PROCESS MONITORING AND CONTROL WITH HIGH-DIMENSIONAL DATA

by

SANGAHN KIM

A dissertation submitted to the

School of Graduate Studies

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

Myong K. Jeong and Elsayed A. Elsayed

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

MAY, 2018

ABSTRACT OF THE DISSERTATION

PROCESS MONITORING AND CONTROL WITH HIGH-DIMENSIONAL DATA

By SANGAHN KIM

Dissertation Directors:

Myong K. Jeong and Elsayed A. Elsayed

The increased accessibility of a large number of data streams makes it possible to use multivariate statistical process control (SPC) in various modern industries. However, as the number of quality characteristics and process parameters to be monitored increases such simultaneous monitoring becomes less sensitive to the out-of-control signals especially when only a few variables are responsible for abnormal situations or changes in the processes output. This dissertation proposes several efficient statistical process control methodologies for monitoring high-dimensional processes such as chemical production processes, semiconductor manufacturing processes and liquefied natural gas (LNG) processes. More specifically, we investigate and develop methodologies for monitoring the shifts in the means of the quality characteristics under sparsity. It is intended to detect these shifts as soon as they occur regardless of their magnitudes.

We first investigate approaches for monitoring high-dimensional processes with an application of the multistage process. Due to the properties of multistage processes such as variance propagation and the specific structure of correlation, monitoring such multistage

processes becomes much more challenging. We introduce a fault diagnosis procedure-integrated SPC chart for monitoring a multistage process, especially with beta-distributed output variables by adopting a model-based approach. For fault diagnosis, we propose a partial regression-based variable selection (VS) approach to choose several "suspicious" variables that might be regarded as the ones causing out-of-control signals. This approach is effective and its performance is compared with other VS-based charts such as forward variable selection approach.

Second, we consider that some high-dimensional processes have grouped patterns of the data structure. In other words, the quality characteristics' pattern can be grouped by the relevance or the correlation structure of these characteristics. Moreover, these processes would possibly shift by changes only in a few relevant quality characteristics. The multistage process is another example where the stages can be considered as groups and a few variables in a group may shift together. In this case, exploiting the grouped patterns would provide more advantages in the VS than choosing the variables individually. Therefore, we develop a sparse group variable selection approach to reflect the grouped behavior of the process shift. We modify the selection procedure appropriately to implement sparsity within a group and between groups. Extensive simulation studies are conducted to demonstrate the numerical performance of the proposed method.

Third, we consider the cases where quality characteristics (or process parameters) are strongly correlated and introduce small size of shifts. In complex modern industries, the highly correlated data structure is present in numerous applications such as monitoring

spatially correlated data streams, surveillance of wafer surface and monitoring connected job-shop manufacturing processes in chemical plants. In such processes, the existing VS-based charts including the proposed partial regression based chart suffer from the detection of small process changes since the strong correlation often confuses the correct selection of the faulty variables. Therefore, we introduce a ridge penalized likelihood in order to improve the efficiency in monitoring processes when small process shifts occur in highly correlated data structures. Accurate probability distributions of the monitoring statistics under null and alternative hypotheses corresponding to in-control and out-of-control situations, respectively, are obtained. In addition, we investigate several theoretical properties of the proposed scheme and present further extensions of the proposed methods to other existing methods. We demonstrate the performance of the proposed chart theoretically and empirically.

Fourth, we investigate a new approach for change detection by utilizing the correlation information among variables. While a traditional multivariate exponentially weighted moving average (MEWMA) chart is an extension of univariate EWMA, we develop a generalized model for the MEWMA that uses appropriate non-diagonal elements in the smoothing matrix based on the correlation among variables. We offer the interpretation of the relationship between the correlation and the non-diagonal elements of the smoothing matrix. We also suggest an optimal design for a proposed method and compare the proposed chart fairly with existing EWMA-based charts.

Finally, we develop a new method for monitoring high-dimensional processes based on the Bayesian approach. The approach sequentially updates a posterior distribution of the process parameter of interest through the Bayesian rule. In particular, a sparsity promoting prior distribution of the parameter is applied properly under sparsity, and is sequentially updated in online processing. A data-driven Bayesian hierarchical model enables the monitoring scheme to be effective to the detection of process shifts and improves the efficiency of the computational complexity in the high-dimensional processes. Comparisons with recently proposed methods for monitoring high-dimensional processes demonstrate the superiority of the proposed method in detecting small shifts. In addition, graphical presentations in tracking the process parameter provide information about decisions regarding whether a process needs to be adjusted before it triggers the alarm.

ACKNOWLEDGEMENTS

I would like to acknowledge and present my sincere thanks to my advisors Professor Myong K. Jeong and Professor Elsayed A. Elsayed for their excellent guidance and encouragement during my dissertation and research. I am very honored to be given the opportunity to have worked with them and been accepted as a member of their academic family. Each of the members of my Dissertation Committee has provided me extensive personal and professional guidance and taught me a great deal about both scientific research and life in general. I would like to thank Professor Susan Albin and Professor Ying Hung who have helped me beyond their expected role. I would also like to thank Professor Melike Baykal-Gursoy, who has been supportive of my career goals and provided me with the protected academic time to pursue those goals.

I am grateful to all of those with whom I have had the pleasure to work during this and other related works. All my friends in Rutgers University have enabled me to enjoy the life of Ph.D., and have shared joys and sorrows. I would like to especially thank my faithful friend, Mehmet Turkoz, for his brotherhood and friendship, for all sleepless nights we were working together, for all the fun we have had in the last five years. I would also like to thank my labmates Mejdal Alqahtani and Jason Baek.

Nobody has been more important to me in the pursuit of this dissertation than the members of my family. My sincere thanks go to my parents, whose constant love and guidance are with me in whatever I pursue. I would like to also thank my brother, Ho-nam, my sister, Mee-kyoung and sister-in-law, Yong-ki for their endless support, help and trust in me. I

would like to extend my gratitude to my friends in Korea, especially Jong-sun, Hyun-suk and Jun-yong.

Most importantly, I wish to thank my loving and supportive wife, Minji, and my wonderful daughter, Anna, who provide unending inspiraiton. This work would not have been possible without their constant support, love and sacrifice. I am also indebted to my parents-in-law in Busan, Korea, who have been sending their unquestioning trust and support.

Finally, I send my special thank to myself, Sangahn. You did a good job and will do better.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1. Motivation of the Work

In modern era, various fields of industries have demanded a great deal of quality improvements. Controlling and monitoring quality characteristics have become significant in business strategy for manufacturers, communication companies, services organizations and health care providers. Statistical process control (SPC) techniques have been widely adopted to monitor the quality characteristics that are crucial for the final products or services. In modern industries, the environment of the manufacturing processes becomes more complex and costly. Accordingly, a number of process monitoring methodologies have been developed for practical purposes to ensure the quality of the products and services.

The control chart is one of the most commonly used techniques for process monitoring. When sources of variability occur, a statistical criterion is calculated based on the observations and is plotted to determine if it falls outside the control limits of the chart. This is a signal that should be investigated and corrective actions may be required to remedy the abnormality. The main purpose of the control chart is to detect assignable causes which are distinguished from common causes inherently embedded in the process as quickly as possible, and to take actions to remove the root cause of process disturbances.

A number of control charts such as Shewhart, exponentially weighted moving average (EWMA), cumulative sum (CUSUM) and their various extensions have been developed (Shewhart, 1931, Page, 1954, Page, 1961, Crowder, 1987, Crowder, 1989, Lucas and Saccucci, 1990, Gan, 1991, Hawkins, 1991, Hawkins, 1993, Woodall and Adams, 1993, Hawkins and Olwell, 1998). Shewhart introduces a statistical method to monitor the process mean change when dealing with single quality characteristic, for say, univariate process. While Shewhart chart is a sequential hypothesis testing with an individual observation independently at each sampling epoch, EWMA and CUSUM integrate the information of previous samples.

The conventional univariate SPC charts are further extended to multivariate processes where the process involves multiple quality characteristics to be monitored. A standard approach for multivariate process monitoring and surveillance is to monitor all of quality characteristics simultaneously by taking correlation into consideration (Bersimis *et al.*, 2007) such as Hotelling's chi-square, multivariate exponentially weighted moving average (MEWMA) and multivariate CUSUM charts (Crosier, 1988, Pignatiello and Runger, 1990, Lowry *et al.*, 1992, Ngai and Zhang, 2001) which are analogous to Shewhart, EWMA and CUSUM in the univariate process monitoring case.

However, as the number of quality characteristics (e.g. process variables, parameters) increases, their simultaneous monitoring becomes less sensitive to the out-of-control signals, which makes the monitoring processes much more challenging. It is called 'curse of dimensionality' (Jiang and Tsui, 2008, Wang and Tsung, 2008, Zhu and Jiang, 2009). A

number of high-dimensional processes that motivate this dissertation is found in the data-rich modern industry such as industrial surveillance via image processing and monitoring of multistage manufacturing processes (Wilcox *et al.*, 2014). Image processing, for example, receives a great attention in various areas such as banks, military and civilian applications. Close-circuit television (CCTV) systems generate thousands image frames to be monitored for various purposes. Machine vision systems (MVS) are used in industrial applications to provide information of product geometry, surface detects. Other examples include multistage processes which commonly exist in process industries with wide applications in the chemical engineering processes such as refineries, liquefied natural gas (LNG) processing and batch production. It can be readily speculated that the number of quality characteristics that should be monitored would geometrically increase as the number of stages or the number of variables in each stage increases. Monitoring such high-dimensional data has become a crucial issue in quality improvements because it is often inhibited from adopting conventional SPC techniques in practice. It has been shown in the literature that the run length to detect abnormal signals increases as the dimensionality increases. In this dissertation, we intend to investigate the monitoring of high-dimensional processes and introduce several methodologies for such processes.

1.2. Problem Description and Assumptions

It is crucial and challenging to monitor such high-dimensional processes especially when only a few variables are responsible for abnormal situations or changes in the processes output because the leading sources are readily concealed by the noises, which results in

monitoring much more challenging (Fan and Li, 2006, Wang and Jiang, 2009, Jeong *et al.*, 2006). As the number of quality characteristics increases, the probability that process mean shifts occur in a large number of variables simultaneously would be rather small, which is called ‘sparsity’ (Friedman *et al.*, 2001, Wang and Jiang, 2009, Zou *et al.*, 2011). Throughout this dissertation, it is practically assumed that only a few variables cause the process out of control simultaneously.

Moreover, we assume that the original data or at least the transformed dataset follows multivariate normal distribution with given parameters. Even though the SPC charts can be applied to historical dataset (Phase I) or to online monitoring (Phase II), we intend to concentrate only on monitoring in Phase II rather than Phase I.

In monitoring and control strategies for quality improvement, several quality characteristics can be monitored such as quantifiable process parameters (e.g., mean, variance and moments) and classifiable attributes (e.g. conforming and nonconforming). In this dissertation, the scope of monitoring is limited to process mean and the terminologies ‘quality characteristic’, ‘quality feature’, ‘process parameter’ and ‘(output) variable’ are used interchangeably as a ‘process mean’. Moreover, only a single observation is sampled and monitored at each sampling epoch.

1.3. Approaches for Monitoring High-Dimensional Processes

1.3.1. A Fault Diagnosis Incorporated Process Monitoring and Control

With the sparsity assumption, it is more efficient and convenient to reduce the dimension because most of the variables are not practically responsible for the process change. Several schemes for monitoring high-dimensional processes considering dimension reduction have been developed as shown in the literature (Runger, 1996, Wang and Jiang, 2009, Zou and Qiu, 2009, Capizzi and Masarotto, 2011, Jiang *et al.*, 2012). Hence, several conventional fault identification procedures as a way of dimension reduction such as decomposition of T^2 and various step-down procedures would be possibly considered prior to applying control charts (Kim *et al.*, 2016). We review several monitoring schemes adopting dimension reduction methods in Chapter 2.

We first investigate approaches for monitoring high-dimensional processes with an application of the multistage process. The variable selection based on partial regression to choose several "suspicious" variables that might be regarded as the ones causing out-of-control signals is applied as a dimension reduction technique. Due to the properties of multistage processes such as variance propagation, monitoring such multistage processes becomes much more challenging. In spite of the wide applications around beta distributed process output, no charting scheme has been developed due to the complexity of the problem. In Chapter 3, we propose a model-based approach to obtain output variables to be monitored by adopting beta regression, and its performance is compared with other variable selection-based charts such as forward variable selection approach.

Although the variable selection based procedure performs generally well, the performance may deteriorate when the data exhibit group patterns and when the shifts tend to occur in a grouped pattern fashion as observed in many industries. In Chapter 4, we propose a sparse group variable selection-based approach to implement 'between groups' and 'within-a-group' sparsity. The performance of the method is compared with individual variable selection-based charts under the certain conditions of group patterns of data and group behavior of the process change.

1.3.2. A Ridge Penalized Likelihood-Based Process Control Chart

The variable selection based charts studied in Chapters 3 and 4 have their limitations upon adopting diagnosis procedures. Even if those charts perform well in many cases, they have still challenges. For example, when small shifts occur in process parameters, the capability of variable selection would possibly decrease resulting in poor performance of detection. Accordingly, in Chapter 5 we investigate a new way of monitoring high-dimensional processes by introducing L_2 regularization in calculating likelihood ratio test statistic, called 'ridge regularization'. The ridge penalized likelihood based chart does not select variable, rather 'shrinks' the variables toward zero.

Based on literature about ridge regression, we provide an effective approach for shrinkage of the process variables in SPC, even though the ridge penalty does not reduce dimensionality. Since the shrinkage is coherently related to the correlation among variables,

the proposed scheme offers several theoretical properties of the charting statistic with respect to the association between correlation and shrinkage.

The closed form of the solution of ridge-based monitoring statistic allows computation time in online monitoring to be substantially effective, and it enables the user to have approximate probability distributions of the monitoring statistics under null and alternative hypotheses corresponding to in and out-of-control situations, respectively. In addition, we obtain several theoretical properties of the proposed ridge regression approach, and present further extensions to other existing methods. Based on the properties of ridge, we show that the proposed chart improves the efficiency of monitoring small process shifts compared to existing dimension reduction-based charts, especially in the cases where quality characteristics are strongly correlated. Finally, we demonstrate the performance of the proposed chart through both theoretical and empirical approaches.

1.3.3. Generalized Smoothing Parameters of a Multivariate EWMA Control Chart

The ridge penalty shrinks the coefficient of variables proportional to their correlations and the performance of the chart can be explained in terms of the 'utilization' of the correlation. While a traditional MEWMA chart is a simple extension of univariate EWMA, we develop a generalized model for the MEWMA that uses appropriate non-diagonal elements in the smoothing matrix based on the correlation among variables in Chapter 6.

We offer the interpretation of the relationship between the utilization of correlation and the non-diagonal elements of the smoothing matrix. We also suggest an optimal design for the proposed method by constructing the optimization problem based on the correlation of the data. Finally, we compare the performance of the proposed chart through extensive simulation studies with other recent variants of an MEWMA control charts and illustrate the chart via a real-life case study of an on-line monitoring of dimensions of bolts using an image processing system.

1.3.4. Bayesian Sequential Update-Based Control Chart

The penalized likelihood based approaches including VS-based methods introduced in Chapters 3 and 4 and ridge penalty-based method introduced in Chapter 5 employ a fixed constant for the penalty parameter. The penalty parameter with L_0 or L_1 norm type penalty function is associated with the sparsity in VS-based methods, and that with L_2 norm type penalty function is related with the level of shrinkage. Since the information of the process shift is generally unknown before operating the process, fixing the penalty as a constant would probably not guarantee the good performance of the detection. In addition, in existing VS-based methods, the result of the VS does not affect the selection to the next VS, since VS procedures are carried out at every sampling point independently of the previous result of the selection. Thus, these methods may not function well especially when small shifts occur because VS may select different variables from one sampling point to another in the out-of-control situations.

In Chapter 7, we develop a new method based on the Bayesian sequential update in which the prior probability for the current process parameter is updated over the sampling period. Specifically, a Bayesian theorem is applied to update the prior distribution of the mean at every sampling point sequentially, which considers the result of previous VS. Moreover, a data-driven Bayesian hierarchical model is utilized for determining the distribution of the scale parameter of the prior results in a dynamic change of the penalty over time. This also enables the penalty values to be different for all variables while existing methods have equal and constant penalties. Compared with the penalized likelihood based methods including VS-based methods, we demonstrate that the proposed monitoring scheme is effective in detecting process shifts and efficient in the computational complexity in the high-dimensional processes. The proposed method is applied to the well-known Tennessee Eastman Industrial Challenge Problem (TE problem) which is formulated as a realistic simulation model of a chemical plant.

1.4. Dissertation Outline

This dissertation is organized as follows. Chapter 2 reviews the literature related to conventional multivariate SPC charts focusing on high-dimensional processes. We also review typical methodologies for monitoring multistage processes. Chapter 3 proposes an individual VS-based chart via partial regression and applies it to monitoring multistage processes where the output variables are distributed between 0 and 1. Chapter 4 extends the work in Chapter 3 based on a sparse group variable selection in that the behavior of the multistage process would possibly occur in a grouped pattern. In Chapter 5, we present a

penalized likelihood based multivariate statistical process control chart via ridge regularization. Chapter 6 studies a generalized MEWMA approach by assigning the appropriate non-diagonal elements in the smoothing matrix based on the correlation among variables. Chapter 7 develops a Bayesian sequential update based high-dimensional process monitoring method. Finally, in Chapter 8 we discuss the conclusions and the future research topics.

CHAPTER 2

LITERATURE REVIEW

In this chapter, we provide a comprehensive review of work related to the research being investigated in this dissertation. We begin by presenting relevant research for monitoring and controlling quality in multistage processes. We then review the research related to monitoring of high-dimensional processes, followed by work related to monitoring of the process with small process changes. We describe the methodologies, their advantages and limitations.

2.1. Methodologies for Monitoring Multistage Processes

Multistage processes commonly exist in process industries such as oil refineries, LNG processing and batch manufacturing processes as shown in Figure 2.1. Monitoring and controlling such processes are based on the engineering process control (EPC) methods which require a process model that relates process output to process input, drift, and random noise.

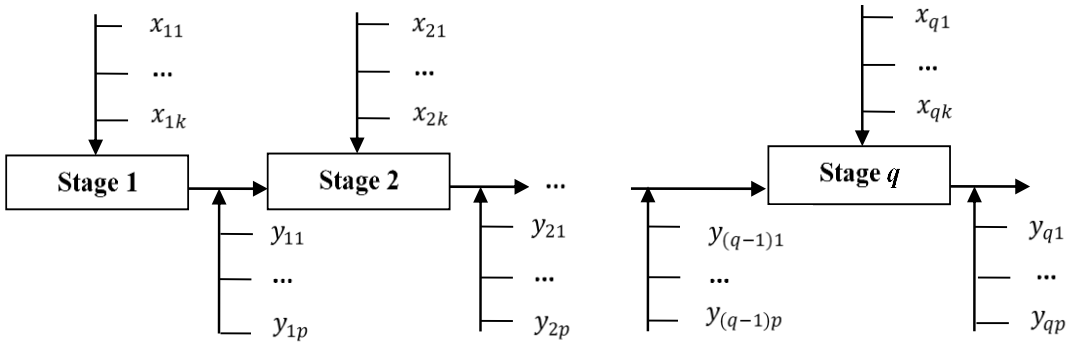


Figure 2.1. An example of multistage processes

Variation propagation from one stage to the next is a major concern in such processes. Sources of output variability can be attributed to technological advances such as the choice of process technology and process automation systems or operational factors such as operator's reaction to output variability at each stage and operational procedures (Rajaram and Robotis, 2004). Thus, monitoring and diagnosis of multistage processes are important and challenging due to both their complexity and the cascade property (outputs from operations at upstream stages may affect the quality of downstream stages), and product variability may propagate throughout the production stages (Kim *et al.*, 2017a). Moreover, it can be readily seen that the number of quality characteristics (parameters of the processes) geometrically increases as the number of stages or the number of characteristics in each stage increases.

Suppose a process has a total of *q* stages as shown in Figure 2.1. A vector $\mathbf{y}_n$ is the quality characteristic measurements to be monitored at the n^{th} stage, and $\mathbf{x}_n$ is an input information characterizing these quality characteristics. In a linear state space model, one of the popular methods to model multistage processes, the n^{th} quality characteristic in the normal process is formulated as (Shi, 1999)

$$\mathbf{x}_n = \mathbf{A}_{n-1}\mathbf{x}_{n-1} + \mathbf{w}_n$$
$$\mathbf{y}_n = \mathbf{C}_n\mathbf{x}_n + \mathbf{v}_n$$

where $\mathbf{v}_n$ is a measurement error for the product quality and $\mathbf{w}_n$ is a common-cause variation which is a process noise. The matrix $\mathbf{C}_n$ relates the quality information $\mathbf{x}_n$ and the measurement $\mathbf{y}_n$, and $\mathbf{A}_{n-1}$ associates the quality information from stage $n-1$ to stage *n.* Both $\mathbf{C}_n$ and $\mathbf{A}_n$ are determined by the information from the engineering knowledge and experience. Based on physics and chemical properties as well as the engineering knowledge of quality information, a variety of literature adopts a linear state space model for modeling multistage processes (Agrawal *et al.*, 1999, Shi, 1999, Lawless *et al.*, 1999, Djurdjanovic and Ni, 2001, Ding *et al.*, 2002, Huang *et al.*, 2002, Zhou *et al.*, 2003).

One may consider to monitoring the quality of the final product in multistage processes by using traditional SPC charts such as Hotelling T^2, MEWMA and MCUSUM. However, in modern complex multistage processes, it might lose its monitoring and diagnosis ability

because it ignores the stage-dynamic nature of the data due to the cascading property of the stages. In order to overcome this issue, Xiang and Tsung (2008) and Zou and Tsung (2008) monitor the one-step ahead forecast error (OSFE) for each stage based on the recursive formulations by adopting time series analysis; see Durbin and Koopman (2012). Note that the OSFEs are independently and identically distributed (i.i.d.) among stages. Thus, conventional SPC techniques can be applied to monitor the OSFEs. Zantek *et al.* (2006) apply the 2*N* individual multivariate control charts by adopting one-sided CUSUM charts for each OSFE. Xiang and Tsung (2008) develop a group EWMA scheme by computing maximum of *N* EWMA statistics for each OSFE. A maximum likelihood estimation procedure is used for phase I analysis. Based on the estimators from phase I, they derive OSFE and make use of them to obtain the monitoring statistic in EWMA scheme. They applied EWMA scheme for individual OSFE and used a Bonferroni-type control limit for the out-of-control signals.

Different from the state space modeling, a model-based approaches known as regression adjustments, is suggested by Hawkins (1991, 1993). These approaches mainly consider cases where output variables are regressed on multiple input variables and preceding output variables as another input (Zhang, 1984, Shu *et al.*, 2004a). Rao *et al.* (1996) investigate regression adjustments in multistage processes under a Bayesian framework. Shu *et al.* (2004b) study the run length distribution of regression control charts with the assumption of the parameters' uncertainty. Considering the collinearity between stages, cause-selecting charts are applied in Zhang (1984, 1985, 1990, 1992) and Wade and Woodall (1993). Zantek *et al.* (2006) propose a method to monitor the residuals of stages through

simultaneous CUSUM charts, and Asadzadeh *et al.* (2008) propose the regression-adjusted CUSUM chart for multistage SPC with censored data. In general, the multiple cause-selecting charts for a multistage process monitor normally distributed residuals based on the ordinary least square approach. However, some data may not follow normal distributions and may violate the equal variance assumption.

Jearkpaporn *et al.* (2003) extend the multiple cause-selecting charts to non-normally distributed data by replacing the multiple linear models with the generalized linear model. In particular, they first derive the deviance residual monitoring schemes for gamma distributed data. Then the derived deviance residuals approximately follow normal distribution. Jearkpaporn *et al.* (2005) also focus on the contaminated data with outliers. Robust GLM with Huber's score function (to identify outliers) is utilized for estimating the parameters of gamma distribution. The deviance residuals from the robust fittings are then used to monitor the two-stage processes. Furthermore, Jearkpaporn *et al.* (2007) consider the three stage processes where the input variables follow uniform distribution and the output variables follow gamma, Poisson, and normal distributions. They show that the deviance residual charts provide out-of-control average run length (ARL) smaller than Shewhart charts and Hotelling's charts.

Although multistage processes with beta distributed data are widely available in industry (Cribari-Neto and Zeileis, 2009), there are no available multistage SPC procedures that monitor and control multistage processes when the output variables follow beta distribution, which motivates our research in Chapter 3.

2.2. Methodologies for Monitoring High-Dimensional Processes

The increased accessibility of large amounts of data stream encourages the use of multivariate statistical process control in various modern industries. Due to the complicated data structure, simultaneous monitoring of several process' parameters or product's characteristics has become considerably crucial. However, as the number of quality characteristics, i.e. process means, to be monitored increases in complex processes and simultaneous monitoring of process means becomes less sensitive to the out-of-control signals, which makes the monitoring processes much more challenging. In particular, when only a few variables are responsible for abnormal situations, the assignable causes are easily concealed by the noises (Fan and Li, 2006, Zantek *et al.*, 2006). Since the probability that process mean shift occurs in a large number of variables simultaneously would be rather small as the number of quality characteristics increases, called 'sparsity'; it is natural to assume that only a few mean shifts occur simultaneously (Friedman *et al.*, 2001, Wang and Jiang, 2009, Zou *et al.*, 2011). Based on the sparsity assumption, several SPC charts have been developed for monitoring both means and variances of the quality characteristics.

It is intuitive to reduce the dimension and forage a parsimonious model for sensitive screening. Hawkins (1991, 1993) and Runger (1996) demonstrate focusing on the specific assignable causes is more sensitive than monitoring the entire set of variables in a full dimension case. Hence, it is natural that many of studies for monitoring high-dimensional processes reasonably consider dimension reduction as their first choice. Several conventional fault identification procedures such as decomposition of T^2 and various step

down procedures would be possibly considered prior to applying control charts (see details about fault identification methodologies in (Hawkins, 1991, Mason *et al.*, 1995, Mason *et al.*, 1997, Sullivan *et al.*, 2007, Li *et al.*, 2008, Zhu and Jiang, 2009, Kim *et al.*, 2016)). Recently, several variable selection (VS) based control charts have been developed and demonstrated that those charts function well in monitoring high-dimensional processes with a quick detection of the process shift. Wang and Jiang (2009) apply the forward variable selection (FVS) algorithm for dimension reduction and Zou and Qiu (2009) propose another VS-based control chart by adopting least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996, Zou, 2006, Wang and Leng, 2007). Jiang *et al.* (2012) extend the work of Wang and Jiang (2009) by adopting EWMA with the same underlying procedure of FVS. Capizzi and Masarotto (2011) develop a combination of the least angle regression (LAR) (Efron *et al.*, 2004) with MEWMA for monitoring not only the mean but also the variability. More variants of VS-based methodologies and relevant discussions can be found in Peres and Fogliatto (2018), Abdella *et al.* (2017), Weese *et al.* (2016), Capizzi (2015), Woodall and Montgomery (2014), Mehmood *et al.* (2012). Taking advantages of VS-based monitoring scheme, we investigate another VS method based on partial linear regression and apply it to beta distributed multistage process monitoring in Chapter 3. In a special case where a grouped pattern of the process is observed, traditional VS-based chart may not perform well, which motivates another research introduced in Chapter 4.

These methods are effective in monitoring high-dimensional processes under the sparsity assumption. However, they have several challenging issues. For example, when the shift size is relatively small, diagnosis procedures may be misleading and may result in poor

performance in detecting process changes. Consequently, if the diagnosis procedure malfunctions, the performance of detection might deteriorate. Therefore, we intend to investigate a different monitoring scheme via ridge penalized likelihood for detecting small shifts as discussed in Chapter 5.

2.3. Methodologies for Monitoring Processes in Detecting Small Process Changes

The EWMA control chart provides better performance in detecting small shifts by adjusting the smoothing parameter according to the shift size (Roberts, 1959). For multivariate extension, Hotelling's T^2 is analogous to Shewhart control charts and shows similar properties in detecting changes in the mean shift. The MEWMA control chart by Lowry *et al.* (1992) is also analogous of the univariate EWMA control chart. Lucas and Saccucci (1990) provide discussion of the properties and extension of EWMA. Prabhu and Runger (1997) provide a table of the average run lengths for various smoothing parameters and the number of variables for an MEWMA chart by using a Markov chain approach as developed by Brook and Evans (1972). Jensen *et al.* (2006) state the necessity of an accurate parameter estimation for MEWMA, and Mahmoud and Maravelakis (2010) implement an MEWMA chart and suggest an optimal strategy for its design.

Variations and analyses of the MEWMA chart have been conducted to improve its performance. Lee (2010) attempts to apply the variation of the sample size, and to show that the adaptive-sample-size MEWMA chart is more efficient than the fixed-sample-size chart and provides the optimal sample size through the Markov chain approach.

Laungrungrong *et al.* (2011, 2014) develop EWMA for multivariate Poisson-distributed data and compare its performance with the traditional MEWMA based on normal-theory control limits. They demonstrate that MEWMA with Poisson variables tends to increase the false alarm rate when the mean of Poisson variables exceeds a certain value. Shamma *et al.* (1991) and Shamma and Shamma (1992) propose a double EWMA (DEWMA) control chart by performing exponential smoothing twice. Zhang *et al.* (2003) apply DEWMA for Poisson data, and Zhang and Chen (2005) and Mahmoud and Woodall (2010) evaluate the performance of the DEWMA chart. Alkahtani and Schaffer (2012) develop a double MEWMA (dMEWMA) control chart analogous to the univariate DEWMA chart. They present several numerical examples according to the various sizes of a smoothing parameter and the number of variables, and show that its out-of-control average run lengths (ARL_1) outperform the conventional MEWMA chart.

Yumin (1996) suggests adopting different values of diagonal elements for the chart smoothing matrix and also develops a methodology for obtaining the optimal smoothing parameters for each variable according to the mean shift scales. Huh *et al.* (2013) demonstrate the benefit of using different smoothing parameters in bivariate EWMA through various simulations and compare its performance to several MEWMA schemes using ARL_1 as a performance measure; they show that the model performs well in particular situations when several variables have a direct effect on the process mean shift. Hawkins *et al.* (2007) propose a full smoothing matrix EWMA (FEWMA) which has non-zero off-diagonals. They set up off-diagonals with equal values having a certain ratio with the diagonal elements and demonstrate the benefit of FEWMA through a practical example.

However, this model is limited due to the assumption of equal values for the off-diagonal elements in the smoothing matrix. Moreover, they do not provide a methodology for determining the off-diagonal elements. Therefore, we intend to investigate a different monitoring scheme by utilizing the off-diagonal elements in smoothing matrix in Chapter 6.

2.4. Methodologies for Monitoring Processes with Sequential Information

It is very common in transactional, manufacturing, and service processes that individual measurements are the only available information. Since the process parameters of interest to be monitored may not be well estimated and may have much variability, the individual measurements may lead to the poor capability of detection of the small shifts in the process. A number of variants of MEWMA (see Section 3.2) and multivariate cumulative sum (MCUSUM) charts have been studied to accommodate the past measurements to reduce the variability and to demonstrate the capability in detecting small shifts of the process (Crosier, 1988, Kim *et al.*, 2017b, Pignatiello and Runger, 1990). From the prediction point of view, these procedures can be interpreted that the process parameter at a certain sampling point is estimated based on the sequential measurements (Montgomery, 2007).

Bayesian methods is a theoretical approach that utilizes probabilistic derivations of the posterior distribution of the process parameter in a sequence of the samples. In the univariate case, Crowder and Eshleman (2001), Tsiamyrtzis and Hawkins (2005), Triantafyllopoulos (2007), and Apley (2012) develop Bayesian methods for updating the

posterior distribution $\mu_t \mid \tilde{x}_t$, where $\tilde{x}_t$ is a set of all measurements, $x_1, ..., x_t$. Crowder and Eshleman (2001) assume random walk model such that $\mu_t = \mu_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim N(0, \sigma^2)$. Tsiamyrtzis and Hawkins (2005) consider the random walk with a nonzero mean occurs with a certain probability. Triantafyllopoulos (2007) further assumes the uncertainty of variance of ε_t; and Apley (2012) provides graphical displays of the mean to monitor the process in Phase II. Consequently, they explore the consecutive estimates of the process parameter, μ and its density in the high speed monitoring of the process. The Bayesian framework of these methods can be similarly extended to the multivariate parameter estimation. For example, Kalman filtering is a widely used technique to estimate the process parameter in a state space model, which has the Bayesian interpretations in that the posterior distribution of the multivariate parameter, $\boldsymbol{\mu}$ is sequentially updated (Apley, 2012, Durbin and Koopman, 2012). Further, Bayesian framework developed for economically optimal decision policies is also found in Nikolaidis and Tagaras (2017), Veeravalli and Banerjee (2014), and Makis (2008).

Although these approaches provide good sequential estimators of the process parameters of interest, they are insufficient to be applied in the high-dimensional process monitoring under sparsity and when the quick detection of the small process change is primarily of importance. Therefore, we intend to develop a monitoring scheme based on the Bayesian approach to sequentially update the process parameter considering the sparsity in high-dimensional processes as shown in Chapter 7.

CHAPTER 3

VARIABLE SELECTION-BASED MULTIVARIATE SPC IN MULTISTAGE PROCESSES

3.1. Introduction

Typical manufacturing processes involve multiple stages of production such as those found in pharmaceutical manufacturing, chemical industry and semiconductor manufacturing. Conventional SPC techniques have been widely applied in process monitoring to detect process mean shift or variance changes (Nomikos and MacGregor, 1994, Nomikos and MacGregor, 1995, MacGregor and Kourti, 1995, Doan and Srinivasan, 2008). However, they are not effective in monitoring multistage processes since such processes display a great degree of output variability due to the propagation and amplification of the variability through the stages. When a number of stages are concatenated sequentially in the process, the variance propagation effect becomes significant when variability is initiated in the upstream stages. Moreover, as the number of stages and the number of quality features (process variables) to be monitored increases, the dimensionality of the monitoring parameters increases geometrically. Thus, conventional SPC techniques are not effective in monitoring such high-dimensional multistage processes.

Statistical process monitoring techniques for multistage processes have received significant interest from different investigators (Asadzadeh *et al.*, 2008, Shi and Zhou, 2009, Fenner *et al.*, 2005) with some limited studies that deal with high-dimensional multistage processes. A cause-selecting chart based on regression adjustment method is

proposed to model two-stage processes using simple linear regression, and monitors the residuals between two consecutive stages (Hawkins, 1991, Hawkins, 1993). The multiple cause-selecting charts for a multistage process generally monitor normally distributed residuals based on the ordinary least square approach. However, the assumption of normality may be violated in many processes, and may also violate the assumption of equal variance for production stages. In this case, the multiple linear models are replaced with the generalized linear model proposed by Jearkpaporn *et al.* (2003, 2005, 2007).

When practitioners consider multiple linear regression models, they commonly face response variables restricted to the interval (0, 1) such as proportions which can be described by a beta distribution. In Prater's gasoline-proportion data, for example, the restricted response variable is the proportion of crude oil converted to gasoline after distillation and fractionation, whereas the explanatory variables are: the crude oil gravity (degrees API); the vapor pressure of the crude oil (lbf/in^2); the crude oil 10% point ASTIM (i.e. the temperature at which 10% of the crude oil becomes vapor); and the temperature (^oF) at which all the gasoline is vaporized (Prater, 1956, Ferrari and Cribari-Neto, 2004). In this case, beta regression can be considered for the restricted response variable. Furthermore, it is common to use beta distribution to describe the output variable in multistage processes such as the proportion of pentane in the multistage oil-refinery process as shown in Figure 3.1 (Kim *et al.*, 2017c).

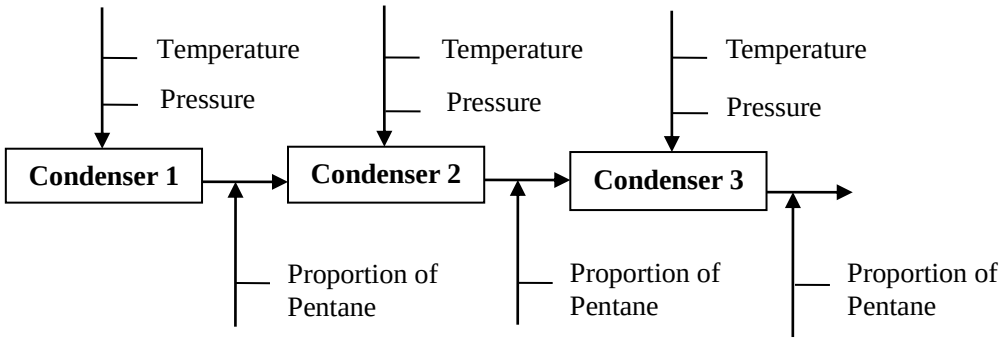


Figure 3.1. A multistage oil refinery process

Although multistage processes with beta distributed data are widely available in industry (Cribari-Neto and Zeileis, 2009), there are limited studies in monitoring multistage SPC procedures when the output variables follow beta distribution because the generalized linear models are not applicable for the estimation of beta parameters (Schmidt, 2002, Ferrari and Cribari-Neto, 2004). Recently, Hwang *et al.* (2014) and Kim *et al.* (2017c) developed monitoring the multistage processes with beta-distributed output variables. As stated, however, the investigation of effective process monitoring and control in high-dimensional multistage processes is still limited. In this chapter, we develop an efficient SPC chart for monitoring high-dimensional multistage processes with beta-distributed output variables via a variable selection approach. To address the challenges in monitoring beta distributed output variables, a beta regression model is adopted (Ferrari and Cribari-Neto, 2004, Hwang *et al.*, 2014, Kim *et al.*, 2017c). Moreover, we propose a simple but more efficient model selection method based on partial multiple linear regression to investigate the challenging issue of high dimensionality.

The remainder of this chapter is organized as follows. Section 3.2 provides a review of the model-based MSPC charts with beta distributed output variables. In Section 3.3, we propose a chart integrating a faulty variable diagnosis based on partial multiple linear regression, followed by simulation studies in Section 3.4. Section 3.5 summarizes the chapter and discusses future work.

3.2. Model-Based Multistage SPC Charts with Beta Distributed Output Variables

Production stages are sequentially concatenated to form multistage processes. A typical multistage process is illustrated in Figure 2.1 (Jearkpaporn *et al.*, 2005). A regression adjustment, named model-based SPC chart, suggested by Hawkins (1991, 1993) is developed to monitor output variables associated with input variables in a multistage process. Non-normality assumption of the output variables is studied by Jearkpaporn *et al.* (2007, 2003, 2005). They apply generalized linear models (GLM) for the estimation of the parameters of Poisson and gamma distributed data. However, the GLM is not applicable for data with beta distribution. Ferrari and Cribari-Neto (2004) investigate a beta regression model instead of GLM and derive a closed form expression for the deviance residual. Hwang *et al.* (2014) develop a deviance residual chart for monitoring multistage processes with beta distributed data.

We consider a multistage process with q stages where each stage has k input variables and p output variables as shown Figure 2.1. We define x_{mi} and y_{ji} as the m^{th} input and the j^{th} output variables at i^{th} stage, respectively, for $m = 1,...,k,\ \ j = 1,...,p$ and $i = 1,...,q.$ The

output from one stage is regarded as an input to the following stage. For example, $x_{i1},...,x_{ik}$ and $y_{(i-1)1},...,y_{(i-1)k}$ become the input variables for stage *i*. Thus, outputs from operations at upstream stages may affect the quality of downstream stages, and product variability may propagate throughout the production stages. Therefore, we intend to obtain the deviance residual for the beta distribution process.

The multistage process model with beta distributed output variables is obtained through the beta regression model (Ferrari and Cribari-Neto, 2004). The cumulative beta distribution function with two shape parameters α_1 and α_2 is expressed as follows:

$$F(y;\alpha_1,\alpha_2)=\frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}\int_0^y t^{\alpha_1-1}(1-t)^{\alpha_2-1}dt$$

By defining $\mu=\alpha_1/(\alpha_1+\alpha_2)$ and $\phi=\alpha_1+\alpha_2,$ the corresponding probability distribution function is written as follows:

$$f(y;\mu,\phi)=\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)}y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}$$

where $0<y<1,\quad 0<\mu<1$, $\phi>0,$ and ϕ is called as a precision parameter. $\Gamma(\cdot)$ represents a gamma function which is defined as $\Gamma(a)=\int_0^\infty x^{a-1}e^{-x}dx.$ Then the modified

beta distribution can be used to estimate beta parameters based on beta regression considering a logit link as follows:

$$\log\left(\frac{\mu}{1-\mu}\right) = \mathbf{X\beta} \tag{3.1}$$

where $\boldsymbol{\beta}$ is a coefficient vector, and $\mathbf{X}$ is a design matrix whose columns are the input variables and the proceeding output variables. From Eq. (3.1), the mean can be found as a function of estimated coefficients as follows:

$$\mu = \frac{\exp(\mathbf{X\beta})}{1+\exp(\mathbf{X\beta})} \tag{3.2}$$

The deviance residual for the beta distribution can be derived as follows (Ferrari and Cribari-Neto, 2004):

$$dr_{ji} = sign\left(y_{ji} - \hat{\mu}_{y_{ji0}}\right)\left[2\left|\log\left(f\left(y_{ji};\tilde{\mu}_{y_{ji1}},\tilde{\phi}\right)\right) - \log\left(f\left(y_{ji};\tilde{\mu}_{y_{ji0}},\tilde{\phi}\right)\right)\right]^{1/2} \tag{3.3}$$

where $i = 1,...,q,\ \ j = 1,...,p.$ $\mu_{y_{ji1}}$ is the mean of the current process for the j^{th} output variable at stage i, $\mu_{y_{ji0}}$ is in-control process mean, and y_{ji} is the current observation. The sign function assumes the value of +1 if $y_{ji} \geq \hat{\mu}_{y_{ji0}}$ and -1 otherwise. The maximum

likelihood estimator of $\mu_{y_{ji1}}$ is denoted by $\tilde{\mu}_{y_{ji1}}$, and $\tilde{\phi}$ can be obtained through beta regression. When ϕ is reasonably large, $\tilde{\mu}_{y_{ji1}} \approx y_{ji}$ (Ferrari and Cribari-Neto, 2004). The deviance residual calculated in Eq. (3.3) is used to construct the control limits

$$\text{LCL}_{ji} = \overline{dr}_{ji} - C_{dr_{ji}}\sqrt{S^2\left(dr_{ji}\right)},\ \ \text{UCL}_{ji} = \overline{dr}_{ji} + C_{dr_{ji}}\sqrt{S^2\left(dr_{ji}\right)}$$

where $i = 1,...,q,\quad j = 1,...,p.$ $\overline{dr}_{ji}$ and $S^2\left(dr_{ji}\right)$ are the sample mean and the sample variance of the deviance residual from the j^{th} output variable at stage i, respectively, and $C_{dr_{ji}}$ is the desired critical value for determining upper and lower control limits for the j^{th} output variable at stage i. Since dr_{ji} for the unimodal beta distributions approximately follow multivariate normal distribution, Hotelling's chi-square control chart can be applied appropriately. On the other hand, the extremely skewed beta distribution approximately follows a multivariate normal distribution but contaminated with some outliers; this may distort the chi-square control boundary. In this dissertation, we assume the unimodal beta distribution for output variables. Therefore, conventional multivariate process monitoring techniques can be applied to multistage processes by approximating the deviance residuals to a normal distribution.

3.3. Proposed Method via Diagnosis of Faulty Variables

Commonly used diagnostic procedures include stepwise variable selection, LASSO, decomposition of T^2 and various stepdown procedures (Mason *et al.*, 1995, Mason *et al.*,

1997, Tibshirani, 1996, Li *et al.*, 2008, Sullivan *et al.*, 2007, Kim *et al.*, 2013). Those diagnostic procedures have advantages and disadvantages. For example, the decomposition of T^2 is significantly intense for computation since it considers $p!$ number of different decompositions. It may be good in a one-time diagnosis, but not practical in online process monitoring. To overcome the computational issue, several control charts have been developed by adopting FVS and an adaptive LASSO (Wang and Jiang, 2009, Jiang *et al.*, 2012, Zou and Qiu, 2009). In the literature of variable selection-based control charts, one parameter representing the number of the selected variables is considered. Thus, when the diagnosis procedure selects the predetermined number of variables in the selection process, the algorithm terminates the selection procedure rather than computing all possible combinations of the selection sets. In the FVS, for example, once the initial variable is selected, the next variable is chosen based on given information of the previously selected variable. Thus, it might have a chance of losing the diagnostic power by choosing the wrong variable in the first step. Since the variables are strongly correlated with each other, the probability of misidentification of the first variable increases. In another example, when the mean shift occurs in variable x_s of a true model, and there is a strong correlation between x_s and x_t, LASSO would probably offer the solution as either $\beta_s = 0$ or $\beta_t = 0$ because the procedure tends to seek a sparse solution. As a result, the LASSO-based procedure may lead to misidentification.

In this chapter, we propose a diagnostic procedure considering the effect of correlation. In partial multiple linear regression model, the vector of regression coefficients is partitioned as:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_i \\ \boldsymbol{\beta}_\Gamma \end{bmatrix},$$

where β_i is i^{th} regression coefficient, and $\boldsymbol{\beta}_\Gamma$ is a vector of the rest of the coefficients, so Γ is a set of all the coefficients excluding i^{th} coefficients. We test the hypotheses $H_0 : \beta_i = 0,\ H_1 : \beta_i \neq 0.$ The model can be written as:

$$\mathbf{y} = \mathbf{X}_i \beta_i + \mathbf{X}_\Gamma \boldsymbol{\beta}_\Gamma + \boldsymbol{\varepsilon}.$$

The design matrix is partitioned with corresponding columns to β_i and $\boldsymbol{\beta}_\Gamma$. From the solution of the least square estimator, β_i, in a reduced model, the regression sum of square is $SS_R(\beta_i) = \hat{\beta}_i' \mathbf{X}_i \mathbf{y},$ and regression sum of squares of β_i given $\boldsymbol{\beta}_\Gamma$ is

$$SS_R(\beta_i \mid \boldsymbol{\beta}_\Gamma) = SS_R(\boldsymbol{\beta}) - SS_R(\boldsymbol{\beta}_\Gamma). \tag{3.4}$$

Hence, Eq. (3.4) can be interpreted as an effect of the i^{th} variable in the model. Let $S^2_{i|\Gamma}$ be a conditional sum of square using the above equation. Then,

$$S_{i|\Gamma}^2 = S^2 - S_{\Gamma}^2$$

where S^2 and S_{Γ}^2 represent the associated sum of squares in right part of Eq. (3.4), respectively. This is the same result as the conditional T^2 given Γ, which is proposed by Runger (1996). The conditional sum of squares, $S_{i|\Gamma}^2$ follows a chi-square distribution with one degree of freedom. We now determine the potentially shifted variables when

$$S_{i|\Gamma}^2 > \chi_{\gamma,1}^2,$$

where γ is a threshold parameter (Kim *et al.*, 2014). Once the variables are identified, we estimate nonzero mean values using the least square solution. As expected, above threshold policy allows more sparse selection with larger γ and vice versa. Thus the proposed methodology does not limit the number of variables, whereas the existing charts apply a fixed number of selected variables. As a result, the proposed methodology is expected to perform better than the existing variable selection-based charts even if the sparsity assumption is violated. Although there are more chances of choosing many of variables, the chart may perform properly by controlling the parameter γ.

One of the crucial issues of variable selection-based SPC charts is the capability of the variable selection procedure in selecting the "true' variables. If the identification ability is insufficient, the chart may perform poorly resulting in longer detection times than using

the original multivariate SPC charts. In order to increase the capability of the variable selection, we apply the EWMA methodology so that we estimate $\boldsymbol{\mu}_t^*$ based on $\mathbf{w}_t$, instead of $\mathbf{x}_t$, as denoted by

$$\mathbf{w}_t = (1-r)\mathbf{w}_{t-1} + r\mathbf{x}_t$$

where r is a smoothing parameter of EWMA. When t approaches infinity, $\mathbf{x}_t$ can be asymptotically replaced with $\mathbf{w}_t$ without loss of generality throughout the methodology. We refer to the proposed chart as CVS-MEWMA, which represents an MEWMA chart using conditional sum of square-based variable selection.

3.4. Performance of CVS-MEWMA

3.4.1. Description of Statistical Process Monitoring

Consider a multistage process having k input variables and that each stage has one output variable. The unimodal beta distribution of the output variable for stage 1 is

$$\text{Beta}\left(\frac{\exp\left(1+\beta_{11}x_{11}+\cdots+\beta_{k1}x_{k1}\right)}{1+\exp\left(1+\beta_{11}x_{11}+\cdots+\beta_{k1}x_{k1}\right)}, \phi_1\right), \tag{3.5}$$

and

$$\text{Beta}\left(\frac{\exp\left(1+\beta_{1i}x_{1i}+\cdots+\beta_{ki}x_{ki}+\beta_{k+1}y_{i-1}\right)}{1+\exp\left(1+\beta_{1i}x_{1i}+\cdots+\beta_{ki}x_{ki}+\beta_{k+1}y_{i-1}\right)},\phi_i\right) \tag{3.6}$$

for stage $i(\neq 1)$ through the logit link. The preceding output variable y_{i-1} in Eq. (3.6) is the input variable for i^{th} stage. For illustration, we consider a five stage production process where each stage has two input variables and two output variables. We use bivariate normal variables since the input variables can be easily normalized. The output variables are fitted to the unimodal-beta distributions via beta regression and the association between output and input variables is expressed as Eqs. (3.5) and (3.6). We assume that the precision parameter ϕ is large enough so that it satisfies $\tilde{\mu}_{y_{ji}} \approx y_{ji}$. At every sampling point, we obtain an $pq \times 1$ dimensional observation vector, where p is the number of output variables, and q is the number of stages. Thus, the multistage process readily becomes high-dimensional problem when either p or q increases. Since the variance propagation effect can be effectively removed by using deviance residual, the stacked observation vector holds the properties of multivariate distributions in probability.

Since the beta distribution does not belong to the over-dispersed exponential family, it is difficult to develop the deviance-based control charts and to determine the mean shift in the out-of-control scenario. In our multistage process model, we consider the shift in the coefficient vector $\boldsymbol{\beta}$ in Eq. (3.1). Since the mean is a function of $\boldsymbol{\beta}$ in Eq. (3.2), shift in the coefficient vector eventually result in mean shifts. The out-of-control mean can be represented as:

$$\mu = \frac{\exp(\mathbf{X\beta} + \delta\mathbf{d})}{1 + \exp(\mathbf{X\beta} + \delta\mathbf{d})} \tag{3.7}$$

where δ is a disturbance magnitude, and $\mathbf{d}$ is a unit vector representing the shift direction. Thus, the multistage statistical process monitoring problem becomes sequential testing of the hypothesis $H_0 : \boldsymbol{\beta}_i = \boldsymbol{\beta}_{i,0}$ from the deviance residual at each stage *i*.

The linear coefficients in Eq. (3.1) can be estimated by beta regression from the retrospective data analysis. Generally, in multistage processes, the output variables are likely to be correlated with each other within a stage and between stages since the model assumes the mean of the one-step ahead variables is affected by the preceding output variables. However, we only consider the inter-correlation within a stage which can be estimated from the preliminary data by adopting the deviance residual. In the experiments, we consider the correlation within the stages and denote it as **W**. We also assume that the magnitude of the correlation within a stage, which is the same for all the stages. Let **B** represent the correlation between stages, then the correlation matrix can be expressed as

$$\boldsymbol{\Sigma} = \mathbf{I}_q \otimes \mathbf{W}_q + (\sim \mathbf{I}_q) \otimes \mathbf{B}_q$$

where $\mathbf{I}_q$ is an identity matrix with dimension *q*, $\sim \mathbf{I}_q$ is a matrix with zero diagonals and all 1's for non-diagonal elements with dimension *q*, and $\otimes$ is a Kronecker product operator. Therefore, the diagonal elements in matrix **B** is the lag-1 autocorrelation of the one output

variable in consecutive stages, and the non-diagonal elements are cross correlations. Both **W** and **B** are *p*-dimensional matrices, so the dimension of the covariance matrix is $pq \times pq$.

3.4.3. Performance Analysis of the CVS-MEWMA Chart

In the experiments, we consider the smoothing parameter $r = 0.2$ and apply an initial simulation period of 100 observations to stabilize the process with EWMA statistic. Thus, we obtain an asymptotic covariance matrix as $r/(2-r)\mathbf{\Sigma}$ for computing the monitoring statistic. Depending on the magnitude δ and the direction $\mathbf{d}$ in Eq. (3.7), the size of out-of-control signal can be determined. Without loss of generality, we calibrate δ and $\mathbf{d}$ appropriately for the size of the shift. In the experiments, we assume that the shift occurs in the first variable in the first stage, and the second variable in the second stage. The additive value of the shift to each direction in the mean of deviance residual is considered. We evaluate the performance of each chart using ARL_1 (the in-control ARL is 200).

In Table 3.1, we compare the CVS-MEWMA chart with conventional multivariate SPC charts such as an individual Shewhart chart, Hotelling T^2, and MEWMA charts. The control limits for the individual Shewhart chart are represented as:

$$\mathrm{CL}_{ji} = \overline{dr}_{ji} \pm C_{dr_{ji}} \sqrt{S^2\left(dr_{ji}\right)}$$

where $\overline{dr}_{ji}$ and $S^2\left(dr_{ji}\right)$ are the mean and the sample variance of the deviance residual from j^{th} variable at stage i, and $C_{dr_{ji}}$ is the desired critical value for in-control ARL. The threshold parameter $\gamma = 0.3$ is used for the proposed chart.

Table 3.1. ARL_1 Performance comparison

Shift	Shewhart	Hotelling T^2	MEWMA	CVS-MEWMA
0.2	189.68	171.07	84.48	**73.04**
0.4	161.98	111.86	24.41	**20.24**
0.6	121.98	63.34	11.14	**9.38**
0.8	86.52	31.86	6.99	**5.95**
1.0	57.48	16.61	5.11	**4.34**
1.5	18.66	4.16	3.18	**2.70**
2.0	7.03	1.70	2.37	**2.03**
2.5	3.23	1.13	2.02	**1.68**
3.0	1.82	1.02	1.88	**1.43**

In Table 3.1, the proposed chart outperforms the other charts. The EWMA type charts perform well in small shifts. The CVS-MEWMA chart outperforms MEWMA for all shifts since it removes noises inherently contained in all variables.

In order to assess the capability of the conditional sum of square-based variable selection methodology, we compare a few variable selection-based control charts such as VSMSPC and VSMEWMA (Jiang *et al.*, 2012, Wang and Jiang, 2009). In their charts, they

demonstrate that they perform well when the number of shifted variables in known in advance. In Table 3.2, we reflect this information and fix the number of selected variables as two (same number as shifted variables). We compare the MEWMA chart with forward variable selection and our proposed variable selection-based chart.

Table 3.2. ARL_1 performance of the variable selection-based control charts

Shift	VSMEWMA	CVS-MEWMA
0.2	77.09	**76.65**
0.4	22.11	**21.16**
0.6	10.15	**9.71**
0.8	6.50	**6.15**
1.0	4.76	**4.51**
1.5	2.97	**2.79**
2.0	2.23	**2.10**
2.5	1.82	**1.74**
3.0	1.56	**1.48**

The Table 3.2 shows that CVS-MEWMA slightly outperforms VSMEWMA. It may be attributed to the stepwise-variable selection of the first-selected variable.

From the results of the CVS-MEWMA in Tables 3.1 and 3.2, the ARL_1 performance depends on the number of selected variables. In Table 3.1, the chart selects about four variables on the average out of ten variables with the given parameter γ. As γ increases,

the chart results in selecting a small number of potentially shifted variables. Table 3.3 illustrates the effect of the parameter γ on the ARL_1 performance.

Table 3.3. Effect of γ on ARL_1 performance

	γ			
Shift	0.01	0.1	0.2	0.3
0.2	84.36	80.14	76.20	73.04
0.4	23.05	22.32	20.89	20.24
0.6	10.36	10.11	9.71	9.38
0.8	6.44	6.32	6.08	5.95
1.0	4.66	4.57	4.45	4.34
1.5	2.86	2.81	2.74	2.70
2.0	2.15	2.10	2.06	2.03
2.5	1.75	1.73	1.71	1.68
3.0	1.52	1.49	1.46	1.43

Since we assume a sparsity of the high-dimensional process, the chart performs better when it selects fewer variables. For example, the chart selects approximately 9, 8, 5 and 4 variables, respectively, on average at each sampling point. As shown in Table 3.2, when selecting only two variables, it performs worse than the last case in Table 3.3 where we select four variables on average. We may not identify the shifted variables correctly due to the strong correlation. In the experiments, the shifts occur in different stages, and each shifted variable has a strong correlation within the stage. Hence, there is a chance of

identifying the wrong variables as shifted ones when we only consider a fixed number of shifted variables.

We can readily conjecture that the performance may be a convex function of γ with regard to the correlation structure. Obviously, it could be explained by the capability of the variable selection. As γ decreases, the chart tends to select more variables at each sampling point and vice versa. Therefore, when γ approaches zero, the ARL_1 becomes closer to that of MEWMA's.

We also utilize other performance measures: the correctness ratio (CR) and the expected error ratio (EER) which are defined as:

$$\text{CR} = \frac{\text{Number of correctly indentified variables}}{\text{Number of indentification}}$$

$$\text{EER} = E\left(\frac{\text{Total number of misidentified variables}}{\text{Total number of variables}}\right).$$

The CR is the ratio of the correctly identified numbers to the total number of identifications, and EER is the ratio of the number of misidentification for each variable and the total number of variables. When the chart selects several variables including the shifted variables, we regard it as a correctly identified case. Both Tables 3.4 and 3.5 show the effect of γ on CR and EER respectively.

Table 3.4. Effect of γ on CR

	γ				
Shift	0.01	0.1	0.2	0.3	γ^*
0.2	0.9581	0.6299	0.3497	0.1646	0.0365
0.4	0.9750	0.7778	0.5651	0.3598	0.0867
0.6	0.9812	0.8306	0.6614	0.4827	0.1283
0.8	0.9843	0.8572	0.7072	0.5444	0.1541
1.0	0.9874	0.8735	0.7390	0.5949	0.1807
1.5	0.9904	0.9132	0.8139	0.6944	0.2284
2.0	0.9949	0.9499	0.8796	0.7798	0.2777
2.5	0.9973	0.9767	0.9322	0.8680	0.3355
3.0	0.9993	0.9901	0.9732	0.9305	0.3881

* The last column represents the case of selecting fixed two variables.

Table 3.5. Effect of γ on EER

	γ				
shift	0.01	0.1	0.2	0.3	γ^*
0.2	0.7820	0.6208	0.4642	0.3436	0.1414
0.4	0.7816	0.6159	0.4529	0.3292	0.1188
0.6	0.7812	0.6146	0.4512	0.3235	0.1082
0.8	0.7819	0.6170	0.4504	0.3237	0.1021
1.0	0.7814	0.6156	0.4541	0.3251	0.0979
1.5	0.7818	0.6172	0.4539	0.3240	0.0877
2.0	0.7806	0.6181	0.4564	0.3292	0.0788
2.5	0.7825	0.6208	0.4592	0.3312	0.0701
3.0	0.7818	0.6222	0.4614	0.3359	0.0628

The CR of the fixed selected variable (last column in Tables 3.4 and 3.5) shows an inability to accurately select the changed variables, resulting in an increased ARL_1. For example, in the large shift case (shift 3.0), its CR is 0.3881, which is significantly lower than 0.9305 when $\gamma = 0.3$. In this case, even though we remove a large amount of the noise by selecting a few variables, we may also remove the assignable causes. Table 3.6 illustrates the capability of the different variable selection methods adopted by the VSMEWMA chart and the CVS-MEWMA chart.

Table 3.6. Capability of the variable selection techniques

	VSMEWMA		CVS-MEWMA	
Shift	CR	EER	CR	EER
0.2	0.0198	0.1525	0.0365	0.1414
0.4	0.0414	0.1408	0.0867	0.1188
0.6	0.0062	0.1334	0.1283	0.1082
0.8	0.0757	0.1290	0.1541	0.1021
1.0	0.0928	0.1236	0.1807	0.0979
1.5	0.1269	0.1149	0.2284	0.0877
2.0	0.1629	0.1062	0.2777	0.0788
2.5	0.2085	0.0970	0.3355	0.0701
3.0	0.2418	0.0896	0.3881	0.0628

In Table 3.6, the CR and EER of the proposed chart present better performance, resulting in the smaller out-of-control ARL shown in Table 3.2.

3.5. Conclusion

In this chapter, we propose a high-dimensional multistage SPC chart for monitoring output variables that follow a beta distribution. Deviance residuals through the model-based multistage SPC are used for removing the cascade property in multistage processes. The vectorized deviance residuals are used as observations, and an appropriate covariance matrix representing correlations and cross correlations is determined. The conditional sum of square-based variable selection technique is used to improve the capability of the detection of the changes in high-dimensional processes.

Extensive simulation studies demonstrate that the proposed chart outperforms existing multivariate SPC charts and recent variable selection-based control charts in terms of out-of-control ARL. The smoothing parameter r affects the capability of diagnosis of shifted variables. Using a small value of r may lead to poor estimation of $\boldsymbol{\mu}_t^*$ in case of large shifts. Determining the optimal values of r and γ simultaneously according to the shift size is challenging and represents a subject of future research. Furthermore, the assumption of unimodal beta distribution can be relaxed for the generalization of the approach.

CHAPTER 4

MONITORING OF HIGH-DIMENSIONAL PROCESSES VIA SPARSE GROUP LASSO

4.1. Introduction

The design of statistical process control and monitoring for multistage processes is challenging due to the ‘curse of dimensionality’ and ‘stage-dependency’ of the datasets. In Chapter 3, we develop approaches to monitor quality characteristics and process parameters in multistage processes by incorporating the individual variable selection technique to resolve the issue of high-dimensionality, and the model based approach is used to analyze the stage-dependency when the process parameters are beta distributed. In particular, this individual variable selection incorporated chart may not perform well under specific structural conditions such as complicated correlation structure and behavior of the process shift. In various manufacturing processes, the clustered data structure and the clustered behavior of process change can be readily observed. Therefore, the need of analyzing groups that exhibit the same pattern in practical applications may prove effective in process monitoring.

Moreover, the individual VS-based charts may function effectively when the number of stages or variables to be monitored is relatively small. However, the complexity of those charts increases exponentially as the dimensionality increases since the computation of VS with respect to each variable would significantly increase. In this chapter, we develop a new method to monitor the high-dimensional process when the clustered structure in both

dataset and behavior of the process are observed. For the high-dimensional feature of the process, a penalized likelihood with mixed regularization parameters is introduced. Finally, we show that the proposed chart outperforms other existing charts in process diagnostics and shift detection throughout various simulation studies under the settings of multistage processes. This chapter is organized as follows. Section 4.2 introduces the variants of LASSO regression for variable selection and proposes a sparse group LASSO based control chart. Section 4.3 presents the performance of the proposed model through extensive simulations, and is followed by conclusion in Section 4.4.

4.2. Methodology

In this section, we describe the proposed methodology for monitoring the high-dimensional processes via the extension of the LASSO criterion. Section 4.2.1 provides a brief introduction of the LASSO regression and its connection with variable selection. Section 4.2.2 demonstrates the utilization of LASSO penalized likelihood in the monitoring of high-dimensional processes, and followed by proposing a new charting methodology based on the EWMA scheme.

4.2.1. Variants of LASSO Regression for Variable Selection

Consider the conventional p-dimensional multiple linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{y}$, $\mathbf{X}$, $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ are the vector of response variables, predictors, regression coefficients, and the errors that are i.i.d. random variables with distribution $N_p(\mathbf{0}, \boldsymbol{\Sigma})$, respectively. In practice, some of the predictors may not contribute to the response variable

and be forced to be deleted from this regression model. Traditional variable selection techniques can be used to delete (or select, on the contrary) several variables via model selection criteria such as AIC and BIC. The penalized least square, or equivalently, the penalized likelihood function with Gaussian error distribution developed as

$$L(\boldsymbol{\beta}) = \min_{\boldsymbol{\beta}} \left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right\|_2^2 + \kappa P(\boldsymbol{\beta}),$$

where κ is a regularization parameter and $P(\cdot)$ is a penalty function. When the penalty function becomes $P(\boldsymbol{\beta}) = \| \boldsymbol{\beta} \|_1$, which is L_1 norm, the corresponding penalized least square equation is called LASSO. When the penalty function is chosen with a positive value of κ, some regression coefficients of insignificant regressors are shrunk towards zero. Especially with the penalty function, $\| \boldsymbol{\beta} \|_1$, $L(\boldsymbol{\beta})$ is not differentiable at $\beta_i = 0$ for some i, so that those coefficients become exactly zero. Therefore, the remaining regressors with nonzero coefficients can be viewed as selected variables as significant ones to explain the response based on the LASSO criterion.

As introduced previously, the clustered data structure is found in wide applications of the high-dimensional dataset. Now we consider the multiple linear regression model with N groups as

$$\mathbf{y} = \sum_{n=1}^{N} \mathbf{X}_n \boldsymbol{\beta}_n + \boldsymbol{\varepsilon},$$

where $\mathbf{y}$ is a $p \times 1$ vector with $p = \sum_{n=1}^{N} p_n$, $\mathbf{X}_n$ is a $p \times p_n$ matrix corresponding to the group n, and $\boldsymbol{\beta}_n$ is a coefficient vector of size p_n, for $n = 1, ..., N$. $\boldsymbol{\varepsilon}$ is a white noise vector that follows multivariate normal distribution with mean zero and covariance $\boldsymbol{\Sigma}$, where the diagonal block is $\boldsymbol{\Sigma}_n$. Denoting $\mathbf{X} = [\mathbf{X}_1, ..., \mathbf{X}_N]$ and $\boldsymbol{\beta}$ as a stacked vector of $\boldsymbol{\beta}_n$ for all N, this becomes a multiple regression model as discussed above.

For the problems with grouped covariates, Yuan and Lin (2006) develop a group LASSO (GLASSO) criterion as

$$\hat{\boldsymbol{\beta}}^{GLASSO} = \arg\min_{\boldsymbol{\beta}} \frac{1}{2} \left\| \mathbf{y} - \sum_{n=1}^{N} \mathbf{X}_n \boldsymbol{\beta}_n \right\|_2^2 + \kappa \sum_{n=1}^{N} \sqrt{p_n} \left\| \boldsymbol{\beta}_n \right\|_2. \tag{4.1}$$

Since it is not differentiable at $\boldsymbol{\beta}_n = \mathbf{0}$, the solution would possibly have exactly zeros for a few groups of coefficients with a positive tuning parameter κ. With specific conditions that all groups are not overlapped and the size of each group is 1, it reduces to the regular LASSO criterion.

4.2.2. Sparse Group LASSO-Based Testing and Monitoring

The hypotheses of interest in our MSPC testing problem are $H_0 : \boldsymbol{\mu} = 0$ and $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$. Particularly from the generalized likelihood ratio test statistic with normality, the following penalized likelihood function can be formulated:

$$L(\boldsymbol{\mu}) = \min_{\boldsymbol{\mu}} || \mathbf{z} - \mathbf{R}\boldsymbol{\mu} ||_2^2 + \kappa P(\boldsymbol{\mu}), \tag{4.2}$$

where $\mathbf{z} = \mathbf{R}\mathbf{x}$, and the matrix, $\mathbf{R}$ is a Cholesky decomposition such that $\boldsymbol{\Sigma}^{-1} = \mathbf{R}^T\mathbf{R}$. Based on the review in the preceding section, Eq. (4.2) can be reformulated as

$$L(\boldsymbol{\mu}) = \min_{\boldsymbol{\mu}} \left\| \mathbf{z} - \sum_{n=1}^{N} \mathbf{R}_n \boldsymbol{\mu}_n \right\|_2^2 + \kappa P(\boldsymbol{\mu}),$$

where $\mathbf{R}_n$ is a submatrix of $\mathbf{R} = [\mathbf{R}_1, ..., \mathbf{R}_N]$, and $\boldsymbol{\mu}_n$ is corresponding coefficient vector in $\boldsymbol{\mu}$. As outlined in this dissertation, it is often reasonable to assume that only a few coefficients in $\boldsymbol{\mu}$ are nonzero when the shift occurs in high-dimensional processes, referred to as sparsity. With the grouped covariates, the sparsity can be viewed in two aspects: 'between groups' and 'within a group' level, i.e., group-wise sparsity and within-a-group sparsity.

Considering two-way sparsity, the combination of two penalties can be applicable (Simon *et al.*, 2013) as

$$L(\boldsymbol{\mu}) = \min_{\boldsymbol{\mu}} \left\| \mathbf{z} - \sum_{n=1}^{N} \mathbf{R}_n \boldsymbol{\mu}_n \right\|_2^2 + \alpha\kappa \left\| \boldsymbol{\mu} \right\|_1 + (1-\alpha)\kappa \sum_{n=1}^{N} \sqrt{p_n} \left\| \boldsymbol{\mu}_n \right\|_2. \tag{4.3}$$

where $0 \le \alpha \le 1$ is a weight parameter to form a convex combination of the LASSO and GLASSO penalties. In this model, κ controls overall sparsity, and the weight, α controls within-a-group or group-wise sparsity. However, under MSPC setting with unknown shift patterns, within-a-group sparsity is not necessarily dependent on a group-wise sparsity. Thus, α is not an appropriate parameter for process control. Moreover, as discussed in literature, determination of the complexity parameter κ is one of the major challenging issues of VS-based control charts (Jiang *et al.*, 2012, Wang and Jiang, 2009). Existing VS-based charts attempt to overcome this issue by introducing another parameter that represents the number of selected variables so that the parameter κ is determined based on this predetermined selection parameter. However, in our model, the number of selected variables cannot be easily determined due to the grouped fashion of sparsity. Moreover, Eq. (4.3) is often infeasible with a constraint of the pre-specified number of selected variables.

To overcome those issues, we propose a following penalized likelihood.

$$L(\boldsymbol{\mu}) = \min_{\boldsymbol{\mu}} \left\| \mathbf{z} - \sum_{n=1}^{N} \mathbf{R}_n \boldsymbol{\mu}_n \right\|_2^2 + c_1 \kappa_1 \left\| \boldsymbol{\mu} \right\|_1 + c_2 \kappa_2 \sum_{n=1}^{N} \sqrt{p_n} \left\| \boldsymbol{\mu}_n \right\|_2, \tag{4.4}$$

where c_1, c_2 are the ratio parameters determined in [0,1], and κ_1, κ_2 are the complexity parameters that control within-a-group sparsity and group-wise sparsity, respectively. One may wonder if κ_1 or κ_2 can be intuitively determined based on the number of variables or

the number of groups to be selected. However, this optimization problem to obtain two parameters simultaneously for the nonzero components in $\boldsymbol{\mu}$ to be pre-specified numbers would result in infinitely many solutions. Thus, we determine κ_1 and κ_2 as follows by adopting the idea of Liu *et al.* (2009b).

$$\kappa_1^* = \arg\inf\left\{\kappa_1 : \left|\Psi_{\kappa_1|c_1=1,\kappa_2 c_2=\kappa_2^* c_2}\right| = 0\right\} \text{ and } \kappa_2^* = \arg\inf\left\{\kappa_2 : \left|\Psi_{\kappa_2|c_2=1,\kappa_1 c_1=0}\right| = 0\right\},$$

where $\Psi_{\kappa_2|c_2=1,\kappa_1 c_1=0}$ is a set of indices of nonzero elements in $\boldsymbol{\mu}^*$ given $c_2 = 1$ and $\kappa_1 c_1 = 0$, $\Psi_{\kappa_1|c_1=1,\kappa_2 c_2=\kappa_2^* c_2}$ is a set of indices of nonzero elements in $\boldsymbol{\mu}^*$ given $c_1 = 1$ and $\kappa_2 c_2 = \kappa_2^* c_2$, and $|A|$ represents a cardinality of the set A. Thus, κ_2^* is the least value of penalty in group-wise, and κ_1^* is the least value of penalty given c_1 and κ_2^*, i.e., given selected groups, to obtain a sparse solution in the those groups. Consequently, both κ_1 and κ_2 are not the parameters because they are obtained when the response, $\mathbf{z}$ and the design matrix, $\mathbf{R}$ are given. Rather c_1 and c_2 control the sparsity as ratios with respect to the constant, κ_1^* and κ_2^*. Even if it does not exactly imply the number of selected groups or variables, it provides a sense of sparsity levels as a ratio; i.e., c_2 implies the ratio of the number of nonzero groups to N; and c_1 implies the ratio of the number of nonzero variables in a group. Thus, by controlling the parameters, c_1 and c_2, the sparsity within and between groups can be achieved practically and intuitively.

Determination of the proper parameter set (c_1, c_2) is substantially dependent on the location and the size of the process shifts. Thus, conventional criteria such as cross validation, AIC and BIC are ineffective in determining these parameters without proper information of the shift. We discuss this issue in later section and present sensitivity of the proposed method to various shift scenarios.

Constructing a chart based on a proper monitoring statistic is also of interest in the SPC problem. Zou and Qiu (2009) adopt maximal $q(\leq p)$ standardized statistics in the nonparametric testing setup for their LASSO-based chart, where q represents the tolerance for the number of nonzero in the estimator. This chart signals when

$$Q = \max_{j=1,\ldots,q} \frac{\tilde{T}_{\kappa_j} - E(\tilde{T}_{\kappa_j})}{\sqrt{Var(\tilde{T}_{\kappa_j})}} > c,$$

where $\tilde{T}_{\kappa_j} = n(\hat{\boldsymbol{\mu}}_{\kappa_j}^T \boldsymbol{\Sigma}^{-1} \bar{\mathbf{X}})^2 / \hat{\boldsymbol{\mu}}_{\kappa_j}^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}}_{\kappa_j}$ for $j = 1, \ldots, q,$ and $\hat{\boldsymbol{\mu}}_{\kappa_j}$ is the adaptive LASSO estimator of $\boldsymbol{\mu}$ in the L_1 norm penalized likelihood with the penalty κ_j. Since the parameter κ is determined in a continuous range, the κ_j is chosen for the estimator to have j nonzero components. Differently, Jiang *et al.* (2012) apply the GLR-based approach and obtain a single monitoring statistic. Both monitoring statistics have their own advantages and disadvantages; see Jiang *et al.* (2012) for the comparison of these two monitoring statistics. In our framework of the likelihood, the specific number of selection

– either in groups or in variables – and corresponding penalty parameters cannot be uniquely obtained. Thus, we use a single monitoring statistic based on GLR as $Q = 2\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}$, where $\hat{\boldsymbol{\mu}}$ is the sparse estimation of $\boldsymbol{\mu}$.

For monitoring the process in Phase II, we construct the control chart based on EWMA statistic using sparse group LASSO (SGL) based likelihood function. Let $\mathbf{x}_t$ be the measurement vector sampled at time t distributed as $N_p(\mathbf{0}, \boldsymbol{\Sigma})$. The sequential vector of EWMA is defined as

$$\mathbf{w}_t = (\mathbf{I} - \boldsymbol{\Gamma})\mathbf{w}_{t-1} + \boldsymbol{\Gamma}\mathbf{x}_t, \ t = 0,1,... \tag{4.5}$$

where $\boldsymbol{\Gamma}$ is a smoothing matrix. $\mathbf{w}$ follows asymptotically multivariate normal distribution with mean zero and covariance, $\boldsymbol{\Sigma}_{\mathbf{w}} = (\mathbf{I} - \boldsymbol{\Gamma})\boldsymbol{\Sigma}_{\mathbf{w}}(\mathbf{I} - \boldsymbol{\Gamma})^T + \boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^T$. The covariance of EWMA vector, $\mathbf{w}$ is an applied form of the discrete time Lyapunov equation (Gajic and Qureshi, 2008). By using its solution, $\boldsymbol{\Sigma}_{\mathbf{w}}$ can be obtained as

$$vec(\boldsymbol{\Sigma}_{\mathbf{w}}) = (\mathbf{I} \otimes \boldsymbol{\Gamma} + \boldsymbol{\Gamma} \otimes \mathbf{I} - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma} vec(\boldsymbol{\Sigma})$$

where $vec(\boldsymbol{\Sigma}_{\mathbf{w}})$ vectorizes $\boldsymbol{\Sigma}_{\mathbf{w}}$ column-wise and $\otimes$ is a Kronecker product operator. Then, our proposed control chart signals if

$$Q_t = 2\mathbf{w}_t^T \boldsymbol{\Sigma}_{\mathbf{w}}^{-1} \hat{\boldsymbol{\mu}}_t - \hat{\boldsymbol{\mu}}_t^T \boldsymbol{\Sigma}_{\mathbf{w}}^{-1} \hat{\boldsymbol{\mu}}_t > H,$$

where

$$\hat{\boldsymbol{\mu}}_t = \arg\min_{\boldsymbol{\mu}} \left\| \tilde{\mathbf{R}}\mathbf{w}_t - \sum_{n=1}^{N} \tilde{\mathbf{R}}_n \boldsymbol{\mu}_n \right\|_2^2 + c_1 \kappa_1^* \|\boldsymbol{\mu}\|_1 + c_2 \kappa_2^* \sum_{n=1}^{N} \sqrt{p_n} \|\boldsymbol{\mu}_n\|_2, \tag{4.6}$$

such that $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} = \tilde{\mathbf{R}}^T \tilde{\mathbf{R}}$, and $H(>0)$ is a control limit chosen to obtain a given in-control ARL through simulations. Hereinafter, we call the proposed chart as an SGL-EWMA chart.

4.3. Performance Analysis

We present the numerical performance of the proposed SGL-EWMA chart through extensive Monte Carlo simulations. In SGL-EWMA, the smoothing matrix, $\boldsymbol{\Gamma}$ should be chosen prior to monitoring. We assume that the diagonal elements of $\boldsymbol{\Gamma}$ are assigned to be r, and non-diagonal elements to be zero. Then, the asymptotic covariance, $\boldsymbol{\Sigma}_{\mathbf{w}} = r(2-r)^{-1}\boldsymbol{\Sigma}$ can be used as having typically used in EWMA literature. In practice, the selection of the smoothing parameter, r mainly depends on a potential shift level. It should be determined small when the size of potential shift is expected to be large and vice versa. In this simulation, we set $r = 0.2,$ and its in-control ARL is fixed at 200. In order to observe the grouping behavior of the chart and compare the proposed chart to other existing

charts, the identity covariance matrix is considered. The number of groups is set as $N = 5$, and each group consists of $p_n = 2$ variables as equal sizes for all groups.

The out-of-control observation $\mathbf{x}_t$ can be seen as

$$\mathbf{x}_t = \boldsymbol{\mu}_0 + \delta \mathbf{d} + \boldsymbol{\varepsilon},$$

where δ is the shift size, $\mathbf{d}$ is the shift direction. For simplicity, we set $\boldsymbol{\mu}_0 = \mathbf{0}$. By splitting $\mathbf{d}$ into N groups, we represent d_{ij} as a j^{th} variable in i^{th} group, for $i = 1, ..., N$ and $j = 1, ..., p$. For the shift scenarios, we consider two representative shifts; 1) $d_{11} = d_{21} = 1$ and $d_{ij} = 0$ for other all i and j; 2) $d_{11} = d_{12} = 1$ and $d_{ij} = 0$ for other all i and j. Case 1) represents that two shifts occur in different groups; and Case 2) represents that two shifts occur in the same group.

We compare the SGL-EWMA chart with the conventional MEWMA chart and VSMEWMA in Table 4.1. The parameter of VSMEWMA for the number of selection is set to $s = 2$. For comparison, the parameters of our proposed chart are chosen for the number of total selection to be close to 2. Shift size is calculated as a noncentrality parameter, and the covariance matrix for each group is set as $\sigma_{(n)i,j} = 0.75$ such that $\boldsymbol{\Sigma}_n = [\sigma_{(n)i,j}]$.

Table 4.1. ARL_1 comparison of SGL-EWMA with MEWMA and VSMEWMA ($N = 5,\ p_n = 2$)

Charts	MEWMA	VSMEWMA		SGL-EWMA					
Parameters	None	$s = 2$		$(c_1, c_2) = (0.5, 0.5)$		$(0.7, 0.3)$		$(0.3, 0.8)$	
Shift\Cases	Dir.inv	1)	2)	1)	2)	1)	2)	1)	2)
0.50	67.25	70.04	70.66	71.90	67.70	70.06	70.02	72.51	65.66
1.00	17.92	17.81	17.61	18.60	16.96	18.18	17.59	18.93	15.94
1.50	8.58	8.24	8.19	8.70	8.09	8.68	8.27	8.81	7.53
2.00	5.60	5.29	5.30	5.64	5.19	5.55	5.36	5.69	4.91
2.50	4.20	3.93	3.92	4.20	3.89	4.16	4.01	4.28	3.68
3.00	3.43	3.15	3.14	3.36	3.12	3.34	3.22	3.45	2.96

Table 4.1 presents out-of-control ARL performances for two cases. From the result, we can observe several findings. First, both charts outperform MEWMA as shift size increases and vice versa. This is expected since the accuracy of VS becomes better as shift size increases. Second, when the shifts occur in the same group (case 2)), the SGL-EWMA charts with all parameter settings illustrate better performances than VSMEWMA. Third, on the contrary, when the shift occurs in different groups, the ARL_1's become smaller as c_1 increases. It is expected that the performance of SGL-EWMA approaches that of the VSMEWMA when $c_2 = 0$ and c_1 are determined for the number of selection to be close to 2. The latter two findings imply that the performance of the proposed chart would perform better the shifts occur in a grouped pattern.

We now implement the higher dimensional case where $N = 5$ and $p_n = 5$ with focus on the grouped pattern of the shift. Under sparsity in both between groups and within a group, we consider $|\mathbf{d}|_0 = 1, 2$ and 3, where $|\mathbf{d}|_0$ is an L_0 norm representing the number of truly shifted variable in $\mathbf{d}$. In addition, we set $c_2 = 0.8$ such that the proposed method selects one or two groups on average and vary c_1 to obtain the ARL performance. The bold character in the table represents the smallest ARL over the charts with the different parameter settings in each scenario. From Table 4.2, when $|\mathbf{d}|_0 = 1$, the proposed method with a large c_1 outperforms VSMEWMA since a large value of the penalty for the term $\|\boldsymbol{\mu}\|_1$ in Eq. (4.6) enforces the estimation more sparse. Similarly, the proposed method presents better performance than VSMEWMA with a moderate value of c_1 to allow more selection in the selected group. Consequently, it is conjectured that the large value of c_1 would perform well in the scenarios of within-a-group sparsity, and vice versa. Thus, if the engineering knowledge is available for the potential shift of the process, the parameters c_1 and c_2 are suggested according to the group-wise sparsity and the within-a-group sparsity.

In order to determine the performance of the charts when the shift occurs in different groups, we consider the two out-of-control scenarios; 3) 1 variable shifts in group 1 and 2; 4) 2 variables shift in group 1 and 2. Thus, the cases 3) and 4) represent 2 and 4 numbers of shifts in two different groups, respectively. Expecting that VSMEWMA would perform overall best among shift sizes when the number of selection is equal to the number of truly shifted variables, we consider $s = 2$ and $s = 4$ for the cases 3) and 4). We also consider

(c_1, c_2) for ARL_1 to be the best among the cases considered in Table 4.2. It is expected in general that SGL-EWMA may perform poorly if shifts occur over the groups due to the fashion of selecting in which the selection procedure would probably exclude several shifted variables. Table 4.3 presents that the proposed method shows slightly better than, but almost similar to VSMEWMA in ARL_1 performance.

Table 4.2. ARL_1 comparison of SGL-EWMA with MEWMA and VSMEWMA ($N = 5,\ p_n = 5$)

Charts	MEWMA	VSMEWMA			SGL-EWMA					
Parameters	None	$s = 2$			$(c_1, c_2) = (0.0, 0.8)$			$(0.2, 0.8)$		
Shift\Cases	Dir.inv	$\|\mathbf{d}\|_0 = 1$	2	3	1	2	3	1	2	3
0.50	93.72	82.61	95.82	100.63	95.43	97.92	99.53	94.20	95.07	98.22
1.00	26.63	17.72	23.42	26.99	27.17	28.11	28.97	24.63	24.21	27.23
1.50	11.90	7.82	9.56	11.09	10.83	11.25	11.69	9.97	9.96	11.20
2.00	7.36	4.96	5.89	6.64	6.52	6.58	6.83	6.14	6.11	6.65
2.50	5.44	3.71	4.28	4.77	4.65	4.70	4.84	4.43	4.41	4.69
3.00	4.38	3.00	3.39	3.80	3.64	3.72	3.81	3.49	3.49	3.71

Charts	SGL-EWMA								
Parameters	$(0.4, 0.8)$			$(0.6, 0.8)$			$(0.8, 0.8)$		
Shift\Cases	$\|\mathbf{d}\|_0 = 1$	2	3	1	2	3	1	2	3
0.50	88.36	89.97	93.63	84.12	92.83	95.98	79.26	98.22	101.00
1.00	21.58	22.41	25.00	19.14	22.69	25.14	17.51	24.70	27.12
1.50	9.07	9.49	10.46	8.39	9.58	10.34	7.76	10.18	11.31
2.00	5.65	5.84	6.33	5.28	5.89	6.40	4.93	6.32	6.86
2.50	4.15	4.26	4.58	3.88	4.32	4.63	3.68	4.56	4.93
3.00	3.30	3.38	3.59	3.13	3.44	3.65	2.96	3.65	3.88

Table 4.3. ARL_1 performances when shifts occur in different groups

Charts	MEWMA	VSMEWMA		SGL-EWMA	
Parameters	None	$s=2$	$s=4$	$(0.8,0.8)$	$(0.2,0.8)$
Shift\Cases	Dir.inv	3)	4)	3)	4)
0.50	93.72	91.48	97.46	90.38	95.67
1.00	26.63	22.03	28.15	21.95	26.42
1.50	11.90	10.16	11.79	9.16	11.39
2.00	7.36	6.19	7.01	5.71	6.92
2.50	5.44	4.50	5.16	4.17	5.01
3.00	4.38	3.60	3.95	3.32	3.92

Although the performances of both charts are generally better than MEWMA as shown in Tables 4.2 and 4.3, there is an increasing pattern of ARL_1 as the number of actual shifted variables increases. It is due to the fact that the variable selection procedure would possibly exclude the shifted variables. Thus, when the sparsity of the process shift is difficult to estimate based on the engineering knowledge, the parameters that control sparsity are suggested to be smaller to select more variables to avoid excluding the shifted variables. Moreover, the chart may underperform MEWMA in detecting small shift, e.g., when $\delta = 0.5$. This is a typical phenomenon of the VS-based methodologies where VS may misidentify the truly shifted variables when the shift size is small. Consequently, when the information of sparsity is not given or when the shift size is small, a moderate or more number of selection is suggested. Figure 4.1 illustrates the mean number of selection (vertical axis) in terms of the parameters (c_1, c_2) when $|\mathbf{d}|_0 = 1, 2$ and 3 (horizontal axis)

with given shift size as $\delta = 2.0$. Moreover Figure 4.2 presents the mean number of selection (vertical axis) in terms of the shift size as a noncentrality with given parameters (0.4, 0.8). Figures 4.1 and 4.2 show that the proposed method selects the variables according to the shift patterns. In Figure 4.1, the proposed method selects more variables as the number of shifted variables increases. In addition, in Figure 4.2, it selects less variables on average as the shift size increases. This is one of the attractive feature of the proposed method in selecting the variables flexibly while the existing VS-based methods such as VSMSPC and VSMEWMA select a fixed number of variables.

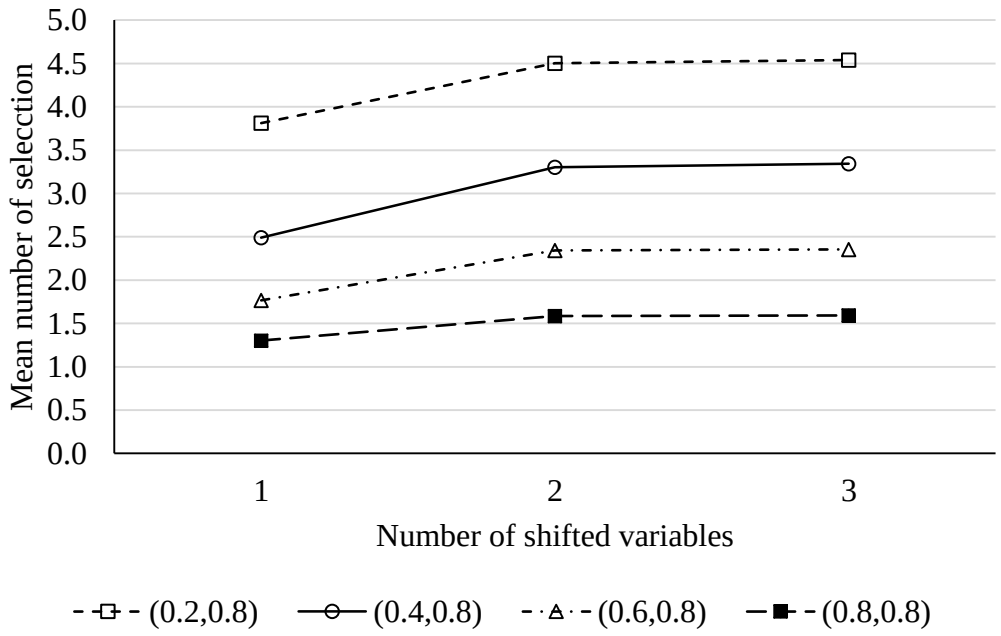


Figure 4.1 Mean number of selection according to the different parameter settings

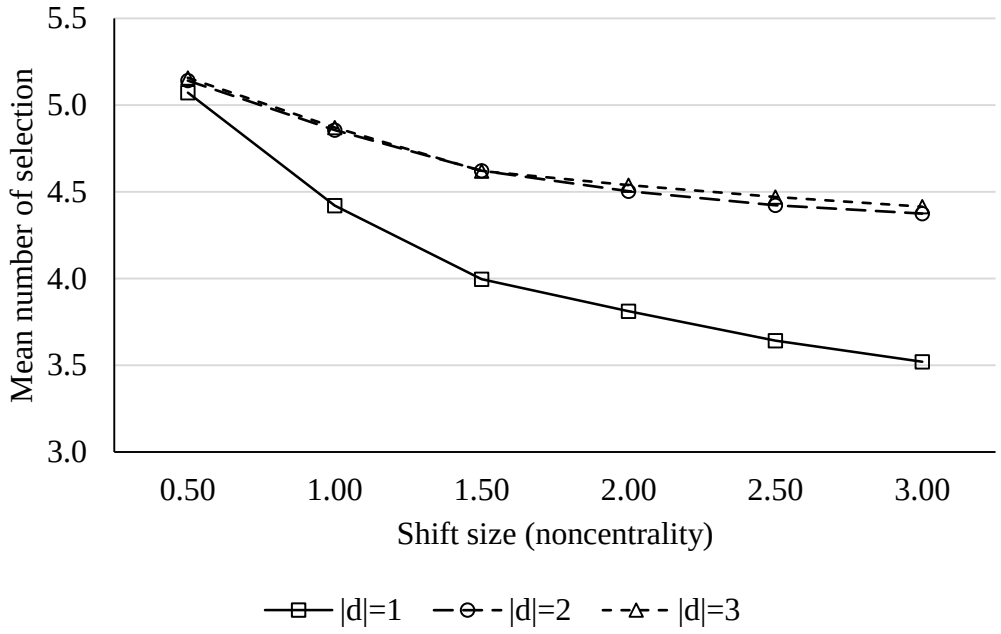


Figure 4.2 Mean number of selection according to the shift size

4.4. Conclusion

In this chapter, we propose a group variable selection based SPC chart for monitoring high dimensional processes especially when the process data have grouped structure and when the behavior of the process shift follows a groued pattern. In multistage processes, for example, we can readily observe the grouped structure of the process data. Under sparsity of both between groups and within-a-group patterns, we apply the concept of sparse group LASSO and modify it appropriately. Specifically, we adjust the parameter settings as a ratio between 0 and 1, rather than allowing large values. Compared to the existing VS-based methodologies such as VSMSPC and VSMEWMA, the proposed method selects the variables adaptively to the shift size and the number of shifted variables, which leads to the flexibility of the chart. We illustrate the proposed chart through the extensive simulation

studies and show that the proposed method outperforms when the shift tends to occur in sparse groups where there is a certain structure of correlation between groups and within a group.

Although the extensive simulation studies demonstrate that the proposed chart generally outperforms existing multivariate SPC charts and recent variable selection-based control charts in terms of out-of-control ARL, there is still the challenging issue of determination of two parameters to control sparsity. Adaptive changing (c_1, c_2) in consecutive hypothesis tests will be an interesting potential work.

CHAPTER 5

RIDGE PENALIZED LIKELIHOOD-BASED SPC CHART

5.1. Introduction

In Chapters 3 and 4, we investigate the monitoring scheme by adopting the diagnosis procedure before monitoring the process referred to VS-based chart. Although the existing VS-based approaches including the proposed method investigated in Chapters 3 and 4 intuitively have advantages in monitoring high-dimensional processes under sparsity, they have several challenges in the diagnosis procedures. The detection power of the VS-based charts may deteriorate when the performance of a diagnosis procedure is poor since VS procedure may exclude suspicious variables out of the selected set (Jiang *et al.*, 2012). For example, as the shift size becomes smaller in highly correlated data structure, the accuracy of a diagnostic procedure decreases resulting in the poor performance of VS-based charts. Further, the VS-based charts using forward VS or LASSO demand extensive computational effort due to the complex form of the optimization problem incorporating L_0 or L_1 norm types of penalties. Specially, when the dimension p is substantially large, the computation becomes much more challenging, which is not practical for online process monitoring.

In this chapter, we present a penalized likelihood-based approach for monitoring high-dimensional processes. Different from the existing VS-based charts, the proposed approach does not select variables, rather "shrinks" all process mean estimates towards zero. Using L_2 regularization, the performance of the chart can be interpreted in terms of the potential

shift direction. For example, it is expected that the chart becomes more sensitive to the out-of-control signals when the shift occurs along with the direction of small shrinkage and vice versa. In addition to the interpretability of the performance of the chart based on the potential shift information, the closed form solution of the proposed approach makes the practical implementation of online monitoring significantly convenient.

In spite of the advantages of the proposed chart, the following challenging issues remain; (a) how to explore a probability distribution of the chart statistic, which enables a practical use of the chart, e.g. to obtain control limits without extensive simulations and to predict the power of the chart with respect to the potential out-of-control scenarios; and (b) how to analyze the run length performances according to the process shift scenarios, i.e., locations and sizes of the shifts, when the chart is not directionally invariant as it occurs in many charts for monitoring high-dimensional processes. We address these issues and explore theoretical properties in the remainder of the chapter.

We compare the performance of the chart with to the existing VS-based charts specially in a highly correlated data structure in detecting small process changes. Moreover, according to the theoretical boundary of the 'performance region', we discuss the improvement of the proposed chart by adjusting the penalty parameter with potential shift direction. Furthermore, we illustrate that the proposed chart can be incorporated into other existing charts including EWMA and VS-based charts and improve its performance.

This chapter is organized as follows. Section 5.2 investigates the framework of our methodology based on the L_2 penalized likelihood and explores the theoretical properties of the proposed chart. Section 5.3 demonstrates the performance, followed by the investigation of several extensions of the proposed chart in Section 5.4. In Section 5.5, we introduce an illustrative example to demonstrate the practical implementation of the proposed chart. The chapter concludes with a further discussion and suggestions of future work in Section 5.6.

5.2. Proposed Approach

5.2.1. Framework of L_2 Regularization-Based Control Chart

With the assumption of normality, a generalized likelihood ratio test (GLRT) statistic is generally used for testing the mean and variance changes. The hypotheses for GLRT are $H_0 : \boldsymbol{\theta} \in \Omega_0$ and $H_1 : \boldsymbol{\theta} \in \Omega_1$, where $\boldsymbol{\theta}$ is a parameter of multivariate normal distribution, and Ω_0, Ω_1 are the parameter spaces for in-control and out-of-control states, respectively. In this chapter, the main interest is detecting change in the process mean, and we particularly assume that the covariance matrix is known and fixed over time. Without loss of generality, we assume $\boldsymbol{\mu} = \mathbf{0}$. We consider a p -dimensional observation vector $\mathbf{x}_t \sim N_p(\boldsymbol{\mu}_t, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}_t = \mathbf{0}$ for $t = 1, 2, ..., \tau$ and $\boldsymbol{\mu}_t \neq \mathbf{0}$ for $t = \tau + 1, ...,$ where τ is a unknown change point. Then the GLRT statistic denoted by $\lambda(\mathbf{x}_t)$ is given as

$$\lambda(\mathbf{x}_t) = \frac{L(\mathbf{x}_t, \mathbf{0})}{\max_{\boldsymbol{\mu}\in\Omega_1} L(\mathbf{x}_t, \boldsymbol{\mu})}$$

where the likelihood of $\mathbf{x}_t$ is given by

$$L(\mathbf{x}_t, \boldsymbol{\mu}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}_t - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_t - \boldsymbol{\mu})\right].$$

The logarithm of the GLRT statistic can be simplified as

$$\log \lambda(\mathbf{x}_t) = \min_{\boldsymbol{\mu}\in\Omega_1}\left(-2\mathbf{x}_t^T \boldsymbol{\Sigma}^{-1}\mathbf{x}_t + (\mathbf{x}_t - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_t - \boldsymbol{\mu})\right). \tag{5.1}$$

Eq. (5.1) is reduced in terms of $\boldsymbol{\mu}$ ignoring the constant term using Cholesky decomposition for the inverse of covariance matrix $\boldsymbol{\Sigma}^{-1} = \mathbf{R}^T\mathbf{R}$ as follows.

$$L_0(\boldsymbol{\mu}) = \min_{\boldsymbol{\mu}\in\Omega_1} \|\mathbf{z}_t - \mathbf{R}\boldsymbol{\mu}\|_2^2 \tag{5.2}$$

where $\mathbf{z}_t = \mathbf{R}\mathbf{x}_t$.

Under H_1, the current observation can be viewed as $\mathbf{x}_t = \boldsymbol{\mu}_0 + \delta\mathbf{d} + \boldsymbol{\varepsilon}_t$ where $\boldsymbol{\mu}_0$ is an in-control mean, δ and $\mathbf{d}$ are the shift size and its direction, respectively. With an

assumption of sparsity of an out-of-control scenario, the direction vector **d** consists of mostly zero values and a few non-zeros which are the true mean shifts. Then, the following penalized likelihood is obtained by introducing the regularization term in Eq. (5.2)

$$L(\boldsymbol{\mu}) = \min_{\boldsymbol{\mu}\in\Omega_1} \left\{ \left\| \mathbf{z}_t - \mathbf{R}\boldsymbol{\mu} \right\|_2^2 + \kappa\, p(\boldsymbol{\mu}) \right\} \tag{5.3}$$

where $p(\bullet)$ is a penalty function and $\kappa(\geq 0)$ is a complexity parameter that controls the amount of penalty. By introducing an L_2 penalty function in Eq. (5.3), we obtain the following penalized likelihood function.

$$L(\boldsymbol{\mu}) = \min_{\boldsymbol{\mu}\in\Omega_1} \left\{ \left\| \mathbf{z}_t - \mathbf{R}\boldsymbol{\mu} \right\|_2^2 + \kappa \left\| \boldsymbol{\mu} \right\|_2^2 \right\},$$

which has an exact solution as

$$\boldsymbol{\mu}_t^* = \arg\min_{\boldsymbol{\mu}} L(\boldsymbol{\mu}) = \left(\mathbf{R}^T\mathbf{R} + \kappa\mathbf{I} \right)^{-1} \mathbf{R}^T \mathbf{z}_t. \tag{5.4}$$

Then, based on the above penalized mean estimate, we can monitor the process mean with the statistic $Q = \mathbf{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^*$ (Jiang *et al.*, 2012, Wang and Jiang, 2009) and use a correlation matrix as covariance matrix for convenience. By replacing $\mathbf{R}^T\mathbf{R} = \boldsymbol{\Sigma}^{-1}$ and $\mathbf{z} = \mathbf{R}\mathbf{x}$, the chart statistic becomes

$$Q(\kappa) = \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^* = \mathbf{x}^T \left(\boldsymbol{\Sigma} + \kappa \boldsymbol{\Sigma}^2 \right)^{-1} \mathbf{x}, \tag{5.5}$$

and the chart signals when $Q(\kappa) > H_\alpha(\kappa)$, where α is a predetermined significance level. Hereinafter, we refer to the chart as a ridge-based multivariate SPC (RMSPC) chart. Since the monitoring statistic of the RMSCP chart does have an explicit form in terms of $\mathbf{x}$ and the matrix $\mathbf{A} = \boldsymbol{\Sigma} + \kappa \boldsymbol{\Sigma}^2$ is positive semi-definite, we obtain the approximate distributions of $Q(\kappa)$ under H_0 and H_1 through the following proposition.

***Proposition* 5.1.** For testing the null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ against the alternative hypothesis $H_1 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 + \delta \mathbf{d}$,

a. the probability distribution of the RMSPC chart statistic under H_0 follows approximately a gamma distribution with a shape parameter k and a scale parameter θ such that

$$k = \frac{\left(\sum_{i=1}^{p} w_i \right)^2}{2 \sum_{i=1}^{p} w_i^2} \text{ and } \theta = \frac{2 \sum_{i=1}^{p} w_i^2}{\sum_{i=1}^{p} w_i}$$

where $w_i = (1 + \kappa \lambda_i)^{-1}$ and λ_i is defined as the eigenvalues of $\boldsymbol{\Sigma}$,

b. the probability distribution of the chart statistic under H_1 with respect to the shift direction $\boldsymbol{\mu}_1$ follows approximately a noncentral chi-square distribution with a noncentrality δ_χ and a degree of freedom l, which holds the following:

$$\Pr\left(Q(\kappa) > L \mid \boldsymbol{\mu}_1\right) \approx \Pr\left(\chi_l^2(\delta_\chi) > v\sigma_\chi + \mu_\chi\right)$$

where $v = (L - \mu_Q)/\sigma_Q$, μ_Q, μ_χ, σ_Q and σ_χ are the means and the standard deviations of $Q(\kappa)$ and $\chi_l^2(\delta_\chi)$, respectively.

The proof of this proposition is shown in Appendix A. When $\kappa = 0$, the distributions under H_0 and H_1 reduce to exactly the central and the noncentral chi-square distributions, respectively. Proposition 5.1 enables the practical use of the chart for the determination of the control limits and the evaluation of the chart performance under various shift scenarios with a specific nonzero value, κ.

5.2.2. Directionally Variant Property of the RMSPC

Hotelling T^2 chart has a directional invariance property, so its performance depends only on a noncentrality parameter regardless of the shift direction. However, the performance of the proposed chart depends on the shift direction due to the nature of shrinkage through the ridge regularization. The direction of shrinkage of the proposed chart occurs proportionally in accordance with the correlation structure. The amount of the shrinkage

of the proposed approach is obtained as follows, through a singular value decomposition of $\mathbf{R} = \mathbf{UDV}^T$,

$$\mathbf{R}\hat{\boldsymbol{\mu}}^{ridge} = \mathbf{R}\left(\mathbf{R}^T\mathbf{R} + \kappa\mathbf{I}\right)^{-1}\mathbf{R}^T\mathbf{z} = \mathbf{UD}\left(\mathbf{D}^2 + \kappa\mathbf{I}\right)^{-1}\mathbf{DU}^T\mathbf{z} = \sum_{j=1}^{p}\mathbf{u}_j\frac{d_j^2}{d_j^2+\kappa}\mathbf{u}_j^T\mathbf{z} \tag{5.6}$$

where d_j is the j^{th} diagonal entry in the matrix $\mathbf{D}$. The diagonal elements are referred to as singular values which satisfy $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$, and $\mathbf{U}$ and $\mathbf{V}$ are complex unitary matrices. We observe that the matrix $\mathbf{V}$ diagonalizes $\boldsymbol{\Sigma}^{-1}$ since $\boldsymbol{\Sigma}^{-1} = \mathbf{R}^T\mathbf{R} = \mathbf{VD}^2\mathbf{V}^T$. Thus, when the eigenvalues of the correlation matrix $\boldsymbol{\Sigma}$ are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ and their corresponding eigenvectors are the columns of $\mathbf{P}$ in order, the identity that $d_j = 1/\sqrt{\lambda_j}$ holds for all $j = 1,..., p$. From the Eq. (5.6), the maximum shrinkage is directly associated with the largest eigenvalue of the covariance matrix $\boldsymbol{\Sigma}$. Therefore, this can be interpreted that the maximum shrinkage takes place along the direction of the eigenvector corresponding to the largest eigenvalue, which is also the direction of maximum variance. It directly implies that the chart performance would be different based on the direction of the mean shift, which means that the RMPC chart is directionally variant.

We also notice that when all variables are independent, the shrinkage occurs equally proportional to the all possible directions of the shifts with the amount of $(1+\kappa)^{-1}$, since the eigenvalues $\lambda_j = 1$ for all j. Then, the monitoring statistic becomes $(1+\kappa)^{-1}\mathbf{x}^T\mathbf{x}$,

which is a simple product of chi-square statistic. Consequently, the performance of RMSPC becomes identical to that of T^2 when all variables are independent.

5.2.3. Theoretical Performance Analysis

While the theoretical performance analysis of existing VS-based charts according to the directions of out-of-control signals is intractable, we perform theoretical analysis of performance of the proposed procedure. Suppose that the mean shift lies on the direction of the maximum variance, which corresponds to the largest eigenvalue. Then, the out-of-control ARL of the proposed chart would deteriorate because the chart shrinks the coefficients of the shifted variables as it considers them as noises. Conversely, when the shift occurs along the direction of a small variance, the chart keeps the shift information but shrinks other coefficients that are mostly random noises. Figure 5.1 shows an example with a bivariate normal distributed data. The performance of the proposed procedure would be better than Hotelling T^2 if the shift direction lies in the shaded area, and vice versa.

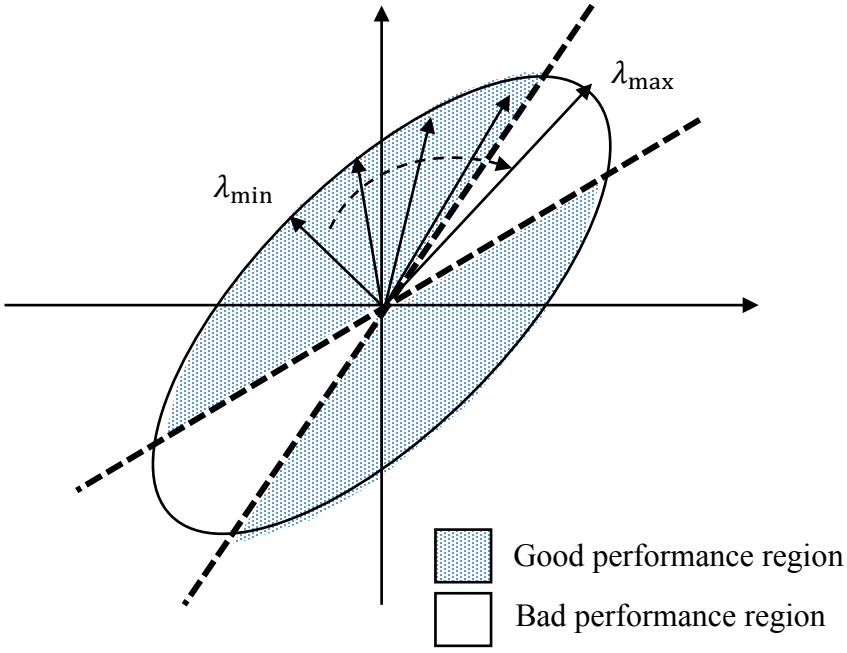


Figure 5.1. An example of performance region in bivariate normal distribution

It is challenging to obtain a theoretical 'boundary' – e.g., two hyperplanes in Figure 5.1 – that exactly distinguishes two performance regions even in a low dimension case. However, we readily speculate that the performance of the chart would become better as the mean is drifted closer to the direction of the last eigenvector corresponding to the smallest eigenvalue, and vice versa. In geometry, the eigenvalues of a correlation matrix imply an Euclidian distance from the center of the hyper-ellipsoid. When a vector **x** satisfies $\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x} = c^2$ where *c* is an arbitrary constant, for example, the lengths of the principle axes are computed as $c\sqrt{\lambda_i}$. Proposition 5.2 shows that when the shift occurs along the direction of the eigenvector with an eigenvalue in a certain range (i.e., a 'good' performance region

in Figure 5.1), the power of the proposed procedure is higher than that of Hotelling T^2 chart.

***Proposition* 5.2.** The power of RMSPC with a positive value of κ is higher than Hotelling T^2 chart when the mean shift occurs along the direction of the eigenvector corresponding to λ_i that satisfies

$$\lambda_i < \frac{1}{\kappa}\left[\frac{H_{T^2,\alpha}}{H_{RMSPC,\alpha}(\kappa)} - 1\right], \text{ for all } i = 1, 2, ..., p, \tag{5.7}$$

where $H_{T^2,\alpha}$ and $H_{RMSPC,\alpha}(\kappa)$ are boundaries with a significance level α for a T^2 chart and an RMSPC, respectively.

5.2.4. Determination of κ

The parameter κ controls the shrinkage and sparsity and is determined via cross-validation, Akaike information criterion (AIC), Bayesian information criterion (BIC), or other model selection criteria (Park *et al.*, 2012). Determination of the optimal κ for estimating $\boldsymbol{\mu}$ is related to location and the size of process shifts, correlation structure and the sparsity of potential shifts. Without precise prior information of the off target process, it may be challenging to obtain the optimal κ analytically. Other penalized likelihood based VS charts attempt to overcome this issue by introducing an additional parameter s which is the number of selections (or maximal number of selections) instead of determining the precise

value of penalty parameter. Moreover, the parameter, s is predetermined by practitioner's know-how and experience and is fixed over time. Therefore, this approach may not be appropriate in detecting shifts when s is not well-determined. Although the setting, $s = p$ is always an option, it can increase the computational time and may not be practical for online monitoring (Zou and Qiu, 2009).

Instead of introducing another parameter, the parameter κ is determined by adopting the approach of minimizing mean squared error as follows (Alkhamisi and Shukur, 2007)

$$\kappa = \frac{p}{\sum_{i=1}^{p} \hat{\mu}_i^2} + \frac{1}{\lambda_{\min}}$$

where $\hat{\boldsymbol{\mu}}$ is an ordinary least square (OLS) estimator, i.e., $\hat{\boldsymbol{\mu}} = \left(\mathbf{R}^T\mathbf{R}\right)^{-1}\mathbf{R}^T\mathbf{z}$, and $\lambda_{\min}$ is the smallest eigenvalue of $\boldsymbol{\Sigma}$. As the covariance matrix becomes more ill-conditioned as in the case of high correlation among variables, $\lambda_{\min}$ decreases further resulting in a large penalty, and vice versa.

Since $\hat{\boldsymbol{\mu}}$ equals the current observation, i.e., $\hat{\boldsymbol{\mu}}_t = \mathbf{x}_t$ for $t = 1, 2, ...,$ then κ is determined as $\kappa = pE\left(\|\mathbf{x}\|_2^{-2}\right) + \lambda_{\min}^{-1}$, where $\|\cdot\|_p$ represents a p -norm. Since $\|\mathbf{x}\|_2^{-2}$ follows approximately an inverse gamma distribution under H_0 from Proposition 5.1,

$E\left(\|\mathbf{x}\|_2^{-2}\right)=\nu_2/(\nu_1-1)$, where a shape parameter, $\nu_1=\|\boldsymbol{\lambda}\|_1^2/\left(2\|\boldsymbol{\lambda}\|_2^2\right)$ and a scale parameter, $\nu_2=\|\boldsymbol{\lambda}\|_1/\left(2\|\boldsymbol{\lambda}\|_2^2\right)$ when $\nu\geq 1$. The expected value of the inverse gamma distribution does not exist when the shape parameter is less than 1, which may occur when $\lambda_{\min}$ is close to zero with a small p. However, in this case, the first term $pE\left(\|\mathbf{x}\|_2^{-2}\right)$ can be ignored since $\lambda_{\min}$ mostly dominates the value of κ. Thus, depending on the estimated value of the scale parameter, κ can be determined explicitly as

$$\kappa=\begin{cases}\dfrac{p\|\boldsymbol{\lambda}\|_1}{\|\boldsymbol{\lambda}\|_1^2-2\|\boldsymbol{\lambda}\|_2^2}+\dfrac{1}{\lambda_{\min}}, & \|\boldsymbol{\lambda}\|_1^2\leq 2\|\boldsymbol{\lambda}\|_2^2\\ \dfrac{1}{\lambda_{\min}}, & \|\boldsymbol{\lambda}\|_1^2>2\|\boldsymbol{\lambda}\|_2^2\end{cases}$$

where $\boldsymbol{\lambda}$ is a vector whose entries are the eigenvalues of $\boldsymbol{\Sigma}$.

5.3. Performance Comparisons

In this section, we illustrate the general performance of RMSPC by comparing the out-of-control average run length (ARL_1) performance of the RMSPC chart with VSMSPC and Hotelling T^2 according to various shift scenarios. A spatial relationship of the variables is considered, i.e., $\rho(x_{ij},x_{kl})=c^{\sqrt{(i-k)^2+(j-l)^2}}$ for $i,k=1,2,...,p_X$ and $j,l=1,2,...,p_Y$, where p_X and p_Y are the horizontal and vertical indices for one frame of spatial data information, respectively, so that $p_X\times p_Y=p$, and $0<c<1$ (Bao *et al.*, 2014, Liu *et al.*, 2015, Wang

et al., 2016). The vectorized data matrix is regarded as an observation vector whose distribution is multivariate normal. To compare with the VSMSPC chart, the cardinality of the set of potentially changed variables to be estimated as nonzero, $s(< p)$, should be priori determined. We choose a moderate size of selection as $s / p = 0.2$. The penalty parameter of the RMSPC is determined as described in Section 5.2, and the in-control ARL (ARL_0) is set to 200 for all charts. Along with the ARL_1, we also assess the performance of the charts through relative mean index (RMI) suggested by Han and Tsung (2006) which is defined as

$$\text{RMI} = \frac{1}{N}\sum_{i=1}^{N}\frac{\text{ARL}_{\boldsymbol{\mu}_i}(X) - \text{ARL}^{*}_{\boldsymbol{\mu}_i}}{\text{ARL}^{*}_{\boldsymbol{\mu}_i}}$$

where N is the number of out-of-control scenarios considered, $\text{ARL}_{\boldsymbol{\mu}_i}(X)$ is the ARL_1 of the given chart X, and $\text{ARL}^{*}_{\boldsymbol{\mu}_i}$ denotes the smallest ARL_1 among the charts under the shift, $\boldsymbol{\mu}_i$. Thus, the RMI provides a relative efficiency with respect to the out-of-control scenarios considered. Table 5.1 shows the ARL_1 performance and RMI for 25 shift scenarios based on 20,000 replications. The dimensions of the data matrix are assumed $p_X = 2$ and $p_Y = 5$, so that $p = 10$, and the correlation parameter is set as $c = 0.9$. The values in the parentheses represent the $(\text{ARL}_{\boldsymbol{\mu}_i}(T) - \text{ARL}^{*}_{\boldsymbol{\mu}_i})/\text{ARL}^{*}_{\boldsymbol{\mu}_i}$. In addition, the shift scenario, $\{1, 3\} = 0.25$, for example, denotes that the 1st and 3rd variables are shifted with a magnitude of mean shift of 0.25.

From Table 5.1, we observe several findings. In most of the shift cases, the proposed RMSPC shows the smallest ARL_1 performance. In a few cases with sparse shifts, the RMSPC underperforms VSMSPC since VSMSPC may construct more sensitive monitoring statistic based on correct identification of potentially changed variables. However, in many cases with sparse shifts and most of the non-spare shifting cases, the proposed chart outperforms VSMSPC. Specially, as the number of shifted variables increases, the VSMSPC performs poorly because the chart is limited by the parameter s, while the RMSPC still shows a good detection power. For example, in the last case, the shift occurs in all variables, but ARL_1 of RMSPC is almost half of that of T^2. In addition to the case-by-case performance, the RMSPC has the smallest RMI, which indicates that the proposed chart performs well in general among three charts. Lastly, we can also observe that the performance of RMSPC is not uniformly better or worse than the T^2. This implies that the performance of RMSPC is not directionally invariant. Thus, two cases where RMSPC underperforms T^2 can be interpreted that the directions of the shifts lie on a 'bad' performance region as in Figure 5.1.

Table 5.1. Performance comparison of RMSPC with T^2 and VSMSPC under various shift cases ($p = 10,$ ARL_0=200)

Shifts	T^2	VSMSPC	RMSPC
$\{1\} = 0.25$	133.54(0.05)	127.49(0.00)	127.54(0.00)
$\{2\} = 0.5$	36.47(0.40)	26.22(0.01)	26.04(0.00)
$\{3\} = 0.75$	10.67(0.49)	7.17(0.00)	7.49(0.05)
$\{5\} = 0.5$	17.84(0.51)	11.81(0.00)	16.2(0.37)

$\{7\}=0.25$	117.03(0.20)	108.85(0.11)	97.81(0.00)
$\{9\}=0.75$	10.72(0.54)	6.96(0.00)	7.58(0.09)
$\{1,3\}=0.25$	84.05(0.33)	78.13(0.23)	63.28(0.00)
$\{1,5\}=0.5$	21.88(0.15)	19.01(0.00)	19.91(0.05)
$\{1,7\}=0.75$	3.97(0.47)	3.19(0.18)	2.69(0.00)
$\{1,9\}=0.5$	16.04(0.25)	13.36(0.04)	12.78(0.00)
$\{2,4\}=0.25$	75.09(0.44)	67.44(0.30)	51.98(0.00)
$\{3,8\}=0.5$	20.68(0.21)	20.52(0.21)	17.02(0.00)
$\{3,9\}=0.75$	3.27(0.41)	2.75(0.19)	2.32(0.00)
$\{1,3,7\}=0.25$	56.6(0.71)	47.09(0.42)	33.16(0.00)
$\{2,4,8\}=0.5$	6.76(0.72)	6.33(0.61)	3.94(0.00)
$\{4,5,10\}=0.75$	11.28(0.00)	14.79(0.31)	11.71(0.04)
$\{1,2,4,5\}=0.25$	79.82(0.00)	90.13(0.13)	80.99(0.01)
$\{3,4,6,7\}=0.5$	10.02(0.10)	14.65(0.61)	9.08(0.00)
$\{3,5,7,9\}=0.75$	1.37(0.20)	1.49(0.31)	1.14(0.00)
$\{1,2,3,5,7\}=0.25$	65.05(0.24)	62.94(0.20)	52.34(0.00)
$\{2,3,4,8,10\}=0.5$	7.76(0.47)	8.19(0.55)	5.29(0.00)
$\{3,5,6,9,10\}=0.75$	1.88(0.22)	2.05(0.33)	1.54(0.00)
$\{1,3,5,7,9\}=0.25$	31.62(0.97)	26.21(0.63)	16.03(0.00)
$\{2,4,6,8,10\}=0.5$	3.23(0.62)	3.41(0.71)	1.99(0.00)
$\{1,3,5,7,9\}=0.5$; $\{2,4,6,8,10\}=0.25$	29.5(0.84)	26.28(0.64)	16.01(0.00)
RMI	0.382	0.269	0.024

Next, we present the performance of the chart in detecting small mean shifts. When the shift size is small, the VSMSPC chart may perform poorly since the VS would possibly

misidentify the potentially changed variables. To assess the performance in such cases, we compute the relative efficiency (RE) of the RMSPC and VSMSPC compared to T^2, under various shift sizes illustrated in Figure 5.2, which is calculated as a ratio of $\Pr(Q_{RMSPC} > H_{RMSPC,\alpha}(\kappa) \mid \boldsymbol{\mu}_1)$ over $\Pr(Q_{T^2} > H_{T^2,\alpha} \mid \boldsymbol{\mu}_1)$. Thus, RE greater than 1 indicates better power than T^2. Note that the RE shows the relative power for a specific shift direction as the shift size varies, while RMI compares overall performance considering multiple shift directions. The shift size is determined as a noncentrality parameter, i.e., $\delta = \sqrt{\mathbf{d}^T \boldsymbol{\Sigma}^{-1} \mathbf{d}}$, where $\mathbf{d}$ is a shift vector. In this experiment, we consider nine different shift cases in terms of the number of shifted variables and dimensions. Denoting s_0 as an index set of shifted variables, the columns of Figure 5.2 represent three shift scenarios with different shifted variables; $s_0 = \{2\}, \{6,8\}$ and $\{1,3,5,7,9\}$. The rows in Figure 5.2 represent different dimensions; $p = 10(2\times5)$, $p = 25(5\times5)$, and $p = 50(5\times10)$. The dotted horizontal line represents the RE of T^2, the solid and dash lines show the RE of RMSPC and VSMSPC compared to T^2, respectively. The results illustrated in Figure 5.2 confirm that the proposed RMSPC performs relatively better than T^2 over all shift sizes, while the VSMSPC underperforms T^2 when shift size is relatively small due to the misidentification issue of VS. Moreover, as the dimension increases, the probability of selecting 'true' shifted variables would decrease resulting in a poor performance of VS-based chart. On the other hand, the RMSPC shows a good performance of detection when the dimension is high.

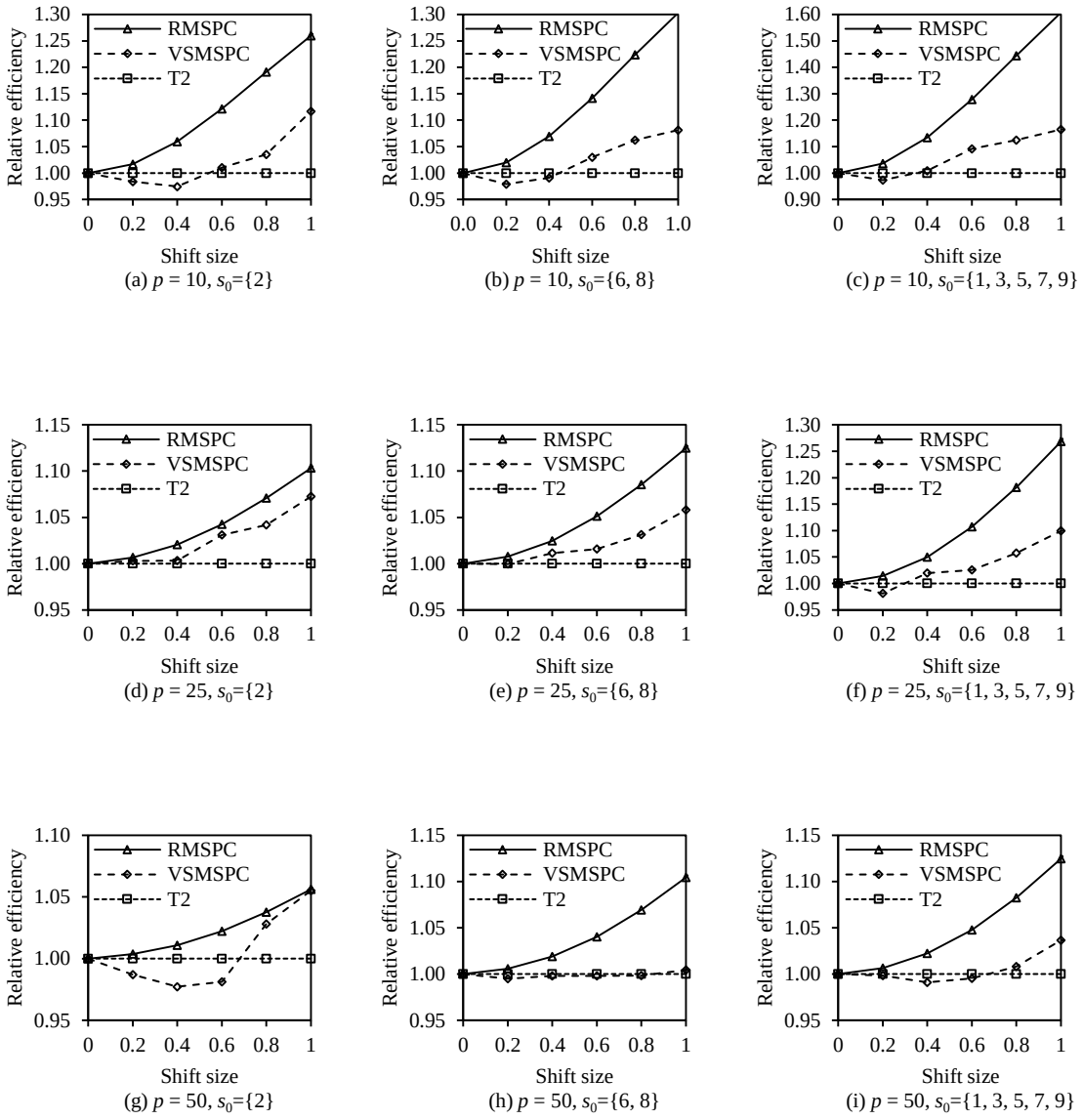


Figure 5.2. Relative efficiency of RMSPC and VSMSPC to T^2

5.4. Extensions of RMSPC

In this section, we provide some extensions of the proposed approach. First, we develop an adaptive RMSPC that changes the parameter κ adaptively based on the estimated shift

direction. We also develop an EWMA version of RMSPC, namely, RMEWMA. Furthermore, we develop VS-based RMSPC/RMEWMA by incorporation the VS algorithm into RMSPC/RMEWMA. We compare the performance of these charts with their corresponding existing charts.

5.4.1. Adaptive RMSPC

The performance of RMSPC depends on the direction of the shift. Thus, depending on the potential shift direction at time *t*, we select κ value adaptively. The chart statistic of an adaptive RMSPC, namely ARMSPC, at time t can be written as

$$Q_{ARMSPC}(\kappa_t) = \mathbf{x}_t^T \left(\boldsymbol{\Sigma} + \kappa_t \boldsymbol{\Sigma}^2 \right)^{-1} \mathbf{x}_t ,$$

The chart signals when $Q_{ARMSPC}(\kappa_t) > H_{ARMSPC,\alpha}(\kappa_t),$ and the control limits, $H_{ARMSPC,\alpha}(\kappa_t)$ can be determined by simulation for the desired ARL_0. The value of κ_t at time *t* can be determined adaptively depending on the estimated shift direction as follows.

$$\kappa_t = \begin{cases} \kappa_0, & \text{RPI}_t \geq \gamma \\ 0, & \text{RPI}_t < \gamma \end{cases}$$

where κ_0 is predetermined penalty described in Section 2.4, and γ is a threshold value, and the RPI introduced in Appendix B represents the relative performance index of RMSPC with $\kappa = \kappa_0$ to T^2.

Based on the analysis of the performance according to the shift direction, we consider two extreme shifting cases that represent directions of the maximum shrinkage and the minimum shrinkage. Considering the same simulation condition as in Table 5.1, the two extreme directions of the shift are $\mathbf{e}_1$ and $\mathbf{e}_{10}$, i.e., the first and the last eigenvectors corresponding to the first and last eigenvalues. Figure 5.3 illustrates the RE of the ARMSPC chart to T^2 as γ varies when $\boldsymbol{\mu}_1 = \mathbf{e}_1$ (dotted line) and $\boldsymbol{\mu}_1 = \mathbf{e}_{10}$ (dashed line). The solid horizontal line represents the reference line referring to the relative efficiency of T^2 itself. When $\gamma = 0,$ it shows the best performance with the shift along $\mathbf{e}_{10}$ and the worst performance with the shift along $\mathbf{e}_1$. As γ increases, the performance with the shift $\mathbf{e}_1$ increases but that with the shift $\mathbf{e}_{10}$ decreases slightly. When γ is reasonably large, as expected, the chart behavior approaches that of $T^2,$ and when $\gamma = 0,$ ARMSPC is identical to RMSPC.

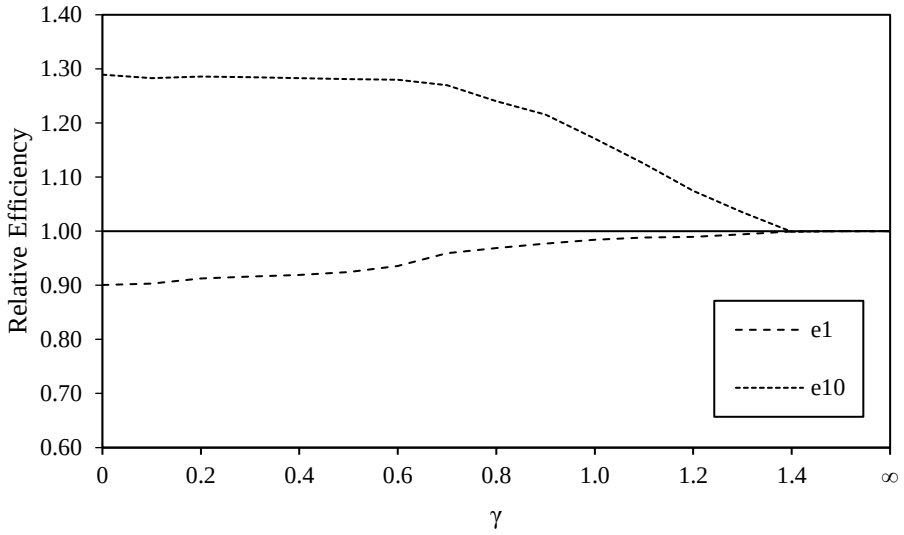


Figure 5.3. Relative efficiency of RMSPC as γ varies when $\boldsymbol{\mu}_1 = \mathbf{e}_1$ and $\boldsymbol{\mu}_1 = \mathbf{e}_{10}$

5.4.2. Extension to EWMA

We now consider the detection of a small mean shift and develop the EWMA version of the proposed chart, namely, RMEWMA. With a constant weight parameter r, the EWMA vector at time t is determined using all past observations as

$$\mathbf{w}_t = (1-r)\mathbf{w}_{t-1} + r\mathbf{x}_t, \quad t = 1, 2, \ldots$$

where $\mathbf{w}_0 = \mathbf{x}_0 = \boldsymbol{\mu}_0$. Letting $\boldsymbol{\Sigma}_{\mathbf{w}}$ as the covariance matrix of MEWMA, the monitoring statistic of RMEWMA can be obtained as

$$Q_{RMEWMA}(\kappa) = \mathbf{w}^T \left(\boldsymbol{\Sigma}_{\mathbf{w}} + \kappa \boldsymbol{\Sigma}_{\mathbf{w}}^2 \right)^{-1} \mathbf{w}.$$

Since the covariance of MEWMA in a steady state is asymptotically $\boldsymbol{\Sigma}_{\mathbf{w}} = r/(2-r)\boldsymbol{\Sigma}$, the identical procedure of inducing the directionally variant property to RMSPC holds for RMEWMA. Moreover, the adaptive RMEWMA (ARMEWMA) can be also derived by modifying RPI with respect to $\mathbf{w}$. Since the vector, $\mathbf{w}_t$, approaches a true shift vector as more past observations are accumulated, the RPI provides more appropriate information of the shift, specially when the process shift is relatively small.

In comparing the proposed approach, we consider the case with $p = 5$, which is initiated by Hawkins (1991) and is used by Jiang *et al.* (2012) and Zou and Qiu (2009) to assess the performance of VSMEWMA and LEWMA. Different from the proposed chart and VSMEWMA, the monitoring statistic of LEWMA is calculated as a maximal statistic among q possible selections. See Jiang *et al.* (2012) for detailed discussion of the difference between single monitoring statistic and maximal monitoring statistic. In this comparison, we illustrate the cases with $q = 3$ and $q = 5$ for LEWMA, and the cases with $s = 1, 2, 3$ and 4 for VSMEWMA. Twenty different types of out-of-control scenarios are taken into account shown in Table 5.2 in which the mean shift $\boldsymbol{\mu}_1$ is added to the process at time $\tau = 25$. The column of δ represents the noncentrality such that $\delta = \sqrt{\boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1}$. For a fair comparison, we consider the same simulation settings as in Jiang *et al.* (2012) and Zou and Qiu (2009); the ARL_0 is set to 500; the smoothing parameter $r = 0.2$; the covariance is set as same as in Hawkins (1991); and the results are based on 20,000

replications. The performance of the charts is assessed by ARL_1 as well as RMI, and the smallest ARL_1 is represented in bold characters.

Several findings can be deduced from Table 5.2. First, we observe that the ARMEWMA outperforms all other charts in most of the cases. It shows the smallest ARL_1 for 13 cases out of 20 shift cases. Considering that the shift sizes of the most cases are greater than 1 as a noncentrality parameter, the proposed chart would possibly show better performance than VS-based charts when the shift size decreases as shown in Figure 5.2. Second, although the proposed chart with $\gamma = 0$ outperforms VS-based charts in many cases, it also shows the poor ARL_1 performance for a few cases; e.g., cases 1) and 19), which leads to the poor overall performance (large RMI). Thus, without any information on the potential shifts such as magnitude and direction, one can achieve more 'balanced' performance by increasing the parameter γ. The proposed chart with $\gamma = 0.7$ shows the best overall performance with the best RMI, even though it does not present the smallest ARL_1 for any cases. With reasonably large γ, the performance of the chart would become identical to that of MEWMA.

Table 5.2. ARL_1 performance comparison of ARMEWMA with MEWMA, LEWMA and VSMEWMA ($p = 5,\ r = 0.2,\ ARL_0 = 500$)

Cases	μ_1	μ_2	μ_3	μ_4	μ_5	δ	MEWMA	LEWMA		VSMEWMA				ARMEWMA			
								$q=3$	$q=5$	$s=1$	$s=2$	$s=3$	$s=4$	$\gamma=0$	$\gamma=0.7$	$\gamma=1$	$\gamma=1.5$
1	0.91	0	0	0	0	1.00	17.3	14.6	14.9	14.4	15.6	16.0	17.0	34.4	21.5	17.4	17.2
2	0	0.36	0	0	0	1.01	17.0	13.9	14.3	13.2	14.4	15.5	16.0	12.5	13.4	15.3	16.9
3	0	0	0.48	0	0	1.00	17.3	14.6	15.0	13.8	15.1	16.1	16.6	14.5	15.9	17.3	17.3
4	0	0	0	0.34	0	1.00	17.2	14.2	14.6	13.4	14.7	15.8	16.5	12.4	13.4	15.2	17.3
5	0	0	0	0	0.46	0.99	17.9	14.9	15.2	13.9	15.3	16.2	16.8	14.6	15.9	17.6	17.7
6	0.36	0.36	0	0	0	1.07	15.0	13.4	13.7	13.4	13.2	13.8	14.3	12.1	12.9	14.1	14.7
7	0.54	0	0.54	0	0	1.15	12.8	12.4	12.4	12.9	12.2	12.3	12.7	12.4	12.9	12.9	12.8
8	0.32	0	0	0.32	0	1.06	15.2	13.4	13.6	13.1	13.3	14.1	14.7	11.9	12.7	14.2	15.2
9	0.49	0	0	0	0.49	1.14	13.2	12.4	12.5	12.4	12.0	12.3	12.7	12.1	13.0	13.0	13.1
10	0	0.54	0.54	0	0	1.41	8.79	8.75	8.88	12.0	8.21	8.25	8.41	7.76	8.20	8.75	8.78
11	0	1.6	0	1.6	0	2.64	3.48	3.59	3.57	8.99	3.89	3.68	3.50	3.93	3.91	3.48	3.48
12	0	0.28	0	0	0.28	1.15	13.0	11.1	11.3	10.3	11.3	12.0	12.4	9.56	10.1	11.3	12.8
13	0	0	0.28	0.28	0	1.14	13.0	11.3	11.4	10.6	11.5	12.2	12.5	9.57	10.2	11.5	12.9
14	0	0	1.26	0	1.26	2.24	4.28	4.41	4.34	8.69	4.41	4.20	4.20	4.75	4.80	4.28	4.26
15	0	0	0	0.56	0.56	1.42	8.60	8.60	8.70	12.2	8.2	8.19	8.28	7.72	8.10	8.64	8.60
16	0.01	−0.15	0.07	0.17	−0.09	1.07	15.0	12.3	12.5	11.5	12.8	13.7	14.3	10.6	11.3	13.0	14.8
17	0.07	−0.13	−0.40	0.19	0.35	1.30	10.1	10.0	10.1	12.3	9.34	9.43	9.66	9.17	9.74	10.4	10.0
18	0.4	0.63	−0.57	0.47	−0.68	2.08	4.74	4.97	4.94	9.30	5.09	4.65	4.67	5.03	5.25	4.75	4.74
19	−1.11	0.26	−0.17	0.34	−0.04	1.20	11.8	12.5	12.2	14.4	13.3	11.3	11.7	30.0	12.8	11.7	11.8
20	2.51	7.11	7.05	7.11	7.08	7.83	1.19	2.88	1.30	7.44	3.02	2.19	1.47	2.39	1.19	1.19	1.19
RMI							0.187	0.161	0.105	0.571	0.178	0.160	0.159	0.227	0.105	0.137	0.180

5.4.3. Extension to VS-Based Control Charts

In this subsection, we consider integrating the L_2 penalized likelihood estimator with other VS-based control charts. Although the VS-based charts may not perform well specially in detecting a small mean shift, this integration would provide the potential capability of the chart for overall performance with unknown shift information.

In the preceding extension, we utilize the current observation, $\mathbf{x}_t$ (or $\mathbf{w}_t$ in RMEMWA) in computing RPI, where the vector $\mathbf{x}_t$ or $\mathbf{w}_t$ is considered as a potential mean shift at time t, and provides the directional information for RPI. Under the sparsity assumption, therefore, more appropriate shift information can be obtained by selecting potentially changed variables via VS algorithms. Then, we propose the procedure of VS-integrated approach as follows. First, the small subset of potentially changed variables are selected through the diagnosis procedures. In this chapter, for simplicity, we use a conditional T^2 statistic as $T^2_{i|\Gamma} = T^2 - T^2_{\Gamma}$ where Γ is a set of all process variables except the i^{th} one, introduced by Runger (1996). The large value of the conditional statistic under H_1 indicates that the corresponding variable is responsible for the process change. Suppose that the $s(< p)$ number of nonzero components in $\boldsymbol{\mu}^*$ is selected accordingly. We then apply L_2 penalizing approach to estimate nonzero values for the selected variables. We denote the nonzero vector, $\tilde{\boldsymbol{\mu}}^*$ with size $s \times 1$ by dividing the column space p into two sets of variables, say suspicious (s) and unsuspicious set of variables ($p - s$). Letting $\tilde{\mathbf{R}}$ be a

matrix with corresponding s columns taken from $\mathbf{R}$ and introducing the penalty parameter, the nonzero values are estimated as

$$\tilde{\boldsymbol{\mu}}^{*}_{s\times 1} = (\tilde{\mathbf{R}}^{T}\tilde{\mathbf{R}} + \kappa \mathbf{I}_{s})^{-1}\tilde{\mathbf{R}}^{T}\mathbf{z}. \tag{5.8}$$

where $\mathbf{z} = \mathbf{R}\mathbf{x}$.

Now, the sparse vector, $\boldsymbol{\mu}^{*}_{t}$ can be utilized to determine the penalty; i.e., it is possible to compute whether this potential shift lies on a 'good' or 'bad' performance regions based on the potential shift direction $\boldsymbol{\mu}^{*}_{t}$. Then RPI given $\boldsymbol{\mu}^{*}$ is proportional to the value, $\mathbf{x}_t^{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^{*}_{t,RMSPC}$ over $\mathbf{x}_t^{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^{*}_{t,T^2}$, where $\boldsymbol{\mu}^{*}_{t,RMSPC}$ is the potential shift vector from Eq. (5.8) with $\kappa = \kappa_0$, and $\boldsymbol{\mu}^{*}_{t,T^2}$ is the one with $\kappa = 0$. According to the RPI and the parameter γ, the nonzero component of $\boldsymbol{\mu}^{*}$ for computing the monitoring statistic can be determined alternatively with $\kappa = \kappa_0$ or $\kappa = 0$.

In addition to the VS-based ARMSPC, we can also derive the EWMA version of VS-based ARMSPC, namely, VS-ARMEWMA. The vector, $\boldsymbol{\mu}^{*}$ is obtained based on an EWMA statistic, and thus it is expected to provide more appropriate directional information when small shift occurs.

Table 5.3 presents the ARL_1 performance of VS-ARMEWMA compared with VSMEWMA when $s = 2$ and $s = 3$. The simulation conditions and shift cases are the same as in Table 5.2. For the VS procedure in VSMEWMA, we use conditional T^2 statistic instead of FVS. Note that the capability of conditional T^2 statistic method in selecting variables is similar to that of FVS under normality. In VS-ARMEWMA, we consider $\gamma = 0$, 0.4 and 0.6. From the results in Table 5.3, we find that the VS-ARMEWMA outperforms VSMEWMA in 14 cases. In the cases where the performance of VS-ARMEWMA is worse than VSMEWMA, VS-ARMEWMA with the large γ shows similar performance as that of VSMEWMA. Figure 5.4 illustrates the 3-dimensional contour plot of the RMI of VS-ARMEWMA in terms of s and γ. The darker region represents the better performance in RMI. The contour surface shapes a hyperboloid with its smallest RMI when $s = 4$ and $\gamma = 0.6$. Thus, similar conclusions to Figure 5.4 are obtained from the Table 5.3; that is, the VS-ARMEWMA performs mostly better than VSMEWMA, but in some cases where the shift lies on a 'bad' performance region, the large γ results in the similar performance as that of VSMEWMA. However, we also notice that the chart may perform relatively worse than MEWMA when VS does not function well, and the shift is not sparse or lies on a 'bad' performance region; e.g., in cases 11 and 17-20.

Table 5.3. ARL_1 comparison of VS-ARMEWMA with VSMEWMA ($p = 5,\ r = 0.2,$ $ARL_0 = 500$)

Cases	VSMEWMA		VS-ARMEWMA					
			$s = 2$			$s = 3$		
	$s = 2$	$s = 3$	$\gamma = 0$	$\gamma = 0.4$	$\gamma = 0.6$	$\gamma = 0$	$\gamma = 0.4$	$\gamma = 0.6$
1	14.3	15.0	37.1	16.5	14.4	34.6	18.2	15.5
2	14.6	15.1	12.1	12.5	13.2	12.4	12.6	13.3
3	15.1	15.7	14.2	14.8	15.5	14.4	15.1	15.5
4	14.7	15.5	12.2	12.6	13.0	12.4	12.7	13.3
5	15.4	16.0	14.2	14.9	15.4	14.4	14.9	15.8
6	13.2	13.7	12.7	12.5	12.7	12.4	12.4	13.0
7	12.8	12.7	12.8	12.6	12.5	12.7	12.5	12.4
8	13.3	13.7	12.0	12.4	12.5	11.9	12.3	12.7
9	12.6	12.3	12.5	12.7	12.3	12.4	12.5	12.3
10	10.4	8.53	10.4	10.2	10.2	9.15	8.55	8.85
11	6.26	4.46	6.43	6.41	6.39	4.87	4.81	4.60
12	11.3	11.5	9.23	9.55	10.0	9.44	9.61	9.99
13	11.3	11.7	9.47	9.68	10.1	9.46	9.74	10.0
14	5.88	4.75	6.54	6.40	6.13	5.31	5.24	4.97
15	10.8	8.57	9.72	10.1	10.4	8.29	8.38	8.65
16	12.7	13.2	10.3	10.6	11.1	10.3	10.6	11.1
17	11.3	9.72	11.6	11.8	12.1	10.0	10.1	10.3
18	6.49	5.35	7.68	7.67	6.92	6.12	6.06	5.64
19	12.6	12.0	35.7	13.5	12.5	30.7	13.1	12.0
20	4.11	2.86	11.0	4.45	4.18	5.61	3.13	2.82

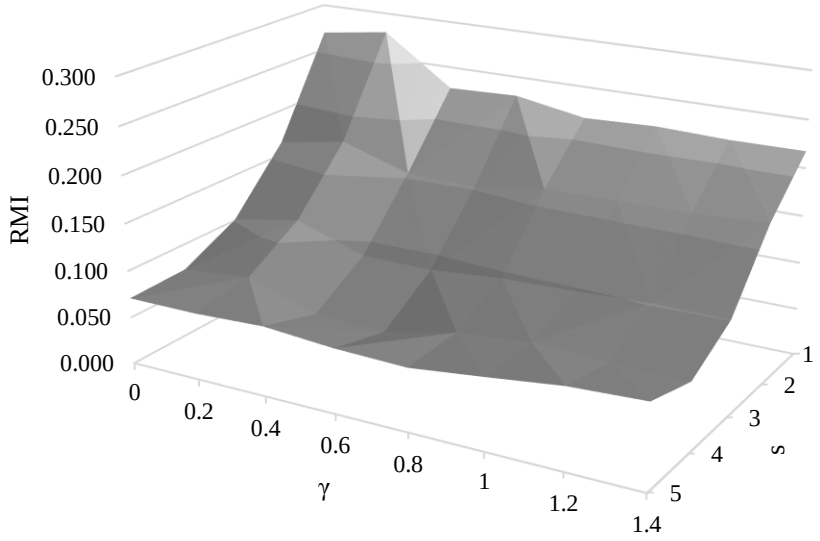


Figure 5.4. Contour plot of RMI of VS-ARMEWMA in terms of s and γ

5.5. Case Study

In this section, we present a case study of a milling process that produces a complex part to be used in the assembly of several devices. The part has multiple characteristics that include several depths, circular and rectangular pockets with high dimensional tolerances. The milling tools are replaced on regular basis to maintain the required high tolerance. However, any changes in the process (cooling rate, speed or variability in the feed rate and tool wear) result in changes in the dimensions. Figure 5.5 shows the final product produced by the milling process and its isometric view. Parts were produced and their dimensions were measured using a highly accurate coordinate measuring machine (CMM). The references for the dimensions to be monitored and its computer-aided design (CAD) drawings for part production are shown in Figure 5.6. Process parameters for production

are denoted as '*d*' and the parameters to be monitored are denoted as '*x*'. In this case study, we monitor ten process parameters, denoted as x_i for $i = 1,...,10$.

a) Produced part b) Isometric view of the part

Figure 5.5. Part features produced using high-speed milling process

Fifty parts are produced; the first twenty parts are produced under the normal conditions of the cutting process parameters. The sample covariance matrix of the in-control process is computed based on these twenty samples as shown in Table 5.4. These samples are then used in Phase II monitoring. The remaining thirty parts are produced under different setting of the machine that assumes the milling process parameters have changed due to one or more of the causes stated earlier leading to the change in some dimensions of the parts. Figure 5.7 shows the measurements for the fifty parts in which moderate positive shifts can be observed in x_7, x_8 and x_9.

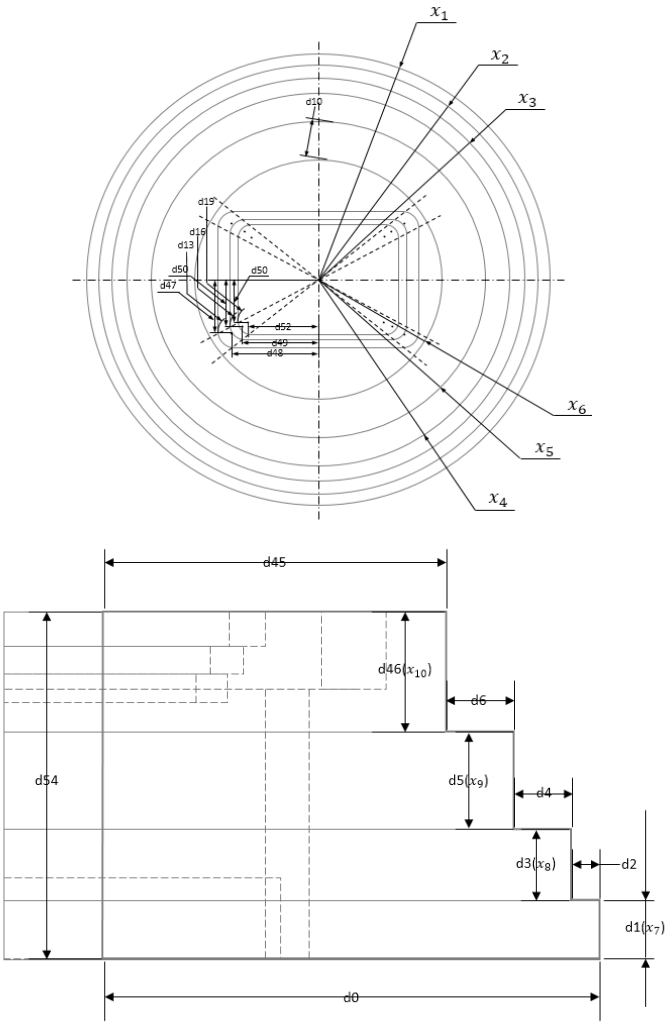


Figure 5.6. Part dimensions, e.g., diameters (left) and heights (right) of the cylinders

Table 5.4. A sample covariance matrix of the dimensions

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1	1.000	-0.093	0.069	-0.421	-0.186	0.431	-0.282	0.064	0.479	-0.011
x_2	-0.093	1.000	-0.058	-0.276	0.270	0.223	-0.237	-0.102	-0.098	-0.260
x_3	0.069	-0.058	1.000	-0.130	-0.252	0.272	-0.146	0.101	0.093	-0.362
x_4	-0.421	-0.276	-0.130	1.000	-0.123	-0.316	0.245	-0.415	-0.057	0.014
x_5	-0.186	0.270	-0.252	-0.123	1.000	-0.225	0.101	0.022	-0.328	0.160
x_6	0.431	0.223	0.272	-0.316	-0.225	1.000	-0.422	0.129	0.059	-0.068
x_7	-0.282	-0.237	-0.146	0.245	0.101	-0.422	1.000	-0.073	-0.008	0.235
x_8	0.064	-0.102	0.101	-0.415	0.022	0.129	-0.073	1.000	-0.085	-0.175
x_9	0.479	-0.098	0.093	-0.057	-0.328	0.059	-0.008	-0.085	1.000	-0.156
x_{10}	-0.011	-0.260	-0.362	0.014	0.160	-0.068	0.235	-0.175	-0.156	1.000

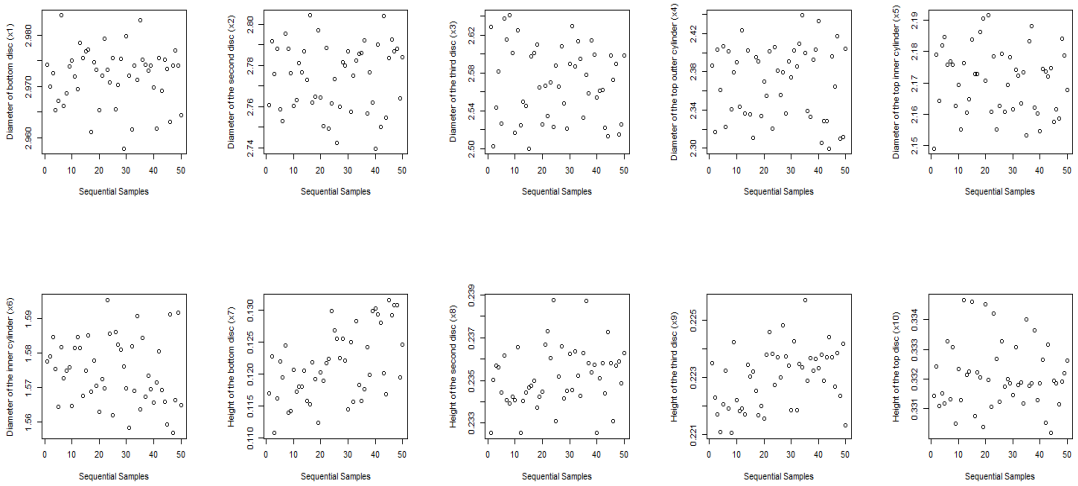


Figure 5.7. Raw data of the samples

We illustrate the implementation of the RMSPC and VSMSPC control charts using this dataset. The assumption of the normality is validated using Mardia's multivariate normality test (Mardia, 1970) with the significance level to 0.05. The parameter, κ for RMSPC is

determined as guided previously, $\kappa = 1.81$; the parameter, s (the number of selection) for VSMSPC chart is set as $s = 3$ assuming that the number of truly shifted variables is known so that VSMPSC results in the best performance in this case study. The control limits of RMSPC and VSMSPC are obtained as 8.02 and 11.73, respectively based on the in-control parameters. Figure 5.8 shows the monitoring paths of the charts over the thirty out-of-control samples; the solid line and dash line represent RMSPC and VSMSPC, respectively; and the control limits are represented by the horizontal lines. RMSPC chart refers to the left-hand-side vertical axis of Figure 5.8 and VMSPC chart refers to the right-hand-side axis.

From Figure 5.8, we observe that the proposed RMSPC successfully alarms for 25 samples out of 30, while VSMSPC does so only for 14 samples. It indicates that RMSPC chart is more sensitive to the out-of-control signals than VSMSPC. Since both charts may fail to detect the out-of-control signal when the shift size is small, the standardized distances of the measurements are shown in Figure 5.9. As shown, the Mahalanobis distances (Friedman *et al.*, 2001) of the samples, 5, 8, 12, 17 and 23 are relatively small, i.e., close to the in-control process mean, so that they are not captured by both charts.

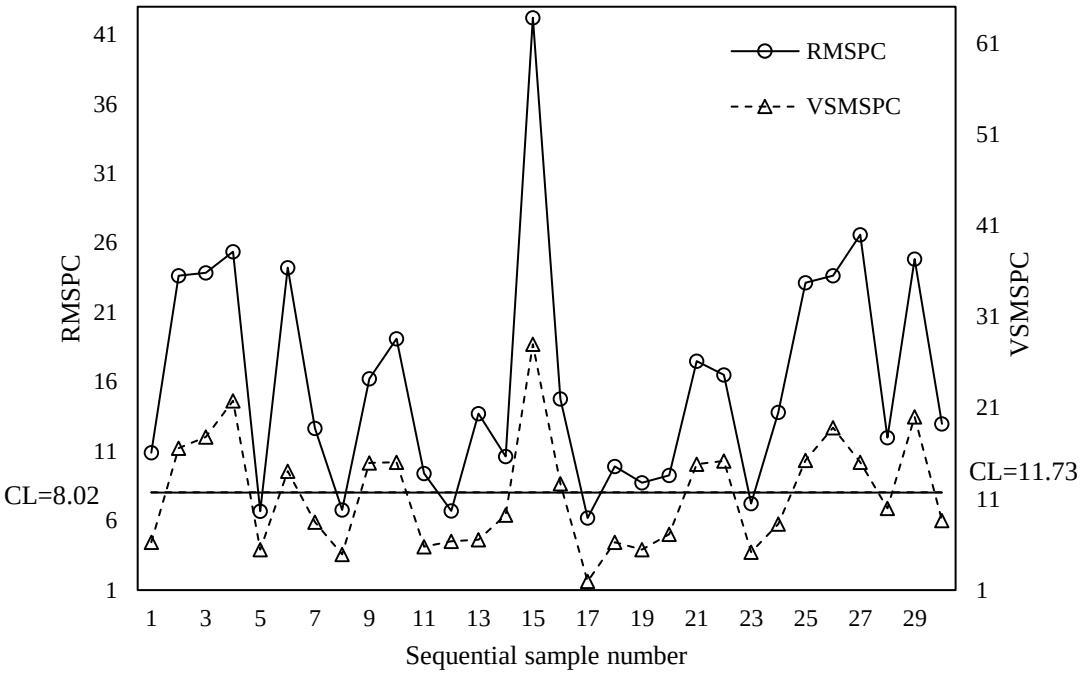


Figure 5.8. RMSPC and VSMSPC charts for the multi-attribute part

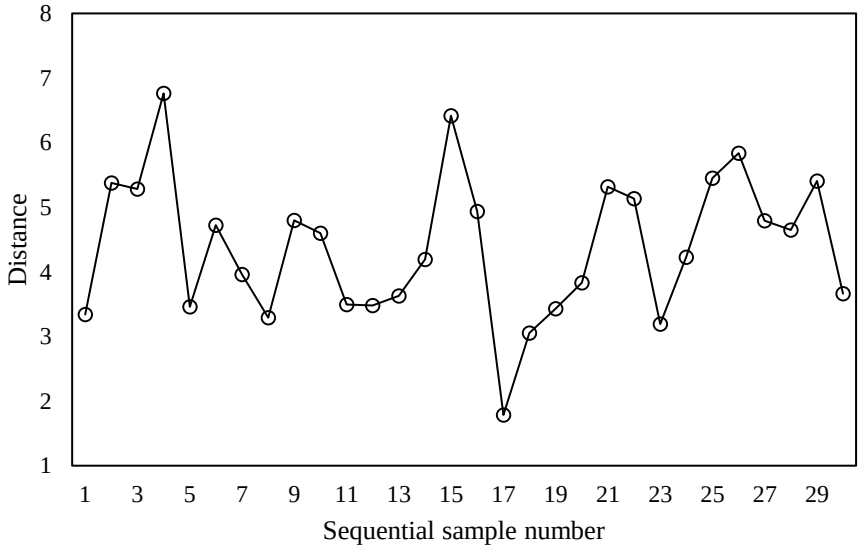


Figure 5.9. Mahalanobis distances of the sample measurements to the in-control mean

There are several samples that are detected by RMSPC but not detected by VSMSPC. Figure 5.10 shows the absolute values of the measurements of the 1st and the 17th samples as examples to explain the difference between shrinkage and variable selection. The 1st sample represents the case where RMSPC alarms while VSMSPC does not; and the 17th sample shows the case where both charts fail to detect. In (a), RMSPC shrinks most of unchanged variables towards zero but does not shrink much the changed variables resulting in a detection of the 1st sample; but in (b), it keeps relatively large portion of the information of unchanged variables such as x_2, x_5 and x_6 leading to the failure of detection. On the other hand, VSMSPC misidentifies the faulty variables for both cases since these samples are close to the in-control mean as shown in Figure 5.9. It identifies x_2, x_4, x_5 and x_2, x_4, x_{10}, respectively, in each case as faulty variables, while true shifts occur in x_7, x_8 and x_9, resulting in the failure to detect the shifts.

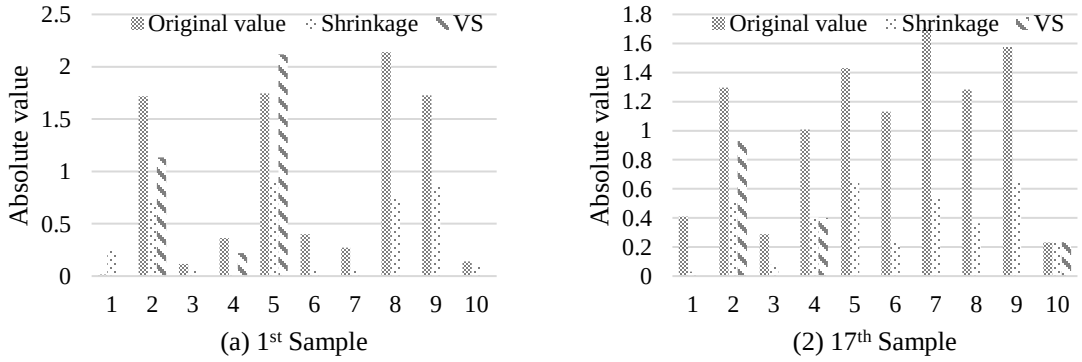


Figure 5.10. Absolute measurement values of 1st and 17th samples

5.6. Concluding Remarks

Phase II monitoring for high-dimensional processes is a challenging task. In this chapter, we focuse on the problem of detecting small mean shifts in high-dimensional processes when the quality features are strongly correlated. We propose an efficient quality control method for such processes based on L_2 norm penalized likelihood approach. This approach incorporates the directional shrinkage property of L_2 regularization based on correlation structure of the data for monitoring various out-of-control scenarios under sparsity. We provide the theoretical boundary for the performance based on a geometrical interpretation and obtained the approximate probability distributions of the monitoring statistic under H_0 and under H_1 with respect to the potential shift directions. We also investigate the improvement of the chart by introducing the relative performance index and controlling the threshold value. Lastly, we have explored the extensions of the proposed chart incorporating EWMA and diagnosis procedure. The results of simulation studies show that the proposed chart and its extended charts perform well in detecting small mean changes in high-dimensional processes and is effective in online monitoring compared to conventional multivariate SPC and recent VS-based charts.

The current study mainly focuses on shrinkage aspect rather than selection of potentially shifted variables. It would be interesting to investigate the integration of our proposed chart with another type of penalty such as the combination of L_1 and L_2 in the likelihood function. It selects suspicious variables like LASSO, and shrinks the correlated covariates together like ridge. It is expected to overcome the drawbacks of the VS-based chart in

highly correlated data structure, and is expected to perform well when the shift tends to occur in a grouped fashion due to the nature of shrinkage of L_2 regularization. In addition, the proposed chart can be extended to the problem of monitoring non-normally distributed high dimensional processes (Chen *et al.*, 2016). Further, the lack of historical dataset is also common in industry. It motivates more research to extend to Phase I analysis, in which the estimation accuracy of covariance would significantly affect the chart performance since our method mainly depends on the correlation structure.

CHAPTER 6

GENERALIZED SMOOTHING PARAMETERS OF A MULTIVARIATE EWMA CONTROL CHART

6.1. Introduction

In preceding chapters, we investigate the monitoring schemes for high-dimensional processes. In particular, for detecting small mean shift, VS-based charts perform poorly due to the misidentification of potentially faulty variables as shown in Chapters 3 and 4. Therefore, we investigated a new methodology using ridge regularization based penalized likelihood rather than adopting the diagnosis procedure. The ridge penalty shrinks the coefficient of variables proportional to the correlation in a regression point of view, it can be explained as the 'utilization of the correlation information' for SPC setting in some sense. In this chapter, we focus on seeking a model to improve the detection ability of the chart by utilizing the correlation information among variables. The traditional multivariate EWMA is extended from univariate EWMA by considering the correlation among variables. Its simplicity and generality stem from the assumption that the smoothing parameters of the variables are given constants and are equally distributed on the diagonal of the smoothing matrix. Recently, the MEWMA model with the full non-diagonal smoothing matrix (FEWMA) is studied by Hawkins *et al.* (2007). However, the model has limited use due to the assumption that the off-diagonal elements are the same which makes it sensitive to the correlation structure of observations. We propose a generalized model of MEWMA and refer to the proposed model as generalized multivariate EWMA (GMEWMA) by assigning the appropriate non-diagonal components in the smoothing

matrix based on the correlations among variables. We offer an interpretation of off-diagonal elements of the smoothing matrix and suggest an optimal design for a proposed MEWMA chart regarding the utilization of the correlation information.

This chapter is organized as follows: In Section 6.2, we briefly review the traditional MEWMA and FEWMA control charts. In Section 6.3, we interpret the meaning of non-diagonal elements in a smoothing matrix and introduce the GMEWMA chart. In Section 6.4, we compare the performance of the GMEWMA chart with MEWMA, FEWMA, and other recent variants of an MEWMA control chart. A real life case study on automatic monitoring of dimensions of bolts is illustrated in Section 6.5. Conclusions and future research are discussed in Section 6.6.

6.2. Review of the MEWMA and FEWMA Control Charts

The MEWMA statistic $\mathbf{w}_t$ is defined for an observation vector $\mathbf{x}_t$ as

$$\mathbf{w}_t = \mathbf{R}\mathbf{x}_t + (\mathbf{I} - \mathbf{R})\mathbf{w}_{t-1} \tag{6.1}$$

for $t = 1, 2, ...,$ where $\mathbf{R}$ is a smoothing matrix, and $\mathbf{I}$ is an identity matrix. Note that the smoothing matrix $\mathbf{R}$ is different from the Cholesky decomposition matrix $\mathbf{R}$ in preceding chapters. $\mathbf{w}_t$ is recursively obtained as

$$\mathbf{w}_t = \sum_{i=0}^{t-1} (\mathbf{I} - \mathbf{R})^i \mathbf{R}\mathbf{x}_{t-i} + (\mathbf{I} - \mathbf{R})^t \mathbf{w}_0.$$

Suppose that $\mathbf{w}_0 = \mathbf{x}_0 = \boldsymbol{\mu}_0$, $\mathbf{x}_i$ are i.i.d. random vectors with p variables, and $\mathbf{x}$ follows a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Since $\mathbf{w}_t$ is also a multivariate normally distributed, its mean vector and covariance matrix are given by

$$E(\mathbf{w}_t) = \boldsymbol{\mu}_0$$

and

$$Cov(\mathbf{w}_t) = \boldsymbol{\Sigma}_{\mathbf{w}_t} = \mathbf{R}\boldsymbol{\Sigma}\mathbf{R}' + (\mathbf{I} - \mathbf{R})\boldsymbol{\Sigma}_{\mathbf{w}_{t-1}}(\mathbf{I} - \mathbf{R})', \tag{6.2}$$

respectively. Assuming that a multivariate process has been operating for a sufficient length of time, it can be analyzed with an asymptotic covariance matrix:

$$\boldsymbol{\Sigma}_{\mathbf{w}} = \mathbf{R}\boldsymbol{\Sigma}\mathbf{R}' + (\mathbf{I} - \mathbf{R})\boldsymbol{\Sigma}_{\mathbf{w}}(\mathbf{I} - \mathbf{R})'.$$

The monitoring statistic of an MEWMA chart is

$$Q_t = (\mathbf{w}_t - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_{\mathbf{w}_t}^{-1} (\mathbf{w}_t - \boldsymbol{\mu}_0).$$

In an MEWMA chart, the smoothing parameters have equal diagonal elements r, i.e., $\mathbf{R} = diag(r)_{p\times p}$. Then, Eq. (6.1) can be written as

$$\mathbf{w}_t = r\mathbf{x}_t + (1-r)\mathbf{w}_{t-1},$$

and the covariance matrix of $\mathbf{w}_t$ is

$$\mathbf{\Sigma}_{\mathbf{w}_t} = \frac{r}{2-r}[1-(1-r)^{2t}]\mathbf{\Sigma}.$$

An asymptotic covariance matrix of $\mathbf{w}_t$ is expressed as a simple form of the covariance of the observation $\mathbf{x}$, $\mathbf{\Sigma}_{\mathbf{w}} = [r/(2-r)]\mathbf{\Sigma}$. It is straightforward that an MEWMA chart is equivalent to the Hotelling T^2 chart when $\mathbf{R} = \mathbf{I}$. If the target mean $\boldsymbol{\mu}_0 = \mathbf{0}$ and the mean shift is $\boldsymbol{\mu}$, the noncentrality parameter is

$$\delta^2 = \boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}.$$

Since the inverse of $\mathbf{\Sigma}_{\mathbf{w}}$ is a constant multiplication of $\mathbf{\Sigma}^{-1}$, the performance of the MEWMA chart is explained only through the noncentrality parameter.

Hawkins *et al.* (2007) attempt to use non-zero off-diagonal elements into the smoothing matrix to develop the FEWMA control chart. Using notations r_{on} and r_{off} as diagonal and non-diagonal elements, respectively, they suggest that all off-diagonal elements are equal such that $r_{off} = c \cdot r_{on}$ with $|c| < 1$. The FEWMA chart can be characterized by a smoothing parameter r, which is defined by

$$r = \sum_{i=1}^{p} r_{ij} = r_{on} + (p-1) r_{off}, \; j = 1, \ldots, p.$$

This implies that the $p-1$ off-diagonal elements in a row (or column) are equally weighted, and sum of each row (column) is r. Even though Hawkins *et al.* (2007) provide the idea of using non-diagonal smoothing parameters, their approach may have limited performance due to the assumption of equal off-diagonal elements.

6.3. Generalized Smoothing in Multivariate EWMA Control Chart

In order to develop a generalized non-diagonal smoothing scheme, we need to interpret the effect of smoothing parameters on the model performance. In this section, we interpret the meaning of a non-diagonal smoothing scheme and present the GMEWMA control chart along with the guidelines for the determination of the smoothing parameters.

6.3.1. Non-Diagonal Smoothing Parameters

We interpret the MEWMA model using a forecast model point of view. The MEWMA statistic can be expanded as follows:

$$\begin{aligned} w_{i,t} &= w_{i,t-1} + r_{i,1}(x_{1,t} - w_{1,t-1}) + \cdots + r_{ii}(x_{i,t} - w_{i,t-1}) + \cdots + r_{i,p}(x_{p,t} - w_{p,t-1}) \\ &= w_{i,t-1} + \sum_{j=1}^{p} r_{ij}(x_{j,t} - w_{j,t-1}) = w_{i,t-1} + \sum_{j=1}^{p} r_{ij} e_{j,t} \end{aligned} \tag{6.3}$$

$w_{i,t}$ can be interpreted as a prediction value of an i^{th} variable's process mean at time $t+1$. By considering $x_{j,t} - w_{j,t-1}$ as the forecast error $e_{j,t}$, we can interpret that the forecast process mean at time $t+1$ consists of the current estimated mean at time t and sum of the weighted forecast errors from all variables. Montgomery (2007) provides an interpretation of forecast viewpoint in a univariate case, and this is mostly analogous in a multivariate EWMA chart. In fact, since $\mathbf{e}_t$ includes all the other variables in multivariate EWMA, the weights r_{ij} should be carefully assigned. This implies that $w_{i,t}$ contains not only its own information of the i^{th} variable, but also the other variables' information. In other words, the meaning of 'using non-zero off-diagonal elements' is 'using the other variables' information' as a contribution to EWMA transformed data in an SPC viewpoint. For the MEWMA chart, Eq. (6.3) can be written as follows:

$$w_{i,t} = w_{i,t-1} + r_{ii}(x_{i,t} - w_{i,t-1}) = w_{i,t-1} + r_{ii} e_{i,t}.$$

Therefore, in case of the MEWMA chart, the prediction of the i^{th} variable, $w_{i,t}$ contains the limited information of only the i^{th} variable and does not consider the information from the other variables. Consequently, assigning proper weights r_{ij} considering the correlation structure of the variables can result in an improved performance of the chart.

From the above interpretation, determining r_{ij} depends on how to incorporate the other variables' information. We conjecture that it might be sensitive to the correlation structure of the original process data because the i^{th} variable is affected by all the other variables according to the weight r_{ij}. The FEWMA chart, for example, shows significantly large variation of ARL performance in accordance with the correlation structure which restricts all off-diagonal elements to be equal. It would probably be obvious that ARL performance of the FEWMA chart is good when the correlation coefficients are all equal. This is due to the fact that when $w_{i,t}$ are referring to the same amount of information from the other variables regardless of their correlation structure. In real life, however, the correlation among variables can hardly be the same. Therefore, our main interest is to consider the correlation effect in the design of the control chart.

6.3.2. Determination of the Smoothing Parameters

Non-diagonal smoothing parameters could have either a positive or a negative impact on the chart performance depending on the value of r_{ij}. We assign the smoothing parameters as a function of correlation coefficients as follows:

$$r_{ji} = (r - r_{ii}) \times \frac{|\rho_{ij}|}{\sum_{i \neq j}^{p} |\rho_{ij}|}, \quad j \neq i,$$

$$r_{ii} = \omega \cdot r, \qquad i = 1, \ldots p$$

where ω is a ratio factor such that $0 < \omega \leq 1$ and ρ_{ij} is the $(i, j)^{\text{th}}$ element in correlation matrix of **x**. It is reasonable that we assign more weight to the highly correlated variable. When the variables are independent we set $\omega = 1$ which results in a chart identical to MEWMA. Likewise, when the variables are equally correlated the resulting GMEWMA can be formulated as an FEWMA with a particular ratio factor.

6.3.3. Guidelines for Setting up an GMEWMA Chart

The effect of smoothing parameters can be interpreted as the accumulation of past data information as previously mentioned. Specifically, using a small smoothing factor means that $\mathbf{w}_t$ uses much more accumulated past data information than current observation information. Moreover, smaller w leads to an assignment of more weight to the off-diagonal elements of **R**, which results in accumulating more of the other variables' marginal information (recall that smaller ω assigns more weight to the forecast error $e_{i,t}$, i.e., $x_{i,t} - w_{i,t-1}$). Thus, in determining ω it is more appropriate to choose a small value so that the chart refers to as much information of all the other variables as possible. Then, the finding the minimum ω value, say ω^* can be formulated as the following constrained optimization problem:

$$
\begin{aligned}
&\min \ \omega \\
&s.t \quad r_{ji} = (r - r_{ii}) \times \frac{|\rho_{ij}|}{\sum_{i \neq j}^{p} |\rho_{ij}|} \quad , j \neq i \\
&\qquad r_{ji} = \omega \cdot r \qquad\qquad , j = i \\
&\qquad r_{ji} < r_{ii} \qquad\qquad , j \neq i \\
&\qquad 0 < \omega \leq 1
\end{aligned}
$$

Then, ω^* is found as

$$
\omega^* = \max_{1 \leq j \leq p} \left\{ \frac{\max_{1 \leq i \leq p, i \neq j} |\rho_{ij}|}{\sum_{i \neq j} |\rho_{ij}| + \max_{1 \leq i \leq p, i \neq j} |\rho_{ij}|} \right\} \tag{6.4}
$$

where ρ_{ij} is a correlation coefficient between i and j (see Appendix C for the detailed derivation). When the variables are equally correlated ($\rho_{ij} = \rho$, $0 < \rho < 1$), the infimum of the ω is $1/p$; i.e. the diagonal elements in the smoothing matrix are r/p. If the diagonal elements are less than r/p, non-diagonal elements can be greater than diagonal, and eventually violate the generality of the chart. Then, ω^* can be expressed as a sum of $1/p$ and an additive value led by unbalanced correlation structure. Consequently, ω^* accounts for a limiting condition for the generality of the model and the additional weights to the diagonal elements of $\mathbf{R}$ according to the correlation structure.

6.4. Performance Comparison

6.4.1. Description of the Experiments

In this section, we illustrate the performance of the GMEWMA chart through numerical examples. In the calculation of the control statistic Q_t to determine the out-of-control signal, an MEWMA chart uses an asymptotic covariance matrix which corresponds to the steady-state process. In some cases, however, the processes go through a start-up period where many out-of-control signals are observed. Once a process is stabilized, the out-of-control signal pattern becomes significantly different from that of the start-up period. These two different states can be expressed in EWMA charts through the covariance matrix. Since determination of out-of-control signal for obtaining ARL_1 is measured by a distance from the target mean, the number of in-control observations before the occurrence of an out-of-control observation depends on how Q_t is calculated. Briefly, the experiments can be summarized in two ways:

- Process starts up in out-of-control state and the exact covariance matrix is applied; and
- Out-of-control signal occurs after process stabilization and the asymptotic covariance matrix is applied.

In this study, we mainly focus on the early stage of the processes when it is out-of-control and apply the exact covariance matrix of EWMA statistic $\mathbf{\Sigma}_{\mathbf{w}_t}$ as given in Eq. (6.2). All

the other charts also adopt the exact covariance matrix in the calculation of the control statistic for performance comparison.

In determination of the control limits, the ideal control limits for the EWMA chart would be varied instantly at a sampling point time t, say $h(t)$. Considering the complexity of obtaining the exact control limits for every sampling point, we assume a fixed control limit h for a desired in-control ARL with the assumption that process starts up in out-of-control state.

In the experiments, type I error is chosen as 0.005, i.e. in-control ARL level is 200 and the smoothing parameter r as 0.1 for all comparable charts. As mentioned previously, the detection of a small process mean shift is of interest. A small smoothing parameter is used for a small shift for the EWMA charts as long as the charts keep the same track of accumulation of the past information. The ratio factor, w is rounded up to two decimal digits. For convenience, we use the covariance matrix $\mathbf{\Sigma_x} = \mathbf{\Gamma}$ which is a correlation matrix and the in-control process mean $\boldsymbol{\mu}_0 = \mathbf{0}$ without loss of generality. Moreover, we consider an individual observation as a subgroup size for all charts for comparison. Since we assume that we have knowledge of a covariance matrix, individual observation would not be a significant issue of the problem. A mean shift size is determined through a noncentrality parameter, δ.

6.4.2. ARL Performance Comparison

We conduct a simulation study for the ARL performance comparison of three different charts: standard MEWMA, dMEWMA and FEWMA. The fundamental concept of the dMEWMA control chart is to smooth twice as shown by the following two equations (Alkahtani and Schaffer, 2012).

$$\mathbf{y}_t = \mathbf{R}\mathbf{x}_t + (\mathbf{I} - \mathbf{R})\mathbf{y}_{t-1}$$
$$\mathbf{w}_t = \mathbf{R}\mathbf{y}_t + (\mathbf{I} - \mathbf{R})\mathbf{w}_{t-1}$$

In their model, the smoothing matrix has the same diagonal elements. Thus we classify MEWMA and dMEWMA into the diagonal scheme and FEWMA and GMEWMA as the non-diagonal scheme.

The simulation is conducted with 200,000 replications in order to achieve a standard error less than 10^{-3} for ARL_0. In the non-diagonal scheme charts, FEWMA and GMEWMA, the diagonal elements in the smoothing matrix are equal for comparison. In Table 6.1 (a), for example, the ratio factor $\omega = 0.36$ and $\gamma = 0.8$ from Eq. (6.4), and $r_{ii} = \omega \times r = 0.036$ for all *i*. When $c = 0.89$ in the FEWMA, the diagonal element in the smoothing matrix is 0.036 from the relationship between r_{on} and r_{off} ($r = r_{on} + (p-1)r_{off} = r_{on} + c(p-1)r_{on}$). Hence, the diagonal elements of the smoothing matrix in both the GMEWMA chart and the FEWMA chart are equal with a value of $0.036,$ and their smoothing matrices are shown below:

$$\mathbf{R}_{GMEWMA} = \begin{bmatrix} 0.0360 & 0.0320 & 0.0284 \\ 0.0356 & 0.0360 & 0.0356 \\ 0.0284 & 0.0320 & 0.0360 \end{bmatrix} \text{ and } \mathbf{R}_{FEWMA} = \begin{bmatrix} 0.0360 & 0.0320 & 0.0320 \\ 0.0320 & 0.0360 & 0.0320 \\ 0.0320 & 0.0320 & 0.0360 \end{bmatrix}.$$

Since the main motive of the GMEWMA chart is to use the correlation information, the experiments are conducted by the level of the correlation imbalance. For more systematic comparison in accordance with the correlation difference, we organize the experiments with one popular correlation structure which can simply be expressed by a descending order form as follows:

$$\begin{bmatrix} 1 & \gamma\rho & \gamma^2\rho & \gamma^3\rho & \cdots & \gamma^{p-1}\rho \\ \gamma\rho & 1 & \gamma\rho & \gamma^2\rho & \ddots & \gamma^{p-2}\rho \\ \gamma^2\rho & \gamma\rho & 1 & \gamma\rho & \ddots & \gamma^{p-2}\rho \\ \gamma^3\rho & \gamma^2\rho & \gamma\rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \gamma\rho \\ \gamma^{p-1}\rho & \gamma^{p-2}\rho & \gamma^{p-3}\rho & \cdots & \gamma\rho & 1 \end{bmatrix}$$

where γ is an adjusting parameter to increase the correlation difference for the experiments. The correlation difference decreases as γ decreases. When $\gamma = 1$, the correlation structure is even with ρ, i.e. $\rho_{ij} = \rho$ for all $i \neq j$. In Table 1, we illustrate the ARL_1 performances for different values of γ. We set $\rho = 0.9$, and conduct experiments with $\gamma = 1$, 0.8, 0.6, 0.4 and 0.2. The ω and c represent the parameters for GMEWMA and FEWMA, respectively. Tables 6.1 (a), (b) and (c) show the ARL_1 performance for

different number of variables; $p = 3,\ 5$ and 10, respectively. The control limits are adjusted with $300,000$ replications for the desired ARL_0.

In Table 6.1, the ARL_1 performance of the GMEWMA control chart is mostly superior to the other control charts. As mentioned, an FEWMA chart is a special case of an GMEWMA by adjusting ω, and therefore the ARL_1 performance of both GMEWMA and the FEWMA chart when $\gamma = 1$ (when all the correlation coefficients are the same) are identical. As shown in Table 6.1 an GMEWMA chart has a better ARL_1 performance compared to the FEWMA chart as the correlation matrix becomes more unbalanced (i.e. smaller γ) regardless of p. In fact, an GMEWMA chart outperforms FEWMA for all cases. The GMEWMA chart assigns the smoothing parameters adaptively considering the correlation structure of variables while the FEWMA chart assigns an equal weight for non-diagonal elements in the smoothing matrix without considering the correlation structure of variables.

The ARL_1 of the GMEWMA chart increases slightly when p becomes large in the extreme case when the correlation is greatly uneven. When p is large, such as $p = 10$, most of the correlation coefficients become close to zero as γ decreases in the correlation structure, which means that most of the variables appear to be independent. Since the GMEWMA chart refers to the correlation information, the chart performance would eventually approach that of MEWMA's when all correlations become zero.

Table 6.1. ARL_1 performance for different correlation structures

(a) $p = 3$

γ	N/A		1		0.8	
	MEWMA	dMEWMA	GMEWMA	FEWMA	GMEWMA	FEWMA
	N/A	N/A	$\omega^* = 0.34$	$c = 0.97$	$\omega^* = 0.36$	$c = 0.89$
Shift	$h = 10.96$	$h = 8.14$	$h = 8.47$	$h = 8.47$	$h = 8.67$	$h = 8.69$
0.00	200.670	200.078	200.349	200.349	199.478	200.219
0.10	165.411	153.162	144.376	144.376	145.973	148.341
0.25	83.035	67.234	58.710	58.710	61.226	62.425
0.50	28.782	23.433	19.871	19.871	20.605	20.709
0.75	14.246	12.204	10.204	10.204	10.541	10.567
1.00	8.716	7.640	6.453	6.453	6.637	6.667
2.00	2.860	2.452	2.288	2.288	2.337	2.342
γ	0.6		0.4		0.2	
	GMEWMA	FEWMA	GMEWMA	FEWMA	GMEWMA	FEWMA
	$\omega^* = 0.39$	$c = 0.78$	$\omega^* = 0.42$	$c = 0.69$	$\omega^* = 0.46$	$c = 0.59$
Shift	$h = 8.89$	$h = 8.97$	$h = 9.06$	$h = 9.16$	$h = 9.25$	$h = 9.41$
0.00	199.412	200.206	200.273	199.068	200.285	199.787
0.10	148.945	152.287	149.706	153.201	150.907	156.267
0.25	63.159	65.971	64.777	67.789	66.253	70.761
0.50	21.311	21.905	22.015	22.668	22.729	23.538
0.75	10.923	11.057	11.148	11.375	11.473	11.776
1.00	6.829	6.927	6.995	7.091	7.166	7.314
2.00	2.385	2.403	2.424	2.448	2.471	2.505

(b) $p = 5$

γ	N/A		1		0.8	
	MEWMA	dMEWMA	GMEWMA	FEWMA	GMEWMA	FEWMA
	N/A	N/A	$\omega^* = 0.21$	$c = 0.94$	$\omega^* = 0.25$	$c = 0.75$
Shift	$h = 10.96$	$h = 11.53$	$h = 10.81$	$h = 10.81$	$h = 11.34$	$h = 11.48$
0.00	200.303	200.047	199.032	199.032	199.550	119.955
0.10	172.625	163.281	145.855	145.855	150.286	156.727
0.25	96.266	78.531	59.677	59.677	65.651	70.383
0.50	34.496	27.626	19.823	19.823	21.887	22.877
0.75	16.737	14.167	10.316	10.316	11.143	11.443
1.00	10.158	8.893	6.521	6.521	6.999	7.161
2.00	3.262	2.831	2.359	2.359	2.478	2.509
γ	0.6		0.4		0.2	
	GMEWMA	FEWMA	GMEWMA	FEWMA	GMEWMA	FEWMA
	$\omega^* = 0.31$	$c = 0.56$	$\omega^* = 0.38$	$c = 0.41$	$\omega^* = 0.44$	$c = 0.32$
Shift	$h = 11.90$	$h = 12.12$	$h = 12.42$	$h = 12.66$	$h = 12.72$	$h = 13.04$
0.00	199.401	200.492	200.621	199.512	200.412	200.711
0.10	154.549	162.493	159.929	164.574	162.157	167.549
0.25	70.391	78.001	76.760	83.199	78.892	86.051
0.50	24.011	25.651	26.084	27.892	27.300	29.231
0.75	12.057	12.609	12.960	13.529	13.476	14.159
1.00	7.558	7.769	8.029	8.302	8.312	8.647
2.00	2.594	2.654	2.722	2.777	2.792	2.867

(c) $p = 10$

γ	N/A		1		0.8	
	MEWMA	dMEWMA	GMEWMA	FEWMA	GMEWMA	FEWMA
	N/A	N/A	$\omega^* = 0.11$	$c = 0.90$	$\omega^* = 0.19$	$c = 0.47$
Shift	$h = 22.91$	$h = 18.98$	$h = 16.52$	$h = 16.52$	$h = 18.24$	$h = 18.52$
0.00	199.378	200.194	199.191	199.191	200.377	200.367
0.10	179.668	171.723	150.822	150.822	164.246	170.338
0.25	114.090	94.372	64.114	64.114	68.167	92.124
0.50	44.413	34.476	20.685	20.685	27.414	30.564
0.75	21.344	17.501	10.509	10.509	13.519	14.311
1.00	12.673	10.961	6.728	6.728	8.298	8.683
2.00	3.967	3.481	2.499	2.499	2.875	2.947
γ	0.6		0.4		0.2	
	GMEWMA	FEWMA	GMEWMA	FEWMA	GMEWMA	FEWMA
	$\omega^* = 0.29$	$c = 0.27$	$\omega^* = 0.38$	$c = 0.18$	$\omega^* = 0.45$	$c = 0.14$
Shift	$h = 19.42$	$h = 19.81$	$h = 20.20$	$h = 20.57$	$h = 20.60$	$h = 21.02$
0.00	199.289	200.205	200.856	200.842	200.651	200.291
0.10	169.910	176.353	175.581	177.403	176.459	178.191
0.25	93.137	102.621	100.966	107.261	104.207	108.613
0.50	32.536	35.983	35.901	38.671	34.581	40.136
0.75	15.668	16.748	17.058	18.074	17.764	18.855
1.00	9.483	9.966	10.238	10.687	10.614	11.136
2.00	3.148	3.235	3.325	3.415	3.414	3.524

Moreover, as discussed in Section 6.3.3, it is evident that using more information about the other variables which results in a small ω, improves the detection power of the chart.

Consequently, the GMEWMA chart outperforms other charts for the cases that are close to balanced correlation structure because we can use a smaller w than those of uneven correlation structures. On the contrary, in the case of uneven correlation, such as when $\gamma = 0.2,$ the ratio factor ω becomes large due to the general condition that non-diagonal elements in the smoothing matrix ought to be smaller than the diagonal elements resulting in a slightly longer detection time.

Tables 6.2 (a), (b) and (c) show the effect of ρ on the performance of GMEMWA chart. The parameter $\gamma = 1$ so that the correlation structure is even with all correlation coefficients $\rho_{ij} = \rho$. Tables 6.2 (a) through (c) are the results for three cases of $\rho = 0.1,$ 0.5, and 0.9. The correlation structures when $p = 3$ for Table 6.2 (a) are given below:

$$\boldsymbol{\Sigma}_{\mathbf{x}_{(a),0.1}} = \begin{bmatrix} 1 & 0.1 & 0.1 \\ 0.1 & 1 & 0.1 \\ 0.1 & 0.1 & 1 \end{bmatrix} \quad \boldsymbol{\Sigma}_{\mathbf{x}_{(a),0.5}} = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix} \quad \boldsymbol{\Sigma}_{\mathbf{x}_{(a),0.9}} = \begin{bmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{bmatrix}$$

Table 6.2 demonstrates that the GMEWMA chart refers to 'relative correlation information' rather than 'absolute amount of correlation'. Thus, the size of ρ does not affect ARL_1, and the chart performance depends on the relative difference of correlation.

Table 6.2. Effect of ρ and p on the performance of a GMEWMA chart

(a) $p = 3,\ \gamma = 1$ and $\rho = 0.1, 0.5, 0.9$

Shift size	$\rho = 0.1$	$\rho = 0.5$	$\rho = 0.9$
0.00	199.534	199.534	200.591
0.10	146.095	146.094	145.993
0.25	60.184	60.181	60.239
0.50	20.239	20.259	20.279
0.75	10.354	10.384	10.358
1.00	6.532	6.572	6.537
2.00	2.314	2.316	2.314

(b) $p = 5,\ \gamma = 1$ and $\rho = 0.1, 0.5, 0.9$

Shift size	$\rho = 0.1$	$\rho = 0.5$	$\rho = 0.9$
0.00	199.831	199.135	199.032
0.10	145.721	146.092	145.855
0.25	60.012	59.806	59.677
0.50	19.901	19.971	19.823
0.75	10.268	10.235	10.316
1.00	6.531	6.524	6.521
2.00	2.359	2.361	2.359

(c) $p = 10,\ \gamma = 1$ and $\rho = 0.1, 0.5, 0.9$

Shift size	$\rho = 0.1$	$\rho = 0.5$	$\rho = 0.9$
0.00	199.992	199.792	199.191
0.10	150.847	150.837	150.822
0.25	63.929	64.201	64.114
0.50	20.581	20.622	20.685
0.75	10.553	10.527	10.509
1.00	6.736	6.741	6.728
2.00	2.499	2.499	2.499

Table 6.3 and Figure 6.1 show the effect of ω on the performance of the chart and confirm that minimum ω results in the best performance for all shift sizes, as suggested in Section 6.3. The experiment is conducted by the setting $p = 3$ and randomly choosing correlation structure that has $\omega^* = 0.4444$. Due to the constraint $r_{ji} < r_{ii}$, the ratio ω has its lower bound as ω^*. When ω is less than 0.4444, it would possibly violate generality of the charts, which means that one variable may have more information about the other variables than its own. Note that as ω approaches 1, GMEWMA approaches MEWMA, accordingly.

Table 6.3. Performance of the GMEWMA chart for different values of ω

ω	0.45	0.55	0.65	0.75	0.85	0.95	1.00
$\delta \setminus h$	h = 9.23	9.80	10.17	10.45	10.68	10.87	10.96
0.00	199.834	199.666	199.296	200.043	199.836	199.241	200.670
0.10	142.847	153.091	157.679	160.981	162.219	165.376	165.411
0.25	61.727	68.489	73.206	77.149	79.835	82.058	83.035
0.50	22.059	24.125	25.567	26.699	27.567	28.339	28.782
0.75	11.383	12.334	12.918	13.401	13.744	14.059	14.246
1.00	7.161	7.703	8.001	8.251	8.457	8.636	8.716
2.00	2.458	2.595	2.688	2.745	2.798	2.845	2.860

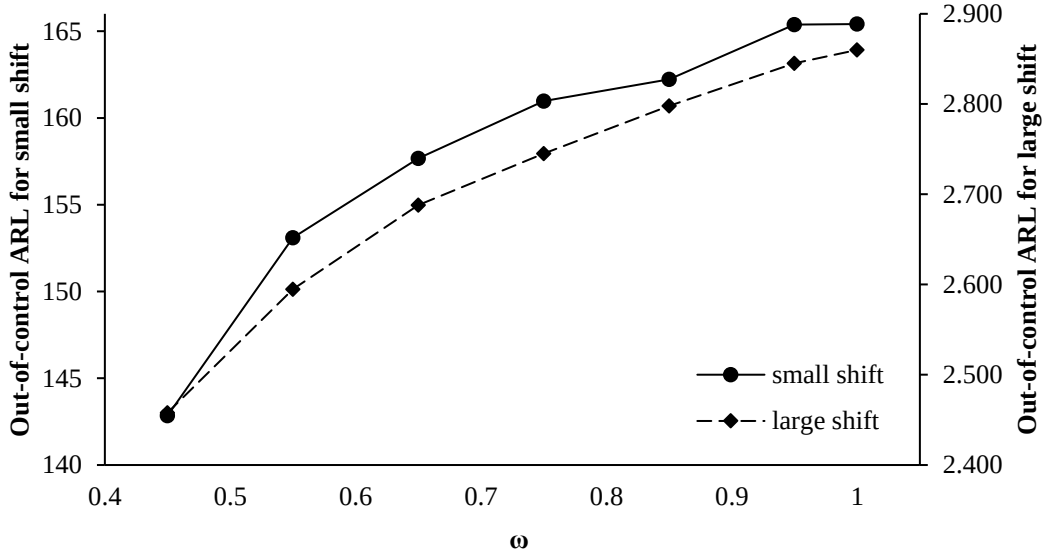


Figure 6.1. Effect of ω on ARL_1

6.4.3. Steady State Behavior

We present the steady state behavior of the charts in Table 6.4. The conditions of the experiments are the same as that in Table 6.3. Calculating the EWMA control statistic Q_t with an asymptotic covariance matrix, we observe the chart behavior when the processes are monitored for a long time.

Table 6.4. Steady state ARL_1 comparison

Shift size	Diagonal Scheme		Non-diagonal Scheme	
	MEWMA	dMEWMA	FEWMA ($c = 0.61$)	GMEWMA ($\omega^* = 0.45$)
	h = 10.78	h = 7.71	h = 8.67	h = 8.24
0.00	200.159	200.667	199.801	200.213
0.10	165.482	157.994	161.481	155.579
0.25	86.259	76.178	80.912	76.154
0.50	31.753	30.401	30.591	29.834
0.75	16.908	18.803	16.448	16.392
1.00	11.246	14.237	10.927	10.918
2.00	4.841	8.422	4.657	4.656

In Table 6.4, all control charts show slightly lower ARLs for all shifts than the start-up out-of-control state. Specifically, there is a greater difference in the large shift than in the small shift. For example, there is less than 10% of the ARL increase in small shift and about 70% or more increase in large shift for all charts. Since all EWMA charts accumulate past

information, it takes a slightly longer time to detect an out-of-control signal in steady state specially for the large shift as Lowry *et al.* (1992) pointed out. This is highly associated with the setting of the control limit as mentioned in Section 6.4.1.

Moreover, the proposed chart and MEWMA chart may show similar behavior for different values of r; smaller r, result in better performance in small shifts and vice versa, because they are based on the weighted average statistic of past information. Table 6.5 shows the ARL_1 performances of both MEWMA and GMEWMA for various r values when $p = 3$.

The ALR_1 of both charts increases in small shifts and decreases in large shifts as r increases. The preliminary work by Prabhu and Runger (1997) presents the out-of-control ARL performances in terms of r and p**.** There is a clear pattern of the performance of the GMEWMA chart with respect to r on the same horizon as EWMA-based charts, which means that the more accumulation of the past information, the better detection power in small shifts. From the observation of the results in Table 6.5, the GMEWMA chart in steady state has the same property as MEWMA in terms of r, which means that it is characterized by a smoothing parameter.

Table 6.5. ARL_1 performances for various r values

r	0.1		0.2		0.4	
Charts	MEWMA	GMEWMA	MEWMA	GMEWMA	MEWMA	GMEWMA
Shift size	$h=10.78$	$h=8.24$	$h=11.88$	$h=9.64$	$h=12.55$	$h=10.75$
0.00	200.159	200.213	200.575	200.666	199.569	199.623
0.25	86.259	75.772	108.662	86.708	136.370	95.762
0.50	31.753	29.834	41.284	35.535	63.453	42.9043
1.00	11.246	10.918	11.488	11.368	15.752	14.1542
2.00	4.841	4.656	4.129	4.151	3.901	4.0691
3.00	3.193	3.068	2.605	2.604	2.214	2.2983
r	0.6		0.8		0.9	
Charts	MEWMA	GMEWMA	MEWMA	GMEWMA	MEWMA	GMEWMA
Shift size	$h=12.76$	$h=11.25$	$h=12.83$	$h=11.53$	$h=12.84$	$h=11.63$
0.00	199.494	199.9517	201.005	199.5962	200.074	199.6514
0.25	155.871	100.6189	168.635	103.7607	173.196	105.6528
0.50	87.123	46.703	109.558	48.8603	119.080	49.6063
1.00	23.660	16.932	35.868	18.8011	43.714	19.5832
2.00	4.388	4.539	5.764	5.4163	7.007	5.9268
3.00	2.081	2.220	2.134	2.3266	2.288	2.4649

6.4.4. Analysis of Directional Variance Property of the GMEWMA Chart

MEWMA and dMEWMA control charts are classified as ‘diagonal scheme’ and FEWMA and GMEWMA charts are classified as ‘non-diagonal scheme.’ In a diagonal scheme, each multivariate chart is an extension of a univariate EWMA chart; MEWMA corresponds to

EWMA; and dMEWMA corresponds to dEWMA. Thus, the diagonal scheme charts are directionally invariant whereas non-diagonal scheme charts are not. Although GMEWMA shows considerably good performance, its directional variant property may not reproduce the same out-of-control ARL with respect to all directions. In practice, shifts can occur in various directions simultaneously; and positive or negative shifts with different magnitudes. From Eq. (6.3), we conjecture that the shift of some variables may take slightly longer to detect than the others. For instance, suppose that the first variable is highly correlated with all the other variables. Then, most r_{1j} $(j \neq 1)$ will be relatively larger than the other r_{ij} $(i \neq 1, j \neq i)$, i.e., it provides relatively more information of the other unchanged variables to ω_{1t}, which results in slightly longer detection time of this mean shift. Table 6.6 shows its relative difference of ARL_1 when the following covariance matrix is used.

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \begin{bmatrix} 1 & 0.2 & 0.8 \\ 0.2 & 1 & 0.5 \\ 0.8 & 0.5 & 1 \end{bmatrix}$$

In order to verify the performance in terms of the directions, we investigate several cases with different out-of-control scenarios in terms of the directions as stated below:

(1) Scenario 1: Shift in single quality characteristic;

(2) Scenario 2: Simultaneous shifts in two or three quality characteristics;

(3) Scenario 3: Simultaneous shifts in positive and negative directions; and

(4) Scenario 4: Larger shift in highly correlated quality characteristics

Since there is a large number of combinations of out-of-control scenarios as *p* increases, we keep the number of variables as three. This might be justified with a strong conjecture that the performance of the chart would be similar although the number of variables varies. We denote the shift direction and its magnitude as a vector form. In Table 6.6, for example, $(-\theta, \theta, 0)$ means that the shift occurs along the first variable and the second variable with the same magnitude $|\theta|$, and (-) sign represent negative direction. The shift size in the first column represents a noncentrality parameter δ**.** Thus, when $\delta = 1$**,** for example, the shifted vector satisfies $\sqrt{\boldsymbol{\delta}_1 \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \boldsymbol{\delta}_1} = 1$ where $\boldsymbol{\delta}_1 = (-\theta, \theta, 0) = (-0.5692, 0.5692, 0)$**.** For the case of scenario 4, the principle directions of given correlation matrix noting $\mathbf{u}_i$ for $i = 1, 2, 3$ are considered noting $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$ below:

$$\mathbf{U} = \begin{bmatrix} 0.6159 & 0.5113 & 0.5993 \\ 0.2832 & -0.8536 & 0.4372 \\ -0.7351 & 0.0996 & 0.6706 \end{bmatrix}$$

If the process mean vector lies on the direction $\mathbf{u}_1 = (0.6159, 0.2832, -0.7351)$, it appears that the shift occurs in the opposite direction while the process variables are positively correlated, and thus suitable for scenario 4. That is, in spite of the positive correlation $\rho_{13} = 0.8$ the mean shift for x_1 occurs with a magnitude of 0.6159, and the mean shift for x_3 occurs with an amount of -0.7351. The direction $\mathbf{u}_3$ represents that the highly and

positively correlated variables have relatively large shifts, 0.5993 and 0.6706, along x_1 and x_3 direction, respectively. The control factor $\omega = 0.45$ is used for GMEWMA by following the guidelines for determining ω in Eq. (6.4).

Table 6.6. Comparison of ARL_1 for directional mean shifts

Shift size(δ)	MEWMA	dMEWMA	GMEWMA: Shift scenario (1)		
	Directionally invariant		$(\theta,0,0)$	$(0,\theta,0)$	$(0,0,\theta)$
0.10	165.411	153.162	136.201	134.181	139.807
0.25	83.035	67.234	59.392	57.617	62.463
0.50	28.782	23.433	22.238	21.801	23.309
0.75	14.246	12.204	11.698	11.534	11.927
1.00	8.716	7.640	7.290	7.261	7.366
2.00	2.860	2.452	2.475	2.472	2.475

Shift size (δ)	GMEWMA: Shift scenario (2)			
	$(\theta,\theta,0)$	$(\theta,0,\theta)$	$(0,\theta,\theta)$	(θ,θ,θ)
0.10	133.875	149.722	148.273	142.847
0.25	58.497	65.043	64.883	61.727
0.50	22.422	22.809	23.050	22.059
0.75	11.662	11.593	11.702	11.383
1.00	7.326	7.212	7.291	7.161
2.00	2.475	2.463	2.476	2.458

Shift size (δ)	GMEWMA: Shift scenario (3)				
	$(-\theta,\theta,0)$	$(0,-\theta,\theta)$	$(-\theta,0,\theta)$	$(-\theta,-\theta,\theta)$	$(-\theta,\theta,-\theta)$
0.10	141.264	135.965	138.306	135.079	140.303
0.25	59.590	59.991	60.640	59.851	60.105
0.50	21.678	22.755	22.791	22.776	22.264
0.75	11.475	11.867	11.830	11.831	11.618
1.00	7.208	7.352	7.336	7.371	7.271
2.00	2.474	2.486	2.475	2.476	2.469

Shift size (δ)	GMEWMA: Shift scenario (4)		
	$\mathbf{u}_1$	$\mathbf{u}_2$	$\mathbf{u}_3$
0.10	136.708	141.946	151.632
0.25	60.519	58.488	66.791
0.50	22.749	21.522	22.715
0.75	11.782	11.362	11.521
1.00	7.382	7.190	7.187
2.00	2.478	2.470	2.463

Table 6.6 also shows that the GMEWMA chart detects changes in the mean shift considerably faster than the MEWMA chart despite the directional variant property. We observe from the experiments that it takes slightly more observations to detect out-of-control signals when the shift occurs among highly correlated variables such as x_3 (correlation 0.8 and 0.5 with x_1 and x_2, respectively). It is also observed that the detection

time increases when the out-of-control process mean lies on directions along highly correlated variables such as $(0,0,\theta)$, $(\theta,0,\theta)$, $(0,\theta,\theta)$, (θ,θ,θ) and $\mathbf{u}_3$. Even though there is a minor performance difference between directional shifts, the proposed chart shows better variant detection power than the other charts. Thus we conclude that the detection power is robust to the directional variant property. Moreover, we conjecture that this directional variant property is offset when the process has more variables.

6.5. Case Study: Continuous Monitoring of Dimensions of Bolts

Quality control of products and components may be monitored using image processing which detects changes in features like shape, color, and any change in the process, since some of these characteristics are difficult to measure with traditional measurement methods (Radke *et al.*, 2005, Qiu, 2005). In this section, we present a case study conducted in a laboratory with an industry sponsor to develop an automatic and continuous monitoring system for bolts' dimensions using image processing techniques. The system is able to detect changes in the bolts' dimensions. This system is described as follows: When infrared sensors detect a bolt as it travels on a conveyer belt, a high resolution camera takes images of the bolt which are instantaneously processed using image processing algorithms. The dimensions (in pixels) are compared with in-control images, and when any of the dimensions exceeds its corresponding threshold value, the bolt is automatically diverted from the conveyor. We consider a bolt to be acceptable when all the following four dimensions are within the acceptable threshold pixel count. The four dimensions of the bolt are: the head diameter (x_1), the head height (x_2), the bolt diameter (x_3), and the length of

the bolt (x_4) as shown in Figure 6.2. It should be noted that the gray scaled image of the bolt is assessed by finding the smallest rectangle containing the bolt and head through the image pixel matrices.

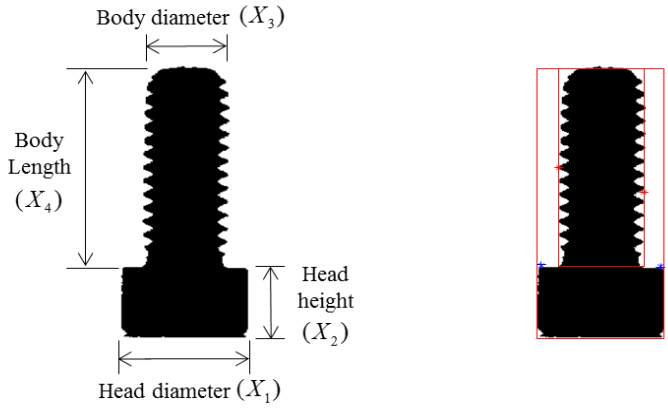


Figure 6.2. Measurements of bolt (left) and image measurement (right)

The summary statistics of bolts from the in-control process are shown in Table 6.7. The large negative correlation between x_2 and x_4 in the correlation matrix demonstrates that these variables have strong interrelationships.

Table 6.7. Multivariate summary statistics of bolt image data

	x_1	x_2	x_3	x_4
Mean	0.3673	0.2449	0.2502	0.7346
Standard deviation	0.0018	0.0063	0.0032	0.0075
		Correlation matrix		
x_1	1.0000	-0.1853	0.3231	0.2026
x_2	-0.1853	1.0000	0.1025	-0.9511
x_3	0.3231	0.1025	1.0000	-0.1516
x_4	0.2026	-0.9511	-0.1516	1.0000

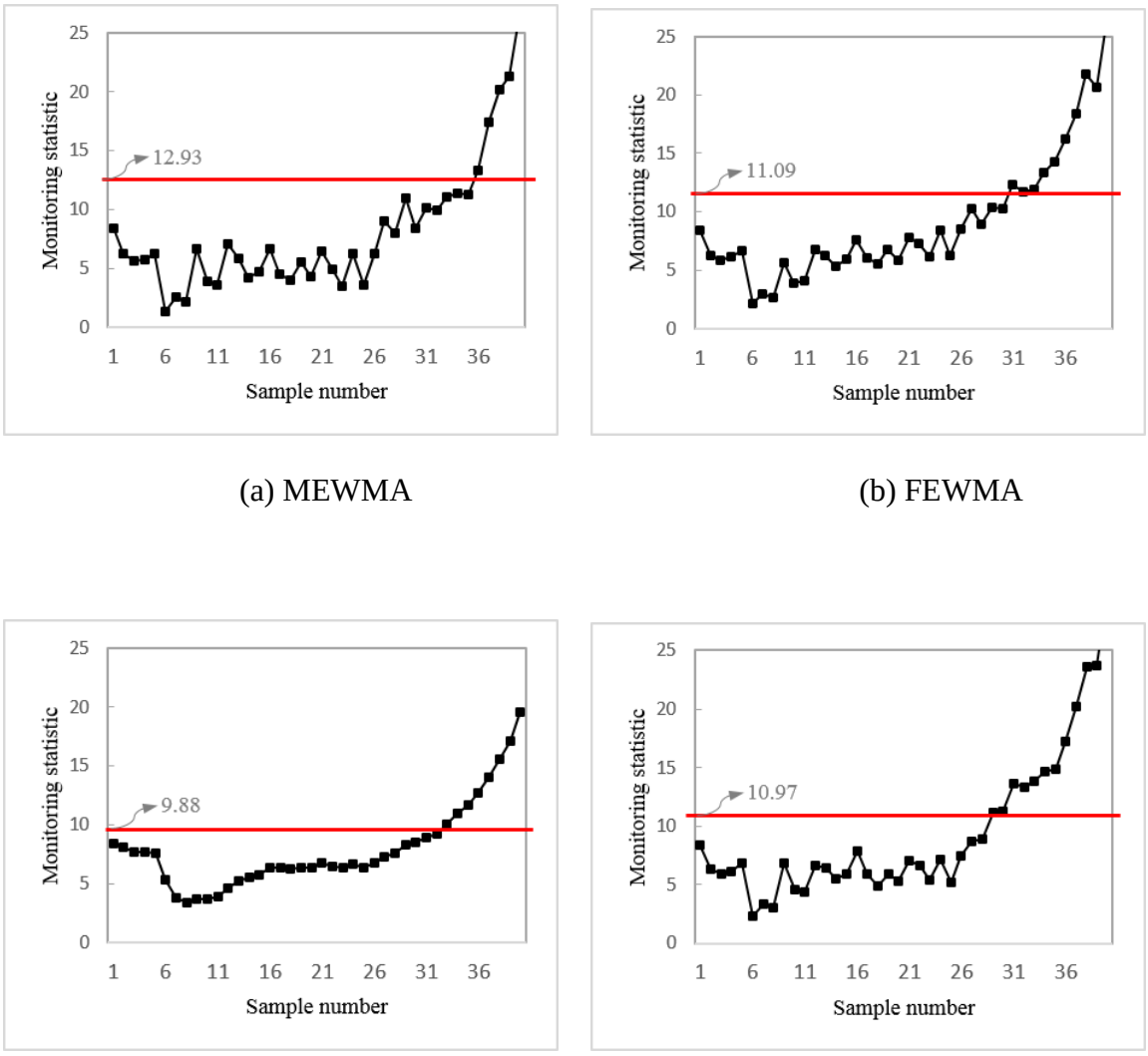
Let $\mathbf{x}_t = (x_{1t}, x_{2t}, x_{3t}, x_{4t})$ be an observation vector at time t, where x_{it} denotes an observation of i^{th} dimension at time t. There are 40 test measurements $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{40}$ in Table 6.8, the first 25 observations are from an in-control process while the other 15 observations are from an out-of-control process which has a negative change for the mean of x_4 of 0.7279 representing change in the bolt dimensions due to tool wear out during machining. Each observation is used to be monitored as a subgroup size equal to one. The MEWMA, FEWMA, dMEWMA and GMEWMA charts are used for performance comparison; the results are shown in Figure 6.3.

Table 6.8. Bolt measurements (inch)

Sample Number	x_1	x_2	x_3	x_4
1	0.370628	0.254686	0.251137	0.723283
2	0.367139	0.247098	0.246640	0.730371
3	0.366800	0.249428	0.247779	0.725521
4	0.367001	0.250363	0.249392	0.725039
5	0.369994	0.233658	0.247023	0.747591
6	0.367196	0.235647	0.255932	0.739443
7	0.369168	0.249010	0.258968	0.725309
8	0.367572	0.249508	0.248305	0.725550
9	0.370375	0.246569	0.252285	0.724183
10	0.365189	0.244316	0.246981	0.731474
11	0.367861	0.248050	0.247536	0.729589
12	0.369859	0.243923	0.246098	0.732441
13	0.368251	0.248883	0.254636	0.726934
14	0.367631	0.244285	0.251243	0.733613
15	0.367906	0.251682	0.250763	0.725142
16	0.370247	0.239627	0.252208	0.735005
17	0.364437	0.239062	0.243603	0.735645
18	0.367602	0.246370	0.246498	0.734107
19	0.368349	0.236231	0.249067	0.738114
20	0.368715	0.240900	0.253322	0.736646
21	0.369529	0.246852	0.249108	0.728854
22	0.361626	0.243270	0.245834	0.728858
23	0.369189	0.249621	0.250636	0.728972
24	0.367542	0.240844	0.245311	0.734529
25	0.365522	0.243802	0.249774	0.734210
26	0.370026	0.242646	0.250071	0.730969
27	0.366348	0.232486	0.247360	0.742526
28	0.368187	0.247120	0.256806	0.725110
29	0.368215	0.241477	0.252176	0.731347
30	0.364544	0.257483	0.248535	0.716462
31	0.364711	0.251427	0.248602	0.721296

32	0.367363	0.250724	0.251839	0.723443
33	0.371126	0.242866	0.251777	0.731991
34	0.366043	0.246791	0.247075	0.727952
35	0.366335	0.250986	0.247080	0.724394
36	0.365567	0.246144	0.249244	0.726775
37	0.368617	0.258029	0.252128	0.713560
38	0.363502	0.250981	0.247695	0.721604
39	0.366565	0.248175	0.255167	0.723879
40	0.366645	0.250225	0.245910	0.721555

Figure 6.3 demonstrates the out-of-control run length performance. The control limits of the charts for Figure 6.3(a) – (d) are calibrated for an ARL_0 of 200 as 12.93, 11.09, 9.88, and 10.97, respectively, drawn as horizontal lines in the Figure 6.3. The vertical axis represents the monitoring statistic value. All charts are characterized by a smoothing parameter of 0.1. The FEWMA and GMEWMA have the ratio parameter c and ω as 0.42 and 0.44, respectively, according to the correlation structure, and they are set for the comparison with the same magnitude of diagonal elements. In Figure 6.3(c), the MEWMA chart shows the out-of-control signal at the 36th observation. The FEWMA chart and the dMEWMA chart detect the shift at the 31st and 33rd observations, respectively, while the GMEWMA chart detects it at the 29th observation. This demonstrates that the proposed chart significantly improves the detection ability when compared with the other charts.



(a) MEWMA (b) FEWMA

(c) dMEWMA (d) GMEWMA

Figure 6.3. Control charts of (a) MEWMA, (b) FEWMA, (c) dEWMA and (d) GMEWMA for detecting a change when ARL_0 is 200.

6.6. Conclusion

In this chapter, we present an interpretation of the non-diagonal elements of the smoothing matrix and develop a new method for assigning these non-diagonal elements. We also

present guidelines to determine the proper ratio factor between diagonals and off-diagonals considering the correlation structure of variables so that the processes can be monitored adaptively.

The GMEWMA chart shows a better ARL performance when compared with the existing EWMA-based approaches, specially in detecting smaller shifts. As the correlation structure becomes more unbalanced, the in-control ARL performance of the GMEWMA chart increases and approaches the performance of the MEWMA chart. Moreover, the proposed GMEWMA chart demonstrates better performance in most of the correlation structures whereas the FEWMA chart performs well when all of the correlation coefficients are identical, which can rarely happen. In addition, we investigate the advantage of the proposed chart in two different cases when: (i) the process starts up with an exposed risk of out-of-control, and (ii) process undergoes the out-of-control after stabilization. The performance of the GMEWMA chart is superior to all charts in both cases. It is more comparable to another accumulation-based chart such as dMEWMA chart when a large shift occurs in steady state.

Similar to the FEWMA chart, the GMEWMA chart is not directionally invariant. We demonstrate the chart performance with cases of multiple shift scenarios. Despite the directional variant property, the chart performance is robust to the directional issue.

Even though we can interpret the meaning of the ratio factor based on the moving average concept, finding an optimal ratio theoretically in terms of the shift magnitude is challenging.

We expect that theoretical derivation of the average run length may provide a better suggestion for the chart design resulting in a better ARL performance.

CHAPTER 7

BAYESIAN SEQUENTIAL UPDATE FOR MONITORING HIGH-DIMENSIONAL PROCESSES

7.1. Introduction

The VS-based methods in Chapter 3 and 4 select several suspicious variables that are believed to be the cause of process abnormality, and to monitor the process based on the selected variables. Specifically, the process parameter of interest such as a mean $\boldsymbol{\mu}_t$ at sampling point t is estimated as a sparse vector, $\hat{\boldsymbol{\mu}}_t$ consisting of mostly zero when the in-control parameter, $\boldsymbol{\mu}_0 = \mathbf{0}$. Although these VS-based methods perform well in monitoring high-dimensional processes under sparsity, they do not perform well in detecting small process changes, which is substantially of significance in modern industries, because VS methods would possibly fail to identify the truly changed variables. Furthermore, they do not involve the sequential information of the process change in performing VS, i.e., the current VS is independent of the result from the previous VS, which makes it inappropriate in high-speed online monitoring. Thus, the location of the nonzero components in $\hat{\boldsymbol{\mu}}_t$ may be sampling specific in existing VS-based methods. Consequently, this would become an issue of the VS-based methods detection when the size of the shift is sufficiently small because VS would possible select different variables at every sampling point due to the large type II error, which is the motivation of this chapter.

In this chapter, we propose an efficient methodology for monitoring high-dimensional processes through the Bayesian sequential update for the posterior distribution of the process mean. The principle of the proposed approach is that the parameter of interest follows a stochastic process; i.e., $\boldsymbol{\mu}_t$ is in a dynamic state and is determined through the path, $\boldsymbol{\mu}_{t-1}, \boldsymbol{\mu}_{t-2}, ..., \boldsymbol{\mu}_1$. The Bayesian approach sequentially updates the posterior distribution of $\boldsymbol{\mu}_t$ and supports the decision making in detecting the process change. Moreover, the analytical solution of the estimation with appropriately defined sparsity-promoting prior distribution of the parameter makes it attractive in online monitoring of the high-dimensional processes. The technical derivation of the Bayesian sequential update for the posterior distribution is described in Section 7.2. Practical significances and advantages of the proposed method over the existing methodologies such as VS-based charts and the Kalman filter model in the viewpoint of updating the posterior are discussed in Section 7.3. In Section 7.4, we demonstrate the power of the proposed method especially in detecting the small shift via various simulations and present the graphical implementation of the path of the estimated parameter. An illustrative example is presented in Section 7.5, and is followed by the conclusion in Section 7.6.

7.2. Bayesian Sequential Update

Let the p -dimensional process observation $\{\mathbf{x}_t : t = 1, 2, ...\}$ be represented by a linear generative model

$$\mathbf{x}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t$$

where $\boldsymbol{\varepsilon}_t$ represents an independent and identically distributed random sequence. The unknown parameter, $\boldsymbol{\mu}_t$ is continuously monitored and can be appropriately described stochastically. A Markovian state dynamics can be applied to the model of the unknown state, $\boldsymbol{\mu}_t$ as

$$\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t + \boldsymbol{\eta}_t,$$

where $\boldsymbol{\eta}_t$ is the model error. Under the sparsity assumption, a few or moderate number of process parameters would be responsible for the process abnormality. Moreover the variability of $\eta_{t,i}$ can be interpreted as a change of the state $\mu_{t,i}$ for the variable i. Details of the prior distribution is discussed in Section 7.2.1.

Process monitoring with a single observation $\mathbf{x}_t$ at each sampling epoch can be carried out statistically through the sequential hypothesis testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ and $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, where $\boldsymbol{\mu}_0$ is an in-control process parameter. As assumed in previous chapters, the covariance remains in control, and $\boldsymbol{\mu}_0 = 0$ without loss of generality. Then the Bayesian decision rule for out-of-control process is $\Lambda(\mathbf{x}_t) = p(H_0 \mid \mathbf{x}_t) / p(H_1 \mid \mathbf{x}_t) < L,$ where L is a predetermined threshold (Jeffreys, 1998, Woodward and Naylor, 1993). Thus, we propose

the following monitoring statistic with respect to the estimated parameter, $\hat{\boldsymbol{\mu}}_t$ by applying the Bayes rule and Jeffrey's prior, i.e., $p(H_0) = p(H_1)$, to the criterion, and refer to it as a Bayesian sequential update (BSU) chart.

$$\Lambda(\mathbf{x}_t) = \frac{p(\mathbf{x}_t \mid H_0)}{\max_{\boldsymbol{\mu}_t} p(\mathbf{x}_t \mid H_1)} < L.$$

Further notice that the Bayesian decision rule becomes identical to the general likelihood ratio test statistic with a single observation available at every sampling point and the assumption of Jeffrey's prior (Kass and Raftery, 1995). The strategy to increase the sensitivity of the monitoring statistic with respect to the high-dimensional vector $\boldsymbol{\mu}_t$ is to update the posterior probability distribution of $\boldsymbol{\mu}_t$ given a set of all sequential observations and the previous estimates. Denoting $\tilde{\mathbf{x}}_t = \{\mathbf{x}_1, ..., \mathbf{x}_t\}$ and $\tilde{\boldsymbol{\mu}}_t = \{\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_t\}$, the Bayesian analysis updates the posterior distribution of $\boldsymbol{\mu}_t \mid \tilde{\mathbf{x}}_t, \tilde{\boldsymbol{\mu}}_{t-1}$ stochastically as

$$\begin{aligned} p\left(\boldsymbol{\mu}_t \mid \tilde{\mathbf{x}}_t, \tilde{\boldsymbol{\mu}}_{t-1}\right) &= \frac{p(\tilde{\mathbf{x}}_t, \tilde{\boldsymbol{\mu}}_t)}{p(\tilde{\mathbf{x}}_t, \tilde{\boldsymbol{\mu}}_{t-1})} \\ &= \frac{p(\tilde{\mathbf{x}}_{t-1})}{p(\tilde{\mathbf{x}}_t)} p(\boldsymbol{\mu}_t \mid \boldsymbol{\mu}_{t-1}) p(\mathbf{x}_t \mid \tilde{\mathbf{x}}_{t-1}, \tilde{\boldsymbol{\mu}}_t) \\ &\propto p(\boldsymbol{\mu}_t \mid \boldsymbol{\mu}_{t-1}) p(\mathbf{x}_t \mid \boldsymbol{\mu}_t) \end{aligned} \tag{7.1}$$

Eq. (7.1) can be interpreted that the posterior distribution given the sequence of observations and the path of the process parameter is proportional to the updated prior

distribution of $\boldsymbol{\mu}_t \mid \boldsymbol{\mu}_{t-1}$ and the likelihood probability represented by $p(\mathbf{x}_t \mid \boldsymbol{\mu}_t)$. The second equality holds due to the Markovian property. (See Appendix D.)

The generality that results from the vast literature of the SPC enables close estimate of the parameter, $\boldsymbol{\mu}_t$ to the true process parameter. This leads to the most sensitive value of the monitoring statistic against the process abnormality. Accordingly, we obtain the most probable process parameter, $\hat{\boldsymbol{\mu}}_t$ by maximizing a posterior distribution (MAP) as

$$\hat{\boldsymbol{\mu}}_t = \arg\max_{\boldsymbol{\mu}_t} p\left(\boldsymbol{\mu}_t \mid \mathbf{x}_t\right) = \arg\max_{\boldsymbol{\mu}_t} \ln\left(p(\boldsymbol{\mu}_t \mid \boldsymbol{\mu}_{t-1})\, p(\mathbf{x}_t \mid \boldsymbol{\mu}_t)\right) \tag{7.2}$$

7.2.1. Prior Probability Distributions

In this section we discuss the specification of the likelihood probability and the prior distribution under sparsity in high-dimensional processes. Given the linear model $\mathbf{x}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t$, the likelihood probability is determined through the distribution of $\boldsymbol{\varepsilon}_t$. When the Gaussian error distribution is considered as $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, the likelihood probability can be derived through the multivariate normal distribution with the mean $\boldsymbol{\mu}_t$ and the covariance $\boldsymbol{\Sigma}$.

Now let $p(\boldsymbol{\mu}_t \mid \boldsymbol{\mu}_{t-1}; \boldsymbol{\theta}, \mathbf{K})$ be an updated prior distribution with a location parameter, $\boldsymbol{\theta}$ and a scale parameter, $\mathbf{K}$, respectively. The posterior distribution of $\boldsymbol{\mu}_{t-1}$ given $\tilde{\mathbf{x}}_{t-1}$ and $\tilde{\boldsymbol{\mu}}_{t-2}$

includes not only the explicit fluctuation of $\boldsymbol{\mu}_{t-1}$ given past information but also the implicit belief of the likelihood of $\boldsymbol{\mu}_t$. The latter advocates the use of a distribution that places more mass at zero for contemporaneously sparse shifts. Moreover, the scale matrix for the prior distribution, $\mathbf{K} = [\kappa_{ij}]$ refers to the variance of $\boldsymbol{\mu}_t$, and plays a key role in the variability of the process change as seen in the state dynamics. Indeed, we can let $\mathbf{K} \equiv f(\boldsymbol{\Sigma})$ as having more observations although the variability of $\boldsymbol{\mu}_t$ does not necessarily to be the same $\boldsymbol{\Sigma}$ over time. Thus, the matrix $\mathbf{K}$ can be rather determined based on the knowledge of engineering and physical structure of the process dynamics than merely be set through f over time. To promote the dynamics of $\boldsymbol{\mu}_t$ and the sparse shift in high-dimensional process, we consider the Laplacian prior of $\mu_{t,i}$ with time-varying parameters, $\theta_{t,i}$ and $\mathbf{K}_{t,ii} \equiv \kappa_{t,i}$. Moreover, we assume $\mathbf{K}_{t,ij} = 0$, for all $i \neq j$ over t because updating the covariance in dynamics is not apparent with non-Gaussian distribution compared to the Gaussian Kalman filter update, and it does not provide the meaningful interpretation in a sparsity perspective (Charles *et al.*, 2016). Thus, the prior distribution for the variable i is determined as follows.

$$p\left(\mu_{t,i} \mid \mu_{t-1,i}, \kappa_{t,i}\right) = \frac{\kappa_{t,i}}{2} e^{-\kappa_{t,i}\left|\mu_{t,i} - \hat{\mu}_{t-1,i}\right|}, \qquad i = 1, \ldots, p. \tag{7.3}$$

The location parameter, $\theta_{t,i}$ of the prior can be set as the parameter estimated in the posterior distribution at $t-1$, i.e., $\theta_{t,i} = \hat{\mu}_{t-1,i}$, which analogies a linear Gaussian Kalman filter recursion.

Based on the Laplacian prior and the Gaussian likelihood probability distributions, a MAP leads to the solution as follows.

$$\hat{\boldsymbol{\mu}}_t = \arg\min_{\boldsymbol{\mu}_t} \left\{ \left\| \mathbf{x}_t - \mathbf{Z}_t \boldsymbol{\mu}_t \right\|_{\boldsymbol{\Sigma},2}^2 + \sum_{i=1}^{p} \kappa_{t,i} \left| \mu_{t,i} - \hat{\mu}_{t-1,i} \right| \right\} \tag{7.4}$$

where $\| \mathbf{x} \|_{\boldsymbol{\Sigma},2}^2$ represents $\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x}$. Eq. (7.4) forms a weighted L_1 optimization which is similar to the fused LASSO introduced by Tibshirani *et al.* (2005). The term $\kappa_{t,i} \left| \mu_{t,i} - \hat{\mu}_{t-1,i} \right|$ encourages sparsity in the deviation of the consecutive process parameters. Thus, it enforces the deviation towards zero if there is no change detected, otherwise it remains as nonzero.

7.2.2. Bayesian Hierarchical Model

In this section, we develop a Bayesian hierarchical model for the dynamics of the scale parameter, $\kappa_{t,i}$ to control sparsity discussed in the previous section. We consider the gamma distribution for $\kappa_{t,i}$ which explains the probabilistic behavior of the variability properly in Bayesian statistics as

$$p(\kappa_{t,i}) = \frac{\beta^{\alpha}\kappa_{t,i}^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta\kappa_{t,i}}$$

where $\Gamma(\cdot)$ is the gamma function and α and β are the hyperparameters. As the process continues to operate in control, the posterior parameters are updated with sufficient amount of the normal observations, which leads to decreasing the variability of $\boldsymbol{\mu}$. Since $\kappa_{t,i}$ is the reciprocal of the variance of the prior distribution, it becomes large as the process remains in control. On the other hand, when several variables cause the process to be out of control, the corresponding $\kappa_{t,i}$'s are expected to become smaller since these variables would have large variabilities.

The posterior distribution can now be maximized with respect to both $\boldsymbol{\mu}_t$ and $\kappa_{t,i}$'s. However, even in a simple case with $\alpha = 1,$ the MAP estimator becomes

$$\{\hat{\boldsymbol{\mu}}_t, \hat{\boldsymbol{\kappa}}_t\} = \arg\min_{\boldsymbol{\mu}_t, \boldsymbol{\kappa}_t} \left\{ \left\| \mathbf{x}_t - \mathbf{Z}_t\boldsymbol{\mu}_t \right\|_{\boldsymbol{\Sigma},2}^2 + \sum_{i=1}^{p} \kappa_{t,i} \left| \mu_{t,i} - \hat{\mu}_{t-1,i} \right| - \sum_{i=1}^{p} \ln \kappa_{t,i} + \beta \left(\sum_{i=1}^{p} \kappa_{t,i} \right) \right\} \qquad (7.5)$$

which is challenging to be optimized for both $\hat{\boldsymbol{\mu}}_t$ and $\hat{\boldsymbol{\kappa}}_t$. Therefore, we propose the following expectation-maximization (EM) algorithm.

$$\begin{aligned} &\text{Expectation step:} && E\left[\kappa_{t,i}\right]^{(n)}_{p(\kappa_{t,i}|\hat{\mu}_{t,i}^{(n)})} \\ &\text{Maximization step:} && \mu_{t,i}^{(n+1)} = \arg\max_{\mu_{t,i}} p(\boldsymbol{\mu}_t \mid \boldsymbol{\kappa}_t^{(n)}) \end{aligned} \tag{7.6}$$

where $E\left[\kappa_{t,i}\right]^{(n)}_{p(\kappa_{t,i}|\hat{\mu}_{t,i}^{(n)})}$ denotes a conditional expectation of $\kappa_{t,i}$ in the n^{th} iteration with respect to the conditional density, $p(\kappa_{t,i} \mid \hat{\mu}_{t,i}^{(n)})$. The algorithm terminates the iterations if

$$E\left[\kappa_{t,i}\right]^{(n)}_{p(\kappa_{t,i}|\hat{\mu}_{t,i}^{(n)})} - E\left[\kappa_{t,i}\right]^{(n-1)}_{p(\kappa_{t,i}|\hat{\mu}_{t,i}^{(n-1)})} < \varepsilon,$$

where ε is a sufficiently small value. The conditional distribution, $p(\kappa_{t,i} \mid \hat{\mu}_{t,i}^{(n)})$ can be obtained through the Bayesian rule as

$$p(\kappa_{t,i} \mid \hat{\mu}_{t,i}^{(n)}) = \frac{p(\hat{\mu}_{t,i}^{(n)} \mid \kappa_{t,i})\, p(\kappa_{t,i})}{p(\hat{\mu}_{t,i}^{(n)})}, \tag{7.7}$$

where $p(\hat{\mu}_{t,i}^{(n)} \mid \kappa_{t,i})$ is a Laplacian distribution and $p(\kappa_{t,i})$ is a gamma distribution. The probability density, $p(\hat{\mu}_{t,i}^{(n)})$ is obtained analytically by marginalizing in terms of $\kappa_{t,i}$ (Appendix E). Then, $p(\kappa_{t,i} \mid \hat{\mu}_{t,i}^{(n)})$ is analytically obtained as a gamma distribution as

$$p(\kappa_{t,i} \mid \hat{\mu}_{t,i}^{(n)}) = \frac{(\beta + \mid \hat{\mu}_{t,i}^{(n)} - \hat{\mu}_{t-1,i} \mid)^{\alpha+1} \kappa_{t,i}^{\alpha}}{\Gamma(\alpha+1)} e^{-(\beta + |\hat{\mu}_{t,i}^{(n)} - \hat{\mu}_{t-1,i}|)\kappa_{t,i}}$$

with parameters, $\alpha+1$ and $\beta+|\hat{\mu}_{t,i}^{(n)}-\hat{\mu}_{t-1,i}|$ for the shape and the inverse scale parameters, respectively, so called conjugacy (Charles *et al.*, 2016, Garrigues and Olshausen, 2010). In particular, the expected value of $\kappa_{t,i}$ conditioned on $\hat{\mu}_{t,i}^{(n)}$ exactly determines the solution of the E-step as a closed form as

$$E\left[\kappa_{t,i}\right]_{p(\kappa_{t,i}|\hat{\mu}_{t,i}^{(n)})}^{(n)}=\frac{\alpha+1}{\beta+|\mu_{t,i}^{(n)}-\hat{\mu}_{t-1,i}|}, \tag{7.8}$$

which enhances the computation of the algorithm significantly.

7.2.3. Determination of Hyperparameters

The hyperparameters in the prior distribution play an important role in controlling the sparsity. Since the scale parameter, $\kappa_{t,i}$ can be seen as a regularization in Eq. (7.4), it is appropriate to be set according to the level of shift in the process variable i. In particular, if the process parameter i is believed to be shifted, $\kappa_{t,i}$ is set to be small, which leads to the nonzero value of μ_i. On the contrary, if the process parameter i is considered as unchanged, $\kappa_{t,i}$ is set as large so that the estimation of μ_i will be most likely a zero. To introduce the dynamics of $\boldsymbol{\mu}$ over time-varying $\boldsymbol{\kappa}$ to each variable, we determine the hyperparameters of $p(\kappa_{t,i})$ for which the conditional expectation controls sparsity based on the path of the estimates, i.e., $\tilde{\hat{\mu}}_{t,i}$ as an informative way as follows.

$$\beta_{t,i} = \alpha\left[f(\hat{\mu}_{t,i}^{(n)}, \hat{\mu}_{t-1,i}) + \zeta \right]$$

where ζ is a sufficiently small positive value to avoid $\beta_{t,i} = 0$. We note that the hyperparameter, α is also of importance for the determination of $E[\kappa_{t,i}]$. However, the effect of α can be approximately cancelled out, by setting β as a linear function of α because the term in the denominator $|\mu_{t,i}^{(n)} - \hat{\mu}_{t-1,i}|$ is mostly small over the sampling period assuming the high-speed online monitoring. Thus, the expectation of $\kappa_{t,i}$ is bounded in a small value regardless of α.

Furthermore, the conditional expectation of $\kappa_{t,i}$ regularizes the solution behavior of MAP in two ways. Firstly, $\kappa_{t,i}$ controls the sparsity against the deviation of the current estimate and the previous estimate from Eq. (7.4). Thus, if a sudden process shift occurs at t, the deviation, $|\mu_{t,i}^{(n)} - \hat{\mu}_{t-1,i}|$ would become large resulting in the nonzero value for $\hat{\mu}_{t,i}$ due to a small value of the conditional expectation of $\kappa_{t,i}$ from Eq. (7.8). Secondly, the hyperparameter, $\beta_{t,i}$ incorporates the information of the entire sequence of the parameter estimates, $\tilde{\hat{\mu}}_{t,i}$ through a function $f(\hat{\mu}_{t,i}^{(n)}, \hat{\mu}_{t-1,i})$ by the Markovian state dynamics. Thus, $f(\hat{\mu}_{t,i}^{(n)}, \hat{\mu}_{t-1,i})$ can be set proportional to the magnitude of successive estimates of the process parameter. For example, if the current and the previous estimates are sufficiently large, we consider the process to be out of control. Similarly, when the consecutive estimates are close to the in-control value, the process would likely be in control.

Throughout this chapter, we consider $f(\hat{\mu}_{t,i}^{(n)}, \hat{\mu}_{t-1,i}) = | c\hat{\mu}_{t,i}^{(n)} + (1-c)\hat{\mu}_{t-1,i} |$, as a weighted average of the consecutive estimates, and $0 \le c \le 1$.

7.3. Discussion

In this section we investigate the relationship of the proposed model with the linear state space model and the penalized likelihood based model in VS-based control charts. We investigate the similarity/dissimilarity and the advantages of the proposed method over the methods through those models.

7.3.1. Kalman Filter Model

The sequential Bayesian update with the Markovian model for the parameter can be written through a state space representation as

$$\begin{aligned} \boldsymbol{\mu}_{t+1} &= \boldsymbol{\mu}_t + \boldsymbol{\eta}_t \\ \mathbf{x}_t &= \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t \end{aligned} \tag{7.9}$$

Assuming that the errors follow normal distribution, we obtain the solution for the estimation of $\hat{\boldsymbol{\mu}}_t$ as

$$\hat{\boldsymbol{\mu}}_t = \arg\min_{\boldsymbol{\mu}_t} \left\{ \left\| \mathbf{x}_t - \mathbf{Z}_t \boldsymbol{\mu}_t \right\|_{\boldsymbol{\Sigma},2}^2 + (\mu_{t,i} - \hat{\mu}_{t-1,i})^T (\mathbf{P}_{t|t-1} + \mathbf{K}_t)^{-1} (\mu_{t,i} - \hat{\mu}_{t-1,i}) \right\} \tag{7.10}$$

where $\mathbf{P}_{t|t-1}$ is the conditional variance of the parameter $\boldsymbol{\mu}_t$ given past observations, $\tilde{\mathbf{x}}_{t-1}$ and $\mathbf{K}_t$ is a scale matrix of the model error, $\boldsymbol{\eta}_t$. This has a closed form of the solution for $\hat{\boldsymbol{\mu}}_t$ known as the Kalman filter update and guarantees the simplicity and optimality in linear Gaussian model assumption (Durbin and Koopman, 2012). However, this model does not cover the sparsity in the estimation even if it may shrink the estimation close to zero as a ridge estimator or a principle component analysis. A number of approaches have been developed in the spirit of the linear state space model for the different purposes such as nonlinear system dynamics and sparse estimation of the high-dimensional parameter (Burgers *et al.*, 1998, Julier and Uhlmann, 1997, Taghvaei *et al.*, 2018, Xiong, 2008).

The proposed model can be modelled through a state space as in (7.9) by assuming $\boldsymbol{\mu}_{t+1} \mid \boldsymbol{\mu}_t$ follows Laplacian, which is identical to the parameter equation with the Laplacian error distribution. While the Gaussian Kalman filter model has closed form expression for the conditional expectation and the variance for the estimator is obtained through propagation from the sequence, the proposed model employs the EM algorithm for the variance of the estimator rather than evolving it through the dynamics. The reason is that updating the variance is not immediately apparent with Laplacian distribution, and it would probably not provide the meaningful update in a sparsity perspective. Another reason is the computational difficulty because when the dimension of the signal, p is sufficiently high, it demands $O(p^3)$ of the complexity for computing the inverse covariance matrix at each sampling point. Therefore, although we do not evolve the covariance matrix for the

estimator, the $\kappa_{t,i}$ determined from the past sequence of the estimates would provide more appropriate solution for the sparse estimation.

7.3.2. VS-Based Control Charts

The recently developed methodologies for monitoring high-dimensional processes can be summarized through the penalized likelihood function.

$$PL(\boldsymbol{\mu}_t) = \min_{\boldsymbol{\mu}_t} \left\{ \left\| \mathbf{x}_t - \boldsymbol{\mu}_t \right\|_{\boldsymbol{\Sigma},2}^2 + \kappa g(\boldsymbol{\mu}) \right\}$$

where $g(\bullet)$ is a penalty function and $\kappa(\geq 0)$ is a complexity parameter that controls the size of penalty. Wang and Jiang (2009) and Jiang *et al.* (2012) apply L_0 type penalty function, and Zou and Qiu (2009) employ L_1 type penalty, which is a LASSO penalty. The methods with L_0 type penalty require another parameter that controls the sparsity, i.e., the number of nonzero in $\boldsymbol{\mu}$. Thus, when the number of true shifted variables is deviated from the number of selected variables, the control chart may lose the power of detection regardless of the choice of VS technique. Zou and Qiu (2009) choose the most likely number of suspicious variables through the maximal type of monitoring statistic that maximizes the value of monitoring statistic according to the different number of selection. Thus, this method is expected to perform better than the one of choosing the fixed number of selection when the true number of changed variables is completely unknown.

The proposed method estimates the process parameter, $\boldsymbol{\mu}_t$ using Eq. (7.4) which is similar to the penalized likelihood. The difference is that the complexity parameter, $\boldsymbol{\kappa}$ in the proposed method is determined dynamically in online processing as a vector while existing methods employ a scalar value of κ. In addition, the maximal statistic by Zou and Qiu (2009) can be seen similarly to the proposed method in that the estimation of $\boldsymbol{\mu}_t$ results from the different κ at every sampling point. However, it may lead to the suboptimal solution since the same κ is applied for all variables regardless of the status of each parameter, i.e., whether the individual variable is in-control or out-of-control at a certain sampling point, while the proposed methods determines the different κ to each variable according to its sequential path of parameter estimations. Therefore, the estimator of the proposed method would be flexibly changing over time and adaptively selecting variables under sparsity based on the potential shift.

7.4. Performance Analysis

In this section, we present simulation results illustrating the performance of the proposed BSU chart. In order to detect small shifts, we transform the observation vector $\mathbf{x}$ to the exponentially weighted moving average, i.e., $\mathbf{w}_t = (1-r)\mathbf{w}_{t-1} + r\mathbf{x}_t, \quad t = 1, 2, ...,$ and consider a small value of $r = 0.2.$ The performance of BSU control chart is compared with the conventional MEWMA chart and VSMEWMA chart (represented as VS in tables) which has been recently proposed for monitoring high-dimensional processes. In addition, the Kalman filter update can be also utilized to monitor the process based on the general

likelihood ratio test, $2\mathbf{x}_t^T\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}_t^{(K)} - \hat{\boldsymbol{\mu}}_t^{(K)T}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}_t^{(K)}$ in the same horizon as the penalized likelihood, where $\hat{\boldsymbol{\mu}}_t^{(K)}$ is the Kalman filter update shown in (7.10).

7.4.1. Out-of-Control ARL Performance

The performance of the chart is measured by the out-of-control average run length (ARL_1) whose in-control ARL (ARL_0) is fixed at 200; i.e., Type I error is set 0.005. The true shift occurs at $t = 100$ so that we assume the process is in control and ignore the alarm before 100 steps of sampling. Further, we consider few or moderate number of the components in $\boldsymbol{\mu}_1$, which is the out-of-control parameter, to be nonzero under sparsity. Denoting $p_0 = \| \boldsymbol{\mu}_1 \|_0$, we consider $p_0 = 2, 4, 6$ and 8, and the magnitude is equal values of δ, e.g., $\boldsymbol{\mu}_1 = [\delta\ \delta\ 0\ 0 \cdots 0]^T$, when $p_0 = 2$.

In VSMEWMA, the parameter, s as the number of selection must be chosen. Throughout the simulation study, we choose $s = 2$, which shows the overall best results for cases considered. In BSU, the weight c and the initial $\boldsymbol{\kappa}$, say $\boldsymbol{\kappa}_0$, to obtain $\hat{\boldsymbol{\mu}}_t^{(1)}$ for the first iteration of EM need to be specified. Likely to the EWMA transformed observation, the $f(\hat{\mu}_{t,i}^{(n)}, \hat{\mu}_{t-1,i})$ can be interpreted as the weighted average of the current estimate and the previous estimates shown in Section 2.3. Then, it is reasonable to set $c = r = 0.2$. Moreover, $\boldsymbol{\kappa}_0$ can be set any reasonable value so that it leads to at least a few nonzero values for $\hat{\boldsymbol{\mu}}_t^{(1)}$. We here choose $\kappa_{0,i} = 3$ for all i to be conservative to the shift. Determining the $\boldsymbol{\kappa}_0$ will be discussed more specifically in Section 7.4.3.

Tables 7.1, 7.2 and 7.3 show the ARL_1 performances according to $p = 25,\ 50$ and 100, respectively. The additive shift, δ is mostly set as a small size ranging from 0.2 to 2.0. The bold case represents the smallest ARL_1. The results demonstrate that the proposed chart outperforms the others in most of the cases. Even in the cases where BSU underperforms VSMEWMA, the ARL_1s are quite close to each other. Specially, the VSMEWMA shows the deterioration of the performance in detecting small shifts, which is one of the major drawbacks of VSMEWMA. Several VS-based control charts such as VSMSPC by Wang and Jiang (2009) and LEWMA by Zou and Qiu (2009) also have the same issue in detecting small process changes and show the worse performance than conventional control charts. For example, the ARL_1's for MEWMA, VSMEWMA and BSU are 134.73, 158.13 and 107.46, respectively, when $p_0 = 6$ with $\delta = 0.2$ in $p = 100$. The average detection time of BSU is almost two third of that of VSMEWMA and about four fifth of that of MEWMA, which is a significant improvement in detecting such a small change of the process. Specifically, the deterioration of the power of VSMEWMA in small shift detection can be figured out in the viewpoint of variable selection. This is possibly because the VS techniques would mostly fail to identify the suspicious variables. Although the EWMA transform incorporates the information of the past, the VS may malfunction due to the small size of the shift. However, BSU takes the previous selection into consideration to select the next suspicious variables by changing $\kappa_{t,i}$. Moreover, since $\kappa_{t,i}$ differs for all variables, BSU takes an advantage of selecting different number of variables according to their mean path so as to increase the chance of selecting the suspicious variables.

Table 7.1. ARL performance when $p = 25$

δ	$p_0 = 2$				$p_0 = 4$			
	MEWMA	VS	Kalman	BSU	MEWMA	VS	Kalman	BSU
0.2	152.52	160.97	150.19	142.15	119.95	132.55	115.52	107.02
0.4	78.62	78.24	77.51	68.00	42.47	48.35	41.21	34.68
0.6	37.83	33.62	36.78	30.70	18.48	18.81	17.21	14.65
0.8	20.71	16.98	19.39	16.42	10.60	10.34	9.62	8.70
1.0	13.15	10.56	12.20	10.53	7.40	6.97	6.64	6.16
1.5	6.81	5.25	6.06	5.41	4.35	3.92	3.84	3.61
2.0	4.68	3.57	4.11	3.72	3.23	2.79	2.77	2.64

δ	$p_0 = 6$				$p_0 = 8$			
	MEWMA	VS	Kalman	BSU	MEWMA	VS	Kalman	BSU
0.2	95.22	111.87	93.86	81.43	78.00	96.86	76.63	66.61
0.4	28.32	34.28	26.67	22.45	20.82	26.57	19.20	16.81
0.6	12.35	14.11	11.32	10.27	9.61	11.58	8.77	8.17
0.8	7.61	8.16	6.87	6.50	6.24	6.93	5.52	5.42
1.0	5.58	5.72	4.95	4.79	4.68	5.04	4.10	4.09
1.5	3.51	3.37	3.03	2.97	3.04	3.07	2.61	2.65
2.0	2.67	2.47	2.27	2.25	2.28	2.27	2.02	2.01

Table 7.2. ARL performance when $p = 50$

δ	$p_0 = 2$				$p_0 = 4$			
	MEWMA	VS	Kalman	BSU	MEWMA	VS	Kalman	BSU
0.2	164.13	170.98	163.60	147.89	137.03	150.67	136.70	116.64
0.4	100.20	100.50	99.36	80.30	59.15	63.93	57.20	42.30
0.6	53.14	43.39	50.83	37.69	25.99	24.65	24.00	17.34
0.8	29.24	20.64	27.44	19.43	14.14	12.49	12.92	9.88
1.0	18.19	12.30	16.57	12.10	9.47	8.02	8.40	6.81
1.5	8.64	5.81	7.64	6.07	5.38	4.36	4.58	3.94
2.0	5.79	3.89	4.92	4.13	3.91	3.04	3.21	2.87

δ	$p_0 = 6$				$p_0 = 8$			
	MEWMA	VS	Kalman	BSU	MEWMA	VS	Kalman	BSU
0.2	116.84	135.71	115.10	95.04	99.89	120.75	99.58	76.94
0.4	39.58	47.30	38.45	26.92	29.10	37.60	27.38	20.07
0.6	16.92	17.78	15.29	11.69	12.66	14.54	11.35	9.26
0.8	9.78	9.82	8.68	7.20	7.82	8.35	6.84	5.95
1.0	6.99	6.60	6.04	5.24	5.81	5.79	4.92	4.46
1.5	4.28	3.74	3.54	3.23	3.70	3.41	3.01	2.84
2.0	3.22	2.70	2.60	2.41	2.88	2.49	2.26	2.15

Table 7.3. ARL performance when $p = 100$

δ	$p_0 = 2$				$p_0 = 4$			
	MEWMA	VS	Kalman	BSU	MEWMA	VS	Kalman	BSU
0.2	180.09	187.67	173.16	156.58	155.26	173.80	152.08	130.53
0.4	125.10	125.13	118.56	91.97	82.53	87.31	77.31	49.19
0.6	75.00	57.15	69.71	44.55	38.42	34.05	34.97	18.78
0.8	43.34	26.62	39.95	22.62	20.11	16.16	17.84	10.33
1.0	26.36	14.91	24.13	13.15	12.83	9.88	11.08	7.18
1.5	11.43	6.72	9.90	6.43	6.79	5.11	5.59	4.20
2.0	7.31	4.41	6.12	4.41	4.84	3.57	3.83	3.06

δ	$p_0 = 6$				$p_0 = 8$			
	MEWMA	VS	Kalman	BSU	MEWMA	VS	Kalman	BSU
0.2	134.73	158.13	131.19	107.46	120.22	147.02	119.07	90.39
0.4	55.28	67.91	53.46	30.45	41.95	54.89	40.22	21.99
0.6	23.63	24.58	22.26	12.50	17.62	19.91	15.79	9.70
0.8	13.14	12.41	11.65	7.53	10.16	10.65	8.81	6.27
1.0	8.96	8.13	7.76	5.51	7.30	7.25	6.08	4.74
1.5	5.28	4.50	4.23	3.42	4.55	4.17	3.57	3.02
2.0	3.95	3.24	3.04	2.57	3.43	3.06	2.62	2.30

In the preceding experiments, the shift occurs at $t = 100$. As shown in several EWMA-based control charts, the behavior of ARL_1 may considerably differ when the process is out-of-control at the start up stage, e.g., due to the setup bias or the ramp-up period, called zero-state shift, from when the shift occurs after the process is stabilized, called steady-

state shift (Jiang *et al.*, 2012, Kim *et al.*, 2017b). For example, double MEWMA introduced by Alkahtani and Schaffer (2012) shows the substantial difference between the zero-state ARL and the steady state ARL (See Table 8 in Kim *et al.* (2017b)). Figure 7.1 shows the relative efficiency (RE) in terms of the power, which calculates the ratio of the power of each chart over that of MEWMA. Thus, the horizontal line represents the RE of MEWMA itself, and the RE greater than 1 indicates better power than MEWMA. The result shows that the steady-state ARL is slightly shorter than the zero-state ARL, but not too much different. Moreover, the RE of BSU presents always greater than 1, which indicates that the detection ability of BSU is always better than that of MEWMA. However, VSMEWMA shows considerably low detection power in small shift cases being worse than that of MEWMA, and the RE increases as the size of the shift increases, which is expected.

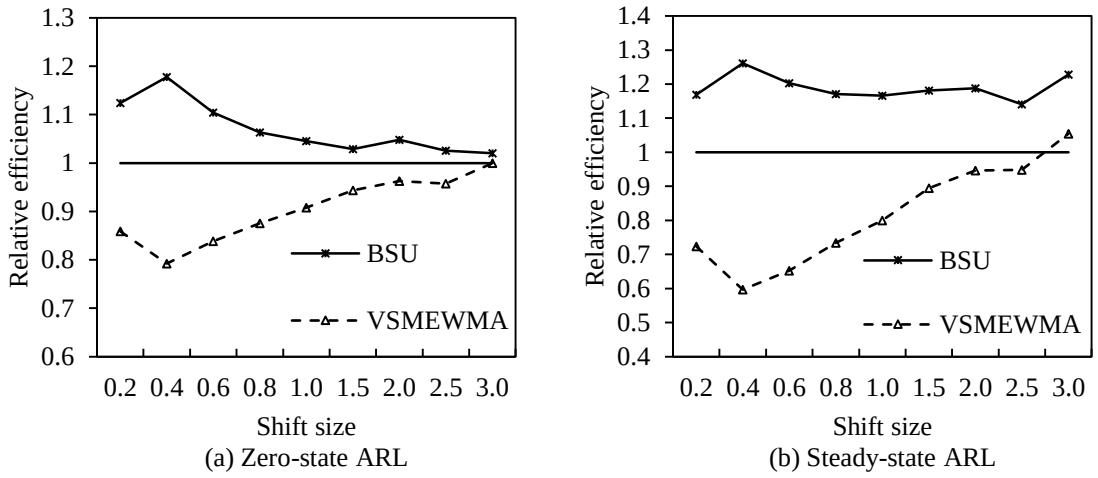


Figure 7.1. Zero-state ARL and steady-state ARL ($p = 25$)

7.4.2. Tracking the Process Mean

The preceding experiments fall into the problem of the hypothesis test, $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ and $H_1 : \boldsymbol{\mu} = \boldsymbol{\mu}_1$. This is a common test in conventional SPC literature in which $\boldsymbol{\mu}_1$ is considered as a sudden jump at certain point of time and remains unchanged after the occurrence. However, in many applications of the manufacturing process the process parameter may change over time by process shifts. For example, in the chemical manufacturing process, some process parameter shows a severe dynamic behavior once the disturbances are occurred, e.g., drift, oscillation and step change (Raich and Cinar, 1996). Here we construct the testing problem as $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ and $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, and the true out-of-control mean after the occurrence of the disturbance is considered as a function of the sampling time, i.e., $\boldsymbol{\mu}_1(t)$. This can be seen as tracking the multivariate sparse mean over time. In following experiments, we measure the mean squared error (MSE), i.e., $1/T \sum_{t=1}^{T} \| \hat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}_1(t) \|_2^2$, where T is the sampling period. We consider $T = 40$, and assume that the process is in control when $t \leq 10$, and out of control when $10 < t \leq T$. Moreover, one thousand replications are applied for computing the MSE. For the faulty mean functions, we consider $\delta_a(t) = 0.5 \arctan(t - \tau)$, where $\tau = 10$, for the drift type of the mean shift; consider $\delta_b(t) = 0.3 \sin(0.15(t - \tau))$, where $\tau = 10$, for the oscillation type of shift; and consider $\delta_c(t) = 0.3$ when $10 < t < 25$, and -0.2 when $t \geq 25$, for the step change of the shift. Figures 7.2, 7.3 and 7.4 illustrate the sample paths of the true mean and estimated means of MEWMA, VSMEWMA, Kalman filter update and BSU for one shifted variable. The

asterisks symbol is used to plot the observation, x_t; the solid line, dashed line, long dashed line, dash-dotted line and the solid line with asterisks represent the true mean path, MEWMA, VS-MESNA, Kalman update and BSU, respectively. The MSE values are shown in Table 7.4.

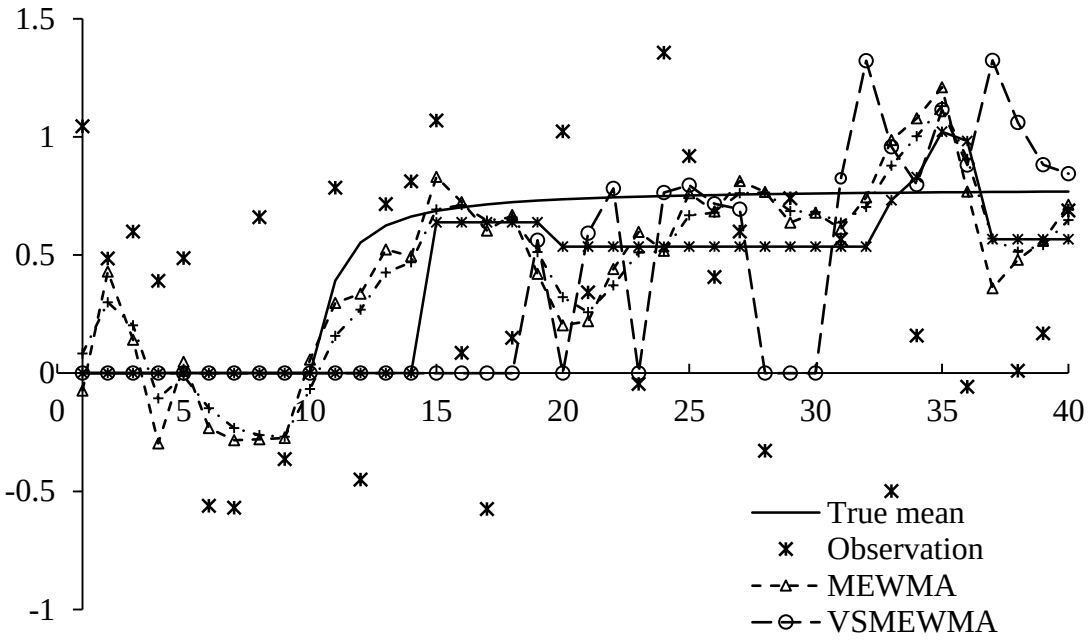


Figure 7.2. Sample paths of the estimated means with $\delta_a(t)$

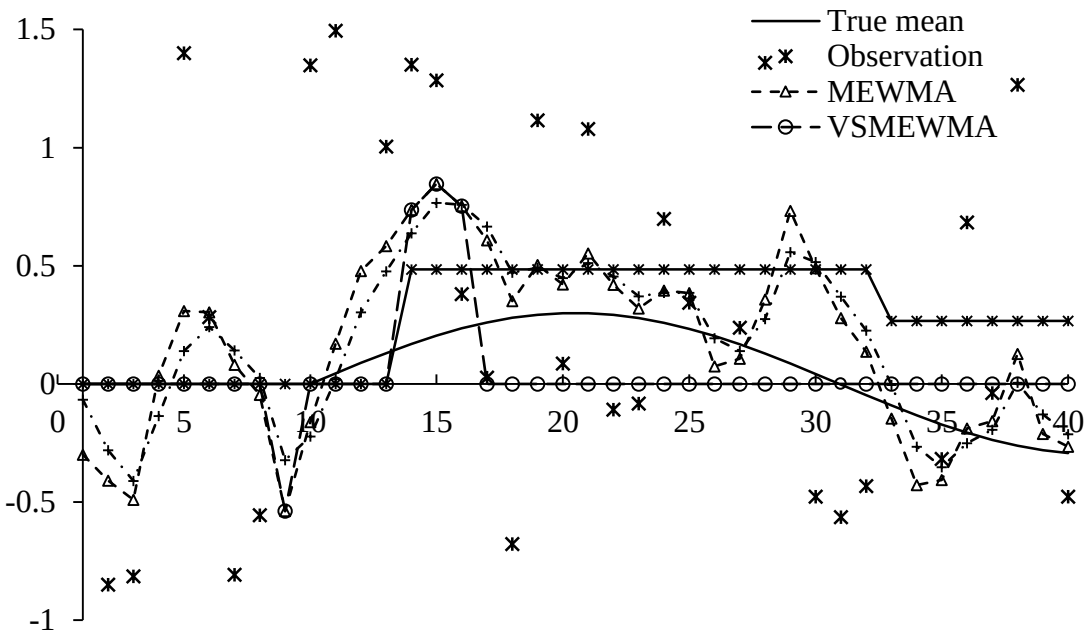


Figure 7.3. Sample paths of the estimated means with $\delta_b(t)$

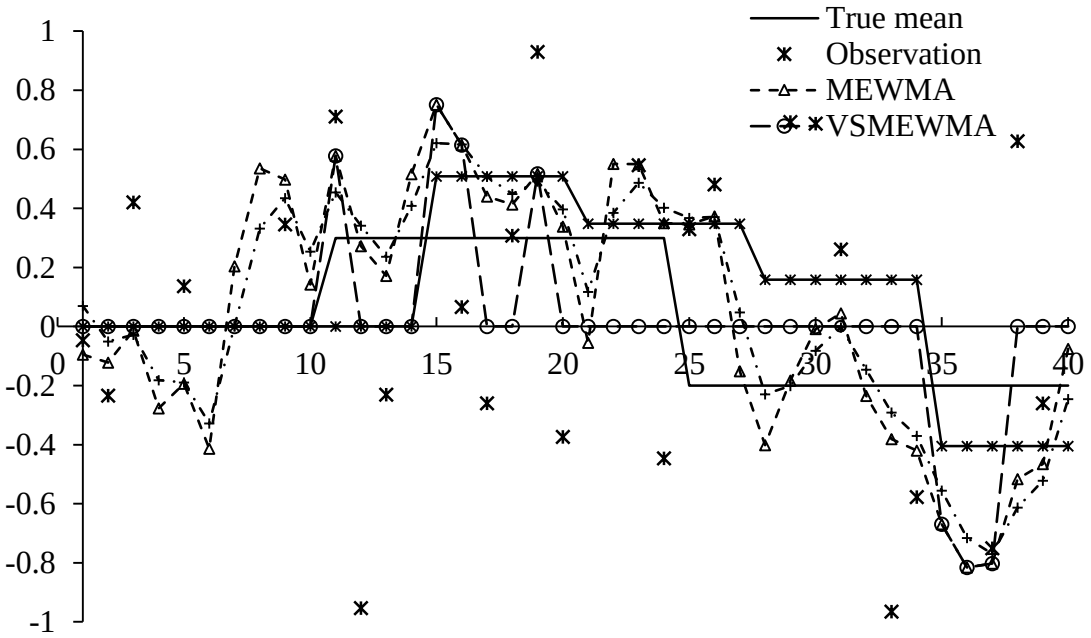


Figure 7.4. Sample paths of the estimated means with $\delta_c(t)$

Table 7.4. MSE of MEWMA, VSMEWMA, Kalman filter update and BSU for $\delta_a(t)$, $\delta_b(t)$ and $\delta_c(t)$

	$\delta_a(t)$	$\delta_b(t)$	$\delta_c(t)$
MEWMA	3.0326	2.8416	2.8902
VSMEWMA	2.3716	2.0198	2.0761
Kalman update	2.5598	2.3677	2.4303
BSU	1.6870	1.3384	1.4544

Figures 7.2, 7.3 and 7.4 show that BSU reacts quickly to the change of the mean. Since the EWMA transform is applied, all charts do not immediately react to the process change due to the inertia effect. However, BSU tends to trace the changing mean quite well after a short delay. On the other hand, it is hard to see whether the VSMEWMA keeps track of the mean. Although VSMEWMA starts plotting consecutive nonzero values from $t = 31$, it is still incapable of confirming that the process mean is accurately estimated. In Figures 7.3 and 7.4, BSU reacts the fluctuating mean and the changing the direction of the mean shift even if the size is small, while VSMEWMA cannot capture the process change over the most sampling points. In addition, the Kalman filter update rather tends to track the MEWMA regardless of the path of the true mean because the update is based solely on the MEWMA transformed observations. Although it provides a good smoothing the fluctuating measurements like a ridge estimator – apparently the Kalman filter update can reduce to the ridge estimator with certain conditions – it does not guarantee the detection ability under sparsity because the estimator (Kalman gain) still includes significant noises

and may contaminate the monitoring statistic. Accordingly, the result of MSE shown in Table 7.4 demonstrate that the BSU outperforms the other methods in tracking the true mean under sparsity conditions.

7.4.3. Determination of Initial Prior Parameter $\boldsymbol{\kappa}_0$

One remaining challenge to apply BSU is to determine the initial scale parameter, $\boldsymbol{\kappa}_0$ for the first iteration of EM. Basically the $\boldsymbol{\kappa}_0$ is directly related to the variance of the process parameter $\boldsymbol{\mu}$ before being updated in EM, where the variance of the Laplacian distribution is $2/\boldsymbol{\kappa}_t^2$. Thus, prior knowledge of the process by engineers or the engineering physics, if available, would enable to set $\boldsymbol{\kappa}_0$ properly based on the variability of the process shift. If there is no prior information available about a process shift, it can be reasonably set $\kappa_{0,i} = \sqrt{2}/\sigma_{ii}$, where σ_{ii} is the standard deviation of the i^{th} variable by setting the variance of μ_i equal to the process variability σ_{ii} (Jain *et al.*, 1993).

Although the parameter $\boldsymbol{\kappa}_0$ can be determined in a certain way, it will be changing dynamically over sampling time and adaptively to each variable at every sampling epoch based only on the observations. Thus, it is expected that the effect of the initial value of $\boldsymbol{\kappa}$ would not affect the performance of the chart significantly. Thus, as long as κ_0 is initially set within an acceptable range of the variation of the process parameter, e.g., l standard deviation (i.e., $l \times \sqrt{2}/\sigma_{ii}$), the ARL performance would be robust over the selection of

κ_0. Table 7.5 shows the effect of κ_0 in ARL performance when $p = 25$. We test the value of $\boldsymbol{\kappa}_0$ as $5 \approx 3\sqrt{2}$ with $\sigma_{ii} = 1$ for the maximum sparsity of the estimator from 1 for the minimum sparsity.

Table 7.5. Effect of κ_0

δ	$p_0 = 2$			$p_0 = 4$		
	$\kappa_0 = 1$	$\kappa_0 = 3$	$\kappa_0 = 5$	$\kappa_0 = 1$	$\kappa_0 = 3$	$\kappa_0 = 5$
0.2	144.63	145.61	146.44	108.05	109.18	109.36
0.4	68.72	69.58	69.50	36.11	35.83	36.01
0.6	32.38	31.87	31.07	15.73	15.81	15.93
0.8	17.45	17.44	17.12	9.56	9.39	9.79
1.0	11.42	11.29	11.33	6.81	6.80	6.95
1.5	6.10	6.06	5.98	4.10	4.10	4.12
2.0	4.24	4.24	4.11	3.00	3.00	3.00

δ	$p_0 = 6$			$p_0 = 8$		
	$\kappa_0 = 1$	$\kappa_0 = 3$	$\kappa_0 = 5$	$\kappa_0 = 1$	$\kappa_0 = 3$	$\kappa_0 = 5$
0.2	83.66	83.96	84.24	68.90	67.03	69.86
0.4	23.79	23.83	24.48	18.02	18.29	18.96
0.6	11.10	11.21	11.69	8.92	9.08	9.52
0.8	7.16	7.18	7.46	6.09	6.10	6.32
1.0	5.40	5.37	5.60	4.66	4.69	4.89
1.5	3.41	3.42	3.49	3.07	3.05	3.16
2.0	2.54	2.54	2.60	2.26	2.27	2.35

Table 7.5 shows that the ARL performance becomes slightly better with a small $\boldsymbol{\kappa}_0$ as p_0 increases and vice versa. This is reasonable because a small $\boldsymbol{\kappa}_0$ allows more nonzero values in the estimate so as to include more shift information, and the opposite case is also interpretable in the same way. Moreover, when δ is small, a small $\boldsymbol{\kappa}_0$ generates slightly better results because it decrease a chance of misidentification of VS by estimating $\boldsymbol{\mu}$ less sparse. However, above all, the difference among the ARLs in terms of $\boldsymbol{\kappa}_0$ is ignorable according to the results, which demonstrates the ARL performance of BSU is robust to the initial setting of $\boldsymbol{\kappa}_0$. This makes the proposed method more attractive to the practitioner because it can be seen as a parameter-free method. Throughout the chapter, we set $\boldsymbol{\kappa}_0 = 3$ which is approximately 2 standard deviation of the inherent variability of the process parameter μ_{ii}, i.e., $\boldsymbol{\kappa}_0 = 2\sqrt{2}/\sigma_{ii}$.

7.5. Case Study

In this section, we illustrate the proposed chart using an industrial process described by Tennessee Eastman (TE) Chemical Company (Downs and Vogel, 1993). The process is composed of five major units: a reactor, a product condenser, a recycle compressor, a vapor-liquid separator and a product stripper as shown in Figure 7.5. It manufactures two products along with a by-product from four reactants, and an inert which is produced separately through the downstream operations. The data set includes ten quality measurements to be monitored listed with their nominal values in Table 7.6. The normality of the error variance in the measurement equation is valid for the dataset from the Mardia's

multivariate normality test (Mardia, 1970). The shift occurs after the normal operation due to the change of one of the reactants' temperature (D feed temperature), which leads to a slight downward shift. The raw data are plotted in Figure 7.6.

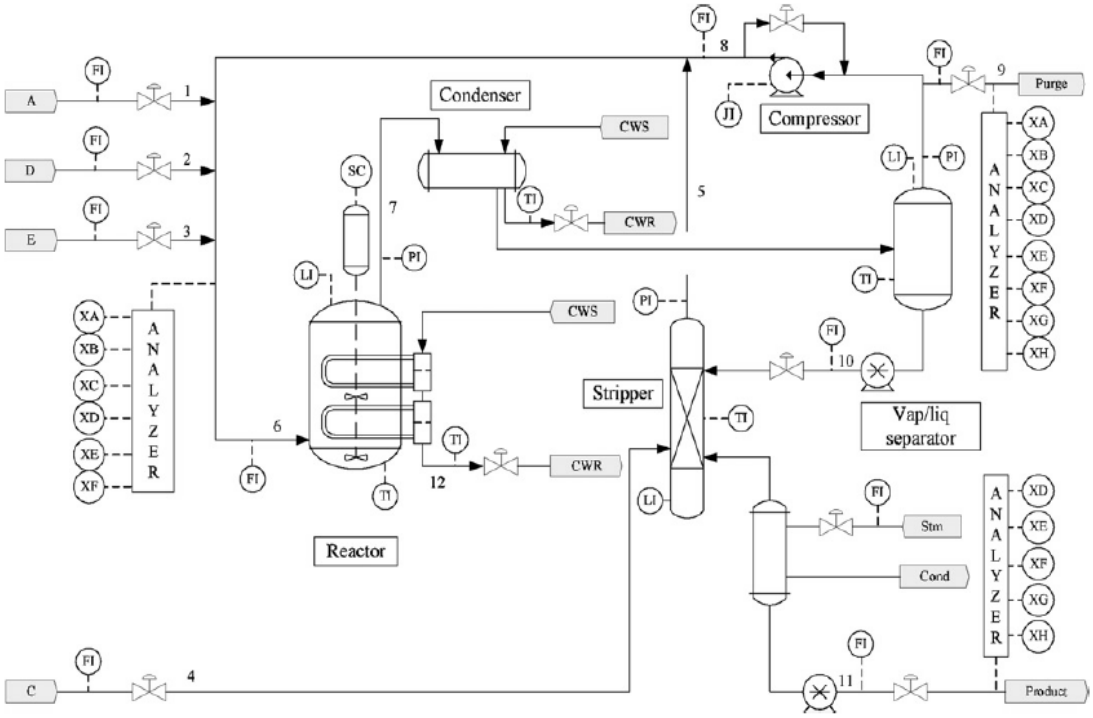


Figure 7.5. A diagram of the TE process (Yin *et al.*, 2012)

Table 7.6. Nominal values of the ten variables

Block name	Variable name	Nominal value	Units
Input feed	E feed flow	700.92	kg/h
	A and C feed flow	7.83	kscmh
Reactor	Reactor feed rate	32.15	kscmh
	Reactor level	65.00	%
	Reactor temperature	121.90	°C
Separator	Separator level	50.00	%
	Separator underflow	17.56	m^3/h
Stripper	Stripper underflow	18.00	m^3/h
	Stripper steam flow	5.31	kg/h
Miscellaneous	Reactor water temperature	101.86	°C

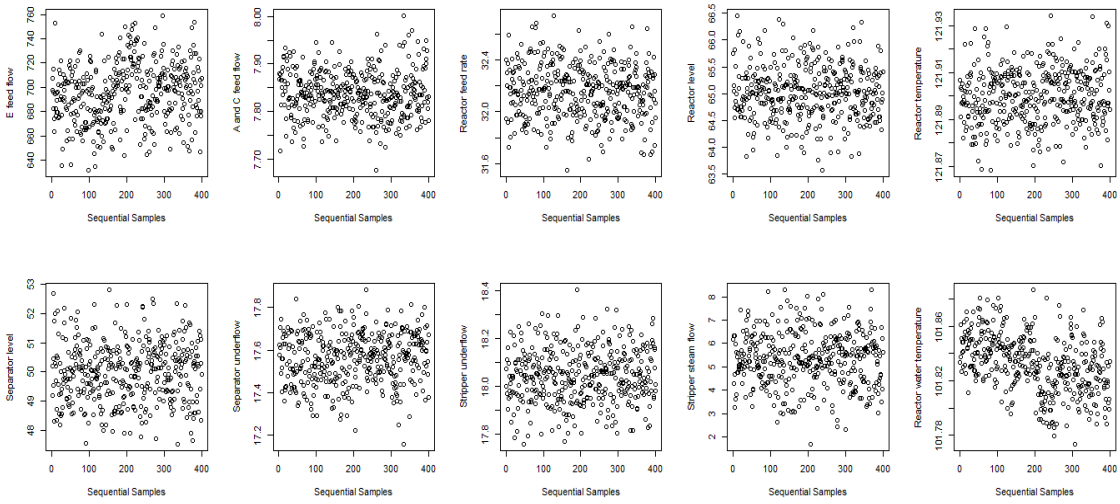


Figure 7.6. Raw data of the TE process

We illustrate the implementation of the BSU, MEWMA, VSMEWMA and Kalman filter update based control charts using this dataset. We set the significance level to 0.005 for all charts; the control limits of BSU, MEWMA, VSMEWMA and Kalman filter update are obtained as 1.8011, 2.6767, 1.7689 and 2.5978 based on the in-control parameters. The number of selection for VSMEWMA chart is set as $s = 2$. Ten observations are sampled before the occurrence of the shift and thirty observations are taken consecutively after the shift. Figures 7.7 shows the monitoring paths of the charts over the time. The solid line with connecting stars represents the BSU; the dashed line with triangles shows the MEWMA chart; the line with long dashes connecting diamonds reveals the VSMEWMA; and the dash-dotted line represents the Kalman filter update based chart. The same styles of the horizontal lines are applied to the corresponding control charts as their control limits. Values of monitoring statistics are shown on the left vertical axis.

The BSU chart triggers the alarm four times at the 29^{th}, 30^{th}, 31^{st} and the 36^{th} observations without adjusting the process after the alarms, while MEWMA, VSMEWMA and Kalman filter update charts triggers 2, 0 and 2 alarms, respectively. Moreover, except commonly alarmed 30^{th} and 31^{st} observations, BSU alarms at the 29^{th} observation as its first alarm, which shows the faster detection than others.

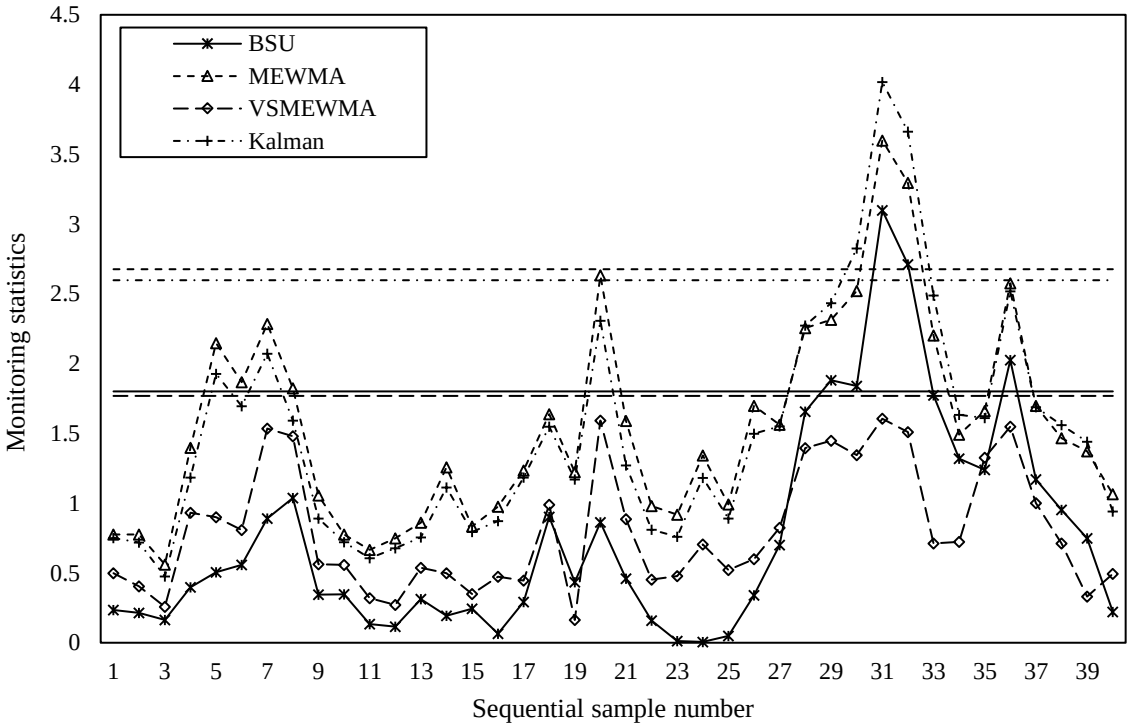


Figure 7.7. BSU, MEWMA, VSMEWMA and Kalman filter update charts for monitoring the TE process

7.6. Conclusion

In this chapter, we develop a Bayesian approach for monitoring high-dimensional processes. The BSU chart considers the stochastic behavior of the process parameter in high speed monitoring systems. Further, the sparsity of the high-dimensional vector is appropriately implemented at each sampling point according to the dynamic path of the process parameter. Several attractive features of the proposed chart are summarized as follows. The BSU chart adaptively selects the suspicious variables based on the recent selections. The hyperparameter of the prior distribution is appropriately determined at each sampling point based on the path of the previous estimates and reflected to update the prior

for the next sampling. This makes the BSU chart significantly useful in Phase II monitoring where the information of the shift, e.g., direction and the size, is completely unknown. Particularly, the BSU controls the sparsity, i.e., the number of nonzero elements in the estimate, and determines the value of nonzero based on the potential possibility of the shift for each variable.

Another significance of the proposed method is that the BSU is free from manual parameter adjustment. Although it requires to set the initial hyperparameter, the effect of the initial value diminishes rapidly in the iteration of EM. Moreover, the ARL_1 results show the robustness to the detection over various considerations of the initial prior parameter. Thus, when it is determined in a reasonable range, e.g., one to three sigma level of the confidence, the practitioner may not need to take actions to adjust the parameter even if the information of the shift is provided.

Furthermore, the closed form analytical expression of the conditional distribution given the history of the estimations makes the practical implementation of the online monitoring significantly convenient. Unlike the Kalman filter model, the BSU updates the posterior based on the empirically determined variation rather than evolving covariance. This simplifies the updating algorithm and reduces the computational complexity compared to the Kalman filtering. Moreover, while Kalman filtering is not immediately apparent in sparsity perspectives, the BSU provides a practical advantage to implement the sparsity.

CHAPTER 8

CONCLUSION AND FUTURE WORK

This dissertation proposes and subsequently develops procedures for monitoring of high-dimensional processes. This chapter presents the summary and conclusions of this dissertation and describes the possible future research related to this dissertation.

8.1. Summary and Conclusions

8.1.1. Variable Selection-Based Multivariate SPC in Multistage Processes

In Chapter 3, we introduce a high-dimensional multistage SPC chart for monitoring output variables that follow a beta distribution. The cascade property in multistage processes is removed by using deviance residuals through the model-based approach accommodating the beta regression. The vectorized deviance residuals from all stages are regarded as observations, and the covariance matrix is properly determined to represent the correlations of variables within a stage and the cross correlations among stages. The partial regression-based model selection is developed to improve the capability of the detection of the changes in high-dimensional processes under sparsity. In addition, the capability of the identification of root cause of fault is compared with the existing VS-based charts. Extensive simulation studies from the setting of the LNG process as a representative multistage process are performed to compare the performance of the proposed method with existing methods.

8.1.2. Monitoring of High-Dimensional Processes via Sparse Group LASSO

In Chapter 4, we propose a group variable selection based SPC chart for monitoring high dimensional processes while Chapter 3 incorporates the individual variable selection into the monitoring scheme. It is expected to perform well especially when the process data have grouped structure and when the behavior of the process shift is grouped. Under sparsity of both between groups and within-a-group, we apply the idea of sparse group LASSO and modify it appropriately in SPC point of view by adjusting the parameter settings as a ratio, which helps practitioner's intuitive access and utilization of the procedure based on their engineering knowledge. The proposed method selects the variables flexibly considering the estimated shift size and number of shifted variables. Extensive simulation studies show that the proposed method outperforms existing ones when the shift tends to occur in sparse groups.

8.1.3. Ridge Penalized Likelihood-Based SPC Chart

The monitoring methodologies introduced in Chapters 3 and 4 integrate the variable selection as a prediagosis procedure. They perform generally well in many cases under sparsity. However, when the shift size is small, the ARL_1 performance tends to decrease due to the misidentification of the changed variables based on the variable selection techniques especially when variables are highly correlated. In Chapter 5, we focus on the problem of detecting small mean shifts in high-dimensional processes when the quality measurements are strongly correlated. We propose an efficient quality control method for such processes based on L_2 norm penalized likelihood approach that does not do 'variable

selection' but does 'shrinkage' considering the correlation structure of the data. We could obtain the accurate probability distributions of the monitoring statistic under H_0 and under H_1 with respect to the potential shift directions and provide the theoretical boundary for the performance based on a geometrical interpretation. The numerical results of simulation demonstrate that the proposed method performs better than existing methods in detecting small mean changes.

8.1.4. Generalized Smoothing Parameters of a Multivariate EWMA Control Chart

In Chapter 6, we focus more on detecting small mean shift by utilizing the correlation structure of the data. Rather than shrinking the measurement vector for the mean estimation at each sampling point, we put additional weight to the EWMA transformed vector according to the correlation. It has been known that the MEWMA control chart is effective in detecting a small process mean shift. Its simplicity and generality stem from the assumption that the smoothing parameters of the variables are given constants and equally distributed on the diagonal of the smoothing matrix. In this chapter, we propose a generalized model for the MEWMA that uses appropriate non-diagonal elements in the smoothing matrix based on the correlation among variables. We also offer an interpretation of off-diagonal elements of the smoothing matrix and suggest an optimal design for a proposed MEWMA chart. The proposed chart shows a better ARL performance when compared with existing approaches, especially in detecting smaller shifts. We demonstrate the chart's performnace with various cases of shift scenarios and show that the chart performance is robust to the shift direction. A case study on the automatic monitoring of

dimensions of bolts using an imaging processing system is presented to illustrate the proposed control chart.

8.1.5. Bayesian Sequential Update for Monitoring High-Dimensional Processes

In this chapter, we develop an efficient statistical process monitoring methodology in high-dimensional processes based on the Bayesian approach. The key idea of this chapter is to sequentially update a posterior distribution of the process parameter of interest through the Bayesian rule. A sparsity promoting prior distribution of the parameter is applied properly under sparsity in high-dimensional data, and is sequentially updated in online processing. A Bayesian hierarchical model with a data-driven way of determining the hyperparameters enables the monitoring scheme to be effective to the detection of process shifts with less computational complexity in the high-dimensional processes. Comparisons with recently proposed methods for monitoring high-dimensional processes demonstrates the superiority of the proposed method in detecting small shifts. In addition, graphical presentations in tracking the process parameter provides the information about decisions regarding whether a process needs to be adjusted before it triggers alarm.

8.2. Future Research

Chapters 3, 4 and 7 implements the sparsity in the high-dimensional processes considering that only small number of variables would possibly be shifted simultaneously. Chapters 3 and 4 directly apply the variable selection technique to the quality characteristics that are measured at every sampling point; and Chapter 7 utilizes sparsity promoting prior

distribution to implement the sparsity. One warrant research, in that the adaptive variable selection over sampling points can be modeled with a state space model, is the extension to the cascade process monitoring. A large pool of the literature exploit the conversion of the state space model to the autoregressive and moving average model (ARMA) for monitoring such cascade processes (Triantafyllopoulos and Bersimis 2016; Du *et al.* 2015; Cheng *et al.* 2014; Psarakis and Papaleonida 2007; Pan and Jarrett 2004). Once the accurate estimations for the model parameters are obtained in the retrospective analysis (Phase I analysis), similar procedures incorporating the adaptive variable selection can be immediately applicable to monitor the cascade process.

Another interesting research topic is to monitor the process variability as well as the centrality. Having n samples at each sampling, it can be shown that $\bar{\mathbf{x}} \sim N(\hat{\boldsymbol{\mu}}, n^{-1}\boldsymbol{\Sigma})$ and $\mathbf{S} \sim W(\boldsymbol{\Sigma}, n-1)$, where $W(\boldsymbol{\Sigma}, n-1)$ denotes a Wishart distribution with the covariance matrix, $\boldsymbol{\Sigma}$ and the degrees of freedom, $n-1$. In a similar vein in updating the mean shown in Chapter 7, the posterior distribution of interest is $p\left(\boldsymbol{\Sigma}_t \mid \tilde{\mathbf{S}}_t, \hat{\boldsymbol{\Sigma}}_{t-1}\right)$ and may be derived as proportional to the updated prior and the likelihood. The difficulties are expected to determine a joint likelihood considering $Cov(\bar{\mathbf{x}}_t, \tilde{S}_{t-1})$ may not be zero, and to promote the sparsity in the covariance matrix. The recursive mechanism to update both mean and covariance, and soft/hard thresholding to the quantity $\mathbf{S}_t - \hat{\boldsymbol{\Sigma}}_{t-1} \mid \mathbf{S}_{t-1}$ may be useful.

Unlike the variable selection based methodologies, Chapters 5 and 6 focus on utilizing the correlation structure of the data. The method in Chapter 5 shrinks the measurement vector based on the correlation; and that in Chapter 6 assigns additional weight to the measurement vector according to the correlation. Thus, it would be interesting to investigate the integration of variable selection and utilization of the correlation. For example, the combination of L_1 and L_2 in the likelihood function can be a trivial procedure sharing two different features with VS-based methods and shrinkage-based method according to the correlation. Consequently, it selects suspicious variables like LASSO, and shrinks the correlated covariates together like ridge. It is expected to overcome the drawbacks of the VS-based chart in highly correlated data structure, and is expected to perform well when the shift tends to occur in a grouped fashion due to the nature of shrinkage of L_2 regularization.

In addition, the proposed methodologies developed in this dissertation can be extended to the problem of monitoring non-normally distributed high dimensional processes (Chen *et al.* 2016). Furthermore, the lack of historical dataset is also common in industry. It motivates more research to extend to Phase I analysis, in which the estimation accuracy of covariance would significantly affect the chart performance since our method is highly dependent on the correlation structure.

Appendix A. Proof of Proposition 4.1

i) *Proof of part a.*

The chart statistic Q can be rewritten as a general quadratic form as

$$Q = \mathbf{x}^T(\boldsymbol{\Sigma} + \kappa\boldsymbol{\Sigma}^2)^{-1}\mathbf{x} = \mathbf{x}^T\mathbf{A}^{-1}\mathbf{x},$$

where $\mathbf{A} = \boldsymbol{\Sigma} + \kappa\boldsymbol{\Sigma}^2$. Using an eigen-decomposition of $\boldsymbol{\Sigma}$ as $\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^T$ where $\mathbf{P}$ represents an eigenvector matrix of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ is a diagonal matrix whose diagonal entries are eigenvalues of $\boldsymbol{\Sigma}$, the matrix $\mathbf{A}$ can be expressed as

$$\mathbf{A} = \boldsymbol{\Sigma} + \kappa\boldsymbol{\Sigma}^2 = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^T + \kappa\mathbf{P}\boldsymbol{\Lambda}^2\mathbf{P}^T = \mathbf{P}(\boldsymbol{\Lambda} + \kappa\boldsymbol{\Lambda}^2)\mathbf{P}^T.$$

Then, the chart statistic can be rewritten as

$$Q = \mathbf{x}^T(\boldsymbol{\Sigma} + \kappa\boldsymbol{\Sigma}^2)^{-1}\mathbf{x} = \mathbf{y}^T\left(\boldsymbol{\Lambda} + \kappa\boldsymbol{\Lambda}^2\right)^{-1}\mathbf{y}$$

where $\mathbf{y} = \mathbf{P}^T\mathbf{x}$ and $\mathbf{y} \sim N_p(0, \boldsymbol{\Lambda})$. It can be expanded as

$$Q = \mathbf{y}^T\left(\boldsymbol{\Lambda} + \kappa\boldsymbol{\Lambda}^2\right)^{-1}\mathbf{y} = \sum_{i=1}^{p}\frac{1}{1+\kappa\lambda_i}z_i^2$$

where z_i's are independent standard normal random variables. Thus, the statistic Q forms a weighted sum of independent chi-square variables with weight $(1+\kappa\lambda_i)^{-1}$. When $\kappa = 0$, it reduces to the sum of chi-square random variables resulting in Hotelling T^2 chart since all weights are identical to 1.

By adopting Welch-Satterthwaite (WS) method (Welch, 1938, Satterthwaite, 1946), we can approximates the statistic Q to the gamma distribution. Since the expected value and the variance of Q under H_0, are $E(Q) = E\left(\sum_{i=1}^{p} w_i z_i^2\right) = \sum_{i=1}^{p} w_i = k\theta$, and $Var(Q) = 2\sum_{i=1}^{p} w_i^2 = k\theta^2$, we obtain two parameters as in Proposition 1 a.

ii) *Proof of part b.*

The chart statistic Q with a certain mean shift $\boldsymbol{\mu}_1$ can be expressed as a weighted sum of non-central chi-square random variables as

$$Q = \sum_{i=1}^{p} \frac{1}{1+\kappa\lambda_i} z_i^2(\delta_{\mathbf{z},i})$$

where $\delta_{\mathbf{z},i} = \mu_{z,i}^2$ and the vector $\boldsymbol{\mu}_\mathbf{z}$ is a projection of $\boldsymbol{\mu}_1$ onto the orthogonal space of $\mathbf{z}$ that satisfies $\boldsymbol{\mu}_\mathbf{z} = \boldsymbol{\Lambda}^{-1/2}\mathbf{P}^T\boldsymbol{\mu}_1$. Thus the chart statistic under H_1 with a shift $\boldsymbol{\mu}_1$ is composed of non-central chi-square random variables with noncentrality $\delta_{\mathbf{z},i}$ and weight w_i.

Letting $c_k = \sum_{i=1}^{p} w_i^k + k\sum_{i=1}^{p} w_i^k \delta_{z,i}$, $s_1 = c_3 / c_2^{3/2}$ and $s_2 = c_4 / c_2^2$, the parameters l and δ_χ can be determined by making the skewness of Q and $\chi_l^2(\delta_\chi)$ to be equal while the difference of kurtoses being minimized (Liu *et al.*, 2009a). That is, if $s_1^2 > s_2$, $\delta_\chi = s_1 a^3 - a^2$ and $l = a^2 - 2\delta_\chi$, where $a = 1/\left(s_1 - \sqrt{s_1^2 - s_2}\right)$. ; otherwise, $\delta_\chi = 0$ and $l = 1/s_1^2$. Then, the tale distribution of the quality characteristic Q under H_1 can be approximated as

$$\Pr\left(Q > L \mid \boldsymbol{\mu}_1\right) = \Pr\left(\frac{Q - \mu_Q}{\sigma_Q} > v \mid \boldsymbol{\mu}_1\right)$$

$$\approx \Pr\left(\frac{\chi_l^2(\delta_\chi) - \mu_\chi}{\sigma_\chi} > v\right) = \Pr\left(\chi_l^2(\delta_\chi) > v\sigma_\chi + \mu_\chi\right)$$

Appendix B. Proof of Proposition 4.2

For a given observation $\mathbf{x}$, the chart statistics of Hotelling T^2 and RMSPC are

$$Q_{T^2} = \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x}, \text{ and } Q_{RMSPC} = \mathbf{x}^T (\boldsymbol{\Sigma} + \kappa \boldsymbol{\Sigma}^2)^{-1} \mathbf{x}.$$

Then, the proportional distances to the boundaries with a significance level α for Hotelling T^2 and RMSPC become $Q_{T^2} / H_{T^2,\alpha}$ and $Q_{RMSPC} / H_{RMSPC,\alpha}(\kappa)$, respectively. Therefore, when the relative performance index (RPI) is larger than 1, RMSPC chart performs better than Hotelling T^2 chart.

$$\text{RPI} = \frac{\mathbf{x}^T(\boldsymbol{\Sigma} + \kappa\boldsymbol{\Sigma}^2)^{-1}\mathbf{x} / H_{RMSPC,\alpha}(\kappa)}{\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x} / H_{T^2,\alpha}} \tag{B.1}$$

By the extension of Cauchy-Schwarz inequality, we obtain the following result for Eq. (B.1) as

$$\max_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^T(\boldsymbol{\Sigma} + \kappa\boldsymbol{\Sigma}^2)^{-1}\mathbf{x} / H_{RMSPC,\alpha}(\kappa)}{\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x} / H_{T^2,\alpha}} = \frac{1}{1+\kappa\lambda_{\min}} \cdot \frac{H_{T^2,\alpha}}{H_{RMSPC,\alpha}(\kappa)}$$

when the equality holds when $\mathbf{x}$ is equal to the corresponding eigenvector to the maximum eigenvalue of the matrix $(\mathbf{I}+\kappa\boldsymbol{\Lambda})^{-1}$, i.e., $(1+\kappa\lambda_{\min})^{-1}$, where $\lambda_{\min}$ is a minimum eigenvalue of the correlation matrix $\boldsymbol{\Sigma}$. Therefore, for all $i = 1, 2, ..., p,$ if

$$\frac{1}{1+\kappa\lambda_i} \cdot \frac{H_{T^2,\alpha}}{H_{RMSPC,\alpha}(\kappa)} > 1,$$

the value of monitoring statistics of the proposed chart is relatively closer to the control limit, resulting in larger power than that of T^2 chart.

Appendix C. Derivation of the Minimum ω^*

The suggestion of ω in Eq. (6.4) can be derived by the constraint that the off-diagonals in the smoothing matrix cannot be larger than the diagonal elements. Let us define r_{on} as a

diagonal element in the smoothing matrix, since all diagonals are equal. From the constraint that $r_{on} > r_{ij},\ i \neq j,$ for each column $j,$ we can obtain the following equation:

$$r_{on} = \omega \cdot r > r_{ij} = (r - r_{on}) \cdot \frac{|\rho_{ij}|}{\sum_{i \neq j} |\rho_{ij}|} = (r - \omega \cdot r) \cdot \frac{|\rho_{ij}|}{\sum_{i \neq j} |\rho_{ij}|}$$

for all i's ($i \neq j$). Note that we do not consider ρ_{ii} because it is always 1 and is not considered to determine the off-diagonals. Then the above equation can be written as:

$$\omega > (1 - \omega) \cdot \frac{\max_{1 \leq i \leq p, i \neq j} |\rho_{ij}|}{\sum_{i \neq j} |\rho_{ij}|}.$$

By rearranging each side with respect to $\omega,$ we obtain

$$\omega > \frac{\max_{1 \leq i \leq p, i \neq j} |\rho_{ij}|}{\sum_{i \neq j} |\rho_{ij}| + \max_{1 \leq i \leq p, i \neq j} |\rho_{ij}|}.$$

Therefore, the minimum ω value, $\omega^*,$ is given by

$$\omega^* = \max_{1\le j\le p}\left\{\frac{\max\limits_{1\le i\le p, i\ne j}|\rho_{ij}|}{\sum\limits_{i\ne j}|\rho_{ij}| + \max\limits_{1\le i\le p, i\ne j}|\rho_{ij}|}\right\}.$$

Appendix D. Proof of Eq. (7.1)

The joint distribution $p(\tilde{\mathbf{x}}_t, \tilde{\boldsymbol{\mu}}_t)$ can be recursively obtained as

$$\begin{aligned}
p(\tilde{\mathbf{x}}_t, \tilde{\boldsymbol{\mu}}_t) &= p(\mathbf{x}_t \mid \tilde{\mathbf{x}}_{t-1}, \tilde{\boldsymbol{\mu}}_t)\, p(\tilde{\mathbf{x}}_{t-1} \mid \tilde{\boldsymbol{\mu}}_t)\, p(\tilde{\boldsymbol{\mu}}_t) \\
&= p(\mathbf{x}_t \mid \tilde{\mathbf{x}}_{t-1}, \tilde{\boldsymbol{\mu}}_t)\, p(\tilde{\mathbf{x}}_{t-1} \mid \tilde{\boldsymbol{\mu}}_{t-1})\, p(\tilde{\boldsymbol{\mu}}_t) \\
&= p(\mathbf{x}_t \mid \tilde{\mathbf{x}}_{t-1}, \tilde{\boldsymbol{\mu}}_t)\, \frac{p(\tilde{\mathbf{x}}_{t-1}, \tilde{\boldsymbol{\mu}}_{t-1})}{p(\tilde{\boldsymbol{\mu}}_{t-1})}\, p(\tilde{\boldsymbol{\mu}}_t) \\
&= p(\mathbf{x}_t \mid \tilde{\mathbf{x}}_{t-1}, \tilde{\boldsymbol{\mu}}_t)\, \frac{p(\mathbf{x}_{t-1} \mid \tilde{\mathbf{x}}_{t-2}, \tilde{\boldsymbol{\mu}}_{t-1})\, \dfrac{p(\tilde{\mathbf{x}}_{t-2}, \tilde{\boldsymbol{\mu}}_{t-2})}{p(\tilde{\boldsymbol{\mu}}_{t-2})}\, p(\tilde{\boldsymbol{\mu}}_{t-1})}{p(\tilde{\boldsymbol{\mu}}_{t-1})}\, p(\tilde{\boldsymbol{\mu}}_t) \\
&= p(\mathbf{x}_t \mid \tilde{\mathbf{x}}_{t-1}, \tilde{\boldsymbol{\mu}}_t)\, p(\mathbf{x}_{t-1} \mid \tilde{\mathbf{x}}_{t-2}, \tilde{\boldsymbol{\mu}}_{t-1})\, \frac{p(\tilde{\mathbf{x}}_{t-2}, \tilde{\boldsymbol{\mu}}_{t-2})}{p(\tilde{\boldsymbol{\mu}}_{t-2})}\, p(\tilde{\boldsymbol{\mu}}_t) \\
&= \cdots \\
&= p(\tilde{\boldsymbol{\mu}}_t) \prod_{i=1}^{t} p(\mathbf{x}_i \mid \tilde{\mathbf{x}}_{i-1}, \tilde{\boldsymbol{\mu}}_i)
\end{aligned} \tag{D.1}$$

where the density function $p(\tilde{\boldsymbol{\mu}}_t)$ is

$$\begin{aligned}
p(\tilde{\boldsymbol{\mu}}_t) &= p(\boldsymbol{\mu}_t \mid \tilde{\boldsymbol{\mu}}_t)\, p(\tilde{\boldsymbol{\mu}}_{t-1}) = p(\boldsymbol{\mu}_t \mid \boldsymbol{\mu}_{t-1})\, p(\tilde{\boldsymbol{\mu}}_{t-1}) \\
&= p(\boldsymbol{\mu}_0) \prod_{i=1}^{t} p(\boldsymbol{\mu}_i \mid \boldsymbol{\mu}_{i-1})
\end{aligned}$$

In the second equality holds the Markovian property. By plugging it into Eq. (D.1), the joint probability density, $p(\tilde{\mathbf{x}}_t, \tilde{\boldsymbol{\mu}}_t)$ can be obtained as

$$p(\tilde{\mathbf{x}}_t, \tilde{\boldsymbol{\mu}}_t) = p(\boldsymbol{\mu}_0)\prod_{i=1}^{t} p(\boldsymbol{\mu}_i \mid \boldsymbol{\mu}_{i-1}) p(\mathbf{x}_i \mid \tilde{\mathbf{x}}_{i-1}, \tilde{\boldsymbol{\mu}}_i).$$

The denominator probability, $p(\tilde{\mathbf{x}}_t, \tilde{\boldsymbol{\mu}}_{t-1})$ is written as

$$\begin{aligned} p(\tilde{\mathbf{x}}_t, \tilde{\boldsymbol{\mu}}_{t-1}) &= p(\tilde{\boldsymbol{\mu}}_{t-1} \mid \tilde{\mathbf{x}}_t)\, p(\tilde{\mathbf{x}}_t) = p(\tilde{\boldsymbol{\mu}}_{t-1} \mid \tilde{\mathbf{x}}_{t-1})\, p(\tilde{\mathbf{x}}_t) \\ &= \frac{p(\tilde{\mathbf{x}}_{t-1}, \tilde{\boldsymbol{\mu}}_{t-1})}{p(\tilde{\mathbf{x}}_{t-1})}\, p(\tilde{\mathbf{x}}_t) = p(\tilde{\mathbf{x}}_{t-1}, \tilde{\boldsymbol{\mu}}_{t-1}) \frac{p(\tilde{\mathbf{x}}_t)}{p(\tilde{\mathbf{x}}_{t-1})} \end{aligned}$$

Then, the posterior is written as

$$\begin{aligned} p\left(\boldsymbol{\mu}_t \mid \tilde{\mathbf{x}}_t, \tilde{\boldsymbol{\mu}}_{t-1}\right) &= \frac{p(\tilde{\mathbf{x}}_t, \tilde{\boldsymbol{\mu}}_t)}{p(\tilde{\mathbf{x}}_t, \tilde{\boldsymbol{\mu}}_{t-1})} \\ &= \frac{p(\tilde{\mathbf{x}}_t, \tilde{\boldsymbol{\mu}}_t)}{p(\tilde{\mathbf{x}}_{t-1}, \tilde{\boldsymbol{\mu}}_{t-1})} \frac{p(\tilde{\mathbf{x}}_{t-1})}{p(\tilde{\mathbf{x}}_t)} = \frac{p(\tilde{\mathbf{x}}_{t-1})}{p(\tilde{\mathbf{x}}_t)}\, p(\boldsymbol{\mu}_t \mid \boldsymbol{\mu}_{t-1})\, p(\mathbf{x}_t \mid \tilde{\mathbf{x}}_{t-1}, \tilde{\boldsymbol{\mu}}_t) \qquad \text{(D.2)} \\ &\propto p(\boldsymbol{\mu}_t \mid \boldsymbol{\mu}_{t-1})\, p(\mathbf{x}_t \mid \tilde{\mathbf{x}}_{t-1}, \tilde{\boldsymbol{\mu}}_t) \end{aligned}$$

When the observations, x_i's are all independent and determined only by the current mean, $\boldsymbol{\mu}_i$, the distribution can be simply written as

$$p\left(\boldsymbol{\mu}_t \mid \tilde{\mathbf{x}}_t, \tilde{\boldsymbol{\mu}}_{t-1}\right) = \frac{1}{p(\mathbf{x}_t)} p(\boldsymbol{\mu}_t \mid \boldsymbol{\mu}_{t-1}) p(\mathbf{x}_t \mid \boldsymbol{\mu}_t)$$
$$\propto p(\boldsymbol{\mu}_t \mid \boldsymbol{\mu}_{t-1}) p(\mathbf{x}_t \mid \boldsymbol{\mu}_t),$$

Q.E.D.

Appendix E. Proof of Eq. (7.8)

The conditional distribution, $p(\kappa_{t,i} \mid \hat{\mu}_{t,i}^{(n)})$ can be obtained through the Bayesian rule as

$$p(\kappa_{t,i} \mid \hat{\mu}_{t,i}^{(n)}) = \frac{p(\hat{\mu}_{t,i}^{(n)} \mid \kappa_{t,i}) p(\kappa_{t,i})}{p(\hat{\mu}_{t,i}^{(n)})} \tag{E.1}$$

where $p(\hat{\mu}_{t,i}^{(n)} \mid \kappa_{t,i})$ is Laplacian distribution and $p(\kappa_{t,i})$ is gamma distribution. The probability density, $p(\hat{\mu}_{t,i}^{(n)})$ can be obtained by marginalizing in terms of $\kappa_{t,i}$ as

$$\begin{aligned} p(\hat{\mu}_{t,i}^{(n)}) &= \int_0^\infty p(\hat{\mu}_{t,i}^{(n)} \mid \kappa_{t,i}) p(\kappa_{t,i}) d\kappa_{t,i} = \int_0^\infty \frac{\kappa_{t,i}}{2} e^{-\kappa_{t,i}|\hat{\mu}_{t,i}^{(n)} - \hat{\mu}_{t-1,i}|} \cdot \frac{\beta^\alpha \kappa_{t,i}^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta\kappa_{t,i}} d\kappa_{t,i} \\ &= \frac{1}{2}\int_0^\infty e^{-\kappa_{t,i}(\beta + |\hat{\mu}_{t,i}^{(n)} - \hat{\mu}_{t-1,i}|)} \frac{(\beta + |\hat{\mu}_{t,i}^{(n)} - \hat{\mu}_{t-1,i}|)^\alpha \kappa_{t,i}^{\alpha-1}}{\Gamma(\alpha)} \frac{\beta^\alpha}{(\beta + |\hat{\mu}_{t,i}^{(n)} - \hat{\mu}_{t-1,i}|)^\alpha} \\ &= \frac{1}{2} \frac{\beta^\alpha}{(\beta + |\hat{\mu}_{t,i}^{(n)} - \hat{\mu}_{t-1,i}|)^\alpha} \frac{\alpha}{\beta + |\hat{\mu}_{t,i}^{(n)} - \hat{\mu}_{t-1,i}|} \\ &= \frac{\alpha\beta^\alpha}{2(\beta + |\hat{\mu}_{t,i}^{(n)} - \hat{\mu}_{t-1,i}|)^{\alpha+1}} \end{aligned} \tag{E.2}$$

By substituting Eq. (E.2) into (E.1), we obtain the density function of (E.1) as a gamma distribution as

$$p(\kappa_{t,i} \mid \hat{\mu}_{t,i}^{(n)}) = \frac{(\beta + \mid \hat{\mu}_{t,i}^{(n)} - \hat{\mu}_{t-1,i} \mid)^{\alpha+1} \kappa_{t,i}^{\alpha}}{\Gamma(\alpha+1)} e^{-(\beta+|\hat{\mu}_{t,i}^{(n)} - \hat{\mu}_{t-1,i}|)\kappa_{t,i}}.$$

Therefore, the expected value can be obtained as

$$E\left[\kappa_{t,i}\right]_{p(\kappa_{t,i}|\hat{\mu}_{t,i}^{(n)})}^{(n)} = \frac{\alpha+1}{\beta + \mid \mu_{t,i}^{(n)} - \hat{\mu}_{t-1,i} \mid}.$$

REFERENCES

ABDELLA, G. M., AL-KHALIFA, K. N., KIM, S., JEONG, M. K., ELSAYED, E. A. and HAMOUDA, A. M. 2017. Variable selection-based multivariate cumulative sum control chart. *Quality and Reliability Engineering International,* 33**,** 565-578.

AGRAWAL, R., LAWLESS, J. and MACKAY, R. 1999. Analysis of variation transmission in manufacturing processes--part II. *Journal of Quality Technology,* 31**,** 143.

ALKAHTANI, S. and SCHAFFER, J. 2012. A double multivariate exponentially weighted moving average (dMEWMA) control chart for a process location monitoring. *Communications in Statistics-Simulation and Computation,* 41**,** 238-252.

ALKHAMISI, M. A. and SHUKUR, G. 2007. A Monte Carlo study of recent ridge parameters. *Communications in Statistics—Simulation and Computation®,* 36**,** 535-547.

APLEY, D. W. 2012. Posterior distribution charts: A Bayesian approach for graphically exploring a process mean. *Technometrics,* 54**,** 279-293.

ASADZADEH, S., AGHAIE, A. and YANG, S.-F. 2008. Monitoring and diagnosing multistage processes: a review of cause selecting control charts. *Journal of Industrial and Systems Engineering,* 2**,** 214-235.

BAO, L., WANG, K. and JIN, R. 2014. A hierarchical model for characterising spatial wafer variations. *International Journal of Production Research,* 52**,** 1827-1842.

BERSIMIS, S., PSARAKIS, S. and PANARETOS, J. 2007. Multivariate statistical process control charts: An overview. *Quality and Reliability Engineering International,* 23**,** 517-543.

BROOK, D. and EVANS, D. 1972. An approach to the probability distribution of CUSUM run length. *Biometrika***,** 539-549.

BURGERS, G., JAN VAN LEEUWEN, P. and EVENSEN, G. 1998. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review,* 126**,** 1719-1724.

CAPIZZI, G. 2015. Recent advances in process monitoring: Nonparametric and variable-selection methods for phase I and phase II. *Quality Engineering,* 27**,** 44-67.

CAPIZZI, G. and MASAROTTO, G. 2011. A least angle regression control chart for multidimensional data. *Technometrics,* 53**,** 285-296.

CHARLES, A. S., BALAVOINE, A. and ROZELL, C. J. 2016. Dynamic filtering of time-varying sparse signals via L_1 minimization. *IEEE Transactions on Signal Processing,* 64**,** 5644-5656.

CHEN, N., ZI, X. and ZOU, C. 2016. A distribution-free multivariate control chart. *Technometrics,* 58**,** 448-459.

CRIBARI-NETO, F. and ZEILEIS, A. 2009. Beta regression in R.

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

CROWDER, S. V. 1987. A simple method for studying run–length distributions of exponentially weighted moving average charts. *Technometrics,* 29**,** 401-407.

CROWDER, S. V. 1989. Design of exponentially weighted moving average schemes. *Journal of Quality Technology,* 21**,** 155-162.

CROWDER, S. V. and ESHLEMAN, L. 2001. Small sample properties of an adaptive filter applied to low volume SPC. *Journal of Quality Technology,* 33**,** 29-46.

DING, Y., SHI, J. and CEGLAREK, D. Diagnosability analysis of multi-station manufacturing processes. ASME 2002 International Mechanical Engineering Congress and Exposition, 2002. American Society of Mechanical Engineers, 475-484.

DJURDJANOVIC, D. and NI, J. 2001. Linear state space modeling of dimensional machining errors. *Transactions-North American Manufacturing Research Institution of SME***,** 541-548.

DOAN, X.-T. and SRINIVASAN, R. 2008. Online monitoring of multi-phase batch processes using phase-based multivariate statistical process control. *Computers & Chemical Engineering,* 32**,** 230-243.

DOWNS, J. J. and VOGEL, E. F. 1993. A plant-wide industrial process control problem. *Computers & Chemical Engineering,* 17**,** 245-255.

DURBIN, J. and KOOPMAN, S. J. 2012. *Time series analysis by state space methods*, OUP Oxford.

EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. 2004. Least angle regression. *The Annals of Statistics,* 32**,** 407-499.

FAN, J. and LI, R. 2006. Statistical challenges with high dimensionality: Feature selection in knowledge discovery. *arXiv preprint math/0602133*.

FENNER, J. S., JEONG, M. K. and LU, J.-C. 2005. Optimal automatic control of multistage production processes. *IEEE Transactions on Semiconductor Manufacturing,* 18**,** 94-103.

FERRARI, S. and CRIBARI-NETO, F. 2004. Beta regression for modelling rates and proportions. *Journal of Applied Statistics,* 31**,** 799-815.

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. 2001. *The elements of statistical learning,* Springer Series in Statistics Springer, Berlin.

GAJIC, Z. and QURESHI, M. T. J. 2008. *Lyapunov matrix equation in system stability and control*, Courier Corporation.

GAN, F. 1991. An optimal design of CUSUM quality control charts. *Journal of Quality Technology,* 23**,** 279-286.

GARRIGUES, P. and OLSHAUSEN, B. A. Group sparse coding with a laplacian scale mixture prior. Advances in Neural Information Processing Systems, 2010. 676-684.

HAN, D. and TSUNG, F. 2006. A reference-free cuscore chart for dynamic mean change detection and a unified framework for charting performance comparison. *Journal of the American Statistical Association,* 101**,** 368-386.

HAWKINS, D. and OLWELL, D. 1998. *Statistics for engineering and physical science*, New York: Springer.

HAWKINS, D. M. 1991. Multivariate quality control based on regression-adiusted 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., CHOI, S. and LEE, S. 2007. A general multivariate exponentially weighted moving-average control chart. *Journal of Quality Technology,* 39**,** 118.

HUANG, Q., ZHOU, S. and SHI, J. 2002. Diagnosis of multi-operational machining processes through variation propagation analysis. *Robotics and Computer-Integrated Manufacturing,* 18**,** 233-239.

HUH, I., VIVEROS-AGUILERA, R. and BALAKRISHNAN, N. 2013. Differential smoothing in the bivariate exponentially weighted moving average chart. *Journal of Quality Technology,* 45**,** 377.

HWANG, W., AL-KHALIFA, K., HAMOUDA, A. M., JEONG, M. K. and ELSAYED, E. A. Multistage statistical process control for parametric variables. International Conference on Quality, Risk, Maintenance, and Safety Engineering, 2014.

JAIN, K., ALT, F. and GRIMSHAW, S. Multivariate quality control-a Bayesian approach. Annual Quality Congress Transactions-American Society for Quality Control, 1993. American Society for Quality Control, 645-645.

JEARKPAPORN, D., BORROR, C., RUNGER, G. and MONTGOMERY, D. 2007. Process monitoring for mean shifts for multiple stage processes. *International Journal of Production Research,* 45**,** 5547-5570.

JEARKPAPORN, D., MONTGOMERY, D., RUNGER, G. and BORROR, C. 2005. Model-based process monitoring using robust generalized linear models. *International Journal of Production Research,* 43**,** 1337-1354.

JEARKPAPORN, D., MONTGOMERY, D. C., RUNGER, G. C. and BORROR, C. M. 2003. Process monitoring for correlated gamma-distributed data using generalized-linear-model-based control charts. *Quality and Reliability Engineering International,* 19**,** 477-491.

JEFFREYS, H. 1998. *The theory of probability*, OUP Oxford.

JENSEN, W. A., JONES-FARMER, L. A., CHAMP, C. W. and WOODALL, W. H. 2006. Effects of parameter estimation on control chart properties: A literature review. *Journal of Quality Technology,* 38**,** 349.

JEONG, M. K., LU, J.-C. and WANG, N. 2006. Wavelet-based SPC procedure for complicated functional data. *International Journal of Production Research,* 44**,** 729-744.

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

JIANG, W., WANG, K. and TSUNG, F. 2012. A variable-selection-based multivariate EWMA chart for process monitoring and diagnosis. *Journal of Quality Technology,* 44**,** 209.

JULIER, S. J. and UHLMANN, J. K. New extension of the Kalman filter to nonlinear systems. Signal Processing, Sensor Fusion, and Target Recognition VI, 1997. International Society for Optics and Photonics, 182-194.

KASS, R. E. and RAFTERY, A. E. 1995. Bayes factors. *Journal of the American Statistical Association,* 90**,** 773-795.

KIM, J., AL-KHALIFA, K., JEONG, M., HAMOUDA, A. and ELSAYED, E. 2014. Multivariate statistical process control charts based on the approximate sequential χ^2 test. *International Journal of Production Research,* 52**,** 5514-5527.

KIM, J., AL-KHALIFA, K., PARK, M., JEONG, M., HAMOUDA, A. and ELSAYED, E. 2013. Adaptive cumulative sum charts with the adaptive runs rule. *International Journal of Production Research,* 51**,** 4556-4569.

KIM, J., JEONG, M. K. and ELSAYED, E. A. 2017a. Process monitoring in multistage processes with autocorrelated observations. *International Journal of Production Research,* 55**,** 2385-2396.

KIM, J., JEONG, M. K., ELSAYED, E. A., AL-KHALIFA, K. and HAMOUDA, A. 2016. An adaptive step-down procedure for fault variable identification. *International Journal of Production Research,* 54**,** 3187-3200.

KIM, S., JEONG, M. K. and ELSAYED, E. A. 2017b. Generalized smoothing parameters of a multivariate EWMA control chart. *IISE Transactions,* 49**,** 58-69.

KIM, S., KIM, J., JEONG, M. K., AL-KHALIFA, K., HAMOUDA, A. M. and ELSAYED, E. A. 2017c. Monitoring and control of beta-distributed multistage production processes. *Quality Technology & Quantitative Management,* Accepted**,** 1-18.

LAUNGRUNGRONG, B., BORROR, C. M. and MONTGOMERY, D. C. 2011. EWMA control charts for multivariate Poisson-distributed data. *International Journal of Quality Engineering and Technology,* 2**,** 185-211.

LAUNGRUNGRONG, B., BORROR, C. M. and MONTGOMERY, D. C. 2014. A one-sided MEWMA control chart for Poisson-distributed data. *International Journal of Data Analysis Techniques and Strategies,* 6**,** 15-42.

LAWLESS, J., MACKAY, R. and ROBINSON, J. 1999. Analysis of variation transmission in manufacturing processes--Part I. *Journal of Quality Technology,* 31**,** 131.

LEE, M. 2010. Multivariate EWMA control chart with adaptive sample sizes. *Communications in Statistics-Simulation and Computation,* 39**,** 1548-1561.

LI, J., JIN, J. and SHI, J. 2008. Causation-based T^2 decomposition for multivariate process monitoring and diagnosis. *Journal of Quality Technology,* 40**,** 46.

LIU, H., TANG, Y. and ZHANG, H. H. 2009a. A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. *Computational Statistics & Data Analysis,* 53**,** 853-856.

LIU, J., JI, S. and YE, J. 2009b. SLEP: Sparse learning with efficient projections. Arizona State University.

LIU, K., MEI, Y. and SHI, J. 2015. An adaptive sampling strategy for online high-dimensional process monitoring. *Technometrics,* 57**,** 305-319.

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**,** 1-12.

MACGREGOR, J. F. and KOURTI, T. 1995. Statistical process control of multivariate processes. *Control Engineering Practice,* 3**,** 403-414.

MAHMOUD, M. A. and MARAVELAKIS, P. E. 2010. The performance of the MEWMA control chart when parameters are estimated. *Communications in Statistics—Simulation and Computation®,* 39**,** 1803-1817.

MAHMOUD, M. A. and WOODALL, W. H. 2010. An evaluation of the double exponentially weighted moving average control chart. *Communications in Statistics—Simulation and Computation®,* 39**,** 933-949.

MAKIS, V. 2008. Multivariate Bayesian control chart. *Operations Research,* 56**,** 487-496.

MARDIA, K. V. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika,* 519-530.

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.

MEHMOOD, T., LILAND, K. H., SNIPEN, L. and SÆBØ, S. 2012. A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems,* 118**,** 62-69.

MONTGOMERY, D. C. 2007. *Introduction to statistical quality control,* John Wiley & Sons.

NGAI, H.-M. and ZHANG, J. 2001. Multivariate cumulative sum control charts based on projection pursuit. *Statistica Sinica,* 747-766.

NIKOLAIDIS, Y. and TAGARAS, G. 2017. New indices for the evaluation of the statistical properties of Bayesian x-bar control charts for short runs. *European Journal of Operational Research,* 259**,** 280-292.

NOMIKOS, P. and MACGREGOR, J. F. 1994. Monitoring batch processes using multiway principal component analysis. *AIChE Journal,* 40**,** 1361-1375.

NOMIKOS, P. and MACGREGOR, J. F. 1995. Multivariate SPC charts for monitoring batch processes. *Technometrics,* 37**,** 41-59.

PAGE, E. 1954. Continuous inspection schemes. *Biometrika,* 41**,** 100-115.

PAGE, E. 1961. Cumulative sum charts. *Technometrics,* 3**,** 1-9.

PARK, M., KIM, J., JEONG, M., HAMOUDA, A., AL-KHALIFA, K. and ELSAYED, E. 2012. Economic cost models of integrated APC controlled SPC charts. *International Journal of Production Research,* 50**,** 3936-3955.

PERES, F. A. P. and FOGLIATTO, F. S. 2018. Variable selection methods in multivariate statistical process control: A systematic literature review. *Computers & Industrial Engineering,* 115**,** 603-619.

PIGNATIELLO, J. J. and RUNGER, G. C. 1990. Comparisons of multivariate CUSUM charts. *Journal of Quality Technology,* 22**,** 173-186.

PRABHU, S. S. and RUNGER, G. C. 1997. Designing a multivariate EWMA control chart. *Journal of Quality Technology,* 29**,** 8.

PRATER, N. 1956. Estimate gasoline yields from crudes. *Petroleum Refiner,* 35**,** 236-238.

QIU, P. 2005. *Image processing and jump regression analysis,* John Wiley & Sons.

RADKE, R. J., ANDRA, S., AL-KOFAHI, O. and ROYSAM, B. 2005. Image change detection algorithms: a systematic survey. *IEEE Transactions on Image Processing,* 14**,** 294-307.

RAICH, A. and CINAR, A. 1996. Statistical process monitoring and disturbance diagnosis in multivariable continuous processes. *AIChE Journal,* 42**,** 995-1009.

RAJARAM, K. and ROBOTIS, A. 2004. Analyzing variability in continuous processes. *European Journal of Operational Research,* 156**,** 312-325.

RAO, S., STROJWAS, A. J., LEHOCZKY, J. P. and SCHERVISH, M. J. 1996. Monitoring multistage integrated circuit fabrication processes. *IEEE Transactions on Semiconductor Manufacturing,* 9**,** 495-505.

ROBERTS, S. 1959. Control chart tests based on geometric moving averages. *Technometrics,* 1**,** 239-250.

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

SATTERTHWAITE, F. E. 1946. An approximate distribution of estimates of variance components. *Biometrics Bulletin,* 2**,** 110-114.

SCHMIDT, K. D. 2002. A note on the overdispersed Poisson family. *Insurance: Mathematics and Economics,* 30**,** 21-25.

SHAMMA, S. E., AMIN, R. W. and SHAMMA, A. K. 1991. A double exponentially weigiited moving average control procedure with variable sampling intervals. *Communications in Statistics-Simulation and Computation,* 20**,** 511-528.

SHAMMA, S. E. and SHAMMA, A. K. 1992. Development and evaluation of control charts using double exponentially weighted moving averages. *International Journal of Quality & Reliability Management,* 9.

SHEWHART, W. A. 1931. *Economic control of quality of manufactured product,* New York, D. Van Nostrand Company, Inc.

SHI, J. 1999. State space modeling of sheet metal assembly for dimensional control. *Ann Arbor,* 1050**,** 48109-2117.

SHI, J. and ZHOU, S. 2009. Quality control and improvement for multistage systems: A survey. *IIE Transactions,* 41**,** 744-753.

SHU, L., TSUNG, F. and KAPUR, K. C. 2004a. Design of multiple cause-selecting charts for multistage processes with model uncertainty. *Quality Engineering,* 16**,** 437-450.

SHU, L., TSUNG, F. and KWOK-LEUNG, T. 2004b. Run-length performance of regression control charts with estimated parameters. *Journal of Quality Technology,* 36**,** 280.

SIMON, N., FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. 2013. A sparse-group lasso. *Journal of Computational and Graphical Statistics,* 22**,** 231-245.

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.

TAGHVAEI, A., DE WILJES, J., MEHTA, P. G. and REICH, S. 2018. Kalman filter and its modern extensions for the continuous-time nonlinear filtering problem. *Journal of Dynamic Systems, Measurement, and Control,* 140**,** 030904.

TIBSHIRANI, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)***,** 267-288.

TIBSHIRANI, R., SAUNDERS, M., ROSSET, S., ZHU, J. and KNIGHT, K. 2005. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 67**,** 91-108.

TRIANTAFYLLOPOULOS, K. 2007. Feedback quality adjustment with Bayesian state-space models. *Applied Stochastic Models in Business and Industry,* 23**,** 145-156.

TSIAMYRTZIS, P. and HAWKINS, D. M. 2005. A Bayesian scheme to detect changes in the mean of a short-run process. *Technometrics,* 47**,** 446-456.

VEERAVALLI, V. V. and BANERJEE, T. 2014. Quickest change detection. *Academic Press Library in Signal Processing.* Elsevier.

WADE, M. and WOODALL, W. 1993. A review and analysis of cause-selecting control charts. *Journal of Quality Technology,* 25**,** 161-169.

WANG, H. and LENG, C. 2007. Unified LASSO estimation by least squares approximation. *Journal of the American Statistical Association,* 102**,** 1039-1048.

WANG, K. and JIANG, W. 2009. High-dimensional process monitoring and fault isolation via variable selection. *Journal of Quality Technology,* 41**,** 247.

WANG, K., JIANG, W. and LI, B. 2016. A spatial variable selection method for monitoring product surface. *International Journal of Production Research,* 54**,** 4161-4181.

WANG, K. and TSUNG, F. 2008. An adaptive T^2 chart for monitoring dynamic systems. *Journal of Quality Technology,* 40**,** 109.

WEESE, M., MARTINEZ, W., MEGAHED, F. M. and JONES-FARMER, L. A. 2016. Statistical learning methods applied to process monitoring: An overview and perspective. *Journal of Quality Technology,* 48**,** 4-24.

WELCH, B. L. 1938. The significance of the difference between two means when the population variances are unequal. *Biometrika,* 29**,** 350-362.

WILCOX, P., HORTON, T. M., YOUN, E., JEONG, M. K., TATE, D., HERRMAN, T. and NANSEN, C. 2014. Evolutionary refinement approaches for band selection of hyperspectral images with applications to automatic monitoring of animal feed quality. *Intelligent Data Analysis,* 18**,** 25-42.

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

WOODALL, W. H. and MONTGOMERY, D. C. 2014. Some current directions in the theory and application of statistical process monitoring. *Journal of Quality Technology,* 46**,** 78-94.

WOODWARD, P. W. and NAYLOR, J. C. 1993. An application to Bayesian methods in SPC. *The Statistician***,** 461-469.

XIANG, L. and TSUNG, F. 2008. Statistical monitoring of multi-stage processes based on engineering models. *IIE Transactions,* 40**,** 957-970.

XIONG, J. 2008. *An introduction to stochastic filtering theory*, Oxford University Press on Demand.

YIN, S., DING, S. X., HAGHANI, A., HAO, H. and ZHANG, P. 2012. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. *Journal of Process Control,* 22**,** 1567-1581.

YUAN, M. and LIN, Y. 2006. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 68**,** 49-67.

YUMIN, L. 1996. An improvement for MEWMA in multivariate process control. *Computers & industrial engineering,* 31**,** 779-781.

ZANTEK, P. F., WRIGHT, G. P. and PLANTE, R. D. 2006. A self-starting procedure for monitoring process quality in multistage manufacturing systems. *IIE Transactions,* 38**,** 293-308.

ZHANG, G. A new type of control charts and a theory of diagnosis with control charts. World Quality Congress Transactions, 1984. American Society for Quality Control Milwaukee, WI, 175-185.

ZHANG, G. 1985. Cause-selecting control charts–a new type of quality control charts. *The QR Journal,* 12**,** 221-225.

ZHANG, G. 1990. A new diagnosis theory with two kinds of quality. *Total Quality Management,* 1**,** 249-258.

ZHANG, G. 1992. *Cause-selecting control chart and diagnosis: Theory and practise,* Aarhus School of Business.

ZHANG, L. and CHEN, G. 2005. An extended EWMA mean chart. *Quality Technology & Quantitative Management,* 2**,** 39-52.

ZHANG, L., GOVINDARAJU, K., LAI, C. and BEBBINGTON, M. 2003. Poisson DEWMA control chart. *Communications in Statistics-Simulation and Computation,* 32**,** 1265-1283.

ZHOU, S., HUANG, Q. and SHI, J. 2003. State space modeling of dimensional variation propagation in multistage machining process using differential motion vectors. *IEEE Transactions on Robotics and Automation,* 19**,** 296-309.

ZHU, Y. and JIANG, W. 2009. An adaptive T^2 chart for multivariate process monitoring and diagnosis. *IIE Transactions,* 41**,** 1007-1018.

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.

ZOU, C. and TSUNG, F. 2008. Directional MEWMA schemes for multistage process monitoring and diagnosis. *Journal of Quality Technology,* 40**,** 407.

ZOU, H. 2006. The adaptive LASSO and its oracle properties. *Journal of the American Statistical Association,* 101**,** 1418-1429.