. In this case, we can conclude all variables not in $\hat{\Gamma}$ are fault variables.

The second decision rule is to keep from selecting a fault variable and adding it into $\hat{\Gamma}$. If $T^2_{\hat{\Gamma} \cup \{\gamma_i\}} > \chi^2_{i,\alpha_2}$, where α_2 is a significance level for the group of unchanged variables and $T^2_{\hat{\Gamma} \cup \{\gamma_i\}}$ follows a chi-square distribution with i df, then the selected variable at the ith step can be a fault variable. To provide early notification if a fault variable is selected, we suggest choosing α_2 with relatively larger value than the Type I error of the control chart used for mean shift detection. In this case, the procedure stops and identifies which variables not in $\hat{\Gamma}$ are fault variables. As commented in the previous section, the conditional $T^2_{j \cdot \hat{\Gamma}}$ for all $j \notin \hat{\Gamma}$ can be powerful test statistics for identifying which variables not in $\hat{\Gamma}$ are fault variables. When both $T^2_{\gamma_i \cdot \hat{\Gamma}} < \chi^2_{1,\alpha_1}$ and $T^2_{\hat{\Gamma} \cup \{\gamma_i\}} < \chi^2_{i,\alpha_2}$ are satisfied, then adds γ_i into $\hat{\Gamma}$ and move to the next step. In this section, we set $\alpha_1 = \alpha_2$ for simplicity.

4.3.1 Initial variable selection

The proposed procedure begins with selecting a variable with statistically strongest evidence of no change. Although different schemes can be used for the initial variable selection, it seems reasonable to select a variable based on correlation information between variables as

$$\gamma_1 = \text{argmin}_{j=1,\dots,p} \, T^2_{j \cdot 1,\dots,j-1,j+1,\dots,p}, \tag{4.7}$$

where $T^2_{j \cdot 1,\dots,j-1,j+1,\dots,p}$ is the conditional T^2 value of jth variable on the remaining $p-1$ variables, $\{1,\dots,j-1,j+1,\dots,p\}$. The conditional value can be easily obtained as $T^2_{j \cdot 1,\dots,j-1,j+1,\dots,p} = T^2 - T^2_{\{1,\dots,j-1,j+1,\dots,p\}}$, where $T^2_{\{1,\dots,j-1,j+1,\dots,p\}}$ is the T^2 of the remaining $p-1$ variables. When the mean of each variable is on-target, $T^2_{j \cdot 1,\dots,j-1,j+1,\dots,p}$ follows a χ^2_1 distribution with one df. When $T^2_{\gamma_1} > \chi^2_{1,\alpha_1}$, we stop and conclude that all variables whose $T^2_j > \chi^2_{1,\alpha_1}$ for $j \in \{1,\dots,p\}$ are fault variables.

4.3.2 Design of parameters

The design of an adaptive step-down procedure involves the choice of parameters α_1 and α_2. The significance level α_1 is to test whether the mean of a variable is changed and α_2 is to maintain $\hat{\Gamma}$ having no fault variables. The choices of α_1 and α_2 are considered in Table 4.1. This compare the identification performance with various α_1 and α_2 when $p = 10$. Let $\boldsymbol{\mu}_0 = 0$ and $\boldsymbol{\Sigma}_0 = [\sigma_{ij}]_{1 \leq i,j \leq p}$, where $\sigma_{ii} = 1$ and $\sigma_{ij} = \rho$. Diagnostic analysis is executed 10,000 times whenever the Hotelling's T^2 chart with $ARL_0 = 200$ detects a shift of the mean vector. The shifted mean vector is defined as $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \boldsymbol{\delta}$, where $\boldsymbol{\delta} = (\delta_1,\dots,\delta_p)$. When the mean of the jth variable is shifted, $\delta_j \neq 0$, while $\delta_j = 0$ when the mean is not changed. The performance measure is the relative frequency

that the proposed procedure identifies fault variables exactly, which means no fault variables are missed and all fault variables are identified.

It is expected that the corresponding variable with large δ_i can be easily identified as a fault variable even with smaller α_1, but may not falsely identify unchanged variables with high probability. So smaller α_1 may provide better correct identification rate when single variable is changed. However, when the means of multiple variables are changed, smaller α_1 can often miss fault variables, while larger α_1 can identify unchanged variables as fault with higher probability.

The basic strategy of selecting α_2, which is a significance level for the group of unchanged variables, is to choose $\alpha_2 \geq \alpha_1$ as large as the detection performance of the adaptive step-down procedure is not affected, since larger α_2 can stop the procedure earlier. However, α_2 is too large, the procedure can stop at the first step and then use p unconditional T^2 for identification. Suppose that the procedure is at ith step after selecting γ_i with $T^2_{\gamma_i \cdot \hat{\Gamma}} \leq \chi^2_{1,\alpha_1}$. If the selected one is a fault variable and $\alpha_1 > \alpha_2$, then a fault variable γ_i can be added into $\hat{\Gamma}$ since $T^2_{\hat{\Gamma} \cup \{\gamma_i\}} = T^2_{\hat{\Gamma}} + T^2_{\gamma_i \cdot \hat{\Gamma}} < \chi^2_{i,\alpha_2}$. However, when $\alpha_1 \leq \alpha_2$, it can be $T^2_{\hat{\Gamma} \cup \{\gamma_i\}} > \chi^2_{i,\alpha_2}$ even $T^2_{\gamma_i \cdot \hat{\Gamma}} \leq \chi^2_{1,\alpha_1}$. In this case, instead adding the variable γ_i into $\hat{\Gamma}$, it stops the procedure and tests using conditional T^2 given $\hat{\Gamma}$, a group of variables having no significant evidence of a change. Table 4.1 shows that the procedure with $\alpha_2 = 0.05$ provides slightly bad performance than those with $\alpha_2 = 0.01$ and 0.005, while the procedures with $\alpha_2 = 0.01$ and 0.005 provides the same results. Hence we suggest using $\alpha_2 = \alpha_1$ for the group of unchanged variables.

Table 4.1 Relative frequencies identifying fault variables exactly with different α_1 and

α_2 when $\rho = 0.75$, 0.5, and 0.25

α_2	$\alpha_1 = 0.05$			$\alpha_1 = 0.01$			$\alpha_1 = 0.005$		
	0.05	0.01	0.005	0.05	0.01	0.005	0.05	0.01	0.005
Shifts				$\rho = 0.75$					
$\delta_1 = 1$	0.21	0.21	0.21	0.54	0.54	0.54	0.61	0.61	0.61
$\delta_1 = 2$	0.50	0.51	0.51	0.84	0.85	0.85	0.91	0.91	0.91
$\delta_1 = 3$	0.63	0.63	0.63	0.91	0.91	0.91	0.95	0.95	0.95
$\delta_1 = 1, \delta_2 = 1$	0.27	0.27	0.27	0.32	0.32	0.32	0.25	0.25	0.25
$\delta_1 = 2, \delta_2 = 2$	0.62	0.62	0.62	0.78	0.78	0.78	0.77	0.77	0.77
$\delta_1 = 3, \delta_2 = 3$	0.70	0.70	0.70	0.92	0.92	0.92	0.96	0.96	0.96
				$\rho = 0.50$					
$\delta_1 = 1$	0.11	0.11	0.11	0.36	0.37	0.36	0.43	0.42	0.42
$\delta_1 = 2$	0.33	0.33	0.33	0.70	0.71	0.71	0.80	0.80	0.80
$\delta_1 = 3$	0.52	0.53	0.53	0.86	0.86	0.86	0.92	0.92	0.92
$\delta_1 = 1, \delta_2 = 1$	0.14	0.14	0.14	0.17	0.17	0.17	0.12	0.12	0.12
$\delta_1 = 2, \delta_2 = 2$	0.46	0.46	0.46	0.55	0.55	0.55	0.47	0.47	0.47
$\delta_1 = 3, \delta_2 = 3$	0.64	0.64	0.64	0.84	0.84	0.84	0.83	0.83	0.83
				$\rho = 0.25$					
$\delta_1 = 1$	0.08	0.08	0.08	0.29	0.29	0.29	0.34	0.34	0.34
$\delta_1 = 2$	0.24	0.25	0.25	0.62	0.63	0.63	0.71	0.71	0.71
$\delta_1 = 3$	0.42	0.42	0.42	0.81	0.81	0.81	0.88	0.88	0.88
$\delta_1 = 1, \delta_2 = 1$	0.10	0.10	0.10	0.10	0.10	0.10	0.07	0.07	0.07
$\delta_1 = 2, \delta_2 = 2$	0.35	0.35	0.35	0.42	0.42	0.42	0.33	0.33	0.33
$\delta_1 = 3, \delta_2 = 3$	0.57	0.57	0.57	0.70	0.70	0.70	0.64	0.64	0.64

4.3.3 Implementation of the adaptive step-down procedure

Practitioners who wish to apply the proposed procedure can follow the steps below. The decision threshold values χ^2_{1,α_1} and χ^2_{i,α_2} for $i = 1, \dots, p$ can be obtained prior from a χ^2 distribution with i degree of freedoms with pre-specified significant level α_1 and α_2.

Let $S = \{1, \dots, p\}$, and set $i = 1$ and $\hat{\Gamma} = \emptyset$.

(1) Compute $\gamma_1 = \text{argmin}_{j \in S} T^2_{j \cdot 1, \dots, j-1, j+1, \dots, p}$.

(2) Repeat step a - d until $T^2_{\hat{\Gamma} \cup \{\gamma_i\}} > \chi^2_{i,\alpha_2}$.

    a. Set $\hat{\Gamma} = \hat{\Gamma} \cup \{\gamma_i\}$.

    b. Increase i by 1, $i = i + 1$.

    c. Compute $\gamma_i = \text{argmin}_{j \notin \hat{\Gamma}} T^2_{j \cdot \hat{\Gamma}}$.

    d. Stop repetition and move to (3) if $T^2_{\gamma_i \cdot \hat{\Gamma}} > \chi^2_{1,\alpha_1}$.

(3) Stop and declare variables whose $T^2_{j \cdot \hat{\Gamma}} > \chi^2_{1,\alpha_1}$ for all $j \notin \hat{\Gamma}$ as fault variables.

The proposed procedure reduces the computation complexity dramatically compared with the approximation procedure of the MTY decomposition (1997). The procedure does not investigate all decompositions to alleviate fault variable identification. In worst case, the number of test statistics is $O(p^2) = p(p + 1)/2$. When all variables are independent, the proposed test statistics are equivalent with individual T^2 test statistics (Doganaksoy *et al.* 1991).

4.3.4 Relationship with MYT decomposition

Suppose that the proposed procedure stops at $(p\text{-}1)$th step. Then the proposed procedure computes all conditional terms of one MYT decomposition (Mason *et al.* 1995) as

$$T^2 = T_{\gamma_1}^2 + T_{\gamma_2 \cdot \gamma_1}^2 + \cdots + T_{\gamma_p \cdot \gamma_1, \ldots, \gamma_{p-1}}^2. \tag{4.8}$$

Note that the diagnostic procedure starts after Hotelling's T^2 chart signals since the value of T^2 statistic is significantly large. In this case, since a group of variables $\{\gamma_1, \gamma_2, \ldots, \gamma_{p-1}\}$ has no evidence of a change, the final variable γ_p is a fault variable with high probability.

4.4. An example

We illustrate our proposed procedure using dataset of switch drums from Flury and Riedwyl (1988) and Hawkins (1991). The dataset contains five variables $(p = 5)$: X_1 is the inside diameter of the drum, and X_2, X_3, X_4, X_5 are distances from the head to the edges of four sectors cut in the drum. The target mean of $\mathbf{X}$ is $\boldsymbol{\mu_0} = (17.96, 10.30, 11.08, 8.26)$ and the standard deviation is $(1.8622, 1.7053, 1.7090, 1.8718, 2.2114)$. The covariance matrix of standardized variables is defined as

$$\boldsymbol{\Sigma_0} = \begin{pmatrix} 1 & 0.1388 & 0.3496 & 0.0829 & 0.2652 \\ 0.1388 & 1 & 0.7324 & 0.9130 & 0.6932 \\ 0.3496 & 0.7324 & 1 & 0.6824 & 0.8214 \\ 0.0829 & 0.9130 & 0.6824 & 1 & 0.7640 \\ 0.2652 & 0.6932 & 0.8214 & 0.7640 & 1 \end{pmatrix}$$

A sequence of 50 observations is presented in the dataset. After first 35 observations, the marginal standard deviation of X_1 is increased by 0.5 and an upward shift of 0.25 standard deviation is given to X_5. All other correlations and means are left unchanged. At

48th observation, $\mathbf{X}$ = (13.065, 11.625, 14.923, 12.589, 12.446), Hotelling's T^2 control chart signals as $T^2 = 18.19 > \chi^2_{5,\alpha}$ = 15.09, where $\alpha_1 = 0.01$ significnat level. In step 1, the 3th variable has provides strongest evidence of no change, where $T^2_{3\cdot1,2,4,5}$ = 0.80. Then $\gamma_1 = 3$ so that $\hat{\Gamma} = \{3\}$ in step 1. In step 2, the conditional T^2 values given X_3 are evaluated, and select $\gamma_2 = 2$ since $T^2_{2\cdot3}$ has the smallest value among $\{T^2_{2\cdot3}, T^2_{4\cdot3}, T^2_{5\cdot3}\}$ so that $\hat{\Gamma} = \{2, 3\}$. Similarly, we select $\gamma_3 = 4$ so that $\hat{\Gamma} = \{2, 3, 4\}$ in step 3. After step 3, the step-down procedure stops since both $T^2_{1\cdot2,3,4}$ and $T^2_{5\cdot2,3,4}$ are significantly large, and identify both X_1 and X_5 are fault variables. The identification of X_1 and X_5 as the variables contributing significantly to the signal is similar to the conclusions reached by Hawkins (1991). Mason *et al.* (1995) show that 31 terms are significant, and X_1, X_4, X_5 are identified as changed variables when the shortened sequential procedure (Mason *et al.* 1997) are used. When we test only unconditional T^2 for individual variables, X_1 is identified as changed variable.

Table 4.2 Conditional and unconditional T^2 values of all steps in the proposed procedure

Step	Test Statistics				
1	$T^2_{1\cdot2,3,4,5} = 10.77$	$T^2_{2\cdot1,3,4,5} = 1.44$	$\mathbf{T^2_{3\cdot1,2,4,5}}= 0.80$	$T^2_{4\cdot1,2,3,5} = 2.98$	$T^2_{5\cdot1,2,3,4} = 8.30$
2	$T^2_{1\cdot3} = 9.36$	$\mathbf{T^2_{2\cdot3}} = 0.17$	$T^2_{4\cdot3} = 0.22$	$T^2_{5\cdot3} = 5.47$	
3	$T^2_{1\cdot2,3} = 9.22$	$\mathbf{T^2_{4\cdot2,3}} = 0.05$	$T^2_{5\cdot2,3} = 5.32$		
4	$T^2_{1\cdot2,3,4} = 9.20$	$T^2_{5\cdot2,3,4} = 6.73$			

4.3 Performance comparisons

We now compare the proposed diagnostic procedure with the existing procedures such as step-down procedure (Sullivan *et al.* (2007)) and LASSO-based procedure (ZOU et al. (2011)). In this study, only mean shifts are presented to the variables and there are no changes in the variable relationships, that is, correlations between variables. For performance comparisons, the relative frequency that the diagnostic procedures identify fault variables exactly (CR) is used to evaluate the performances of a diagnostic procedure.

The diagnostic procedures are executed 10,000 times after Hotelling's T^2 chart with $ARL_0 = 200$ signals. Significance level for step-down procedures is set to 0.05, where Sullivan *et al.* (2007) use the value about optimal. For compositions, the significance level is set to 0.005 for the proposed adaptive step-down (ASD) procedure.

In this simulations, we set $\mathbf{\mu}_0 = \mathbf{0}$ and $\mathbf{\Sigma}_0 = \left[\sigma_{ij}\right]_{1 \le i,j \le p}$, where $\sigma_{ii} = 1$ and $\sigma_{ij} = \rho$, for simplicity. The shifted mean vector is $\mathbf{\mu}_1 = \mathbf{\mu}_0 + \mathbf{\delta}$, where $\mathbf{\delta} = \left(\delta_1, \dots, \delta_p\right)$. For performance comparisons, relative frequencies are considered to evaluate the performances of diagnostic procedures. A diagnostic procedure performs better if relative frequency is larger. We assume that a Hotelling's SPC detects a shift at time T, and then the diagnosis procedures start.

4.3.1 Performance comparisons using only the last observation responsible for the OC signal

Figure 4.2 compares relative frequencies of the adaptive step-down (ASD) and MYT procedures using only the last point $\mathbf{X}_T$ that generate an OC signal. Two fault

scenarios are used for comparisons: (a) $\Omega = \{5\}$ and (b) $\Omega = \{1,5\}$. Because we assume only mean shifts, unconditional T^2 statistics for individual variables in the MYT procedure are evaluated for identifying mean shifts as suggested by Mason *et al.* (1995, 1997). For both procedures, we use 0.005 as the Type I error probability. When shift size is small, performances of both procedures are poor in Figure 4.2 (a) with single fault variable. However, the CR of ASD increases fast with the shift size increased by 5.0. In most situations, the ASD procedure outperforms the MYT procedure by a large difference, especially when shift sizes are between 1.0 and 4.0.

4.3.2 Performance comparisons using OC observations based on estimated change point

As shown in Zou *et al.* (2011), the performance of LASSO-based approach strongly depends on the number of OC observations. The LASSO-based procedures (Zou and Qui (2009) and Zou *et al.* (2011)) estimates the shift location first, and then do diagnostic procedures using the mean of all observations after that the location. For fair comparisons, we follow the change point estimation based on the MLE-based multivariate change point estimator (Zamba and Hawkins (2006)). The location of change point τ is defined as

$$\hat{\tau} = \text{argmin}_{0 \leq t < T}(T - t)(\overline{\mathbf{X}}_t - \boldsymbol{\mu_0})\boldsymbol{\Sigma}_0^{-1}(\overline{\mathbf{X}}_t - \boldsymbol{\mu_0}),$$

where T is the time that a SPC chart detects a shift and $\overline{\mathbf{X}}_t = \frac{1}{T-t}\sum_{i=t+1}^{T}\mathbf{X}_i$. In this comparison, the mean $\overline{\mathbf{X}}_{\hat{\tau}}$ is used for fault identification, instead of $\mathbf{X}_T$.
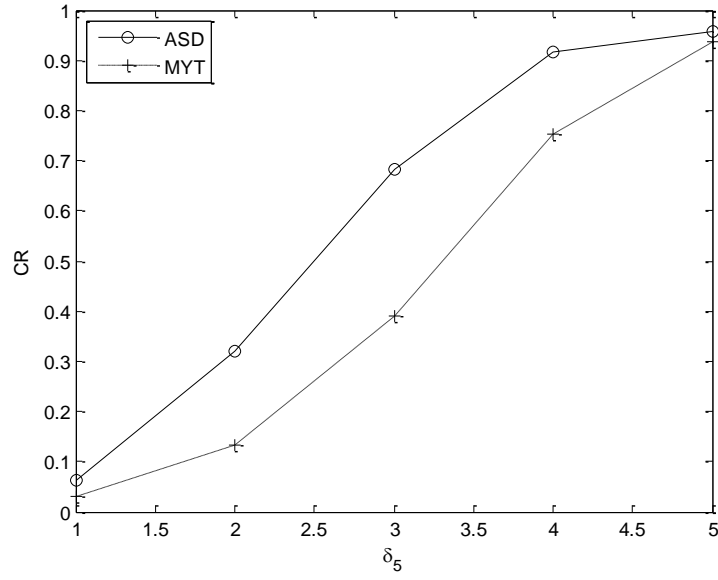
(a)
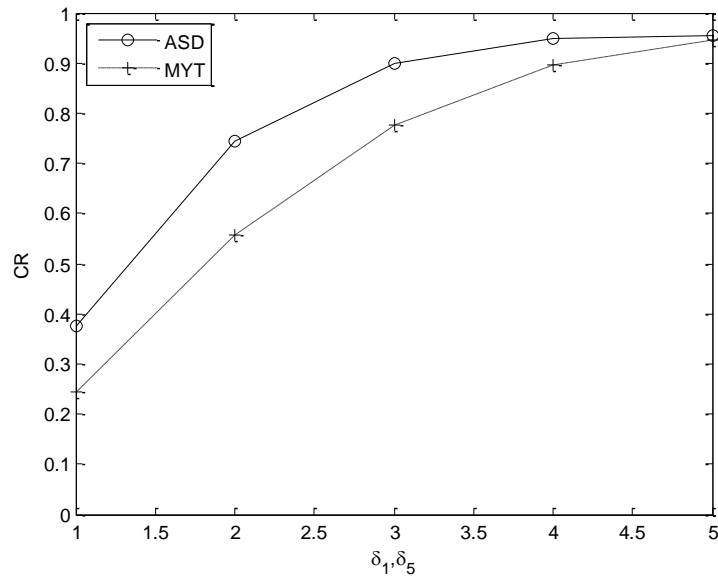
(b)

Figure 4.2 Performance comparisons with various fault variables when only the last

observation is used with (a) $\Omega = \{5\}$ and (b) $\Omega = \{1,5\}$

Figure 4.3 compare relative frequencies of the adaptive step-down (ASD), the

step-down, and LASSO procedures with various shift sizes when the mean of the 5th

variable is changed. Compared with Table 4.2, the simulation results show that relative frequencies of both procedures are increasing when shift sizes are less than 4.0. When sift size is large, the SPC detects the shift immediately, so that the last observation can be used for diagnosis in most case. Because of estimation error, the performances of the ASD in (a) and (b) of Figure 4.3 are slightly worse than those of Figure 4.2 when sift size is large. As expected, the performance of LASSO procedure is improved when the shift is detected later.

For performance comparisons, we also compare the expected error rates in mean shift decisions (EER) defined as

$$\text{EER} = E\left(\frac{\text{Number of Error}}{\text{Number of Variables}}\right) = E\left(\sum_{i=1}^{p} |I_{\{i \in \Omega\}} - I_{\{i \in \hat{\Omega}\}}|\right)/p \,,$$

where number of errors is the number of missed fault variables and falsely identified variables. When a diagnostic procedure identifies fault variables exactly, $CR = 1$ and $EER = 0$, while $CR=0$ and $EER = 1$ when all identified fault variables are false and all fault variables are missed. Thus a diagnostic procedure performs better if its value in column "CR" is larger and its value in column "EER" is smaller. The results in Table 4.3 and 4.4 show that the proposed approach has comparable diagnostic ability to other procedures.
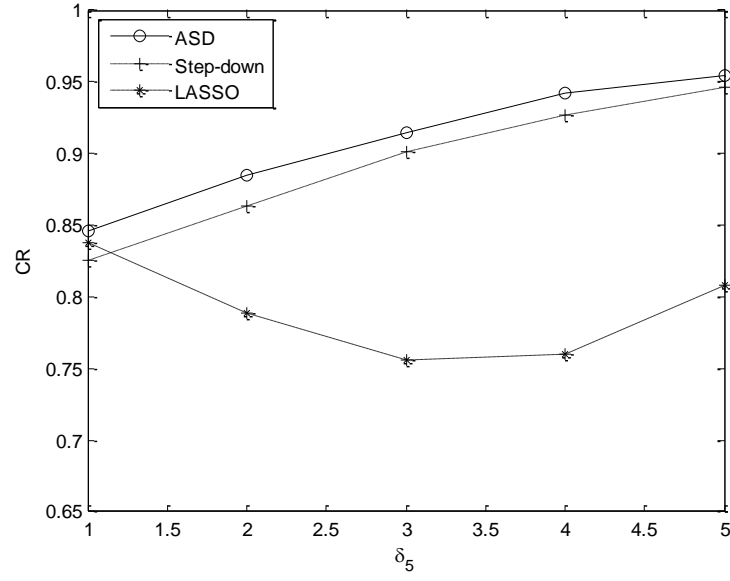
(a)
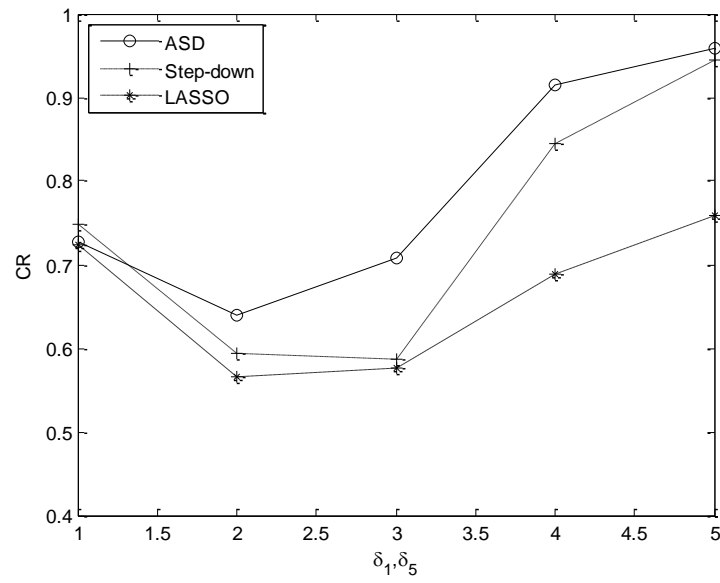
(b)

Figure 4.3 Performance comparisons of ASD, step-down, and LASSO procedures with

(a) $\Omega = \{5\}$ and (b) $\Omega = \{1,5\}$

Table 4.3 shows the effects of the locations and the number of fault variables. Overall performance decreases as the number of fault variables increases. The performance of step-down and ASD procedure are more improved than LASSO when the locations of fault variables are closer (more correlated) in both 2 and 3 fault variables.

Table 4.3 Performance comparisons of ASD, step-down, and LASSO procedures with various location of shifts when $\delta_i = 3.0$, for $i \in \Omega$

	Step-down		LASSO		ASD	
Ω	CR	EER	CR	EER	CR	EER
{1}	0.881	0.017 (0.052)	0.760	0.035 (0.074)	0.906	0.013 (0.050)
{5}	0.901	0.014 (0.048)	0.756	0.035 (0.074)	0.915	0.012 (0.046)
{1,10}	0.581	0.051 (0.070)	0.577	0.062 (0.087)	0.697	0.038 (0.066)
{5,6}	0.794	0.032 (0.077)	0.611	0.059 (0.089)	0.839	0.026 (0.077)
{1,5,6}	0.410	0.093 (0.099)	0.463	0.087 (0.101)	0.508	0.079 (0.098)
{4,5,6}	0.570	0.076 (0.113)	0.452	0.091 (0.105)	0.590	0.097 (0.158)

Table 4.4 Performance comparisons of ASD, step-down, and LASSO procedures with $\rho = 0.25, 0.5, 0.75$

	Step-down		LASSO		ASD	
ρ	CR	EER	CR	EER	CR	EER
0.25	0.294	0.100 (0.087)	0.434	0.087 (0.095)	0.372	0.098 (0.094)
0.50	0.410	0.093 (0.099)	0.463	0.087 (0.101)	0.508	0.079 (0.098)
0.75	0.774	0.038 (0.090)	0.534	0.082 (0.118)	0.854	0.029 (0.089)

Table 4.4 shows the effects of ρ when $\delta_i = 3.0$, for $i \in \Omega = \{1,5,6\}$ and $p = 10$. As shown in Table 4.4, the proposed procedures outperform the LASSO procedures by a large margin when $\rho = 0.75$. However, when ρ is small, CRs of LASSO procedure are better than those of the ASD and step-down procedures. This is because the penalty functions of LASSO do not use correlation information. By simulations, ρ is increasing, relative frequencies of ASD tend to superior to those of step-down and LASSO procedures.

Note: Because the step-down procedure requires 2^{p-1} computations, it takes significant computation time when p is large. Average computation times and standard deviations of the step-down and the proposed adaptive step-down procedures for single diagnosis with $p = 10$, 15, 20 and 25 are shown in Table 4.5. When $p = 25$, the average execution time of the step-down procedure is about 44 minutes, while that of the proposed procedure is about 0.02 seconds. Therefore the step-down procedure is not practical for automatic processes that adjust fault variables automatically after identification when p is large. The experiments were executed on Window 7 (64 bits) desktop with 8GB RARM and 2.10 GHz Dual-Core CPU.

Table 4.5 Average computation time (standard deviations) in seconds for single diagnosis

p	Step-down	Adaptive step-down
10	0.0810 (0.0379)	0.0035 (0.0012)
15	2.1800 (0.0459)	0.0068 (0.0013)
20	75.2758 (0.6705)	0.0112 (0.0013)
25	2643.5394 (4.6083)	0.0180 (0.0011)

4.4. Conclusions

When a shift in the mean vector is detected, identifying which variable or a group of variables causes an out-of-control signal is very crucial. Conventional diagnosis approaches such as step-down and MYT decomposition are theoretically sound for diagnosing root-causes of the process change, but computationally impractical for a large number of variables.

In this chapter we develop an adaptive step-down procedure using conditional T^2 statistics for fault variable identification. By selecting a variable having no significant evidence of a change based on the variables that are selected in previous steps, we can construct single decomposition among p! MYT decompositions.

The proposed procedure provides reasonable computational complexity in high-dimensional processes and enhances diagnostic power in identifying the shifted components of the mean vector when a shift occurs only in a few variables. As shown in simulations, the proposed procedure outperforms the MYT and Step-down procedures. Moreover, the proposed procedure is superior to the LASSO-based procedure when shift sizes are not small. In future work, we may extend our procedure for both monitoring and diagnosis in multistage process.

CHAPTER 5

PROCESS MONITORING IN MULTISTAGE PROCESSES WITH AUTOCORRELATED OBSERVATIONS

5.1 Introduction

As modern manufacturing industries become more sophisticated, it is common to find a production process involving multiple stages such as those found in pharmaceutical manufacturing, chemical industry and semiconductor manufacturing. Most of the multistage SPC approaches treat the multistage system as a whole and lack the capability of discriminate among changes at different stages (Montgomery and Woodall 1997, Shi and Zhou 2009). A regression adjustment method named cause-selecting chart was proposed to model two-stage processes using simple linear regression and then monitoring the residual of the current stage by subtracting the impacts from the previous stage (Zhang 1992, Hawkins 1991, 1993).

Recently, some multistage SPC approaches are developed to exploit the detailed structure of multistage systems to achieve high detection power and diagnostic capability. For example, an exponentially weighted moving average scheme under the static state space model is proposed as a monitoring method for multistage systems (Xiang and Tsung 2008, Zou and Tsung 2008). Methodologies for identifying in-control samples and adjusting the detection power for multistage systems are reported in (Zou et al. 2008, Li and Tsung 2009).

Figure 5.1 shows the complex data relationships in a multistage manufacturing process (MMP) where the *X*, *Y*, and *Z* axes represent the manufacturing stages, the time, and the quality attributes, respectively, and M_i identifies the quality features. There are three types of correlations among such data streams in an MMP (Shi 2007): (i) the quality attributes are correlated in terms of the stages along the production line (e.g., M_2 along the *X* axis); (ii) the quality attributes are correlated among them within the same stage ($[M_1,…, M_m]$ at stage *N* along the *Z* axis); and (iii) each quality attribute is also auto-correlated in terms of time due to the degradation or wear of production tooling over time (M_i, *i*=1, 2, …,*m* along the *Y* axis). Those three types of correlations, observed as a stream of data, introduce significant challenges in variation modeling, analysis, and control. However, there are no available multistage process monitoring procedures that address the autocorrelation along the *Y* axis. All existing approaches assume that the data generated within the same stage are independently distributed (Li and Tsung 2009, Fenner et al. 2009, Zhou et al. 2004, Jin and Tsung 2009, Shi 2007).
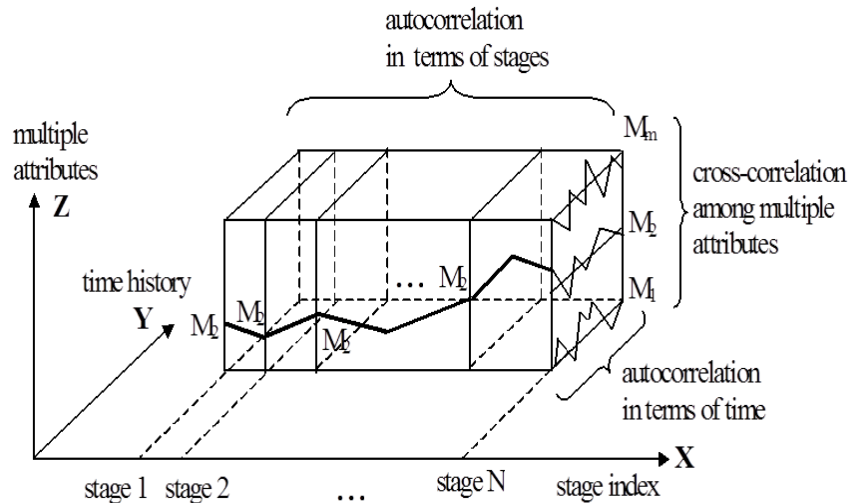

Figure 5.1 Complex data relationships in an MMP (Shi 2007)

The assumption of uncorrelated or independent observations is not even approximately satisfied in most of continuous and batch manufacturing processes such as chemical processes (e.g., liquefied natural gas (LNG) processes) where consecutive measurements on product characteristics are often highly correlated due to automated test and inspection procedure as well as the chemical reaction processes, where the quality characteristics are measured on every unit in time order of production (Montgomery 2008, Rosotowski and Schmid 2006, Jarrett and Pan 2007). As an example, a simplified LNG process is composed of five stages is shown in Figure 5.2. Natural gas is a naturally occurring hydrocarbon gas mixture consisting primarily of methane (CH4), with other hydrocarbons (usually ethane) as well as small amounts of impurities such as carbon dioxide. Through multiple stages natural gas is treated to remove dust, water, hydrogen sulfide, carbon dioxide and other components to increase a percentage of methane (CH4). Figure 5.3 shows a plot of 100 observations of the percentage increase of methane by the removal of certain components (i.e., CO_2, Hg, heavy hydrocarbons, respectively), of the first three stages of the multistage LNG process shown in Figure 5.2 when all three processes are in control. Close examination of this plot reveals that the behaviors of the three state variables ($x_i(t)$, i = 1, 2, 3) over sampling time t are not independent because a value of $x_i(t)$, that is above (or below) the long-term average tends to be followed by other similar values.

Most of continuous-time multistage processes use the dynamic state space model while discrete-time multistage processes use the static state space model to explain the input and output relationships of the processes. Under such conditions traditional multistage SPC techniques may be inappropriate for monitoring process quality, and thus

more appropriate correlated models should be considered. A thorough review of the literature shows that there is no prior work dealing with these autocorrelated models. This chapter investigates the continuous state multistage processes with correlated observation. Approaches for monitoring the process means and variances are developed and their performance is evaluated.
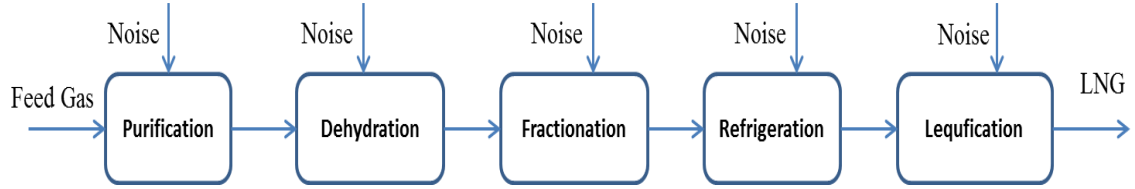
Figure 5.2 An example of a multistage model in an LNG process

(a)   (b)   (c)
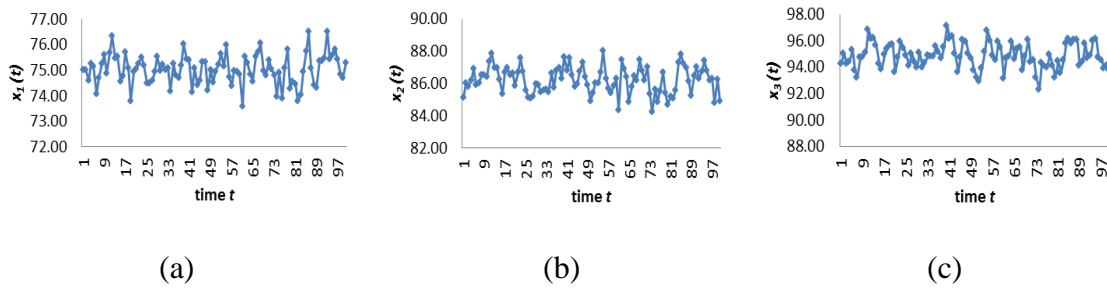
Figure 5.3 State variables in a three-stage model at (a) stage 1, (b) stage 2, and (c) stage 3

5.2 Multistage models and variability propagation

5.2.1 Discrete-time multistage process model

A multistage process model is described by incorporating engineering knowledge with statistical model-based methods to explore the relationship among stages (Jin and Shi 1999, Basseville and Nikiforov 1993, Ding *et al.* 2002, Huang *et al.* 2002, Fenner *et al.* 2005, Agrawal *et al.* 1999). In order to characterize the propagation of variation of quality characteristics in multistage systems, the following linear model of quality

measurement at the kth stage of a process with n stages is proposed (Jin and Shi 1999, Ding *et al* . 2002, Huang *et al* . 2000, Zhou *et al* . 2003). For stations $k = 1, \ldots, l,$

$$Y_k = \beta_{k-1} Y_{k-1} + \varepsilon_k, \tag{5.1}$$

where Y_k is the quality characteristic variable of the kth stage, ε_k is the process noise, $\varepsilon_k \sim N(0, \sigma_{\varepsilon_k}^2)$, and β_{k-1} is the regression coefficient of Y_k on Y_{k-1}, which is assumed to be a known constant. In the linear model, $\beta_{k-1} X_{k-1}$ represents the transformation of quality information from stage k-1 to stage k. Cause-selecting charts (Zhang 1984, Shu *et al*. 2004) is based on the two stage linear model dealing with the cases where an output variable under the normality assumption is a linear function of an input variable.

When the mean of the state variable at stage k is shifted at some time, the out-of-control (OC) process model can be represented (Lawless *et al*. 1996, Lui 2010) as

$$Y_k = \beta_{k-1} Y_{k-1} + \varepsilon_k + \delta_k \tag{5.2}$$

where $\delta_k \neq 0$ is a shift level induced by the process variation sources at stage k. The popular state space models of equation (5.1) are used for the process monitoring of discrete-time multistage processes such as discrete assembly and machining processes where a static linear state space model is assumed. However, in many manufacturing systems such as LNG production process, the system is a continuous-time multistage processes where the dynamic state space model is used. A thorough literature review reveals that the variability propagation models for first-order or higher-order dynamic state space have not been investigated.

5.2.2 Continuous-time multistage process model

We develop models to capture the propagation of mean and variance shifts of autocorrelated multistage processes. The propagation models can capture the correlations

between two stages and autocorrelations of observations over time within the same stage while existing state space model can capture only the correlations between stages. Using the proposed propagation models, the mean and variance of the state variable for each stage can be estimated, which serve as the basis to construct the observations-based control charts.

An important characteristic of multistage manufacturing systems is that the quality of a product at the end of stage k depends on the performance of stage k as well as the input to stage k from previous stages. Also, the shifts of either mean or variance at a certain stage can affect the mean and variance of both that stage and subsequent stages. The state space models of equation (5.1) are used to explain this relationship that characterizes the propagation of variation of quality characteristics in multistage systems. However, as we mentioned in Section 5.1.1, the state space model is developed for discrete-time multistage processes where the static linear state space model is used and observations of a given state variable over time at the same stage are assumed to be independently distributed. However, in many manufacturing systems such as LNG systems, the system is a continuous-time multistage process where the first-order dynamic linear state space model is used and observations of a given state variable within the same stage are autocorrelated over time. In this section, we develop new variation propagation models under the linear dynamic state space for continuous processes to monitor the mean and variance of autocorrelated multistage systems.

To model a continuous-time multistage process, we consider an l stage, continuous flow production process, in which stages are indexed by $k \in \{1, \dots, l\}$. Each stage has its own characteristics that transform the input into an output for the

downstream stage. To model an autocorrelated process, the following series of first order

differential equations are used (English *et al*. 1991, Runger 2002):

$$\frac{dY_k(t)}{dt} + a_k Y_k(t) = b_k Y_{k-1}(t) + \varepsilon_k(t)$$

where a_k and b_k are state parameters, $Y_k(t)$ is a univariate state variable with continuous

time Gaussian white noise $\varepsilon_k(t) \sim N(0, \sigma_k^2)$ of the ith stage at time t, which propagates to

the next stage. When $Y_k(t)$'s are observed at small, equally spaced intervals of time Δt,

where $\tau_0 = 0$ and $\tau_j = \tau_{j-1} + \Delta t$ for $j = 1,2,\dots$, we can approximate the general

solution $Y_k(\tau_j)$ of equation (5.2) as follows

$$Y_k(\tau_j) = Y_k(\tau_{j-1})e^{-a_k \Delta t} + \frac{b_k}{a_k} Y_{k-1}(\tau_j)(1 - e^{-a_k \Delta t}) + \frac{1}{a_k} \varepsilon_k(\tau_j)(1 - e^{-a_k \Delta t}) \qquad (5.3)$$

Then, the process can be characterized by the following state space model with

autoregressive process of order one, AR(1),

$$Y_k(\tau_j) = \Phi_k Y_k(\tau_{j-1}) + A_k Y_{k-1}(\tau_j) + W_k(\tau_j) \qquad (5.4)$$

where $A_k = \frac{b_k}{a_k}(1 - e^{-a_k \Delta t})$ and $\Phi_k = e^{-a_k \Delta t}$ are one dimensional, $W_k(\tau_j) =$

$\frac{1}{a_k} \varepsilon_k(\tau_j)(1 - e^{-a_k \Delta t})$ is a Gaussian white noise with mean 0 and variance $\sigma_{W_k}^2 =$

$\frac{1}{a_k}(1 - e^{-a_k \Delta t})\sigma_k^2$.

5.2.3 The propagation of variability from a stage to the subsequent stages

The following explains how the shift of the mean of stage k is propagated to the

means of subsequent stages. Suppose that the mean of a state variable $Y_k(t)$ of stage k is

changed from μ_{k1} to $\mu_{k2} = \mu_{k1} + \delta_k$, where $\delta_k \neq 0$ at time t_1. The mean of a state

variable $Y_{k+i}(t)$ of stage $k + i$ is given by

$$E\big(Y_{k+i}(t)\big) = \begin{cases} \prod_{j=1}^{k+i} \frac{b_j}{a_j} y_0, & t \le t_1 \\ \prod_{j=1}^{k+i} \frac{b_j}{a_j} x_0 + \delta_k \prod_{j=k+1}^{k+i} \frac{b_j}{a_j}\big(1 - e^{-a_k(t-t_1)}\big), & t > t_1 \end{cases},$$ (5.5)

where y_0 is an initial state. Based on the equation (5.5), the expected value of a state

variable $Y_n(t)$ of the final stage n, $E\big(Y_n(t)\big)$ at time t is given by

$$E\big(Y_n(t)\big) = \begin{cases} \prod_{j=1}^{n} \frac{b_j}{a_j} y_0, & t \le t_1 \\ \prod_{j=1}^{n} \frac{b_j}{a_j} y_0 + \delta_k \prod_{j=k+1}^{n} \frac{b_j}{a_j}\big(1 - e^{-a_k(t-t_1)}\big), & t > t_1 \end{cases}.$$

For example, in two-stage model, when the mean of a state variable at stage 1 is shifted,

then $E\big(Y_1(t)\big)$, shifts from μ_1 to $\mu_1 + \delta_1$, where $\delta_1 \ne 0$ at time t_1 under a two-stage

process as

$$E\big(Y_1(t)\big) = \begin{cases} \frac{b_1}{a_1} y_0, & t \le t_1 \\ \frac{b_1}{a_1} y_0 + \delta_1, & t > t_1 \end{cases},$$

then the mean of the a state variable $Y_2(t)$ at stage 2 is given by

$$E\big(Y_2(t)\big) = \begin{cases} \frac{b_2 b_1}{a_2 a_1} y_0, & t \le t_1 \\ \frac{b_2 b_1}{a_2 a_1} y_0 + \delta_1 \frac{b_2}{a_2}\big(1 - e^{-a_2(t-t_1)}\big), & t > t_1 \end{cases}.$$

Figure 5.4 shows how the expected values of state variables at each stage can change

over time. As shown in Figure 5.4 (a), a mean shift of a state variable at stage 1 occurs at

time t_1, which will be propagated to the mean of a state variable at stage 2. As shown in

Figure 5.4 (b), the mean of the state variable at stage 2 shows (i) a step change in case of

an existing static state space model ($\Phi_k=0$), or (ii) a nonlinear change pattern with faster

increase over time for a higher autocorrelation in the proposed dynamic state space

model, which results in an earlier detection of a mean shift at stage 1.

(a) A mean shift at stage 1 to stage 2

(b) The propagation of a mean shift at stage 1
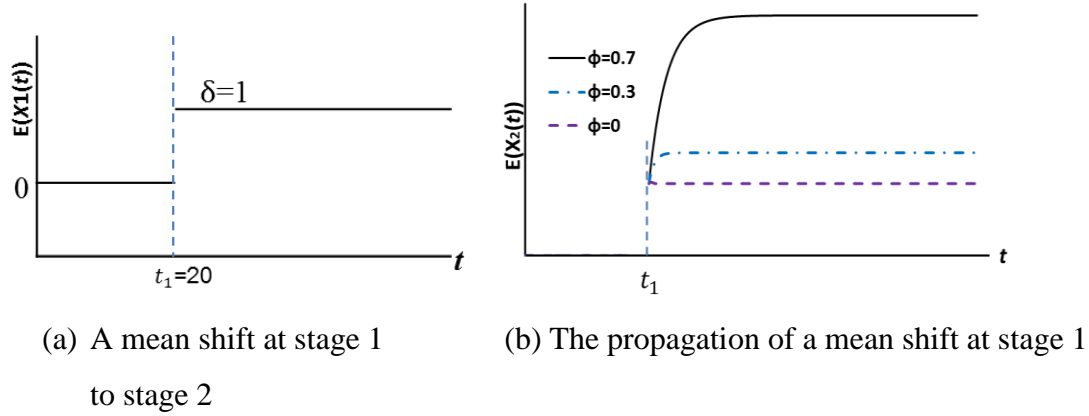
Figure 5.4 Effect of an autocorrelation on the propagation of a mean shift at stage 1 to 2

5.3 Detection of process changes of the mean in multistage processes

When one of the stages of the multistage processes experiences a malfunction or does not maintain the levels of its parameters, a consequence of such a change will be reflected in the final product or downstream of intermediate products (Li and Tsung 2009, Lawless *et al*. 1999, Jin and Tsung 2009). However, most of existing approaches do not consider this cascade property of multistage processes that quality measurements at a certain stage are affected by the output quality from preceding stages (Zhou *et al*. 2003, Zhou *et al*. 2004, Ding *et al*. 2005, Shi and Zhou 2009). In this section, we derive a statistical monitoring procedure for quality measurements in autocorrelated multistage processes with VAR(1) model.

5.3.1 Representation of an autocorrelated multistage process as a VAR(1) model

In section 5.2.2, we show that continuous-time multistage model can be represented as autocorrelated multistage model. In this section, we show that the autocorrelated multistage process can be represented as a VAR(1) model, which is a vector autoregressive model. For simplicity, we use simple form of the equation (5.4) as

$$Y_{it} = a_i + \beta_{i-1}Y_{(i-1)t} + \phi_i Y_{i(t-1)} + \varepsilon_{it},$$

where ϕ_i is the autocorrelation factor for $i = 1,2,\dots,l$. Let $\mathbf{Y_t} = (Y_{1t},\dots,Y_{lt})'$, $\mathbf{A} = (a_1,\dots,a_l)'$, $\mathbf{\varepsilon_t} = (\varepsilon_{1t},\dots,\varepsilon_{lt})'$, then the observation vector can be written as

$$\mathbf{BY_t} = \mathbf{A} + \mathbf{\Phi Y_{t-1}} + \mathbf{\varepsilon_t},$$

where $\mathbf{\Phi} = \text{diag}(\phi_1,\dots,\phi_l)$ and

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\beta_1 & 1 & \ddots & \vdots \\ 0 & \ddots & 1 & 0 \\ 0 & 0 & -\beta_{l-1} & 1 \end{bmatrix}.$$

The white noise vector, $\mathbf{\varepsilon_t}$, follows multivariate normal distribution with $E(\mathbf{\varepsilon_t}) = \mathbf{0}$ and $\text{Cov}(\mathbf{\varepsilon_t}, \mathbf{\varepsilon_t}) = \text{diag}(\sigma_{\varepsilon_1}^2,\dots,\sigma_{\varepsilon_l}^2)$ when all stages are in-control. After multiplying both side by $\mathbf{B}^{-1}$, we obtain the following VAR(1) model

$$\mathbf{Y_t} = \mathbf{\widetilde{A}_0} + \mathbf{\widetilde{\Phi} Y_{t-1}} + \mathbf{\widetilde{\varepsilon}_t}, \tag{5.6}$$

where $\mathbf{\widetilde{A}} = \mathbf{B}^{-1}\mathbf{A}$, $\mathbf{\widetilde{\Phi}} = \mathbf{B}^{-1}\mathbf{\Phi}$, $\mathbf{\widetilde{\varepsilon}_t} = \mathbf{B}^{-1}\mathbf{\varepsilon_t}$, and

$$\mathbf{B}^{-1} = \begin{bmatrix} 1 & 0 & \cdots\cdots & 0 & 0 \\ b_{11} & 1 & 0 & \cdots & 0 \\ b_{21} & b_{22} & \ddots & 0 & \vdots \\ \vdots & \vdots & \ddots & 1 & 0 \\ b_{l-1,1} & b_{l-1,2} & \cdots & b_{l-1,l-1} & 1 \end{bmatrix},$$

where $b_{kj} = \beta_k \cdots \beta_j$ for $k > j$ and $b_{kk} = \beta_k$. For example, the VAR(1) model in a two stage model can be expressed as

$$\mathbf{Y_t} = \begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} a_1 \\ \beta_1 a_1 + a_2 \end{bmatrix} + \begin{bmatrix} \phi_1 & 0 \\ \beta_1\phi_1 & \phi_2 \end{bmatrix}\begin{bmatrix} Y_{1(t-1)} \\ Y_{2(t-1)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \beta_1\varepsilon_{1t} + \varepsilon_{2t} \end{bmatrix},$$

where $Y_{1t} = a_1 + \phi_1 Y_{1(t-1)} + \varepsilon_{1t}$ and $Y_{2t} = a_2 + \beta_1 Y_{1t} + \phi_2 Y_{2(t-1)} + \varepsilon_{2t}$.

The properties of the stationary VAR(1) are studied in various time-series areas (Hamilton 1994). The stationary VAR(1) model can be rewritten as

$$\mathbf{Y_t} = (\mathbf{I} - \mathbf{\widetilde{\Phi}})^{-1}\mathbf{\widetilde{A}_0} + \sum_{i=0}^{\infty} \mathbf{\widetilde{\Phi}}^i \mathbf{\widetilde{\varepsilon}}_{t-i}, \tag{5.7}$$

Stationary mean and covariance matrix can be obtained as

$$E(\mathbf{Y_t}) = \boldsymbol{\mu}_t = \left(\mathbf{I} - \tilde{\boldsymbol{\Phi}}\right)^{-1}\tilde{\mathbf{A}}_\mathbf{0} \text{ and } Cov(\mathbf{Y_t}, \mathbf{Y_t}) = \boldsymbol{\Sigma}_Y = \sum_{i=0}^{\infty} \tilde{\boldsymbol{\Phi}}^i \boldsymbol{\Sigma}_\varepsilon \tilde{\boldsymbol{\Phi}}^{i\prime},$$

where $\boldsymbol{\Phi}^i\boldsymbol{\Sigma}_\varepsilon\boldsymbol{\Phi}^i \to 0$ as $i \to \infty$ and $\text{Cov}(\tilde{\boldsymbol{\varepsilon}}_\mathbf{t}, \tilde{\boldsymbol{\varepsilon}}_\mathbf{t}) = \boldsymbol{\Sigma}_\varepsilon$. The covariance matrix of $\mathbf{Y_t}$ can be approximated as $\boldsymbol{\Sigma}_Y \cong \sum_{i=0}^{25} \tilde{\boldsymbol{\Phi}}^i\boldsymbol{\Sigma}_\varepsilon\tilde{\boldsymbol{\Phi}}^{i\prime}$ (Zhang 1998).

5.3.2 Multivariate SPC approaches for detecting mean changes

5.3.2.1 Observation-based T^2 charts

Assume that a multistage process consists of l stages and the measurements, $\mathbf{Y} = (Y_1, \dots, Y_l)$ follows a multivariate normal distribution, $N_l(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. When the process is in-control (IC), the mean vector is $\boldsymbol{\mu} = \boldsymbol{\mu_0}$, where $\boldsymbol{\mu_0} = (\mu_1, \dots, \mu_l)$, and the covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$, where $\boldsymbol{\Sigma}_0 = \left[\sigma_{ij}\right]_{1 \le i,j \le l}$, is known and constant over time. When the process is out-of-control, the process mean vector is changed to $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_0$. The Hotelling's T^2 statistic can be decomposed to identify the fault variables when the control charts generate an out-of-control signal, and it is defined as

$$T_{Yt}^2 = \left(\mathbf{Y}_t - \boldsymbol{\mu}_0\right)' \boldsymbol{\Sigma}_0^{-1} \left(\mathbf{Y}_t - \boldsymbol{\mu}_0\right),$$

where T^2 statistic follows a χ^2 distribution with l degrees of freedom when the process is in control. With a given false alarm rate α_0, it signals if $T_{Yt}^2 > \chi_{p,\alpha_0}^2$. Although T^2 is the optimal test statistic for a general multivariate shift of the known process mean vector, it is not optimal when some of the process variables are known to be unchanged in high-dimensional processes (Lowry and Montgomery 1995).

5.3.2.2 Residual-based T^2 charts

A basic assumption in most classical SPC chats is that the observations are independent. However this assumption is not true in autocorrelated multistage processes. If it is assumed that observations are independent, but they are correlated in fact, high frequency of false alarms can be generated due to larger variance of the process caused by autocorrelation. In order to overcome this problem, residuals are often used for monitoring autocorrelated observations. If residuals are independent, then residual-based control charts such as Shewhart, EWMA, and CUSUM provide the same detection performances with these charts for i.i.d. observations.

Under the assumption that all parameters are known, we can obtain independent residuals using the linear relationship in equation (5.6). Residual vectors can be defined as

$$\mathbf{e}_t = \mathbf{Y}_t - \widehat{\mathbf{Y}}_t,$$

where $\widehat{\mathbf{Y}}_t = \widetilde{\mathbf{A}}_0 + \widetilde{\boldsymbol{\Phi}}\mathbf{Y}_{t-1}$ for $t = 1,2,\ldots,$ and $\mathbf{Y}_0 = \mathbf{0}$. When $\widetilde{\mathbf{A}}_0$ and $\widetilde{\boldsymbol{\Phi}}$ are known, the residual vector follows a multivariate normal distribution with $(\mathbf{e}_t) = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{e}_t, \mathbf{e}_t) = \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ and. Based on the residual vectors, we can construct T^2 statistic as

$$T^2_{e_t} = \mathbf{e}_t' \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1} \mathbf{e}_t,$$

where $T^2_{e_t}$ follows a χ^2 distribution with l degrees of freedom.

5.3.3 New multistage SPC approaches using unchanged stages information

Runger (1996) propose a conditional T^2 chart for process monitoring of multivariate processes which are known a priori to be stable in control. Let $P = \{1, \ldots, l\}$ is the index set of all variables and $S \subset P$ be a subset of unchanged variables. In this case, all fault variables belong to $\bar{S} = P \backslash S$, which is the complementary set of S. Then the

conditional T^2 can be derived as $T_{P|S}^2 = T_P^2 - T_S^2$ for detecting changes in variables in the set $\bar{S}$, where T_S^2 is the T^2 of all variables in S. It is easy to see that the conditional T^2 follows a χ^2 distribution with degrees of freedom $|\bar{S}|$, where $|\bar{S}|$ denotes the number of elements in $\bar{S}$. The conditional T^2 chart is shown to be more efficient than the conventional T^2 of all variables variables in P or variables in $\bar{S}$.

Table 5.1 shows the efficiency of the conditional T^2 when some information of unchanged variables are known. We evaluate the performance of control charts in terms of ARL, which is defined as the average number of observations $\mathbf{Y}_t$ until the first out-of-control signal is triggered by the control chart for a given mean shift. In this section, we consider $ARL_0 = 200$ for the conventional T^2 and fix the number of variables in the process to $n = 10$. For simplicity, we set $\boldsymbol{\mu}_0 = \mathbf{0}$ and $\boldsymbol{\Sigma}_0 = [\rho_{ij}]_{1 \le i,j \le l}$, where $\rho_{ii} = 1$ and $\rho_{ij} = 0.5$, and the means of 3rd and 7th variables are shifted with various magnitudes $\delta = 1,2,3,4,5$. Let S^* be the index set of all unchanged variables, $S^* = \{1,2,4,5,6,7,9,10\}$ in Table 5.1. When $S = S^*$, the conditional T^2 provides always the best performance, while the conventional T^2 of all 10 variables always performs the worst when $S = \emptyset$. Even only two variables are known such as $S = \{2,4\}$, its performance is superior to the conventional T^2.

Similar performance improvements in multistage processes are shown in Table 5.2, which provides the performance comparisons with various S when $S = \{1, \dots, 7, 9, 10\}$. For simplicity, we set the parameters $\sigma_{\varepsilon_k} = 1$ and $a_k = 0$ for $k = 1, \dots, 10$, and only the mean of stage 8 is changed. Similarly, when all unchanged stage information is known, $S = S^*$, it provides the best performance. It is clear that the detection performance is better when S is closer to S^* in multistage processes.

Table 5.1 Performance comparisons with various S in multivariate processes

δ_3, δ_7	$T^2_{P|S}$				
	$S = \emptyset$	$S = \{2,4\}$	$S = \{2,4,6,8\}$	$S = \{1,2,..,6\}$	$S = S^*$
0.0	201.07	200.01	201.51	199.21	200.7
0.5	105.82	94.64	85.84	76.18	51.10
1.0	27.81	24.30	19.53	14.90	9.36
1.5	7.53	6.29	5.11	3.98	2.84
2.0	2.73	2.47	2.12	1.75	1.44
2.5	1.49	1.38	1.28	1.17	1.09
3.0	1.11	1.07	1.04	1.03	1.01

Table 5.2 Performance comparisons with various S in multistage processes

δ_8	$T^2_{P|\hat{S}}$				
	$S = \emptyset$	$S = \{1,2\}$	$S = \{1,2,3,4\}$	$S = \{1,2,..,6\}$	$S = S^*$
0.0	199.98	197.02	200.29	200.18	201.71
0.5	131.26	114.38	108.84	98.63	89.79
1.0	49.86	44.48	36.30	28.42	23.18
1.5	16.84	14.50	12.07	8.84	7.30
2.0	6.26	5.44	4.50	3.49	3.04
2.5	2.92	2.61	2.27	1.88	1.70
3.0	1.78	1.59	1.45	1.31	1.25

5.3.3.1 Unchanged stages identification procedures in multistage processes

The conditional T^2 statistic assumes that the set of in-control stages S is known a priori. In practice, the priori information is unknown. In this case, a search among all possible subsets can be conducted to select unchanged stages. However, when the number of stage is large, searching all subsets can also be impractical. In this case, we may need a procedure for unchanged stage identification. Let $\hat{S}$ be the subset of P that is selected by an identification procedure.

If the data itself can provide information on the process fault status, then the partial information can be used to construct the conditional T^2 statistic. In autocorrelated multistage processes, cascading property between stages and autocorrelations within stages can make it challenging to identify unchanged stages correctly. In this section, we propose a procedure for identifying unchanged stages based on residuals. For a stage $i \in P$, we can obtain a residual to select the stage has an evidence of unchanged as

$$e_{it} = Y_{it} - \widehat{Y_{it}},$$

where $\widehat{Y_{it}} = a_i + \beta_{i-1}Y_{i-1,t} + \phi_i Y_{i,t-1}$ is the estimate of Y_{it} given the observation of the previous stage $Y_{i-1,t}$ and the observation of the one step before the current time $Y_{i,t-1}$. However, using single residual e_{it} at time t can generate considerably many false selections, EWMA scheme can be used for more accurate unchanged stages identification. For $i = 1, \ldots, l$, the EWMA statistic based on e_{it} is given by

$$\xi_{it} = (1 - r)\xi_{it-1} + re_{it}, \tag{5.8}$$

where $0 < r \leq 1$ and $\frac{\xi_{it}}{\sqrt{var(\xi_{it})}} \sim N(0,1)$. Let p_{it} be the p-value of $\frac{\xi_{it}}{\sqrt{var(\xi_{it})}}$, then we can select an estimated unchanged stage set $\hat{S}$ as

$$\hat{S} = \{i \mid p_{it} > \gamma, i = 1, \dots, l\}, \tag{5.9}$$

where γ is a predefined threshold value. The effect of γ is presented in Table 5.2 in terms of ARL.

5.3.3.2 Multistage SPC charts using unchanged stages information

Because of the i.i.d. property of residuals, we do not consider autocorrelation factors and regression factors in selecting unchanged stages. After the selection, we can construct conditional T^2 statistics to monitor the mean vector of the autocorrerated multistage process. Due to the cascade property, when a shift occurs in early stages such as stage 1 or 2, observation-based T^2 can detect the shift quickly, while it may be insensitive when a shift occurs at later stages such as stage 9 or 10 as shown in Table 5.4. Residual-based T^2 statistics are robust regardless of the location where a shift occurs. Therefore, we construct the conditional T^2 statistics based on residuals as

$$Te_{P|\hat{S}}^2 = \mathbf{e}_t' \mathbf{\Sigma}_\varepsilon^{-1} \mathbf{e}_t - \mathbf{e}_t'(\hat{S}) \mathbf{\Sigma}_\varepsilon^{-1}(\hat{S}, \hat{S}) \mathbf{e}_t(\hat{S}), \tag{5.10}$$

where $\mathbf{e}_t(\hat{S})$ is a subset of variables in vector $\mathbf{e}_t$ and $\mathbf{\Sigma}_\varepsilon(\hat{S}, \hat{S})$ is a covariance matrix of $\mathbf{e}_t(\hat{S})$. It is easy to show that $Te_{P|\hat{S}}^2$ follows a χ^2 distribution with degrees of freedom $|P \backslash \hat{S}|$.

Although we consider a Hotelling's T^2-type control statistic using individual residual vectors, EWMA-type control charts can be developed in a similar way to monitor small shifts. Let $\Xi_t = (\xi_{1t}, \dots, \xi_{lt})'$, where ξ_{it} is the residual-based EWMA statistic at the ith stage defined in equation (5.8), be the observed residual vector collected over time t. In this case, we propose a conditional multivariate EWMA

(CMEMWA) chart based on the MEWMA sequence of statistics (Lowry *et al*. 1992), which is defined as

$$T^2_{\Xi,P|\hat{S}} = \Xi'_t \Sigma_\varepsilon^{-1} \Xi_t - \Xi'_t\left(\hat{S}\right)\Sigma_\varepsilon^{-1}(\hat{S},\hat{S})\Xi_t(\hat{S}), \qquad (5.11)$$

where $\Sigma_\Xi = \dfrac{r}{2-r}\Sigma_\varepsilon$, $\Xi_0 = \mathbf{0}$, and $r \in (0,1]$ is a weighting parameter. It signals when the test statistic $T^2_{\Xi,P|\hat{S}}$ is a greater than a control limit chosen to satisfy a given in-control ARL.

5.3.4 Accuracy of unchanged stages identification procedures in multistage processes

Let $\hat{S}$ be a set of selected unchanged stages, where $\hat{S} = \{i|\ p_{it} > \gamma, i = 1, ..., l\}$, with a given threshold value γ. The design of unchanged stages identification involves the choice of γ. With the larger γ, the smaller number of unchanged stages are selected with higher probability of $\hat{S} \subset S^*$. However selecting smaller number of unchanged stages causes a smaller performance improvement. When smaller γ is chosen, it is expected that larger number of unchanged stages are selected, but it increases false selection rate, that is $\hat{S} \not\subset S^*$. Selecting larger number of unchanged stages causes a larger performance improvement when only a few stages are fault stages.

In this section, we propose a performance measure of accuracy to suggest a guideline how to find a proper γ value. Basic goal is to select as many as unchanged stages without false changed stage selection. It seems reasonable that a performance measure (PM) of accuracy satisfies two properties:

(i)   If $\hat{S} \subset S^*$, then $PM \rightarrow 1$ as $|\widehat{S}| \rightarrow |S^*|$,

(ii)   If $\hat{S} \not\subset S^*$, then $PM = 0$.

Although other performance measures can be used, we propose a novel measure of accuracy as

$$PM = \frac{E(|S^* \cap \hat{S}|)}{|S^*|} P(\hat{S} \subset S^*) + 0 \cdot P(\hat{S} \not\subset S^*), \qquad (5.12)$$

where $|S^*|$ is the number of elements of S^*. The measure can be reformed as

$$PM = \frac{|S^* \cap \hat{S}|}{|S^*|} I_{\{\hat{S} \subset S^*\}},$$

where $I_{\{\}}$ is the indicator function.

For the multistage process model, shifts of magnitude δ occur at the stages only indexed in $\bar{S} = P \backslash S^*$. For instance, when $\bar{S} = \{2,8\}$, the means of stages 2 and 8 are changed by δ, where the shifts initially occur. Table 5.3 compares the values of PMs with various γ for two different fault scenarios: (i) single fault stage, $\bar{S} = \{5\}$; (ii) two fault stages $\bar{S} = \{2,8\}$. In both scenarios, the smaller γ provides poor accuracy when the magnitude of the mean shift is smaller, while the larger γ is poor when the magnitude larger. In general, since most SPC charts detect shift quickly when a shift magnitude is large, identifying unchanged stages with the large shift size may improve a little of the performance of SPC charts. Thus it may reasonable to give more credits for higher accuracy in identifying unchanged stages when the magnitudes of shifts are smaller.

Table 5.3 shows that $\gamma = 0.8$ provides the lowest accuracy overall, and $\gamma = 0.5$ and 0.1 provides low accuracy when the magnitude of shifts is large and high, respectively. When δ is unknown, it is reasonable to select $\gamma = 0.25$, since it provides overall good accuracy for a wide range of small shifts. Table 5.4 compares ARL performance of the proposed CMEWMA chart with various $\gamma = 0.5, 0.25, 0.1$ when a shift occurs only at stage 5 and both stages 2 and 8. The EWMA parameter used in this

simulation is $r = 0.2$ and $ARL_0 = 200$. Generally, the CMEMWA with the larger γ is sensitive to the smaller shifts, while the CMEMWA with the smaller γ is sensitive to the larger shifts. Notice that the proposed CMEWMA with $\gamma = 0.5$ is better than $\gamma = 0.25$ for signaling shifts $\delta \leq 0.5$, and the CMEWMA with $\gamma = 0.25$ is better than $\gamma = 0.5$ and $\gamma = 0.1$ for signaling shifts $0.5 < \delta < 3.0$. The table reveals that CMEWMA with $\gamma = 0.25$ is sensitive to a wide range of small shifts. Hence it is reasonable to select $\gamma = 0.25$ when δ is unknown.

Table 5.3 PMs for accuracy with various γ in multistage processes

δ	Single fault stage ($\bar{S} = \{5\}$)				Two fault stages ($\bar{S} = \{2,8\}$)			
	0.80	0.50	0.25	0.10	0.80	0.50	0.25	0.10
0.5	0.16	0.28	0.23	0.13	0.14	0.15	0.07	0.02
1.0	0.18	0.34	0.34	0.24	0.16	0.23	0.16	0.06
1.5	0.19	0.41	0.48	0.40	0.18	0.32	0.33	0.18
2.0	0.19	0.45	0.60	0.57	0.19	0.42	0.48	0.36
2.5	0.20	0.48	0.68	0.72	0.20	0.47	0.62	0.58
3.0	0.20	0.50	0.73	0.80	0.20	0.49	0.70	0.75

Table 5.4 ARL performances of CMEWMA with various $\gamma = 0.5,\ 0.25, 0.1$

δ	Single fault stage ($\bar{S} = \{5\}$)			Two fault stages ($\bar{S} = \{2,8\}$)		
	$\gamma = 0.50$	0.25	0.10	0.50	0.25	0.10
0.0	201.71	201.99	199.72	201.28	202.60	200.46
0.5	61.37	61.44	64.65	35.72	35.84	37.75
1.0	16.21	15.50	16.69	8.97	8.83	9.24
1.5	7.80	7.57	7.76	5.02	4.95	5.10
2.0	5.16	4.95	5.05	3.47	3.40	3.51
2.5	3.86	3.74	3.84	2.75	2.68	2.72
3.0	3.14	3.02	3.00	2.27	2.24	2.23

5.4 Performance Comparisons

In this section, we investigate the detection performance of the proposed scheme through Monte Carlo simulation and compare the results with existing approaches under various combinations of the multistage parameters. We assume individual observations. The in-control process is assumed with a mean of zero and a standard deviation of one for each individual variable.

We evaluate the performance of control charts in terms of ARL, which is defined as the average number of observations until the first out-of-control signal is triggered by the control chart for a given mean shift. In this section, we consider MEWMA-type charts to detect small mean shifts quickly. Here we compare the proposed CMEWMA with residual-based MEMWA (RMEWMA) charts and observation-based MEWMA (OMEWMA). Similar to the previous simulations, we consider $ARL_0 = 200$ for all MEWMA-type charts and set the number of stages in the process to $n = 10$ with parameters $\sigma_{\varepsilon_k} = 1$ and $a_k = 0$ for $k = 1, \dots, l$. The proposed CMEMWA chart based on the conditional $T^2_{\Xi,P|\hat{S}}$ statistic using selected unchanged stages information defined in equation (5.11), is compared with OMEWMA and RMEWMA. The smoothing constants r for the proposed chart is specified as 0.2 and $\gamma = 0.25$ is selected based on the results in Table 5.2, which presents the effect of γ. The numerical results show that overall ARL performances are almost best when $\gamma = 0.25$, while performances with $\gamma = 0.1$ and $\gamma = 0.50$ are slightly better when $\delta = 3.0$ and $\delta = 0.5$, respectively.

In this section, the ARLs for all charts are determined through Monte Carlo simulation with 20,000 replications. For the multistage process model, shifts of magnitude δ occur at stages indexed in $\bar{S} = P \backslash S^*$. For instance, when $\bar{S} = \{3, 5\}$, the

means of stages 3 and 5 are changed by δ, where the shifts initially occur. In the simulations, various levels of $\delta = 0.5, 1.0, 1.5, 2.0, 2.5, 3.0$ are considered for single fault stage, $\bar{S} = \{2\}$, $\{5\}$, $\{9\}$, and two fault stages $\bar{S} = \{4,6\}$ and $\bar{S} = \{1,10\}$.

Table 5.5 shows the proposed CMEWMA is almost uniformly superior to the other two MEWMA charts, OMEWMA and RMEWMA, in detecting small magnitude of shifts. The observation-based OMWMA chart exhibits the worst sensitivity in the autocorrelated multistage process when shifts occur at downstream stages such as stages later than stage 5. The residual-based charts such as RMEWMA and CMEWMA provide considerably robust results regardless of locations. Further, CMEWMA performs superior to RMEWMA and OMEWMA charts in detecting overall shifts.

Table 5.5 ARL comparisons of procedures for shifts with various fault locations

	$\bar{S} = \{9\}$			$\bar{S} = \{5\}$			$\bar{S} = \{2\}$		
δ	RM	OM	CM	RM	OM	CM	RM	OM	CM
0.0	209.11	200.42	200.89	203.15	199.49	201.99	204.09	202.30	200.73
0.5	68.12	123.98	63.17	68.33	103.58	61.04	66.04	58.41	62.01
1.0	17.61	39.31	15.47	17.87	30.78	15.50	17.96	17.35	15.57
1.5	8.70	15.63	7.56	8.63	12.74	7.57	8.42	8.93	7.47
2.0	5.55	8.92	4.93	5.60	7.64	4.95	5.67	6.17	4.92
2.5	4.22	6.23	3.72	4.22	5.56	3.74	4.25	4.80	3.72
3.0	3.42	4.89	3.00	3.42	4.40	3.00	3.42	3.96	3.01

* RM: RMEMWA, OM: OMEWMA, CM: CMEWMA

In Table 5.6, we consider the combinations of two different fault stages scenarios: (i) fault stages are located relatively close together such as $\bar{S} = \{4,6\}$; (ii) fault stages are located relatively far from each other such as $\bar{S} = \{1,10\}$. Control limits of OMEWMA, RMEWMA, CMEWMA are $h = 1.5$, 0.96, 1.4, respectively, for $ARL_0 = 200$. The

experimental results in Table 5.5 and 5.6 reveal that the proposed CMEWMA chart perform almost best when shifts occur at the stages that are not too early with reasonable level of autocorrelation. Based on simulation results, we conclude that the proposed conditional chart, by taking advantages of incorporating unchanged stage information, provides considerably large improvements in detecting small shifts in terms of ARL.

Table 5.6 Performance comparisons with MEWMA-type procedures

δ	$\bar{S} = \{4,6\}$			$\bar{S} = \{1,10\}$		
	RMEWMA	OMEWMA	CMEWMA	RMEWMA	OMEWMA	CMEWMA
0.0	201.80	202.90	201.02	200.25	201.78	201.16
0.5	36.29	56.51	35.60	37.15	32.13	36.78
1.0	9.53	13.75	9.05	9.38	9.98	9.17
1.5	5.21	6.97	4.92	5.15	6.19	4.97
2.0	3.65	4.79	3.43	3.67	4.60	3.48
2.5	2.89	3.75	2.67	2.89	3.73	2.73
3.0	2.38	3.14	2.24	2.41	3.17	2.26

5.5 Concluding Remarks

Multistages processes are sequentially concatenated. Generally, each stage consists of input variables and output variables where model-based SPC charts can monitor the output variables which can be explained by the input variables. In this chapter we develop advanced SPC methodologies and associated tools for autocorrelated multistage processes. We model an autocorrelated multistage process as VAR(1) model and derive the propagation models of mean shifts to subsequent stages under the state space model. Further, we propose a new procedure CMEWMA chart to detect the shift

of mean in a multistage process by incorporating unchanged stage information. The experimental results show that the proposed CMEWMA chart is efficient in detecting a wide range of small mean shifts compared with the observation-based and the residual-based MEWMA charts.

  Our initial results show this approach is promising and lead to efficient approaches for shift detection in the mean of multistage processes when shifts occur at the stages that are not too early with reasonable level of autocorrelation. However when shifts occur at the early stage(s) with larger autocorrelation, observation-based charts can provide better detection performances. To overcome this drawback under these situations, we intend to explore procedures dealing with autocorrelation information.

CHAPTER 6

CONCLUDING REMARKS AND FUTURE RESEARCH

6.1 Concluding remarks

In this dissertation, we propose and subsequently develop several methodolgies for SPC for univariate and multivariate processes. In chapter 2, we propose an adaptive runs rule, which is motivated by the concept of supplementary runs rule, in order to make control charts more sensitive to small mean shifts. The adaptive runs rule assigns scores to consecutive runs based on the estimated shift size of the mean. We supplement the ACUSUM chart with the adaptive runs rule to enhance its sensitivity in detecting small mean shifts. The average run length performance of the ACUSUM chart with the adaptive runs rule is compared with those of CUSUM and variants of adaptive charts including ACUSUM. The experimental results reveal that the ACUSUM chart with the adaptive runs rule achieves superior detection performance over a wide range of mean shifts.

In chapter 3, we propose an MSPC chart based on a sequential test having an optimal property for testing shift vectors with a specific noncentrality parameter. Due to difficulty of having a closed form for the test statistics using log-likelihood ratios of the sequential test, which makes them impractical for real applications, we drive an approximate log-likelihood ratio which is integrated into an MSPC chart for detecting

shift vectors having a specific noncentrality parameter. Further, we adopt an adaptive scheme by adjusting its reference value based on an MEWMA estimate for detecting a broader range of mean shifts. The statistical properties of the proposed test statistic are explored. The ARL performance of the proposed chart is compared with other MSPC charts for process mean monitoring. The simulation results reveal that the proposed MSPC chart achieves superior detection performance over a wide range of mean shifts, especially when the dimension of measurements is large.

In chapter 4, we propose an adaptive step-down procedure using conditional T^2 statistics for fault variable identification. By selecting a variable having no significant evidence of a change based on the p variables that are selected in previous steps, we can construct single decomposition among $p!$ MYT decompositions. The proposed procedure provides reasonable computational complexity in high-dimensional processes and enhances diagnostic power in identifying the shifted components of the mean vector when a shift occurs only in a few variables.

Finally, in chapter 5, we propose advanced SPC methodologies and associated tools for multistage processes. We model an autocorrelated multistage process and derive the propagation models of mean shifts to subsequent stages under the state space model. Further, we develop methods to detect the shift of mean in a multistage process by incorporating unchanged stage information. The experimental results show that the proposed CMEWMA procedure performs consistently better than existing observation-based and residual-based MEWMA charts.

.

6.2 Future research

Although the dissertation intends to develop a monitoring procedure, identifying OC stages when the shift initially occurs is critical to many industries such as LNG processes and semiconductor manufacturing. Our ongoing research effort is to develop a diagnosis procedure considering autocorrelations and unchanged stage information. In future work, we may extend our procedures for both monitoring and diagnosis in multistage process.

Moreover, most of SPC procedures detect the mean shifts of multistage processes under the assumption that process variability is constant (i.e., not changed) over time (Zantek et al. 2002, 2006). Various charting approaches have been proposed to detect variability changes at a single stage process (Montgomery and Wadsworth 1972; Alt and Smith 1988; Aparisi et al. 1999). Some recently developed multistage variation monitoring approaches do not consider the variation propagation and do not discriminate between local and propagated variations (Zeng and Zhou 2008). The monitoring procedures may consider the propagated variation as a local variation of a given stage, but this may increase the number of false alarms. Thus, multistage monitoring of process variability is a challenging problem due to the variation propagation of multistage processes. Therefore, it deserves further attention to develop a new variability monitoring procedure that considers propagated variations from preceding stages.

Finally, simultaneous monitoring of the mean and variance changes in an autocorrelated multistage process is also open for further research. We believe that more methodologies will bring improvements in this research area.

Appendix A. MARKOV CHAIN APPROXIMATION OF ACUSUM CHARTS

For simplicity we develop a Markov chain model for an upper ACUSUM chart. Similar to the Markov chain model suggested by Shu and Jiang (2006), the upper ACUSUM chart has a two-dimensional state space of $(\hat{\delta}_t^+, Z_t^+)$. The range of possible $\hat{\delta}_t^+$ is $U = [\delta_{min}^+, \delta_{max}^+]$ so that U is partitioned into $m_1 + 1$ subintervals, $U_0 = \{\delta_{min}^+\}$ and $U_i = (u_i - \Delta_1, u_i + \Delta_1)$ for $i = 1, ..., m_1$, where $u_i = u_{i-1} + 2\Delta_1$ and $\Delta_1 = (\delta_{max}^+ - \delta_{min}^+)/2m_1$.

Similarly, the range of possible Z_t^+ values is $S = [0, h]$ that is partitioned into $m_2 + 1$ subintervals, $S_0 = \{0\}$ and $S_i = (s_i - \Delta_2, s_i + \Delta_2)$ for $i = 1, ..., m_2$, where $s_i = s_{i-1} + 2\Delta_2$ and $\Delta_2 = h/2m_2$. When $\hat{\delta}_t^+ \in U_i$ and $Z_t^+ \in S_j$, $(\hat{\delta}_t^+, Z_t^+)$ is in a transient state (i, j) at time t. To approximate the transition probability, we assume that the control statistics $\hat{\delta}_t^+ \in U_i$ and $Z_t^+ \in S_j$ are equal to the center points u_i and s_j, respectively. The transition probability $P_{(i_0, j_0), (i_1, j_1)}$ from state (i_0, j_0) to state (i_1, j_1), where $i_1, j_1 \neq 0$, can be obtained by

$$
\begin{aligned}
P_{(i_0, j_0), (i_1, j_1)} &= \Pr\{\hat{\delta}_t^+ \in U_{i_1}, Z_t^+ \in S_{j_1} \mid \hat{\delta}_{t-1}^+ = u_{i_0}, Z_{t-1}^+ = s_{j_0}\} \\
&= \Pr\{u_{i_1} - \Delta_1 < \hat{\delta}_t^+ < u_{i_1} + \Delta_1, s_{j_1} - \Delta_2 < Z_t^+ < s_{j_1} + \Delta_2 \mid \hat{\delta}_{t-1}^+ = u_{i_0}, Z_{t-1}^+ = s_{j_0}\} \\
&= \Pr\{a_1 < X_t < a_2, b_1 < X_t < b_2\} \\
&= \Pr\{\min[b_2, \max(a_1, b_1)] < X_t < \min[b_1, \max(a_2, b_2)]\}
\end{aligned}
$$

where

$$
a_1 = \begin{cases} \left(u_{i_1} - \Delta_1 - (1-\lambda)u_{i_0}\right)\lambda^{-1}, & \text{if } j_0 \neq 0 \\ u_{i_1} - \Delta_1, & \text{if } j_0 = 0 \end{cases},
$$

$$a_2 = \begin{cases} \left(u_{i_1} + \Delta_1 - (1-\lambda)u_{i_0}\right)\lambda^{-1}, & \text{if } j_0 \neq 0 \\ u_{i_1} + \Delta_1, & \text{if } j_0 = 0 \end{cases},$$

$$b_1 = \left(s_{j_1} - \Delta_2 - s_{j_0}\right)h\left(u_{i_1}/2\right) + u_{i_1}/2,$$

$$b_2 = \left(s_{j_1} + \Delta_2 - s_{j_0}\right)h\left(u_{i_1}/2\right) + u_{i_1}/2.$$

Note that $\hat{\delta}_t^+$ is reset to $\hat{\delta}_0^+$ whenever $Z_t^+ \leq 0$ so that $\hat{\delta}_t^+ = \hat{\delta}_0^+$ if $j_0 = 0$, otherwise

$\hat{\delta}_t^+ = \lambda X_t + (1-\lambda)\hat{\delta}_{t-1}^+$. Similar to the transition probability by Shu and Jiang (2006), the

transition probability $P_{(i_0,j_0),(i_1,j_1)}$ from state (i_0, j_0) and (i_1, j_1) can be obtained as

$$P_{(i_0,j_0),(i_1,j_1)} = \Pr\{\hat{\delta}_t^+ \in U_{i_1}, Z_t^+ \in S_{j_1} \mid \hat{\delta}_{t-1}^+ = u_{i_0}, Z_{t-1}^+ = s_{j_0}\}$$

$$= \begin{cases} \Pr\{\min[b_2, \max(a_1, b_1)] < X_t < \max[b_1, \min(a_2, b_2)]\}, & \text{if } i_1 \neq 0, j_1 \neq 0 \\ \Pr\{a_1 < X_t < \max[a_1, \min(a_2, b_2)]\}, & \text{if } i_1 \neq 0, j_1 = 0 \\ \Pr\{b_1 < X_t < \max[b_1, \min(a_2, b_2)]\}, & \text{if } i_1 = 0, j_1 \neq 0 \\ \Pr\{-\infty < X_t < \min(a_2, b_2)\}, & \text{if } i_1 = 0, j_1 = 0 \end{cases}$$

$$= \Pr\{c_1 < X_t < c_2\},$$

where $P_{(i_0,j_0),(i_1,j_1)} = \Pr\{c_1 < X_t < c_1\}$ is a general form of the transition probability. When

$i_1 = 0$ and $j_1 = 0$, we set $c_1 = -\infty$.

We can obtain the approximated ARL based on the transition probability.

Suppose that **1** is a column vector of ones and **I** is the identity matrix. The transition

matrix is formed as

$$\mathbf{P} = \begin{bmatrix} \mathbf{R} & \mathbf{Q} \\ 0,...,0 & 1 \end{bmatrix},$$

where $\mathbf{Q} = (\mathbf{I} - \mathbf{R})\mathbf{1}$ is a vector that contains the probabilities from one transient state to the OC state and the submatrix $\mathbf{R}$ includes the probabilities of going from one transient state to another. The approximated zero-state ARL is computed as

$$ARL = \mathbf{P}_0^T (\mathbf{I} - \mathbf{R})^{-1} \mathbf{1},$$

where $\mathbf{P}_0^T$ is the initial probability vector. Further, the steady-state ARL can be evaluated using a cyclic steady-state probability vector (Lucas and Saccucci 1990).

Some experiments with different values of m suggest that satisfactory results can be obtained by choosing $m = m_1 = m_2$ to be greater than 30. As an example, the following table shows the approximations to the IC ARL of the ACUSUM chart computed using different values of m with $[\delta_{\min}^+, \delta_{\max}^+] = [0.5, 4.0], h = 1.025, \lambda = 0.2$.

m	15	20	25	30	35	40	45
ARL	**467.96**	**479.27**	**486.54**	**493.47**	**498.92**	**499.23**	**499.73**

We extend a two-dimensional Markov chain model of the improved ACUSUM charts to three-dimensional model for ACUSUM-ACR charts using the random vector $(\hat{\delta}_t^+, Z_t^+, N_t^+)$. The discretized state spaces of $\hat{\delta}_t^+$ and Z_t^+ are the same as those of the ACUSUM charts. The IC state space of N_t^+ is $V=[0,l]$ that is partitioned into $o+1$ subintervals, $V_0 = \{0\}$ and $V_i = (v_i - \Delta_3, v_i + \Delta_3)$ for $i = 1, ..., o$, where $v_i = v_{i-1} + 2\Delta_3$ and $\Delta_3 = l/2o$. When $N_t^+ \in V_i$, we approximate N_t^+ by the midpoint v_i of V_i. Note that $N_t^+ = \max\{0, N_0^+ + \hat{\delta}_t^+(X_t - \hat{\delta}_t^+ / 2)\}$ whenever $Z_t^+ > 0$. For simplicity, we define

$$d_1 = \begin{cases} (v_{j_0} - \Delta_3)(u_{k_1})^{-1} - u_{k_1} / 2, & \text{if } k_1 \neq 0 \\ -\infty, & \text{if } k_1 = 0 \end{cases},$$

$$d_2 = \begin{cases} (v_{j_0} + \Delta_3)(u_{k_1})^{-1} - u_{k_1}/2, & \text{if } k_1 \neq 0 \\ -u_{k_1}/2, & \text{if } k_1 = 0 \end{cases}$$

.

The transition probability $P_{(i_0,j_0,k_0),(i_1,j_1,k_1)}$ from state (i_0, j_0, k_0) to state (i_1, j_1, k_1) when $j_1 \neq 0$ is

$$P_{(i_0,j_0,k_0),(i_1,j_1,k_1)} = \Pr\{\hat{\delta}_t^+ \in U_{i_1}, Z_t^+ \in S_{j_1}, N_t^+ \in V_{k_1} \mid \hat{\delta}_{t-1}^+ = u_{i_0}, Z_{t-1}^+ = s_{j_0}, N_{t-1}^+ = u_{k_0}\}$$

$$= \Pr\{c_1 < X_t < c_2, d_1 < X_t < d_2\}$$

$$= \begin{cases} \Pr\{\min[d_2, \max(c_1, d_1)] < X_t < \min[d_1, \max(c_2, d_2)]\}, & \text{if } k_1 \neq 0. \\ \Pr\{c_1 < X_t < \min(c_2, d_2)\}, & \text{if } k_1 = 0. \end{cases}$$

Note that N_t^+ is reset to zero when $j_1 = 0$. In this case, the transition probability is

$$P_{(i_0,j_0,k_0),(i_1,0,0)} = \Pr\{\hat{\delta}_t^+ \in U_{i_1}, Z_t^+ \in S_0, N_t^+ \in V_0 \mid \hat{\delta}_{t-1}^+ = u_{i_0}, Z_{t-1}^+ = s_{j_0}, N_{t-1}^+ = u_{k_0}\}$$

$$= \Pr\{c_1 < X_t < c_2\}$$

$$= \begin{cases} \Pr\{a_1 < X_t < \max[a_1, \min(a_2, b_2)]\}, & \text{if } i_1 \neq 0 \\ \Pr\{-\infty < X_t < \min(a_2, b_2)\}, & \text{if } i_1 = 0 \end{cases}.$$

Appendix B. DERIVATION OF APPROXIMATE LOG-LIKELIHOOD RATIO

For simplicity, we assume that $\boldsymbol{\mu}_0 = \mathbf{0}$ in Appendix. Patnaik's approximation consists of replacing a noncentral χ^2 by a central χ^2 as

$$\chi^2\left(p, \lambda^2\right) \approx \frac{p + 2\lambda^2}{p + \lambda^2} \chi^2\left(p + w, 0\right),$$

where $w = \lambda^4/(p + 2\lambda^2)$ and $\chi^2(p + w, 0)$ is a central χ^2 distribution with $p + w$ df.

When $\lambda = \lambda_1$, $n\left\|\bar{\mathbf{X}}_n\right\|_{\Sigma_0}$ follows a noncentral χ^2 distribution $\chi^2(p, \lambda_n^2)$, where $\lambda_n = \sqrt{n}\lambda_1$

. Then, the log-likelihood ratio Λ_n can be approximated as follows

$$\Lambda_n = \log \frac{f(n\left\|\bar{\mathbf{X}}_n\right\| \mid \lambda = \lambda_1)}{f(n\left\|\bar{\mathbf{X}}_n\right\| \mid \lambda = 0)} \approx \tilde{\Lambda}_n = \log \frac{\dfrac{p + 2\lambda_n^2}{p + \lambda_n^2} \chi^2\left(p + w, 0\right)}{\chi^2\left(p, 0\right)}.$$

Since the pdf of $\chi^2\left(p, 0\right)$ is $f_X(x) = \dfrac{1}{2^{p/2}\Gamma\left(p/2\right)} x^{\frac{p}{2}-1} e^{-\frac{x}{2}}$, we can obtain

$$\tilde{\Lambda}_n = 0.5w_n\log\left(n\left\|\bar{\mathbf{X}}_n\right\|\right) + \log\frac{p + 2\lambda_n^2}{p + \lambda_n^2} + \log\Gamma\left(\frac{p}{2}\right) - \left(0.5w_n\log 2 + \log\Gamma\left(\frac{p + w_n}{2}\right)\right),$$

where $w_n = \dfrac{\lambda_n^4}{p + 2\lambda_n^2}$ and $\Gamma(\cdot)$ is a gamma function. By rescaling, the approximated ratio can be expressed as

$$\tilde{\Lambda}_n = w_n\log\left(n\left\|\bar{\mathbf{X}}_n\right\|\right) - k_n,$$

where

$$k_n = \left(w_n \log 2 + 2 \log \Gamma \left(\frac{p+w_n}{2} \right) \right) - 2 \left(\log \frac{p+2\lambda_n^2}{p+\lambda_n^2} + \log \Gamma \left(\frac{p}{2} \right) \right).$$

The test statistic of the MC1 chart is based on the statistic $\sqrt{\mathbf{D}_n' \boldsymbol{\Sigma}_0^{-1} \mathbf{D}_n} - 0.5 n \lambda_1$, where $\mathbf{D}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n$. Jackson and Bradley (1961) define the log-likelihood ratio Λ_n as

$$\Lambda_n = \log G \left(\frac{p}{2}, \frac{\lambda_1 n^2 \| \bar{\mathbf{X}}_n \|}{4} \right) - \frac{n \lambda_1}{2}$$

where

$$G(b,x) = 1 + \frac{x}{b} + \frac{x}{b(b+1)2!} + \cdots + \frac{x^n}{b(b+1)...(b+n-1)n!} + \cdots$$

is the generalized hypergeometirc function. Assuming that

$$\log G \left(\frac{p}{2}, \frac{\lambda_1 n^2 \| \bar{\mathbf{X}}_n \|}{4} \right) \approx \sqrt{\mathbf{D}_n' \boldsymbol{\Sigma}_0^{-1} \mathbf{D}_n},$$

where $\sqrt{\mathbf{D}_n' \boldsymbol{\Sigma}_0^{-1} \mathbf{D}_n} - 0.5 n \lambda_1$ can be an approximation of Λ_n. That is, under the above assumption, the MC1 statics are based on likelihood ratios of a sequential χ^2 test. In this case, the test statistics of the MC1 and the MASC charts are equivalent, and both charts may provide similar performance.

Appendix C. DERIVATION OF THE NONCENTRALITY PARAMETER OF $T_{i \cdot \Gamma}^2$

For simplicity, we assume that $\boldsymbol{\mu_0} = \mathbf{0}$, and $\sigma_{ii} = 1$ and $0 \leq \sigma_{ij} < 1$ for $i, j \in \{1, \dots, p\}$. Suppose that the mean of X_i, where $i \in S$, is shifted to $\delta_i \neq 0$, and the means of all elements in Γ are zero, $E(X_j) = 0$ for all $j \in \Gamma$. Then the expected value of $T_{i \cdot \Gamma}$ is

$$E(T_{i \cdot \Gamma}) = \frac{\delta_i}{\sqrt{1 - \boldsymbol{\beta}^{(i)} \boldsymbol{\Sigma}_{\Gamma S}^{(i)}}},$$

where $\boldsymbol{\beta}^{(i)} = \boldsymbol{\Sigma}_{\Gamma S}^{(i)'} \boldsymbol{\Sigma}_{\Gamma \Gamma}^{-1}$. Therefore the noncentrality parameter of $T_{i \cdot \Gamma}^2$ is

$$\lambda_i = \frac{\delta_i^2}{1 - \boldsymbol{\beta}^{(i)} \boldsymbol{\Sigma}_{\Gamma S}^{(i)}}.$$

Let δ_i^2 be the noncentality parameter of T_i^2, which follows a χ^2 distribution with one degree of freedom. Since $0 \leq 1 - \boldsymbol{\beta}^{(i)} \boldsymbol{\Sigma}_{21}^{(i)} < 1$, it can be shown that $\lambda_i \geq \delta_i^2$. Let $t > 0$ be a constant value, then difference between cdf's of noncentral chi-squares, T_i^2 and $T_{i \cdot \Gamma}^2$, is given by (Johnson and Kotz 1970)

$$Pr(T_i^2 \leq t) - \Pr(T_{i \cdot \Gamma}^2 \leq t) \geq 0.$$

From the above equation, we can obtain

$$\Pr(T_{i \cdot \Gamma}^2 > t) \geq Pr(T_i^2 > t).$$

REFERENCES

Acosta-Mejia, C. A., 2007. Two sets of runs rules for the X chart. *Quality Engineering,* 19, 129-136.

Alt, F. B., 1985. Multivariate quality control. In *Ency-clopedia of Statistical Sciences*, Kotz, S., Johnson, N. L. and Read, C. R. E., 111-122.

Alwan, L. C., 1986. CUSUM quality control - multivariate approach. *Communications in Statistics—Theory and Methods,* 15, 3531-3543.

Apley, D. and Shi, J., 1998. Diagnostics in disassembly unscrewing operations. *International Journal of Flexible Manufacturing*, 10, 111–128.

Bakker, K. D., 2006. A step change in LNG operations through advanced process control. *23rd World Gas Conference.* Amsterdam, Netherlands.

Basseville, M. and Nikiforov, I. V., 1993. *Detection of abrupt changes: theory and application*, Prentice Hall, Englewood Cliffs, NJ.

Brook, D. and Evans, D. A., 1972. An approach to the probability distribution of CUSUM run length. *Biometrika,* 59, 539-549.

Capizzi, G. and Masarotto, G., 2003. An adaptive exponentially weighted moving average control chart. *Technometrics*, 45, 199-207.

Casella, G. and Berger, R. L., 2002. *Statistical Inference*, *2nd Ed*. Thomson Learning, Pacific Grove, CA.

Ceglarek, D. and Shi, J., 1996. Fixture failure diagnosis for auto body assembly using pattern recognition. *ASME Transactions, Journal of Engineering for Industry*, 118, 55–65.

Champ, C. W. and Woodall, W. H., 1987. Exact results for Shewhart control charts with supplementary runs rules. *Technometrics,* 29, 393-399.

Chua, M. K. and Montgomery, D. C., 1992. Investigation and characterization of a control scheme for multivariate quality control. *Quality and Reliability Engineering International,* 8, 37–44.

Cox, D. R. and Reid, N., 1987. Approximations to noncentral distributions. *The Canadian Journal of Statistics,* 15, 105-114.

Crosier, R. B., 1988. Multivariate generalization of cumulative sum quality control schemes. *Technometrics,* 30, 291-303.

Dai, Y., Luo, Y., Li, Z. and Wang, Z., 2011. A new adaptive CUSUM control chart for detecting the multivariate process mean. *Quality and Reliability Engineering International,* 27, 877-884.

Ding, Y., Ceglarek, D. and Shi, J., 2002a. Fault diagnosis of multi-station manufacturing processes by using state space approach. *ASME Transactions, Journal of Manufacturing Science and Engineering*, 124, 313–322.

Ding, Y., Shi, J. and Ceglarek, D., 2002b. Diagnosability analysis of multi-station manufacturing processes. *ASME Transactions, Journal of Dynamic Systems, Measurement and Control*, 124, 1–33.

Doganaksoy, N, Faltin, F. W. and Tucker, W. T., 1991. Identification of out-of-control multivariate characteristic in a multivariable manufacturing environment. *Communications in Statistics—Theory and Methods,* 20, 2775–2790.

Fenner, J. S., Jeong,, M. K., and Lu, J. C., 2005. Optimal automatic control of multistage production processes. *IEEE Transaction on Semiconductor Manufacturing System,* 18, 94-103.

Fenner, J. S., Jeong, Y. S., Jeong,, M. K. , and Lu, J. C., 2009. A Bayesian parallel site methodology with an application to uniformity modeling in semiconductor manufacturing. *IIE Transaction on Quality and Reliability*, 41(9), 754-763.

Golosnoy, V., Ragulin, S. and Schmid, W., 2009. Multivariate CUSUM chart: properties and enhancements. *Advances in Statistical Analysis,* 93, 263-279.

Hawkins, D. M., 1991. Multivariate quality control based on regression-adjusted variables. *Technometrics,* 33, 61–75.

Hawkins, D. M., 1993. Regression adjustment for variables in multivariate quality control. *Journal of Quality Technology,* 25, 170–182.

Hawkins, D. M. and Maboudou-Tchao, E. M., 2008. Multivariate exponentially weighted moving covariance matrix. *Technometrics*, 50, 155–166.

Hawkins, D. M. and Wixley, R. A. J., 1986. A note on the transformation of chi-squared variables to normality. *The American Statistician,* 40, 296-298.

Healy, J. D., 1987. A note on multivariate CUSUM procedures. *Technometrics,* 29, 409-412.

Jackson, J. E. and Bradley, R. A., 1961. Sequential χ^2 and T^2 tests. *Annals Mathematical Statistics,* 32, 1063-1077.

Jaehn, A. H., 1987. Zone control charts—SPC made easy. *Quality and Reliability Engineering International,* 26, 51-53.

Jeong, M. K., Lu, J. C., Huo, X., Vidakovic, B. and Chen, D., 2006. Wavelet-based data reduction techniques for fault detection. *Technometrics*, 48, 26-40.

Jiang, W., Shu, L and Apley, D. W., 2008. Adaptive CUSUM procedures with EWMA-based shift estimators. *IIE Transactions,* 40, 992-1003.

Jiang, W. and Tsui, K. L., 2008. A theoretical framework and efficiency study of multivariate statistical process control charts. *IIE Transactions,* 40, 650–663.

Johnson, N. L. and Kotz, S., 1970. *Continuous Univariate Distributions*. *Volume 2*. Houghton-Mifflin, New York.

Kallenberg, W. C. M., 1990. Inequalities for noncentral chi-square distributions. *Statistics & Probability Letters,* 9, 273–278.

Keene, O. N., 1995. The log transformation is special. *Statistics in Medicine,* 14, 811-819.

Khoo, M. B. C., 2003. Design of runs rules schemes. *Quality Engineering,* 16, 27–43.

Koutras, M. V., Bersimis, S. and Maravelakis, P. E., 2007. Statistical process control using Shewhart control charts with supplementary runs rules. *Methodology and Computing in Applied Probability,* 9, 207-224.

Lai, T. L., 1981. Asymptotic optimality of invariant sequential probability ratio tests. *The Annals of Statistics,* 9, 318-333.

Li, J., Jin, J. H. and Shi, J. J., 2008. Causation-based T^2 decomposition for multivariate process monitoring and diagnosis. *Journal of Quality Technology,* 40, 46–58.

Lim, T.-J. and Cho, M., 2009. Design of control charts with m-of-m runs rules. *Quality and Reliability Engineering International,* 25, 1085-1101.

Lorden, G., 1971. Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics,* 42, 1897-1908.

Lowry, C. A. and Montgomery, D. C., 1995. A review of multivariate control charts. *IIE Transactions,* 27, 800-810.

Lowry, C. A., Woodall,W. H., Champ, C. W. and Rigdon, S. E., 1992. A multivariate exponentially weighted moving average control chart. *Technometrics,* 34, 46-53.

Lucas, J. M. and Saccucci, M. S., 1990. Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics,* 32, 1-29.

Mahmoud, M. A., Woodall, W. H. and Davis, R. E., 2008. Performance comparison of some likelihood ratio-based statistical surveillance methods. *Journal of Applied Statistics,* 35, 783–798.

Mason, R. L., Tracy, N. D. and Young, J. C., 1995. Decomposition of T^2 for multivariate control chart interpretation. *Journal of Quality Technology,* 27, 99–108.

Mason, R. L., Tracy, N. D. and Young, J. C., 1997. A practical approach for interpreting multivariate T^2 control chart signals. *Journal of Quality Technology,* 29, 396–406.

Moustakides, G. V., 1986. Optimal stopping time for detecting changes in distributions. *Annals of Statistics,* 14, 1379-1387.

Moyne, J., Del Castillo, E. and Hurwitz, A.M., 2001. *Run-to-run control in semiconductor manufacturing*: CRC.

Murphy, B. J., 1987. Selecting out-of-control variables with T^2 multivariate quality procedures. *The Statistician,* 36, 571–583.

Montgomery, D. C., 2005. *Introduction to statistical quality control*, *5th ed.,* John Wiley & Sons, New York, NY.

Nishina, K. and Nishiyuki, S., 2003. False alarm probability function of CUSUM charts. *Economic Quality Control,* 18, 101-112.

Ou, Y., Wu, Z. and Tsung, F., 2012. A comparison study of effectiveness and robustness of control charts for monitoring process mean. *International Journal of Production Economics,* 135, 479-490.

Page, E. S., 1954. Continuous inspection schemes. *Biometrika,* 42, 243-254.

Patnaik, P. B., 1949. The noncentral χ^2 and *F* distributions and their applications. *Biometrika,* 36, 202-232.

Pignatiello, J. J. and Runger, G. C., 1990. Comparisons of multivariate CUSUM charts. *Journal of Quality Technology,* 22, 173–186.

Pollack, M., 1985. Optimal detection of a change in distribution. *Annals of Statistics,* 13, 206-227.

Reynolds, M. R., Jr. and Lou, J., 2010. An evaluation of a GLR control chart for monitoring the process mean. *Journal of Quality Technology,* 42, 287-310.

Riaz, M., Abbasa, N. and Doesb, R. J. M. M., 2010. Improving the performance of CUSUM charts. *Quality and Reliability Engineering International.* DOI: 10.1002/qre.1124.

Runger, G. C., 1996. Projections and the U^2 multivariate control chart. *Journal of Quality Technology,* 28, 313–319.

Runger, G. C., Alt, F. B. and Montgomery, D. C., 1996. Contributors to a multivariate statistical process control chart signal. *Communications in Statistics—Theory and Methods,* 25, 2203–2213.

Saxena, L. and Alam, K., 1982. Estimation of the non-centrality parameter. *Annals of Statistics,* 10, 1012-1016.

Shu, L. and Jiang, W., 2006. A Markov chain model for the adaptive CUSUM control chart. *Journal of Quality Technology,* 38, 135-147.

Shu, L., Jiang, W. and Wu, S., 2007. A one-sided EWMA control chart for monitoring process means. *Communications in Statistics-Simulation and Computation*, 36, 901-920.

Shu, L., Jiang, W. and Wu, Z., 2008. Adaptive CUSUM procedures with Markovian mean estimation. *Computational Statistics & Data Analysis,* 52(9), 4395-4409.

Sparks, R., 2000. CUSUM charts for signaling varying location shifts. *Journal of Quality Technology,* 32, 157-171.

Sparks, R., Carter, C., Graham, P., Muscatello, D., Churches, T., Kaldor, J., Turner, R., Zheng, W. and Ryan, L., 2010. Understanding sources of variation in syndromic surveillance for early warning of natural or intentional disease outbreaks. *IIE Transactions,* 42, 613-631.

Sullivan, J. H. and Jones, L. A., 2002. A self-starting control chart for multivariate individual observations. *Technometrics,* 44, 24–33.

Sullivan, J. H., Stoumbos, Z. G., Mason, R. L. and Young, J. C., 2007. Step-down analysis for changes in the covariance matrix and other parameters. *Journal of Quality Technology,* 39, 66–84.

Tan, M. H. and Shi, J., 2012. A Bayesian approach for interpreting mean shifts in multivariate quality control. *Technometrics,* 54(3), 294–307.

Tsui, K. L., Han, S. W., Jiang, W. and Woodall, W. H., 2012. A review and comparison of likelihood-based charting methods. *IIE Transactions,* 44, 724-743.

Wald, A., 1947. *Sequential Analysis*. John Wiley & Sons, New York, NY.

Wang, K. and Jiang, W., 2009. High-dimensional process monitoring and fault isolation via variable selection. *Journal of Quality Technology,* 41, 247–258.

Woodall, W. H. and Adams, B. M., 1993. The statistical design of CUSUM charts. *Quality Engineering,* 5, 559-570.

Woodall, W. H. and Mahmoud, M. A., 2005. Inertial properties of quality control charts. *Technometrics,* 47, 425-436.

Woodall, W. H. and Ncube, M. M., 1985. Multivariate CUSUM quality control procedures. *Technometrics,* 27, 285-292.

Wu, J., Jiao, J., Yang, M., Liu, Y. and Wang, Z., 2009. An enhanced adaptive CUSUM control chart. *IIE Transactions,* 41, 642-653.

Yashchin, E., 1987. Some aspects of the theory of statistical control schemes. *IBM Journal of Research and Development,* 31, 199-205.

Yashchin, E., 1995. Estimating the current mean of a process subject to abrupt changes. *Technometrics*, 37, 311-323.

Zhang, S. and Wu, Z., 2005. Designs of control charts with supplementary runs rules. *Computers and Industrial Engineering,* 49, 76-97.

Zhou, S., Chen, Y. and Shi, J., 2004. Statistical estimation and testing for variation root causes determination of multistage manufacturing processes. *IEEE Transactions on Robotics and Automation*, 1, 73–83.

Zhou, S., Ding, Y., Chen, Y. and Shi J., 2003. Diagnosability study of multistage manufacturing processes based on linear mixed-effects models. *Technometrics*, 45, 312–325.

Zou, C., Jiang, W. and Tsung, F., 2011. A LASSO-based diagnostic framework for multivariate statistical process control. *Technometrics*, 53, 297–309.

Zou, C. and Qiu, P., 2009. Multivariate statistical process control using LASSO. *Journal of the American Statistical Association*, 104, 1586–1596.